



ΣΤΡΑΤΙΩΤΙΚΗ ΣΧΟΛΗ ΕΥΕΛΠΙΔΩΝ  
Τμήμα Στρατιωτικών Επιστημών

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ  
ΔΙΔΡΥΜΑΤΙΚΟ ΔΙΑΤΜΗΜΑΤΙΚΟ  
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ  
ΑΚΑΔΗΜΑΪΚΟΥ ΕΤΟΥΣ 2024-25

ΕΠΙΧΕΙΡΗΣΙΑΚΗ ΕΡΕΥΝΑ ΚΑΙ  
ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ –  
MASTER OF SCIENCE IN OPERATIONAL  
RESEARCH AND DECISION MAKING

(ΠΔ 59 /2021/ΦΕΚ 145Α'/17.08.2021)



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ  
Σχολή Μηχανικών Παραγωγής & Διοίκησης

# ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

## Αλγόριθμοι μηχανικής μάθησης για την ανάλυση των μέσων κοινωνικής δικτύωσης

Διατριβή που υπεβλήθη για την μερική ικανοποίηση των απαιτήσεων  
για την απόκτηση Μεταπτυχιακού Διπλώματος Ειδίκευσης

Υπό:

ΓΕΩΡΓΙΟΣ ΒΑΣΙΛΕΙΑΔΗΣ

A.M.: 2024018111

ΟΚΤΩΒΡΙΟΣ 2025



Η Μεταπτυχιακή Διατριβή του Γεωργίου Βασιλειάδη εγκρίνεται:

### **ΤΡΙΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ**

Καθηγητής ΟΝΟΜΑΤΕΠΩΝΥΜΟ (Επιβλέπων)

Νικόλαος Παπαδάκης

*Νικόλαος Παπαδάκης*

Καθηγητής ΟΝΟΜΑΤΕΠΩΝΥΜΟ

Νικόλαος Δάρας



Καθηγητής ΟΝΟΜΑΤΕΠΩΝΥΜΟ

Νικόλαος Ματσατσίνης

ΣΕΛΙΔΑ ΣΚΟΠΙΜΑ ΚΕΝΗ

© Copyright υπό Γεώργιο Βασιλειάδη

10/10/2025

## Αφιερώσεις

Η παρούσα διπλωματική εργασία αφιερώνεται με την προσδοκία να αποτελέσει μια μικρή, αλλά ουσιαστική συμβολή στην έρευνα και στην επιστήμη, ενισχύοντας τη γνώση και ανοίγοντας τον δρόμο για περαιτέρω μελέτη και πρόοδο στο πεδίο της Μηχανικής Μάθησης και της ανάλυσης κοινωνικών δεδομένων.

ΣΕΛΙΔΑ ΣΚΟΠΙΜΑ ΚΕΝΗ

## ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω τις ειλικρινείς μου ευχαριστίες στα μέλη της τριμελούς εξεταστικής επιτροπής, τον Καθηγητή Νικόλαο Παπαδάκη, τον Καθηγητή Νικόλαο Δάρα και τον Καθηγητή Νικόλαο Ματσατσίνη, για τη συμβολή και την καθοριστική βοήθειά τους κατά τη διάρκεια των μεταπτυχιακών μου σπουδών. Η καθοδήγησή τους, οι πολύτιμες παρατηρήσεις τους και η αδιάκοπη στήριξή τους συνέβαλαν ουσιαστικά στην ολοκλήρωση της παρούσας εργασίας και στην επιστημονική μου εξέλιξη.

ΣΕΛΙΔΑ ΣΚΟΠΙΜΑ ΚΕΝΗ

## Περιεχόμενα

<b>ΠΕΡΙΛΗΨΗ.....</b>	<b>XII</b>
<b>ABSTRACT.....</b>	<b>XIII</b>
<b>ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ.....</b>	<b>XIV</b>
1.1 ΣΚΟΠΟΣ ΚΑΙ ΑΝΤΙΚΕΙΜΕΝΟ ΤΗΣ ΈΡΕΥΝΑΣ .....	XIV
1.2 ΕΡΕΥΝΗΤΙΚΑ ΕΡΩΤΗΜΑΤΑ ΚΑΙ ΣΤΟΧΟΙ .....	XVI
1.3 ΜΕΘΟΔΟΛΟΓΙΑ ΚΑΙ ΔΟΜΗ ΤΗΣ ΕΡΓΑΣΙΑΣ.....	XVIII
1.4 ΠΕΡΙΟΡΙΣΜΟΙ ΚΑΙ ΟΡΙΟΘΕΤΗΣΗ ΤΗΣ ΜΕΛΕΤΗΣ .....	XX
<b>ΚΕΦΑΛΑΙΟ 2: ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ.....</b>	<b>XXII</b>
2.1 ΟΡΙΣΜΟΣ ΚΑΙ ΙΣΤΟΡΙΚΗ ΕΞΕΛΙΞΗ ΤΩΝ ΜΕΣΩΝ ΚΟΙΝΩΝΙΚΗΣ ΔΙΚΤΥΩΣΗΣ.....	XXII
2.2 Η ΑΝΑΛΥΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΣΤΑ ΜΕΣΑ ΚΟΙΝΩΝΙΚΗΣ ΔΙΚΤΥΩΣΗΣ (SOCIAL MEDIA ANALYTICS).....	XXIV
2.3 Η ΈΝΝΟΙΑ ΤΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ (MACHINE LEARNING) .....	XXVII
2.4 ΔΙΑΦΟΡΑ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ – ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ – ΣΤΑΤΙΣΤΙΚΗΣ .....	XXXI
2.5 ΕΦΑΡΜΟΓΕΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΣΤΟΝ ΠΡΑΓΜΑΤΙΚΟ ΚΟΣΜΟ.....	XXXIV
<b>ΚΕΦΑΛΑΙΟ 3: ΚΑΤΗΓΟΡΙΕΣ ΑΛΓΟΡΙΘΜΩΝ.....</b>	<b>XXXVIII</b>
3.1 ΕΠΙΒΛΕΠΟΜΕΝΗ ΜΑΘΗΣΗ (SUPERVISED LEARNING).....	XXXVIII
3.2 ΜΗ ΕΠΙΒΛΕΠΟΜΕΝΗ ΜΑΘΗΣΗ (UNSUPERVISED LEARNING) .....	XLII
3.3 ΕΝΙΣΧΥΤΙΚΗ ΜΑΘΗΣΗ (REINFORCEMENT LEARNING).....	XLV
<b>ΚΕΦΑΛΑΙΟ 4: ΒΑΣΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ.....</b>	<b>XLVIII</b>
4.1 ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ ΚΑΙ ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ.....	XLVIII
4.1.1 Απλή Γραμμική Παλινδρόμηση.....	xlvi
4.1.3 Απλό Λογιστικό Μοντέλο .....	lxxi
4.1.4 Απλή Λογιστική Παλινδρόμηση .....	lxxiv
4.1.5 Ελεγχουσυναρτήσεις.....	lxxvii
4.1.6 Υπόλοιπα .....	lxxxii
4.1.7 Κριτήρια επιλογής μοντέλου λογιστικής παλινδρόμησης .....	lxxxvi
4.1.8 Καμπύλη ROC.....	lxxxviii
4.1.9 Πολλαπλή Λογιστική Παλινδρόμηση.....	xc
4.2 ΑΠΟΦΑΣΗ ΔΕΝΤΡΩΝ ΚΑΙ ΤΥΧΑΙΑ ΔΑΣΗ (DECISION TREES, RANDOM FORESTS) .....	XCIII
4.2.1 Σύγκριση με Δέντρα Απόφασης και Άλλους Αλγορίθμους.....	xcix
4.2.2 Πλεονεκτήματα και Μειονεκτήματα του Random Forest.....	ciii
4.2.3 Παραδείγματα Χρήσης του Random Forest.....	cv
4.2.4 Παράδειγμα Υλοποίησης σε Python (scikit-learn).....	cix
4.3 ΥΠΟΣΤΗΡΙΞΗ ΔΙΑΝΥΣΜΑΤΙΚΩΝ ΜΗΧΑΝΩΝ (SVM) .....	CXII

4.3.1 Μαθηματική Θεμελίωση των SVM .....	cxiii
4.3.2 Πλεονεκτήματα και Μειονεκτήματα των SVM.....	cxix
4.3.3 Case study: Εφαρμογή των SVM σε πραγματικά δεδομένα Social Media.....	cxv
4.4 K-NEAREST NEIGHBORS (K-NN) .....	CXXXVII
4.4.1 Αρχιτεκτονική και Βασική Λογική του k-NN.....	cxvii
4.4.2 Κανονικοποίηση Δεδομένων (Feature Scaling).....	cxlii
4.4.3 Εφαρμογές του k-NN σε Δεδομένα Social Media.....	cxlvii
4.5 CLUSTERING (K-MEANS - DBSCAN) .....	CLII
4.5.1 Αλγόριθμος K-Means .....	clii
4.5.2 Αλγόριθμος DBSCAN.....	clx
4.6 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΒΑΘΙΑ ΜΑΘΗΣΗ (DEEP LEARNING) .....	CLXXVII
4.6.1 Τεχνητά Νευρωνικά Δίκτυα.....	clxxvii
4.6.2 Κυριότερες κατηγορίες ANN.....	clxxx
4.6.3 Ιστορικά στοιχεία για τα ΤΝΔ.....	clxxvii
4.6.4 Εφαρμογές των ΤΝΔ.....	clxxviii
4.6.5 Παραδείγματα επίλυσης προβλημάτων με ΤΝΔ.....	clxxxix
4.6.6 Βαθιά Μάθηση και Νευρωνικά Δίκτυα Βαθιάς Μάθησης.....	cxv
4.6.7 Συμπεράσματα.....	ccv

## ΚΕΦΑΛΑΙΟ 5: ΜΕΣΑ ΚΟΙΝΩΝΙΚΗΣ ΔΙΚΤΥΩΣΗΣ ΚΑΙ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝCCVII

5.1 ΔΟΜΗ ΚΑΙ ΙΔΙΑΙΤΕΡΟΤΗΤΕΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΤΩΝ ΚΟΙΝΩΝΙΚΩΝ ΔΙΚΤΥΩΝ.....	CCVII
5.2 ΠΗΓΕΣ ΚΑΙ ΤΥΠΟΙ ΔΕΔΟΜΕΝΩΝ (ΚΕΙΜΕΝΟ, ΕΙΚΟΝΑ, ΒΙΝΤΕΟ, HASHTAGS, ΧΡΟΝΙΚΑ ΔΕΔΟΜΕΝΑ) .....	CCX
5.3 ΠΡΟ-ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΚΟΙΝΩΝΙΚΑ ΔΙΚΤΥΑ (DATA CLEANING & FEATURE ENGINEERING) .....	CCXII
5.4 ΑΝΑΛΥΣΗ ΣΥΝΑΙΣΘΗΜΑΤΟΣ (SENTIMENT ANALYSIS) .....	CCXV
5.5 ΑΝΙΧΝΕΥΣΗ ΘΕΜΑΤΩΝ ΚΑΙ ΤΑΣΕΩΝ (TOPIC MODELING & TREND DETECTION).....	CCXVII
5.6 ΑΝΑΛΥΣΗ ΔΙΚΤΥΩΝ ΚΑΙ ΚΟΙΝΟΤΗΤΩΝ (NETWORK ANALYSIS & COMMUNITY DETECTION) .....	CCXIX
5.7 ΠΡΟΒΛΕΨΕΙΣ ΚΑΙ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΕΙΣ (CLASSIFICATION & PREDICTION).....	CCXXI

## ΚΕΦΑΛΑΙΟ 6: ΕΦΑΡΜΟΓΕΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΣΤΑ ΚΟΙΝΩΝΙΚΑ ΔΙΚΤΥΑCCXXIV

6.1 ΑΝΙΧΝΕΥΣΗ FAKE NEWS ΚΑΙ ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗΣ.....	CCXXIV
6.2 ΣΤΟΧΕΥΜΕΝΗ ΔΙΑΦΗΜΙΣΗ ΚΑΙ ΠΡΟΒΛΕΨΕΙΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΧΡΗΣΤΩΝ .....	CCXXVII
6.3 ΑΝΙΧΝΕΥΣΗ ΑΣΥΝΗΘΙΣΤΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΚΑΙ SPAM.....	CCXXIX
6.4 ΑΝΑΛΥΣΗ ΕΠΙΡΡΟΗΣ (INFLUENCE ANALYSIS) .....	CCXXXII
6.5 ΑΝΑΛΥΣΗ ΚΑΜΠΑΝΙΩΝ ΚΑΙ SOCIAL MEDIA MARKETING.....	CCXXXIV
6.6 DEEP LEARNING ΚΑΙ ΑΝΑΛΥΣΗ ΠΟΛΥΜΕΣΙΚΩΝ ΠΕΡΙΕΧΟΜΕΝΩΝ (ΕΙΚΟΝΕΣ, ΒΙΝΤΕΟ).....	CCXXXVII

## ΚΕΦΑΛΑΙΟ 7: ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΚΑΙ ΠΡΟΚΛΗΣΕΙΣ..... CCXL

7.1 ΟΦΕΛΗ ΤΗΣ ΧΡΗΣΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΣΤΑ ΜΕΣΑ ΚΟΙΝΩΝΙΚΗΣ ΔΙΚΤΥΩΣΗΣ .....	CCXL
7.2 ΤΕΧΝΙΚΕΣ ΚΑΙ ΘΕΩΡΗΤΙΚΕΣ ΠΡΟΚΛΗΣΕΙΣ .....	CCXLII

7.3 ΖΗΤΗΜΑΤΑ ΙΔΙΩΤΙΚΟΤΗΤΑΣ ΚΑΙ ΗΘΙΚΗΣ .....	CCXLIV
7.4 ΕΡΜΗΝΕΥΣΙΜΟΤΗΤΑ ΚΑΙ ΔΙΑΦΑΝΕΙΑ ΤΩΝ ΜΟΝΤΕΛΩΝ .....	CCXLVI
7.5 ΔΕΔΟΜΕΝΑ ΕΚΤΟΣ ΠΛΑΙΣΙΟΥ ΚΑΙ ΣΦΑΛΜΑΤΑ ΠΡΟΒΛΕΨΗΣ .....	CCXLVIII
<b>ΚΕΦΑΛΑΙΟ 8: ΑΝΟΙΧΤΑ ΖΗΤΗΜΑΤΑ ΚΑΙ ΚΑΤΕΥΘΥΝΣΕΙΣ ΓΙΑ ΜΕΛΛΟΝΤΙΚΗ ΈΡΕΥΝΑ</b>	
<b>CCL</b>	
8.1 ΑΛΓΟΡΙΘΜΟΙ ΝΕΑΣ ΓΕΝΙΑΣ (Π.Χ. TRANSFORMERS, BERT, LLMs) .....	CCL
8.2 ΣΥΝΔΥΑΣΜΟΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΜΕ ΜΕΤΑ-ΑΝΑΛΥΣΗ ΚΑΙ BIG DATA .....	CCLXVI
8.3 ΧΡΗΣΗ ΤΩΝ SOCIAL MEDIA ΩΣ ΕΡΓΑΛΕΙΑ ΚΟΙΝΩΝΙΚΗΣ ΚΑΙ ΠΟΛΙΤΙΚΗΣ ΈΡΕΥΝΑΣ .....	CCLXXI
8.4 ΕΠΙΚΑΙΡΟΠΟΙΗΣΗ ΜΟΝΤΕΛΩΝ ΣΕ ΔΥΝΑΜΙΚΑ ΠΕΡΙΒΑΛΛΟΝΤΑ .....	CCLXXIX
8.5 ΕΝΣΩΜΑΤΩΣΗ ΜΕ ΤΕΧΝΟΛΟΓΙΕΣ METAVERSE ΚΑΙ AUGMENTED SOCIAL MEDIA .....	CCLXXXIV
<b>ΚΕΦΑΛΑΙΟ 9: ΣΥΜΠΕΡΑΣΜΑΤΑ.....CCLXXXVII</b>	
9.1 ΣΥΝΟΠΤΙΚΗ ΕΠΙΣΚΟΠΗΣΗ .....	CCLXXXVII
9.2 ΑΠΑΝΤΗΣΕΙΣ ΣΤΑ ΕΡΕΥΝΗΤΙΚΑ ΕΡΩΤΗΜΑΤΑ .....	CCXCIV
9.2.1 Ερώτημα 1: Κύριοι Αλγόριθμοι Μηχανικής Μάθησης και Χαρακτηριστικά .....	ccxciv
9.2.2 Ερώτημα 2: Κυριότερες Εφαρμογές των Αλγορίθμων στα Κοινωνικά Δίκτυα και Επιλυόμενα Προβλήματα.....	ccxcvi
9.2.3 Ερώτημα 3: Τεχνικά, Ηθικά και Κοινωνικά Ζητήματα από την Εφαρμογή Μηχανικής Μάθησης.....	ccxcviii
9.2.4 Ερώτημα 4: Μελλοντικές Τάσεις και Ανοιχτά Ερευνητικά Ζητήματα στην Ανάλυση Κοινωνικών Μέσων.....	ccc
9.3 ΣΥΜΠΕΡΑΣΜΑΤΙΚΑ ΣΧΟΛΙΑ .....	CCCII
<b>ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ.....CCCIV</b>	

## Περίληψη

Η ραγδαία εξάπλωση των μέσων κοινωνικής δικτύωσης (Social Media) τα τελευταία χρόνια έχει δημιουργήσει τεράστιες ποσότητες δεδομένων, οι οποίες μπορούν να αξιοποιηθούν για την εξαγωγή γνώσης και την κατανόηση της ανθρώπινης συμπεριφοράς, των τάσεων και της κοινωνικής δυναμικής. Η ανάλυση αυτών των δεδομένων προϋποθέτει εξελιγμένες τεχνικές, με τη Μηχανική Μάθηση (Machine Learning) να διαδραματίζει πρωταγωνιστικό ρόλο σε αυτό το πεδίο. Η παρούσα διπλωματική εργασία αποσκοπεί στη μελέτη των αλγορίθμων μηχανικής μάθησης που χρησιμοποιούνται για την ανάλυση των μέσων κοινωνικής δικτύωσης, παρέχοντας μια θεωρητική και τεχνολογική επισκόπηση, καθώς και μια κριτική αξιολόγηση των σύγχρονων προσεγγίσεων.

Αρχικά, θα παρουσιαστεί συνοπτικά το θεωρητικό υπόβαθρο των μέσων κοινωνικής δικτύωσης και της μηχανικής μάθησης, με ιδιαίτερη έμφαση στις βασικές κατηγορίες και λειτουργίες των αλγορίθμων, όπως η επιβλεπόμενη και μη επιβλεπόμενη μάθηση, η ενισχυτική μάθηση, τα νευρωνικά δίκτυα και οι αλγόριθμοι βαθιάς μάθησης. Στη συνέχεια, θα πραγματοποιηθεί μια λεπτομερής ανάλυση των προσεγγίσεων μηχανικής μάθησης που εφαρμόζονται στην ανάλυση των κοινωνικών δικτύων, όπως η ανάλυση συναισθήματος (sentiment analysis), η ανίχνευση παραπληροφόρησης, η πρόβλεψη συμπεριφοράς χρηστών, η ανάλυση κοινοτήτων και η επεξεργασία φυσικής γλώσσας.

Η εργασία εξετάζει επίσης τα πλεονεκτήματα και τις προκλήσεις της χρήσης των αλγορίθμων αυτών, αναδεικνύοντας θέματα όπως η ακρίβεια των μοντέλων, η ανάγκη για μεγάλα και ποιοτικά δεδομένα, τα ζητήματα ηθικής και ιδιωτικότητας, καθώς και τη δυσκολία ερμηνείας των αποτελεσμάτων. Τέλος, αναλύονται τα ανοιχτά ερευνητικά ζητήματα του πεδίου και προτείνονται κατευθύνσεις για μελλοντική έρευνα, όπως η αξιοποίηση νέων αλγορίθμων βαθιάς μάθησης (π.χ. transformers, LLMs), η προσαρμογή μοντέλων σε δυναμικά περιβάλλοντα και η ενσωμάτωση των κοινωνικών δικτύων σε καινοτόμες τεχνολογίες όπως το metaverse.

Η εργασία φιλοδοξεί να προσφέρει ένα ολοκληρωμένο και ερευνητικά τεκμηριωμένο πλαίσιο κατανόησης της συμβολής της μηχανικής μάθησης στην ανάλυση των μέσων κοινωνικής δικτύωσης, συμβάλλοντας τόσο στη θεωρητική εμβάθυνση όσο και στην πρακτική αξιοποίηση των εργαλείων αυτών σε ένα συνεχώς εξελισσόμενο ψηφιακό περιβάλλον.

**Λέξεις-Κλειδιά:** Μηχανική Μάθηση - Μέσα Κοινωνικής Δικτύωσης - Αλγόριθμοι Μάθησης - Ανάλυση Συναισθήματος - Επεξεργασία Φυσικής Γλώσσας - Βαθιά Μάθηση - Ανάλυση Δεδομένων - Επιβλεπόμενη και Μη Επιβλεπόμενη Μάθηση - Ανίχνευση Παραπληροφόρησης - Ανάλυση Κοινοτήτων - Προβλέψεις Συμπεριφοράς Χρηστών - Big Data - Τεχνητή Νοημοσύνη - Δεδομένα Χρηστών - Τάσεις και Θεματολογία

## *Abstract*

The rapid expansion of social media in recent years has generated vast amounts of data that can be utilized for knowledge extraction and for understanding human behavior, trends, and social dynamics. The analysis of this data requires advanced techniques, with Machine Learning playing a central role in this domain. This thesis aims to study the machine learning algorithms used in the analysis of social media, offering a theoretical and technological overview, as well as a critical evaluation of contemporary approaches.

Initially, the theoretical background of social media and machine learning will be briefly presented, with a particular emphasis on the main categories and functions of algorithms, such as supervised and unsupervised learning, reinforcement learning, neural networks, and deep learning algorithms. Subsequently, a detailed analysis will be conducted on the machine learning approaches applied to social media analysis, such as sentiment analysis, misinformation detection, user behavior prediction, community detection, and natural language processing (NLP).

The study also examines the benefits and challenges of using these algorithms, highlighting issues such as model accuracy, the need for large and high-quality datasets, ethical and privacy concerns, and the difficulty of interpreting results. Finally, open research challenges in the field are discussed, and directions for future research are proposed, such as the utilization of new deep learning algorithms (e.g., transformers, LLMs), model adaptation in dynamic environments, and the integration of social media with innovative technologies like the metaverse.

This thesis aspires to provide a comprehensive and academically grounded framework for understanding the contribution of machine learning to social media analysis, contributing both to theoretical insight and to the practical use of these tools in an ever-evolving digital landscape.

**Keywords:** Machine Learning – Social Media – Learning Algorithms – Sentiment Analysis – Natural Language Processing (NLP) – Deep Learning – Data Analysis – Supervised and Unsupervised Learning – Misinformation Detection – Community Detection – User Behavior Prediction – Big Data – Artificial Intelligence – User Data – Trend & Topic Detection

# Κεφάλαιο 1: Εισαγωγή

## 1.1 Σκοπός και Αντικείμενο της Έρευνας

Η ραγδαία εξάπλωση των μέσων κοινωνικής δικτύωσης τα τελευταία χρόνια έχει οδηγήσει στη δημιουργία τεράστιων ποσοτήτων δεδομένων, τα οποία αποτελούν πολύτιμη πηγή πληροφορίας για την κατανόηση της ανθρώπινης συμπεριφοράς, των κοινωνικών τάσεων και της δυναμικής των κοινοτήτων. Η ανάλυση αυτών των δεδομένων παρουσιάζει σημαντικές προκλήσεις λόγω του όγκου, της ποικιλομορφίας και της ταχύτητας με την οποία παράγονται. Στο πλαίσιο αυτό, η Μηχανική Μάθηση (Machine Learning) αναδεικνύεται ως ένα ισχυρό εργαλείο για την επεξεργασία και ανάλυση των δεδομένων των κοινωνικών δικτύων (Kapur & Harikrishnan, 2022), επιτρέποντας την εξαγωγή χρήσιμων συμπερασμάτων και την ανάπτυξη εφαρμογών που καλύπτουν ένα ευρύ φάσμα τομέων, από το μάρκετινγκ έως την ανίχνευση ψευδών ειδήσεων Alassafi, et al., 2023).

Η παρούσα διπλωματική εργασία έχει ως σκοπό τη διερεύνηση και ανάλυση των αλγορίθμων μηχανικής μάθησης που χρησιμοποιούνται για την ανάλυση δεδομένων από μέσα κοινωνικής δικτύωσης. Ειδικότερα, επιδιώκεται η κατανόηση των μεθόδων και τεχνικών που εφαρμόζονται για την εξόρυξη γνώσης από τα δεδομένα αυτά, καθώς και η αξιολόγηση της αποτελεσματικότητάς τους σε διάφορες εφαρμογές. Μέσω αυτής της μελέτης, στοχεύουμε στην ανάδειξη των δυνατοτήτων και των περιορισμών των εν λόγω αλγορίθμων, συμβάλλοντας στην περαιτέρω ανάπτυξη του πεδίου (Acikara et al., 2023).

Το αντικείμενο της έρευνας εστιάζει σε διάφορες πτυχές της εφαρμογής της μηχανικής μάθησης στα μέσα κοινωνικής δικτύωσης. Αρχικά, εξετάζεται η ανάλυση συναισθήματος (sentiment analysis), η οποία αφορά στην κατηγοριοποίηση των συναισθημάτων που εκφράζονται σε αναρτήσεις χρηστών. Η ανάλυση συναισθήματος επιτρέπει την κατανόηση της στάσης των χρηστών απέναντι σε προϊόντα, υπηρεσίες ή γεγονότα, παρέχοντας πολύτιμες πληροφορίες για επιχειρήσεις και οργανισμούς (Rodríguez-Ibáñez et. al., 2023).

Επιπλέον, η εργασία εξετάζει την ανίχνευση ψευδών ειδήσεων (fake news detection), ένα κρίσιμο ζήτημα στην εποχή της πληροφορίας. Η διάδοση αναληθών πληροφοριών μέσω των κοινωνικών δικτύων μπορεί να έχει σοβαρές κοινωνικές και πολιτικές επιπτώσεις. Η εφαρμογή αλγορίθμων μηχανικής μάθησης για την αναγνώριση και φιλτράρισμα τέτοιου περιεχομένου αποτελεί αντικείμενο έντονης έρευνας (Shu et. al., 2017).

Μια άλλη σημαντική πτυχή είναι η πρόβλεψη συμπεριφοράς χρηστών (user behavior prediction). Με την ανάλυση ιστορικών δεδομένων, οι αλγόριθμοι μηχανικής μάθησης μπορούν να προβλέψουν μελλοντικές ενέργειες των χρηστών, όπως η πιθανότητα αλληλεπίδρασης με συγκεκριμένο περιεχόμενο ή η αποχώρησή τους από μια πλατφόρμα. Αυτές οι προβλέψεις είναι ιδιαίτερα χρήσιμες για τον σχεδιασμό στρατηγικών εμπλοκής και διατήρησης χρηστών (Lin, 2016).

Επιπροσθέτως, η εργασία διερευνά την ανάλυση κοινοτήτων (community detection) στα κοινωνικά δίκτυα. Η αναγνώριση ομάδων χρηστών με κοινά ενδιαφέροντα ή συμπεριφορές μπορεί να βοηθήσει στην κατανόηση της δομής και της δυναμικής των κοινωνικών δικτύων, καθώς και στη στοχευμένη προώθηση περιεχομένου (Barabási, 2016).

Τέλος, εξετάζεται η επεξεργασία φυσικής γλώσσας (Natural Language Processing - NLP) σε δεδομένα κοινωνικών δικτύων. Η ανάλυση κειμένου που παράγεται από χρήστες απαιτεί προηγμένες τεχνικές NLP για την εξαγωγή νοήματος, την αναγνώριση οντοτήτων και τη σύνοψη πληροφοριών (Cambria et. al., 2013).

Συνολικά, η εργασία στοχεύει στην παροχή μιας ολοκληρωμένης εικόνας των εφαρμογών της μηχανικής μάθησης στην ανάλυση των μέσων κοινωνικής δικτύωσης, αναδεικνύοντας τις δυνατότητες, τις προκλήσεις και τις μελλοντικές κατευθύνσεις της έρευνας σε αυτό το πεδίο.

## ***1.2 Ερευνητικά Ερωτήματα και Στόχοι***

Η διαμόρφωση σαφών και τεκμηριωμένων ερευνητικών ερωτημάτων αποτελεί ακρογωνιαίο λίθο κάθε επιστημονικής έρευνας, καθώς καθορίζει το ερευνητικό πλαίσιο και τη μεθοδολογική προσέγγιση. Στο πεδίο της μηχανικής μάθησης και της εφαρμογής της στα μέσα κοινωνικής δικτύωσης, η ανάγκη για στοχευμένη και δομημένη έρευνα είναι επιτακτική, δεδομένης της συνεχούς τεχνολογικής εξέλιξης, της πολυπλοκότητας των δεδομένων και των διεπιστημονικών εφαρμογών (Dritsas & Trigka, 2025). Η παρούσα εργασία επιχειρεί να εξετάσει το αντικείμενο αυτό υπό το πρίσμα της θεωρητικής κατανόησης, της τεχνολογικής υλοποίησης και της κοινωνικής επίδρασης.

### **Ερευνητικά Ερωτήματα**

Βάσει της φύσης και της στόχευσης της μελέτης, διαμορφώνονται τα ακόλουθα βασικά ερευνητικά ερωτήματα:

**Ερ. 1: Ποιοι αλγόριθμοι μηχανικής μάθησης εφαρμόζονται κυρίως στην ανάλυση δεδομένων από μέσα κοινωνικής δικτύωσης και ποια είναι τα χαρακτηριστικά, τα πλεονεκτήματα και οι περιορισμοί τους;**

Το πρώτο ερώτημα αποσκοπεί στην καταγραφή και ταξινόμηση των πιο διαδεδομένων αλγορίθμων που χρησιμοποιούνται για την εξαγωγή πληροφορίας από τα κοινωνικά μέσα, όπως οι αλγόριθμοι επιβλεπόμενης και μη επιβλεπόμενης μάθησης, τα νευρωνικά δίκτυα, τα μοντέλα ενισχυτικής μάθησης και οι τεχνικές βαθιάς μάθησης (Imran et al., 2021). Η συγκριτική ανάλυσή τους αποτελεί απαραίτητη προϋπόθεση για την κατανόηση της αποδοτικότητας και της αξιοπιστίας τους σε διαφορετικά πλαίσια εφαρμογής.

**Ερ. 2: Ποιες είναι οι κυριότερες εφαρμογές των αλγορίθμων μηχανικής μάθησης στα κοινωνικά δίκτυα και ποια προβλήματα επιχειρούν να επιλύσουν;**

Το δεύτερο ερώτημα αφορά την εννοιολογική και εμπειρική χαρτογράφηση των πρακτικών εφαρμογών των αλγορίθμων, όπως η ανάλυση συναισθήματος (sentiment analysis), η πρόβλεψη χρηστικής συμπεριφοράς, η ανίχνευση παραπληροφόρησης και η κατηγοριοποίηση περιεχομένου (Pang & Lee, 2008). Μέσω αυτής της ανάλυσης, επιχειρείται η ανάδειξη του ρόλου της μηχανικής μάθησης ως εργαλείου στρατηγικού σχεδιασμού και λήψης αποφάσεων.

**Ερ. 3: Ποια τεχνικά, δεοντολογικά και κοινωνικά ζητήματα προκύπτουν από την εφαρμογή της μηχανικής μάθησης στα δεδομένα των μέσων κοινωνικής δικτύωσης;**

Η χρήση αλγοριθμικών μοντέλων για την ανάλυση δεδομένων που περιέχουν προσωπικές πληροφορίες, συναισθήματα και απόψεις, εγείρει σημαντικά ηθικά και κοινωνικά ερωτήματα, όπως η προστασία της ιδιωτικότητας, η αλγοριθμική μεροληψία, η διαφάνεια και η ερμηνευσιμότητα των μοντέλων (Binns, 2018). Το παρόν ερώτημα εξετάζει το πλαίσιο αυτών των ζητημάτων και τονίζει την ανάγκη ενσωμάτωσης αρχών ηθικής τεχνολογίας.

#### **Ερ. 4: Ποιες είναι οι τάσεις εξέλιξης και τα ανοιχτά ερευνητικά ζητήματα στον τομέα της μηχανικής μάθησης για την ανάλυση κοινωνικών μέσων την επόμενη δεκαετία;**

Η συνεχιζόμενη ανάπτυξη τεχνολογιών όπως τα μεγάλα γλωσσικά μοντέλα (LLMs), τα πολυτροπικά δίκτυα (multimodal networks) και η ενσωμάτωση των social media στο περιβάλλον του metaverse καθιστούν απαραίτητη τη διερεύνηση των μελλοντικών κατευθύνσεων του πεδίου (Zhang & Lu, 2021). Η κατανόηση των ερευνητικών αναγκών και των δυνατοτήτων εφαρμογής αποτελεί καθοριστικό παράγοντα για τη διαμόρφωση στρατηγικών και επιστημονικών προτάσεων.

#### **Στόχοι της Έρευνας**

Οι στόχοι της παρούσας εργασίας καθορίζονται από την ανάγκη για θεωρητική εμβάθυνση, τεχνολογική αποσαφήνιση και κοινωνική ενσυναίσθηση. Συγκεκριμένα, διαμορφώνονται ως εξής:

**Στόχος 1: Η συστηματική καταγραφή και ταξινόμηση των αλγορίθμων μηχανικής μάθησης που εφαρμόζονται στην ανάλυση κοινωνικών μέσων**

Ο στόχος αυτός υπηρετείται μέσω της αναλυτικής παρουσίασης των κυριότερων κατηγοριών και της αποσαφήνισης των θεωρητικών τους βάσεων, προκειμένου να εδραιωθεί ένα επιστημονικό υπόβαθρο για την περαιτέρω μελέτη του πεδίου (Mitchell, 1997).

**Στόχος 2: Η ενδεδειγμένη μελέτη των εφαρμογών μηχανικής μάθησης σε περιβάλλοντα κοινωνικής δικτύωσης**

Προσεγγίζονται πρακτικά παραδείγματα, στα οποία αξιοποιούνται μοντέλα για τη διεξαγωγή αναλύσεων, την υποστήριξη επιχειρηματικών αποφάσεων και την κατανόηση της κοινωνικής δυναμικής (Cambria et al., 2017).

**Στόχος 3: Η ανάδειξη των πλεονεκτημάτων και των προκλήσεων από την εφαρμογή των εν λόγω τεχνικών**

Η εργασία εξετάζει την αποδοτικότητα, την επεκτασιμότητα και τη διαφάνεια των αλγορίθμων, καθώς και τις δυσκολίες που προκύπτουν από την ερμηνεία, τη γενίκευση των μοντέλων και τη διαχείριση των ευαίσθητων δεδομένων (Shah, 2024).

**Στόχος 4: Η διατύπωση προτάσεων και η ενίσχυση της ερευνητικής ατζέντας για το μέλλον**

Η εργασία επιδιώκει να εντοπίσει τα ανοιχτά ζητήματα του πεδίου και να προτείνει στρατηγικές κατευθύνσεις που μπορούν να ενισχύσουν την καινοτομία και την ηθική χρήση των αλγορίθμων στο πλαίσιο της ανάλυσης κοινωνικών μέσων (Dwivedi et al., 2021).

### *1.3 Μεθοδολογία και Δομή της Εργασίας*

Η επιλογή της κατάλληλης μεθοδολογίας αποτελεί κρίσιμο παράγοντα για την επιτυχία κάθε επιστημονικής μελέτης, καθορίζοντας το πλαίσιο ανάλυσης, τα εργαλεία έρευνας και τις στρατηγικές επεξεργασίας των δεδομένων. Η παρούσα διπλωματική εργασία, που επικεντρώνεται στη χρήση αλγορίθμων μηχανικής μάθησης για την ανάλυση των μέσων κοινωνικής δικτύωσης, ακολουθεί ποιοτική και αναλυτική μεθοδολογική προσέγγιση, βασισμένη κυρίως σε δευτερογενή έρευνα (desk research) και κριτική βιβλιογραφική ανασκόπηση (literature review), με στόχο τη θεωρητική εμβάθυνση και την αποτίμηση σύγχρονων εφαρμογών του πεδίου (Creswell & Creswell, 2018).

Η εργασία υιοθετεί συστηματική βιβλιογραφική μεθοδολογία, εντοπίζοντας, καταγράφοντας και αναλύοντας ακαδημαϊκές πηγές, επιστημονικά άρθρα, διεθνείς μελέτες και εφαρμογές πραγματικών σεναρίων. Το επίκεντρο της ανάλυσης δεν είναι η πειραματική εφαρμογή αλγορίθμων σε πρωτογενή δεδομένα, αλλά η θεωρητική κατανόηση και αξιολόγηση του πλαισίου χρήσης τους.

Ακολουθούνται τα εξής μεθοδολογικά βήματα:

- Εντοπισμός και επιλογή έγκριτης βιβλιογραφίας, με έμφαση σε πρόσφατες δημοσιεύσεις (τελευταίας δεκαετίας) σχετικές με μηχανική μάθηση, επεξεργασία φυσικής γλώσσας και ανάλυση δεδομένων κοινωνικών δικτύων (Kitchenham & Charters, 2007).
- Ανάλυση των κυριότερων αλγορίθμων ανά κατηγορία (επιβλεπόμενη, μη επιβλεπόμενη, ενισχυτική, βαθιά μάθηση).
- Επισκόπηση εφαρμογών και ερευνητικών μελετών, που περιγράφουν την υλοποίηση αλγορίθμων σε πραγματικά δεδομένα από Facebook, Twitter (X), Instagram, TikTok και YouTube.
- Κριτική αποτίμηση των τεχνικών, ηθικών και λειτουργικών ζητημάτων, βάσει μελετών περίπτωσης, αλλά και θεωρητικών μοντέλων.
- Η έρευνα δεν επιχειρεί την εξαγωγή γενικεύσιμων συμπερασμάτων με στατιστικές μεθόδους, αλλά τη θεωρητική θεμελίωση και τη συνθετική προσέγγιση του φαινομένου μέσα από την αξιοποίηση πολλαπλών πηγών (Webster & Watson, 2002).

Η εργασία εντάσσεται στο διεπιστημονικό πεδίο της Επιστήμης των Υπολογιστών, με ισχυρές συνδέσεις με την Κοινωνική Πληροφορική, τη Διαχείριση Πληροφορίας, τη Γλωσσική Τεχνολογία και τη Διοίκηση Επιχειρήσεων. Η έρευνα επιχειρεί να γεφυρώσει την τεχνική προσέγγιση με τη στρατηγική χρήση των κοινωνικών μέσων, ιδιαίτερα στο πλαίσιο της επιχειρηματικής ανάλυσης, του marketing και της κοινωνικής επιρροής (Taherdoost, 2023).

Παρόλο που η μελέτη εστιάζει σε δευτερογενή δεδομένα, λαμβάνει υπόψη μεθοδολογικά εργαλεία που χρησιμοποιούνται σε εμπειρικές μελέτες (όπως precision, recall, accuracy, F1-score), προκειμένου να αξιολογήσει κριτικά τις τεχνικές προσεγγίσεις άλλων ερευνητών.

Η εργασία είναι οργανωμένη σε έξι κεφάλαια που εξελίσσονται λογικά και θεματικά:

### **Κεφάλαιο 1: Εισαγωγή**

Περιλαμβάνει την παρουσίαση του ερευνητικού πεδίου, τον σκοπό, τα ερευνητικά ερωτήματα, τους στόχους, τη μεθοδολογία και τη δομή του έργου.

### **Κεφάλαιο 2: Θεωρητικό Πλαίσιο**

Αναλύονται οι βασικές έννοιες των κοινωνικών μέσων, η φύση των δεδομένων που παράγουν, καθώς και οι βασικές αρχές της μηχανικής μάθησης.

### **Κεφάλαιο 3: Αλγόριθμοι Μηχανικής Μάθησης**

Παρουσιάζονται οι βασικές κατηγορίες αλγορίθμων (supervised, unsupervised, deep learning, reinforcement learning), με τεχνική ανάλυση και παραδείγματα εφαρμογών.

### **Κεφάλαιο 4: Εφαρμογές σε Μέσα Κοινωνικής Δικτύωσης**

Εξετάζονται πρακτικές εφαρμογές των αλγορίθμων, όπως η ανάλυση συναισθήματος, η ανίχνευση παραπληροφόρησης, η πρόβλεψη συμπεριφοράς και η ανάλυση κοινοτήτων.

### **Κεφάλαιο 5: Προκλήσεις, Περιορισμοί και Προοπτικές**

Συζητούνται τα ηθικά και τεχνικά ζητήματα, όπως η ερμηνευσιμότητα, η ιδιωτικότητα, η διαφάνεια και οι αλγοριθμικές προκαταλήψεις.

### **Κεφάλαιο 6: Συμπεράσματα και Προτάσεις Μελλοντικής Έρευνας**

Γίνεται σύνοψη των βασικών ευρημάτων, απαντώνται τα ερευνητικά ερωτήματα και κατατίθενται προτάσεις για μελλοντική μελέτη του πεδίου.

## *1.4 Περιορισμοί και Οριοθέτηση της Μελέτης*

Η οριοθέτηση και η αναγνώριση των περιορισμών μιας επιστημονικής εργασίας αποτελεί θεμελιώδες στοιχείο για την ενίσχυση της εγκυρότητας, της αξιοπιστίας και της διαφάνειας της ερευνητικής διαδικασίας. Η παρούσα μελέτη, μολονότι επιχειρεί μια ευρεία και αναλυτική προσέγγιση των εφαρμογών της μηχανικής μάθησης στα μέσα κοινωνικής δικτύωσης, δεν εξαιρείται από εγγενείς περιορισμούς που απορρέουν τόσο από τη φύση του ερευνητικού αντικειμένου όσο και από τη μεθοδολογία που υιοθετεί (Bryman, 2016).

Η εργασία βασίζεται αποκλειστικά σε δευτερογενή βιβλιογραφική έρευνα, υιοθετώντας μία ποιοτική και συνθετική προσέγγιση για τη μελέτη του ερευνητικού αντικειμένου. Δεν περιλαμβάνει πρωτογενή εμπειρική ανάλυση, δηλαδή δεν εφαρμόζει άμεσα αλγορίθμους σε σύνολα δεδομένων μέσω πειραματικής διαδικασίας. Αυτό σημαίνει ότι η αξιολόγηση της αποδοτικότητας των αλγορίθμων βασίζεται σε δευτερογενείς πηγές (π.χ. δημοσιευμένες ερευνητικές εργασίες) και όχι σε συγκριτικά πειράματα που διεξάγονται από τον ερευνητή (Webster & Watson, 2002).

Αν και αυτή η προσέγγιση επιτρέπει τη σφαιρική θεώρηση του πεδίου, περιορίζει τις δυνατότητες εμπειρικής επικύρωσης των θεωρητικών συμπερασμάτων. Ωστόσο, ενισχύει τη θεμελίωση της γνώσης και συνεισφέρει στη θεωρητική εδραίωση του γνωστικού πεδίου (Okoli & Schabram, 2010).

Η παρούσα μελέτη επικεντρώνεται στα μέσα κοινωνικής δικτύωσης (social media) ως πεδίο ανάλυσης και, ειδικότερα, σε πλατφόρμες όπως το Facebook, το Twitter (X), το Instagram, το TikTok και το YouTube. Δεν εξετάζει εφαρμογές μηχανικής μάθησης σε άλλα είδη κοινωνικών δικτύων ή διαδικτυακών κοινοτήτων (π.χ. Reddit, forums, messaging apps), ούτε ασχολείται με παραδοσιακά μέσα μαζικής ενημέρωσης.

Επιπλέον, η μελέτη επικεντρώνεται στον τομέα της ανάλυσης περιεχομένου και συμπεριφοράς και όχι σε άλλους τομείς εφαρμογής της μηχανικής μάθησης, όπως η ρομποτική, η ιατρική διάγνωση ή η βιοπληροφορική (Jordan & Mitchell, 2015). Η εν λόγω οριοθέτηση είναι αναγκαία για τη διατήρηση της εστίασης και της ερευνητικής σαφήνειας.

Καθώς το πεδίο της μηχανικής μάθησης και των κοινωνικών μέσων εξελίσσεται ραγδαία, ο χρονικός παράγοντας αποτελεί ουσιαστικό περιορισμό. Τα μοντέλα, οι τεχνικές και οι αλγόριθμοι που καταγράφονται στη μελέτη είναι επικαιροποιημένα μέχρι τη στιγμή συγγραφής της, γεγονός που ενδέχεται να περιορίσει τη διαχρονικότητα των ευρημάτων (Dwivedi et al., 2021).

Επιπλέον, οι γλωσσικοί περιορισμοί της εργασίας (π.χ. έμφαση σε αγγλόφωνη βιβλιογραφία και εφαρμογές σε δεδομένα στα αγγλικά) αποκλείουν πιθανές προσεγγίσεις που αφορούν γλωσσικά και πολιτισμικά συμφραζόμενα σε άλλες γλώσσες, όπως η ελληνική ή η αραβική (Cambria et al., 2017).

Ένας σημαντικός περιορισμός αφορά τη δεοντολογική διάσταση της χρήσης των δεδομένων κοινωνικών δικτύων. Η εργασία, αν και θεωρητική, αντιμετωπίζει εννοιολογικά ζητήματα που σχετίζονται με την ιδιωτικότητα, τη συγκατάθεση και τη διαχείριση ευαίσθητων πληροφοριών, καθώς η μηχανική μάθηση συχνά βασίζεται σε δεδομένα που περιέχουν προσωπικές απόψεις, συναισθήματα ή κοινωνικά χαρακτηριστικά (Binns, 2018).

Το πρόβλημα εντείνεται όταν τα αλγοριθμικά μοντέλα χρησιμοποιούνται χωρίς επαρκή διαφάνεια ή χωρίς ερμηνευσιμότητα, γεγονός που ενδέχεται να ενισχύσει φαινόμενα αλγοριθμικής μεροληψίας ή κοινωνικού αποκλεισμού (Floridi et al., 2018).

Τέλος, η φύση των αλγορίθμων μηχανικής μάθησης και η πολυπλοκότητα των δεδομένων κοινωνικών μέσων συνεπάγεται περιορισμούς στη γνωσιακή κατανόηση και στην ερμηνεία των αποτελεσμάτων. Πολλές τεχνικές (ιδίως τα deep learning μοντέλα) χαρακτηρίζονται από «μαύρο κουτί» (black box) συμπεριφορά, καθιστώντας δύσκολη την κατανόηση του τρόπου με τον οποίο προκύπτουν τα συμπεράσματα (Doshi-Velez & Kim, 2017).

Αυτός ο περιορισμός δεν υποβαθμίζει την αξία της μεθόδου, αλλά απαιτεί κριτική προσέγγιση και συνεχή επανεκτίμηση των αλγοριθμικών εργαλείων.

## *Κεφάλαιο 2: Θεωρητικό Υπόβαθρο*

### *2.1 Ορισμός και Ιστορική Εξέλιξη των Μέσων Κοινωνικής Δικτύωσης*

Τα μέσα κοινωνικής δικτύωσης (ΜΚΔ) αποτελούν διαδικτυακές πλατφόρμες που επιτρέπουν στους χρήστες να δημιουργούν, να μοιράζονται και να ανταλλάσσουν πληροφορίες, ιδέες και περιεχόμενο σε εικονικές κοινότητες και δίκτυα. Σύμφωνα με τους Kaplan και Haenlein (2010), τα ΜΚΔ ορίζονται ως «μια ομάδα διαδικτυακών εφαρμογών που βασίζονται στα ιδεολογικά και τεχνολογικά θεμέλια του Web 2.0 και επιτρέπουν τη δημιουργία και την ανταλλαγή περιεχομένου που δημιουργείται από τους χρήστες». Αυτή η έννοια διαφοροποιεί τα ΜΚΔ από τις παραδοσιακές μορφές επικοινωνίας, καθώς δίνουν έμφαση στη συμμετοχή των χρηστών και στη δημιουργία περιεχομένου από αυτούς. Η εξέλιξη των ΜΚΔ είναι άρρηκτα συνδεδεμένη με την ανάπτυξη του Διαδικτύου και των τεχνολογιών επικοινωνίας. Η ιστορία τους μπορεί να διακριθεί σε διάφορες φάσεις:.

Οι απαρχές των ΜΚΔ εντοπίζονται στις πρώτες μορφές διαδικτυακής επικοινωνίας, όπως το ηλεκτρονικό ταχυδρομείο (email) και τα συστήματα συζητήσεων (chat systems) που εμφανίστηκαν τη δεκαετία του 1970. Το 1979, δημιουργήθηκε το USENET, ένα σύστημα που επέτρεπε στους χρήστες να δημοσιεύουν και να λαμβάνουν μηνύματα σε θεματικές ομάδες συζήτησης (newsgroups), αποτελώντας πρόδρομο των σύγχρονων φόρουμ (Britannica, 2025). Αυτές οι πρώιμες μορφές επικοινωνίας έθεσαν τα θεμέλια για την ανάπτυξη πιο εξελιγμένων πλατφορμών κοινωνικής δικτύωσης.

Με την εξάπλωση του Διαδικτύου τη δεκαετία του 1990, εμφανίστηκαν οι πρώτες πλατφόρμες κοινωνικής δικτύωσης. Το 1997, το SixDegrees.com θεωρείται το πρώτο σύγχρονο κοινωνικό δίκτυο, επιτρέποντας στους χρήστες να δημιουργούν προφίλ και να συνδέονται με φίλους. Παρά την καινοτομία του, το SixDegrees έκλεισε το 2001 λόγω έλλειψης βιωσιμότητας (Jones, 2015). Παράλληλα, το 1995, το Classmates.com δημιουργήθηκε με στόχο την επανασύνδεση παλιών συμμαθητών, θέτοντας τις βάσεις για την έννοια της διαδικτυακής κοινωνικής δικτύωσης (Britannica, 2025).

Η δεκαετία του 2000 σηματοδότησε την εμφάνιση και καθιέρωση των ΜΚΔ όπως τα γνωρίζουμε σήμερα. Το 2003, το MySpace έγινε δημοφιλές, επιτρέποντας στους χρήστες να δημιουργούν εξατομικευμένα προφίλ και να μοιράζονται μουσική και φωτογραφίες. Το 2004, το Facebook ιδρύθηκε από τον Mark Zuckerberg, αρχικά ως πλατφόρμα για φοιτητές του Harvard και σύντομα επεκτάθηκε παγκοσμίως, αλλάζοντας τον τρόπο με τον οποίο οι άνθρωποι αλληλεπιδρούν διαδικτυακά (Steinhauer, 2021). Το 2006, το Twitter εισήγαγε την έννοια των "tweets", επιτρέποντας στους χρήστες να δημοσιεύουν σύντομα μηνύματα 140 χαρακτήρων, δημιουργώντας μια νέα μορφή επικοινωνίας σε πραγματικό χρόνο (Fournier-Tombs, 2020).

Η επόμενη δεκαετία χαρακτηρίστηκε από την εμφάνιση ποικίλων πλατφορμών με εξειδικευμένες λειτουργίες. Το 2010, το Instagram εισήγαγε την κοινή χρήση φωτογραφιών με φίλτρα, ενώ το Snapchat, που κυκλοφόρησε το 2011, επέτρεψε την αποστολή εφήμερων μηνυμάτων που εξαφανίζονται μετά από λίγο χρόνο. Το 2016, το TikTok έκανε την εμφάνισή του, εστιάζοντας σε σύντομα βίντεο με μουσική και εφέ, κερδίζοντας γρήγορα δημοτικότητα, ιδιαίτερα μεταξύ των νεότερων ηλικιών (Our World in Data, 2019).

## 2.2 Η Ανάλυση των Δεδομένων στα Μέσα Κοινωνικής Δικτύωσης (Social Media Analytics)

Η Ανάλυση Δεδομένων στα Μέσα Κοινωνικής Δικτύωσης (Social Media Analytics) αναφέρεται στη διαδικασία συλλογής, μέτρησης και ανάλυσης δεδομένων από πλατφόρμες κοινωνικής δικτύωσης με σκοπό την εξαγωγή πολύτιμων πληροφοριών για τη λήψη στρατηγικών επιχειρηματικών αποφάσεων. Σύμφωνα με τον ορισμό της IBM, η ανάλυση αυτή επιτρέπει τη συλλογή και την εξαγωγή νοήματος από δεδομένα που προέρχονται από κοινωνικά δίκτυα, υποστηρίζοντας έτσι τις επιχειρηματικές αποφάσεις και μετρώντας την απόδοση των ενεργειών που βασίζονται σε αυτές τις αποφάσεις μέσω των κοινωνικών μέσων (IBM, χ.χ.).

Η σημασία της ανάλυσης δεδομένων στα μέσα κοινωνικής δικτύωσης έγκειται στην ικανότητά της να παρέχει στις επιχειρήσεις και τους οργανισμούς βαθύτερη κατανόηση της συμπεριφοράς και των προτιμήσεων του κοινού τους. Μέσω της ανάλυσης αυτής, οι επιχειρήσεις μπορούν να εντοπίσουν τάσεις, να αξιολογήσουν την αποτελεσματικότητα των καμπανιών τους και να προσαρμόσουν τις στρατηγικές τους για τη βελτίωση της αλληλεπίδρασης με το κοινό τους (TechTarget, 2025).

Η εξέλιξη της ανάλυσης δεδομένων στα μέσα κοινωνικής δικτύωσης είναι στενά συνδεδεμένη με την ανάπτυξη των ίδιων των πλατφορμών κοινωνικής δικτύωσης και της τεχνολογίας ανάλυσης δεδομένων. Αρχικά, οι επιχειρήσεις βασίζονταν σε απλές μετρήσεις, όπως ο αριθμός των "likes" ή των "followers", για να αξιολογήσουν την απόδοση τους στα κοινωνικά μέσα. Ωστόσο, με την πάροδο του χρόνου και την αύξηση της πολυπλοκότητας των δεδομένων, αναπτύχθηκαν πιο προηγμένες τεχνικές ανάλυσης που επιτρέπουν βαθύτερη κατανόηση της συμπεριφοράς των χρηστών και των τάσεων της αγοράς (Bounteous, 2010).

Η εμφάνιση εργαλείων όπως το Google Analytics και το Facebook Insights επέτρεψε στις επιχειρήσεις να παρακολουθούν λεπτομερώς την απόδοση των καμπανιών τους και να λαμβάνουν αποφάσεις βάσει δεδομένων. Παράλληλα, η ανάπτυξη τεχνολογιών τεχνητής νοημοσύνης και μηχανικής μάθησης έχει ενισχύσει την ικανότητα ανάλυσης μεγάλων όγκων δεδομένων, επιτρέποντας την εξαγωγή πιο ακριβών και χρήσιμων συμπερασμάτων (Grazitti Interactive, 2023).

Η διαδικασία ανάλυσης δεδομένων στα μέσα κοινωνικής δικτύωσης περιλαμβάνει διάφορα στάδια:

1. **Συλλογή Δεδομένων:** Αφορά τη συγκέντρωση δεδομένων από διάφορες πλατφόρμες κοινωνικής δικτύωσης, όπως Facebook, Twitter, Instagram κ.λπ. Τα δεδομένα μπορεί να περιλαμβάνουν αναρτήσεις, σχόλια, αντιδράσεις, κοινοποιήσεις και άλλα.
2. **Επεξεργασία και Καθαρισμός Δεδομένων:** Τα συλλεγόμενα δεδομένα συχνά περιέχουν θόρυβο ή μη σχετικές πληροφορίες. Σε αυτό το στάδιο, τα δεδομένα καθαρίζονται και οργανώνονται για να είναι κατάλληλα για ανάλυση.

3. **Ανάλυση Δεδομένων:** Εφαρμόζονται διάφορες τεχνικές ανάλυσης, όπως η ανάλυση συναισθήματος, η ανάλυση δικτύων και η εξόρυξη κειμένου, για την εξαγωγή χρήσιμων πληροφοριών από τα δεδομένα.
4. **Οπτικοποίηση και Ερμηνεία Αποτελεσμάτων:** Τα αποτελέσματα της ανάλυσης παρουσιάζονται μέσω γραφημάτων, πινάκων και άλλων οπτικοποιήσεων, διευκολύνοντας την κατανόηση και την ερμηνεία τους από τους ενδιαφερόμενους φορείς.

Η αποτελεσματική ανάλυση δεδομένων απαιτεί τη χρήση κατάλληλων εργαλείων και τεχνικών, καθώς και την κατανόηση του πλαισίου και των στόχων της επιχείρησης (Sprout Social, 2024).

Υπάρχουν διάφορες τεχνικές και εργαλεία που χρησιμοποιούνται στην ανάλυση δεδομένων των μέσων κοινωνικής δικτύωσης:

- **Ανάλυση Συναισθήματος (Sentiment Analysis):** Αναφέρεται στην αξιολόγηση των συναισθημάτων που εκφράζονται σε αναρτήσεις, σχόλια ή hashtags, βοηθώντας στην κατανόηση της στάσης του κοινού απέναντι σε προϊόντα, γεγονότα ή μάρκες (Cambria et al., 2022). Οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούνται για να διαχωρίσουν θετικά, αρνητικά ή ουδέτερα συναισθήματα.
- **Ανάλυση Περιεχομένου και Θέματος (Content and Topic Analysis):** Περιλαμβάνει τη θεματική κατηγοριοποίηση του περιεχομένου που δημοσιεύεται στα μέσα, ώστε να εντοπιστούν τάσεις, ενδιαφέροντα ή κρίσιμα ζητήματα που απασχολούν συγκεκριμένες ομάδες χρηστών (Weller et al., 2014).
- **Ανάλυση Κοινοτήτων (Community Detection):** Χρησιμοποιείται για να εντοπιστούν ομάδες χρηστών που αλληλεπιδρούν μεταξύ τους σε υψηλό βαθμό. Αυτή η τεχνική αποκαλύπτει κοινωνικές δομές και επιρροές στο δίκτυο (Liu & Liu, 2018).
- **Social Network Analysis (SNA):** Εξετάζει τις σχέσεις μεταξύ χρηστών, τους κόμβους και τις συνδέσεις, προσδιορίζοντας επιρροές, ροές πληροφορίας και την κοινωνική δυναμική του δικτύου (Scott, 2017).
- **Χρήση εργαλείων όπως τα Hootsuite, Google Analytics, Brandwatch, Talkwalker, Sprout Social κ.ά.,** επιτρέπει την αυτοματοποιημένη συλλογή και ανάλυση δεδομένων από πολλαπλές πλατφόρμες, συνδυάζοντας ποσοτικές και ποιοτικές μεθόδους (TechTarget, 2025).

Η ανάλυση δεδομένων στα μέσα κοινωνικής δικτύωσης βρίσκει εφαρμογή σε ευρύ φάσμα τομέων:

- **Μάρκετινγκ και Διαφήμιση:** Οι επιχειρήσεις χρησιμοποιούν δεδομένα για στοχευμένες καμπάνιες, κατανόηση του κοινού, αξιολόγηση προϊόντων και διαχείριση φήμης.

- **Πολιτική Επικοινωνία:** Οι πολιτικοί οργανισμοί αναλύουν τη συμπεριφορά ψηφοφόρων, τις απόψεις του κοινού και τη διάχυση πολιτικών μηνυμάτων.
- **Δημοσιογραφία και Μέσα Μαζικής Ενημέρωσης:** Αναλύουν τις απόψεις του κοινού σε θέματα επικαιρότητας και την εξάπλωση ειδήσεων.
- **Κοινωνική έρευνα και Ανθρωπιστικές επιστήμες:** Μελέτες γύρω από κοινωνικές ανισότητες, τάσεις, πολιτιστικές μετατοπίσεις, μέσω των κοινωνικών δικτύων.
- **Διαχείριση κρίσεων:** Οι οργανισμοί μπορούν να εντοπίζουν έγκαιρα καταστάσεις κρίσης ή αρνητικές καμπάνιες, λαμβάνοντας γρήγορα αντίμετρα.

Παρά τις δυνατότητες της ανάλυσης, υπάρχουν σημαντικές προκλήσεις:

- **Όγκος και Πολυπλοκότητα των Δεδομένων:** Τα big data που προκύπτουν από τα κοινωνικά μέσα απαιτούν ισχυρές υπολογιστικές υποδομές και αλγορίθμους που μπορούν να αντεπεξέλθουν σε ετερογενή, αδόμητα και πολυμορφικά δεδομένα.
- **Προβλήματα Ιδιωτικότητας και Ηθικής:** Η συλλογή και ανάλυση δεδομένων από προσωπικούς λογαριασμούς εγείρει ζητήματα ηθικής, όπως η συγκατάθεση, η αλγοριθμική προκατάληψη και η διαφάνεια.
- **Ψευδείς ειδήσεις και παραπληροφόρηση:** Η ανάλυση κοινωνικών μέσων χρησιμοποιείται πλέον και για τον εντοπισμό bots ή ψευδών ειδήσεων, αλλά οι τεχνικές δεν είναι ακόμη απολύτως ακριβείς.
- **Ερμηνεία και Εξήγηση:** Η εξαγωγή συμπερασμάτων απαιτεί συνδυασμό τεχνικής γνώσης και εννοιολογικής κατανόησης. Η σημασία του συμφραζομένου (context) παραμένει κρίσιμη για την ορθή ερμηνεία των αποτελεσμάτων.

Η Ανάλυση των Δεδομένων στα Μέσα Κοινωνικής Δικτύωσης αποτελεί ένα ταχέως εξελισσόμενο πεδίο, το οποίο συνδυάζει τεχνολογίες όπως η μηχανική μάθηση, η εξόρυξη δεδομένων και η επεξεργασία φυσικής γλώσσας, με εφαρμογές σε επιχειρήσεις, κοινωνικές επιστήμες και πολιτικές στρατηγικές. Παρά τις προκλήσεις που αντιμετωπίζει, αποτελεί βασικό εργαλείο κατανόησης της κοινωνικής δυναμικής στον ψηφιακό κόσμο.

## 2.3 Η Έννοια της Μηχανικής Μάθησης (Machine Learning)

Η Μηχανική Μάθηση (Machine Learning) αποτελεί κλάδο της Τεχνητής Νοημοσύνης (Artificial Intelligence), ο οποίος επικεντρώνεται στην ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν στους υπολογιστές να μαθαίνουν από δεδομένα και να βελτιώνουν την απόδοσή τους σε συγκεκριμένες εργασίες χωρίς να είναι ρητά προγραμματισμένοι. Σύμφωνα με τον Mitchell (1997), ένα σύστημα θεωρείται ότι μαθαίνει από εμπειρία **E** σε σχέση με κάποια τάξη εργασιών **T** και απόδοση **P**, εάν η απόδοσή του στις εργασίες της τάξης **T**, όπως μετράται από την **P**, βελτιώνεται με την εμπειρία **E**.

Η βασική ιδέα της μηχανικής μάθησης είναι η δημιουργία μοντέλων που μπορούν να γενικεύσουν από παραδείγματα, επιτρέποντας έτσι την πρόβλεψη ή την κατηγοριοποίηση νέων, μη γνωστών δεδομένων. Αυτό επιτυγχάνεται μέσω της αναγνώρισης προτύπων και σχέσεων στα δεδομένα εκπαίδευσης, τα οποία στη συνέχεια εφαρμόζονται σε νέα δεδομένα.

Η εξέλιξη της μηχανικής μάθησης μπορεί να εντοπιστεί στις αρχές του 20ού αιώνα, με τις πρώτες μελέτες στον τομέα της στατιστικής και της θεωρίας της πληροφορίας. Ωστόσο, η πραγματική ανάπτυξη ξεκίνησε τη δεκαετία του 1950, όταν οι ερευνητές άρχισαν να εξετάζουν την πιθανότητα δημιουργίας μηχανών που μπορούν να μαθαίνουν από δεδομένα.

Το 1959, ο Arthur Samuel εισήγαγε τον όρο "machine learning" και ανέπτυξε ένα πρόγραμμα που μπορούσε να παίζει το παιχνίδι του ντάμα, μαθαίνοντας από τις προηγούμενες παρτίδες του. Αυτή η εργασία θεωρείται μία από τις πρώτες εφαρμογές της μηχανικής μάθησης.

Τη δεκαετία του 1980, η ανάπτυξη των νευρωνικών δικτύων και των αλγορίθμων όπως ο αλγόριθμος αναδρομικής διάδοσης (backpropagation) έδωσε νέα ώθηση στον τομέα, επιτρέποντας την εκπαίδευση πολυεπίπεδων νευρωνικών δικτύων.

Στις αρχές του 21ου αιώνα, με την αύξηση της υπολογιστικής ισχύος και τη διαθεσιμότητα μεγάλων συνόλων δεδομένων, η μηχανική μάθηση γνώρισε ραγδαία ανάπτυξη, οδηγώντας σε σημαντικές προόδους σε τομείς όπως η αναγνώριση εικόνας, η επεξεργασία φυσικής γλώσσας και η ρομποτική.

Η μηχανική μάθηση διακρίνεται σε διάφορες κατηγορίες, ανάλογα με τον τρόπο εκπαίδευσης και τον τύπο των δεδομένων που χρησιμοποιούνται:

1. **Επιβλεπόμενη Μάθηση (Supervised Learning):** Σε αυτήν την κατηγορία, το μοντέλο εκπαιδεύεται σε ένα σύνολο δεδομένων που περιλαμβάνει εισόδους και τις αντίστοιχες επιθυμητές εξόδους. Στόχος είναι το μοντέλο να μάθει τη σχέση μεταξύ εισόδων και εξόδων, ώστε να μπορεί να προβλέπει την έξοδο για νέες εισόδους. Παραδείγματα περιλαμβάνουν την ταξινόμηση (classification) και την παλινδρόμηση (regression).

2. **Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning):** Εδώ, το μοντέλο εκπαιδεύεται σε δεδομένα που δεν περιλαμβάνουν επιθυμητές εξόδους. Στόχος είναι η ανακάλυψη κρυφών δομών ή προτύπων στα δεδομένα. Παραδείγματα περιλαμβάνουν την ομαδοποίηση (clustering) και τη μείωση διαστάσεων (dimensionality reduction).
3. **Ημι-επιβλεπόμενη Μάθηση (Semi-supervised Learning):** Αυτή η προσέγγιση χρησιμοποιεί ένα μικρό σύνολο επισημασμένων δεδομένων μαζί με ένα μεγαλύτερο σύνολο μη επισημασμένων δεδομένων για την εκπαίδευση του μοντέλου. Είναι ιδιαίτερα χρήσιμη όταν η επισημάνση δεδομένων είναι δαπανηρή ή χρονοβόρα.
4. **Ενισχυτική Μάθηση (Reinforcement Learning):** Σε αυτήν την κατηγορία, ένας πράκτορας μαθαίνει να λαμβάνει αποφάσεις μέσω αλληλεπίδρασης με ένα περιβάλλον, λαμβάνοντας ανταμοιβές ή ποινές για τις ενέργειές του. Στόχος είναι η μεγιστοποίηση της συνολικής ανταμοιβής.

Υπάρχουν πολλοί αλγόριθμοι που χρησιμοποιούνται στη μηχανική μάθηση, μερικοί από τους πιο κοινούς περιλαμβάνουν:

- **Γραμμική Παλινδρόμηση (Linear Regression):** Χρησιμοποιείται για την πρόβλεψη αριθμητικών τιμών με βάση τη γραμμική συσχέτιση μεταξύ μεταβλητών. Αποτελεί έναν από τους απλούστερους και πιο κατανοητούς αλγορίθμους, ιδανικό για ερμηνεύσιμα μοντέλα.
- **Λογιστική Παλινδρόμηση (Logistic Regression):** Χρησιμοποιείται για δυαδική ταξινόμηση (π.χ. ναι/όχι, αληθές/ψευδές). Παρά το όνομά της, δεν πρόκειται για μέθοδο παλινδρόμησης αλλά για μέθοδο ταξινόμησης.
- **Δέντρα Απόφασης και Τυχαία Δάση (Decision Trees & Random Forests):** Τα decision trees διαμορφώνουν ένα δενδρικό μοντέλο βασισμένο σε ερωτήσεις-κλειδιά για την ταξινόμηση ή παλινδρόμηση. Τα random forests συνδυάζουν πολλά δέντρα, βελτιώνοντας τη σταθερότητα και την ακρίβεια.
- **k-Πλησιέστεροι Γείτονες (k-Nearest Neighbors):** Είναι ένας αλγόριθμος που βασίζεται στην απόσταση μεταξύ των σημείων δεδομένων για να προβλέψει την κλάση ή τιμή μιας νέας παρατήρησης.
- **Νευρωνικά Δίκτυα και Βαθιά Μάθηση (Neural Networks & Deep Learning):** Τα νευρωνικά δίκτυα αποτελούν τη βάση της βαθιάς μάθησης, με τη χρήση πολλαπλών στρωμάτων που μαθαίνουν πολυεπίπεδες αναπαραστάσεις των δεδομένων. Οι τεχνικές deep learning (όπως τα CNNs και RNNs) έχουν οδηγήσει σε εξαιρετικές επιδόσεις στην αναγνώριση εικόνας, φωνής και γλώσσας.

Η μηχανική μάθηση αποτελεί βασικό εργαλείο στην επιστήμη των δεδομένων (data science), καθώς επιτρέπει την αξιοποίηση των μεγάλων δεδομένων (Big Data) για την εξαγωγή προβλέψεων, συσχετίσεων και τάσεων. Είναι η καρδιά της γνωστικής υπολογιστικής (cognitive computing) και ενσωματώνεται πλέον σχεδόν σε κάθε πληροφοριακό σύστημα που απαιτεί "έξυπνη" συμπεριφορά (Provost & Fawcett, 2013).

Η σύζευξη της μηχανικής μάθησης με τεχνολογίες όπως το cloud computing, η επεξεργασία φυσικής γλώσσας (NLP), το IoT και τα γραφήματα γνώσης, καθιστούν τη Μηχανική Μάθηση απαραίτητη για την επεξεργασία και κατανόηση περίπλοκων, δυναμικών και υψηλού όγκου δεδομένων (Bennett et. al., 2024).

Η Μηχανική Μάθηση έχει αναδειχθεί ως μια από τις πιο ισχυρές και επαναστατικές τεχνολογίες της σύγχρονης εποχής, προσφέροντας μια σειρά από αξιοσημείωτα πλεονεκτήματα που ενισχύουν την αποτελεσματικότητα και την καινοτομία σε πολλούς τομείς. Πρώτα απ' όλα, ένα από τα πιο σημαντικά οφέλη της είναι η αυτοματοποίηση διαδικασιών και λήψης αποφάσεων. Τα μοντέλα μηχανικής μάθησης μπορούν να μάθουν από δεδομένα και να λαμβάνουν αποφάσεις ή να εκτελούν ενέργειες χωρίς ανθρώπινη παρέμβαση, εξοικονομώντας χρόνο και κόπο σε επιχειρήσεις, οργανισμούς και συστήματα πληροφορικής. Αυτή η δυνατότητα είναι ιδιαίτερα σημαντική σε εφαρμογές όπως η αυτόματη ταξινόμηση ηλεκτρονικών μηνυμάτων, η ανίχνευση απάτης στις τραπεζικές συναλλαγές, ή η διαχείριση ροών εργασιών σε βιομηχανικές αλυσίδες (Mitchell, 1997).

Ένα δεύτερο πλεονέκτημα είναι η ικανότητα των μοντέλων να ανακαλύπτουν μοτίβα και σχέσεις μέσα σε δεδομένα που δεν είναι εμφανή μέσω κλασικών στατιστικών τεχνικών. Σε περιβάλλοντα όπου τα δεδομένα είναι υψηλής διάστασης ή περιέχουν μη γραμμικές σχέσεις, η μηχανική μάθηση μπορεί να εντοπίσει πολύπλοκες εξαρτήσεις και να υποστηρίξει πολύτιμα συμπεράσματα, συμβάλλοντας σε βελτιωμένες στρατηγικές πρόβλεψης και λήψης αποφάσεων (James et al., 2021).

Επιπλέον, η μηχανική μάθηση έχει την ικανότητα χειρισμού πολύπλοκων, αδόμητων και μεγάλου όγκου δεδομένων, όπως κείμενα, εικόνες, βίντεο ή ήχοι. Σε αντίθεση με τις παραδοσιακές μεθόδους που προϋποθέτουν προκαθορισμένα χαρακτηριστικά εισόδου, τα σύγχρονα μοντέλα –ιδίως τα βαθιά νευρωνικά δίκτυα (deep learning)– είναι σε θέση να εξάγουν αναπαραστάσεις από τα δεδομένα και να προσαρμόζονται δυναμικά στις μεταβολές του περιβάλλοντος (Goodfellow et al., 2016).

Ένα ακόμη πλεονέκτημα αφορά τη βελτίωση της εμπειρίας χρήστη σε ψηφιακές πλατφόρμες. Η μηχανική μάθηση ενσωματώνεται σε εφαρμογές κοινωνικής δικτύωσης (όπως Facebook και TikTok), ηλεκτρονικού εμπορίου (Amazon, eBay) και συστήματα σύστασης (Netflix, Spotify), επιτρέποντας την προσωποποιημένη εμπειρία, την πρόβλεψη ενδιαφερόντων και τη βελτιστοποίηση περιεχομένου με βάση τη συμπεριφορά του χρήστη (Chaffey & Ellis-Chadwick, 2019).

Ωστόσο, παρά τα πλεονεκτήματα, η μηχανική μάθηση δεν στερείται περιορισμών. Ένας από τους βασικότερους είναι η ανάγκη για μεγάλους όγκους δεδομένων και υπολογιστικούς πόρους. Τα μοντέλα μηχανικής μάθησης, ιδίως τα βαθιά νευρωνικά δίκτυα, απαιτούν τεράστια σύνολα εκπαίδευσης και υπολογιστική ισχύ (όπως GPUs), γεγονός που μπορεί να

καθιστά απαγορευτική τη χρήση τους σε περιβάλλοντα περιορισμένων πόρων (Hastie et al., 2009).

Ένας ακόμη σημαντικός περιορισμός είναι ο κίνδυνος υπερπροσαρμογής (overfitting) ή υποεκπαίδευσης (underfitting). Στην πρώτη περίπτωση, το μοντέλο "μαθαίνει υπερβολικά καλά" τα δεδομένα εκπαίδευσης, με αποτέλεσμα να έχει κακή απόδοση σε νέα δεδομένα. Στη δεύτερη, αποτυγχάνει να συλλάβει τα μοτίβα των δεδομένων, οδηγώντας σε αναποτελεσματικές προβλέψεις. Η ισορροπία ανάμεσα στη γενίκευση και την ακρίβεια παραμένει μια από τις βασικές προκλήσεις στη σχεδίαση αλγορίθμων (Bishop, 2006).

Επιπλέον, η έλλειψη ερμηνευσιμότητας, ιδίως στα πιο περίπλοκα μοντέλα όπως τα deep learning, αποτελεί σοβαρό πρόβλημα, ειδικά σε τομείς όπου η διαφάνεια είναι απαραίτητη (όπως η ιατρική, η νομική ή η δημόσια διοίκηση). Οι τεχνικές αυτές λειτουργούν συχνά ως "μαύρα κουτιά", καθιστώντας δύσκολη τη δικαιολόγηση των αποφάσεων τους (Doshi-Velez & Kim, 2017).

Τέλος, προκύπτουν ζητήματα δεοντολογίας, προκατάληψης και ιδιωτικότητας. Η μηχανική μάθηση συχνά εκπαιδεύεται σε δεδομένα που περιέχουν ενσωματωμένες προκαταλήψεις (bias), οι οποίες αναπαράγονται ή και ενισχύονται από το μοντέλο. Παράλληλα, η χρήση ευαίσθητων προσωπικών δεδομένων εγείρει ηθικά και νομικά ερωτήματα, ειδικά όταν γίνεται χωρίς συναίνεση ή με αδιαφανείς πρακτικές (Binns, 2018). Για να διασφαλιστεί η υπεύθυνη χρήση της μηχανικής μάθησης, είναι κρίσιμο να ενσωματώνονται αρχές δεοντολογίας, διαφάνειας και ελέγχου σε όλα τα στάδια του κύκλου ζωής των αλγορίθμων.

Η μηχανική μάθηση αποτελεί μία από τις πιο καθοριστικές τεχνολογικές και επιστημονικές καινοτομίες του 21ου αιώνα. Η ικανότητά της να προσδίδει στις μηχανές τη δυνατότητα να "μαθαίνουν" και να "προβλέπουν" καθιστά αυτή την τεχνολογία θεμελιώδη για την επιχειρηματικότητα, την επιστημονική έρευνα και την καθημερινή ζωή. Παρά τις προκλήσεις, οι προοπτικές εξέλιξης είναι τεράστιες, ειδικά με την ανάπτυξη νέων μοντέλων όπως τα large language models (LLMs) και τα transformer-based δίκτυα, που αλλάζουν ριζικά τον τρόπο με τον οποίο αντιλαμβανόμαστε την αλληλεπίδραση ανθρώπου-μηχανής.

## 2.4 Διαφορά Μηχανικής Μάθησης – Τεχνητής Νοημοσύνης – Στατιστικής

Κατά την ενασχόληση με την ανάλυση δεδομένων, την αυτοματοποίηση και τη δημιουργία «έξυπνων» συστημάτων, τρεις όροι εμφανίζονται διαρκώς: Τεχνητή Νοημοσύνη (Artificial Intelligence – AI), Μηχανική Μάθηση (Machine Learning – ML) και Στατιστική (Statistics). Οι τρεις αυτοί όροι συχνά συγχέονται μεταξύ τους ή θεωρούνται εναλλακτικοί, όμως στην πραγματικότητα συνιστούν διακριτούς αλλά αλληλοσυμπληρούμενους επιστημονικούς κλάδους. Η αποσαφήνιση των διαφορών τους είναι κρίσιμη για την κατανόηση των θεωρητικών τους βάσεων, της μεθοδολογίας και των εφαρμογών τους (Coursera, 2025).

Η τεχνητή νοημοσύνη αποτελεί έναν ευρύ επιστημονικό τομέα που έχει ως στόχο την προσομοίωση της ανθρώπινης νοημοσύνης μέσω συστημάτων υπολογιστών. Περιλαμβάνει τεχνικές και μεθόδους που επιτρέπουν σε έναν υπολογιστή να εκτελεί εργασίες όπως λήψη αποφάσεων, επίλυση προβλημάτων, επεξεργασία φυσικής γλώσσας και αυτονομία κινήσεων, οι οποίες μέχρι πρόσφατα θεωρούνταν προνόμιο του ανθρώπινου εγκεφάλου (Unite.AI, 2022). Η τεχνητή νοημοσύνη χρησιμοποιεί διάφορες τεχνικές: από αλγορίθμους λογικής και κανόνων, μέχρι μοντέλα μηχανικής μάθησης και νευρωνικά δίκτυα (Google Cloud, 2025).

Η μηχανική μάθηση αποτελεί έναν υποκλάδο της τεχνητής νοημοσύνης. Επικεντρώνεται στη δημιουργία αλγορίθμων που έχουν τη δυνατότητα να «μαθαίνουν» από δεδομένα και να βελτιώνουν σταδιακά την απόδοσή τους χωρίς να απαιτείται ρητός προγραμματισμός (Mitchell, 1997). Ενώ η AI μπορεί να περιλαμβάνει καθαρά προγραμματιστικά συστήματα (π.χ. expert systems), η ML απαιτεί την ύπαρξη δεδομένων για την εκπαίδευση του συστήματος. Μάλιστα, όσο μεγαλύτερο και πιο ποιοτικό είναι το dataset, τόσο πιο αποτελεσματικό μπορεί να είναι το μοντέλο μηχανικής μάθησης (Nasteski, 2017).

Η στατιστική αποτελεί επιστήμη που προϋπάρχει των δύο προηγούμενων πεδίων. Εστιάζει στη συλλογή, περιγραφή, ανάλυση και ερμηνεία ποσοτικών δεδομένων, προκειμένου να εξαχθούν αξιόπιστα συμπεράσματα. Χρησιμοποιεί θεωρίες πιθανοτήτων, στατιστικούς ελέγχους υποθέσεων, παραμετρικά και μη παραμετρικά μοντέλα, επιτρέποντας στους ερευνητές να αξιολογούν την εγκυρότητα των υποθέσεων τους (DataRobot, 2021). Η στατιστική χαρακτηρίζεται από αυστηρότητα, ερμηνευσιμότητα και ακρίβεια.

Ο παρακάτω πίνακας συνοψίζει τις κύριες διαφορές μεταξύ των τριών εννοιών:

Κριτήριο	Τεχνητή Νοημοσύνη (AI)	Μηχανική Μάθηση (ML)	Στατιστική
Ορισμός	Κλάδος που στοχεύει στην προσομοίωση ανθρώπινης νοημοσύνης.	Υποσύνολο της AI που επιτρέπει στις μηχανές να μαθαίνουν από δεδομένα.	Κλάδος των μαθηματικών που ασχολείται με την ανάλυση και ερμηνεία δεδομένων.
Κύριος Σκοπός	Ανάπτυξη «έξυπνων» μηχανών που προσομοιώνουν την ανθρώπινη συμπεριφορά.	Βελτίωση απόδοσης συστημάτων μέσω εμπειρίας και προβλέψεων.	Εξαγωγή συμπερασμάτων για πληθυσμούς από δείγματα.
Μεθοδολογία	Κανόνες, λογική, αναπαραστάσεις γνώσης, μάθηση.	Αλγόριθμοι όπως decision trees, SVM, νευρωνικά δίκτυα.	Παραμετρικές και μη παραμετρικές μέθοδοι, έλεγχοι υποθέσεων.
Εξάρτηση από Δεδομένα	Μέτρια – μπορεί να βασίζεται σε κανόνες ή σε δεδομένα.	Υψηλή – απαιτείται μεγάλος όγκος δεδομένων για εκπαίδευση.	Χαμηλή έως μέτρια – λειτουργεί και με μικρά δείγματα.
Ανάγκη για Ερμηνευσιμότητα	Μέτρια – ανάλογα με την εφαρμογή (σε κρίσιμες εφαρμογές απαιτείται).	Χαμηλή – συχνά λειτουργεί ως «μαύρο κουτί» (ιδίως deep learning).	Πολύ υψηλή – τα μοντέλα πρέπει να είναι πλήρως κατανοητά και ερμηνεύσιμα.
Βαθμός Αυτονομίας	Υψηλός – συχνά αφορά αυτόνομες αποφάσεις και δράση (π.χ. ρομπότ, AI agents).	Μεσαίος – εξαρτάται από την ενσωμάτωσή του σε συστήματα.	Χαμηλός – χρησιμοποιείται ως υποστηρικτικό εργαλείο για ανθρώπινη ερμηνεία.
Εφαρμογές	Ρομποτική, αυτοοδηγούμενα οχήματα, ιατρική διάγνωση, φωνητικοί βοηθοί.	Predictive analytics, chatbots, φίλτρα spam, συνιστώμενα προϊόντα.	Έρευνες αγοράς, μελέτες αξιοπιστίας, δοκιμές προϊόντων, δημοσκοπήσεις.
Υποθέσεις για τα Δεδομένα	Δεν απαιτεί αυστηρές στατιστικές υποθέσεις.	Δεν απαιτεί αυστηρές παραδοχές – ευέλικτο με noisy ή αδόμητα δεδομένα.	Συνήθως απαιτεί σαφείς υποθέσεις (κανονικότητα, ανεξαρτησία, ισότητα διασποράς).

Η τεχνητή νοημοσύνη, η μηχανική μάθηση και η στατιστική δεν είναι αντίπαλες επιστήμες, αλλά τρεις συμπληρωματικοί τομείς που συνεργάζονται στενά στη σύγχρονη επιστημονική και τεχνολογική παραγωγή. Η στατιστική προσφέρει θεωρητική αυστηρότητα και

ερμηνευσιμότητα, η μηχανική μάθηση προσδίδει προσαρμοστικότητα και απόδοση σε μεγάλα δεδομένα, ενώ η τεχνητή νοημοσύνη συνιστά το πλαίσιο μέσα στο οποίο αυτά τα εργαλεία αποκτούν δυναμική αυτονομίας και έξυπνης συμπεριφοράς. Η κατανόηση των διαφορών και των συνδέσεων μεταξύ αυτών των τριών πεδίων είναι απαραίτητη για την ορθή τους αξιοποίηση, ειδικά σε εφαρμογές όπως η ανάλυση κοινωνικών δικτύων, η ανίχνευση παραπληροφόρησης και η επεξεργασία φυσικής γλώσσας.

## 2.5 Εφαρμογές Μηχανικής Μάθησης στον Πραγματικό Κόσμο

Η Μηχανική Μάθηση (Machine Learning - ML) έχει αναδειχθεί ως ένας από τους πιο σημαντικούς κλάδους της Τεχνητής Νοημοσύνης (Artificial Intelligence - AI), επιτρέποντας στα συστήματα να μαθαίνουν από δεδομένα και να βελτιώνουν την απόδοσή τους χωρίς ρητή προγραμματιστική καθοδήγηση. Η εφαρμογή της ML έχει επηρεάσει ποικίλους τομείς, από την υγεία και τη βιομηχανία μέχρι τις χρηματοοικονομικές υπηρεσίες και την καθημερινή ζωή. Σκοπός αυτού του κεφαλαίου είναι η παρουσίαση και ανάλυση των κυριότερων εφαρμογών της μηχανικής μάθησης στον πραγματικό κόσμο, υπογραμμίζοντας τον τρόπο με τον οποίο αυτές οι τεχνολογίες μετασχηματίζουν διάφορους κλάδους και βελτιώνουν τις διαδικασίες και τις υπηρεσίες.

### 1. Υγειονομικός Τομέας

Η μηχανική μάθηση έχει επιφέρει σημαντική πρόοδο στον υγειονομικό τομέα, λειτουργώντας ως καταλύτης για τη βελτίωση της διάγνωσης, της πρόγνωσης και της παροχής εξατομικευμένης ιατρικής φροντίδας. Ειδικότερα, τα συστήματα μηχανικής μάθησης χρησιμοποιούνται ευρέως στην ανάλυση ιατρικών εικόνων, όπως ακτινογραφιών και μαγνητικών τομογραφιών, για την έγκαιρη ανίχνευση ανωμαλιών και την υποστήριξη της διάγνωσης παθήσεων. Ένα χαρακτηριστικό παράδειγμα αποτελεί η αναγνώριση καρδιακών αλλοιώσεων σε πρώιμα στάδια, χάρη στην ικανότητα των αλγορίθμων να εντοπίζουν πρότυπα που δεν είναι ορατά με το ανθρώπινο μάτι (Chinta et. al., 2023).

Παράλληλα, μέσω της ανάλυσης γονιδιωματικών δεδομένων, αλλά και του ιστορικού υγείας των ασθενών, τα μοντέλα μηχανικής μάθησης προσφέρουν δυνατότητες πρόβλεψης της ανταπόκρισης του οργανισμού σε συγκεκριμένα θεραπευτικά σχήματα. Αυτό επιτρέπει την εφαρμογή της λεγόμενης «εξατομικευμένης ιατρικής», δηλαδή τη σχεδίαση θεραπειών προσαρμοσμένων στις ιδιαίτερες ανάγκες κάθε ασθενή. Επιπλέον, η χρήση αλγορίθμων ML στη διαχείριση υγειονομικών πόρων συμβάλλει στην καλύτερη πρόβλεψη της ροής ασθενών και στη βελτιστοποίηση της λειτουργίας νοσοκομείων, ενισχύοντας την αποδοτικότητα και την ποιότητα της παρεχόμενης φροντίδας (IBM, 2025).

### 2. Βιομηχανία και Παραγωγή

Στον βιομηχανικό τομέα, η εφαρμογή της μηχανικής μάθησης έχει αναβαθμίσει σημαντικά τις παραγωγικές διαδικασίες, συμβάλλοντας στην αυτοματοποίηση, την αποτελεσματική συντήρηση του εξοπλισμού και τη βελτιστοποίηση της εφοδιαστικής αλυσίδας. Μέσω της προγνωστικής συντήρησης, τα συστήματα ML αναλύουν δεδομένα που προέρχονται από αισθητήρες μηχανημάτων και προβλέπουν την πιθανότητα εμφάνισης βλάβης. Αυτό επιτρέπει την έγκαιρη παρέμβαση, ελαχιστοποιώντας τους χρόνους διακοπής της παραγωγής και το κόστος των επισκευών (Riccio et. al., 2024).

Επιπροσθέτως, η μηχανική μάθηση συμβάλλει στον αυτοματοποιημένο έλεγχο ποιότητας, καθώς η αναγνώριση ελαττωμάτων ή αστοχιών σε προϊόντα μπορεί να πραγματοποιηθεί

μέσω ανάλυσης εικόνων ή άλλων τύπων δεδομένων, αυξάνοντας την ακρίβεια του ελέγχου και περιορίζοντας τα σφάλματα ανθρώπινου παράγοντα. Επιπροσθέτως, η πρόβλεψη της ζήτησης και η διαχείριση των αποθεμάτων ενισχύεται σημαντικά από τους αλγορίθμους ML, οι οποίοι επιτρέπουν σε επιχειρήσεις να προσαρμόζουν εγκαίρως την παραγωγή και να μειώνουν την αβεβαιότητα στην εφοδιαστική αλυσίδα, οδηγώντας σε χαμηλότερο λειτουργικό κόστος και υψηλότερη αποδοτικότητα (NetSuite, 2024).

### **3. Χρηματοοικονομικές Υπηρεσίες**

Ο χρηματοοικονομικός τομέας αποτελεί έναν από τους πρώτους κλάδους που υιοθέτησε ευρέως τη μηχανική μάθηση, αξιοποιώντας τη για την επεξεργασία μεγάλων ποσοτήτων δεδομένων, την πρόβλεψη χρηματοοικονομικών τάσεων και τη βελτίωση της διαχείρισης κινδύνων. Ιδιαίτερη έμφαση δίνεται στην ανίχνευση απάτης, καθώς τα μοντέλα μηχανικής μάθησης είναι ικανά να αναγνωρίζουν μοτίβα συμπεριφοράς που αποκλίνουν από το φυσιολογικό και να εντοπίζουν έγκαιρα ύποπτες συναλλαγές, προστατεύοντας τόσο τους πελάτες όσο και τους ίδιους τους οργανισμούς από οικονομικές απώλειες.

Επιπλέον, η δυνατότητα ανάλυσης πολύπλοκων και ετερογενών δεδομένων από τις χρηματοοικονομικές αγορές επιτρέπει στα μοντέλα ML να προβλέπουν με μεγαλύτερη ακρίβεια τάσεις, ενισχύοντας τη διαμόρφωση στρατηγικών επενδύσεων και τη λήψη τεκμηριωμένων αποφάσεων. Επιπλέον, η αξιολόγηση της πιστοληπτικής ικανότητας των πελατών έχει αυτοματοποιηθεί σε μεγάλο βαθμό, καθώς οι αλγόριθμοι ML λαμβάνουν υπόψη πληθώρα μεταβλητών (π.χ. ιστορικό συναλλαγών, συμπεριφορικά χαρακτηριστικά) και παρέχουν ταχύτατα αξιόπιστες εκτιμήσεις, μειώνοντας τον χρόνο έγκρισης δανείων και ενισχύοντας την αποτελεσματικότητα των χρηματοπιστωτικών ιδρυμάτων (Patil & Mailcontractor, 2024).

### **4. Λιανικό Εμπόριο και Ηλεκτρονικό Εμπόριο**

Στο σύγχρονο περιβάλλον του λιανικού και ηλεκτρονικού εμπορίου, η μηχανική μάθηση έχει καταστεί καθοριστικός παράγοντας ενίσχυσης της ανταγωνιστικότητας και βελτίωσης της εμπειρίας του καταναλωτή. Οι επιχειρήσεις αξιοποιούν αλγορίθμους ML για να παρέχουν εξατομικευμένες εμπειρίες στους πελάτες τους, να βελτιστοποιούν τις εφοδιαστικές αλυσίδες και να λαμβάνουν αποφάσεις μαρκετινγκ βασισμένες σε δεδομένα. Οι μηχανισμοί συστάσεων (recommender systems) αποτελούν μία από τις πιο ευρέως χρησιμοποιούμενες εφαρμογές, επιτρέποντας σε πλατφόρμες όπως η Amazon, το Netflix και το Spotify να αναλύουν τις προτιμήσεις και τη συμπεριφορά των χρηστών, προσφέροντας προσωποποιημένες προτάσεις για προϊόντα, ταινίες ή μουσική. Η στοχευμένη αυτή προσέγγιση οδηγεί όχι μόνο σε αύξηση των πωλήσεων, αλλά και σε ενίσχυση της ικανοποίησης και διατήρησης πελατών (Chaffey & Ellis-Chadwick, 2019).

Ακόμα, η μηχανική μάθηση ενισχύει την πρόβλεψη της ζήτησης, χρησιμοποιώντας ιστορικά δεδομένα και εξωτερικούς παράγοντες, όπως η εποχικότητα και οι καιρικές συνθήκες. Αυτό

επιτρέπει την πιο αποδοτική διαχείριση αποθεμάτων και την αποφυγή προβλημάτων όπως ελλείψεις ή υπερπαραγωγή, προσφέροντας ακριβέστερες προβλέψεις συγκριτικά με τις παραδοσιακές στατιστικές μεθόδους (NetSuite, 2024). Επιπλέον, η εξυπηρέτηση πελατών έχει βελτιωθεί σημαντικά μέσω της υιοθέτησης chatbots που βασίζονται σε τεχνικές μηχανικής μάθησης και επεξεργασίας φυσικής γλώσσας (NLP). Τα συστήματα αυτά είναι ικανά να απαντούν αυτόματα σε συχνές ερωτήσεις και να καθοδηγούν τους χρήστες κατά την πλοήγηση τους στις πλατφόρμες, προσφέροντας άμεση και αποτελεσματική υποστήριξη (IBM, 2025).

## **5. Κοινωνικά Δίκτυα και Ανάλυση Περιεχομένου**

Η μηχανική μάθηση έχει καταστεί βασικός πυλώνας της λειτουργίας και της εξέλιξης των πλατφορμών κοινωνικής δικτύωσης, όπως το Facebook, το Twitter (νυν X), το TikTok και το Instagram, μετασχηματίζοντας ριζικά τον τρόπο με τον οποίο διανέμεται και καταναλώνεται το ψηφιακό περιεχόμενο. Οι αλγόριθμοι ταξινόμησης και συστάσεων αξιοποιούν δεδομένα συμπεριφοράς των χρηστών, όπως τα likes, τα shares, τα σχόλια και ο χρόνος παραμονής σε συγκεκριμένες δημοσιεύσεις, ώστε να προβλέπουν τα ενδιαφέροντά τους και να προτείνουν σχετικό περιεχόμενο, περιλαμβανομένων προσωποποιημένων διαφημίσεων και προτάσεων επαφών. Η λειτουργία αυτή ενισχύει τη διατήρηση της προσοχής του χρήστη και την αλληλεπίδραση με την πλατφόρμα, αποτελώντας παράλληλα εργαλείο ψηφιακού μάρκετινγκ υψηλής απόδοσης.

Επιπλέον, μέσω της επεξεργασίας φυσικής γλώσσας (Natural Language Processing - NLP), οι πλατφόρμες είναι σε θέση να αναλύουν το συναίσθημα (sentiment) και τον τόνο των δημοσιεύσεων, ταξινομώντας τις ως θετικές, αρνητικές ή ουδέτερες. Η πληροφορία αυτή χρησιμοποιείται από εταιρείες και οργανισμούς για τη διαχείριση της φήμης τους, την κατανόηση των αναγκών του κοινού και την πρόβλεψη νέων κοινωνικών ή καταναλωτικών τάσεων (Salazar et al., 2022). Συγχρόνως, η μηχανική μάθηση συμβάλλει καθοριστικά στην ανίχνευση αυτοματοποιημένων λογαριασμών (bots), spammers και δικτύων παραπληροφόρησης, μέσω της ανάλυσης γλωσσικών μοτίβων, χρονικών συσχετίσεων και συμπεριφορικών ανωμαλιών. Αν και η αντιμετώπιση των ψευδών ειδήσεων και των κακόβουλων χρηστών παραμένει μια πρόκληση με τεχνικά και δεοντολογικά εμπόδια, η χρήση αλγορίθμων ML σε αυτόν τον τομέα αποτελεί έναν από τους πιο ενεργούς και κρίσιμους χώρους έρευνας (Zhou et al., 2020).

## **6. Δημόσια Διοίκηση και Έξυπνες Πόλεις**

Η μηχανική μάθηση βρίσκει ολοένα και ευρύτερη εφαρμογή στον τομέα της δημόσιας διοίκησης, στο πλαίσιο της ανάπτυξης πολιτικών για τις λεγόμενες «έξυπνες πόλεις» (smart cities), οι οποίες αξιοποιούν τεχνολογίες πληροφορικής για την ενίσχυση της αποδοτικότητας, της βιωσιμότητας και της ποιότητας ζωής των πολιτών. Οι κρατικοί φορείς και οι δημοτικές αρχές εφαρμόζουν αλγορίθμους ML για την πρόβλεψη κυκλοφοριακής συμφόρησης, τη ρύθμιση της κυκλοφορίας μέσω έξυπνων φωτεινών σηματοδοτών και τη δυναμική διαχείριση των μετακινήσεων, ειδικά με την ανάλυση δεδομένων GPS, αισθητήρων

και δεδομένων από εφαρμογές αστικής κινητικότητας (Google Cloud, 2025). Αυτές οι εφαρμογές οδηγούν σε μείωση του κυκλοφοριακού φόρτου, εξοικονόμηση ενέργειας και μείωση των εκπομπών ρύπων, ενώ ταυτόχρονα βελτιώνουν την καθημερινότητα των πολιτών.

Επιπλέον, η μηχανική μάθηση χρησιμοποιείται για τον εντοπισμό ανωμαλιών και προβληματικών μοτίβων σε οικονομικά και διοικητικά δεδομένα. Μέσω αυτής της διαδικασίας είναι δυνατή η πρόληψη και η έγκαιρη ανίχνευση φαινομένων κακοδιαχείρισης, σπατάλης πόρων ή ακόμη και απάτης, ενισχύοντας τη διαφάνεια, τη λογοδοσία και την αποτελεσματικότητα του δημόσιου τομέα. Η εφαρμογή της μηχανικής μάθησης σε αυτό το πεδίο αποτελεί ένα σημαντικό βήμα προς την ψηφιακή μεταρρύθμιση της διοίκησης, ενισχύοντας την εμπιστοσύνη των πολιτών στους θεσμούς και προωθώντας την αποδοτικότητα των δημόσιων υπηρεσιών.

Η μηχανική μάθηση δεν είναι πλέον μια θεωρητική έννοια περιορισμένη στα ερευνητικά εργαστήρια. Αποτελεί αναπόσπαστο εργαλείο σε πλήθος εφαρμογών του πραγματικού κόσμου, προσδίδοντας νοημοσύνη, προσαρμοστικότητα και αποτελεσματικότητα σε πολύπλοκα συστήματα. Από τη διάγνωση ασθενειών έως τις προβλέψεις αγοράς και τη διαχείριση της κυκλοφορίας, οι αλγόριθμοι μηχανικής μάθησης ενσωματώνονται σταδιακά σε όλες τις πτυχές της σύγχρονης ζωής. Παρότι υπάρχουν ακόμα προκλήσεις –ιδίως σε θέματα δεοντολογίας, ιδιωτικότητας και ερμηνευσιμότητας– το μέλλον δείχνει ότι οι εφαρμογές της μηχανικής μάθησης θα επεκταθούν ακόμη περισσότερο, ενισχύοντας τη λήψη αποφάσεων, αυτοματοποιώντας πολύπλοκες διαδικασίες και μετασχηματίζοντας την καθημερινότητα σε επίπεδο τόσο ατομικό όσο και συλλογικό. Με την κατάλληλη ρύθμιση, διαφάνεια και ηθικό πλαίσιο, η μηχανική μάθηση μπορεί να αποτελέσει μοχλό βιώσιμης και κοινωνικά υπεύθυνης καινοτομίας.

## Κεφάλαιο 3: Κατηγορίες Αλγορίθμων

### 3.1 Επιβλεπόμενη Μάθηση (Supervised Learning)

Η επιβλεπόμενη μάθηση (supervised learning) αποτελεί έναν από τους θεμελιώδεις πυλώνες της μηχανικής μάθησης, όπου τα μοντέλα εκπαιδεύονται χρησιμοποιώντας δεδομένα που περιλαμβάνουν τόσο τις εισόδους όσο και τις αντίστοιχες επιθυμητές εξόδους. Ο στόχος είναι η εκμάθηση μιας συνάρτησης που, δεδομένων νέων εισόδων, θα μπορεί να προβλέπει με ακρίβεια την αντίστοιχη έξοδο.

Η διαδικασία της επιβλεπόμενης μάθησης ακολουθεί μια σειρά από καλά καθορισμένα στάδια, τα οποία διασφαλίζουν τη συστηματική εκπαίδευση και αξιολόγηση των μοντέλων. Αρχικά, απαιτείται η συλλογή δεδομένων, τα οποία πρέπει να περιλαμβάνουν αντιστοιχίες μεταξύ εισόδων και εξόδων. Για παράδειγμα, σε ένα πρόβλημα αναγνώρισης εικόνων, οι εισοδοί μπορεί να είναι αρχεία εικόνων και οι αντίστοιχες εξοδοί ετικέτες που περιγράφουν το περιεχόμενό τους, όπως "γάτα" ή "σκύλος" (Géron, 2019).

Στη συνέχεια, το σύνολο δεδομένων διαιρείται σε δύο υποσύνολα: το σύνολο εκπαίδευσης (training set), το οποίο χρησιμοποιείται για να εκπαιδευτεί το μοντέλο και το σύνολο δοκιμής (test set), με σκοπό την αξιολόγηση της απόδοσής του σε νέα, άγνωστα δεδομένα. Η συγκεκριμένη διαίρεση είναι κρίσιμη για την αποτίμηση της γενικευσιμότητας του μοντέλου, δηλαδή της ικανότητάς του να αποδίδει καλά σε δεδομένα που δεν έχει "δει" κατά την εκπαίδευση (James et al., 2021).

Η εκπαίδευση του μοντέλου πραγματοποιείται με τη χρήση του training set, κατά την οποία οι παράμετροι του αλγορίθμου προσαρμόζονται σταδιακά ώστε να ελαχιστοποιηθεί η διαφορά μεταξύ των προβλεπόμενων τιμών και των πραγματικών τιμών εξόδου. Οι τεχνικές που χρησιμοποιούνται περιλαμβάνουν τη βελτιστοποίηση συναρτήσεων κόστους (loss functions) με μεθόδους όπως η καθοδική κλίση (gradient descent) (Goodfellow et. al., 2016).

Τέλος, ακολουθεί η αξιολόγηση και βελτιστοποίηση, κατά την οποία το μοντέλο εφαρμόζεται στο test set προκειμένου να εκτιμηθεί η ικανότητά του να γενικεύει σε νέα δεδομένα. Αν η απόδοσή του κρίνεται ανεπαρκής, ενδέχεται να απαιτηθεί αναπροσαρμογή των υπαρχόντων παραμέτρων, χρήση διαφορετικού αλγορίθμου ή εμπλουτισμός του συνόλου εκπαίδευσης. Η διαδικασία αυτή είναι κυκλική και επαναληπτική, με στόχο τη συνεχή βελτίωση της απόδοσης του μοντέλου (Alpaydin, 2020).

Η επιβλεπόμενη μάθηση εφαρμόζεται κυρίως σε δύο κατηγορίες προβλημάτων:

- **Ταξινόμηση (Classification):** Το μοντέλο προβλέπει την κατηγορία στην οποία ανήκει μια είσοδος. Παραδείγματα περιλαμβάνουν την αναγνώριση ανεπιθύμητων μηνυμάτων (spam detection) και την αναγνώριση χειρόγραφων ψηφίων.

- **Παλινδρόμηση (Regression):** Το μοντέλο προβλέπει μια συνεχόμενη αριθμητική τιμή. Παραδείγματα περιλαμβάνουν την πρόβλεψη της τιμής ενός ακινήτου βάσει χαρακτηριστικών όπως η τοποθεσία και το μέγεθος.

Μερικοί από τους πιο διαδεδομένους αλγορίθμους περιλαμβάνουν:

- **Γραμμική Παλινδρόμηση (Linear Regression):** Χρησιμοποιείται για την πρόβλεψη συνεχών τιμών, υποθέτοντας γραμμική σχέση μεταξύ εισόδων και εξόδων.
- **Λογιστική Παλινδρόμηση (Logistic Regression):** Παρά το όνομά της, χρησιμοποιείται κυρίως για δυαδικά προβλήματα ταξινόμησης, προβλέποντας την πιθανότητα μιας εισόδου να ανήκει σε μια κατηγορία.
- **Δέντρα Απόφασης (Decision Trees):** Δομές που λαμβάνουν αποφάσεις βασισμένες σε κανόνες που εξάγονται από τα δεδομένα εκπαίδευσης.
- **Τυχαία Δάση (Random Forests):** Συνδυασμός πολλαπλών δέντρων απόφασης για βελτίωση της ακρίβειας και της γενίκευσης.
- **Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines - SVM):** Αλγόριθμοι που βρίσκουν το υπερεπίπεδο που διαχωρίζει τις κατηγορίες με το μέγιστο περιθώριο.
- **Νευρωνικά Δίκτυα (Neural Networks):** Εμπνευσμένα από τη δομή του ανθρώπινου εγκεφάλου, χρησιμοποιούνται για πολύπλοκα προβλήματα, όπως η αναγνώριση εικόνων και η επεξεργασία φυσικής γλώσσας.

Η επιβλεπόμενη μάθηση έχει κατακτήσει κεντρική θέση στην εφαρμοσμένη μηχανική μάθηση, καθώς αξιοποιείται σε πλήθος τομέων της επιστήμης, της βιομηχανίας και της καθημερινής ζωής. Ένα ενδεικτικό παράδειγμα αποτελεί η αναγνώριση ομιλίας, όπου τα μοντέλα επιβλεπόμενης μάθησης χρησιμοποιούνται για τη μετατροπή ακουστικών σημάτων σε γραπτό λόγο, ενισχύοντας σημαντικά εφαρμογές όπως οι εικονικοί βοηθοί (π.χ. Siri, Alexa) και τα συστήματα μεταγραφής. Στον χρηματοοικονομικό τομέα, η ανίχνευση απάτης βασίζεται σε τέτοιους αλγορίθμους για τον εντοπισμό ύποπτων συναλλαγών που αποκλίνουν από τη συνήθη συμπεριφορά του χρήστη, συμβάλλοντας στην πρόληψη οικονομικών εγκλημάτων και την προστασία των πελατών.

Στην ιατρική διάγνωση, οι επιβλεπόμενοι αλγόριθμοι χρησιμοποιούνται για την πρόβλεψη της πιθανότητας εμφάνισης ασθενειών ή παθολογικών καταστάσεων, αξιοποιώντας δεδομένα όπως ιατρικό ιστορικό, εργαστηριακά αποτελέσματα ή ιατρικές εικόνες. Παράλληλα, η ανάλυση συναισθήματος (sentiment analysis) αξιοποιεί επιβλεπόμενα μοντέλα για την αυτόματη κατηγοριοποίηση κειμένων — όπως αναρτήσεις στα κοινωνικά δίκτυα, κριτικές προϊόντων ή ερωτηματολόγια — σε θετικές, αρνητικές ή ουδέτερες κατηγορίες, προσφέροντας πολύτιμη πληροφόρηση σε εταιρείες, οργανισμούς και πολιτικούς φορείς. Οι εφαρμογές αυτές αποδεικνύουν την πρακτική αξία της επιβλεπόμενης μάθησης σε ένα δυναμικά εξελισσόμενο ψηφιακό περιβάλλον.

Παρότι η επιβλεπόμενη μάθηση έχει καθιερωθεί ως μία από τις πιο ισχυρές και ευέλικτες μεθόδους μηχανικής μάθησης, παρουσιάζει ορισμένες σημαντικές προκλήσεις, τόσο τεχνικές όσο και ηθικές. Αυτές οι προκλήσεις σχετίζονται με τη διαθεσιμότητα και ποιότητα των δεδομένων, την πολυπλοκότητα των μοντέλων, την ανάγκη για ερμηνευσιμότητα και τις ενδεχόμενες κοινωνικές επιπτώσεις που προκύπτουν από τη χρήση τους.

Μία από τις σημαντικότερες τεχνικές προκλήσεις είναι η ανάγκη για μεγάλα, καλά ετικεταρισμένα και ποιοτικά δεδομένα. Η επιτυχία της επιβλεπόμενης μάθησης βασίζεται στην εκπαίδευση των μοντέλων με δεδομένα που περιλαμβάνουν την "σωστή" απάντηση (label), γεγονός που απαιτεί ανθρώπινη παρέμβαση και συχνά εξειδικευμένη γνώση. Ειδικά σε τομείς όπως η ιατρική ή η νομική, η σήμανση των δεδομένων απαιτεί την εμπλοκή ειδικών, γεγονός που αυξάνει σημαντικά το κόστος και τον χρόνο της διαδικασίας. Παράλληλα, η ύπαρξη σφαλμάτων στις ετικέτες μπορεί να οδηγήσει σε μειωμένη ακρίβεια και γενικευσιμότητα των μοντέλων (Alpaydin, 2020).

Ένα ακόμα τεχνικό πρόβλημα είναι η υπερπροσαρμογή (overfitting) ή η υποεκπαίδευση (underfitting) του μοντέλου. Στην πρώτη περίπτωση, το μοντέλο μαθαίνει υπερβολικά καλά τα δεδομένα εκπαίδευσης, συμπεριλαμβανομένων των θορυβωδών ή μη γενικεύσιμων μοτίβων, με αποτέλεσμα να παρουσιάζει φτωχή απόδοση σε νέα δεδομένα. Στη δεύτερη περίπτωση, αποτυγχάνει να εντοπίσει ουσιαστικές συσχετισμούς και μοτίβα, οδηγώντας σε χαμηλή ακρίβεια. Η ισορροπία μεταξύ των δύο αποτελεί διαρκές ζητούμενο κατά την ανάπτυξη επιβλεπόμενων αλγορίθμων (Géron, 2019).

Επιπλέον, οι πιο προηγμένοι αλγόριθμοι επιβλεπόμενης μάθησης, όπως τα βαθιά νευρωνικά δίκτυα, συνεπάγονται υψηλό υπολογιστικό κόστος και πολυπλοκότητα. Η ανάγκη για επεξεργασία τεράστιων όγκων δεδομένων και η χρήση εξελιγμένων τεχνικών εκπαίδευσης, όπως το stochastic gradient descent ή η ρύθμιση υπερπαραμέτρων, προϋποθέτουν σημαντικούς υπολογιστικούς πόρους, όπως GPUs ή TPUs, καθιστώντας την υιοθέτηση τέτοιων μοντέλων δύσκολη για μικρότερους οργανισμούς (Goodfellow et. al., 2016).

Ιδιαίτερη σημασία έχει και η πρόκληση της ερμηνευσιμότητας των μοντέλων. Πολλές μέθοδοι επιβλεπόμενης μάθησης και ιδίως τα νευρωνικά δίκτυα, λειτουργούν ως "μαύρα κουτιά", καθιστώντας δύσκολη τη δικαιολόγηση των αποφάσεών τους. Η έλλειψη διαφάνειας εγείρει ζητήματα λογοδοσίας, ιδιαίτερα σε τομείς όπου οι αποφάσεις επηρεάζουν ανθρώπινες ζωές, όπως η υγειονομική περίθαλψη ή η δημόσια διοίκηση. Η ανάγκη για ερμηνεύσιμη μηχανική μάθηση (interpretable ML) είναι σήμερα πιο επίκαιρη από ποτέ (Doshi-Velez & Kim, 2017).

Τέλος, ένα από τα σοβαρότερα ζητήματα που σχετίζονται με την επιβλεπόμενη μάθηση είναι η ηθική προκατάληψη (bias) που ενδέχεται να ενσωματωθεί στα μοντέλα. Εφόσον τα μοντέλα μαθαίνουν από ιστορικά δεδομένα, η ύπαρξη κοινωνικών ή θεσμικών προκαταλήψεων στα δεδομένα εκπαίδευσης μπορεί να οδηγήσει σε διάκριση εις βάρος συγκεκριμένων πληθυσμιακών ομάδων. Χαρακτηριστικά παραδείγματα περιλαμβάνουν την πρόσληψη εργαζομένων, την αξιολόγηση της πιστοληπτικής ικανότητας και τη χρήση

συστημάτων πρόβλεψης εγκληματικότητας. Η μεροληψία αυτή ενδέχεται να ενισχυθεί και να αναπαραχθεί από τα μοντέλα, οδηγώντας σε ενίσχυση κοινωνικών ανισοτήτων. Η αντιμετώπιση της algorithmic bias αποτελεί σήμερα αντικείμενο ενεργούς επιστημονικής και πολιτικής συζήτησης (Zliobaite, 2017).

Συνοψίζοντας, η επιβλεπόμενη μάθηση δεν αποτελεί απλώς μία από τις βασικές μεθοδολογίες της μηχανικής μάθησης, αλλά έναν ουσιαστικό καταλύτη για την πρόοδο της τεχνητής νοημοσύνης σε κρίσιμους τομείς της επιστήμης, της κοινωνίας και της οικονομίας. Η ικανότητά της να προσφέρει ακριβείς προβλέψεις και λειτουργικές λύσεις την καθιστά απαραίτητο εργαλείο σε περιβάλλοντα όπου η λήψη αποφάσεων βασίζεται σε δεδομένα. Παρά τις προκλήσεις που σχετίζονται με την ποιότητα των δεδομένων, το υπολογιστικό κόστος και τα ζητήματα δεοντολογίας, η διαρκής εξέλιξη των τεχνικών, η ενίσχυση της διαφάνειας και η διεπιστημονική προσέγγιση ενδυναμώνουν τη βιωσιμότητα και τη χρηστικότητά της. Η συνδυαστική χρήση της επιβλεπόμενης μάθησης με άλλες μορφές μάθησης, όπως η μη επιβλεπόμενη και η ενισχυτική, ανοίγει νέους επιστημονικούς και τεχνολογικούς δρόμους, επιτρέποντας την περαιτέρω εμβάθυνση στην κατανόηση σύνθετων φαινομένων και την ανάπτυξη ευφυών, προσαρμοστικών συστημάτων.

### 3.2 Μη Επιβλεπόμενη Μάθηση (*Unsupervised Learning*)

Η μη επιβλεπόμενη μάθηση (*Unsupervised Learning*) αποτελεί ένα από τα βασικά πεδία της μηχανικής μάθησης και της τεχνητής νοημοσύνης, με ολοένα και αυξανόμενη εφαρμογή σε πολλαπλούς τομείς της επιστήμης, της τεχνολογίας και της καθημερινής ζωής. Σε αντίθεση με την επιβλεπόμενη μάθηση, η οποία βασίζεται σε δεδομένα με προκαθορισμένες ετικέτες εξόδου (labels), η μη επιβλεπόμενη μάθηση στοχεύει στην ανάλυση και κατανόηση της εγγενούς δομής των δεδομένων χωρίς να υπάρχει κάποια αναφορά στο τι θεωρείται «σωστό» ή «λάθος» αποτέλεσμα (Alpaydin, 2020).

Η θεμελιώδης αρχή της μη επιβλεπόμενης μάθησης έγκειται στην ικανότητα των αλγορίθμων να εντοπίζουν μοτίβα, συσχετίσεις, ομαδοποιήσεις ή ασυνήθιστα σημεία (outliers) μέσα στα ίδια τα δεδομένα. Η ανάγκη για τέτοιες μεθόδους προκύπτει κυρίως σε περιπτώσεις όπου η ετικετοποίηση των δεδομένων είτε είναι εξαιρετικά δαπανηρή, είτε τεχνικά αδύνατη, όπως για παράδειγμα στη γονιδιακή έκφραση, σε δεδομένα συναλλαγών ή στα κοινωνικά δίκτυα, όπου η ποσότητα πληροφορίας είναι τεράστια και η ερμηνεία της μεταβλητή (James et al., 2021).

Ένας από τους πιο θεμελιώδεις μηχανισμούς της μη επιβλεπόμενης μάθησης είναι η συσταδοποίηση (clustering). Η συσταδοποίηση επιχειρεί να ομαδοποιήσει ένα σύνολο δεδομένων σε κατηγορίες, ή αλλιώς «συστάδες», βάσει της ομοιότητας ή της εγγύτητας των παρατηρήσεων μεταξύ τους. Ο δημοφιλέστερος αλγόριθμος της κατηγορίας αυτής είναι ο K-means, ο οποίος επιχειρεί να χωρίσει τα δεδομένα σε k ομάδες, επιλέγοντας αρχικά τυχαία τα κέντρα των ομάδων και επαναυπολογίζοντάς τα μέχρι να ελαχιστοποιηθεί η εσωτερική διακύμανση (Géron, 2019). Παράλληλα, σημαντικές εναλλακτικές μέθοδοι περιλαμβάνουν τον ιεραρχικό αλγόριθμο συσταδοποίησης (Hierarchical Clustering), ο οποίος δημιουργεί ένα δενδροειδές σύστημα κατηγοριοποίησης και τον DBSCAN, έναν αλγόριθμο βασισμένο στην έννοια της πυκνότητας, κατάλληλο για την αναγνώριση συστάδων ακανόνιστου σχήματος.

Η τεχνική της συσταδοποίησης έχει τεράστια σημασία στην πρακτική εφαρμογή. Στον επιχειρηματικό τομέα, επιτρέπει την τμηματοποίηση πελατών (customer segmentation) βάσει κοινών χαρακτηριστικών, διευκολύνοντας τις στρατηγικές μάρκετινγκ. Στη βιοπληροφορική, χρησιμοποιείται για την αναγνώριση ομάδων γονιδίων με παρόμοια έκφραση, ενώ στην κοινωνιολογία συμβάλλει στην αναγνώριση κοινοτήτων μέσα σε κοινωνικά δίκτυα.

Ένα δεύτερο σημαντικό πεδίο είναι η μείωση διαστασιμότητας (dimensionality reduction). Πολλά σύνολα δεδομένων στις μέρες μας χαρακτηρίζονται από υψηλή διαστασιμότητα, δηλαδή διαθέτουν μεγάλο αριθμό χαρακτηριστικών (features), γεγονός που δημιουργεί προβλήματα τόσο στον υπολογισμό όσο και στην ερμηνεία. Η Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis - PCA) είναι η πιο διαδεδομένη μέθοδος, καθώς επιτρέπει τη μετατροπή ενός μεγάλου συνόλου μεταβλητών σε ένα μικρότερο,

διατηρώντας παράλληλα το μεγαλύτερο ποσοστό της διακύμανσης των δεδομένων (James et al., 2021).

Εκτός από την PCA, άλλες μέθοδοι μείωσης διαστασιμότητας είναι η t-Distributed Stochastic Neighbor Embedding (t-SNE) και η Uniform Manifold Approximation and Projection (UMAP), οι οποίες είναι κατάλληλες για την οπτικοποίηση υψηλής διάστασης δεδομένων σε δύο ή τρεις διαστάσεις. Οι τεχνικές αυτές χρησιμοποιούνται σε εφαρμογές όπως η χαρτογράφηση πληθυσμιακών δεδομένων, η ανάλυση εικόνων και η κατανόηση της κατανομής δεδομένων σε βάσεις μη επισημασμένων εγγραφών (Aggarwal, 2016).

Άλλος βασικός τομέας εφαρμογής της μη επιβλεπόμενης μάθησης είναι η ανίχνευση ανωμαλιών (anomaly detection). Εδώ, ο στόχος είναι να αναγνωριστούν οι παρατηρήσεις που αποκλίνουν από τον γενικό κανόνα ή συμπεριφέρονται διαφορετικά. Πρόκειται για ιδιαίτερα χρήσιμη τεχνική σε περιοχές όπως η ανίχνευση χρηματοοικονομικής απάτης, η ασφάλεια στον κυβερνοχώρο, η προληπτική συντήρηση εξοπλισμού, ή ακόμα και στην ιατρική διάγνωση (Zhang et al., 2021).

Η μη επιβλεπόμενη μάθηση έχει διαδραματίσει επίσης καθοριστικό ρόλο στον τομέα της επεξεργασίας φυσικής γλώσσας (Natural Language Processing - NLP) και της ανάλυσης κειμένου. Οι τεχνικές topic modeling, όπως ο Latent Dirichlet Allocation (LDA), επιτρέπουν την αυτόματη αναγνώριση θεματικών ενοτήτων μέσα σε μεγάλα σώματα κειμένων, χωρίς ανθρώπινη ετικετοποίηση (Blei et al., 2003). Η τεχνική αυτή χρησιμοποιείται από οργανισμούς για τη διαχείριση της φήμης τους, την κατανόηση του κοινωνικού λόγου και την κατηγοριοποίηση της πληροφορίας.

Ένα ακόμα ενδιαφέρον πεδίο εφαρμογής είναι τα συστήματα σύστασης (recommender systems). Πολλές φορές, αυτά τα συστήματα βασίζονται σε τεχνικές μη επιβλεπόμενης μάθησης για να αναγνωρίσουν λανθάνουσες προτιμήσεις χρηστών βάσει των κοινών μοτίβων συμπεριφοράς, ακόμα και χωρίς ρητή δήλωση προτιμήσεων. Πλατφόρμες όπως το Netflix, το Spotify, ή το YouTube, χρησιμοποιούν τέτοιες τεχνικές για να προβλέπουν ποιες ταινίες, τραγούδια ή βίντεο ενδέχεται να ενδιαφέρουν τον κάθε χρήστη, ενισχύοντας την εξατομίκευση (Aggarwal, 2016).

Παρά τα πλεονεκτήματα, η μη επιβλεπόμενη μάθηση παρουσιάζει και σημαντικές προκλήσεις. Η αξιολόγηση των αποτελεσμάτων είναι δυσχερής, καθώς δεν υπάρχει αντικειμενικό μέτρο σύγκρισης χωρίς τις ετικέτες. Αυτό σημαίνει ότι η επιτυχία ενός μοντέλου εξαρτάται συχνά από την ερμηνεία του χρήστη ή από έμμεσα κριτήρια, όπως η συνοχή εντός συστάδων ή η απόσταση μεταξύ τους (Goodfellow et al., 2016).

Επιπλέον, η ερμηνευσιμότητα των αποτελεσμάτων μπορεί να είναι περιορισμένη. Οι ομάδες ή τα πρότυπα που εντοπίζονται μπορεί να μην έχουν προφανή σημασιολογική αξία ή να απαιτούν ανθρώπινη παρέμβαση για την ερμηνεία τους. Οι αλγόριθμοι μη επιβλεπόμενης μάθησης είναι συχνά ευαίσθητοι στις αρχικές παραμέτρους, όπως ο αριθμός των συστάδων ή

ο ορισμός της απόστασης. Αυτή η εξάρτηση καθιστά τη διαδικασία του σχεδιασμού του μοντέλου απαιτητική, ενώ επηρεάζει και τη γενικευσιμότητα των αποτελεσμάτων (Alpaydin, 2020).

Συνοψίζοντας, η μη επιβλεπόμενη μάθηση αποτελεί έναν από τους πιο δημιουργικούς και πολυσχιδείς κλάδους της μηχανικής μάθησης. Παρά τις προκλήσεις που φέρει, όπως η δυσκολία αξιολόγησης και η ανάγκη για ενστικτώδη ερμηνεία, προσφέρει εξαιρετικά ισχυρά εργαλεία για την εξερεύνηση του άγνωστου. Η εξέλιξή της, σε συνδυασμό με την πρόοδο σε τομείς όπως τα deep learning και η ερμηνεύσιμη τεχνητή νοημοσύνη, υπόσχεται να ξεκλειδώσει νέες διαστάσεις στη χρήση των δεδομένων χωρίς την ανάγκη για ανθρώπινη καθοδήγηση.

### 3.3 Ενισχυτική Μάθηση (Reinforcement Learning)

Η ενισχυτική μάθηση (Reinforcement Learning - RL) αποτελεί έναν από τους πλέον καινοτόμους και ταχέως αναπτυσσόμενους τομείς της μηχανικής μάθησης. Σε αντίθεση με την επιβλεπόμενη (supervised) και τη μη επιβλεπόμενη μάθηση (unsupervised), η RL δεν βασίζεται στη χρήση δεδομένων που φέρουν προκαθορισμένες ετικέτες, αλλά λειτουργεί στη βάση της αλληλεπίδρασης ενός πράκτορα (agent) με ένα περιβάλλον (environment), μέσω του οποίου ο πράκτορας μαθαίνει να λαμβάνει αποφάσεις με στόχο τη μεγιστοποίηση μιας ποσοτικής ανταμοιβής (reward). Αυτή η αλληλεπιδραστική προσέγγιση κάνει τη RL κατάλληλη για την επίλυση προβλημάτων λήψης αποφάσεων σε δυναμικά και αβέβαια περιβάλλοντα.

Η RL εμπνέεται από τις αρχές της συμπεριφορικής ψυχολογίας και συγκεκριμένα από τη θεωρία της ενίσχυσης (reinforcement theory), σύμφωνα με την οποία οι πράξεις που επιβραβεύονται είναι πιθανότερο να επαναληφθούν. Η βασική δομή της RL στηρίζεται στο πλαίσιο των Μαρκοβιανών Διαδικασιών Απόφασης (Markov Decision Processes - MDPs). Ένα MDP χαρακτηρίζεται από τα εξής συστατικά: (α) ένα σύνολο καταστάσεων (S), (β) ένα σύνολο ενεργειών (A), (γ) μια συνάρτηση μεταβάσεων (T), η οποία περιγράφει την πιθανότητα μετάβασης από μία κατάσταση σε άλλη μέσω μιας ενέργειας και (δ) μια συνάρτηση ανταμοιβής (R), η οποία καθορίζει την ποσοτική επιβράβευση του πράκτορα μετά από κάθε ενέργεια (Sutton & Barto, 2018).

Ο στόχος του πράκτορα είναι η εύρεση μιας πολιτικής (policy)  $\pi$ , η οποία αντιστοιχεί σε ένα σύνολο κανόνων ή πιθανών επιλογών ενεργειών που μεγιστοποιούν τη συνολική αθροιστική ανταμοιβή (cumulative reward) μακροπρόθεσμα. Η απόδοση μιας πολιτικής μετριέται μέσω μιας συνάρτησης αξίας (value function), η οποία εκτιμά την προσδοκώμενη ανταμοιβή από μία συγκεκριμένη κατάσταση ή ζεύγος κατάστασης-ενέργειας. Υπάρχουν δύο βασικές στρατηγικές προσέγγισης: (α) value-based methods, που επιχειρούν να υπολογίσουν απευθείας τη συνάρτηση αξίας (π.χ., Q-learning) και (β) policy-based methods, που επιδιώκουν την άμεση βελτιστοποίηση της πολιτικής  $\pi$  (Kaelbling et al., 1996).

Η πιο ευρέως διαδεδομένη τεχνική value-based είναι το Q-learning, μια μέθοδος η οποία επιτρέπει στον πράκτορα να μαθαίνει την αναμενόμενη ανταμοιβή κάθε ενέργειας σε δεδομένες καταστάσεις, ακόμη και χωρίς πλήρη γνώση του περιβάλλοντος. Σε περιβάλλοντα που ενσωματώνουν μεγάλη διαστασιμότητα ή πολύπλοκες εισόδους (όπως εικόνες ή σήματα), η RL επεκτείνεται μέσω τεχνικών βαθιάς μάθησης, δημιουργώντας τον υποτομέα Deep Reinforcement Learning (Deep RL). Η πιο εμβληματική εφαρμογή της Deep RL ήταν το Deep Q-Network (DQN), που αναπτύχθηκε από την εταιρεία DeepMind και επέτρεψε σε έναν πράκτορα να μάθει να παίζει παιχνίδια Atari καλύτερα από ανθρώπους, χρησιμοποιώντας μόνο τα pixel της οθόνης ως είσοδο (Mnih et al., 2015).

Πέρα από την ακαδημαϊκή έρευνα, η RL έχει βρει πρακτικές εφαρμογές σε πλήθος πεδίων. Στον τομέα της ρομποτικής, η RL εφαρμόζεται για την εκπαίδευση ρομπότ ώστε να εκτελούν σύνθετες κινήσεις, να αλληλεπιδρούν με άγνωστα αντικείμενα και να συνεργάζονται

σε ομάδες. Στον τομέα των τηλεπικοινωνιών, χρησιμοποιείται για τη βελτιστοποίηση των δικτύων μετάδοσης, ενώ στην αυτονομία οχημάτων, τα συστήματα RL διδάσκουν τα αυτόνομα οχήματα να προσαρμόζουν τη συμπεριφορά τους ανάλογα με τις συνθήκες κυκλοφορίας, επιδιώκοντας ασφάλεια και αποτελεσματικότητα (Ding et al., 2022).

Ακόμα και στην υγειονομική περίθαλψη, η ενισχυτική μάθηση αρχίζει να παίζει σημαντικό ρόλο, καθώς μπορεί να υποστηρίξει την επιλογή εξατομικευμένων θεραπευτικών στρατηγικών μέσω συνεχούς μάθησης από ιατρικά δεδομένα ασθενών. Στον τραπεζικό και χρηματοοικονομικό τομέα, πράκτορες RL χρησιμοποιούνται για τη βελτιστοποίηση στρατηγικών διαπραγμάτευσης, την πρόβλεψη της αγοράς και την αυτόματη κατανομή κεφαλαίων.

Ωστόσο, η RL συνοδεύεται από πλήθος προκλήσεων. Πρωτίστως, η απαίτηση για εκτεταμένη αλληλεπίδραση με το περιβάλλον καθιστά δύσκολη την εφαρμογή της σε σενάρια όπου οι δοκιμές έχουν υψηλό κόστος, όπως στην ιατρική ή στην αεροναυπηγική. Για το λόγο αυτό, αναπτύσσονται τεχνικές όπως η προσομοιωμένη εκπαίδευση (simulated learning) και η μεταφορά μάθησης (transfer learning), ώστε τα μοντέλα να εκπαιδεύονται πρώτα σε εικονικά περιβάλλοντα.

Μια άλλη πρόκληση είναι η διαχείριση της εξερεύνησης (exploration) έναντι της εκμετάλλευσης (exploitation). Ο πράκτορας καλείται να επιλέξει πότε να ακολουθήσει ήδη γνωστές πολιτικές που προσφέρουν υψηλές ανταμοιβές και πότε να πειραματιστεί με νέες ενέργειες, γεγονός που εμπλέκει τη χρήση στρατηγικών όπως ε-greedy ή softmax policies για την εξισορρόπηση της απόφασης.

Επιπλέον, τα συστήματα RL συχνά υποφέρουν από προβλήματα όπως η υπερεκπαίδευση (overfitting) και η ευαισθησία στις αλλαγές του περιβάλλοντος. Στην πράξη, αυτό σημαίνει ότι ένα μοντέλο RL που έχει εκπαιδευτεί επιτυχώς σε ένα συγκεκριμένο περιβάλλον μπορεί να αποτύχει παταγωδώς αν μεταφερθεί σε ένα ελαφρώς τροποποιημένο περιβάλλον. Η γενίκευση παραμένει ένα από τα μεγαλύτερα ανοιχτά ζητήματα στον τομέα της RL.

Ακόμη πιο ανησυχητικό είναι το ζήτημα της ερμηνευσιμότητας. Τα περισσότερα σύγχρονα συστήματα RL, ειδικά εκείνα που ενσωματώνουν βαθιά νευρωνικά δίκτυα, λειτουργούν ως «μαύρα κουτιά», καθιστώντας δύσκολη την κατανόηση του τρόπου με τον οποίο καταλήγουν στις αποφάσεις τους. Αυτό είναι ιδιαίτερα προβληματικό σε τομείς όπως η ιατρική, όπου απαιτείται διαφάνεια και λογοδοσία.

Τέλος, υπάρχουν σημαντικά ζητήματα ηθικής και ασφάλειας. Ένα πράκτορας RL μπορεί να αναπτύξει στρατηγικές που, αν και αποδοτικές ως προς τη συνάρτηση ανταμοιβής, δεν είναι αποδεκτές από ηθική ή νομική σκοπιά. Οι ερευνητές έχουν εντοπίσει περιπτώσεις όπου αλγόριθμοι RL "εκμεταλλεύτηκαν" τυχαία σφάλματα στους μηχανισμούς ανταμοιβής για να αποκτήσουν μέγιστη βαθμολογία χωρίς να επιτελέσουν το καθήκον τους πραγματικά. Ως εκ τούτου, η ενσωμάτωση αρχών ηθικής τεχνητής νοημοσύνης (Ethical AI) στο σχεδιασμό RL συστημάτων θεωρείται επιβεβλημένη (François-Lavet et al., 2018).

Συνοψίζοντας, η ενισχυτική μάθηση είναι ένα εξαιρετικά υποσχόμενο πεδίο της μηχανικής μάθησης, με τη δυνατότητα να μεταμορφώσει ποικίλους τομείς της ανθρώπινης δραστηριότητας. Αν και συνοδεύεται από σημαντικές τεχνικές, θεωρητικές και ηθικές προκλήσεις, η συνεχιζόμενη πρόοδος στους αλγορίθμους, στη μαθηματική μοντελοποίηση και στην υπολογιστική ισχύ καθιστά την RL έναν από τους σημαντικότερους παράγοντες διαμόρφωσης του μέλλοντος της τεχνητής νοημοσύνης.

## **Κεφάλαιο 4: Βασικοί Αλγόριθμοι Μηχανικής Μάθησης**

### **4.1 Γραμμική Παλινδρόμηση και Λογιστική Παλινδρόμηση**

#### **4.1.1 Απλή Γραμμική Παλινδρόμηση**

Η παλινδρόμηση αποτελεί ένα εξαιρετικά χρήσιμο εργαλείο για τον αναλυτή, καθώς επιτρέπει την εξέταση των οικονομικών και χρηματοοικονομικών φαινομένων. Στο επίκεντρο της προσοχής βρίσκεται η περιγραφή και αξιολόγηση των σχέσεων μεταξύ μιας μεταβλητής, η οποία ονομάζεται εξαρτημένη ή μεταβλητή απόκρισης ή προβλέψιμη και μιας ή περισσότερων μεταβλητών, που αναφέρονται ως ανεξάρτητες, προβλεπτικές ή επεξηγηματικές (Gujarati & Porter, 2009).

Η ανεξάρτητη μεταβλητή παίρνει το όνομά της, διότι υπόκειται σε μετρήσεις που πραγματοποιεί ο ερευνητής. Τα αποτελέσματα αυτών των μετρήσεων αναμένεται να επηρεάσουν την εξαρτημένη μεταβλητή, η οποία εξαρτάται άμεσα από αυτές τις τιμές. Όταν υπάρχει αυτή η εξαρτημένη σχέση μεταξύ δύο μεταβλητών, τότε αυτό ονομάζεται παλινδρόμηση. Ειδικότερα, όταν εμπλέκονται μόνο δύο μεταβλητές, αναφερόμαστε σε απλή παλινδρόμηση.

Σε πολλές περιπτώσεις της στατιστικής, αντιμετωπίζουμε το πρόβλημα της ανάλυσης της σχέσης μεταξύ δύο ή περισσότερων τυχαίων μεταβλητών. Παραδείγματα αυτού του είδους σχέσης περιλαμβάνουν τη μελέτη της σχέσης μεταξύ ύψους και βάρους σε μια ομάδα ανθρώπων, του εισοδήματος και της κατανάλωσης εργαζομένων σε μια εταιρεία κ.λπ. Το βασικό ερώτημα είναι να διαπιστώσουμε αν υπάρχει κάποια τέτοια σχέση και, αν ναι, πώς μπορούμε να την περιγράψουμε.

Η σημασία αυτής της μελέτης είναι ιδιαίτερα έντονη σε εφαρμογές που σχετίζονται με επιχειρήσεις και οικονομία, καθώς οι εν λόγω σχέσεις συχνά χρησιμοποιούνται για προβλέψεις. Είναι προφανές ότι εταιρείες, είτε ιδιωτικές είτε κρατικές, χρειάζονται συχνά να προβλέψουν ποσότητες ή χαρακτηριστικά όπως η ζήτηση, τα επιτόκια, ο πληθωρισμός, οι τιμές πρώτων υλών, το εργατικό κόστος κ.λπ. Επομένως, είναι ενδιαφέρον να εξεταστούν οι επιδράσεις που ορισμένες μεταβλητές έχουν σε άλλες μεταβλητές (Stock & Watson, 2019).

Η (γραμμική) παλινδρόμηση είναι μια στατιστική μέθοδος που χρησιμοποιείται για τον προσδιορισμό ενός μαθηματικού μοντέλου που εξηγεί, ερμηνεύει και προβλέπει τις τιμές ενός χαρακτηριστικού (μεταβλητής) βάσει των τιμών πολλών άλλων χαρακτηριστικών (μεταβλητών). Στην αρχή εξετάζουμε την απλούστερη μορφή παλινδρόμησης, την απλή γραμμική παλινδρόμηση, όπου υπάρχει μόνο μια ανεξάρτητη μεταβλητή  $X$  και μια εξαρτημένη μεταβλητή  $Y$  που μπορεί να προσεγγιστεί ικανοποιητικά από μια γραμμική συνάρτηση του  $X$ .

Στην απλή γραμμική παλινδρόμηση, η σχέση μεταξύ των δύο μεταβλητών χαρακτηρίζεται ως αιτιώδης, καθώς οι τιμές της ανεξάρτητης μεταβλητής ερμηνεύουν τις τιμές της εξαρτημένης. Απλούστευση αυτής της σχέσης παλινδρόμησης μπορεί να εκφραστεί ως  $Y = f(X)$ .

Ανεξαρτήτως των λόγων για τους οποίους η ανάλυση της σχέσης μεταξύ δύο ή περισσότερων μεταβλητών είναι χρήσιμη, το πρώτο βήμα είναι η κατασκευή ενός μαθηματικού μοντέλου που να περιγράφει αυτήν τη σχέση. Αυτό το μοντέλο αποσκοπεί στην ερμηνεία, περιγραφή και πρόβλεψη των τιμών μιας εξαρτημένης μεταβλητής σε σχέση με τις τιμές πολλών άλλων μεταβλητών.

Η διαδικασία κατασκευής ενός μοντέλου μπορεί να είναι πολύπλοκη, καθώς απαιτεί γνώση της φύσης της σχέσης μεταξύ των μεταβλητών. Για παράδειγμα, ένας επενδυτής που ενδιαφέρεται να προβλέψει την μελλοντική τιμή μετοχών μπορεί να εξετάζει παράγοντες όπως τα καθαρά έσοδα μιας εταιρείας και τη ζήτηση. Αν ο επενδυτής υποθέσει μια θετική γραμμική σχέση, αυτό σημαίνει ότι μια αύξηση στα καθαρά έσοδα μπορεί να οδηγήσει σε αντίστοιχη αύξηση στην τιμή των μετοχών.

Η σύνδεση μεταξύ της εξαρτημένης μεταβλητής και των ανεξάρτητων μεταβλητών είναι στατιστική και όχι συναρτησιακή. Σε μια στατιστική σχέση, υπολογίζεται μια θεωρητική τιμή της εξαρτημένης μεταβλητής για κάθε τιμή της ανεξάρτητης, ενώ η πραγματική τιμή βρίσκεται εντός ενός εύρους τιμών που περιλαμβάνει τη θεωρητική. Αντίθετα, σε μια συναρτησιακή σχέση, η εξαρτημένη μεταβλητή δίνει πάντα την ίδια τιμή για κάθε τιμή της ανεξάρτητης (στη μορφή  $Y=f(X)$ ).

Παρόλο που χρησιμοποιούμε τον όρο "εξισώσεις παλινδρόμησης" για ευκολία, πρόκειται για ένα στατιστικό μοντέλο και όχι για μια ακριβή εξίσωση. Στο παράδειγμα δύο μεταβλητών,  $X$  και  $Y$ , η συνάρτηση είναι γραμμική, όπως  $Y = \beta_0 + \beta_1 X$  στο γραμμικό υπόδειγμα. Το "γραμμικό" αναφέρεται στις παραμέτρους του μοντέλου, όχι στις μεταβλητές.

Σε πολλούς τομείς, όπως η οικονομία, εμφανίζονται συχνά σχέσεις μεταξύ διάφορων χαρακτηριστικών. Για παράδειγμα, στον τομέα της οικονομίας συναντούμε σχέσεις όπως:

- Κέρδος = Εισπράξεις – Κόστος
- Συνολικό κόστος = Σταθερό κόστος + (Μεταβλητό κόστος · αριθμό μονάδων που παρήχθησαν)

Ωστόσο, στην πραγματική ζωή, σπάνια έχουμε ιδανικές μαθηματικές σχέσεις μεταξύ δύο μεγεθών. Η ύπαρξη σφαλμάτων μέτρησης καθιστά αναπόφευκτες τις αποκλίσεις από τη θεωρητική σχέση. Αυτά τα μετρικά σφάλματα προέρχονται από αβεβαιότητες στη μέτρηση των μεταβλητών. Στην πλειονότητα των πρακτικών περιπτώσεων, χρειάζεται να χρησιμοποιήσουμε μοντέλα που λαμβάνουν υπόψη το στοιχείο της τυχαιότητας, το οποίο αποτελεί ουσιαστικό μέρος της καθημερινής πραγματικότητας.

Όπως έχει αναφερθεί, η απλούστερη περίπτωση παλινδρόμησης είναι η απλή γραμμική παλινδρόμηση, όπου υπάρχει μόνο μία ανεξάρτητη μεταβλητή, συμβολίζεται ως  $X$  και μία εξαρτημένη μεταβλητή, συμβολίζεται ως  $Y$ . Αυτή η κατηγορία εμφανίζεται τόσο σε πειραματικές όσο και σε μη πειραματικές μελέτες.

Στις πειραματικές μελέτες, ο ερευνητής καθορίζει προκαθορισμένα τα επίπεδα της ανεξάρτητης μεταβλητής, όπως οι δόσεις ενός φαρμάκου και μετρά τις αντιδράσεις, που είναι η εξαρτημένη μεταβλητή. Εδώ, η παλινδρόμηση χρησιμεύει για τον προσδιορισμό μιας δυναμικής σχέσης δόσης-αντίδρασης.

Στις μη πειραματικές μελέτες ή δειγματοληψίες, όπου γίνονται μετρήσεις σε διάφορα χαρακτηριστικά για κάθε μονάδα του δείγματος, η ανάλυση εξαρτάται από το ενδιαφέρον του ερευνητή. Αν, για παράδειγμα, ενδιαφέρει πώς το ύψος επηρεάζει το βάρος, τότε το ύψος θεωρείται ως ανεξάρτητη μεταβλητή (X) και το βάρος ως εξαρτημένη μεταβλητή (Y). Έτσι, πραγματοποιείται παλινδρόμηση του βάρους πάνω στο ύψος. Η επιλογή εξαρτημένης και ανεξάρτητης μεταβλητής καθορίζεται από το ερευνητικό ερώτημα και το πεδίο ενδιαφέροντος.

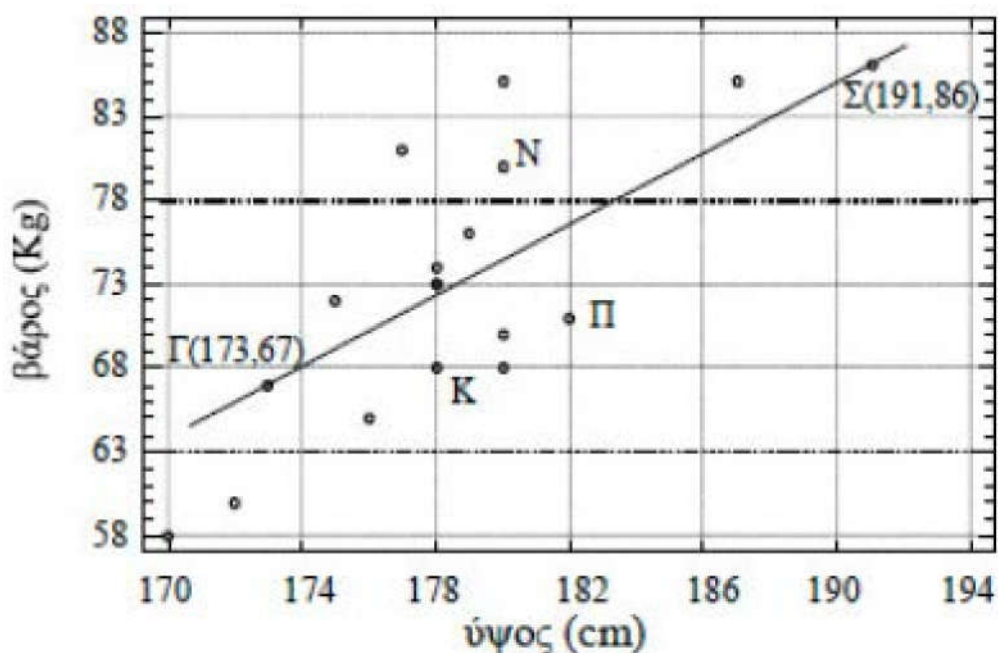
Ο παρακάτω πίνακας δίνει τα ύψη X (σε cm) και τα βάρη Y (σε kg) των 18 μαθητών. Οι τιμές του ύψους δίνονται σε αύξουσα σειρά.

Μαθητής	Ύψος X	Βάρος Y	Μαθητής	Ύψος X	Βάρος Y
A	170	58	K	178	68
B	172	60	Λ	179	76
Γ	173	67	M	180	68
Δ	175	72	N	180	80
E	176	65	Ξ	180	70
Z	177	81	O	180	85
H	178	73	Π	182	71
Θ	178	74	P	187	85
I	178	73	Σ	191	86

Σε αυτήν την περίπτωση, προβαίνουμε σε δύο μετρήσεις για κάθε άτομο (μαθητή) σχετικά με δύο συνεχείς μεταβλητές, το ύψος (X) και το βάρος (Y). Αυτό σημαίνει ότι το δείγμα αποτελείται από τα ζεύγη των τιμών (x, y) που αντιστοιχούν στους μαθητές που εξετάζονται.

Εάν αναπαραστήσουμε αυτά τα ζεύγη παρατηρήσεων σε ένα δισδιάστατο σύστημα αξόνων, παρατηρούμε ότι δημιουργείται μια "διασπορά" των σημείων, τα οποία αντιστοιχούν στους μαθητές. Αυτή η αναπαράσταση στον άξονα x απεικονίζει το ύψος (X) ως ανεξάρτητη μεταβλητή, ενώ στον άξονα y απεικονίζεται το βάρος (Y) ως εξαρτημένη μεταβλητή.

Το γράφημα αυτό, που απεικονίζει τα δεδομένα, ονομάζεται διάγραμμα διασποράς, όπως φαίνεται στο παρακάτω σχήμα. Έτσι, με τη χρήση του διαγράμματος διασποράς, μπορούμε να οπτικοποιήσουμε τη σχέση μεταξύ του ύψους και του βάρους των μαθητών που εξετάζουμε.



Με προσεκτική παρατήρηση του διαγράμματος διασποράς, μπορούμε να αντλήσουμε σημαντικές πληροφορίες σχετικά με την ενδεχόμενη σχέση εξάρτησης ανάμεσα στις εξεταζόμενες μεταβλητές. Η εμπειρία μας υποδεικνύει ότι υπάρχει μια συγκεκριμένη σχέση μεταξύ του ύψους και του βάρους και συγκεκριμένα μια θετική σχέση. Αυτό σημαίνει ότι όταν το ύψος αυξάνεται, το βάρος τείνει να αυξάνεται επίσης.

Αναλύοντας το διάγραμμα διασποράς, παρατηρούμε γενικά ότι ψηλοί μαθητές εμφανίζουν τάση να έχουν και μεγαλύτερο βάρος. Για παράδειγμα, στη σύγκριση μεταξύ των μαθητών Ν, Κ και Π, παρατηρούμε ότι ο Ν, που είναι ψηλότερος, είναι και βαρύτερος από τον Κ, ενώ ο Π, παρότι ψηλότερος από τον Κ, είναι επίσης βαρύτερος από τον Κ.

Η παρατήρηση του διαγράμματος διασποράς στο προηγούμενο παράδειγμα αποκαλύπτει σαφώς την ύπαρξη μιας σχέσης μεταξύ του ύψους (X) και του βάρους (Y) των 18 μαθητών. Τα σημεία (x, y) σχηματίζουν περίπου μια ευθεία, προτείνοντας ότι η σχέση μεταξύ του ύψους και του βάρους είναι περίπου γραμμική.

Για να αναλύσουμε αυτή τη σχέση, μπορούμε να θεωρήσουμε το ύψος (X) ως την ανεξάρτητη μεταβλητή και το βάρος (Y) ως την εξαρτημένη μεταβλητή. Σε αυτήν την περίπτωση, η ευθεία που εφαρμόζεται καλύτερα στα παρατηρούμενα σημεία είναι η ευθεία παλινδρόμησης του βάρους (Y) πάνω στο ύψος (X).

Όπως γνωρίζουμε, η εξίσωση μιας ευθείας δίνεται από τη σχέση:

$$y = \alpha + \beta x \quad (1)$$

όπου  $\alpha$  και  $\beta$  είναι παράμετροι τις οποίες θέλουμε να υπολογίσουμε ή, όπως λέμε, να “εκτιμήσουμε”, έτσι ώστε η ευθεία που θα προκύψει να μας δίνει όσο το δυνατόν την καλύτερη περιγραφή της σχέσης (εξάρτησης) που υπάρχει μεταξύ των μεταβλητών X και Y.

Η παράμετρος  $\alpha$  μας δίνει το σημείο  $(0, \alpha)$ , όπου η ευθεία αυτή τέμνει τον άξονα  $y'y$ , ενώ η παράμετρος  $\beta$  παριστάνει το συντελεστή διεύθυνσης της ευθείας.

Ο πιο εύκολος τρόπος χάραξης της ευθείας είναι αυτός που γίνεται “με το μάτι”. Μια τέτοια ευθεία έχουμε φέρει και στο διάγραμμα διασποράς του παραπάνω σχήματος. Για να βρούμε τα  $\alpha$  και  $\beta$ , εργαζόμαστε ως εξής:

- Επιλέγουμε δύο σημεία, έστω τα  $\Gamma(173,67)$  και  $\Sigma(191,86)$  πάνω στην ευθεία που φέραμε “με το μάτι”.
- Αντικαθιστούμε τις συντεταγμένες  $(x, y)$  των σημείων αυτών στην (1), οπότε προκύπτει το σύστημα:

$$\begin{cases} y_1 = \alpha + \beta x_1 \\ y_2 = \alpha + \beta x_2 \end{cases} \Leftrightarrow \begin{cases} 67 = \alpha + 173\beta \\ 86 = \alpha + 191\beta \end{cases}$$

- Επιλύοντας το σύστημα αυτό βρίσκουμε  $\alpha = -115,6$  και  $\beta = 1,06$  οπότε η εξίσωση της ευθείας (1) γίνεται:  
 $y = -115,6 + 1,06x$  (2)

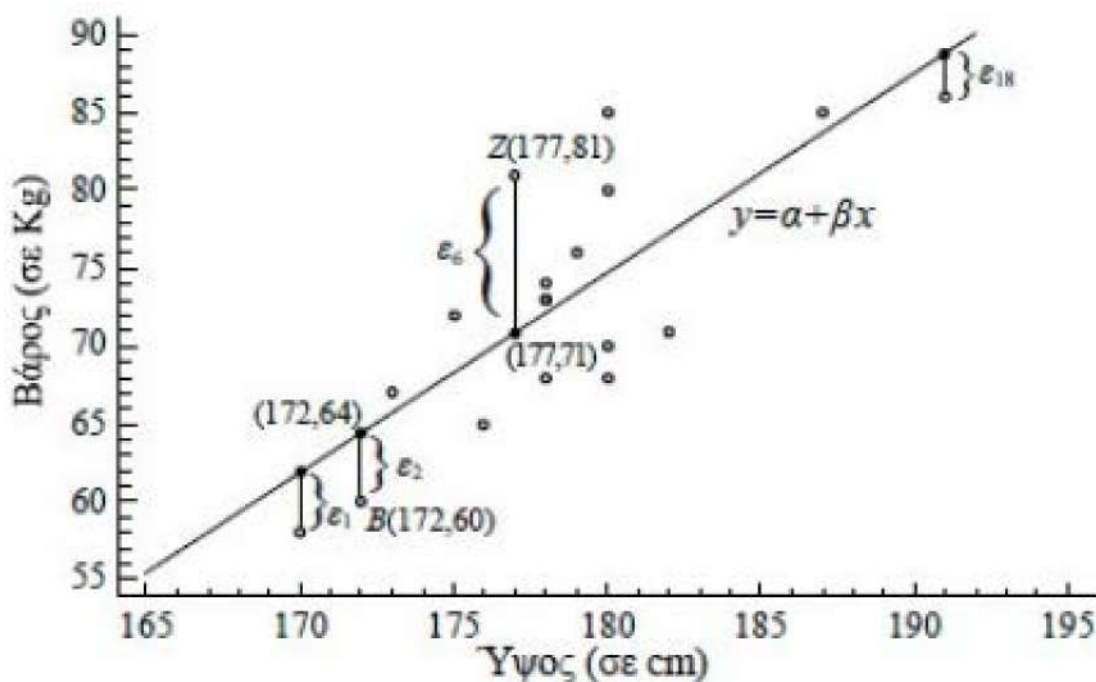
Επομένως, η ευθεία που προσαρμόζεται καλύτερα στα σημεία του διαγράμματος διασποράς διέρχεται από το σημείο  $(0, -115,6)$  και έχει συντελεστή διεύθυνσης 1,06.

### **Εκτίμηση της Ευθείας Παλινδρόμησης με τη Μέθοδο των Ελαχίστων Τετραγώνων**

Παρατηρήσαμε ότι η απλούστερη μέθοδος προσαρμογής μιας ευθείας γραμμής σε ένα διάγραμμα διασποράς είναι η οπτική εκτίμηση, όμως αυτή έχει πολλά μειονεκτήματα παρά την απλότητά της. Κυριότερο από αυτά είναι η έλλειψη αντικειμενικότητας, καθώς διάφορα άτομα μπορούν να σχηματίσουν διαφορετικές ευθείες, ακόμα και το ίδιο άτομο μπορεί να παράγει διαφορετικά αποτελέσματα κάθε φορά.

Γι' αυτό, χρειαζόμαστε μια πιο ακριβή μέθοδο για την προσαρμογή μιας ευθείας σε τέτοια δεδομένα. Μια διαδεδομένη και αξιόπιστη μέθοδος είναι η “μέθοδος ελαχίστων τετραγώνων.” Αυτή η μέθοδο χρησιμοποιείται για την εκτίμηση των παραμέτρων  $\alpha$  (άλφα) και  $\beta$  (βήτα), καθώς και για τον εντοπισμό της εξίσωσης της βέλτιστης ευθείας που προσαρμόζεται στα δεδομένα.

Η μέθοδος αυτή είχε την πρώτη αναφορά της στο έργο του Γάλλου μαθηματικού Legendre το 1805 και στη συνέχεια εφαρμόστηκε από τον Γερμανό μαθηματικό Gauss στην αστρονομική για τον προσδιορισμό της τροχιάς του πλανήτη Δήμητρα.



Πράγματι, ο Gauss αναφέρει ότι χρησιμοποίησε την αρχή των ελαχίστων τετραγώνων πριν από το 1794, καταφέροντας να προηγηθεί του Legendre στην ανακάλυψη αυτής της μεθόδου και αυτό σε ηλικία μόλις 17 ετών. Ας επανεξετάσουμε το διάγραμμα διασποράς του προηγούμενου παραδείγματος, που απεικονίζει τα ύψη (X) και τα βάρη (Y) των 18 μαθητών. Στο εν λόγω διάγραμμα διασποράς, επιφέρουμε μια ευθεία της μορφής  $y = \alpha + \beta x$ , υποθέτοντας ότι αυτή η ευθεία προσαρμόζεται βέλτιστα στα παρατηρούμενα σημεία  $(x_i, y_i)$  για τους 18 μαθητές.

Επομένως, για παράδειγμα, για τον μαθητή B με ύψος  $x^2 = 172$  cm και πραγματικό βάρος  $y_2 = 60$  kg, η προβλεπόμενη τιμή σύμφωνα με την ευθεία παλινδρόμησης είναι περίπου 64 kg. Επομένως, το σφάλμα  $\varepsilon_2 = 60 - 64 = -4$  kg, προκλώντας μια απόκλιση του βάρους κατά 4 kg κάτω από την προβλεπόμενη τιμή. Ομοίως για το μαθητή Z, σημείο Z(177,81), το βάρος του που μετρήθηκε ήταν  $y_6 = 81$ kg, ενώ το αναμενόμενο βάρος του σύμφωνα με την ευθεία που φέραμε είναι 71kg, έχουμε δηλαδή ένα σφάλμα  $\varepsilon_6 = 81 - 71 = 10$ , δηλαδή βάρος 10kg περισσότερο από το αναμενόμενο. Ανάλογα σφάλματα υπολογίζονται και για τους άλλους μαθητές. Θα θέλαμε λοιπόν να βρούμε με κάποια μέθοδο εκείνη την ευθεία  $y = \alpha + \beta x$ , έτσι ώστε τα σφάλματα που προκύπτουν να είναι όσο το δυνατόν μικρότερα.

Η μέθοδος των ελαχίστων τετραγώνων συνίσταται στον προσδιορισμό των παραμέτρων  $\alpha, \beta$ , έτσι ώστε να ελαχιστοποιείται το άθροισμα των τετραγώνων των κατακόρυφων αποστάσεων των σημείων  $(x_i, y_i)$  από την ευθεία  $y = \alpha + \beta x$ , δηλαδή το

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \quad (3)$$

να γίνεται ελάχιστο.

Οι τιμές των παραμέτρων  $\alpha$  και  $\beta$ , που ελαχιστοποιούν την (3), καλούνται εκτιμήτριες ελαχίστων τετραγώνων, συμβολίζονται με  $\hat{\alpha}$  (“α καπέλο”) και  $\hat{\beta}$  (“β καπέλο”), αντιστοίχως και αποδεικνύεται (η απόδειξη εδώ παραλείπεται) ότι δίνονται από τις σχέσεις:

$$\hat{\beta} = \frac{n \sum_{i=1}^n x_i y_i - \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}$$
(4)

Όπου:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Η παρακάτω ευθεία, καλείται ευθεία ελαχίστων τετραγώνων ή ευθεία παλινδρόμησης της  $Y$  (πάνω) στη  $X$ :

$$\hat{y} = \hat{\alpha} + \hat{\beta} x$$
(5)

Αντικαθιστώντας το  $\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}$  στη σχέση (5) βρίσκουμε την

$$\hat{y} - \bar{y} = \hat{\beta} (x - \bar{x})$$

η οποία φανερώνει ότι η ευθεία ελαχίστων τετραγώνων  $\hat{y} = \hat{\alpha} + \hat{\beta} x$  διέρχεται από το σημείο με συντεταγμένες  $(\bar{x}, \bar{y})$  και έχει συντελεστή διεύθυνσης το  $\hat{\beta}$ . Αντικαθιστώντας τις τιμές  $x_i$  και  $y_i$  από τον πίνακα στις σχέσεις (4) βρίσκουμε:

$$\hat{\beta} = 1,28 \text{ και } \hat{\alpha} = -156,1$$

οπότε η ευθεία ελαχίστων τετραγώνων που προσαρμόζεται καλύτερα στα δεδομένα είναι από τη σχέση (5), η  $\hat{y} = -156,1 + 1,28x$  και παρατηρούμε ότι υπάρχει σημαντική διαφορά από την ευθεία  $y = -115,6 + 1,06x$ .

### Ερμηνεία και Ιδιότητες Εκτιμητών Ελαχίστων Τετραγώνων

Στην εξίσωση ελαχίστων τετραγώνων  $\hat{y} = \hat{\alpha} + \hat{\beta} x$  η τιμή της εκτιμήτριας  $\hat{\alpha}$  της παραμέτρου  $\alpha$  παριστάνει την τεταγμένη του σημείου στο οποίο η ευθεία τέμνει τον άξονα  $y'$ , δηλαδή την τιμή της εξαρτημένης μεταβλητής  $Y$  όταν  $x = 0$ . Όταν το

$\hat{\alpha} = 0$  τότε η ευθεία διέρχεται από την αρχή των αξόνων. Έστω τώρα δυο τιμές  $x_1$  και  $x_2 = x_1 + 1$  της ανεξάρτητης μεταβλητής. Τότε λαμβάνοντας τη διαφορά των αντίστοιχων προβλεπόμενων τιμών της εξαρτημένης μεταβλητής βρίσκουμε

$$\hat{y}_2 - \hat{y}_1 = (\hat{\alpha} + \hat{\beta}x_2) - (\hat{\alpha} + \hat{\beta}x_1) = \hat{\alpha} + \hat{\beta}(x_1 + 1) - (\hat{\alpha} + \hat{\beta}x_1) = \hat{\beta}$$

Δηλαδή  $\widehat{y}_2 = \widehat{y}_1 + \hat{\beta}$ . Συνεπώς ο συντελεστής διεύθυνσης  $\hat{\beta}$  της ευθείας  $\hat{y} = \hat{\alpha} + \hat{\beta}x$  παριστάνει τη μεταβολή της εξαρτημένης μεταβλητής  $Y$  όταν το  $X$  μεταβληθεί κατά μια μονάδα. Έτσι, όταν το  $x$  αυξηθεί κατά μια μονάδα τότε το  $\hat{y}$  αυξάνεται κατά  $\hat{\beta}$  μονάδες όταν  $\hat{\beta} > 0$  ή ελαττώνεται κατά  $\hat{\beta}$  μονάδες όταν  $\hat{\beta} < 0$ .

Σύμφωνα με το θεώρημα των Gauss-Markov, για το απλό γραμμικό μοντέλο, οι εκτιμήτριες ελαχίστων τετραγώνων  $\hat{\alpha}$  και  $\hat{\beta}$  είναι άριστες, γραμμικές συναρτήσεις των παρατηρήσεων της εξαρτημένης μεταβλητής  $Y$ , αμερόληπτες και μεταξύ όλων των γραμμικών αμερόληπτων εκτιμητών, έχουν την μικρότερη διακύμανση.

### Συντελεστής Προσδιορισμού

Η διασπορά της μεταβλητής  $Y$  εκφράζεται με τις αποκλίσεις  $y_i - \bar{y}$  των διαφόρων τιμών από τη μέση τιμή τους. Αν όλες οι τιμές ήταν ίσες μεταξύ τους δεν θα υπήρχε μεταβλητότητα στα δεδομένα και κάθε απόκλιση  $y_i - \bar{y}$  θα ήταν ίση με το μηδέν. Όσο μεγαλύτερες είναι οι αποκλίσεις  $(y_i - \bar{y})$ , τόσο μεγαλύτερη θα είναι και η διασπορά των δεδομένων. Η ολική μεταβλητότητα (διασπορά) των παρατηρήσεων εκφράζεται σαν το άθροισμα των τετραγώνων των αποκλίσεων  $(y_i - \bar{y})$  και συμβολίζεται με  $SST = \sum (y_i - \bar{y})^2$ , που ονομάζεται Ολικό άθροισμα τετραγώνων (Total Sum of Squares).

Ένα νέο μέτρο της μεταβλητότητας των  $y_i$  γύρω από την ευθεία παλινδρόμησης, είναι το άθροισμα τετραγώνων των σφαλμάτων (Error Sum of Squares)  $\sum e_i^2$  και συμβολίζεται με  $SSE = \sum (y_i - \hat{y}_i)^2$ . Επομένως η διαφορά των  $SST$  και  $SSE$  συμβολίζεται  $SSR = SST - SSE$ . Το  $SSR = \sum (\hat{y}_i - \bar{y})^2$  καλείται άθροισμα τετραγώνων παλινδρόμησης (Regression Sum of Squares) και εκφράζει την επίδραση της σχέσης παλινδρόμησης των δύο μεταβλητών στη μείωση της μεταβλητότητας των παρατηρήσεων  $y_i$ .

Γενικά, μπορούμε να αποδείξουμε ότι ισχύει η σχέση  $SST = SSR + SSE$ , δηλαδή η συνολική μεταβλητότητα των τιμών εκφράζεται ως το άθροισμα δύο όρων: της μεταβλητότητας που ερμηνεύεται από την παλινδρόμηση ( $SSR$ ) και της μεταβλητότητας που παραμένει ανερμήνευτη από τη μεταβλητή  $X$ , γνωστή ως υπόλοιπο ή σφάλμα ( $SSE$ ). Το άθροισμα τετραγώνων που οφείλεται σε σφάλματα ( $SSE$ ) μπορεί να ερμηνευθεί ως η ποσότητα μεταβλητότητας της  $Y$  που δεν εξηγείται από το γραμμικό μοντέλο. Ισχύει  $SSE \leq SST$ , με την ισότητα να ισχύει όταν δεν υπάρχουν σφάλματα, δηλαδή όταν το μοντέλο παλινδρόμησης περιγράφει με ακρίβεια την πληθυσμιακή εξίσωση παλινδρόμησης. Ο συντελεστής προσδιορισμού  $R^2$  αναφέρεται στην αναλογία της συνολικής διασποράς της εξαρτημένης μεταβλητής  $Y$ , που ερμηνεύεται από την παλινδρόμηση ( $SSR$ ), στη συνολική

διασπορά SST. Ο  $R^2$  εκφράζει το ποσοστό της εξήγησης της μεταβλητότητας της  $Y$  από την ανεξάρτητη μεταβλητή  $X$ . Συνήθως, κυμαίνεται μεταξύ 0 και 1, με υψηλότερες τιμές να υποδηλώνουν πιο αποτελεσματική εξήγηση. Δηλαδή:

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Η τιμή του συντελεστή προσδιορισμού  $R^2$  κυμαίνεται από 0 έως 1. Όταν το  $R^2$  πλησιάζει το 1, υπονοεί ότι το μοντέλο παλινδρόμησης έχει υψηλή δυνατότητα εξήγησης της εξαρτημένης μεταβλητής και τα σφάλματα είναι μικρά. Αντίθετα, χαμηλές τιμές του  $R^2$  υποδηλώνουν ότι το μοντέλο δεν είναι αποτελεσματικό στην ερμηνεία της εξαρτημένης μεταβλητής.

Είναι σημαντικό να σημειωθεί ότι ο συντελεστής προσδιορισμού παράγει αξιόπιστα αποτελέσματα όταν ο αριθμός των παρατηρήσεων είναι σημαντικά υψηλότερος από τον αριθμό των μεταβλητών. Μπορούμε επίσης να εξετάζουμε την ποσότητα  $1 - R^2$ , η οποία εκφράζει το ποσοστό της συνολικής μεταβλητότητας που οφείλεται σε σφάλματα. Αυτό μας επιτρέπει να εστιάσουμε στη μη εξηγημένη μεταβλητότητα. Παρόλο που ο  $R^2$  είναι ένα χρήσιμο μέτρο, πρέπει να είμαστε προσεκτικοί, καθώς είναι μια σημειακή εκτίμηση και τείνει να υπερεκτιμά τον πραγματικό συντελεστή προσδιορισμού, ειδικά όταν το δείγμα είναι μικρό.

Για να μετρήσουμε το βαθμό συσχέτισης μεταξύ δύο τυχαίων μεταβλητών, χρησιμοποιούμε τον συντελεστή συσχέτισης  $\rho$ . Αυτός ο συντελεστής ορίζεται ως το κλάσμα με αριθμητή τη συνδιακύμανση των μεταβλητών  $X$  και  $Y$  και παρονομαστή το γινόμενο των τυπικών αποκλίσεων των  $X$  και  $Y$ . Ο συντελεστής  $\rho$  αναδεικνύει τη στενή σχέση μεταξύ των δύο μεταβλητών και κυμαίνεται μεταξύ -1 και 1. Ο συντελεστής  $\rho$  είναι μια άγνωστη παράμετρος του πληθυσμού, αναφέρεται δηλαδή σε ένα χαρακτηριστικό της συνολικής κατανομής των μεταβλητών  $X$  και  $Y$ . Με άλλα λόγια, ο  $\rho$  περιγράφει τη στενή συσχέτιση μεταξύ των δύο μεταβλητών, με υψηλές τιμές να υποδηλώνουν ισχυρή συσχέτιση, ενώ οι χαμηλές τιμές υποδηλώνουν ασθενή συσχέτιση ή την απουσία της. Ως εκτιμητής του  $\rho$  χρησιμοποιείται ο συντελεστής συσχέτισης του δείγματος, που ορίζεται ως:

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2} \sqrt{\sum (Y - \bar{Y})^2}}$$

Ισχύει ότι  $r = \pm\sqrt{R^2}$ , ωστόσο υπάρχει μεγάλη διαφορά στην ερμηνεία τους.

Ο συντελεστής συσχέτισης του δείγματος  $r$  λειτουργεί ως εκτιμητής του συντελεστή συσχέτισης στον πληθυσμό  $\rho$ , ενώ είναι ανεξάρτητος από τις μονάδες μέτρησης των μεταβλητών  $X$  και  $Y$ . Λαμβάνει τιμές μεταξύ -1 και 1, όπου η -1 υποδεικνύει πλήρη αρνητική γραμμική συσχέτιση, το 1 πλήρη θετική γραμμική συσχέτιση και το 0 απουσία γραμμικής συσχέτισης. Η τιμή  $\rho$  καθορίζεται από τη συνδιακύμανση και τις τυπικές

αποκλίσεις των μεταβλητών. Ο συντελεστής συσχέτισης  $r$  μας δείχνει τη δύναμη και την κατεύθυνση της γραμμικής συσχέτισης, ενώ ο συντελεστής προσδιορισμού  $R^2$  αναλαμβάνει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής που ερμηνεύεται από τη μεταβλητή  $X$ . Επίσης, προσδιορίζοντας το  $1 - R^2$ , εξετάζουμε το ποσοστό της συνολικής μεταβλητότητας που παραμένει ανεξήγητη από τη μεταβλητή  $X$ . Παρόλο που ο  $r$  μας δίνει μια συγκεκριμένη μέτρηση της γραμμικής συσχέτισης, ο  $R^2$  μας παρέχει επιπλέον πληροφορίες για το πόσο καλά το γραμμικό μοντέλο προσαρμόζεται στα δεδομένα.

### **Τα Σφάλματα Εκτίμησης ή Κατάλοιπα**

Μετά την δημιουργία του γραφήματος με τις μεταβλητές, η Μέθοδος Ελαχίστων Τετραγώνων χρησιμοποιείται για να επιλεγεί μια κατάλληλη γραμμή που να περνά από τα σημεία του γραφήματος. Στην πραγματικότητα, οι σχέσεις μεταξύ των μεταβλητών δεν είναι πάντα ακριβείς λόγω τυχαίων ή μη παρατηρήσιμων διακυμάνσεων στα παρατηρηθέντα δεδομένα.

Ενώ η αυστηρή μαθηματική σχέση μεταξύ των μεταβλητών δεν πάντα επαληθεύεται λόγω αυτών των διακυμάνσεων, η Μέθοδος Ελαχίστων Τετραγώνων προσαρμόζει ένα μαθηματικό μοντέλο παλινδρόμησης στα δεδομένα. Για να ληφθούν υπόψη οι τυχαίες ή μη παρατηρήσιμες διακυμάνσεις, προστίθεται ένα στοχαστικό (τυχαίο) τμήμα στο μοντέλο παλινδρόμησης. Αν γίνει χρήση της  $X$  για την επεξήγηση της συμπεριφοράς της  $Y$ , οποιαδήποτε ευθεία γραμμή μπορεί να αποδοθεί με τη μορφή:

$$y_i = \beta_0 + \beta_i x_i + \varepsilon_i$$

Το  $\beta_0 + \beta_1 X$  αντιπροσωπεύει το συστηματικό τμήμα της εξίσωσης, ενώ το είναι το τυχαίο τμήμα, γνωστό και ως διαταρακτικός όρος ή σφάλμα. Τα σφάλματα παρουσιάζονται στα πειράματα λόγω λαθών κατά τη μέτρηση της εξαρτημένης μεταβλητής ή λόγω ελλιπούς προσδιορισμού του μοντέλου.

Η πρώτη αιτία είναι αντιληπτή, καθώς σφάλματα μπορεί να προκύψουν από ανακρίβειες στις μετρήσεις. Η δεύτερη αιτία είναι πιο περίπλοκη. Για παράδειγμα, η κατανάλωση ενός αγαθού εξαρτάται από πολλούς παράγοντες, όπως οι τιμές των συναγωνιστικών προϊόντων, το εισόδημα, τα επιτόκια, κ.λπ. Εάν το μοντέλο παλινδρόμησης προσδιοριστεί ως γραμμική συνάρτηση μόνο της τιμής ή του εισοδήματος, αυτό μπορεί να είναι ελλιπώς προσδιορισμένο. Συνεπώς, ο διαταρακτικός όρος  $\varepsilon_i$  προστίθεται για να λάβει υπόψη τις μη παρατηρήσιμες διακυμάνσεις ή τους ανεξερεύνητους παράγοντες που επηρεάζουν τα δεδομένα.

Για να εκτιμηθεί ένα υπόδειγμα, πρέπει να συλλεχθεί ένα δείγμα στοιχείων για την εξαρτημένη και την ανεξάρτητη μεταβλητή που μας ενδιαφέρει. Συγκεκριμένα, αν έχουμε ένα τυχαίο δείγμα παρατηρήσεων της μεταβλητής εξαρτημένης  $Y$  και της μεταβλητής ανεξάρτητης  $X$  (π.χ.,  $Y_1, Y_2, \dots, Y_n$  και  $X_1, X_2, \dots, X_n$ ), στόχος της ανάλυσης παλινδρόμησης είναι να εξαχθούν εκτιμήσεις για τις άγνωστες παραμέτρους  $\beta_0$  και  $\beta_1$  του πληθυσμού.

Ωστόσο, πρακτικά, δεν είναι δυνατό να προβλεφθούν εκ των προτέρων οι επιδράσεις στο τυχαίο τμήμα της παραπάνω εξίσωσης. Ο διαταρακτικός όρος, ή το "σφάλμα," αποτελεί το μη προβλέψιμο μέρος που περιλαμβάνει λάθη μέτρησης ή άλλες απρόβλεπτες επιδράσεις που δεν μπορούν να ληφθούν υπόψη κατά τη διαμόρφωση του μοντέλου παλινδρόμησης.

Για να είναι δυνατή η εκτίμηση ενός υποδείγματος παλινδρόμησης, απαιτείται η καθορισμός μιας κατανομής για τον διαταρακτικό όρο (ή "σφάλμα") και η υπόθεση των παρακάτω σημείων:

1. Για οποιαδήποτε τιμή της ανεξάρτητης μεταβλητής  $X$ , ο διαταρακτικός όρος  $\varepsilon_i$  είναι μια τυχαία μεταβλητή που κατανέμεται με μέσο τιμή 0 και διακύμανση  $\sigma^2$ . Δηλαδή,  $E(\varepsilon_i) = 0$  και  $Var(\varepsilon_i) = \sigma^2$  για κάθε  $i$ . Αυτό σημαίνει ότι ο διαταρακτικός όρος είναι μια τυχαία μεταβλητή που λαμβάνει θετικές και αρνητικές τιμές, με τον μέσο όρο της να είναι μηδέν.
2. Οι δειγματικές τιμές του  $\varepsilon_i$  είναι ανεξάρτητες, δηλαδή τα σφάλματα δεν συσχετίζονται μεταξύ τους. Αυτό σημαίνει ότι για δύο διαφορετικές παρατηρήσεις  $i$  και  $j$  με  $i \neq j$ , η αναμενόμενη τιμή  $E(\varepsilon_i, \varepsilon_j) = 0$  και η συνδιακύμανση τους  $Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i - E(\varepsilon_i)) \cdot (\varepsilon_j - E(\varepsilon_j)) = E(\varepsilon_i, \varepsilon_j) = 0$
3. Κάθε δειγματικό σφάλμα κατανέμεται κανονικά για κάθε  $i$ .
4. Οι τιμές της ανεξάρτητης μεταβλητής  $X$  θεωρούνται σταθερές και για μια συγκεκριμένη τιμή της  $X$  αντιστοιχεί μια ολόκληρη κατανομή της εξαρτημένης μεταβλητής  $Y$ . Συνεπώς, κάθε διαφοροποίηση της  $Y$  οφείλεται στους παράγοντες που συμπεριλαμβάνονται στον διαταρακτικό όρο.

Αυτές οι υποθέσεις επιβάλλουν την ορθή προσδιορισμό του υποδείγματος παλινδρόμησης όσον αφορά τη συναρτησιακή του μορφή και τις συμπεριληφθείσες μεταβλητές. Ο στόχος της ανάλυσης παλινδρόμησης είναι να εκτιμήσει τις παραμέτρους του υποδείγματος, έτσι ώστε η εξήγηση της μεταβολής της εξαρτημένης μεταβλητής  $Y$ , όπως ορίζεται από το υπόλοιπο (residual), να είναι μικρή και να μην εμφανίζει συστηματικά σφάλματα.

Η παραβίαση αυτών των υποθέσεων μπορεί να οδηγήσει σε προβλήματα όπως αυτοσυσχέτιση και ετεροσκεδαστικότητα. Τα παραπάνω προβλήματα αποτελούν σημαντικές πτυχές που απαιτούν ειδική ανάλυση, ιδίως όταν σκοπός είναι η σωστή μοντελοποίηση οικονομικών και κοινωνικών φαινομένων με σκοπό τις προβλέψεις. Οι προαναφερθείσες υποθέσεις σημαίνουν ότι τα σφάλματα είναι ανεξάρτητες μεταβλητές που κατανέμονται κανονικά ως  $N(0, \sigma^2)$ . Το άθροισμα των τετραγώνων των σφαλμάτων, όπως προαναφέρθηκε, συμβολίζεται συνήθως με SSE. Μια ισοδύναμη έκφραση του SSE που διευκολύνει σημαντικά τους υπολογισμούς είναι η:

$$SSE = \sum \hat{\varepsilon}_i^2 = \sum Y_i^2 - \hat{b}_0 \sum Y_i - \hat{b}_1 \sum X_i Y_i$$

Τα σφάλματα σε μια ανάλυση παλινδρόμησης έχουν ορισμένες βασικές ιδιότητες. Ας εξετάσουμε τις κύριες ιδιότητες αυτών των σφαλμάτων:

- **Ανεξαρτησία:** Τα σφάλματα θεωρούνται ανεξάρτητα μεταβλητές. Αυτό σημαίνει ότι η τιμή ενός σφάλματος δεν επηρεάζεται από τις τιμές άλλων σφαλμάτων. Η ανεξαρτησία είναι σημαντική για τη σωστή εκτίμηση των παραμέτρων.
- **Ετεροσκεδαστικότητα (Homoscedasticity):** Τα σφάλματα έχουν σταθερή διακύμανση σε όλα τα επίπεδα των ανεξάρτητων μεταβλητών. Αυτό σημαίνει ότι η διακύμανση των σφαλμάτων είναι σταθερή σε όλο το εύρος των τιμών των μεταβλητών.
- **Κανονική Κατανομή:** Συνήθως, υποθέτουμε ότι τα σφάλματα ακολουθούν κανονική κατανομή. Αυτό διευκολύνει τη στατιστική ανάλυση και προκύπτει από το Κεντρικό Όριο.
- **Μέσο Τιμή Μηδενικό:** Το άθροισμα των σφαλμάτων έχει μέσο τιμή μηδέν. Αυτή η ιδιότητα διασφαλίζει ότι η εκτιμητική συνάρτηση περνά από τα παρατηρηθέντα σημεία.

Οι παραπάνω προϋποθέσεις βοηθούν στη διασφάλιση της έγκυρης ερμηνείας των αποτελεσμάτων της ανάλυσης παλινδρόμησης. Ωστόσο, πρέπει να είμαστε προσεκτικοί και να ελέγχουμε αυτές τις προϋποθέσεις προτού εφαρμόσουμε τη μέθοδο των ελαχίστων τετραγώνων.

### **Εφαρμογή της Ανάλυσης της Παλινδρόμησης στον Επενδυτικό Κίνδυνο**

Η Γραμμική Παλινδρόμηση αποτελεί ισχυρό εργαλείο για τη δημιουργία τάσεων ή προβλέψεων μελλοντικών εξελίξεων, εισάγοντας στοιχεία από παρελθοντικά δεδομένα. Στο επιχειρησιακό περιβάλλον, η ανάλυση παλινδρόμησης χρησιμοποιείται κυρίως για να προβλέψει την εξέλιξη οικονομικών παραγόντων ή των χαρακτηριστικών προϊόντων με την πάροδο του χρόνου. Στον επενδυτικό τομέα, η ανάλυση παλινδρόμησης χρησιμοποιείται για να εξεταστούν, παραδείγματος χάριν, οι τιμές των μετοχών, του πετρελαίου ή των προδιαγραφών προϊόντων, προσφέροντας έτσι μια προοπτική σχετικού κινδύνου σε μια επένδυση. Επομένως, η εφαρμογή της γραμμικής παλινδρόμησης θεωρείται καίρια για την αξιολόγηση του κινδύνου που συνδέεται με πολλές επενδυτικές αποφάσεις.

Για αυτόν τον λόγο, η ανάλυση της γραμμικής παλινδρόμησης ανέπτυξε το μοντέλο Capital Asset Pricing, το οποίο χρησιμοποιείται για τον υπολογισμό ενός κοινού μέτρου μεταβλητότητας για μια μετοχή ή επένδυση, γνωστό ως beta. Το beta καθορίζεται μέσω της γραμμικής παλινδρόμησης και περιγράφει τη σχέση της απόδοσης μιας μετοχής με αυτήν της χρηματοπιστωτικής αγοράς στο σύνολό της.

Ένα θετικό beta υποδεικνύει ότι οι αποδόσεις των περιουσιακών στοιχείων ακολουθούν γενικά τις αποδόσεις της αγοράς, είτε και οι δύο τείνουν να είναι ανώτερες είτε κάτω από τους αντίστοιχους μέσους όρους τους. Αντίθετα, ένα αρνητικό beta υπονοεί αντικρουόμενες κινήσεις, όπου οι αποδόσεις των περιουσιακών στοιχείων έχουν αντίθετη κίνηση από τις αποδόσεις της αγοράς.

Ο συντελεστής beta αποτελεί μια κρίσιμη παράμετρο στο πλαίσιο του Capital Asset Pricing Model (CAPM), το οποίο αναλύει την αναμενόμενη απόδοση ενός περιουσιακού στοιχείου. Αυτός ο συντελεστής δεν είναι περιττός καθώς δεν μπορεί να αναιρεθεί από τη

διαφοροποίηση που παρέχεται από τη συνύπαρξη πολλών υψηλού κινδύνου περιουσιακών στοιχείων σε ένα χαρτοφυλάκιο. Αυτό συμβαίνει εξαιτίας της σχέσης των αποδόσεων του περιουσιακού στοιχείου με αυτές των υπόλοιπων στοιχείων που περιλαμβάνονται στο χαρτοφυλάκιο.

Ο συντελεστής beta μπορεί να υπολογιστεί για μεμονωμένες εταιρείες μέσω της ανάλυσης παλινδρόμησης, λαμβάνοντας υπόψη τη σύγκριση των αποδόσεων τους έναντι ενός χρηματιστηριακού δείκτη. Ο τύπος για την beta ενός περιουσιακού στοιχείου σε ένα χαρτοφυλάκιο είναι:

$$\beta_a = \frac{\text{Cov}(r_a, r_p)}{\text{Var}(r_p)}$$

όπου το  $r_a$  μετρά το ποσοστό απόδοσης του περιουσιακού στοιχείου,  $r_p$  μετρά το ποσοστό απόδοσης του χαρτοφυλακίου και  $\text{Cov}(r_a, r_p)$  είναι η συνδιακύμανση μεταξύ των ποσοστών της επιστροφής. Το χαρτοφυλάκιο του ενδιαφέροντος στις διατυπώσεις του CAMP είναι το χαρτοφυλάκιο της αγοράς που περιλαμβάνει όλα τα επικίνδυνα στοιχεία ενεργητικών μετοχών και έτσι οι όροι  $r_p$  του άνωθεν τύπου αντικαθίστανται από  $r_m$  (δηλαδή το ποσοστό απόδοσης της αγοράς).

Αναφορικά με τον συντελεστή beta στο πλαίσιο της επενδυτικής αγοράς, πρέπει να σημειωθεί ότι κατά προσδιορισμό ο συντελεστής beta της αγοράς ορίζεται ως 1.0. Οι μεμονωμένες μετοχές ταξινομούνται ανάλογα με το πόσο αποκλίνουν από τη μακροοικονομική αγορά, συχνά χρησιμοποιώντας τον δείκτη S&P 500 ως αντιπρόσωπο της αγοράς. Η αξιολόγηση γίνεται με βάση την απόδοση τους σε σχέση με τον μέσο όρο της απόδοσης της αγοράς στη διάρκεια του χρόνου.

Μια μετοχή με απόδοση υψηλότερη από το μέσο όρο της αγοράς θα έχει έναν συντελεστή beta με απόλυτη τιμή μεγαλύτερη από 1.0. Αντίστοιχα, μια μετοχή με χαμηλότερη απόδοση θα έχει έναν συντελεστή beta με απόλυτη τιμή μικρότερη από 1.0. Ο συντελεστής beta μετρά το πόσο συχνά οι αποδόσεις μιας μετοχής αλλάζουν σε σύγκριση με τις αποδόσεις της αγοράς. Συγκεκριμένα, ένας συντελεστής beta 2 σημαίνει ότι η απόδοση της μετοχής τείνει να αυξάνεται ή να μειώνεται περίπου διπλάσια συχνότερα από τον μέσο όρο της αγοράς.

Επιπλέον, ο συντελεστής beta μπορεί να είναι και αρνητικός, υποδηλώνοντας ότι οι αποδόσεις της μετοχής κινούνται αντίθετα προς τις αποδόσεις της αγοράς. Για παράδειγμα, ένας συντελεστής beta -3 υποδεικνύει ότι, κατά μέσο όρο, η μετοχή θα έχει μείωση των αποδόσεων της κατά 9% όταν η αγορά αυξάνεται κατά 3% και αντιστρόφως, θα έχει αύξηση των αποδόσεων της κατά 9% όταν η αγορά μειώνεται κατά 3%.

Είναι κοινά αποδεκτό ότι, παρότι το γεγονός ότι μετοχές υψηλού beta μπορεί να προσφέρουν υψηλότερες αποδόσεις, τείνουν να είναι πιο ασταθείς και συνεπώς, πιο επιρρεπείς σε ρίσκο. Αντίστοιχα, οι μετοχές με χαμηλότερους συντελεστές beta θεωρούνται λιγότερο επικίνδυνες, αλλά συνήθως προσφέρουν χαμηλότερες αποδόσεις. Παρότι ταύτα, η ιδέα αυτή έχει αμφισβητηθεί, καθώς υπάρχουν επιχειρήματα που υποστηρίζουν ότι δεν υπάρχει

συγκεκριμένη σχέση μεταξύ του συντελεστή beta και των πιθανών αποδόσεων. Συχνά, οι μετοχές με χαμηλότερους συντελεστές beta φαίνεται να είναι λιγότερο επικίνδυνες και πιο επικερδείς.

Τέλος, ο συντελεστής beta δεν αποτελεί μόνο ένδειξη για τη σχέση μιας μετοχής με τις αλλαγές στην αγορά, αλλά είναι επίσης ένα εργαλείο για την εκτίμηση της απαιτούμενης απόδοσης των επενδύσεων (ROI). Για παράδειγμα, λαμβάνοντας υπόψη ένα επιτόκιο μηδενικού κινδύνου ύψους 2%, μια μετοχή με συντελεστή beta 1,5 θα πρέπει να έχει απόδοση 11% ( $= 2\% + 1,5 * (8\% - 2\%)$ ), ενώ η αγορά με beta 1 αναμένεται να έχει απόδοση 8%.

#### 4.1.2 Πολλαπλή Γραμμική Παλινδρόμηση

Σε πολλές περιπτώσεις, είναι αναγκαίο να χρησιμοποιήσουμε δύο ή περισσότερες ανεξάρτητες μεταβλητές προκειμένου να αναλύσουμε εκτενέστερα ένα φυσικό φαινόμενο και να προβλέψουμε με μεγαλύτερη ακρίβεια. Για παράδειγμα, στο πλαίσιο της πρόβλεψης της ζήτησης ενός προϊόντος σε διάφορες πόλεις, είναι σκόπιμο να συμπεριλάβουμε διάφορες κατηγορίες μεταβλητών, όπως κοινωνικοοικονομικές (μέσο οικογενειακό εισόδημα, μόρφωση), δημογραφικές (αριθμός μελών οικογένειας, αριθμός συνταξιούχων) και περιβαλλοντολογικές (μέση ημερήσια θερμοκρασία) κ.ά. (Montgomery et.al., 2012).

Όταν αντιμετωπίζουμε περισσότερες από μία ανεξάρτητες μεταβλητές, χρησιμοποιούμε το μοντέλο πολλαπλής παλινδρόμησης για να εξετάσουμε τη συμπεριφορά της εξαρτημένης μεταβλητής  $Y$ . Ειδικότερα, όταν η σχέση μεταξύ της εξαρτημένης μεταβλητής είναι γραμμική συνάρτηση των ανεξάρτητων μεταβλητών, τότε το μοντέλο περιγράφεται ως πολλαπλή γραμμική παλινδρόμηση.

Η πολλαπλή παλινδρόμηση αποτελεί ευρέως αποδεκτή επιστημονική πρακτική, καθώς θεωρείται ισχυρό και ευέλικτο στατιστικό εργαλείο που εφαρμόζεται σε διάφορα ερευνητικά πεδία (Kutner et. al., 2004). Υπάρχουν πολλές εφαρμογές αυτού του μοντέλου, μερικές εκ των οποίων περιλαμβάνουν:

- Διοίκηση επιχειρήσεων και έρευνα αγοράς: Χρησιμοποιείται για την εκτίμηση της απόδοσης του προσωπικού μιας εταιρείας και τη διαχείριση των παραπόνων των πελατών.
- Προβλήματα οδικής συγκοινωνίας: Εφαρμόζεται για τη διαχείριση του είδους του οδοστρώματος και των μέσων μεταφοράς προκειμένου να βελτιωθεί ο χρόνος μετακίνησης.
- Υπέρβαση στον αθλητισμό: Χρησιμοποιείται για τη βελτίωση των αθλητικών επιδόσεων και την προσαρμογή της διατροφής.
- Ατμοσφαιρική και υδροβία ρύπανση: Συμβάλλει στη διαφύλαξη της δημόσιας υγείας με την ανάλυση της ρύπανσης.

- Ερευνητική εξέταση της νοημοσύνης: Εξετάζει τη συμπεριφορά του δείκτη νοημοσύνης σε διαγωνιστικό επίπεδο.
- Εκτίμηση της δράσης χημικών συστατικών σε τρόφιμα: Χρησιμοποιείται για να αναλύσει την επίδρασή τους στις οργανοληπτικές ιδιότητες τους.

Συνολικά, η ανάλυση παλινδρόμησης χρησιμοποιείται για να εξηγήσει τις ειδικές σχέσεις μεταξύ των μεταβλητών, να επιβεβαιώσει θεωρητικές υποθέσεις, να προβλέψει αποτελέσματα πειραματικών δεδομένων και να αναπτύσσει και να επαληθεύει εξισώσεις πολλαπλής παλινδρόμησης.

### **Μελέτη του Μοντέλου της Πολλαπλής Γραμμικής Παλινδρόμησης**

Όπως προαναφέρθηκε, όταν μιλάμε για μοντέλα παλινδρόμησης που περιλαμβάνουν δύο ή περισσότερες ανεξάρτητες μεταβλητές, αναφερόμαστε σε μοντέλα πολλαπλής παλινδρόμησης. Για να το απλοποιήσουμε, ας εξετάσουμε αρχικά ένα μοντέλο με δύο ανεξάρτητες μεταβλητές, το οποίο αποτελεί μια επέκταση της απλής γραμμικής παλινδρόμησης. Σε αυτό το σενάριο, ερευνούμε τις σχέσεις μεταξύ δύο ανεξάρτητων μεταβλητών, ονομαστικά  $X_1$  και  $X_2$ . Έτσι θα έχουμε:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \quad i = 1, 2, \dots, n$$

όπου:

- $Y_i$  είναι η τιμή της εξαρτημένης μεταβλητής στην  $i$  παρατήρηση.
- $x_{i1}, x_{i2}$  είναι οι τιμές των ανεξαρτήτων μεταβλητών  $X_1$  και  $X_2$  στην  $i$  παρατήρηση, οι οποίες υποτίθεται ότι είναι γνωστές.
- $\beta_0, \beta_1$  και  $\beta_2$  είναι οι παράμετροι του μοντέλου.
- $\varepsilon_i$  είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κατανομή  $N(0, \sigma^2)$ .

Επομένως, η συνάρτηση παλινδρόμησης (regression function) ή αλλιώς συνάρτηση ανταπόκρισης (response function) του παραπάνω μοντέλου είναι

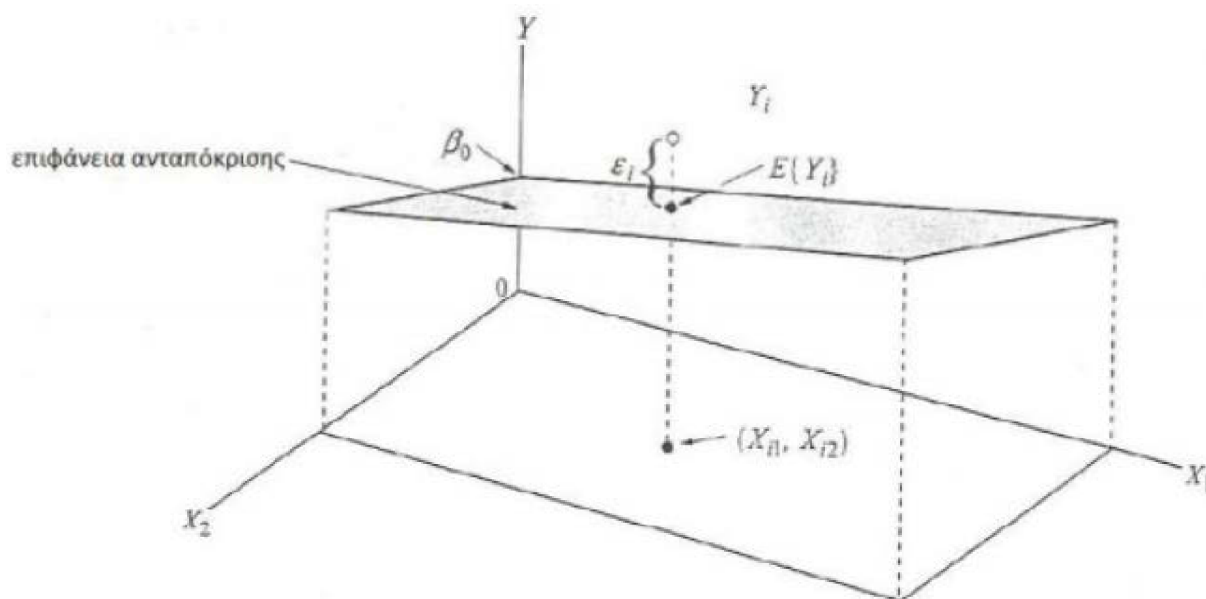
$$E(Y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Σε αυτό το σημείο αξίζει να σημειωθεί ότι η συνάρτηση αυτή ονομάζεται αριστερές φορές και επιφάνεια παλινδρόμησης (regression surface) ή επιφάνεια ανταπόκρισης (response surface).

Όπως και στην απλή γραμμική παλινδρόμηση, έτσι και στην πολλαπλή, οι παράμετροι έχουν ανάλογες ερμηνείες. Έτσι, στην επιφάνεια παλινδρόμησης:

- Το  $\beta_0$  αντιστοιχεί στο σημείο τομής του άξονα του  $Y$  από την επιφάνεια (επίπεδο) παλινδρόμησης.
- Το  $\beta_1$  δείχνει την μεταβολή της  $E(Y)$  όταν το  $x_1$  μεταβάλλεται κατά μια μονάδα ενώ το  $x_2$  παραμένει σταθερό.
- Αντίστοιχα, το  $\beta_2$  δείχνει την μεταβολή της  $E(Y)$  όταν το  $x_2$  μεταβάλλεται κατά μία μονάδα ενώ το  $x_1$  παραμένει σταθερό.

Μια απεικόνισή της με την χρήση δύο ανεξάρτητων μεταβλητών, θα μπορούσε να είναι η παρακάτω:



Αντίστοιχο του μοντέλου με δύο ανεξάρτητες μεταβλητές, είναι το μοντέλο παλινδρόμησης με  $p$  ανεξάρτητες μεταβλητές, το οποίο θα έχει τη μορφή:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{i,p-1} + \varepsilon_i \quad i = 1, \dots, n$$

όπου:

- $Y_i$  είναι η τιμή της εξαρτημένης μεταβλητής στην  $i$  παρατήρηση.
- $x_{i1}, x_{i2}, \dots, x_{i,p-1}$  είναι οι τιμές των ανεξαρτήτων μεταβλητών στην  $i$  παρατήρηση, οι οποίες υποτίθεται ότι είναι γνωστές.
- $\beta_i$  αντιπροσωπεύει την μεταβολή στην  $Y$  που προέρχεται από μια μεταβολή στην  $X_i$  κατά μία μονάδα, όταν όλες οι άλλες ανεξάρτητες μεταβλητές παραμένουν σταθερές.
- $\varepsilon_i$  είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κατανομή  $N(0, \sigma^2)$ .

Η ύπαρξη υπολοίπων  $\varepsilon_i$ , όπως και στην απλή γραμμική παλινδρόμηση, είναι αναγκαία επειδή κανένα μοντέλο δεν μπορεί πραγματικά να αποτυπώσει το σύνολο των πληροφοριών που περιέχει ένα σύνολο δεδομένων. Ακόμη και όταν η γραμμή πολλαπλής παλινδρόμησης προσαρμόζεται εξαιρετικά στα δεδομένα, πάντα θα υπάρχει ένα μέρος των πληροφοριών που δεν μπορεί να εξηγηθεί από το μοντέλο. Αυτό το μη εξηγήσιμο μέρος ονομάζεται λάθος της παλινδρόμησης.

Οπότε, η συνάρτηση παλινδρόμησης ή συνάρτηση ανταπόκρισης (η οποία μερικές φορές ονομάζεται και επιφάνεια παλινδρόμησης ή επιφάνεια ανταπόκρισης) είναι η:

$$E(Y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{i,p-1}$$

για  $i=1, \dots, n$ .

### Ανάλυση Διασποράς

Όπως και στο απλό γραμμικό μοντέλο, η δειγματική διασπορά των παρατηρήσεων  $Y_i$  χωρίζεται σε δύο αθροίσματα:

$$\sum_{i=1}^n \left( Y_i - \bar{Y} \right)^2 = \sum_{i=1}^n \left( \hat{Y}_i - \bar{Y} \right)^2 + \sum_{i=1}^n \left( Y_i - \hat{Y}_i \right)^2$$

τα οποία συμβολίζονται και πάλι με SST, SSR και SSE αντίστοιχα. Το SST εκφράζει τη συνολική παρατηρούμενη μεταβλητότητα των  $Y_i$ , το SSR εκφράζει τη μεταβλητότητα των προσαρμοσμένων τιμών, ενώ το SSE εκφράζει τη μεταβλητότητα των  $Y_i$  σε σχέση με τις αντίστοιχες προσαρμοσμένες τιμές.

Ο πίνακας ανάλυσης διασποράς ANOVA είναι:

Πηγή Μεταβλητότητας	Αθροίσματα Τετραγώνων	Βαθμοί Ελευθερίας	Μέσο Αθροισμα Τετραγώνων	Έλεγχος F
Παλινδρόμηση	$SSR = \sum_{i=1}^n \left( \hat{Y}_i - \bar{Y} \right)^2$	p-1	$MSR = \frac{SSR}{p-1}$	$F = \frac{MSR}{MSE}$
Υπόλοιπα	$SSE = \sum_{i=1}^n \left( Y_i - \hat{Y}_i \right)^2$	n-p	$MSE = \frac{SSE}{n-p}$	
Σύνολο	$SST = \sum_{i=1}^n \left( Y_i - \bar{Y} \right)^2$	n-1		

Σημαντική διαφορά από τον πίνακα ανάλυσης διασποράς στην απλή παλινδρόμηση, είναι οι βαθμοί ελευθερίας. Η ποσότητα SST εξακολουθεί να έχει n-1 βαθμούς ελευθερίας, ενώ η ποσότητα SSE έχει πλέον n-p βαθμούς ελευθερίας λόγω του ότι γίνεται εκτίμηση p μερικών συντελεστών παλινδρόμησης. Επίσης, η ποσότητα SSR έχει p-1 βαθμούς ελευθερίας που αντιπροσωπεύουν το πλήθος των μεταβλητών  $X_1, X_2, \dots, X_p - 1$ .

Στην πολλαπλή παλινδρόμηση το άθροισμα των τετραγώνων των υπολοίπων μπορεί να εκφραστεί ως:

$$\begin{aligned}
SSE &= \sum_{i=1}^n \left( Y_i - \hat{Y}_i \right)^2 = \sum_{i=1}^n e_i^2 = \underline{e}' \underline{e} \\
&= \left( \underline{Y} - \underline{X} \underline{b} \right)' \left( \underline{Y} - \underline{X} \underline{b} \right) = \left( \underline{Y} - \underline{H} \underline{Y} \right)' \left( \underline{Y} - \underline{H} \underline{Y} \right) \\
&= \underline{Y}' \underline{Y} - \underline{Y}' \underline{H} \underline{Y} - \underline{Y}' \underline{H}' \underline{Y} + \underline{Y}' \underline{H}' \underline{H} \underline{Y} \\
&= \underline{Y}' \underline{Y} - \underline{Y}' \underline{H} \underline{Y} = \underline{Y}' \left( \underline{I}_n - \underline{H} \right) \underline{Y}.
\end{aligned}$$

Ανάλογα:

$$\begin{aligned}
SSR &= \sum_{i=1}^n \left( \hat{Y}_i - \bar{Y} \right)^2 = \sum_{i=1}^n \hat{Y}_i^2 - 2\bar{Y} \sum_{i=1}^n \hat{Y}_i + n\bar{Y}^2 \\
&= \sum_{i=1}^n \hat{Y}_i^2 - n\bar{Y}^2 = \underline{Y}' \underline{\hat{Y}} - \frac{1}{n} \underline{Y}' \underline{J}_n \underline{Y} = \left( \underline{H} \underline{Y} \right)' \left( \underline{H} \underline{Y} \right) - \frac{1}{n} \underline{Y}' \underline{J}_n \underline{Y} \\
&= \underline{Y}' \underline{H} \underline{Y} - \frac{1}{n} \underline{Y}' \underline{J}_n \underline{Y} = \underline{Y}' \left( \underline{H} - \frac{1}{n} \underline{J}_n \right) \underline{Y}
\end{aligned}$$

όπου  $\underline{J}_n = 1 \cdot 1^T$  και  $\underline{1}$  ο  $n \times 1$  πίνακας ή πίνακας στήλη με στοιχεία του μονάδες. Επειδή  $SST = SSR + SSE$ , έχουμε πως:

$$SST = \underline{Y}' \left( \underline{H} - \frac{1}{n} \underline{J}_n \right) \underline{Y} + \underline{Y}' \left( \underline{I}_n - \underline{H} \right) \underline{Y} = \underline{Y}' \left( \underline{I}_n - \frac{1}{n} \underline{J}_n \right) \underline{Y}$$

Αυτές οι εκφράσεις αποτελούν τις τετραγωνικές μορφές.

Όσον αφορά την ποσότητα MSE, όπως και στην απλή παλινδρόμηση, αποτελεί αμερόληπτο εκτιμητή της διακύμανσης  $\sigma^2$ .

Στην πολλαπλή παλινδρόμηση, προστίθενται επιπλέον όροι που αντιπροσωπεύουν τα άθροισμα των τετραγώνων. Αυτοί οι όροι αποτυπώνουν την περιθώρια αύξηση στα σύνολα τετραγώνων της παλινδρόμησης όταν ενσωματώνονται μία ή περισσότερες επιπλέον μεταβλητές στο μοντέλο παλινδρόμησης. Η περιθώρια αύξηση προσθέτοντας την  $X_2$  σε ένα μοντέλο που ήδη έχει την  $X_1$ , θα συμβολίζεται ως:

$$SSR(X_2 \setminus X_1) = SSR(X_1, X_2) - SSR(X_1)$$

το οποίο είναι ισοδύναμο με την έκφραση:

$$SSR(X_2 \setminus X_1) = SSE(X_1) - SSE(X_1, X_2)$$

Όταν στο μοντέλο υπάρχουν k μεταβλητές X, τότε θα υπάρχουν k αποσυνθέσεις των μεταβλητών X.

### Μέθοδος Εκτίμησης Παραμέτρων: Μέθοδος Ελάχιστων Τετραγώνων

Στην πολλαπλή παλινδρόμηση προσδιορίζονται περισσότερες παράμετροι με τρόπο ανάλογο όπως και για την απλή παλινδρόμηση. Όπως και εκεί, η γραμμή παλινδρομήσεως στον πληθυσμό είναι άγνωστη, εφόσον είναι άγνωστες οι τιμές των παραμέτρων  $\beta_i$  για  $i=0, 1, \dots, p-1$ . Για τον λόγο αυτό, θα πρέπει να εκτιμήσουμε τις τιμές των συντελεστών από ένα δείγμα παρατηρήσεων για τις μεταβλητές  $Y_i$  και  $X_{ij}$  (όπου η  $X_{ij}$  είναι η  $i$  (για  $i=1, \dots, n$ ) παρατήρηση της  $j$  (για  $j=1, \dots, p-1$ ) ανεξάρτητης μεταβλητής). Με αυτό τον τρόπο κάνουμε μια εκτίμηση του πληθυσμιακού επιπέδου παλινδρόμησης από το δειγματικό

$$SSR(X_2 \setminus X_1) = SSE(X_1) - SSE(X_1, X_2)$$

για  $i=1, \dots, n$  με τα κατάλοιπα να ορίζονται ως το διάνυσμα  $e = Y - \hat{Y} = Y - Xb$ . Η μέθοδος ελαχίστων τετραγώνων για την εκτίμηση των παραμέτρων  $\beta_i$  για  $i=1, \dots, n$ , βασίζεται όπως και στο απλό γραμμικό μοντέλο στην ελαχιστοποίηση της παράστασης  $\sum_{i=1}^n e_i^2$ . Έχουμε λοιπόν ότι:

$$\begin{aligned} \sum_{i=1}^n e_i^2 &= \underline{\varepsilon}' \underline{\varepsilon} = (\underline{Y} - \underline{X} \underline{\beta})' (\underline{Y} - \underline{X} \underline{\beta}) \\ &= (\underline{Y}' - \underline{\beta}' \underline{X}') (\underline{Y} - \underline{X} \underline{\beta}) \\ &= \underline{Y}' \underline{Y} - \underline{Y}' \underline{X} \underline{\beta} - \underline{\beta}' \underline{X}' \underline{Y} + \underline{\beta}' \underline{X}' \underline{X} \underline{\beta} \\ &= \underline{Y}' \underline{Y} - 2 \underline{\beta}' \underline{X}' \underline{Y} + \underline{\beta}' \underline{X}' \underline{X} \underline{\beta}. \end{aligned}$$

Ορίζουμε τις μερικές παραγώγους της σχέσης αυτής ως προς το διάνυσμα  $\beta$ , οπότε:

$$\frac{\partial \left( \sum_{i=1}^n e_i^2 \right)}{\partial \underline{\beta}} = \frac{\partial \left( \underline{Y}' \underline{Y} - 2 \underline{\beta}' \underline{X}' \underline{Y} + \underline{\beta}' \underline{X}' \underline{X} \underline{\beta} \right)}{\partial \underline{\beta}} = -2 \underline{X}' \underline{Y} + 2 \underline{X}' \underline{X} \underline{\beta}$$

Θέτοντας τις μερικές παραγώγους ίσες με το μηδέν, οι κανονικές εξισώσεις ισορροπίας για το μοντέλο πολλαπλής γραμμικής παλινδρόμησης είναι οι:  $X'Xb=X'Y$

Σημειώνουμε ότι οι παράμετροι του διανύσματος  $\beta$  αντικαθίστανται από τις παραμέτρους του διανύσματος  $b$ , καθώς οι δεύτερες αποτελούν εκτιμήτριες των άλλοτε κανονικών εξισώσεων που ορίσαμε θέτοντας τις μερικές παραγώγους ίσες με μηδέν. Δηλαδή προκύπτει ότι οι εκτιμητές των συντελεστών  $\beta$  για το πολλαπλό γραμμικό μοντέλο παλινδρόμησης είναι οι:  $b=(X'X)^{-1} X'Y$ .

Οι κανονικές εξισώσεις  $X'Xb=X'Y$  στην αλγεβρική τους μορφή είναι:

$$\begin{pmatrix} 1 & 1 \dots & 1 \\ X_{11} & X_{21} \dots & X_{n1} \\ \vdots & \vdots & \vdots \\ X_{1,p-1} & X_{2,p-1} \dots & X_{n,p-1} \end{pmatrix} \begin{pmatrix} 1 & X_{11} & X_{1,p-1} \\ 1 & X_{21} & X_{2,p-1} \\ \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n,p-1} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \dots & 1 \\ X_{11} & X_{21} \dots & X_{n1} \\ \vdots & \vdots & \vdots \\ X_{1,p-1} & X_{2,p-1} \dots & X_{n,p-1} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} n & \sum_{i=1}^n X_{i1} \dots & \sum_{i=1}^n X_{i,p-1} \\ \sum_{i=1}^n X_{i1} & \sum_{i=1}^n X_{i1}^2 \dots & \sum_{i=1}^n X_{i1} X_{i,p-1} \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^n X_{i,p-1} & \sum_{i=1}^n X_{i,p-1} X_{i1} & \sum_{i=1}^n X_{i,p-1}^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_{i1} Y_i \\ \vdots \\ \sum_{i=1}^n X_{i,p-1} Y_i \end{pmatrix}$$

$$\Rightarrow \left\{ \begin{array}{l} \sum_{i=1}^n Y_i = nb_0 + b_1 \sum_{i=1}^n X_{i1} + \dots + b_{p-1} \sum_{i=1}^n X_{i,p-1} \\ \sum_{i=1}^n X_{i1} Y_i = b_0 \sum_{i=1}^n X_{i1} + b_1 \sum_{i=1}^n X_{i1}^2 + \dots + b_{p-1} \sum_{i=1}^n X_{i1} X_{i,p-1} \\ \vdots \\ \sum_{i=1}^n X_{i,p-1} Y_i = b_0 \sum_{i=1}^n X_{i,p-1} + b_1 \sum_{i=1}^n X_{i,p-1} X_{i1} + \dots + b_{p-1} \sum_{i=1}^n X_{i,p-1}^2 \end{array} \right\}$$

Το διάνυσμα των εκτιμητών των παραμέτρων για να υπολογιστεί, χρειάζεται τον πίνακα  $(X'X)^{-1}$ . Για να αντιστρέφεται όμως ένας πίνακας, πρέπει να δειχθεί ότι η ορίζουσά του είναι διαφορετική του μηδενός.

### Ιδιότητες των Εκτιμητών

Το θεώρημα Gauss-Markov, που επικεντρώνεται στην αποτελεσματικότητα του εκτιμητή ελαχίστων τετραγώνων στο γραμμικό μοντέλο παλινδρόμησης, δηλώνει ότι υπό συγκεκριμένες προϋποθέσεις, αυτός ο εκτιμητής είναι ανεκτίμητος και αποτελεί τον βέλτιστο

γραμμικό εκτιμητή για τους συντελεστές του γραμμικού μοντέλου παλινδρόμησης. Συνεπώς, σύμφωνα με το θεώρημα Gauss-Markov στο πλαίσιο του κλασικού γραμμικού προτύπου, οι εκτιμητές των συντελεστών  $\beta$  είναι γραμμικοί, αμερόληπτοι και προσφέρουν τη βέλτιστη δυνατή ακρίβεια στην εκτίμηση του μοντέλου γραμμικής παλινδρόμησης.

### Συντελεστής Προσδιορισμού

Ο συντελεστής προσδιορισμού είναι η αναλογία της συνολικής διασποράς που εξηγείται από το μοντέλο παλινδρόμησης και αναφέρεται επίσης ως  $R^2$ . Αυτός ο συντελεστής μετρά το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής που μπορεί να εξηγηθεί από τη γραμμή παλινδρόμησης και είναι ουσιαστικά το τετράγωνο του συντελεστή συσχέτισης του Pearson ( $r$ ). Το  $R^2$  αποτελεί ένα μέτρο για τον βαθμό προσαρμογής του μοντέλου παλινδρόμησης στα παρατηρούμενα δεδομένα. Με άλλα λόγια, μετράει πόσο καλά η εξίσωση παλινδρόμησης προσαρμόζεται στις παρατηρήσεις του δείγματος και παρέχει μια ενδεικτική εκτίμηση για την ερμηνευτική ικανότητα του μοντέλου και υπολογίζεται από τον τύπο:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Ο συντελεστής πολλαπλού προσδιορισμού  $R^2$  αποτελεί ένα βασικό μέτρο για την αξιολόγηση της προσαρμογής ενός μοντέλου παλινδρόμησης στα δεδομένα. Παρόλα αυτά, ένα σημαντικό του μειονέκτημα είναι ότι η τιμή του αυξάνεται σχεδόν πάντοτε με την προσθήκη νέων ανεξάρτητων μεταβλητών, ανεξαρτήτως της πραγματικής συμβολής τους στο μοντέλο. Αυτό συμβαίνει διότι, με κάθε νέα μεταβλητή, το υπόδειγμα αποκτά περισσότερη ευελιξία να προσαρμοστεί στα δεδομένα, μειώνοντας το άθροισμα τετραγώνων των υπολοίπων (SSE), ενώ η συνολική μεταβλητότητα (SST) της εξαρτημένης μεταβλητής παραμένει αμετάβλητη.

Ωστόσο, η συνεχής αύξηση των ανεξάρτητων μεταβλητών μπορεί να οδηγήσει σε υπερπροσαρμογή (overfitting) και απώλεια βαθμών ελευθερίας, χωρίς ουσιαστική βελτίωση της προβλεπτικής ικανότητας του μοντέλου. Για να αντιμετωπιστεί αυτό το πρόβλημα, χρησιμοποιείται ο διορθωμένος συντελεστής προσδιορισμού  $R_a^2$  (adjusted R-squared), ο οποίος τροποποιεί την τιμή του  $R^2$  λαμβάνοντας υπόψη το πλήθος των παραμέτρων του μοντέλου και τους αντίστοιχους βαθμούς ελευθερίας. Έτσι, το  $R_a^2$  αυξάνεται μόνο όταν η νέα μεταβλητή που προστίθεται στο μοντέλο συμβάλλει ουσιαστικά στη μείωση του σφάλματος πρόβλεψης και όχι απλώς λόγω του πλήθους των μεταβλητών.

Ο τύπος για τον υπολογισμό του διορθωμένου συντελεστή προσδιορισμού είναι:

$$R_a^2 = 1 - \left( \frac{SSE/(n - k - 1)}{SST/(n - 1)} \right)$$

Έτσι, στο πολλαπλό μοντέλο παλινδρόμησης με  $i$  μερικούς συντελεστές παλινδρόμησης, ο μερικός συντελεστής προσδιορισμού ανάμεσα στην  $Y$  και

την  $X_1$ , υπό τον περιορισμό ότι οι υπόλοιπες  $X_i$  παραμένουν σταθερές, είναι:

$$r_{Y,1}^2 = \frac{\beta_1^2 \sum_{i=1}^n (X_{i1} - \bar{X}_1)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

Σε ένα μοντέλο με δύο ανεξάρτητες μεταβλητές  $X$ , μπορούμε να ορίσουμε τον μερικό συντελεστή προσδιορισμού μεταξύ της  $Y$  και της  $X_1$  ενώ η  $X_2$  είναι ήδη στο μοντέλο, να είναι:

$$r_{Y1,2}^2 = \frac{SSR(X_1 \setminus X_2)}{SSE(X_2)}$$

Ανάλογα, ο μερικός συντελεστής προσδιορισμού μεταξύ της  $Y$  και της  $X_2$  ενώ η  $X_1$  είναι ήδη στο μοντέλο, είναι:

$$r_{Y2,1}^2 = \frac{SSR(X_2 \setminus X_1)}{SSE(X_1)}$$

Επιπλέον, σε ένα μοντέλο με τρεις ή περισσότερες μεταβλητές  $X$  μπορούμε να ορίσουμε μερικούς από τους παρακάτω μερικούς συντελεστές προσδιορισμού:

$$r_{Y1,23}^2 = \frac{SSR(X_1 \setminus X_2, X_3)}{SSE(X_2, X_3)}$$

$$r_{Y2,13}^2 = \frac{SSR(X_2 \setminus X_1, X_3)}{SSE(X_1, X_3)}$$

$$r_{Y4,123}^2 = \frac{SSR(X_4 \setminus X_1, X_2, X_3)}{SSE(X_1, X_2, X_3)}$$

### Επιλογή Μεταβλητών

Στην ανάλυση πολλαπλής παλινδρόμησης, είναι συνηθισμένο να υπάρχουν πολλές ανεξάρτητες μεταβλητές που ενδεχομένως επηρεάζουν την εξαρτημένη μεταβλητή. Σε τέτοιες περιπτώσεις, ένας από τους βασικούς στόχους είναι η επιλογή του καταλληλότερου

υποσυνόλου μεταβλητών, ώστε να περιγραφεί και να προβλεφθεί το φαινόμενο με τη μέγιστη δυνατή ακρίβεια, χωρίς να συμπεριληφθούν περιττές ή λιγότερο χρήσιμες μεταβλητές. Η διαδικασία αυτή ονομάζεται επιλογή μεταβλητών (variable selection).

Αν και θεωρητικά μπορεί να εξεταστεί κάθε δυνατός συνδυασμός των μεταβλητών ώστε να εντοπιστεί το πιο αποτελεσματικό μοντέλο βάσει κάποιου κριτηρίου όπως ο διορθωμένος συντελεστής προσδιορισμού  $R_a^2$ , η πρακτική εφαρμογή αυτής της μεθόδου είναι σπάνια λόγω του τεράστιου αριθμού συνδυασμών, ειδικά όταν το πλήθος των μεταβλητών είναι μεγάλο.

Για να διευκολυνθεί η επιλογή, έχουν αναπτυχθεί βηματικές μέθοδοι επιλογής μεταβλητών, όπως:

1. **Οπισθοδρομική Απαλοιφή (Backward Elimination):** Ξεκινάμε με το πλήρες μοντέλο, που περιλαμβάνει όλες τις διαθέσιμες μεταβλητές και σταδιακά αφαιρούμε εκείνες που δεν συνεισφέρουν σημαντικά στην πρόβλεψη, συνήθως με βάση τη στατιστική σημαντικότητα ή την επίδρασή τους στο  $R_a^2$ .
2. **Πρόσθεση προς τα Εμπρός (Forward Selection):** Ξεκινά με ένα κενό μοντέλο και προσθέτει σε κάθε βήμα τη μεταβλητή που έχει τη μεγαλύτερη θετική επίδραση στην ακρίβεια του μοντέλου.
3. **Βηματική Μέθοδος (Stepwise Regression):** Πρόκειται για έναν συνδυασμό των δύο παραπάνω τεχνικών. Σε κάθε στάδιο ελέγχεται τόσο αν κάποια νέα μεταβλητή πρέπει να προστεθεί όσο και αν κάποια ήδη ενσωματωμένη πρέπει να αφαιρεθεί.

Παρότι οι τεχνικές επιλογής μεταβλητών είναι πολύ χρήσιμες, μπορούν να οδηγήσουν στην εμφάνιση του φαινομένου της πολυσυγγραμμικότητας (multicollinearity). Πρόκειται για την περίπτωση κατά την οποία μία ή περισσότερες ανεξάρτητες μεταβλητές συσχετίζονται έντονα με άλλες. Η συνέπεια είναι ότι καθίσταται δύσκολος ο διαχωρισμός της επίδρασης κάθε μεμονωμένης μεταβλητής στην εξαρτημένη.

Η πολυσυγγραμμικότητα δημιουργεί στατιστικά προβλήματα, καθώς ο πίνακας των ανεξάρτητων μεταβλητών ( $X'X$ ) γίνεται είτε αδύνατο να αντιστραφεί (στην περίπτωση ισχυρής πολυσυγγραμμικότητας) είτε δύσκολα αντιστρέψιμος (σε περίπτωση ασθενούς πολυσυγγραμμικότητας). Αυτή η κατάσταση επηρεάζει άμεσα την ακρίβεια των εκτιμήσεων των συντελεστών παλινδρόμησης και αυξάνει τα τυπικά σφάλματα, γεγονός που μειώνει την αξιοπιστία των συμπερασμάτων (Montgomery et. al., 2012).

Για την ανίχνευση της πολυσυγγραμμικότητας, χρησιμοποιείται ευρέως ο εκτιμητής διόγκωσης διακύμανσης (Variance Inflation Factor – VIF). Ο VIF αξιολογεί κατά πόσο η διακύμανση της εκτιμηθείσας τιμής ενός συντελεστή παλινδρόμησης αυξάνεται εξαιτίας της συσχέτισής του με τις υπόλοιπες μεταβλητές. Τυπικά, αν ο VIF είναι μεγαλύτερος από 10, υποδηλώνεται σοβαρή πολυσυγγραμμικότητα (Kutner et. al., 2004).

Η ασθενής πολυσυγγραμμικότητα αναφέρεται σε καταστάσεις όπου, παρότι δεν υπάρχει απόλυτη γραμμική εξάρτηση μεταξύ μεταβλητών, εντούτοις υπάρχει σημαντικός βαθμός συσχέτισης που επηρεάζει την ευστάθεια των εκτιμήσεων. Μπορεί να προκύψει από σφάλματα στρογγυλοποίησης ή διαφορετικές κλίμακες μέτρησης μεταξύ μεταβλητών. Η παρουσία πολυσυγγραμμικότητας έχει επίσης επιπτώσεις στις στατιστικές δοκιμές και στα διαστήματα εμπιστοσύνης, δυσχεραίνοντας την ορθή ερμηνεία των αποτελεσμάτων του μοντέλου.

Συνεπώς, η ορθή επιλογή μεταβλητών και η συνεχής παρακολούθηση της ύπαρξης πολυσυγγραμμικότητας είναι κρίσιμες διαδικασίες για την ανάπτυξη αξιόπιστων και ερμηνεύσιμων μοντέλων παλινδρόμησης.

### 4.1.3 Απλό Λογιστικό Μοντέλο

Το λογιστικό μοντέλο ανήκει στην κατηγορία των μη γραμμικών μοντέλων, όπου τα σφάλματα δεν ακολουθούν κανονική κατανομή, ενώ η μεταβλητή απόκριση είναι διακριτή. Η λογιστική παλινδρόμηση χρησιμοποιείται για να προβλέψει την παρουσία ή απουσία ενός χαρακτηριστικού ή ενός συμβάντος. Αποτελεί γενίκευση της απλής γραμμικής παλινδρόμησης, προσαρμοσμένη για δίτιμες εξαρτημένες μεταβλητές που παίρνουν τιμές 0 ή 1.

Η λογιστική παλινδρόμηση αναπτύχθηκε ως εναλλακτική λύση στη γραμμική διακριτική ανάλυση για την ταξινόμηση και έχει ευρεία εφαρμογή σε πολλά επιστημονικά πεδία, κυρίως στην ιατρική και τις κοινωνικές επιστήμες. Χρησιμοποιείται, για παράδειγμα, για την πρόβλεψη:

- Εμφάνισης νόσου βάσει διαφορετικών χαρακτηριστικών του ατόμου, όπως η ηλικία και το φύλο.
- Επιλογής πολιτικού κόμματος με βάση δημογραφικά στοιχεία όπως η ηλικία, το φύλο, η φυλή και η προηγούμενη ψηφοφορία.
- Πιθανότητα αποτυχίας διεργασίας παραγωγής προϊόντος σε εργοστάσιο τροφίμων.
- Πρόθεσης αγοράς αγαθού από καταναλωτή (έρευνα αγοράς). Πιθανότητα αθέτησης αποπληρωμής δανείου από δανειολήπτη.

Όπως είπαμε και πιο πάνω η λογιστική παλινδρόμηση είναι η γενίκευση της απλής γραμμικής παλινδρόμησης, οπότε θα ξεκινήσουμε από το απλό γραμμικό μοντέλο:  $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$  με  $i=1,2,\dots,n$ , όπου η  $Y_i$  είναι δυαδική, δηλαδή παίρνει ή την τιμή 0 ή την 1. Μπορεί τα σφάλματα να μην κατανέμονται κανονικά, ωστόσο η μέση τιμή τους είναι μηδενική, δηλαδή  $E(\varepsilon_i)=0$  και έτσι προκύπτει ότι:

$$E(Y_i) = E(\beta_0 + \beta_1 X_i + \varepsilon_i) = E(\beta_0 + \beta_1 X_i) + E(\varepsilon_i) = \beta_0 + \beta_1 X_i$$

Επίσης, αφού η  $Y_i$  είναι μια δίτιμη μεταβλητή θα είναι μια μεταβλητή Bernoulli, οπότε ορίζουμε τις πιθανότητες ως εξής:

- Όταν το  $Y_i = 1$  έχουμε  $P(Y_i = 1) = \pi_i$
- Όταν το  $Y_i = 0$  έχουμε  $P(Y_i = 0) = 1 - \pi_i$ , με  $\pi_i$  να είναι η πιθανότητα επιτυχίας.

Από τον ορισμό της μέσης τιμής εξασφαλίζουμε ότι:

$$E(Y_i) = 1 * \pi_i + 0 * (1 - \pi_i) = \pi_i$$

Εξισώνοντας τις παραπάνω σχέσεις, βρίσκουμε:

$$E(Y_i) = \beta_0 + \beta_1 X_i = \pi_i$$

$$Var(Y_i) = \pi_i (1 - \pi_i)$$

Άρα ο αποκρινόμενος μέσος  $E(Y_i)$  δηλώνει την πιθανότητα ότι η  $Y_i$  παίρνει την τιμή 1 όταν η προβλέπουσα μεταβλητή  $X = X_i$ .

Όταν η μεταβλητή  $Y$  είναι δίτιμη, δεν μπορεί να χρησιμοποιηθεί το γραμμικό μοντέλο και αυτό γιατί πέραν του γεγονότος ότι τα σφάλματα δεν είναι κανονικά κατανομημένα, έχουν άνισες διασπορές. Επιπλέον, ένα ακόμη πρόβλημα εντοπίζεται καθώς όπως έχουμε αναφέρει η εξαρτημένη μεταβλητή  $Y$  είναι δυαδική και παίρνει τις τιμές 0 και 1 με αποτέλεσμα να υπάρχει ο παρακάτω περιορισμός:

$$0 \leq E(Y) = \pi \leq 1$$

Η αντιμετώπιση των προβλημάτων που προκύπτουν από άνισες διασπορές είναι δυνατή με τη χρήση σταθμισμένων ελαχίστων τετραγώνων. Επιπρόσθετα, με το να λάβουμε ένα μεγάλο μέγεθος δείγματος, η μέθοδος ελαχίστων τετραγώνων παρέχει εκτιμήσεις που είναι ασυμπτωτικά κανονικές, ακόμη και όταν τα σφάλματα δεν ακολουθούν την κανονική κατανομή.

Ωστόσο, ένα κύριο πρόβλημα παραμένει στον περιορισμό της τιμής της συνάρτησης απόκρισης, καθώς η συγκεκριμένη συνάρτηση είναι μη γραμμική. Αυτό αποτελεί καίρια δυσκολία και είναι η αιτία που η συνάρτηση απόκρισης παραμένει μη γραμμική.

Το μοντέλο που χρησιμοποιούμε όταν η  $Y_i$  είναι δίτιμη είναι το λογιστικό, το οποίο ορίζεται ως εξής:  $Y_i = E(Y_i) + \varepsilon_i$ , όπου  $Y_i$  είναι ανεξάρτητη τυχαία μεταβλητή Bernoulli και:

$$E(Y_i) = \pi_i = \frac{e^{(\beta_0 + \beta_1 X_i)}}{1 + e^{(\beta_0 + \beta_1 X_i)}} = \left[ \frac{1 + e^{(\beta_0 + \beta_1 X_i)}}{e^{(\beta_0 + \beta_1 X_i)}} \right]^{-1} = \left[ \frac{1}{e^{(\beta_0 + \beta_1 X_i)}} + 1 \right]^{-1} = \left[ 1 + e^{(-\beta_0 - \beta_1 X_i)} \right]^{-1}$$

Η μορφή αυτή της αναμενόμενης συνάρτησης προκύπτει από την σχέση  $E(Y_i) = \beta_0 + \beta_1 x_i = \pi_i$ . Κάνοντας την διαπίστωση ότι η συνάρτηση αυτή λαμβάνει τιμές από  $-\infty$  έως  $+\infty$  ενώ η αναμενόμενη συνάρτηση πρέπει να παίρνει τιμές στο διάστημα  $[0, 1]$ , προβαίνουμε στην χρήση των παρακάτω μετασχηματισμών. Ο πρώτος μετασχηματισμός επιτρέπει τη μετατροπή του ποσοστού σε ένα μέτρο, η τιμή του οποίου κυμαίνεται από 0 έως  $+\infty$  και όχι

από 0 έως 1 όπως συμβαίνει με το ποσοστό.

Ο μετασχηματισμός αυτός επιτυγχάνεται λαμβάνοντας το λόγο συμπληρωματικών πιθανοτήτων (odds) του ποσοστού  $\pi_i/1-\pi_i$  και όχι το ποσοστό αυτό καθαυτό. Το odds μιας πιθανότητας είναι ο λόγος των συμπληρωματικών πιθανοτήτων, οπότε εάν η πιθανότητα εμφάνισης ενός ενδεχομένου συμβολιστεί με  $\pi_i$  και η πιθανότητα μη εμφάνισης με  $1-\pi_i$ , τότε το odds υπέρ του ενδεχομένου είναι  $\pi_i/1-\pi_i$  ως προς ένα. Όταν η τιμή του ποσοστού πλησιάζει το μηδέν, τότε η τιμή της ποσότητας σχεδόν ταυτίζεται με την τιμή του ποσοστού, ενώ όταν η τιμή του ποσοστού πλησιάζει το ένα, τότε ο παρονομαστής της ποσότητας  $\pi_i/1-\pi_i$  πλησιάζει το μηδέν και η τιμή της ποσότητας  $\pi_i/1-\pi_i$  προσεγγίζει το  $+\infty$ .

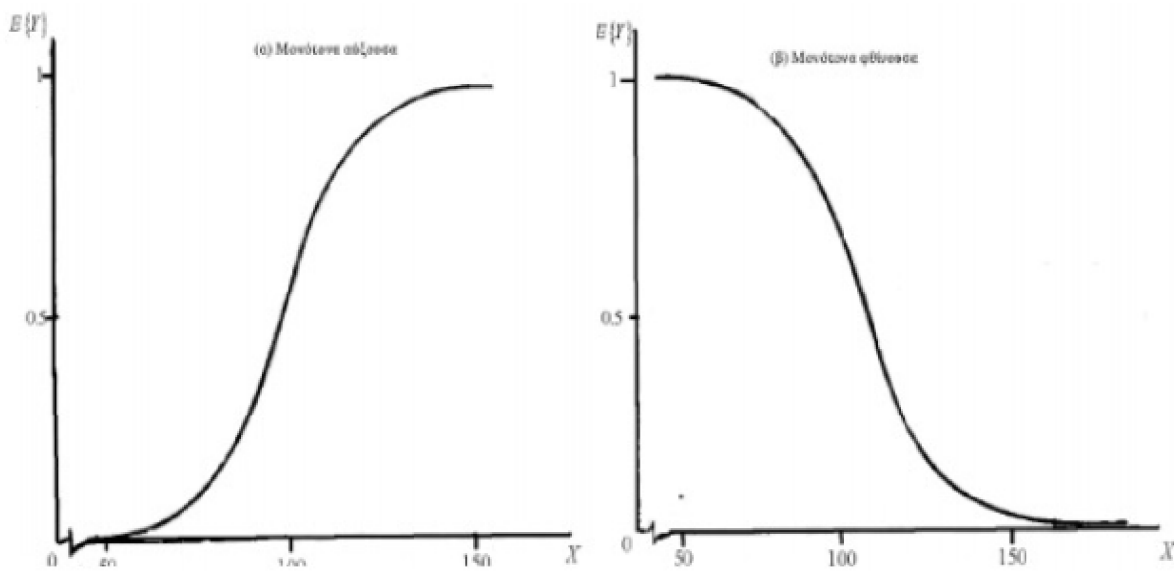
Ο δεύτερος μετασχηματισμός μετατρέπει το λόγο των συμπληρωματικών πιθανοτήτων του ποσοστού σε ένα μέτρο, η τιμή του οποίου κυμαίνεται από 0 έως 1. Στον μετασχηματισμό αυτό λαμβάνεται ο λογάριθμος του λόγου των συμπληρωματικών πιθανοτήτων. Το μέτρο που προκύπτει έπειτα από τους δύο αυτούς μετασχηματισμούς είναι το  $\ln(\pi_i/1-\pi_i)$  και ονομάζεται logit μετασχηματισμός της πιθανότητας  $\pi_i$ . Επομένως το logit είναι ο λογάριθμος του λόγου των συμπληρωματικών πιθανοτήτων. Ο μετασχηματισμός που περιλαμβάνει τα δύο παραπάνω βήματα ονομάζεται λογιστικός μετασχηματισμός.

Τελικά θα έχουμε ότι:

$$\begin{aligned}\ln\left(\frac{\pi_i}{1-\pi_i}\right) &= \beta_0 + \beta_1 X_i \\ \Leftrightarrow \frac{\pi_i}{1-\pi_i} &= e^{\beta_0 + \beta_1 X_i} \\ \Leftrightarrow \pi_i &= e^{\beta_0 + \beta_1 X_i} (1-\pi_i) \\ \Leftrightarrow \pi_i + \pi_i e^{\beta_0 + \beta_1 X_i} &= e^{\beta_0 + \beta_1 X_i} \\ \Leftrightarrow \pi_i (1 + e^{\beta_0 + \beta_1 X_i}) &= e^{\beta_0 + \beta_1 X_i} \\ \Leftrightarrow \pi_i &= \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \\ \Leftrightarrow E(Y_i) &= \frac{e^{(\beta_0 + \beta_1 X_i)}}{1 + e^{(\beta_0 + \beta_1 X_i)}}.\end{aligned}$$

Η αναμενόμενη λογιστική συνάρτηση είναι:

1. Είτε μονότονα αύξουσα συνάρτηση είτε μονότονα φθίνουσα,
2. Είναι σχεδόν γραμμική στην περιοχή  $[0.2, 0.8]$ ,
3. Πλησιάζει το 0 και 1 στις ακραίες τιμές της εμβέλειας του  $X$  όπως βλέπουμε και στην εικόνα 1.



#### 4.1.4 Απλή Λογιστική Παλινδρόμηση

Λόγω της δυαδικής φύσης της εξαρτημένης μεταβλητής στο μοντέλο της λογιστικής παλινδρόμησης, η εφαρμογή της μεθόδου των ελαχίστων τετραγώνων για την εκτίμηση των παραμέτρων δεν είναι εφικτή, όπως θα ήταν στην περίπτωση της γραμμικής παλινδρόμησης. Ωστόσο, μπορούμε να ξεπεράσουμε αυτήν τη δυσκολία χρησιμοποιώντας τη μέθοδο της μέγιστης πιθανοφάνειας. Η μέθοδος της μέγιστης πιθανοφάνειας επιδιώκει να επιλέξει τις τιμές των παραμέτρων του μοντέλου που μεγιστοποιούν τη συνάρτηση πιθανοφάνειας. Στην ουσία, επιλέγουμε αυτές τις παραμέτρους που καθιστούν τα παρατηρούμενα δεδομένα πιθανότερα, λαμβάνοντας υπόψη τη δυαδική φύση της εξαρτημένης μεταβλητής.

Με άλλα λόγια, προσαρμόζουμε το μοντέλο στα δεδομένα επιλέγοντας τις παραμέτρους που μεγιστοποιούν την πιθανότητα εμφάνισης των παρατηρούμενων αποτελεσμάτων.

Αφού τα  $Y_i$  είναι τυχαίες μεταβλητές Bernoulli όπου  $P(Y_i=1)=\pi_i$  και  $P(Y_i=0)=1-\pi_i$  η συνάρτηση πυκνότητας πιθανότητας είναι:

$$f_i(Y_i) = \pi_i^{Y_i} (1 - \pi_i)^{1-Y_i}$$

$Y_i=0,1$  και  $i=1,\dots,n$

Οι παρατηρήσεις  $Y_i$  είναι ανεξάρτητες οπότε η από κοινού συνάρτησης πιθανότητας θα είναι:

$$g(Y_1, \dots, Y_n) = \prod_{i=1}^n f_i(Y_i) = \prod_{i=1}^n \pi_i^{Y_i} (1 - \pi_i)^{1-Y_i}$$

Λογαριθμίζοντας αυτή την σχέση θα έχουμε:

$$\begin{aligned} \ln g(Y_1, \dots, Y_n) &= \ln \prod_{i=1}^n \pi_i^{Y_i} (1 - \pi_i)^{1-Y_i} \\ &= \sum_{i=1}^n \left[ Y_i \ln \left( \frac{\pi_i}{1 - \pi_i} \right) \right] + \sum_{i=1}^n \ln(1 - \pi_i) \end{aligned}$$

Μπορούμε να αντικαταστήσουμε το  $\ln(\pi_i/1 - \pi_i)$  και το  $\pi_i/1 - \pi_i$  οπότε θα έχουμε τη λογαριθμική συνάρτηση πιθανοφάνειας των εκτιμώμενων παραμέτρων:

$$\begin{aligned} \ln L(\beta_0, \beta_1) &= \sum_{i=1}^n Y_i (\beta_0 + \beta_1 X_i) + \sum_{i=1}^n \ln \left( 1 - \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right) \\ &= \sum_{i=1}^n Y_i (\beta_0 + \beta_1 X_i) + \sum_{i=1}^n \ln \left( \frac{1 + e^{\beta_0 + \beta_1 X_i} - e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right) \\ &= \sum_{i=1}^n Y_i (\beta_0 + \beta_1 X_i) + \sum_{i=1}^n \ln \left( 1 + e^{\beta_0 + \beta_1 X_i} \right)^{-1} \\ &= \sum_{i=1}^n Y_i (\beta_0 + \beta_1 X_i) - \sum_{i=1}^n \ln \left( 1 + e^{\beta_0 + \beta_1 X_i} \right) \end{aligned}$$

Εδώ να αναφέρουμε πως δεν μπορούν να βρεθούν οι εκτιμήσεις των συντελεστών  $\beta_0$  και  $\beta_1$  από την μέθοδο μέγιστης πιθανοφάνειας, γιατί δεν υπάρχουν λύσεις κλειστής μορφής των τιμών των συντελεστών που να μεγιστοποιούν την παραπάνω σχέση. Αφού βρεθούν οι εκτιμητές παλινδρομήσεως  $b_0$  και  $b_1$  μέσω επαναληπτικών αριθμητικών μεθόδων, έχουμε την προσαρμοσμένη λογιστική συνάρτηση απόκρισης:

$$\hat{\pi} = \frac{e^{b_0 + b_1 X}}{1 + e^{b_0 + b_1 X}}$$

Χρησιμοποιώντας το μετασχηματισμό logit θα έχουμε  $\hat{\pi}' = \ln \frac{\hat{\pi}}{1-\hat{\pi}}$  και οπότε προκύπτει η προσαρμοσμένη αναμενόμενη λογιστική συνάρτηση (fitted logit)  $\hat{\pi}' = b_0 + b_1 X$ . Ιδιαίτερα σημαντική είναι η ερμηνεία των συντελεστών της λογιστικής παλινδρόμησης. Όταν η εξαρτημένη μεταβλητή είναι ενδεικτική με  $Y=1$  αν έχουμε την εμφάνιση ενός χαρακτηριστικού και  $Y=0$  αν απουσιάζει η εμφάνιση του ίδιου χαρακτηριστικού, τότε ο συντελεστής  $b_1$  ισούται με τον λόγο των logits αυτών που έχουν το χαρακτηριστικό σε σχέση με αυτούς που δεν το έχουν, δηλαδή:

$$\ln \left( \frac{\pi_1}{1-\pi_1} \right) - \ln \left( \frac{\pi_0}{1-\pi_0} \right) = \ln \left( \frac{\frac{\pi_1}{1-\pi_1}}{\frac{\pi_0}{1-\pi_0}} \right) = \ln \left( \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} \right) = \ln(e^{\beta_1}) = b_1$$

Έτσι στη λογιστική παλινδρόμηση, ο αντιλογάριθμος  $e^{b_1}$  του συντελεστή παλινδρόμησης  $b_1$  μιας ενδεικτικής ανεξάρτητης μεταβλητής  $X_1$ , αποτελεί εκτίμηση του λόγου των odds αυτών που έχουν κάποιο χαρακτηριστικό σε σχέση με αυτούς που δεν το εμφανίζουν και ονομάζεται λόγος των εκτιμώμενων πιθανοτήτων:

$$e^{b_1} = e^{\ln \left( \frac{\frac{\pi_1}{1-\pi_1}}{\frac{\pi_0}{1-\pi_0}} \right)} = \left( \frac{\frac{\pi_1}{1-\pi_1}}{\frac{\pi_0}{1-\pi_0}} \right) = \frac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}$$

Αν το  $B_1$  είναι θετικό, ο παράγοντας  $e^{b_1}$  είναι μεγαλύτερος της μονάδας και έτσι ο εκτιμώμενος λόγος πιθανοτήτων αυξάνεται. Ανάλογα αν το  $b_1$  είναι αρνητικό, ο παράγοντας  $e^{b_1}$  είναι μικρότερος της μονάδας και ο λόγος πιθανοτήτων μειώνεται.

Οι παράμετροι της λογιστικής παλινδρόμησης μπορούν να εκφραστούν και μέσα από το σχετικό λόγο των συμπληρωματικών πιθανοτήτων, δηλαδή το λόγο των odds που ονομάζεται odds ratio. Ο λόγος των odds ενός ατόμου με τιμές συμμεταβλητών  $X_1$  σε σχέση με ένα άτομο με τιμές  $X_2$  των ίδιων συμμεταβλητών δίνεται από την παρακάτω σχέση:

$$\text{Odds ratio} = \frac{\frac{\hat{p}_1}{1-\hat{p}_1}}{\frac{\hat{p}_2}{1-\hat{p}_2}} = e^{(x_1 - x_2)' \hat{\beta}}$$

Αν το odds ratio είναι ίσο με τη μονάδα σημαίνει ότι τα odds των δύο ομάδων είναι ίσα. Πιο γενικά το odds ratio μας δείχνει πόσες φορές μεγαλύτερο ή μικρότερο είναι το ένα odds από το άλλο.

#### 4.1.5 Ελεγχοςυναρτήσεις

##### Ελεγχοςυναρτηση Deviance

Η Ελεγχοςυναρτηση Deviance (Deviance Residuals) είναι ένας τρόπος αξιολόγησης της προσαρμογής ενός λογιστικού μοντέλου. Αυτή η μέτρηση χρησιμοποιείται σε μοντέλα λογιστικής παλινδρόμησης, όπου η εξαρτημένη μεταβλητή είναι δυαδική (διακριτική).

Η Ελεγχοςυναρτηση Deviance βασίζεται στον υπολογισμό της αποκλίνουσας απόκρισης (Deviance) για κάθε παρατηρούμενη παρατήρηση. Η Deviance είναι μια μέτρηση της απόκλισης των παρατηρούμενων τιμών από τις προβλεπόμενες τιμές στο μοντέλο. Τα Deviance Residuals είναι οι αποκλίσεις αυτές.

Η Ελεγχοςυναρτηση Deviance μπορεί να χρησιμοποιηθεί για να ελέγξει την προσαρμογή του μοντέλου. Αν οι Deviance Residuals είναι κοντά στο μηδέν και δεν υπάρχει κάποιο παρατηρούμενο πρότυπο στα υπόλοιπα των Deviance Residuals, τότε το μοντέλο πιθανόν να είναι καλά προσαρμοσμένο.

Συγκεκριμένα, αν οι Deviance Residuals είναι μεγάλες για κάποιες παρατηρήσεις, αυτό μπορεί να υπονοεί προβλήματα προσαρμογής σε αυτές τις περιοχές. Επίσης, η σύγκριση της Deviance μεταξύ διαφορετικών μοντέλων μπορεί να βοηθήσει στην επιλογή του καλύτερου μοντέλου.

Η Ελεγχοςυναρτηση Deviance αποτελεί σημαντικό εργαλείο στη στατιστική αξιολόγηση των λογιστικών μοντέλων, καθιστώντας δυνατή την ανίχνευση περιοχών όπου το μοντέλο ενδέχεται να χρειάζεται βελτίωση.

Ο έλεγχος Deviance λοιπόν, συγκρίνει το λόγο των πιθανοφάνειών των παρατηρούμενων και των αναμενόμενων τιμών της μεταβλητής απόκρισης από μοντέλα με και χωρίς την εν λόγω μεταβλητή.

$$D = -2 \ln \left[ \frac{\text{πιθανοφάνεια προσαρμοσμένου μοντέλου}}{\text{πιθανοφάνεια κορεσμένου μοντέλου}} \right]$$

Για να εξάγουμε την τυποποιημένη συνάρτηση Deviance χρειάζεται αρχικά να κατασκευάσουμε τη μεγιστοποιημένη τιμή του λογαρίθμου της πιθανοφάνειας στο κορεσμένο μοντέλο:

$$\tilde{l}_i = \sum_{i=1}^n [y_i \ln(\tilde{p}_i) + (1 - y_i) \ln(1 - \tilde{p}_i)]$$

Η μεγιστοποιημένη τιμή του λογαρίθμου της πιθανοφάνειας:

$$\hat{l}_i = \sum_{i=1}^n [y_i \ln(\hat{p}_i) + (1 - y_i) \ln(1 - \hat{p}_i)]$$

Αφαιρώντας τις παραπάνω σχέσεις παίρνουμε:

$$\begin{aligned} \hat{l}_i - \tilde{l}_i &= \sum_{i=1}^n [y_i \ln(\hat{p}_i) + (1 - y_i) \ln(1 - \hat{p}_i) - y_i \ln(\tilde{p}_i) - (1 - y_i) \ln(1 - \tilde{p}_i)] \\ &= \sum_{i=1}^n [y_i (\ln \hat{p}_i - \ln \tilde{p}_i) + (1 - y_i) [\ln(1 - \hat{p}_i) - \ln(1 - \tilde{p}_i)]] \\ &= \sum_{i=1}^n \left[ y_i \ln \left( \frac{\hat{p}_i}{\tilde{p}_i} \right) + (1 - y_i) \ln \left( \frac{1 - \hat{p}_i}{1 - \tilde{p}_i} \right) \right] \end{aligned}$$

Επομένως ορίζουμε την τυποποιημένη συνάρτηση Deviance ως:

$$\begin{aligned} D(\hat{\beta}) &= -2 \{ \hat{l}_i - \tilde{l}_i \} \\ &= -2 \sum_{i=1}^n \left[ y_i \ln \left( \frac{\hat{p}_i}{\tilde{p}_i} \right) + (1 - y_i) \ln \left( \frac{1 - \hat{p}_i}{1 - \tilde{p}_i} \right) \right] \end{aligned}$$

Αξιζει να σημειωθεί ότι η ελεγχοσυνάρτηση Deviance έχει τον ίδιο ρόλο με εκείνον που έχει το άθροισμα τετραγώνων των σφαλμάτων στη γραμμική παλινδρόμηση.

Όσον αφορά την περίπτωση στην οποία η μεταβλητή απόκρισης είναι δίτιμη η πιθανοφάνεια του κορεσμένου μοντέλου κα είναι ίση είτε με 1 είτε με 0. Από τον ορισμό του κορεσμένου μοντέλου θα έχουμε  $\hat{p}_i = y_i$  και πιθανοφάνεια ίση με 1, οπότε η συνάρτηση Deviance γίνεται:

$$D = -2 \ln (\text{πιθανοφάνεια προσαρμοσμένου μοντέλου})$$

Τέλος μπορούμε να εφαρμόσουμε ένα έλεγχο για να συγκρίνουμε τη καταλληλότητα του μοντέλου με και χωρίς την ανεξάρτητη μεταβλητή. Για να γίνει αυτό συγκρίνουμε την τιμή του D στα δύο μοντέλα:

$$G = D (\text{το μοντέλο χωρίς τη μεταβλητή}) - D (\text{το μοντέλο με τη μεταβλητή})$$

Σο G έχει τον ίδιο ρόλο με το F-test στη γραμμική παλινδρόμηση και μπορεί να εκφραστεί ως:

$$G = -2 \ln \left[ \frac{\text{πιθανοφάνεια χωρίς τη μεταβλητή}}{\text{πιθανοφάνεια με τη μεταβλητή}} \right]$$

Γνωρίζοντας ότι το G ακολουθεί ασυμπτωτικά την κατανομή  $\chi^2$  μπορούμε να εφαρμόσουμε ελέγχους υποθέσεων, ώστε να εξετάσουμε τη σημαντικότητα του.

### Ελεγχοςυνάρτηση Wald

Η Ελεγχοςυνάρτηση Wald είναι μια μέθοδος στατιστικής ανάλυσης που χρησιμοποιείται για τον έλεγχο υποθέσεων σε ένα στατιστικό πλαίσιο. Πρόκειται για έναν τύπο στατιστικού τεστ που εξετάζει αν μια παρατήρηση ή μια σειρά παρατηρήσεων πληροί ή όχι ένα κριτήριο.

Ας εξετάσουμε τη διαδικασία αυτήν πιο αναλυτικά:

1. Υπόθεση Μηδενικής Υπόθεσης ( $H_0$ ): Η υπόθεση που εξετάζεται και την οποία θέλουμε να απορρίψουμε ή να μην απορρίψουμε. Στην περίπτωση της Ελεγχοςυνάρτησης Wald, η  $H_0$  συνήθως σχετίζεται με την εκτίμηση ενός παραμέτρου.
2. Υπόθεση Εναλλακτικής Υπόθεσης ( $H_1$ ): Η υπόθεση που εκφράζει την αντίθετη πρόταση από την  $H_0$ . Στην περίπτωση της Ελεγχοςυνάρτησης Wald, η  $H_1$  συνήθως αφορά τη μη-ισότητα ή κάποια άλλη σχέση που θέλουμε να ελέγξουμε.
3. Εκτίμηση Παραμέτρου: Κατά τον καθορισμό της  $H_0$ , εκτιμούμε μια παράμετρο του πληθυσμού (όπως η μέση τιμή ή η πιθανότητα επιτυχίας).
4. Στατιστικό Τεστ Wald: Υπολογίζουμε το στατιστικό τεστ Wald, το οποίο είναι η αναλογία της εκτιμηθείσας παραμέτρου στο σφάλμα του εκτιμητή. Η σχέση είναι συνήθως της μορφής: 
$$Wald\ Stat = \frac{(\hat{\theta} - \theta_0)^2}{Var(\hat{\theta})}$$
 όπου  $\hat{\theta}$  είναι η εκτιμηθείσα παράμετρος,  $\theta_0$  είναι η τιμή που υποθέτουμε ως μηδενική και  $Var(\hat{\theta})$  είναι η διακύμανση της εκτιμηθείσας παραμέτρου.
5. Απόφαση Αποδοχής ή Απόρριψης της  $H_0$ : Σε αυτό το στάδιο, αποφασίζουμε αν έχουμε επαρκείς αποδείξεις ώστε να απορρίψουμε την  $H_0$ . Χρησιμοποιούμε ένα κατώφλι απόφασης (π.χ., επίπεδο εμπιστοσύνης) για να αποφασίσουμε.

Η Ελεγχοςυνάρτηση Wald συνήθως χρησιμοποιείται σε πλαίσια όπου οι παράμετροι μπορούν να εκτιμηθούν με μεγάλη ακρίβεια και το μοντέλο είναι γραμμικό. Είναι σημαντικό να ληφθεί υπόψη ότι η μέθοδος αυτή έχει κάποιες προϋποθέσεις που πρέπει να πληρούνται για να είναι αξιόπιστη.

Για να ελέγξουμε τη σημαντικότητα των συντελεστών του μοντέλου, οι οποίες αποτελούν σημειακές εκτιμήσεις, θα χρησιμοποιήσουμε τον έλεγχο του Wald. Με τον έλεγχο του Wald μπορούμε να εξετάσουμε τις εξής υποθέσεις

$H_0: \beta_i=0$ , δηλαδή η μεταβλητή  $x_i$  δεν συμβάλει στο μοντέλο

$H_1: \beta_i \neq 0$ , δηλαδή η μεταβλητή  $x_i$  συμβάλει στο μοντέλο

Ο έλεγχος γίνεται σε  $(1-\alpha)\%$  επίπεδο σημαντικότητας και η στατιστική συνάρτηση υπό την  $H_0$  είναι η ακόλουθη:

$$\frac{\widehat{\beta}_1}{se\widehat{\beta}_1}$$

Η οποία καλείται Wald, ακολουθεί ασυμπτωτικά την κανονική κατανομή, ενώ το τετράγωνο της ακολουθεί την  $\chi^2$  κατανομή με 1 βαθμό ελευθερίας. Αν η p-value του ελέγχου Wald κριθεί ότι έχει λάβει μικρή τιμή τότε απορρίπτεται η μηδενική υπόθεση και έτσι ο συντελεστής  $\beta_i$  δεν μηδενίζεται οπότε η μεταβλητή  $x_i$  θεωρείται στατιστικά σημαντική. Σε αντίθετη περίπτωση, δηλαδή όταν η τιμή της p θεωρηθεί μεγάλη δεν απορρίπτουμε τη μηδενική υπόθεση.

Με τη βοήθεια της στατιστικής συνάρτησης Wald μπορούμε να κατασκευάσουμε διαστήματα εμπιστοσύνης για τις παραμέτρους του μοντέλου. Ένα  $100(1-\alpha)\%$  διάστημα εμπιστοσύνης για την παράμετρο  $\beta_i$  δίνεται ως εξής:

$$\exp \left[ \hat{\beta}_i \pm z_{\alpha/2} \text{se}(\hat{\beta}_i) \right]$$

### Ελεγχοςυνάρτηση Pearson

Η Ελεγχοςυνάρτηση Pearson είναι ένα στατιστικό τεστ που χρησιμοποιείται για να αξιολογήσει την εξάρτηση μεταξύ δύο κατηγοριών μεταβλητών. Πρόκειται για ένα τεστ χι-τετραγωνικότητας, που επικεντρώνεται στον υπολογισμό ενός δείκτη γνωστού ως  $\chi^2$ . Ας αναλύσουμε τη διαδικασία αυτήν λεπτομερώς:

1. Υποθέσεις:

Η  $H_0$  (Μηδενική Υπόθεση): Δεν υπάρχει εξάρτηση μεταξύ των δύο κατηγοριών μεταβλητών.

$H_1$  (Εναλλακτική Υπόθεση): Υπάρχει εξάρτηση μεταξύ των δύο κατηγοριών μεταβλητών.

2. Σχεδιασμός Πίνακα Συχνοτήτων:

Δημιουργούμε έναν πίνακα συχνοτήτων (πίνακα ελέγχου) που καταγράφει τον αριθμό των παρατηρήσεων για κάθε συνδυασμό των δύο κατηγοριών μεταβλητών.

3. Υπολογισμός  $\chi^2$ :

Υπολογίζουμε το  $\chi^2$  χρησιμοποιώντας τον τύπο:  $\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

όπου  $O_{ij}$  είναι η παρατήρηση και  $E_{ij}$  είναι η αναμενόμενη συχνότητα για τον συνδυασμό  $i, j$ . Οι αναμενόμενες συχνότητες υπολογίζονται ως το γινόμενο των αντίστοιχων συνολικών συχνοτήτων διαιρούμενο με τον συνολικό αριθμό παρατηρήσεων.

4. Κατανομή  $\chi^2$ :

Σε μεγάλα δείγματα, το  $\chi^2$  ακολουθεί κατανομή  $\chi^2$  με  $df = (r-1)(c-1)$ , όπου  $r$  είναι ο αριθμός των γραμμών και  $c$  είναι ο αριθμός των στηλών στον πίνακα ελέγχου.

5. Απόφαση Αποδοχής ή Απόρριψης της  $H_0$ :

Ελέγχουμε εάν ο υπολογισμός του  $\chi^2$  ταιριάζει με τις τιμές από τον πίνακα κατανομής. Εάν το υπολογισμένο  $\chi^2$  βρίσκεται στο εύρος που προκαθορίζεται για απόρριψη, τότε απορρίπτουμε την μηδενική υπόθεση ( $H_0$ ).

Η Ελεγχοςυνάρτηση Pearson είναι χρήσιμη για να εξεταστεί η σχέση μεταξύ δύο κατηγοριών μεταβλητών. Πρέπει να ληφθεί υπόψη ότι υπάρχουν περιορισμοί στη χρήση της, όπως η ανεξαρτησία των παρατηρήσεων και οι αναμενόμενες συχνότητες που δεν πρέπει να είναι πολύ χαμηλές.

Αυτός ο έλεγχος παρέχει χρήσιμες πληροφορίες για την προσαρμογή του μοντέλου, ιδίως όταν η εξαρτημένη μεταβλητή ακολουθεί διωνυμική κατανομή. Αυτό συμβαίνει καθώς, σε περιπτώσεις όπου η μεταβλητή απόκρισης είναι δίτιμη (παίρνει τις τιμές 0 και 1), η ασυμπτωτική θεωρία δεν είναι ισχύουσα όταν οι συχνότητες είναι χαμηλές (με  $n_i = 1$ ). Γι' αυτό το λόγο, είναι προτιμότερο να χρησιμοποιούμε εναλλακτικούς ελέγχους, όπως αυτόν των Hosmer-Lemeshow, όταν αντιμετωπίζουμε τέτοιες διωνυμικές κατανομές. Η ελεγχοσυνάρτηση του Pearson δίνεται ως:

$$X^2 = \sum_{i=1}^n \frac{(y_i - n_i \hat{p}_i)^2}{n_i \hat{p}_i (1 - \hat{p}_i)}$$

Αυτή η στατιστική συνάρτηση είναι ασυμπτωτικά ισοδύναμη με την ελεγχοσυνάρτηση Deviance και ακολουθεί την ίδια κατανομή με τη Deviance δηλαδή την  $\chi^2$ , εφόσον το μοντέλο είναι το σωστό, δηλαδή  $\chi^2 \sim \chi_{n-p}^2$  ασυμπτωτικά με  $n$  να είναι ο αριθμός των ομάδων των παρατηρήσεων για τις οποίες οι συμμεταβλητές παίρνουν τις ίδιες τιμές και  $p=k+1$  ο αριθμός των παραμέτρων στο μοντέλο.

Συνήθως, οι τιμές των δύο προαναφερθέντων στατιστικών, του Pearson και της Deviance, δεν διαφέρουν σημαντικά. Το πλεονέκτημα της ελεγχοσυνάρτησης Pearson είναι ότι δεν επηρεάζεται υπερβολικά από πολύ μικρές συχνότητες. Αντίθετα, το πλεονέκτημα της Deviance είναι ότι μπορεί να χρησιμοποιηθεί για τη σύγκριση δύο εμφωλευμένων μοντέλων. Αυτό συμβαίνει γιατί η Deviance μετρά τη διαφορά μεταξύ των δύο μοντέλων, επιτρέποντας μας να αξιολογήσουμε τη σημαντικότητα των επιπρόσθετων όρων που εισάγονται.

### **Ελεγχοσυνάρτηση Hosmer-Lemeshow**

Η Ελεγχοσυνάρτηση Hosmer-Lemeshow είναι ένας στατιστικός έλεγχος που χρησιμοποιείται για να αξιολογήσει την προσαρμογή (goodness-of-fit) ενός δυαδικού (δίτιμου) μοντέλου παλινδρόμησης, όπως το λογιστικό μοντέλο. Στόχος του είναι να εξετάσει κατά πόσο το προβλεπόμενο μοντέλο ταιριάζει καλά με τα πραγματικά δεδομένα.

Ας αναλύσουμε τα βασικά βήματα και τα χαρακτηριστικά αυτού του ελέγχου:

1. Υποθέσεις:  
 $H_0$  (Μηδενική Υπόθεση): Το μοντέλο προσαρμόζεται καλά στα δεδομένα, δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των προβλεπόμενων και των πραγματικών τιμών.  
 $H_1$  (Εναλλακτική Υπόθεση): Υπάρχει στατιστικά σημαντική διαφορά μεταξύ των προβλεπόμενων και των πραγματικών τιμών.
2. Διαίρεση Δεδομένων σε Ομάδες:  
 Διαχωρίζουμε τα δεδομένα μας σε ομάδες (bins) με βάση τις προβλεπόμενες πιθανότητες που παράγει το μοντέλο. Συνήθως, χρησιμοποιούνται δεκάδες ή πεντάδες για τις ομάδες.
3. Υπολογισμός Συγκεντρωτικής Στατιστικής:

Υπολογίζουμε μια συγκεντρωτική στατιστική, συνήθως χρησιμοποιώντας το  $\chi^2$  ή το λογάριθμο της Deviance, συγκρίνοντας τις πραγματικές και προβλεπόμενες συχνότητες σε κάθε ομάδα.

4. Υπολογισμός Πιθανοφάνειας (P-value):

Υπολογίζουμε το πιθανοφάνεια (p-value) από τη στατιστική συγκεντρωτικής μέτρησης. Αν το p-value είναι μικρότερο από το επιλεγμένο επίπεδο σημαντικότητας (συνήθως 0.05), τότε απορρίπτουμε τη μηδενική υπόθεση ότι το μοντέλο προσαρμόζεται καλά.

Η Ελεγχοςυνάρτηση Hosmer-Lemeshow είναι χρήσιμη για την αξιολόγηση της προσαρμογής του λογιστικού μοντέλου σε δυαδικά δεδομένα. Πρέπει να ληφθεί υπόψη ότι υπάρχουν κριτικές πτυχές στην αποτελεσματικότητά της και σε ορισμένες περιπτώσεις, είναι προτιμότερο να χρησιμοποιηθούν εναλλακτικοί έλεγχοι προσαρμογής.

Λόγω της αδυναμίας των ελεγχοςυναρτήσεων Deviance και Pearson να ανταπεξέλθουν ως μέτρα καταλληλότητας του μοντέλου σε δυαδικά δεδομένα με  $n_i = 1$  προκύπτει η ανάγκη χρήσης μίας άλλης ελεγχοςυνάρτησης που προτάθηκε από τους Hosmer και Lemeshow το 1980. Για τον υπολογισμό του ελέγχου αυτού οι παρατηρήσεις χωρίζονται σε ομάδες η καθεμία από τις οποίες έχει περίπου ίδιο αριθμό παρατηρήσεων. Η ομαδοποίηση αυτή γίνεται με βάση την τιμή της εκτιμημένης πιθανότητας  $\hat{p}_i$ , δηλαδή διατάσσουμε τις παρατηρήσεις σε αύξουσα σειρά σύμφωνα με την τιμή της  $\hat{p}_i$  και έπειτα τις χωρίζουμε σε ομάδες.

Ας υποθέσουμε ότι δημιουργήσαμε g τέτοιες ομάδες και μία από αυτές, έστω η i-οστή, έχει  $m_i$  παρατηρήσεις με  $o_i$  να είναι ο συνολικός αριθμός των επιτυχιών και  $e_i$  ο αναμενόμενος αριθμός των επιτυχιών. Αξίζει να σημειωθεί ότι η συχνότητα  $o_i$  προκύπτει από το άθροισμα των  $y_i$  για  $j=1,2,\dots,m$  ενώ η συχνότητα  $e_i$  από το άθροισμα  $\hat{\mu}_j$  για  $j=1,2,\dots,m$ . Η μέση πιθανότητα επιτυχίας της i-οστής ομάδας δίνεται από τον τύπο  $\hat{\pi}_i = \frac{e_i}{m_i}$ .

Η Ελεγχοςυνάρτηση Hosmer-Lemeshow δίνεται από τον τύπο:

$$X_{HL}^2 = \sum_{i=1}^g \frac{(o_i - m_i \hat{\pi}_i)^2}{m_i \hat{\pi}_i (1 - \hat{\pi}_i)}$$

Ο εν λόγω έλεγχος δεν είναι ευρέως διαδεδομένος, καθώς η αξία του εξαρτάται από τον τρόπο με τον οποίο γίνεται ο χωρισμός των παρατηρήσεων σε ομάδες, καθώς και από τον αριθμό των παρατηρήσεων σε κάθε μία από αυτές τις ομάδες.

#### 4.1.6 Υπόλοιπα

Ο έλεγχος με τη Deviance παρέχει μια ενδεικτική αξιολόγηση για την καταλληλότητα του μοντέλου, αλλά προσφέρει περιορισμένες πληροφορίες. Για περισσότερες λεπτομέρειες, απευθυνόμαστε στα υπόλοιπα, τα οποία αντιπροσωπεύουν τη συμφωνία μεταξύ των παρατηρήσεων  $y_i$  και των αντίστοιχων προσαρμοσμένων τιμών  $\hat{y}_i$  ή  $\hat{\mu}_i$ . Τα υπόλοιπα είναι ιδιαίτερα χρήσιμα, καθώς μας επιτρέπουν να κατασκευάσουμε διαγνωστικά γραφήματα για

το μοντέλο. Τα γραφήματα των τυποποιημένων υπολοίπων σε σχέση με τις εκτιμώμενες τιμές εξυπηρετούν τον έλεγχο της υπόθεσης της ανεξαρτησίας των παρατηρήσεων, ενώ μας βοηθούν να εντοπίσουμε παρατηρήσεις που ενδέχεται να ξεφεύγουν από τη συνήθη συμπεριφορά.

Προτού εξετάσουμε τα υπόλοιπα Pearson, τα υπόλοιπα Deviance και τα υπόλοιπα πιθανοφάνειας που χρησιμοποιούνται στη λογιστική παλινδρόμηση, ας παρουσιάσουμε τη σχέση που καθορίζει τα υπόλοιπα (residuals) στο απλό γραμμικό μοντέλο. Αυτή η σχέση εκφράζεται ως εξής:

$$e_i = y_i - \hat{y}_i$$

η οποία εκφράζει την κατακόρυφη απόκλιση του  $y_i$  από την ευθεία της προσαρμοσμένης ή εκτιμώμενης συνάρτησης παλινδρόμησης.

### **Υπόλοιπα Pearson**

Στη λογιστική παλινδρόμηση, τα υπόλοιπα Pearson χρησιμοποιούνται για να εκτιμήσουν την ποιότητα της προσαρμογής ενός λογιστικού μοντέλου στα δεδομένα. Τα υπόλοιπα Pearson είναι οι κανονικοποιημένες διαφορές μεταξύ των παρατηρούμενων και των προβλεπόμενων συχνοτήτων στις δυαδικές μεταβλητές. Εδώ είναι ένας λεπτομερής τύπος για τα υπόλοιπα Pearson:

Το υπόλοιπο Pearson για κάθε παρατήρηση υπολογίζεται ως η διαφορά μεταξύ της πραγματικής τιμής της απόκρισης (0 ή 1) και της προβλεπόμενης πιθανότητας επιτυχίας (προβλεπόμενη από το λογιστικό μοντέλο).

Το άθροισμα των τετραγώνων των υπολοίπων Pearson συνολικά για όλες τις παρατηρήσεις χρησιμοποιείται για τον υπολογισμό της στατιστικής δοκιμής  $\chi^2$ , η οποία αξιολογεί την προσαρμογή του μοντέλου στα δεδομένα. Μια μικρή τιμή του  $\chi^2$  και μια υψηλή p-τιμή υποδεικνύουν καλή προσαρμογή του μοντέλου.

Τα υπόλοιπα Pearson στα γενικευμένα γραμμικά μοντέλα δίνονται από την παρακάτω σχέση:

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{v(\hat{\mu}_i)}} \quad i=1,2,3,\dots$$

Επειδή  $v(\hat{\mu}_i) = n_i \hat{p}_i (1 - \hat{p}_i)$ , θα έχουμε:

$$r_i^P = \frac{y_i - n_i \hat{p}_i}{\sqrt{n_i \hat{p}_i (1 - \hat{p}_i)}} \quad i=1,2,3,\dots$$

Τα τυποποιημένα υπόλοιπα Pearson ορίζονται:

$$r_i^{PS} = \frac{y_i - n_i \hat{p}_i}{\sqrt{n_i \hat{p}_i (1 - \hat{p}_i) (1 - \hat{h}_{ii})}} = \frac{r_i^P}{\sqrt{1 - \hat{h}_{ii}}}$$

όπου  $\hat{h}_{ii}$  (hat values) είναι το διαγώνιο στοιχείο του  $n \times n$  πίνακα

$$\hat{H} = \hat{W}^{1/2} X (X' \hat{W} X)^{-1} X' \hat{W}^{1/2}$$

με  $X$  ο  $n \times p$  πίνακας σχεδιασμού και  $\hat{W}$  ο  $n \times n$  διαγώνιος  $W$  πίνακας, του οποίου το κάθε στοιχείο είναι το  $n_i \hat{p}_i (1 - \hat{p}_i)$  που αποτελεί την εκτιμημένη διασπορά της απόκρισης.

### Υπόλοιπα Deviance

Στη λογιστική παλινδρόμηση, τα υπόλοιπα Deviance χρησιμοποιούνται για να αξιολογήσουν την προσαρμογή ενός λογιστικού μοντέλου στα δεδομένα. Το υπόλοιπο Deviance εκτιμά την ανωμαλία των παρατηρούμενων δεδομένων σε σχέση με τις προβλεπόμενες τιμές του μοντέλου.

Μετά τον υπολογισμό του υπολοίπου Deviance, μπορεί να χρησιμοποιηθεί για τον υπολογισμό του τεστ  $\chi^2$ . Ένα χαμηλό υπόλοιπο Deviance και μια υψηλή  $p$ -τιμή υποδεικνύουν ότι το λογιστικό μοντέλο είναι στατιστικά σημαντικό και καλά προσαρμοσμένο στα δεδομένα.

Τα υπόλοιπα Deviance στο μοντέλο της λογιστικής παλινδρόμησης δίνεται από την ακόλουθη σχέση:

$$r_i^D = \text{sgn}(y_i - \hat{\mu}_i) \left[ 2y_i \ln \left( \frac{y_i}{\hat{\mu}_i} \right) + 2(n_i - y_i) \ln \left( \frac{n_i - y_i}{n_i - \hat{\mu}_i} \right) \right]^{1/2}$$

όπου  $\text{sgn}(y_i - \hat{y}_i)$  είναι το πρόσημο της διαφοράς  $y_i - \hat{y}_i$  οπότε η ποσότητα

$$\left[ 2y_i \ln \left( \frac{y_i}{\hat{\mu}_i} \right) + 2(n_i - y_i) \ln \left( \frac{n_i - y_i}{n_i - \hat{\mu}_i} \right) \right]^{1/2}$$

κα λαμβάνει θετικό πρόσημο όταν  $y_i \geq \hat{\mu}_i$  ενώ θα έχει αρνητικό πρόσημο όταν  $y_i \leq \hat{\mu}_i$ .

Το  $\text{sgn}(y_i - \hat{y}_i)$  μας βεβαιώνει ότι το  $r_i^D$  θα έχει ίδιο πρόσημο με το  $r_i^P$  αφού  $\hat{\mu}_i = n_i \hat{p}_i$ .

Αν τα δεδομένα είναι δυαδικά θα έχουμε  $\hat{\mu}_i = \hat{p}_i$ , έτσι η σχέση γίνεται:

$$r_i^D = \text{sgn}(y_i - \hat{\mu}_i) \{-2[y_i \ln \hat{\mu}_i + (1 - y_i) \ln(1 - \hat{\mu}_i)]\}^{1/2}$$

Επομένως τα τυποποιημένα υπόλοιπα υπολογίζονται από την σχέση:

$$r_i^{DS} = \frac{r_i^D}{\sqrt{1 - \hat{h}_{ii}}}$$

### Υπόλοιπα πιθανοφάνειας

Τα υπόλοιπα πιθανοφάνειας (likelihood residuals) στη λογιστική παλινδρόμηση χρησιμοποιούνται για να αξιολογήσουν την προσαρμογή ενός λογιστικού μοντέλου στα δεδομένα. Τα υπόλοιπα πιθανοφάνειας μετρούν τον βαθμό στον οποίο οι παρατηρούμενες τιμές αποκλίνουν από τις προβλεπόμενες πιθανότητες του μοντέλου.

Τα υπόλοιπα πιθανοφάνειας χρησιμοποιούνται επίσης για τον υπολογισμό του τετραγώνου των υπολοίπων πιθανοφάνειας, το οποίο συνολικά μπορεί να χρησιμοποιηθεί για τον υπολογισμό του τεστ  $\chi^2$ , αξιολογώντας την προσαρμογή του μοντέλου στα δεδομένα. Ένα χαμηλό τεστ  $\chi^2$  και μια υψηλή p-τιμή υποδεικνύουν καλή προσαρμογή του μοντέλου.

Τα υπόλοιπα πιθανοφάνειας (likelihood residuals) υπολογίζονται εύκολα μέσω των υπολοίπων Pearson και των υπολοίπων Deviance και δίνονται από τον τύπο:

$$(r_i^L)^2 = \hat{h}_{ii} (r_i^{PS})^2 + (1 - \hat{h}_{ii}) (r_i^{DS})^2$$

και λαμβάνουν πρόσημο κατά αντίστοιχο τρόπο με τα υπόλοιπα Deviance δηλαδή με τη βοήθεια της διαφοράς  $y_i - \hat{\mu}_i$ .

### Απόσταση Cook

Για να διερευνήσουμε τον αντίκτυπο μιας παρατήρησης στην προσαρμογή ενός μοντέλου, χρησιμοποιούμε διάφορες μεθόδους, συμπεριλαμβανομένων των υπολοίπων πιθανοφάνειας που περιεγράφηκαν προηγουμένως. Αυτά τα υπόλοιπα πιθανοφάνειας μπορούν να αξιοποιηθούν για αυτόν τον σκοπό. Για την ανάλυση των σημείων που επηρεάζουν την εκτίμηση ενός μοντέλου, μπορούμε να δημιουργήσουμε γραφήματα. Έτσι, μπορούμε να παρουσιάσουμε τα διαγράμματα υπολοίπων πιθανοφάνειας σε σχέση με τα  $\hat{h}_{ii}$  που εκτιμώνται, καθώς επίσης και τα δείκτη (index plots) των υπολοίπων πιθανοφάνειας των εκτιμηθέντων  $\hat{h}_{ii}$  και των αποστάσεων Cook (Cook's distance).

Ένα μέτρο που δείχνει κατά πόσο η αφαίρεση μίας συγκεκριμένης παρατήρησης θα επηρεάσει τις εκτιμήσεις των παραμέτρων ενός μοντέλου είναι η απόσταση του Cook. Η απόσταση Cook είναι ένα μέτρο που χρησιμοποιείται στην ανάλυση της επίδρασης μιας συγκεκριμένης παρατήρησης στην εκτίμηση των παραμέτρων ενός στατιστικού μοντέλου.

Αυτό το μέτρο είναι σχεδιασμένο για να ανιχνεύει παρατηρήσεις που έχουν μεγάλη επίδραση στα αποτελέσματα του μοντέλου και ενδέχεται να είναι ανωμαλίες ή εξαιρέσεις.

Η απόσταση Cook μετράει την αλλαγή στις εκτιμήσεις των παραμέτρων του μοντέλου όταν μια συγκεκριμένη παρατήρηση αφαιρεθεί από το σύνολο δεδομένων. Συγκεκριμένα, για κάθε παρατήρηση, υπολογίζονται οι εκτιμήσεις των παραμέτρων του μοντέλου χωρίς την παρατήρηση αυτή και συγκρίνονται με τις εκτιμήσεις που προκύπτουν όταν η παρατήρηση συμπεριλαμβάνεται. Η απόσταση Cook είναι μια σύνθετη μέτρηση που λαμβάνει υπόψη την αλλαγή σε όλες τις παραμέτρους του μοντέλου.

Όταν η απόσταση Cook για μια συγκεκριμένη παρατήρηση είναι υψηλή, αυτό υποδεικνύει ότι η παρατήρηση αυτή έχει μεγάλη επίδραση στην εκτίμηση των παραμέτρων του μοντέλου. Αυτό μπορεί να οφείλεται σε ανωμαλίες στα δεδομένα, σε παραβιάσεις των υποθέσεων του μοντέλου ή σε άλλα σημαντικά φαινόμενα.

Οι τιμές της απόστασης Cook κυμαίνονται συνήθως μεταξύ μηδενός και μικρών αριθμών. Συνήθως, παρατηρήσεις με απόσταση Cook που είναι πολύ μεγαλύτερη από τη μέση τιμή ή ένα συγκεκριμένο κατώφλι θεωρούνται ως ύποπτες και αξίζει να εξεταστούν προσεκτικά.

Συνολικά, η απόσταση Cook παρέχει σημαντική πληροφορία σχετικά με το πώς μια συγκεκριμένη παρατήρηση επηρεάζει την προσαρμογή του μοντέλου, καθιστώντας τη χρήσιμη στο πλαίσιο της ανάλυσης δεδομένων και της εκτίμησης μοντέλων.

Η στατιστική συνάρτηση που χρησιμοποιείται δίνεται από τον παρακάτω τύπο:

$$CD_i = \frac{1}{p} (\hat{\beta}_{(i)} - \hat{\beta})' I(\hat{\beta}) (\hat{\beta}_{(i)} - \hat{\beta}) \text{ για } i = 1, 2, \dots, n$$

όπου  $\hat{\beta}_{(i)}$  οι εκτιμήσεις των συντελεστών του μοντέλου από το οποίο παραλείπεται i-οστή παρατήρηση από την ανάλυση ενώ  $\hat{\beta}$  οι εκτιμήσεις των συντελεστών του μοντέλου στο οποίο χρησιμοποιείται όλο το δείγμα. Ακόμα,  $p = k+1$  είναι ο αριθμός των παραμέτρων στο μοντέλο και  $I(\hat{\beta}) = X'WX$  είναι η παρατηρούμενη πληροφορία κατά Fisher με  $\hat{V}(\hat{\beta}) = 1^{-1}(\hat{\beta})$ .

Εκτός από τον παραπάνω τύπο, μπορούμε να χρησιμοποιήσουμε έναν πιο απλό τύπο, ο οποίος παρουσιάζεται παρακάτω:

$$CD_i = \frac{\hat{h}_{ii}(r_i^{PS})^2}{p(1 - \hat{h}_{ii})}$$

#### 4.1.7 Κριτήρια επιλογής μοντέλου λογιστικής παλινδρόμησης

##### Κριτήριο AIC

Το κριτήριο AIC (Akaike Information Criterion) είναι ένα κριτήριο επιλογής μοντέλου που χρησιμοποιείται στη στατιστική και την επιστήμη των υπολογιστών για την επιλογή

μοντέλων. Αναπτύχθηκε από τον Ιάπωνα στατιστικό Hirotugu Akaike και είναι βασισμένο στην έννοια της πληροφορίας. Ο στόχος του κριτηρίου AIC είναι να επιλέξει το μοντέλο που παρέχει την καλύτερη ισορροπία μεταξύ καλής προσαρμογής στα δεδομένα και απλότητας του μοντέλου.

Το κριτήριο AIC έχει τη μορφή μιας συνάρτησης κόστους, όπου χαμηλότερες τιμές υποδηλώνουν καλύτερη προσαρμογή του μοντέλου στα δεδομένα. Η επιλογή ενός μοντέλου με χαμηλότερη τιμή AIC σημαίνει ότι παρέχει καλή προσαρμογή στα δεδομένα με ελάχιστη πολυπλοκότητα. Ωστόσο, είναι σημαντικό να σημειωθεί ότι το κριτήριο AIC δεν μπορεί να διακρίνει μεταξύ μοντέλων που παρουσιάζουν παρόμοιες τιμές και επομένως είναι συνήθως χρήσιμο να συγκρίνεται με άλλα κριτήρια επιλογής μοντέλου.

Στα γενικευμένα γραμμικά μοντέλα το κριτήριο AIC δίνεται από τον ακόλουθο τύπο:

$$AIC = -2I(\hat{\beta}) + 2p$$

όπου,  $I(\hat{\beta})$  η μεγιστοποιημένη τιμή της συνάρτησης πιθανοφάνειας και  $p$  ο αριθμός των επεξηγηματικών μεταβλητών που περιέχει το μοντέλο προς εκτίμηση (Akaike, 1974).

Το Κριτήριο Πληροφορίας του Ακριβέστερου Ελάχιστου (AIC) είναι ένα κριτήριο που χρησιμοποιείται για την επιλογή του βέλτιστου μοντέλου με τον ελάχιστο αριθμό παραμέτρων. Από τον τύπο του AIC, παρατηρούμε ότι η προσθήκη παραμέτρων στο μοντέλο, είτε είναι σημαντικές είτε όχι, αυξάνει τον αριθμό των παραμέτρων και την εκτιμώμενη μεταβλητότητα, επομένως ο πρώτος όρος στον τύπο του AIC μειώνεται, ενώ ο δεύτερος αυξάνεται. Κατά συνέπεια, η εισαγωγή περισσότερων παραμέτρων στο μοντέλο, ανεξαρτήτως της σημαντικότητάς τους, μπορεί να οδηγήσει σε μείωση του AIC. Αυτό σημαίνει ότι επιλέγουμε το μοντέλο με τον μικρότερο δυνατότερο όρο AIC, καθώς αυτό υποδεικνύει μια καλύτερη ισορροπία μεταξύ της προσαρμογής του μοντέλου και της πολυπλοκότητάς του.

Το κριτήριο AIC στη λογιστική παλινδρόμηση έχει την εξής μορφή:

$$\begin{aligned} AIC &= -2 \sum_{i=1}^n \left[ \ln \left( \frac{n_i}{y_i} \right) + y_i \ln \hat{p}_i + (n_i - y_i) \ln(1 - \hat{p}_i) \right] + 2p \\ &= 2 \sum_{i=1}^n \left[ n_i \ln(1 + e^{x_i' \hat{\beta}}) - y_i x_i' \hat{\beta} - \ln \left( \frac{n_i}{y_i} \right) \right] + 2p \end{aligned}$$

### **Κριτήριο BIC**

Το Κριτήριο Πληροφορίας του Bayesian (BIC), γνωστό επίσης και ως Κριτήριο Σλεκτιβιστικής Πληροφορίας του Bayesian, είναι ένα κριτήριο που χρησιμοποιείται για την επιλογή του βέλτιστου μοντέλου με τη χρήση της ανακρίβειας σε μοντέλα που εκτιμούνται βάσει δεδομένων. Στο BIC, όπως και στο AIC, ο στόχος είναι να επιλεγεί ένα μοντέλο που

είναι ικανό να προβλέψει τα δεδομένα καλά, ενώ ταυτόχρονα να έχει λιγότερες παραμέτρους για να αποφευχθεί η υπερπροσαρμογή.

Συνολικά, το BIC επιβάλλει μεγαλύτερη ποινή για περισσότερες παραμέτρους σε σχέση με το AIC, προτιμώντας τα πιο απλά μοντέλα. Αυτό σημαίνει ότι το BIC τιμωρεί περισσότερο την υπερπροσαρμογή από το AIC, καθιστώντας το ένα χρήσιμο κριτήριο επιλογής μοντέλων, ειδικά όταν ο αριθμός των παρατηρήσεων είναι σχετικά μικρός σε σχέση με τον αριθμό των παραμέτρων. Συνεπώς, επιλέγουμε το μοντέλο με το μικρότερο BIC, καθώς αυτό υποδεικνύει ένα ισορροπημένο μοντέλο με καλή προσαρμογή στα δεδομένα και χαμηλότερο κίνδυνο υπερπροσαρμογής.

Στα γενικευμένα γραμμικά μοντέλα το κριτήριο BIC δίνεται από τον ακόλουθο τύπο:

$$BIC = -2I(\hat{\beta}) + p \ln(n)$$

όπου,  $I(\hat{\beta})$  η μεγιστοποιημένη τιμή της συνάρτησης πιθανοφάνειας,  $p$  ο αριθμός των εξηγηματικών μεταβλητών που περιέχει το μοντέλο προς εκτίμηση και  $n$  ο αριθμός των παρατηρήσεων.

Η προσθήκη παραμέτρων επηρεάζει την τιμή του κριτηρίου BIC με παρόμοιο τρόπο όπως και την τιμή του κριτηρίου AIC. Πιο συγκεκριμένα, η εισαγωγή παραμέτρων στο μοντέλο, είτε αυτές είναι στατιστικά σημαντικές είτε όχι, προκαλεί αύξηση του  $p$  και του  $(\hat{\beta})$  επομένως ο πρώτος όρος στον τύπο του BIC μειώνεται και ο δεύτερος αυξάνεται. Τελικά η εισαγωγή επιπλέον παραμέτρων στο μοντέλο οι οποίες οδηγούν σε βελτίωση της προσαρμογής του μοντέλου οδηγεί σε μείωση της τιμής του κριτηρίου BIC. Συνεπώς, επιλέγουμε το μοντέλο εκείνο με τη μικρότερη τιμή του κριτηρίου BIC. Η μόνη διαφορά μεταξύ του AIC και του BIC είναι ότι στην περίπτωση του BIC η εισαγωγή επιπρόσθετων παραμέτρων αποθαρρύνεται σε μεγαλύτερο βαθμό από το AIC.

#### 4.1.8 Καμπύλη ROC

Αν  $\hat{p} = \hat{p}(y = 1)$  η εκτιμημένη πιθανότητα επιτυχίας για κάθε μονάδα και  $p_0$  ένα κατώφλι διακρίνουμε τις εξής δύο περιπτώσεις:

- Αν  $\hat{p} > p_0$ , προβλέπεται  $Y=1$  για την μονάδα αυτή
- Αν  $\hat{p} < p_0$ , προβλέπεται  $Y=0$  για την μονάδα αυτή

Έτσι μπορούμε να κατασκευάσουμε τον παρακάτω πίνακα συνάφειας 2 x 2:

		Πραγματική κατάσταση	
		Y=1	Y=0
Πρόβλεψη	Y=1	a	b
	Y=0	c	d

Από αυτόν τον πίνακα ορίζονται οι ακόλουθες ποσότητες:

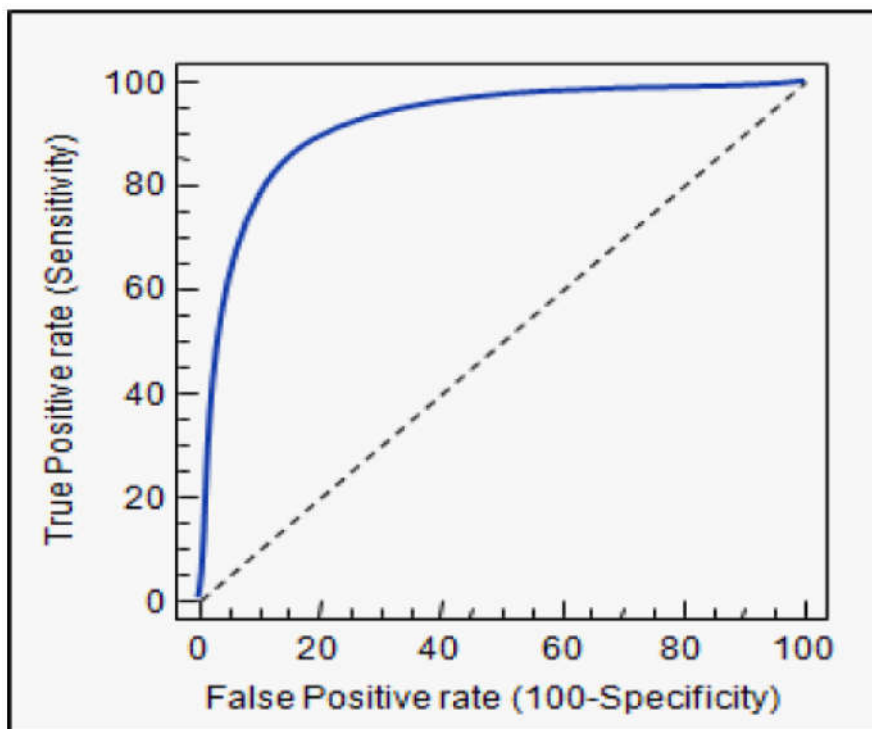
- Ευαισθησία (sensitivity)  $= \frac{a}{a+c}$  , η ευαισθησία εκφράζει το πόσο συχνά προβλέπουμε σωστά ότι είναι  $Y=1$  (true positive rate)
- Ειδικότητα (specificity)  $= \frac{d}{b+d}$  , η ειδικότητα εκφράζει το πόσο συχνά προβλέπουμε σωστά ότι είναι  $Y=0$  (true negative rate)

Με τη βοήθεια της ειδικότητας μπορούμε εύκολα να ορίσουμε μία νέα ποσότητα η οποία καλείται 1-ειδικότητα (false positive rate) και εκφράζει το πόσο συχνά προβλέπουμε λάθος ότι είναι  $Y=1$ .

$$1\text{-Ειδικότητα} = \frac{b}{b+d}$$

Όταν η ευαισθησία λαμβάνει μεγάλες τιμές η 1-ειδικότητα κα λαμβάνει μικρές τιμές και τότε το μοντέλο μας θα έχει υψηλή προβλεπτική ικανότητα.

Ένας τρόπος για να αξιολογήσουμε άμεσα την ικανότητα πρόβλεψης ενός μοντέλου είναι μέσω της κατασκευής της καμπύλης ROC (Receiver Operating Characteristic). Στη γραφική αναπαράσταση αυτής της καμπύλης, ο άξονας x αντιπροσωπεύει την 1-ειδικότητα, ενώ ο άξονας y αντιπροσωπεύει την ευαισθησία. Η καμπύλη ROC είναι η γραμμή που εμφανίζεται στο σχήμα, ενώ η διακεκομμένη γραμμή είναι αυτή όπου ισχύει ότι το ποσοστό αληθώς θετικών αποτελεσμάτων είναι ίσο με το ποσοστό ψευδώς θετικών αποτελεσμάτων. Το εμβαδόν που βρίσκεται κάτω από την καμπύλη ROC ονομάζεται Area Under the Curve (AUC). Όσο πιο κοντά είναι η τιμή του AUC στο 1, τόσο καλύτερη είναι η προβλεπτική ικανότητα του μοντέλου. Το εμβαδόν που δημιουργείται από τη γκρι γραμμή είναι πάντα 0,5. Συνεπώς, η τιμή του AUC κυμαίνεται από 0,5 έως 1.



### 4.1.9 Πολλαπλή Λογιστική Παλινδρόμηση

Πολλές φορές έχουμε περισσότερες από μια ανεξάρτητη μεταβλητή οπότε θα πρέπει να προεκτείνουμε το απλό λογιστικό μοντέλο σε πολλαπλό μοντέλο.

Το πολλαπλό λογιστικό μοντέλο είναι:

$$Y_i = E(Y_i) + \varepsilon_i = \frac{e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1}}} + \varepsilon_i$$

για να διευκολυνθούμε με τις πράξεις αλλά και για την απλοποίηση των σχέσεων θα χρησιμοποιήσουμε πίνακες και διανύσματα:

$$\underset{p \times 1}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{pmatrix}, \quad \underset{p \times 1}{X} = \begin{pmatrix} 1 \\ X_1 \\ X_2 \\ \vdots \\ X_{p-1} \end{pmatrix} \quad \text{και} \quad \underset{p \times 1}{X_i} = \begin{pmatrix} 1 \\ X_{i1} \\ X_{i2} \\ \vdots \\ X_{i,p-1} \end{pmatrix}.$$

οπότε θα έχουμε:

$$\beta' X = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} \quad \text{και} \quad \beta' X_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i,p-1}$$

Με τη βοήθεια των πιο πάνω και με τα  $Y_i$  να είναι ανεξάρτητες μεταβλητές Bernoulli μπορούμε να γράψουμε την αναμενόμενη λογιστική ως εξής:

$$E(Y_i) = \pi_i = \frac{e^{\beta' X_i}}{1 + e^{\beta' X_i}}$$

Παρατηρήσεις:

1. Είναι σημαντικό να τονίσουμε εδώ ότι όλες οι σχέσεις που είχαμε παρατηρήσει στο απλό λογιστικό μοντέλο επεκτείνονται και στο πολλαπλό λογιστικό μοντέλο.
2. Επιπλέον, αξίζει να αναφέρουμε ότι και σε αυτό το πλαίσιο, η αναμενόμενη λογιστική συνάρτηση παραμένει μονότονη, με την καμπύλη να λαμβάνει σχήμα "s" ή ανάποδο "s" ως προς το διάνυσμα  $X$ . Επιπλέον, η καμπύλη είναι σχεδόν γραμμική όταν η πιθανότητα  $E(Y)$  βρίσκεται εντός του διαστήματος 0.2 έως 0.8.

Οι μεταβλητές  $X$  μπορεί να είναι είτε ποσοτικές είτε ποιοτικές. Στην περίπτωση των ποιοτικών μεταβλητών, αναπαρίστανται ως δυαδικές μεταβλητές. Σε περίπτωση που όλες οι μεταβλητές είναι ποιοτικές, μπορούμε επίσης να αναφερόμαστε σε αυτό ως λογαριθμικό μοντέλο.

Όπως και στο απλό λογιστικό μοντέλο έτσι και εδώ η εκτίμηση παραμέτρων θα γίνει με τη βοήθεια της μεθόδου μέγιστης πιθανοφάνειας. Οπότε η λογαριθμική συνάρτηση πιθανοφάνειας είναι:

$$\ln L(\beta) = \sum_{i=1}^n Y_i(\beta' X_i) - \sum_{i=1}^n \ln(1 + e^{\beta' X_i})$$

Για να βρούμε τους εκτιμητές θα πρέπει και πάλι να χρησιμοποιήσουμε κάποια αριθμητική μέθοδο η οποία θα μας δίνει τις τιμές των  $\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}$  που θα μεγιστοποιούν την παραπάνω σχέση. Τις τιμές αυτές θα τις συμβολίζουμε με το μοναδιαίο διάνυσμα

$$b_{p \times 1} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix}$$

Οπότε η προσαρμοσμένη λογιστική συνάρτηση απόκρισης θα έχει την μορφή:

$$\hat{\pi} = \frac{e^{\beta' X}}{1 + e^{\beta' X}} = \left[ 1 + e^{-\beta' X} \right]^{-1}$$

Παρατηρήσεις:

1. Προηγουμένως αναφερθήκαμε στο γεγονός ότι οι εκτιμητές υπολογίζονται με βάση αριθμητικές μεθόδους, με δυνατότητα πιθανού προβλήματος σύγκλισης. Αυτό ενδέχεται να συμβεί όταν οι μεταβλητές πρόβλεψης είναι υπερβολικά πολλές ή όταν κάποιες από αυτές έχουν υψηλή συσχέτιση. Σε τέτοιες περιπτώσεις, είναι αναγκαίο να μειώσουμε τον αριθμό των μεταβλητών πρόβλεψης.
2. Οι εκτιμητές μέγιστης πιθανοφάνειας μπορούν να υπολογιστούν επίσης με τη χρήση της μεθόδου των επαναλαμβανόμενων σταθμισμένων ελαχίστων τετραγώνων.
3. Όταν η λογιστική συνάρτηση δεν εμφανίζει μονοτονία ή δεν ακολουθεί τη μορφή "s", απαιτείται η μετατροπή όλων των μεταβλητών πρόβλεψης σε κατηγορικές. Σε αυτήν την περίπτωση, είναι προτιμητέο να χρησιμοποιηθεί το λογαριθμικό μοντέλο.

### Ερμηνεία των Συντελεστών Παλινδρόμησης

Όσον αφορά τις ερμηνείες των συντελεστών παλινδρόμησης, ο συντελεστής  $\beta_0$  αναπαριστά το επίπεδο βάσης ή το αρχικό επίπεδο πιθανότητας για την επιτυχημένη έκβαση ενός γεγονότος. Κάθε συντελεστής  $\beta_i$ , αντιστοιχεί στην ανεξάρτητη μεταβλητή  $i$  και εκφράζει το μέγεθος της συνεισφοράς της στην πιθανότητα της επιτυχημένης έκβασης.

1. Η θετική τιμή του συντελεστή δηλώνει ότι η ανεξάρτητη μεταβλητή αυξάνει την πιθανότητα επιτυχούς έκβασης ενός γεγονότος, ενώ η αρνητική τιμή υποδεικνύει μείωση της πιθανότητας αυτής της έκβασης.

2. Η υψηλή τιμή του συντελεστή υποδεικνύει ότι η σχετική ανεξαρτησία μεταβλητή έχει ισχυρή επίδραση στην πιθανότητα εμφάνισης του γεγονότος ή όχι, ενώ η χαμηλή τιμή υποδεικνύει μικρή επίδραση της ανεξάρτητης μεταβλητής σε αυτήν την πιθανότητα.

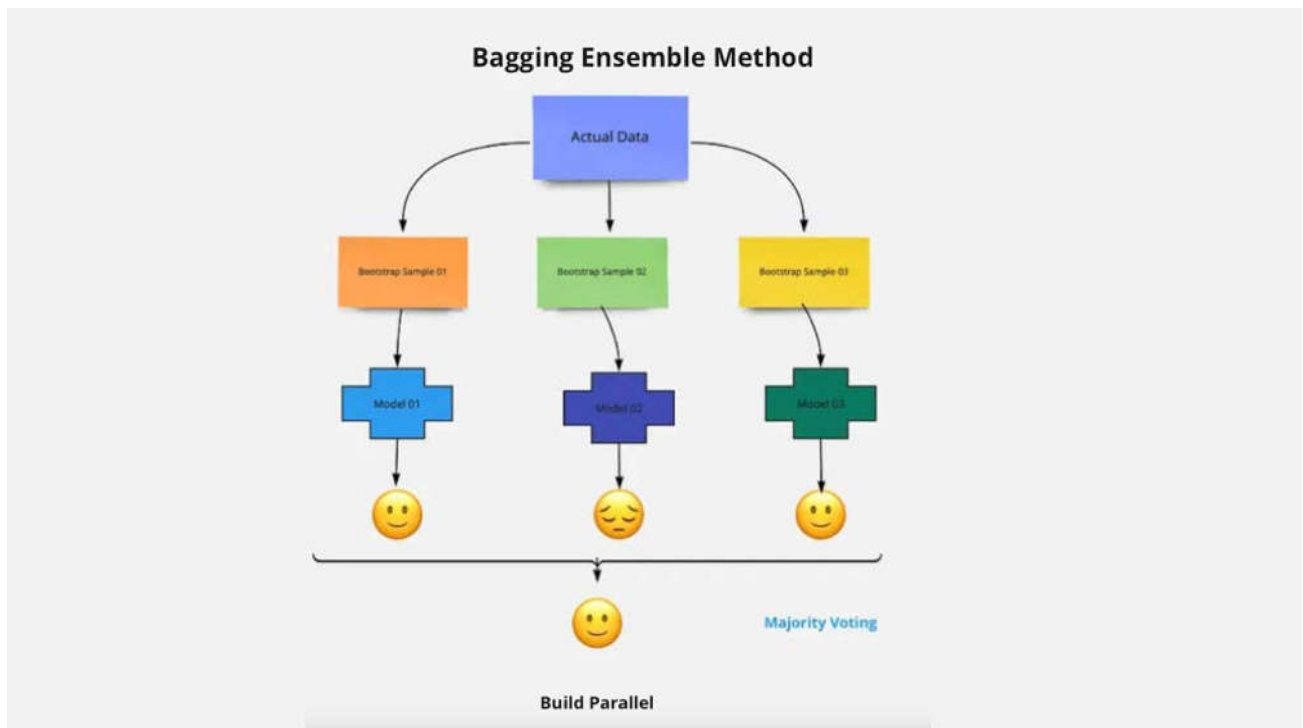
Ισχύει και πάλι πως  $e^{bk}$  είναι ο εκτιμώμενος λόγος πιθανοτήτων για την μεταβλητή  $X_k$ , υπό την προϋπόθεση ότι οι άλλες μεταβλητές πρόβλεψης είναι σταθερές και ότι το πολλαπλό μοντέλο παλινδρόμησης είναι πρώτης τάξεως και δεν περιέχει τετραγωνικούς ή μεγαλύτερου βαθμού όρους για τις μεταβλητές πρόβλεψης. Όμως, όταν το πολλαπλό μοντέλο λογιστικής παλινδρόμησης δεν είναι ένα πρώτης τάξεως αλλά περιέχει τετραγωνικούς ή μεγαλύτερου βαθμού όρους για τις μεταβλητές πρόβλεψης, οι εκτιμώμενοι συντελεστές παλινδρόμησης δεν έχουν πλέον μια απλή ερμηνεία.

## 4.2 Απόφαση Δέντρων και Τυχαία Δάση (Decision Trees, Random Forests)

Ο Random Forest είναι ένας από τους πιο διαδεδομένους αλγόριθμους μηχανικής μάθησης για εποπτευόμενα προβλήματα ταξινόμησης και παλινδρόμησης. Πρόκειται για μια μέθοδο ensemble learning (μεθόδου συνόλου), η οποία συνδυάζει πολλαπλά δέντρα απόφασης σε ένα ενιαίο μοντέλο για να βελτιώσει την ακρίβεια και τη γενικευσιμότητα των προβλέψεων. Η βασική ιδέα πίσω από το Random Forest είναι ότι μια «επιτροπή» πολλών ασυσχέτιστων μοντέλων μπορεί να παράγει πιο αξιόπιστες προβλέψεις από ό,τι ένα μόνο μοντέλο – ένα φαινόμενο γνωστό ως “wisdom of crowds”. Ο αλγόριθμος αναπτύχθηκε από τον Leo Breiman το 2001, βασισμένος στην τεχνική του bagging και στην τυχαία επιλογή χαρακτηριστικών. Έκτοτε, ο Random Forest έχει γίνει ιδιαίτερα δημοφιλής λόγω της αποτελεσματικότητάς του και της ευκολίας χρήσης του (είναι συχνά «έτοιμος από το κουτί» με ελάχιστη ρύθμιση υπερπαραμέτρων), καθώς και της ευελιξίας του να αντιμετωπίζει τόσο προβλήματα ταξινόμησης όσο και παλινδρόμησης.

Ένα βασικό πλεονέκτημα των τυχαίων δασών είναι ότι τείνουν να μην υπερεφαρμόζουν (overfit) όσο αυξάνεται ο αριθμός των δέντρων. Αντίθετα με ένα απλό δέντρο απόφασης που μπορεί εύκολα να υπερεφαρμόσει στα δεδομένα εκπαίδευσης, ένα Random Forest με αρκετά δέντρα συγκλίνει σε ένα ορισμένο σφάλμα γενίκευσης καθώς προστίθενται περισσότερα δέντρα, χωρίς να χειροτερεύει η επίδοση (Breiman, 2001). Αυτό το χαρακτηριστικό τα καθιστά ιδιαίτερα ελκυστικά για πρακτικές εφαρμογές, όπου η αξιοπιστία και η σταθερότητα των προβλέψεων είναι κρίσιμης σημασίας.

Ο αλγόριθμος Random Forest ανήκει στην κατηγορία των μεθόδων ensemble learning – τεχνικές όπου εκπαιδεύονται πολλαπλά μοντέλα και οι προβλέψεις τους συνδυάζονται. Ο συγκεκριμένος αλγόριθμος βασίζεται ειδικότερα στη μέθοδο Bagging (Bootstrap Aggregating) που πρότεινε ο Breiman (1996). Στο bagging, δημιουργούμε πολλά σύνολα εκπαίδευσης αντλώντας τυχαιοποιημένα δείγματα με επανάθεση (bootstrap samples) από το αρχικό σύνολο δεδομένων. Κάθε τέτοιο δείγμα χρησιμοποιείται για να εκπαιδευτεί ένα ξεχωριστό βασικό μοντέλο (στην περίπτωσή μας, ένα δέντρο απόφασης). Στη συνέχεια, για να παραχθεί η τελική πρόβλεψη, συνδυάζονται οι προβλέψεις όλων των μοντέλων της επιτροπής – για ταξινόμηση με πλειοψηφική ψήφο (majority voting) και για παλινδρόμηση με μέσο όρο των αριθμητικών προβλέψεων. Με τον τρόπο αυτό, το μοντέλο εκμεταλλεύεται την «σοφία του πλήθους», όπου τα σφάλματα των επιμέρους δέντρων (που είναι αδύναμοι μαθητές ή weak learners) τείνουν να ακυρώνονται μεταξύ τους, οδηγώντας σε πιο σταθερές και ακριβείς προβλέψεις συνολικά (Breiman, 2001).



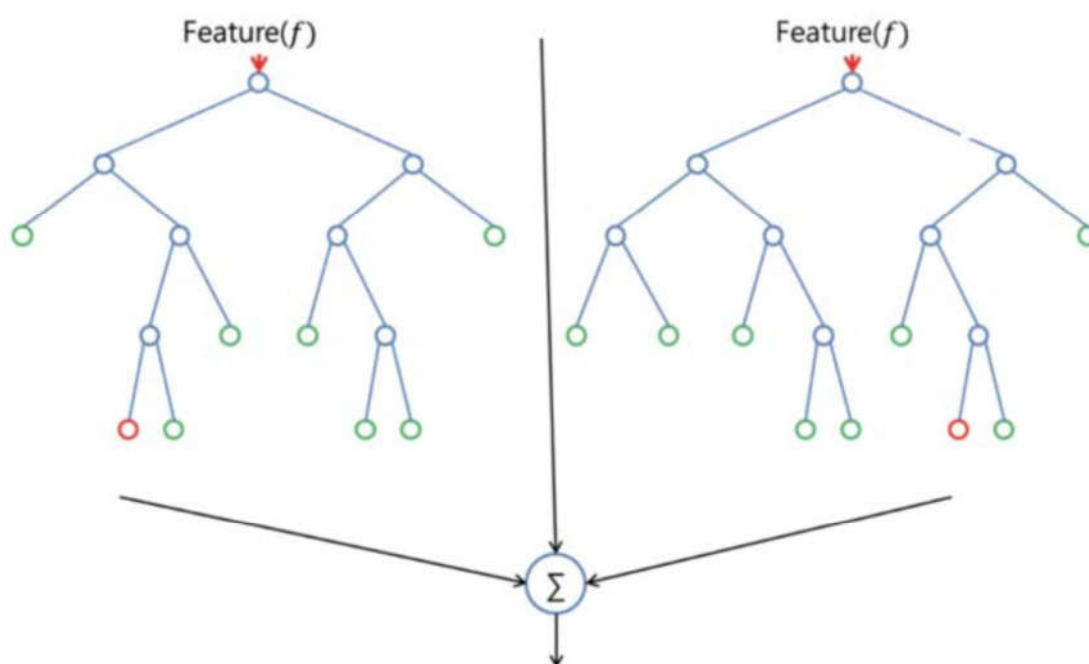
Αρχή του bagging σε ένα ensemble δέντρων. Το «Πραγματικό Σύνολο Δεδομένων» δίνει τρία bootstrap δείγματα (Bootstrap Sample 01, 02, 03), πάνω στα οποία εκπαιδεύονται ανεξάρτητα τρία διαφορετικά μοντέλα (Model 01, 02, 03). Κάθε μοντέλο παράγει μια πρόβλεψη (εδώ εικονίζεται ως χαρούμενο ή λυπημένο emoticon). Τελικά, μέσω παράλληλης εκπαίδευσης και πλειοψηφικής ψήφου (Majority Voting), το ensemble δίνει την τελική πρόβλεψη (χαρούμενο emoticon). Η συλλογική απόφαση των πολλών μοντέλων είναι πιο αξιόπιστη από μιας μεμονωμένης πρόβλεψης.

Στον Random Forest, τα μεμονωμένα μοντέλα του ensemble είναι δέντρα απόφασης τύπου CART (Classification and Regression Trees). Τα δέντρα απόφασης από τη φύση τους παρουσιάζουν υψηλή διασπορά: μικρές αλλαγές στα δεδομένα εκπαίδευσης μπορούν να οδηγήσουν σε διαφορετική δομή δέντρου. Η χρήση του bagging μειώνει δραστηρικά αυτή τη διασπορά, συνδυάζοντας πολλά διαφορετικά δέντρα. Όμως, ένα κρίσιμο στοιχείο για την επιτυχία του Random Forest είναι να μεγιστοποιηθεί η ποικιλία (απουσία συσχέτισης) μεταξύ των δέντρων του δάσους. Εάν όλα τα δέντρα ήταν πανομοιότυπα, δεν θα προσέφεραν κάτι περισσότερο από ένα μόνο δέντρο. Η τυχαιοποίηση των δεδομένων μέσω bootstrapping βοηθά στην διαφοροποίηση, αλλά ο Breiman εισήγαγε και μια επιπλέον πηγή τυχειότητας: την τυχαία δειγματοληψία χαρακτηριστικών σε κάθε διαχωρισμό του δέντρου (Breiman, 2001).

Ένα από τα στοιχεία-κλειδιά του αλγορίθμου Random Forest είναι ότι, κατά την ανάπτυξη κάθε δέντρου, σε κάθε κόμβο διαχωρισμού δεν εξετάζονται όλοι οι διαθέσιμοι μεταβλητές/χαρακτηριστικά, αλλά ένα τυχαίο υποσύνολο χαρακτηριστικών. Δηλαδή, όταν ένα δέντρο αποφασίζει πώς θα διασπάσει (split) έναν κόμβο, αντί να υπολογίσει το βέλτιστο διαχωρισμό ανάμεσα σε όλα τα χαρακτηριστικά, επιλέγει τυχαία ένα μικρό αριθμό από αυτά (π.χ.  $\sqrt{d}$  χαρακτηριστικά αν υπάρχουν  $d$  συνολικά) και βρίσκει τον καλύτερο διαχωρισμό μόνο εντός αυτού του υποσυνόλου. Αυτή η προσέγγιση, γνωστή και ως random subspace method (τυχαίος υποχώρος χαρακτηριστικών), προτάθηκε αρχικά από τον Tin Kam Ho (1998) για την κατασκευή τυχαίων δασών απόφασης και αποδείχθηκε κρίσιμη για τη μείωση

της συσχέτισης μεταξύ των δέντρων (Ho, 1998). Με τον τυχαίο περιορισμό των χαρακτηριστικών που βλέπει κάθε διαχωρισμός, διασφαλίζεται ότι τα δέντρα θα είναι πιο διαφορετικά μεταξύ τους, ακόμη και αν κάποια χαρακτηριστικά είναι ιδιαίτερα ισχυρά στο να προβλέπουν το αποτέλεσμα.

Για παράδειγμα, αν σε ένα σύνολο δεδομένων ένα συγκεκριμένο χαρακτηριστικό είναι το πιο ισχυρό προγνωστικό, ένα απλό δέντρο ή το bagging χωρίς τυχαία επιλογή χαρακτηριστικών πιθανόν να το χρησιμοποιεί σχεδόν πάντα στον ριζικό κόμβο, κάνοντας όλα τα δέντρα παρόμοια. Με την τυχαία δειγματοληψία, κάποια δέντρα δεν θα βλέπουν αυτό το χαρακτηριστικό σε ορισμένους κόμβους, αναγκάζοντάς τα να χρησιμοποιήσουν εναλλακτικά χαρακτηριστικά. Το αποτέλεσμα είναι πιο ασυσχέτιστα δέντρα, τα οποία σε συνδυασμό δίνουν καλύτερη απόδοση. Σύμφωνα με τον Breiman (2001), ο συνδυασμός υψηλής «ισχύος» (strength) στα επιμέρους δέντρα και χαμηλής συσχέτισης μεταξύ τους είναι που οδηγεί σε χαμηλό γενικό σφάλμα του δάσους.



Απλοποιημένη απεικόνιση ενός Random Forest με δύο δέντρα. Κάθε δέντρο (αριστερά και δεξιά) εκπαιδεύεται σε ένα διαφορετικό δείγμα και πιθανόν διαφορετικό υποσύνολο χαρακτηριστικών, παράγοντας τις δικές του προβλέψεις (π.χ. οι πράσινοι και κόκκινοι τετραγωνικοί κόμβοι αναπαριστούν διαφορετικές ταξινομήσεις). Οι προβλέψεις των δύο δέντρων συνδυάζονται τελικά (π.χ. μέσω ενός κόμβου που εδώ συμβολίζει το άθροισμα ή τον μέσο όρο  $\Sigma$ ), δίνοντας μια συνολική πρόβλεψη. Το τυχαίο δάσος συγκεντρώνει έτσι τις αποφάσεις πολλών δέντρων, αυξάνοντας την ακρίβεια και τη σταθερότητα του μοντέλου.

Συνοψίζοντας τις παραπάνω έννοιες, η διαδικασία εκπαίδευσης ενός Random Forest για ταξινόμηση έχει ως εξής:

1. **Δημιουργία Δειγμάτων Bootstrap:** Από το αρχικό σύνολο εκπαίδευσης που περιέχει  $N$  παρατηρήσεις, δημιουργούμε  $B$  νέα σύνολα δεδομένων (όπου  $B$  ο αριθμός των δέντρων που θα φτιάξουμε), καθένα από τα οποία προκύπτει επιλέγοντας  $N$  παρατηρήσεις τυχαία με επαναληπτική δειγματοληψία (bootstrap). Κάθε τέτοιο δείγμα θα περιέχει περίπου το 63% των μοναδικών αρχικών παρατηρήσεων κατά μέσο όρο (καθώς κάποιες επιλέγονται πολλές φορές και κάποιες καθόλου).
2. **Εκπαίδευση Δέντρων:** Για κάθε bootstrap δείγμα, εκπαιδεύεται ένα δέντρο απόφασης. Το δέντρο κατασκευάζεται χωρίς κλάδεμα (unpruned), επιτρέποντας πλήρες βάθος ή μέχρι να ικανοποιηθεί κάποιο όριο ελάχιστων δειγμάτων ανά φύλλο. Κατά την κατασκευή κάθε κόμβου του δέντρου, επιλέγεται τυχαία ένα υποσύνολο  $m$  χαρακτηριστικών από τα  $d$  διαθέσιμα (με  $m < d$ , τυπικά  $m = \sqrt{d}$  για ταξινόμηση ή  $m = d/3$  για παλινδρόμηση, αν και αυτές οι τιμές μπορεί να διαφοροποιούνται). Από αυτά τα  $m$  υποψήφια χαρακτηριστικά, βρίσκουμε εκείνο και το αντίστοιχο κατώφλι διαχωρισμού που μεγιστοποιεί τον διαχωρισμό των δεδομένων στον κόμβο (σύμφωνα με κάποιο κριτήριο ποιότητας διαχωρισμού, π.χ. Gini ή εντροπία – βλέπε επόμενη ενότητα). Χρησιμοποιούμε αυτόν τον διαχωρισμό για να δημιουργήσουμε δύο θυγατρικούς κόμβους και επαναλαμβάνουμε αναδρομικά τη διαδικασία σε κάθε νέο κόμβο μέχρι το δέντρο να ολοκληρωθεί (π.χ. όταν κάθε φύλλο είναι "καθαρό" ή όταν έχει φτάσει το μέγιστο επιτρεπτό βάθος).
3. **Συνδυασμός Προβλέψεων (Voting/Averaging):** Μετά την εκπαίδευση όλων των δέντρων, για να ταξινομηθεί μια νέα παρατήρηση, αυτή περνάει από κάθε δέντρο του δάσους, όπου λαμβάνεται η πρόβλεψη του κάθε δέντρου (η κλάση-έξοδος του φύλλου στο οποίο καταλήγει). Οι προβλέψεις όλων των δέντρων συλλέγονται και η τελική απόφαση του Random Forest δίνεται με πλειοψηφία των ψήφων (η κλάση που προβλέφθηκε συχνότερα ανάμεσα στα δέντρα). Στην περίπτωση παλινδρόμησης, αντί για πλειοψηφία, υπολογίζεται ο μέσος όρος των αριθμητικών τιμών που προβλέπουν τα δέντρα (Breiman, 2001). Αυτός ο τρόπος συνδυασμού διασφαλίζει ότι αν τα περισσότερα δέντρα έχουν μάθει σωστά τα μοτίβα, το συνολικό αποτέλεσμα θα είναι σωστό, ενώ μεμονωμένα λάθη κάποιων δέντρων δεν θα επηρεάσουν σημαντικά.

Η παραπάνω διαδικασία παράγει ένα ισχυρό μοντέλο με χαμηλό σφάλμα γενίκευσης. Θεωρητικά, όπως ανέλυσε ο Breiman (2001), το γενικό σφάλμα ενός Random Forest εξαρτάται από την μέση ακρίβεια των μεμονωμένων δέντρων (ονομάζεται "strength") και από τη μέση συσχέτισή τους: όσο υψηλότερη η ακρίβεια των δέντρων και όσο χαμηλότερη η μεταξύ τους συσχέτιση, τόσο χαμηλότερο το συνολικό σφάλμα. Η τυχαία δειγματοληψία (στα δείγματα και στα χαρακτηριστικά) στοχεύει ακριβώς σε αυτό – στην κατασκευή αριετών ισχυρών δέντρων που όμως να είναι διαφορετικά μεταξύ τους.

Κατά την ανάπτυξη κάθε δέντρου στο δάσος, απαιτείται ένα μέτρο για την επιλογή του βέλτιστου διαχωρισμού σε κάθε κόμβο. Τα πιο διαδεδομένα κριτήρια διάσπασης στα δέντρα

απόφασης είναι δύο: η αντιβύθιση Gini (Gini impurity) και το κριτήριο της Εντροπίας (Entropy) με βάση το Information Gain.

### **Gini Impurity**

Η Gini impurity χρησιμοποιείται παραδοσιακά στον αλγόριθμο CART (Classification and Regression Trees) (Breiman et al., 1984). Εκφράζει την πιθανότητα ένα τυχαία επιλεγμένο στοιχείο από τον κόμβο να ταξινομηθεί λανθασμένα, εφόσον βασιστούμε στην τυχαία επιλογή της ετικέτας σύμφωνα με την κατανομή των ετικετών στον κόμβο. Ο τύπος της Gini impurity για έναν κόμβο είναι:

$$\text{Gini} = 1 - \sum_{k=1}^K p_k^2$$

Όπου:

- K είναι ο αριθμός των κατηγοριών,
- $p_k$  είναι η σχετική συχνότητα (πιθανότητα) των δειγμάτων που ανήκουν στην κατηγορία k στον συγκεκριμένο κόμβο.

Η Gini παίρνει τιμές:

- 0, όταν όλα τα δείγματα του κόμβου ανήκουν στην ίδια κατηγορία (πλήρης καθαρότητα),
- Μέγιστη τιμή (έως και  $1 - (1/K)$ ) όταν τα δείγματα κατανέμονται ισομερώς μεταξύ όλων των κατηγοριών (μέγιστη ακαθαρσία).

Στόχος του αλγορίθμου είναι να επιλέγει διαχωρισμούς που μειώνουν την συνολική Gini impurity στους θυγατρικούς κόμβους. Δηλαδή, ένας καλός διαχωρισμός επιτυγχάνει μεγάλη μείωση της Gini, γνωστή και ως Gini Gain.

### **Entropy / Information Gain**

Η Εντροπία (Entropy) αποτελεί ένα μέτρο αβεβαιότητας ή καθαρότητας, το οποίο χρησιμοποιείται στους αλγορίθμους ID3 και C4.5 (Quinlan, 1986). Όπως και η Gini, μετρά την "ανάμειξη" των κατηγοριών μέσα σε έναν κόμβο. Ο μαθηματικός τύπος της εντροπίας είναι:

$$\text{Entropy} = - \sum_{k=1}^K p_k \log_2 p_k$$

Όπου:

- $p_k$  είναι η πιθανότητα της κατηγορίας k στον κόμβο.

Η εντροπία είναι:

- 0, όταν όλα τα δείγματα ανήκουν στην ίδια κατηγορία,
- Μέγιστη όταν όλες οι κατηγορίες έχουν ίσες πιθανότητες (ισοκατανομή:  $p_k=1/K$ ).

Η Πληροφοριακή Απόδοση (Information Gain) μετρά τη μείωση της εντροπίας μετά από έναν διαχωρισμό και ορίζεται ως:

$$\text{Information Gain} = \text{Entropy}_{\text{parent}} - \sum_{i=1}^n \frac{N_i}{N} \cdot \text{Entropy}_i$$

Όπου:

- $N$  είναι το πλήθος των δειγμάτων στον αρχικό (γονικό) κόμβο,
- $N_i$  είναι το πλήθος των δειγμάτων στον  $i$ -οστό θυγατρικό κόμβο,
- $\text{Entropy}_i$  είναι η εντροπία του  $i$ -οστού θυγατρικού κόμβου.

Ο διαχωρισμός που έχει το υψηλότερο Information Gain θεωρείται ο βέλτιστος, καθώς μειώνει την αβεβαιότητα και αυξάνει την "καθαρότητα" του δέντρου.

Στην πράξη, τα κριτήρια Gini και εντροπίας συχνά οδηγούν σε παρόμοιες αποφάσεις διαχωρισμού. Η διαφορά τους είναι ότι η εντροπία δίνει ελαφρώς μεγαλύτερο βάρος στα σπάνια αποτελέσματα σε σχέση με το Gini. Το Gini, από την άλλη, είναι υπολογιστικά πιο απλό (δεν περιλαμβάνει λογαρίθμους). Στο scikit-learn και σε πολλές άλλες υλοποιήσεις Random Forest, το προεπιλεγμένο κριτήριο για ταξινόμηση είναι το Gini, ενώ υπάρχει δυνατότητα επιλογής της εντροπίας ως εναλλακτική. Για προβλήματα παλινδρόμησης, το ανάλογο μέτρο διασποράς είναι συνήθως το μέσο τετραγωνικό σφάλμα (MSE) σε κάθε κόμβο, αλλά στην περίπτωση του Random Forest παλινδρόμησης, ο μηχανισμός παραμένει παρόμοιος (μέση τιμή στις φύλλοι, μείωση του MSE κ.λπ.).

### **Σφάλμα Εκτός Δείγματος (Out-of-Bag Error)**

Μια σημαντική πτυχή του Random Forest είναι ότι προσφέρει έναν ενσωματωμένο τρόπο εκτίμησης της απόδοσης του, χωρίς την ανάγκη ξεχωριστού συνόλου επικύρωσης. Αυτό γίνεται μέσω του σφάλματος εκτός δείγματος (Out-of-Bag, OOB error). Θυμίζουμε ότι κατά το σχηματισμό κάθε δέντρου, χρησιμοποιείται ένα bootstrap δείγμα από τα δεδομένα εκπαίδευσης. Κατά μέσο όρο, περίπου το 1/3 των αρχικών παρατηρήσεων δεν επιλέγονται σε ένα συγκεκριμένο bootstrap δείγμα – αυτές οι παρατηρήσεις αποτελούν το "out-of-bag" σύνολο για το αντίστοιχο δέντρο. Έτσι, για κάθε δέντρο μπορούμε να έχουμε μια ανεξάρτητη δοκιμή: χρησιμοποιούμε το δέντρο για να προβλέψουμε τις ετικέτες των παρατηρήσεων που δεν είδε κατά την εκπαίδευσή του και καταγράφουμε την ακρίβειά του.

Συγκεντρώνοντας τα αποτελέσματα όλων των δέντρων, μπορούμε να υπολογίσουμε την συνολική απόδοση OOB του δάσους ως εξής: για κάθε παρατήρηση του συνόλου εκπαίδευσης, λαμβάνουμε τις προβλέψεις μόνο από τα δέντρα που δεν την είδαν στο training

τους και αποφασίζουμε με πλειοψηφία (ή μέσο όρο για παλινδρόμηση) την πρόβλεψη OOB για αυτήν την παρατήρηση. Συγκρίνοντας με την πραγματική τιμή, μπορούμε να χαρακτηρίσουμε την παρατήρηση ως σωστή ή λάθος. Το συνολικό OOB accuracy (ή OOB error rate) είναι το ποσοστό των OOB προβλέψεων που είναι σωστές (ή λανθασμένες).

Το σφάλμα OOB αποτελεί μια πολύ χρήσιμη εκτίμηση του σφάλματος γενίκευσης του μοντέλου. Στην πράξη έχει βρεθεί ότι το OOB error προσεγγίζει πολύ καλά την απόδοση σε ένα ανεξάρτητο σύνολο test, καθιστώντας περιττή την ανάγκη για ξεχωριστή διαδικασία cross-validation σε πολλά cases (Breiman, 1996). Έτσι, ο Random Forest “ενσωματώνει” έναν τρόπο επικύρωσης κατά την ίδια την εκπαίδευσή του. Επιπλέον, το OOB επιτρέπει τον υπολογισμό σπουδαιότητας χαρακτηριστικών (feature importance) μέσω Permutation Importance: μπορούμε να μετρήσουμε πόσο αυξάνεται το OOB error όταν τυχαιοποιούμε (κάνουμε permute) τις τιμές ενός συγκεκριμένου χαρακτηριστικού, κρατώντας όλα τα άλλα ίδια. Μια μεγάλη αύξηση του σφάλματος δηλώνει ότι το χαρακτηριστικό ήταν σημαντικό για τις σωστές προβλέψεις (Breiman, 2001).

Εν κατακλείδι, ο αλγόριθμος Random Forest συνδυάζει: (α) πολλαπλά δέντρα απόφασης εκπαιδευμένα σε τυχαία δείγματα (bagging) των δεδομένων, (β) τυχαία υποσύνολα χαρακτηριστικών σε κάθε διαχωρισμό και (γ) συνδυασμό προβλέψεων μέσω ψήφου ή μέσου όρου. Αυτές οι αρχές ensemble learning προσδίδουν στο μοντέλο υψηλή ακρίβεια, ανθεκτικότητα στον θόρυβο και μειωμένο κίνδυνο υπερεφαρμογής. Στην επόμενη ενότητα, θα συγκρίνουμε το Random Forest με έναν απλό ταξινομητή δέντρου καθώς και με άλλους αλγορίθμους μηχανικής μάθησης, όπως τις Υποστηρικτικές Διανυσματικές Μηχανές (SVM) και τα μοντέλα boosting.

#### ***4.2.1 Σύγκριση με Δέντρα Απόφασης και Άλλους Αλγορίθμους***

##### **Random Forest vs. Απόμονο Δέντρο Απόφασης**

Τα δέντρα απόφασης (Decision Trees) αποτελούν τη βάση πάνω στην οποία χτίζεται το Random Forest. Ένα απόμονο δέντρο απόφασης εκπαιδεύεται χωρίς καμία διαδικασία ensemble: προσπαθεί να μάθει απευθείας από τα δεδομένα, κάνοντας διασπάσεις που βελτιώνουν όσο το δυνατόν περισσότερο την καθαρότητα στους κόμβους. Το πλεονέκτημα των δέντρων είναι η απλότητα και η ερμηνευσιμότητά τους – το μοντέλο μπορεί να απεικονιστεί και να εξηγηθεί εύκολα (οι κανόνες των διασπάσεων είναι ανθρώπινα κατανοητοί). Ωστόσο, τα δέντρα έχουν την τάση να εμφανίζουν υψηλή διασπορά: μπορούν να υπερεφαρμόσουν αν δεν κλαδευτούν σωστά, απομνημονεύοντας ιδιομορφίες του training set και αποτυγχάνοντας σε νέα δεδομένα. Ένα επίσης μειονέκτημα είναι ότι η απόδοσή τους (accuracy) συχνά υπολείπεται άλλων πιο εξελιγμένων μεθόδων σε περίπλοκα προβλήματα.

Το Random Forest υπερνικά αυτά τα προβλήματα συνδυάζοντας πολλά δέντρα. Ακριβεία: Σε γενικές γραμμές, ένα Random Forest σχεδόν πάντα θα παρουσιάσει καλύτερη ακρίβεια γενίκευσης από ένα μεμονωμένο δέντρο στο ίδιο πρόβλημα, διότι μειώνει το πρόβλημα της διασποράς μέσω του averaging (Breiman, 2001). Υπερεφαρμογή: Ενώ ένα βαθύ δέντρο μπορεί να υπερεφαρμόσει, το Random Forest είναι πολύ πιο ανθεκτικό στο overfitting. Όπως αναφέρθηκε, προσθέτοντας περισσότερα δέντρα, το σφάλμα τείνει σε ένα όριο αντί να αυξάνεται – κάτι που δεν ισχύει αν προσθέτουμε περισσότερη πολυπλοκότητα σε ένα μόνο δέντρο (Breiman, 2001). Ερμηνευσιμότητα: Το αντίτιμο βέβαια είναι η μειωμένη ερμηνευσιμότητα. Ένα δάσος με εκατοντάδες δέντρα είναι δύσκολο να ερμηνευτεί συνολικά, αν και παρέχει εργαλεία όπως οι σπουδαιότητες χαρακτηριστικών. Για λόγους ερμηνείας, συχνά μπορεί να εξετάσει κανείς μεμονωμένα δέντρα του δάσους ή να χρησιμοποιήσει μεθόδους όπως μερικές εξαρτήσεις (partial dependence plots) για να κατανοήσει τη συμπεριφορά του μοντέλου.

Επιπλέον, το Random Forest, σε αντίθεση με ένα μόνο δέντρο, παρέχει το ενδεικτικό σφάλμα OOB και μπορεί να χρησιμοποιηθεί για ανίχνευση ανωμαλιών μέσω της έννοιας της μέσης απόστασης (proximity) μεταξύ δειγμάτων στα δέντρα – κάτι που ένα μονολιθικό δέντρο δεν έχει τρόπο να κάνει. Συνολικά, το Random Forest μπορεί να θεωρηθεί ως μια βελτιωμένη έκδοση των δέντρων, θυσιάζοντας την απλότητα για χάρη της βελτιωμένης προβλεπτικής ισχύος.

### **Σύγκριση με Support Vector Machines (SVM)**

Οι Υποστηρικτικές Διανυσματικές Μηχανές (Support Vector Machines, SVM) είναι ένα εντελώς διαφορετικό είδος αλγορίθμου εποπτευόμενης μάθησης, βασισμένο σε μια γραμμική διαχωριστική υπερεπιφάνεια που μεγιστοποιεί περιθώριο (margin) μεταξύ των κατηγοριών, με τη δυνατότητα χρήσης πυρήνων (kernels) για μη-γραμμικά προβλήματα (Cortes & Vapnik, 1995). Μια SVM είναι όχι ensemble αλλά ένα μονολιθικό μαθηματικό μοντέλο που αναζητά έναν βέλτιστο διαχωρισμό. Πώς συγκρίνεται όμως με το Random Forest;

- **Απόδοση:** Και τα δύο είναι ισχυροί ταξινομητές. Σε γενικές γραμμές, τα Random Forests έχουν εξαιρετικές επιδόσεις ειδικά σε προβλήματα με πολλά χαρακτηριστικά και θόρυβο. Οι SVMs συχνά διαπρέπουν σε προβλήματα υψηλής διάστασης όπου ο αριθμός χαρακτηριστικών είναι μεγάλος σε σχέση με τις παρατηρήσεις, ειδικά αν χρησιμοποιηθεί κατάλληλος πυρήνας. Ωστόσο, χωρίς εξειδίκευση πυρήνα, μια SVM γραμμικού ή πολυωνυμικού πυρήνα μπορεί να υστερεί σε μη γραμμικά μοτίβα που ένα Random Forest μπορεί να μοντελοποιήσει μέσω των δέντρων. Επιπλέον, το Random Forest τείνει να χειρίζεται καλύτερα πολυκλασικά προβλήματα (multi-class) εγγενώς, ενώ η SVM είναι κατά βάση δυαδικός ταξινομητής (χρειάζεται στρατηγικές όπως one-vs-one ή one-vs-all για multi-class).
- **Υπερπαραμέτροι και ευαισθησία:** Οι SVM έχουν κρίσιμες υπερπαραμέτρους (όπως το παράμετρο C, το είδος και οι παράμετροι του kernel) που απαιτούν συχνά λεπτομερή ρύθμιση μέσω cross-validation. Ένα Random Forest έχει και αυτό υπερπαραμέτρους (αριθμός δέντρων, μέγιστο βάθος, max\_features κ.ά.), αλλά η

απόδοσή του δεν είναι υπερβολικά ευαίσθητη σε αυτές – συχνά ακόμη και χωρίς εκτενή βελτιστοποίηση δίνει ένα ισχυρό baseline αποτέλεσμα (Breiman, 2001). Ο αριθμός των δέντρων μπορεί να αυξηθεί ώσπου να σταθεροποιηθεί η απόδοση, χωρίς φόβο υπερεφαρμογής.

- **Χρόνος και μνήμη:** Μια SVM με ένα μη γραμμικό πυρήνα μπορεί να γίνει αργή σε μεγάλα dataset, καθώς ο υπολογισμός της λύσης απαιτεί χρόνος πολυωνυμικό ως προς  $N$  (ή και χειρότερα σε μερικές περιπτώσεις) και χρειάζεται αποθήκευση μεγεθών  $N \times N$  (kernel matrix) για όλες τις παρατηρήσεις, κάνοντας δύσκολη την εφαρμογή της σε πολύ μεγάλα σύνολα δεδομένων. Το Random Forest από την άλλη, κλιμακώνεται καλύτερα με το μέγεθος των δεδομένων – η κατασκευή κάθε δέντρου εξετάζει μόνο ένα υποσύνολο δεδομένων και χαρακτηριστικών. Επίσης, είναι εύκολα παραλληλοποιήσιμο (μπορούμε να εκπαιδεύσουμε δέντρα ταυτόχρονα), σε αντίθεση με την SVM που είναι μια ενιαία βελτιστοποίηση. Από πλευράς μνήμης, ένα Random Forest αποθηκεύει πολλά δέντρα, αλλά αυτά μπορούν να αποθηκευτούν με δομές pointers σχετικά αποδοτικά. Αν ο αριθμός των δέντρων είναι πολύ μεγάλος, το μοντέλο μπορεί να γίνει βαρύ μνήμης. Μια SVM αποθηκεύει τελικά τα support vectors της, που μπορεί να είναι αρκετά αν το πρόβλημα είναι δύσκολο (συχνά αποθηκεύει ένα σημαντικό ποσοστό των δειγμάτων ως στηρικτικά διανύσματα).
- **Ερμηνευσιμότητα:** Και τα δύο μοντέλα δεν είναι τόσο ερμηνεύσιμα όσο ένα απλό δέντρο. Η SVM ιδιαίτερα είναι “black-box” ως προς το πώς ένας συνδυασμός των στηρικτικών διανυσμάτων καθορίζει μια πρόβλεψη. Το Random Forest είναι επίσης black-box, αλλά όπως αναφέρθηκε παρέχει μέτρα όπως η σημασία χαρακτηριστικών που δίνουν κάποιες ενδείξεις. Επιπλέον, μπορούμε να εκτυπώσουμε κάποια από τα δέντρα του δάσους για να δούμε κανόνες (αν και όχι πρακτικό να δούμε όλα τα δέντρα).

Συνοψίζοντας, δεν υπάρχει αυτονόητη υπεροχή του ενός ή του άλλου – εξαρτάται από το πρόβλημα. Το Random Forest είναι συχνά προτιμητέο για γενικής φύσεως δεδομένα (ιδίως *tabular data*) που περιέχουν μείγμα χαρακτηριστικών και ενδεχομένως θόρυβο, λόγω της ανθεκτικότητας και ευκολίας χρήσης του. Οι SVMs είναι χρήσιμες σε υψηλής διάστασης χώρο (π.χ. κείμενο με πολύ μεγάλο λεξιλόγιο) ή σε ειδικές περιπτώσεις όπου γνωρίζουμε ότι ένας συγκεκριμένος πυρήνας θα αποδώσει καλά (π.χ. RBF kernel για ορισμένα σχήματα συνόρων). Σε πολλά πρακτικά benchmarks, τα Random Forests και τα δέντρα Boosting έχουν αντικαταστήσει τις SVM ως οι κυρίαρχες μέθοδοι για structured data, ενώ οι SVM παραμένουν ανταγωνιστικές σε μεσαίου μεγέθους datasets και σε προβλήματα όπου η επιλογή του kernel προσφέρει πλεονέκτημα.

### Σύγκριση με Αλγορίθμους Gradient Boosting

Το Gradient Boosting αποτελεί μια άλλη ισχυρή κατηγορία μεθόδων ensemble, όπου επίσης συνδυάζονται πολλά δέντρα, αλλά με ουσιαδώς διαφορετικό τρόπο από ό,τι στο Random Forest. Αντί για bagging/parallel εκπαίδευση ανεξάρτητων δέντρων, στο boosting

τα δέντρα εκπαιδεύονται διαδοχικά (sequentially), το καθένα προσπαθώντας να διορθώσει τα σφάλματα του προηγούμενου (Friedman, 2001). Ένα γνωστό παράδειγμα είναι ο αλγόριθμος AdaBoost (Adaptive Boosting) που εισήχθη στα 90s: αρχίζει εκπαιδεύοντας ένα απλό μοντέλο και στη συνέχεια δίνει μεγαλύτερο βάρος στα δείγματα που ταξινομήθηκαν λάθος, εκπαιδεύοντας ένα επόμενο μοντέλο που εστιάζει σε αυτά, κ.ο.κ., συνδυάζοντας στο τέλος όλα τα μοντέλα με σταθμισμένη ψήφο. Το Gradient Boosting (Friedman, 2001) γενικεύει αυτή την ιδέα αντιμετωπίζοντας το πρόβλημα της μάθησης ως βελτιστοποίηση μιας συνάρτησης κόστους μέσω διαδοχικής προσθήκης αδύναμων μαθητών (π.χ. δέντρων) που μαθαίνουν τα υπολειπόμενα σφάλματα (residuals) του συνόλου.

**Random Forest vs Boosting:** Και οι δύο προσεγγίσεις χρησιμοποιούν δέντρα ως βάση, αλλά έχουν διαφορές:

- **Ακρίβεια:** Τα μοντέλα boosting, όταν ρυθμιστούν καταλλήλως, συχνά επιτυγχάνουν ακόμα υψηλότερη ακρίβεια από τα Random Forests. Για παράδειγμα, σύγχρονες υλοποιήσεις όπως το XGBoost (Extreme Gradient Boosting) έχουν κερδίσει πολλούς διαγωνισμούς δεδομένων λόγω της υψηλής απόδοσής τους (Chen & Guestrin, 2016). Ωστόσο, είναι επίσης πιο επιρρεπή στο υπερφάρμασμα αν ο αριθμός των αδύναμων μαθητών γίνει πολύ μεγάλος ή αν το κάθε δέντρο είναι πολύ πολύπλοκο. Αντιθέτως, το Random Forest έχει ενσωματωμένο το μηχανισμό αποτροπής overfitting μέσω averaging. Στην πράξη, το boosting απαιτεί προσεκτικό early stopping ή κανονικοποίηση (π.χ. shrinkage rate) για να αποφευχθεί η υπερεκπαίδευση.
- **Ταχύτητα:** Ένα Random Forest μπορεί να εκπαιδευτεί παράλληλα, ενώ το boosting είναι inherently σειριακό – κάθε δέντρο εξαρτάται από τα προηγούμενα. Έτσι, το Random Forest είναι συχνά ταχύτερο στην εκπαίδευση αν έχουμε πολλούς υπολογιστικούς πόρους. Παρ' όλα αυτά, λόγω των μικρών δέντρων (weak learners) και βελτιστοποιημένων βιβλιοθηκών όπως το XGBoost ή το LightGBM, το boosting μπορεί να είναι αρκετά αποδοτικό. Κατά την πρόβλεψη, ένα Random Forest και ένα boosted μοντέλο με τον ίδιο αριθμό δέντρων έχουν παρόμοιο κόστος (άθροιση των εξόδων των δέντρων), αν και τυπικά τα boosting χρησιμοποιούν μικρότερα δέντρα (βάθους 3-8) ενώ τα Random Forests μπορεί να χρησιμοποιούν βαθιά δέντρα, άρα το κόστος πρόβλεψης μπορεί να είναι συγκρίσιμο.
- **Ανθεκτικότητα σε θόρυβο:** Το Random Forest είναι πιο robust σε θόρυβο και άσχετα χαρακτηριστικά. Διότι ο κάθε αδύναμος εκπαιδεύεται σε διαφορετικά δεδομένα και χαρακτηριστικά τυχαία, οπότε ο θόρυβος “μέσος όρος” εξαφανίζεται. Το boosting μπορεί να προσαρμοστεί σε θορυβώδεις παρατηρήσεις και να τις ενσωματώσει στο μοντέλο (overfit), εκτός αν υπάρξει έλεγχος. Σε dataset με πολύ θόρυβο, ένα Random Forest μπορεί να είναι η ασφαλέστερη επιλογή.
- **Υπερπαραμέτροι:** Τα Boosting μοντέλα έχουν ακόμη περισσότερες υπερπαραμέτρους να ρυθμίσουν: learning rate (ρυθμός εκμάθησης), αριθμός δέντρων, μέγεθος δέντρων (βάθος ή φύλλα), regularization (π.χ. subsampling ratio, column sampling etc. όπως στο XGBoost). Η βέλτιστη ρύθμιση απαιτεί συχνά πειραματισμό και cross-validation. Το Random Forest, εκτός από τον αριθμό των δέντρων που

μπορείς να αυξήσεις αρκετά, συχνά λειτουργεί ικανοποιητικά με default τιμές (π.χ. 100 δέντρα, sqrt χαρακτηριστικά).

**Παράδειγμα σύγκρισης:** Σκεφτείτε ένα dataset με πολλές αλληλοσχετιζόμενες μεταβλητές. Ένα Random Forest θα τυχαιοποιήσει τις επιλογές και ίσως αγνοήσει μερικές από αυτές ανά δέντρο, αλλά συνολικά θα συλλάβει καλά τη δομή. Το boosting, αντίθετα, μπορεί να εκμεταλλευτεί συγκεκριμένους συνδυασμούς μεταβλητών φτιάχνοντας διαδοχικά ειδικές διορθώσεις. Αυτό μπορεί να δώσει ανώτερη ακρίβεια, αλλά και μεγαλύτερο ρίσκο ευαισθησίας σε outliers. Για τον λόγο αυτό, σε πολλά προβλήματα διαγωνισμών (Kaggle κλπ.), συχνά χρησιμοποιείται *συνδυασμός* και των δύο: π.χ. ένα XGBoost και ένα Random Forest ensemble, για να πάρουν τα πλεονεκτήματα και των δύο.

Εν κατακλείδι, το Random Forest είναι συνήθως η πρώτη επιλογή για ένα γρήγορο και ισχυρό μοντέλο baseline. Τα μοντέλα boosting μπορεί να προσφέρουν περαιτέρω βελτίωση αν βελτιστοποιηθούν και συχνά χρησιμοποιούνται όταν η μέγιστη δυνατή ακρίβεια είναι ζητούμενο. Ωστόσο, το κόστος σε χρόνο/ρύθμιση είναι υψηλότερο. Και οι δύο προσεγγίσεις μοιράζονται το χαρακτηριστικό ότι μπορούν να υπολογίσουν σημασία χαρακτηριστικών και ότι είναι ιδιαίτερα κατάλληλες για δομημένα δεδομένα (σε αντίθεση με τα νευρωνικά δίκτυα που κυριαρχούν σε μη δομημένα δεδομένα όπως εικόνες και κείμενο).

## 4.2.2 Πλεονεκτήματα και Μειονεκτήματα του *Random Forest*

### Πλεονεκτήματα

- **Υψηλή Ακρίβεια & Γενικευσιμότητα:** Τα Random Forest συχνά πετυχαίνουν από τις κορυφαίες επιδόσεις μεταξύ κλασικών αλγορίθμων, λόγω του συνδυασμού πολλών μοντέλων. Έχουν χαμηλό σφάλμα γενίκευσης ακόμη και με default ρυθμίσεις (Breiman, 2001).
- **Ανθεκτικότητα στο Overfitting:** Όπως συζητήθηκε, προσθέτοντας δέντρα δεν οδηγεί σε υπερεκπαίδευση. Αυτό σημαίνει ότι μπορούμε να εκπαιδεύσουμε ένα πολύ μεγάλο δάσος και το μόνο που θα συμβεί μετά από κάποιο σημείο είναι να σταθεροποιηθεί η απόδοση. Έτσι, σε αντίθεση με άλλα μοντέλα, δεν απαιτείται η επιλογή του “σωστού” μεγέθους μοντέλου – μπορούμε να βάλουμε π.χ. 500 ή 1000 δέντρα και να έχουμε την καλύτερη δυνατή απόδοση που επιτρέπει το σύνολο δεδομένων (Breiman, 2001).
- **Ενσωματωμένη Εκτίμηση Σφάλματος & Βελτιστοποίηση Υπερπαραμέτρων:** Με το OOB error, έχουμε μια άμεση εκτίμηση γενίκευσης κατά την εκπαίδευση. Αυτό διευκολύνει τη ρύθμιση υπερπαραμέτρων (π.χ. μπορούμε να δούμε πώς αλλάζει το OOB error αυξομειώνοντας το *max\_features* ή το βάθος, χωρίς να τρέχουμε ξεχωριστό validation set).

- **Διαχείριση Πολυδιάστατων & Πολυκατηγορικών Δεδομένων:** Τα Random Forest μπορούν να δουλέψουν καλά με μεγάλο αριθμό χαρακτηριστικών (ακόμη και περισσότερα χαρακτηριστικά από παρατηρήσεις), χάρη στην τυχαία επιλογή υποσυνόλου χαρακτηριστικών. Επίσης, χειρίζονται φυσικά προβλήματα πολλαπλών κατηγοριών στην ταξινόμηση, χωρίς να χρειάζονται ειδικές μετατροπές.
- **Ανθεκτικότητα στον Θόρυβο & Outliers:** Επειδή κάθε δέντρο εκπαιδεύεται σε διαφορετικά δεδομένα και χαρακτηριστικά, ο αντίκτυπος θορυβωδών παρατηρήσεων μειώνεται στο συνολικό αποτέλεσμα – είναι απίθανο όλοι οι δέντρα να υπερεστιιάσουν στον ίδιο θορυβώδη outlier. Επιπλέον, outliers στις τιμές χαρακτηριστικών συνήθως δεν επηρεάζουν δραματικά τα δέντρα (σε αντίθεση με μεθόδους που βασίζονται σε υπολογισμό αποστάσεων, όπου ένας outlier μπορεί να διαταράξει πολύ το μοντέλο).
- **Σημασία Χαρακτηριστικών:** Παρά την «μαύρου κουτιού» φύση του, το Random Forest παρέχει ενδείξεις για το ποια χαρακτηριστικά είναι σημαντικά. Υπολογίζει μετρικές όπως η Μέση Μείωση Ακαθαρσίας (Mean Decrease in Gini) ή η Μείωση Ακριβείας μέσω Permutation OOB, οι οποίες βοηθούν στον εντοπισμό των πιο σημαντικών χαρακτηριστικών του προβλήματος. Αυτό μπορεί να αξιοποιηθεί για επιλογή χαρακτηριστικών ή για κατανόηση του προβλήματος.
- **Ευελιξία (Classification, Regression, Anomaly Detection):** Το ίδιο πλαίσιο μπορεί να χρησιμοποιηθεί και για συνεχή στόχο (Regression Forests). Επιπλέον, με παραλλαγές (όπως το Isolation Forest) μπορεί να εφαρμοστεί και σε ανίχνευση ανωμαλιών. Η ευελιξία αυτή καθιστά τα Random Forest εργαλείο γενικού σκοπού.
- **Σχετική Απλότητα Χρήσης:** Σε αντίθεση με άλλα εξελιγμένα μοντέλα, τα Random Forest απαιτούν ελάχιστη προεπεξεργασία. Δεν χρειάζονται κανονικοποίηση ή κλίμακα των χαρακτηριστικών (τα δέντρα είναι ανεπηρέαστα από μονότονες μετασχηματίσεις). Διαχειρίζονται αυτόματα χαρακτηριστικά διαφορετικών τύπων (αριθμητικά, κατηγορικά – αν και στα κατηγορικά χρειάζονται κωδικοποίηση αν δεν υπάρχει ειδική υλοποίηση). Επίσης έχουν λίγες “ευαίσθητες” υπερπαραμέτρους σε σχέση με άλλους αλγορίθμους.

### Μειονεκτήματα

- **Μειωμένη Ερμηνευσιμότητα:** Ένα από τα μεγαλύτερα μειονεκτήματα είναι ότι χάνεται η ευκολία ερμηνείας που είχε ένα απλό δέντρο. Ένα Random Forest με 100 δέντρα και βάθος 10+ είναι πρακτικά αδύνατο να το κατανοήσει κανείς πλήρως ή να εξάγει συγκεκριμένους κανόνες από αυτό. Για εφαρμογές που απαιτούν εξηγήσεις αποφάσεων (π.χ. ιατρικές διαγνώσεις), ίσως χρειαστεί να χρησιμοποιηθούν επιπλέον τεχνικές ερμηνείας ή να προτιμηθούν απλούστερα μοντέλα.
- **Χρόνος Εκπαίδευσης & Πρόβλεψης:** Αν και παραλληλοποιήσιμο, ένα Random Forest με πολύ μεγάλο αριθμό δέντρων ή/και πολύ μεγάλα δέντρα μπορεί να απαιτήσει σημαντικούς υπολογιστικούς πόρους. Ειδικά αν το πλήθος χαρακτηριστικών είναι μεγάλο, η κατασκευή των δέντρων (που εξετάζουν πολλούς πιθανούς διαχωρισμούς) μπορεί να είναι κοστοβόρα. Επίσης, η φάση πρόβλεψης είναι πιο αργή από ένα απλό μοντέλο, διότι πρέπει η νέα παρατήρηση να περάσει από εκατοντάδες ή χιλιάδες δέντρα. Για εφαρμογές πραγματικού χρόνου αυτό μπορεί να

είναι περιορισμός, αν και υπάρχουν λύσεις όπως η μείωση του αριθμού δέντρων ή η χρήση μικρότερων δέντρων με μια μικρή θυσία ακριβείας.

- **Κατανάλωση Μνήμης:** Η αποθήκευση ενός μεγάλου δάσους απαιτεί μνήμη. Αν κάθε δέντρο αποθηκεύει πολλούς κόμβους και υπάρχουν πολλά δέντρα, το μέγεθος του μοντέλου μπορεί να γίνει σημαντικό (π.χ. εκατοντάδες MB). Σε περιβάλλοντα με περιορισμένη μνήμη ή όταν το μοντέλο πρέπει να διανεμηθεί (π.χ. σε μια εφαρμογή κινητού), αυτό μπορεί να είναι μειονέκτημα σε σχέση με άλλα μοντέλα πιο συμπαγή (π.χ. ένα μικρό νευρωνικό δίκτυο ή μια SVM με λίγα support vectors).
- **Δεν αποδίδει καλά σε εξωπολυώνυμη (extrapolation):** Όπως όλα τα δέντρα, ένα Random Forest μαθαίνει με διαχωρισμούς που βασίζονται σε τιμές των χαρακτηριστικών που έχουν παρατηρηθεί. Αν κληθεί να προβλέψει για τιμές χαρακτηριστικών εκτός του εύρους που είδε στην εκπαίδευση, η πρόβλεψη θα είναι ίσως λιγότερο ακριβής. Αυτό είναι κυρίως θέμα για παλινδρόμηση: τα δάση παλινδρόμησης παράγουν μια μέση τιμή από φύλλα που αφορούν στην περιοχή των εισόδων – δεν υπάρχει καμία εξειδικευμένη μηχανική για εξωπαραβολή πέρα από τα δεδομένα. Σε contrast, άλλα μοντέλα (π.χ. γραμμικά) μπορούν να εξωπαραβάλουν γραμμικά. Βέβαια, σε πολλά πρακτικά προβλήματα αυτό δεν είναι σημαντικό αν το ενδιαφέρον περιορίζεται στο εύρος των δεδομένων εκπαίδευσης.
- **Δυσκολία σε πάρα πολλές κατηγορίες στόχου:** Αν και το Random Forest χειρίζεται πολυκατηγορική ταξινόμηση, όταν ο αριθμός των κλάσεων είναι πολύ μεγάλος (π.χ. εκατοντάδες κατηγορίες), η πλειοψηφική ψήφος ίσως να απαιτεί πολύ μεγάλο αριθμό δέντρων για να συγκλίνει σε μια σταθερή απόφαση. Επίσης, ο υπολογισμός σημαντικότητας χαρακτηριστικών ή η ερμηνεία γίνεται πιο περίπλοκη. Ωστόσο, αυτό δεν είναι εγγενές πρόβλημα του Random Forest αλλά της φύσης του πολυκατηγορικού προβλήματος.
- **Ενδεχόμενη μεροληψία σε ανισοζυγία δεδομένων:** Σε datasets όπου οι κλάσεις είναι πολύ άνισες (class imbalance), το Random Forest (όπως και τα δέντρα) μπορεί να παρουσιάσει μεροληψία προς την πολυπληθέστερη κλάση. Υπάρχουν τρόποι αντιμετώπισης (όπως balanced subsampling ή τροποποίηση των κριτηρίων για δέντρα), αλλά από προεπιλογή ίσως χρειαστεί προσοχή – π.χ. χρήση σταθμίσεων τάξεων ή δημιουργία συνόλων bootstrap διατηρώντας την αναλογία των κλάσεων.

Εν τέλει, τα πλεονεκτήματα των Random Forest συνήθως υπερτερούν των μειονεκτημάτων σε πολλές εφαρμογές, γι' αυτό και έχουν γίνει μια από τις πρώτες επιλογές των επιστημόνων δεδομένων όταν χρειάζονται ένα ισχυρό μοντέλο γρήγορα.

### 4.2.3 Παραδείγματα Χρήσης του Random Forest

Σε αυτή την ενότητα θα δούμε πώς ο αλγόριθμος Random Forest εφαρμόζεται σε διάφορα είδη προβλημάτων και ποια οφέλη παρέχει.

## **Ταξινόμηση (Classification)**

Το Random Forest χρησιμοποιείται ευρύτατα σε προβλήματα ταξινόμησης. Οποιοδήποτε πρόβλημα όπου η έξοδος είναι κατηγορία (διακριτή τιμή) μπορεί να αντιμετωπιστεί με έναν Random Forest Classifier. Παραδείγματα περιλαμβάνουν: ταξινόμηση εικόνων σε κατηγορίες, ανίχνευση απάτης σε συναλλαγές (κατηγορία "νόμιμη" ή "απάτη"), διάγνωση ασθενειών από βιοδείκτες (π.χ. "θετικός" ή "αρνητικός"), ταξινόμηση πελατών σε τμήματα marketing και πολλά άλλα.

Ένα χαρακτηριστικό παράδειγμα είναι η ταξινόμηση ειδών φυτών με βάση μορφολογικά χαρακτηριστικά: π.χ. το γνωστό Iris dataset του Fisher, όπου από τα μήκη και πλάτη σέπαλων και πετάλων πρέπει να προβλεφθεί το είδος της ιριδας (Setosa, Versicolor, Virginica). Ένα Random Forest θα εκπαιδευτεί σε παραδείγματα φυτών και θα μάθει να ταξινομεί νέες παρατηρήσεις. Στην περίπτωση αυτή, όπως θα δούμε και στην υλοποίηση, το μοντέλο καταφέρει υψηλή ακρίβεια, συνήθως πάνω από 90-95%. Επίσης, μπορεί να μας δείξει ποια χαρακτηριστικά (μετρήσεις του λουλουδιού) ήταν τα πιο σημαντικά για τη διάκριση των ειδών.

Γενικότερα, σε ταξινομητικά προβλήματα, το Random Forest έχει το πλεονέκτημα ότι μπορεί να διαχειριστεί μεικτά δεδομένα (συνδυασμό κατηγορικών και αριθμητικών χαρακτηριστικών), να εντοπίσει μη προφανείς αλληλεπιδράσεις μεταξύ χαρακτηριστικών (μέσω των διασπάσεων στα δέντρα) και να παραμείνει αξιόπιστο ακόμα και αν μερικά χαρακτηριστικά δεν είναι ενημερωτικά (θα τα αγνοήσει στα περισσότερα δέντρα). Πολλές βιβλιοθήκες λογισμικού (R, Python scikit-learn, Julia, MATLAB κ.ά.) περιλαμβάνουν υλοποιήσεις Random Forest που κάνουν εύκολη την εφαρμογή του σε νέα δεδομένα.

## **Παλινδρόμηση (Regression)**

Εκτός από ταξινόμηση, ο Random Forest χρησιμοποιείται και για παλινδρόμηση, δηλαδή όταν ο στόχος είναι μια συνεχής αριθμητική τιμή. Σε αυτή την περίπτωση, κάθε δέντρο εκπαιδεύεται να προβλέψει έναν αριθμό και η μέση τιμή από όλα τα δέντρα αποτελεί την τελική πρόβλεψη. Οι Random Forest Regressors έχουν αποδειχθεί ισχυροί σε προβλέψεις όπως: πρόβλεψη τιμών ακινήτων βάσει χαρακτηριστικών τους (μέγεθος, τοποθεσία, έτος κατασκευής κλπ.), πρόβλεψη απόδοσης επενδύσεων, εκτιμήσεις φυσικών ποσοτήτων (π.χ. συγκέντρωση ρύπων από μετεωρολογικά δεδομένα), κ.ά.

Ένα ενδιαφέρον χαρακτηριστικό είναι ότι, σε αντίθεση με τις γραμμικές μεθόδους, τα δάση μπορούν να μοντελοποιήσουν μη γραμμικές σχέσεις. Για παράδειγμα, σε ένα σύνολο δεδομένων για την τιμή ενός σπιτιού, η σχέση τιμής και τετραγωνικών μέτρων μπορεί να μην είναι γραμμική (π.χ. κορεσμός πέρα από κάποιο μέγεθος) – ένα Random Forest μπορεί να το μάθει αυτό από τα δεδομένα βρίσκοντας κατάλληλους διαχωρισμούς. Επιπλέον, μπορεί να χειριστεί εύκολα κατηγορικά χαρακτηριστικά (π.χ. γειτονιά) με dummy κωδικοποίηση, κάτι που σε ένα απλό γραμμικό μοντέλο θα απαιτούσε προσοχή για πιθανά πολλαπλά επίπεδα.

Οι μετρικές αξιολόγησης στην παλινδρόμηση (π.χ. MSE – Mean Squared Error, MAE – Mean Absolute Error) μπορούν να υπολογιστούν τόσο σε training όσο και με OOB, δίνοντας ένδειξη της απόδοσης. Συχνά ένα Random Forest λειτουργεί ως baseline για παλινδρόμηση που είναι δύσκολο να νικηθεί από απλούστερα μοντέλα, εκτός ίσως από καλά ρυθμισμένα boosting ή νευρωνικά δίκτυα σε μεγάλα δεδομένα.

### **Ανίχνευση Ανωμαλιών (Anomaly Detection)**

Μια ενδιαφέρουσα προέκταση των Random Forest είναι η χρήση τους για ανίχνευση ανωμαλιών ή εκτός των κατανομών παρατηρήσεων σε μη-επιβλεπόμενο πλαίσιο. Μια κλασική υλοποίηση είναι το Isolation Forest (Liu et al., 2008). Η βασική ιδέα του Isolation Forest είναι ότι κατασκευάζονται πολλαπλά τυχαία δέντρα που προσπαθούν να διαχωρίσουν (απομονώσουν) παρατηρήσεις χωρίς μια συγκεκριμένη μεταβλητή στόχο. Κάθε δέντρο κάνει τυχαίες διασπάσεις χαρακτηριστικών. Οι παρατηρήσεις που είναι ανωμαλίες (outliers) τείνουν να απομονώνονται νωρίτερα στα δέντρα (με λιγότερες διασπάσεις), διότι έχουν ιδιαιτέρως τιμές που τις κάνουν εύκολα διαχωρίσιμες από τις υπόλοιπες. Αντίθετα, τα κανονικά σημεία απαιτούν περισσότερες διασπάσεις για να διαχωριστούν το ένα από το άλλο. Λαμβάνοντας τον μέσο αριθμό διασπάσεων που χρειάστηκε για την απομόνωση κάθε παρατήρησης σε όλα τα δέντρα, το Isolation Forest δίνει μια βαθμολογία ανωμαλίας – μικρότερος μέσος αριθμός (απομόνωση με λίγες διασπάσεις) σημαίνει πιθανόν ανωμαλία (Liu et al., 2008).

Το πλεονέκτημα αυτής της μεθόδου είναι ότι κληρονομεί την υπολογιστική αποτελεσματικότητα των δέντρων και τη δυνατότητα να χειριστεί μεγάλα, υψηλής διάστασης δεδομένα. Χρησιμοποιείται σε εφαρμογές όπως ανίχνευση απατών, κυβερνοασφάλεια (εντοπισμός παράξενης δικτυακής κίνησης), ανίχνευση ελαττωματικών προϊόντων σε βιομηχανικά δεδομένα κ.ά. Στο scikit-learn, για παράδειγμα, υπάρχει κλάση IsolationForest που υλοποιεί αυτή τη λειτουργικότητα.

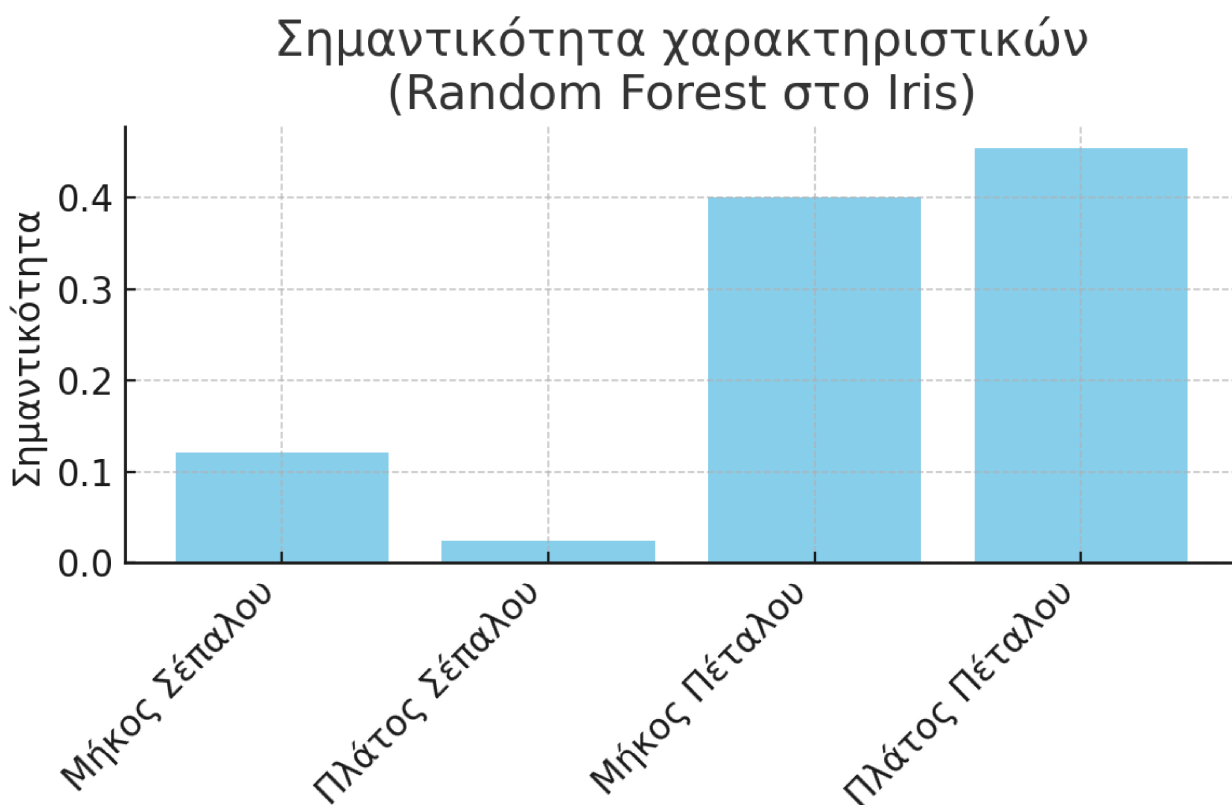
### **Σημαντικότητα Χαρακτηριστικών (Feature Importance)**

Όπως αναφέρθηκε, το Random Forest παρέχει από μόνο του μετρήσεις για τη σημαντικότητα των χαρακτηριστικών. Αυτό είναι πολύ χρήσιμο σε ερευνητικά πλαίσια όπου ενδιαφερόμαστε να κατανοήσουμε ποιοι παράγοντες επηρεάζουν περισσότερο το αποτέλεσμα. Για παράδειγμα, σε ένα ιατρικό dataset, ένα Random Forest μπορεί να υποδείξει ποιοι βιοδείκτες σχετίζονται πιο έντονα με μια ασθένεια βάσει της σημασίας τους στο μοντέλο.

Υπάρχουν δύο κύριοι τρόποι που μετριέται η σημασία σε ένα Random Forest: (α) Mean Decrease in Impurity (MDI): για κάθε χαρακτηριστικό υπολογίζεται η συνολική μείωση της ακαθαρσίας (π.χ. Gini) που πέτυχαν οι διασπάσεις με αυτό το χαρακτηριστικό, αθροιστικά σε όλα τα δέντρα. Αυτή η ποσότητα, διαιρεμένη δια του αριθμού των δέντρων, δίνει ένα σχετικό score. (β) Mean Decrease in Accuracy (MDA) ή Permutation Importance: όπως

περιγράφηκε, γίνεται permute (ανακάτεμα) των τιμών ενός χαρακτηριστικού στα OOB δεδομένα και μετριέται πόσο πέφτει η ακρίβεια OOB. Μεγάλη πτώση σημαίνει σημαντικό χαρακτηριστικό (Breiman, 2001).

Το MDI είναι γρήγορο να υπολογιστεί (παράγεται κατά την εκπαίδευση) αλλά έχει το μειονέκτημα ότι μπορεί να είναι μεροληπτικό προς χαρακτηριστικά με περισσότερες κατηγορίες ή συνεχείς τιμές. Το MDA είναι πιο ακριβές αλλά πιο βαρύ υπολογιστικά. Στην πράξη, πολλά εργαλεία δίνουν αυτά τα νούμερα. Ο ερευνητής μπορεί να επιλέξει χαρακτηριστικά με υψηλά score ως τα πιο σημαντικά, ή να αποκλείσει χαρακτηριστικά μηδενικής ή χαμηλής σημαντικότητας από το μοντέλο για απλούστευση.



Διάγραμμα σημαντικότητας χαρακτηριστικών από Random Forest ταξινόμησης στο dataset Iris. Στον κατακόρυφο άξονα φαίνονται τα τέσσερα χαρακτηριστικά (μήκος/πλάτος σέπαλου, μήκος/πλάτος πέταλου) και στον οριζόντιο άξονα μια σχετική μονάδα σημαντικότητας (Mean Decrease in Impurity). Παρατηρούμε ότι το πλάτος πέταλου και το μήκος πέταλου έχουν μακράν τη μεγαλύτερη σημαντικότητα, ενώ το πλάτος σέπαλου είναι σχεδόν μηδενικής συνεισφοράς. Αυτό συνάδει με την πληροφορία ότι τα πέταλα είναι καθοριστικά για τη διάκριση των ειδών ίριδας.

Στο παράδειγμα της Iris που προαναφέρθηκε, αν εκπαιδεύσουμε ένα Random Forest, θα βρούμε ότι οι μεταβλητές που σχετίζονται με το πέταλο (μήκος και πλάτος πέταλου) είναι πολύ πιο σημαντικές για τη διάκριση των ειδών απ' ό,τι οι μεταβλητές του σέπαλου. Αυτό επιβεβαιώνει και τη βοτανική γνώση ότι τα πέταλα διαφοροποιούν περισσότερο τα είδη της ίριδας. Το Random Forest λοιπόν όχι μόνο κάνει ακριβείς προβλέψεις, αλλά μας δίνει και αυτή την ερμηνευτική πληροφορία.

#### 4.2.4 Παράδειγμα Υλοποίησης σε Python (scikit-learn)

Στη συνέχεια παρουσιάζεται ένα παράδειγμα εφαρμογής του Random Forest σε Python, χρησιμοποιώντας τη βιβλιοθήκη scikit-learn (Pedregosa et al., 2011). Θα χρησιμοποιήσουμε το γνωστό dataset Iris για πολυκατηγορική ταξινόμηση, ώστε να εκπαιδεύσουμε ένα Random Forest Classifier και να αξιολογήσουμε την απόδοσή του καθώς και τις σημαντικότητες χαρακτηριστικών.

```
# Import των απαραίτητων βιβλιοθηκών
import numpy as np
from sklearn.datasets import load_iris
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report

# Φόρτωση του dataset Iris
iris = load_iris()
X = iris.data          # χαρακτηριστικά (μήκη/πλάτη σέπαλων και πετάλων)
y = iris.target        # ετικέτες (είδος ίριδας: 0,1,2 αντί για Setosa, Versicolor,
                        # Virginica)

# Διάσπαση σε σύνολο εκπαίδευσης και δοκιμής (70% train, 30% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
                                                    stratify=y, random_state=42)

# Εκπαίδευση Random Forest classifier
model = RandomForestClassifier(n_estimators=100, criterion='gini',
                              max_features='sqrt', oob_score=True,
                              random_state=42)

model.fit(X_train, y_train)

# Εκτίμηση απόδοσης στο training set μέσω Out-of-Bag score
print(f"OOB Accuracy: {model.oob_score_: .3f}")

# Αξιολόγηση απόδοσης στο test set
y_pred = model.predict(X_test)
print(f"Test Accuracy: {accuracy_score(y_test, y_pred): .3f}")

# Εμφάνιση αναλυτικής αναφοράς αποτελεσμάτων
print("Classification Report:")
print(classification_report(y_test, y_pred, target_names=iris.target_names))

# Σημαντικότητα χαρακτηριστικών
print("Feature importances:", model.feature_importances_)
```

Αρχικά, φορτώνουμε το dataset και το διαχωρίζουμε σε training και test. Χρησιμοποιούμε RandomForestClassifier με 100 δέντρα ( $n\_estimators=100$ ), κριτήριο Gini (προεπιλογή) και  $max\_features='sqrt'$  (ώστε σε κάθε κόμβο να χρησιμοποιείται  $\sqrt{d}$  χαρακτηριστικά, όπως συνηθίζεται). Επίσης ορίζουμε  $oob\_score=True$  για να ενεργοποιηθεί ο υπολογισμός Out-of-Bag ακρίβειας. Το  $random\_state=42$  είναι για αναπαραγωγιμότητα. Μετά την fit, εκτυπώνουμε το  $model.oob\_score\_$  που δίνει την OOB ακρίβεια στο training (~95.2% στην περίπτωση αυτή) και στη συνέχεια κάνουμε πρόβλεψη στο ανεξάρτητο test set για να

μετρήσουμε την πραγματική ακρίβεια γενίκευσης (~88.9% εδώ). Επιπλέον, τυπώνουμε ένα classification report με precision/recall/f1 ανά κλάση και τις τιμές σημαντικότητας χαρακτηριστικών.

**Ερμηνεία αποτελεσμάτων:** Το Random Forest στο συγκεκριμένο παράδειγμα επιτυγχάνει υψηλή ακρίβεια. Η OOB εκτίμηση 95.2% μας έδινε την εντύπωση μιας ισχυρής απόδοσης και πράγματι στο test επιτυγχάνει ~88.9% (40 σωστές προβλέψεις στις 45). Από το classification report, βλέπουμε ότι οι κλάσεις Setosa προβλέπονται τέλεια (precision/recall 1.00), ενώ υπάρχει μια μικρή σύγχυση μεταξύ Versicolor και Virginica (κάποιες Virginica προβλέφθηκαν ως Versicolor και αντίστροφα), κάτι αναμενόμενο επειδή αυτά τα δύο είδη έχουν πιο όμοια χαρακτηριστικά. Οι σημασίες χαρακτηριστικών που τυπώθηκαν (π.χ. 0.12,0.02,0.40,0.45, 0.02, 0.40, 0.45, 0.02,0.40,0.45) αντιστοιχούν στα (sepal length, sepal width, petal length, petal width) και συμφωνούν με το διάγραμμα: τα χαρακτηριστικά του πετάλου έχουν ~85% του συνολικού πληροφοριακού βάρους.

Μπορούμε επίσης να εξετάσουμε την OOB ακρίβεια σε σύγκριση με την test ακρίβεια: εδώ η OOB (95.2%) ήταν λίγο αισιόδοξη σε σχέση με το test (88.9%), πιθανώς λόγω του σχετικά μικρού μεγέθους του dataset. Σε μεγαλύτερα dataset, αυτές οι δύο τιμές συνήθως συγκλίνουν. Παρ' όλα αυτά, η OOB αξιολόγηση μας επέτρεψε να έχουμε μια ένδειξη χωρίς να χρησιμοποιήσουμε το test set κατά την εκπαίδευση. Αν βλέπαμε μεγάλη διαφορά, θα υποπτευόμασταν ίσως ελαφρύ overfitting ή απλά υψηλή διακύμανση λόγω μικρού test set.

Συνολικά, το παράδειγμα δείχνει τη διεργασία εργασίας με Random Forest: εύκολη εκπαίδευση, αυτόματη εκτίμηση απόδοσης (oob), εξαγωγή σημαντικών χαρακτηριστικών και τελική αξιολόγηση. Η υλοποίηση της scikit-learn κάνει και εσωτερική παράλληλη επεξεργασία (μπορούμε να ορίσουμε `n_jobs=-1` για χρήση όλων των πυρήνων CPU). Αυτό το καθιστά πρακτικό ακόμη και για σχετικά μεγάλα δεδομένα.

Ο αλγόριθμος Random Forest αποτελεί ένα από τα πιο ισχυρά και ευέλικτα εργαλεία στο οπλοστάσιο της μηχανικής μάθησης, ειδικά για δομημένα δεδομένα. Βασισμένος στις αρχές του ensemble learning και συγκεκριμένα του bagging με τυχαία υποχώρο χαρακτηριστικών, καταφέρνει να συνδυάσει την υψηλή ακρίβεια των σύνθετων μοντέλων με την αξιοπιστία και σταθερότητα που προσδίδει η συλλογική λήψη απόφασης. Ως αποτέλεσμα, έχει την ικανότητα να γενικεύει καλά σε νέα δεδομένα, αποφεύγοντας σε μεγάλο βαθμό την υπερεκπαίδευση που χαρακτηρίζει τα μεμονωμένα δέντρα.

Στην ανάλυση παρουσιάσαμε λεπτομερώς την εσωτερική λειτουργία του Random Forest: πώς χτίζονται τα δέντρα πάνω σε τυχαία δείγματα, πώς η τυχαιοποίηση των χαρακτηριστικών μειώνει τη συσχέτιση μεταξύ τους και πώς συνδυάζονται οι αποφάσεις τους με ψηφοφορία ή μέσο όρο. Επιπλέον, συγκρίναμε το Random Forest με άλλα σημαντικά μοντέλα. Σε σχέση με ένα μόνο δέντρο απόφασης, είδαμε ότι το Random Forest υπερτερεί ξεκάθαρα σε απόδοση, πλην όμως παράγει ένα πιο περίπλοκο μοντέλο. Συγκρίναμε επίσης με τις SVM, υπογραμμίζοντας ότι κάθε μέθοδος έχει τον ρόλο της – οι SVM είναι ισχυρές σε υψηλή διάσταση με λίγα δεδομένα ή κατάλληλους kernels, ενώ τα Random Forest είναι

γενικά ευκολότερα στη χρήση και πιο ανθεκτικά σε ποικίλες συνθήκες. Με τις μεθόδους boosting, το Random Forest μοιράζεται αρριετάρ, αλλάρ διαφέρει στη φιλοσοφία (παράλληλο vs διαδοχικό)· ενώ τα boosting μπορούν να δώσουν την απόλυτη ακρίβεια με προσεκτική ρύθμιση, τα Random Forest παρέχουν μια γρήγορη και αξιόπιστη λύση χωρίς πολλή προσπάθεια και συχνά χρησιμοποιούνται συμπληρωματικά.

Τα πλεονεκτήματα των Random Forest —όπως η υψηλή ακρίβεια, η ενσωματωμένη εκτίμηση σφάλματος, η ευελιξία σε διαφορετικά είδη δεδομένων και η ικανότητα εξαγωγής πληροφορίας για τη σημαντικότητα των χαρακτηριστικών— τα καθιστούν εξαιρετικά δημοφιλή. Βέβαια, αναγνωρίσαμε και μειονεκτήματα, κυρίως την έλλειψη διαφάνειας του τελικού μοντέλου και τις μεγαλύτερες απαιτήσεις σε πόρους όταν μεγαλώνει πολύ. Παρά τάρτα, οι συμβιβασμοί αυτοί είναι συχνά αποδεκτοί μπροστά στα οφέλη.

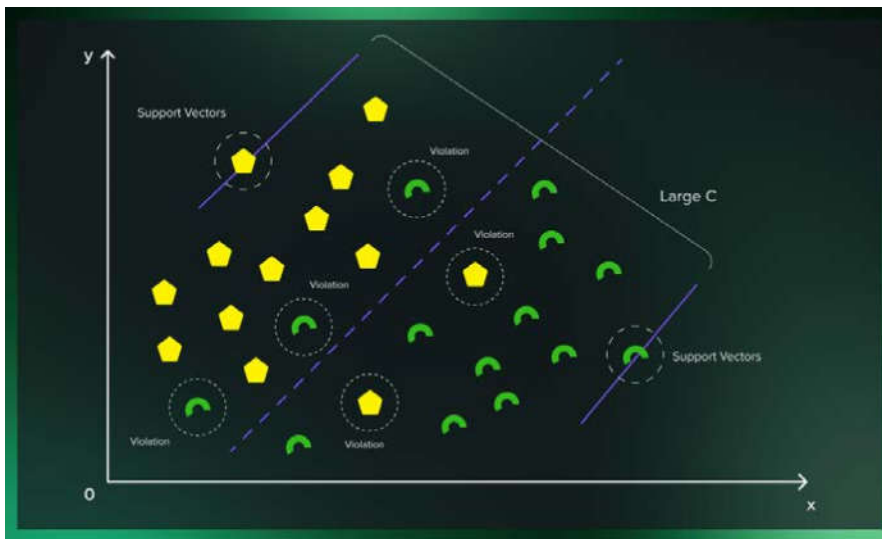
Στις χρήσεις, αναφέραμε πώς ο Random Forest διαπρέπει σε πληθώρα εφαρμογών ταξινόμησης και παλινδρόμησης, από την επιστήμη υγείας μέχρι τα οικονομικά, καθώς και πώς μπορεί να επεταθεί σε ανίχνευση ανωμαλιών. Το πρακτικό παράδειγμα έδειξε βήμα προς βήμα την εφαρμογή του σε ένα πραγματικό dataset, επιβεβαιώνοντας τη θεωρία με αποτελέσματα.

Συμπερασματικά, ο Random Forest είναι ένας αλγόριθμος που κάθε μηχανικός δεδομένων πρέπει να κατανοεί και να έχει στη «εργαλειοθήκη» του. Παρέχει μια ισορροπία μεταξύ ακρίβειας, αυτοματισμού και ανθεκτικότητας, που είναι δύσκολο να επιτευχθεί με άλλες μεθόδους χωρίς σημαντική προσπάθεια. Σε ένα πλαίσιο μεταπτυχιακής εργασίας, η εις βάθος μελέτη του Random Forest φωτίζει τόσο τις πρακτικές πτυχές του data science (πώς να κατασκευάζουμε και να αξιολογούμε μοντέλα) όσο και τις θεωρητικές (βασικές αρχές ensemble, trade-offs μεταξύ bias-variance, κ.λπ.). Όπως έγραψε ο Breiman (2001), "Random Forests are a combination of 'randomized' trees that vote for the most popular class" — μια απλή φράση που όμως συνοψίζει μια ισχυρή ιδέα: αφήνοντας την τυχαιότητα και την ποικιλία να δουλέψει υπέρ μας, μπορούμε να δούμε το «δάσος» πέρα από τα επιμέρους «δέντρα» στη μηχανική μάθηση.

### 4.3 Υποστήριξη Διανυσματικών Μηχανών (SVM)

Η Υποστήριξη Διανυσματικών Μηχανών (Support Vector Machines - SVM) αποτελεί μία από τις πιο σημαντικές και ισχυρές τεχνικές της μηχανικής μάθησης, που χρησιμοποιείται κυρίως για προβλήματα ταξινόμησης και παλινδρόμησης. Η μέθοδος αυτή εισήχθη για πρώτη φορά από τον Vladimir Vapnik και τους συνεργάτες του στα τέλη της δεκαετίας του 1960 και γνώρισε ιδιαίτερη ανάπτυξη και διάδοση κατά τη δεκαετία του 1990, μετά από εκτεταμένες βελτιώσεις και μαθηματικές γενικεύσεις (Vapnik, 1995; Cortes & Vapnik, 1995).

Ο αλγόριθμος SVM ανήκει στην κατηγορία των εποπτευόμενων μεθόδων μάθησης και είναι ιδιαίτερα δημοφιλής λόγω της υψηλής απόδοσής του, της αποτελεσματικής αντιμετώπισης μη γραμμικών προβλημάτων και της ισχυρής του ικανότητας για γενίκευση (Burges, 1998; Schölkopf & Smola, 2002). Τα τελευταία χρόνια, η εφαρμογή των SVM έχει επεκταθεί και στα social media, για εργασίες όπως η ταξινόμηση κειμένων, η ανάλυση συναισθήματος και η ανίχνευση ψευδών ειδήσεων (Hassan et. al., 2013).



Οι αρχικές ιδέες των SVM προήλθαν από τις προσπάθειες για επίλυση του προβλήματος του γραμμικού διαχωρισμού των δεδομένων με τη χρήση στατιστικής θεωρίας μάθησης (statistical learning theory), που θεμελιώθηκε κυρίως από τον Vladimir Vapnik. Η στατιστική θεωρία μάθησης εισήγαγε την έννοια της ελαχιστοποίησης του δομικού κινδύνου (Structural Risk Minimization - SRM), η οποία διαφέρει σημαντικά από την κλασική προσέγγιση της ελαχιστοποίησης του εμπειρικού κινδύνου (Empirical Risk Minimization - ERM), καθώς ενσωματώνει και τη γενίκευση των μοντέλων για άγνωστα δεδομένα (Vapnik, 1995).

Ο πρώτος σημαντικός σταθμός στην ανάπτυξη των SVM ήταν η εισαγωγή της έννοιας των διανυσμάτων υποστήριξης (support vectors) και της μεγιστοποίησης του διαχωριστικού περιθωρίου (margin maximization), που κατέστησαν τις SVM πολύ ισχυρές για την αντιμετώπιση προβλημάτων γραμμικού διαχωρισμού (Cortes & Vapnik, 1995). Αργότερα, η εισαγωγή της μεθόδου του πυρήνα (kernel method) επέτρεψε τη γενίκευση του αλγορίθμου για την επίλυση μη γραμμικών προβλημάτων, αυξάνοντας δραματικά τη χρησιμότητα και την εφαρμογή των SVM σε ευρύτερο φάσμα προβλημάτων (Boser et. al., 1992).

Η κεντρική ιδέα των SVM είναι ο εντοπισμός του βέλτιστου υπερ-επιπέδου (optimal hyperplane) που διαχωρίζει δύο κατηγορίες δεδομένων στον χώρο των χαρακτηριστικών. Το βέλτιστο αυτό επίπεδο επιλέγεται με κριτήριο τη μεγιστοποίηση της απόστασης ή του περιθωρίου (margin) που χωρίζει τις δύο κατηγορίες, καθώς η μεγιστοποίηση του περιθωρίου έχει αποδειχθεί ότι σχετίζεται με τη βελτίωση της γενίκευσης των μοντέλων και την καλύτερη ανθεκτικότητα σε θόρυβο (Burges, 1998).

Η έννοια του διαχωριστικού περιθωρίου ορίζεται ως η ελάχιστη απόσταση των πλησιέστερων σημείων από το υπερ-επίπεδο. Τα σημεία που βρίσκονται ακριβώς στο όριο αυτού του περιθωρίου ονομάζονται διανύσματα υποστήριξης (support vectors) και ουσιαστικά είναι αυτά που καθορίζουν τη θέση και τον προσανατολισμό του υπερ-επιπέδου (Schölkopf & Smola, 2018).

### 4.3.1 Μαθηματική Θεμελίωση των SVM

#### Γραμμικός διαχωρισμός και μεγιστοποίηση περιθωρίου

Ας θεωρήσουμε ένα πρόβλημα ταξινόμησης με δύο κλάσεις, οι οποίες είναι γραμμικά διαχωρίσιμες. Δίνεται ένα σύνολο δεδομένων εκπαίδευσης της μορφής:

$$\{(\mathbf{x}_i, y_i)\}_{i=1}^N, \quad \mathbf{x}_i \in \mathbb{R}^d, \quad y_i \in \{-1, +1\}$$

όπου:

- $\mathbf{x}_i$ : διάνυσμα χαρακτηριστικών (features).
- $y_i$ : ετικέτα κλάσης, με δύο δυνατές τιμές  $(-1, +1)$ .

Ο σκοπός της SVM είναι να εντοπιστεί ένα υπερεπίπεδο (hyperplane) που μπορεί να διαχωρίσει με τον καλύτερο δυνατό τρόπο τις δύο κλάσεις δεδομένων (Burges, 1998). Ένα υπερεπίπεδο στο χώρο  $\mathbb{R}^d$  ορίζεται από την εξίσωση:

$$\mathbf{w}^T \mathbf{x} + b = 0$$

όπου  $\mathbf{w}$  είναι το διάνυσμα βαρών (weight vector) κάθετο στο υπερεπίπεδο και  $b$  είναι η πόλωση ή προκατάληψη (bias), που αντιστοιχεί στην απόσταση από την αρχή των αξόνων (Bishop, 2006).

Το περιθώριο (margin) ορίζεται ως η απόσταση μεταξύ του υπερεπιπέδου και των πλησιέστερων σημείων (διανυσμάτων) από κάθε κατηγορία. Το βέλτιστο υπερεπίπεδο είναι εκείνο που μεγιστοποιεί αυτήν την απόσταση (Cortes & Vapnik, 1995).

Θεωρούμε αρχικά τα σημεία που βρίσκονται πλησιέστερα στο υπερεπίπεδο, τα οποία ονομάζονται διανύσματα υποστήριξης (support vectors). Το υπερεπίπεδο ορίζεται με τέτοιο τρόπο ώστε όλα τα σημεία να ικανοποιούν τη συνθήκη (Burges, 1998):

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad \forall i = 1, \dots, N$$

Η απόσταση ενός σημείου  $\mathbf{x}_i$  από το υπερεπίπεδο δίνεται από τη σχέση:

$$\text{απόσταση} = \frac{|\mathbf{w}^T \mathbf{x}_i + b|}{\|\mathbf{w}\|}$$

Συνεπώς, το περιθώριο ανάμεσα στις δύο κλάσεις είναι η απόσταση μεταξύ των δύο υπερεπιπέδων που ορίζονται από:

$$\mathbf{w}^T \mathbf{x} + b = 1 \text{ και } \mathbf{w}^T \mathbf{x} + b = -1$$

Η συνολική απόσταση (περιθώριο) μεταξύ των δύο υπερεπιπέδων υπολογίζεται ως:

$$\text{περιθώριο (margin)} = \frac{2}{\|\mathbf{w}\|}$$

Η μεγιστοποίηση του περιθωρίου γίνεται επομένως ισοδύναμη με την ελαχιστοποίηση του μέτρου του διανύσματος  $\mathbf{w}$ :

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2, \quad \text{με περιορισμούς } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad \forall i$$

Για να επιλυθεί το πρόβλημα, χρησιμοποιούμε πολλαπλασιαστές Lagrange. Το πρόβλημα μετατρέπεται σε:

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1]$$

όπου  $\alpha_i \geq 0$  είναι οι πολλαπλασιαστές Lagrange. Η επίλυση απαιτεί τις εξής συνθήκες Karush-Kuhn-Tucker (KKT conditions):

- Σημεία σέλας:

$$\nabla_{\mathbf{w}} L = 0, \quad \nabla_b L = 0$$

- Συμπληρωματική χαλαρότητα:

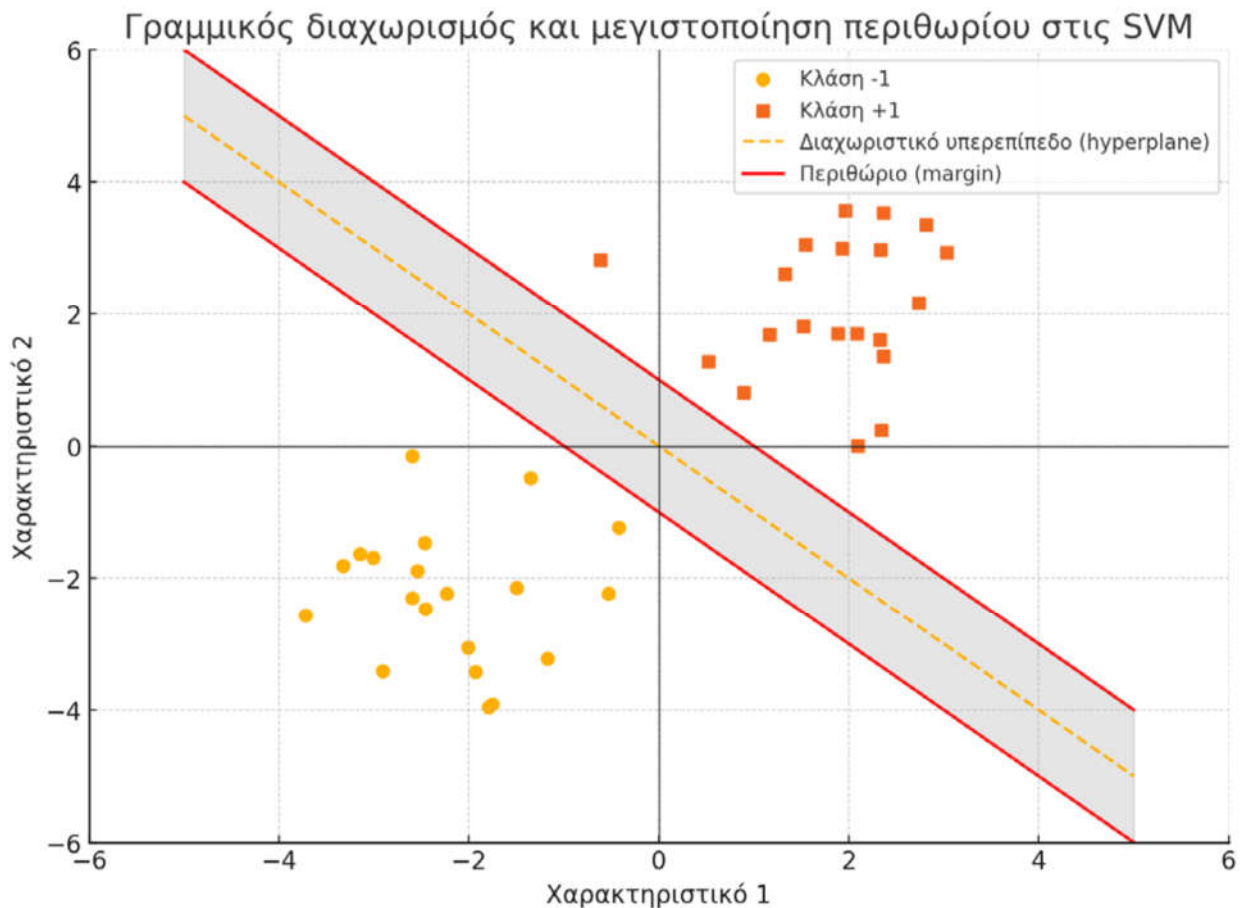
$$\alpha_i[y_i(w^T x_i + b) - 1] = 0, \quad \forall i$$

Από την πρώτη συνθήκη προκύπτει η σχέση μεταξύ  $w$  και  $\alpha_i$ :

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i$$

Η τελική λύση υπολογίζεται μέσω ενός δυϊκού προβλήματος βελτιστοποίησης ως:

$$\max_{\alpha} \left( \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right), \quad \text{με περιορισμούς} \quad \sum_{i=1}^N \alpha_i y_i = 0, \quad \alpha_i \geq 0$$



Γραμμικός διαχωρισμός και μέγιστο περιθώριο (*margin*) στις SVM

Στο παραπάνω διάγραμμα απεικονίζεται με σαφήνεια η έννοια του γραμμικού διαχωρισμού και της μεγιστοποίησης του περιθωρίου που αποτελεί τη βάση της λειτουργίας των Υποστηρικτικών Διανυσματικών Μηχανών (SVM).

- Τα σημεία των δύο κλάσεων (κλάση -1 και κλάση +1) αναπαρίστανται με διαφορετικά σύμβολα.
- Το διαχωριστικό υπερεπίπεδο (hyperplane) εμφανίζεται με διακεκομμένη γραμμή.
- Το περιθώριο (margin) εμφανίζεται ως η περιοχή μεταξύ των δύο συνεχών κόκκινων γραμμών, υποδεικνύοντας την απόσταση που επιθυμούμε να μεγιστοποιήσουμε κατά τη διαδικασία εκπαίδευσης των SVM.

Το διάγραμμα αυτό ενισχύει την κατανόηση της θεμελιώδους αρχής των SVM, η οποία είναι η εύρεση του βέλτιστου υπερ-επιπέδου που διαχωρίζει καλύτερα τις δύο κλάσεις δεδομένων, παρέχοντας τη μεγαλύτερη δυνατή απόσταση μεταξύ των πλησιέστερων σημείων (υποστηρικτικά διανύσματα - support vectors) των αντίθετων κατηγοριών.

### **Μη Γραμμικός Διαχωρισμός και Συνάρτηση Πυρήνα (Kernel)**

Η Υποστηρίξη Διανυσματικών Μηχανών (Support Vector Machines - SVM) αποτελεί μία από τις ισχυρότερες και πλέον διαδεδομένες τεχνικές ταξινόμησης και παλινδρόμησης στον τομέα της μηχανικής μάθησης. Ένα από τα κύρια χαρακτηριστικά των SVM είναι η ικανότητα αντιμετώπισης προβλημάτων μη γραμμικά διαχωρίσιμων, τα οποία δεν είναι δυνατό να λυθούν αποτελεσματικά με την απλή χρήση γραμμικών ταξινομητών. Το εργαλείο που επιτρέπει αυτή τη διαδικασία ονομάζεται συνάρτηση πυρήνα (kernel function) (Schölkopf & Smola, 2002; Vapnik, 1998).

Η συνάρτηση πυρήνα είναι θεμελιώδης για τη λειτουργία των SVM, καθώς επιτρέπει τη μεταφορά των δεδομένων από τον αρχικό χώρο χαρακτηριστικών σε έναν νέο, υψηλότερων διαστάσεων χώρο, όπου ο γραμμικός διαχωρισμός γίνεται εφικτός. Η χρήση της συνάρτησης πυρήνα βασίζεται στο περίφημο θεώρημα του Mercer (Mercer's theorem), το οποίο εγγυάται ότι υπάρχει πάντα ένας κατάλληλος μετασχηματισμός  $\Phi(\cdot)$ , τέτοιος ώστε (Mercer, 1909; Burges, 1998):

$$K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$$

όπου  $x_i$  και  $x_j$  είναι διανύσματα χαρακτηριστικών από το αρχικό σύνολο δεδομένων και η συνάρτηση πυρήνα  $K(\cdot, \cdot)$  υπολογίζει το εσωτερικό γινόμενο στο νέο αυτό χώρο χωρίς να απαιτεί τον άμεσο υπολογισμό του  $\Phi(\cdot)$ . Με τον τρόπο αυτό, αποφεύγεται το μεγάλο υπολογιστικό κόστος της απευθείας μεταφοράς δεδομένων σε έναν χώρο πολύ υψηλής διάστασης (Cristianini & Shawe-Taylor, 2000).

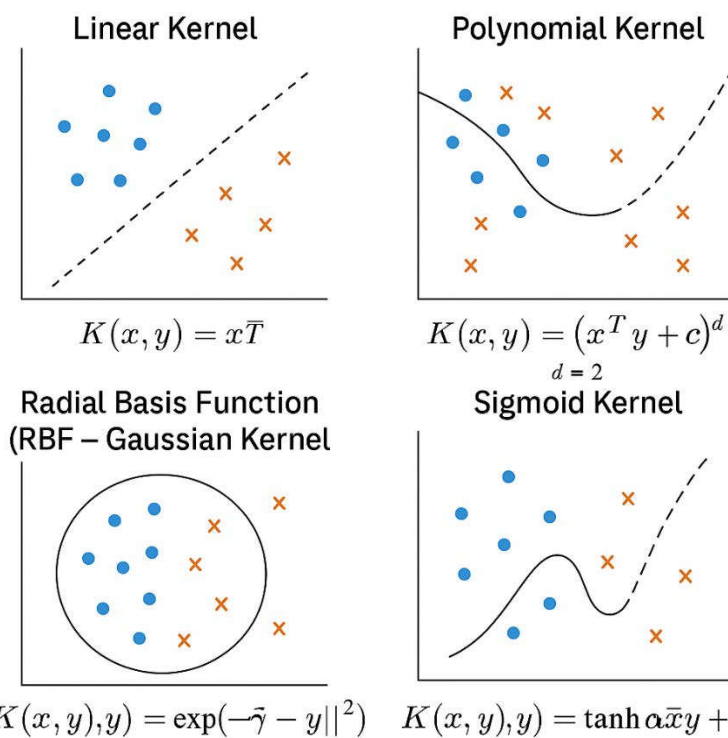
Η ιδέα πίσω από τη συνάρτηση πυρήνα είναι η εξής: Αν τα δεδομένα δεν μπορούν να διαχωριστούν με ένα απλό γραμμικό υπερεπίπεδο στον αρχικό τους χώρο, μπορούμε να τα απεικονίσουμε (mapping) σε έναν νέο, συνήθως υψηλότερων διαστάσεων χώρο, όπου

ενδεχομένως ο διαχωρισμός να είναι γραμμικά εφικτός. Η μαθηματική βάση αυτής της προσέγγισης δίνεται από την εξής σχέση (Schölkopf & Smola, 2002):

$$f(x) = w^T \phi(x) + b$$

όπου  $\phi(x)$  είναι η απεικόνιση (mapping) που μεταφέρει το διάνυσμα  $x$  σε ένα νέο χώρο χαρακτηριστικών,  $w$  είναι το διάνυσμα βαρών στο νέο αυτό χώρο και  $b$  η πόλωση (bias). Ωστόσο, η απευθείας χρήση της  $\phi(x)$  μπορεί να είναι υπολογιστικά ανέφικτη ή εξαιρετικά δαπανηρή λόγω της αύξησης των διαστάσεων. Το θεώρημα του Mercer επιτρέπει την

αντιμετώπιση αυτού του προβλήματος μέσω της χρήσης της συνάρτησης πυρήνα, η οποία υπολογίζει το αποτέλεσμα του εσωτερικού γινομένου στον νέο χώρο χωρίς τον άμεσο υπολογισμό της απεικόνισης (Mercer, 1909; Cristianini & Shawe-Taylor, 2000):



$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$$

Οι πιο διαδεδομένοι τύποι συναρτήσεων πυρήνα που χρησιμοποιούνται στις SVM είναι οι εξής (Boser, Guyon & Vapnik, 1992; Schölkopf & Smola, 2002):

### 1. Γραμμικός Πυρήνας (Linear Kernel):

Η πιο απλή μορφή

πυρήνα, η οποία δεν απαιτεί ιδιαίτερο μετασχηματισμό, είναι ο γραμμικός πυρήνας, ο οποίος δίνεται από τη σχέση:

$$K(x_i, x_j) = x_i^T x_j$$

Ο γραμμικός πυρήνας είναι κατάλληλος για προβλήματα που ήδη διαχωρίζονται γραμμικά ή που οι διαστάσεις των δεδομένων είναι πολύ μεγαλύτερες από τον αριθμό των δειγμάτων.

### 2. Πολυωνυμικός Πυρήνας (Polynomial Kernel):

Ο πολυωνυμικός πυρήνας είναι μία επέκταση του γραμμικού πυρήνα, όπου εισάγεται ένας βαθμός  $d$  και μία σταθερά  $c$ :

$$K(x_i, x_j) = (x_i^T x_j + c)^d$$

Ο πυρήνας αυτός επιτρέπει τη δημιουργία πολυωνυμικών ορίων απόφασης και είναι κατάλληλος για δεδομένα με σύνθετες, μη γραμμικές σχέσεις.

**3. Πυρήνας Ακτινικής Βάσης - RBF (Radial Basis Function Kernel):** Ο πυρήνας RBF είναι από τους πλέον δημοφιλείς και ευέλικτους, δίνεται από τη σχέση:

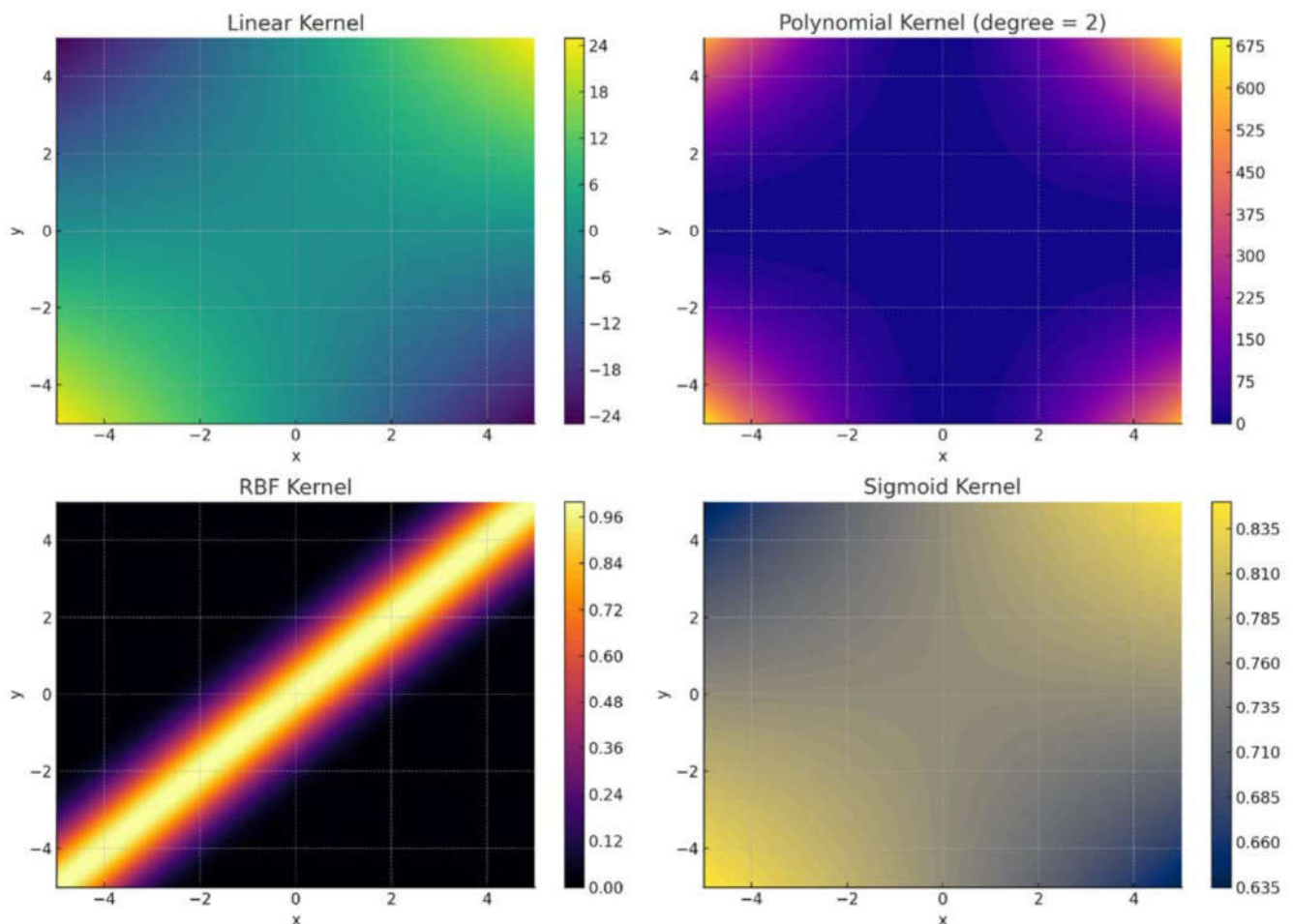
$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

Ο παράμετρος  $\gamma$  καθορίζει πόσο γρήγορα μειώνεται η επίδραση ενός δείγματος καθώς αυξάνει η απόσταση από το κέντρο. Είναι ιδιαίτερα αποτελεσματικός σε προβλήματα με πολύπλοκες γεωμετρικές δομές.

**4. Σιγμοειδής Πυρήνας (Sigmoid Kernel):** Ο σιγμοειδής πυρήνας, ο οποίος είναι εμπνευσμένος από τις νευρωνικές δομές, εκφράζεται ως εξής:

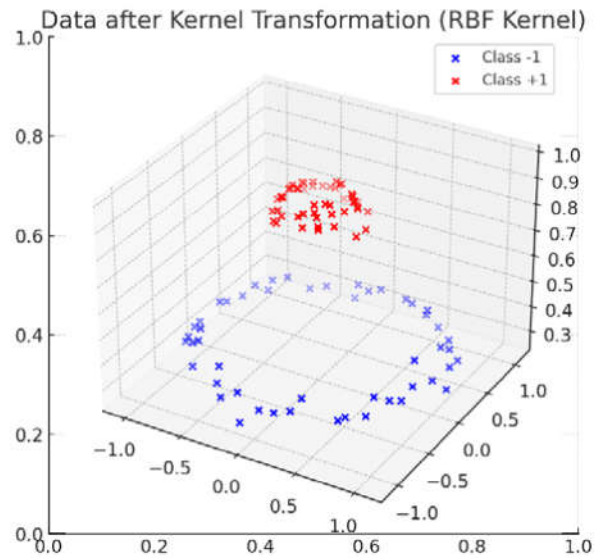
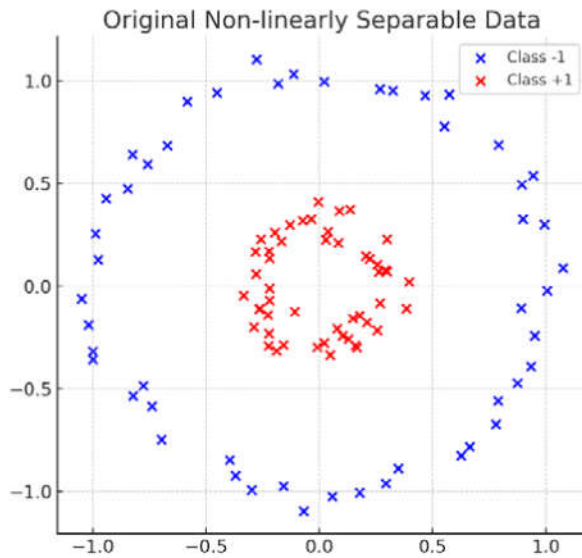
$$K(x_i, x_j) = \tanh(\alpha x_i^T x_j + c)$$

Όπου οι  $\alpha, c$  είναι παράμετροι που προσαρμόζονται.



*Sigmoid Kernel : Η μακέτα με τα τέσσερα διαγράμματα των συναρτήσεων πυρήνα*

Η επιλογή της κατάλληλης συνάρτησης πυρήνα είναι κρίσιμη για την αποτελεσματικότητα των SVM. Συνήθως, η επιλογή αυτή πραγματοποιείται μέσω τεχνικών διασταυρούμενης επικύρωσης (cross-validation), όπου διάφοροι πυρήνες και τιμές παραμέτρων δοκιμάζονται για τη βέλτιστη απόδοση (Hsu et al., 2010).



Η διαδικασία μετασχηματισμού των δεδομένων μέσω της συνάρτησης πυρήνα

Το παραπάνω διάγραμμα παρουσιάζει ένα παράδειγμα της μετατροπής μη γραμμικά διαχωρίσιμων δεδομένων μέσω μιας συνάρτησης πυρήνα (kernel).

- **Αριστερά:** Αρχικά δεδομένα που δεν μπορούν να διαχωριστούν γραμμικά στον δισδιάστατο χώρο.
- **Δεξιά:** Τα δεδομένα μετασχηματισμένα με τη χρήση του πυρήνα RBF (Radial Basis Function) σε έναν χώρο τριών διαστάσεων, όπου ο διαχωρισμός τους γίνεται εφικτός με ένα γραμμικό υπερεπίπεδο.

### Μέθοδοι Βελτιστοποίησης των SVM

Η υποστήριξη διανυσματικών μηχανών (Support Vector Machines - SVM) αποτελεί μία από τις πιο αποτελεσματικές και ευρέως χρησιμοποιούμενες τεχνικές μηχανικής μάθησης, η οποία βασίζεται στην επίλυση προβλημάτων βελτιστοποίησης, ειδικότερα τετραγωνικού προγραμματισμού (Quadratic Programming) (Cristianini & Shawe-Taylor, 2000; Vapnik, 1998). Στην πράξη, η υλοποίηση ενός SVM απαιτεί τη χρήση εξειδικευμένων αλγορίθμων βελτιστοποίησης για να επιτευχθεί γρήγορη και αποτελεσματική επίλυση του προβλήματος, το οποίο γίνεται υπολογιστικά απαιτητικό καθώς αυξάνεται ο αριθμός των δεδομένων εκπαίδευσης (Platt, 1998).

Η διαδικασία εκπαίδευσης ενός SVM είναι ουσιαστικά ένα πρόβλημα περιορισμένης βελτιστοποίησης (constrained optimization problem). Το αρχικό πρόβλημα βελτιστοποίησης για ένα γραμμικά διαχωρίσιμο SVM εκφράζεται ως εξής (Burges, 1998; Cortes & Vapnik, 1995):

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2, \quad \text{με περιορισμούς} \quad y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad \forall i = 1, \dots, N$$

όπου  $w$  είναι το διάνυσμα βαρών και  $b$  η πόλωση. Σε περιπτώσεις που τα δεδομένα δεν είναι γραμμικά διαχωρίσιμα, το πρόβλημα τροποποιείται εισάγοντας τις μεταβλητές χαλαρότητας (slack variables)  $\xi_i$ :

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i, \quad \text{με περιορισμούς} \quad y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i = 1, \dots, N$$

Το πρόβλημα αυτό είναι γνωστό ως soft-margin SVM και έχει την παράμετρο CCC, η οποία ελέγχει την ανοχή σε λάθη ταξινόμησης (Burges, 1998; Schölkopf & Smola, 2002).

**1. Τετραγωνικός Προγραμματισμός (Quadratic Programming - QP):** Ο τετραγωνικός προγραμματισμός αποτελεί την πιο κοινή μαθηματική μορφή που παίρνει το πρόβλημα βελτιστοποίησης των SVM. Γενικά, ένα πρόβλημα QP μπορεί να εκφραστεί ως:

$$\min_x \frac{1}{2} x^T Q x + c^T x$$

υπό τους περιορισμούς:

$$Ax \leq b, \quad Ex = d$$

όπου  $x$  είναι το διάνυσμα των μεταβλητών,  $Q$  ένας πίνακας ημιδιαθετικός θετικός (positive semidefinite) και οι υπόλοιποι όροι περιγράφουν γραμμικούς περιορισμούς (Cristianini & Shawe-Taylor, 2000).

Για τα SVM, το πρόβλημα QP προκύπτει μετά την εφαρμογή των πολλαπλασιαστών Lagrange, μετατρέποντας το αρχικό πρόβλημα σε μια δυϊκή μορφή που είναι πιο εύκολο να επιλυθεί υπολογιστικά:

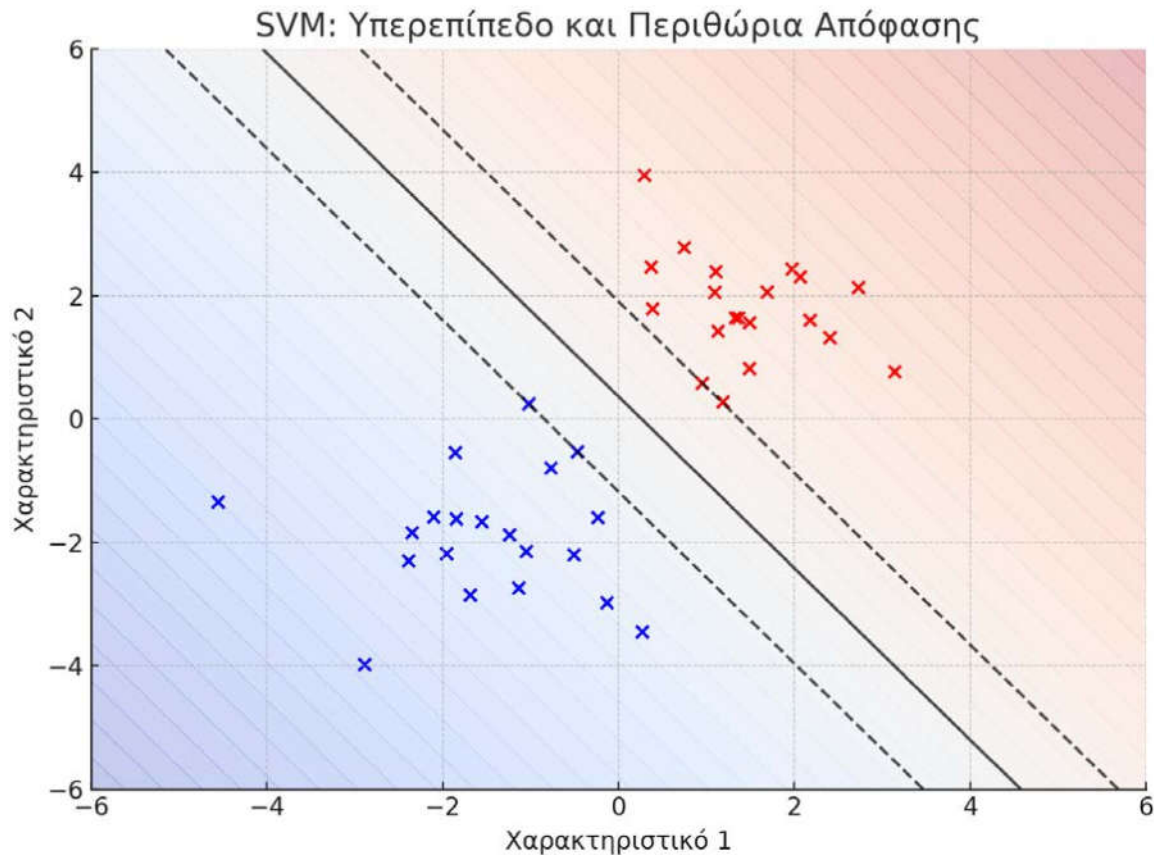
$$\max_{\alpha} \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (x_i^T x_j)$$

με περιορισμούς:

$$\sum_{i=1}^N \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad \forall i$$

Η λύση στο πρόβλημα αυτό δίνει τους πολλαπλασιαστές Lagrange ( $\alpha_i$ ), από τους οποίους προκύπτουν οι διανυσματικοί φορείς υποστήριξης (Cristianini & Shawe-Taylor, 2000). Η διαδικασία εκπαίδευσης των Υποστηρικτικών Διανυσματικών Μηχανών (Support Vector Machines - SVM) έχει ως βασικό στόχο την εξεύρεση του υπερεπιπέδου (hyperplane) που διαχωρίζει καλύτερα τις κατηγορίες των δεδομένων. Αυτή η διαδικασία

βασίζεται σε μεθόδους κυρτής βελτιστοποίησης και ειδικότερα, σε προβλήματα τετραγωνικού προγραμματισμού (Quadratic Programming - QP) (Vapnik, 1995).



Η παραπάνω εικόνα απεικονίζει το αποτέλεσμα της διαδικασίας εκπαίδευσης ενός γραμμικού SVM (Support Vector Machine):

- **Μαύρη ευθεία:** το υπερεπίπεδο απόφασης  $w^T x + b = 0$ ,
- **Διακεκομμένες γραμμές:** τα περιθώρια  $w^T x + b = \pm 1$ ,
- **Σημεία με κύκλο:** τα support vectors που καθορίζουν το περιθώριο.

**2. Αλγόριθμος Sequential Minimal Optimization (SMO):** Ο αλγόριθμος Sequential Minimal Optimization (SMO) είναι μια από τις πιο δημοφιλείς μεθόδους επίλυσης προβλημάτων βελτιστοποίησης των SVM, ειδικά λόγω της υψηλής αποδοτικότητας και της δυνατότητας κλιμάκωσης σε μεγάλα datasets (Platt, 1998).

Η βασική ιδέα του SMO είναι να επιλύει αναλυτικά μικρά υποπροβλήματα QP δύο μεταβλητών τη φορά. Συγκεκριμένα, αντί να επιλύεται το πρόβλημα QP για όλες τις μεταβλητές ταυτόχρονα, ο SMO επιλέγει δυο πολλαπλασιαστές Lagrange (αί) σε κάθε βήμα και επιλύει το πρόβλημα περιορισμένης βελτιστοποίησης για αυτές τις δύο μεταβλητές, κρατώντας τις υπόλοιπες σταθερές (Keerthi et al., 2001).

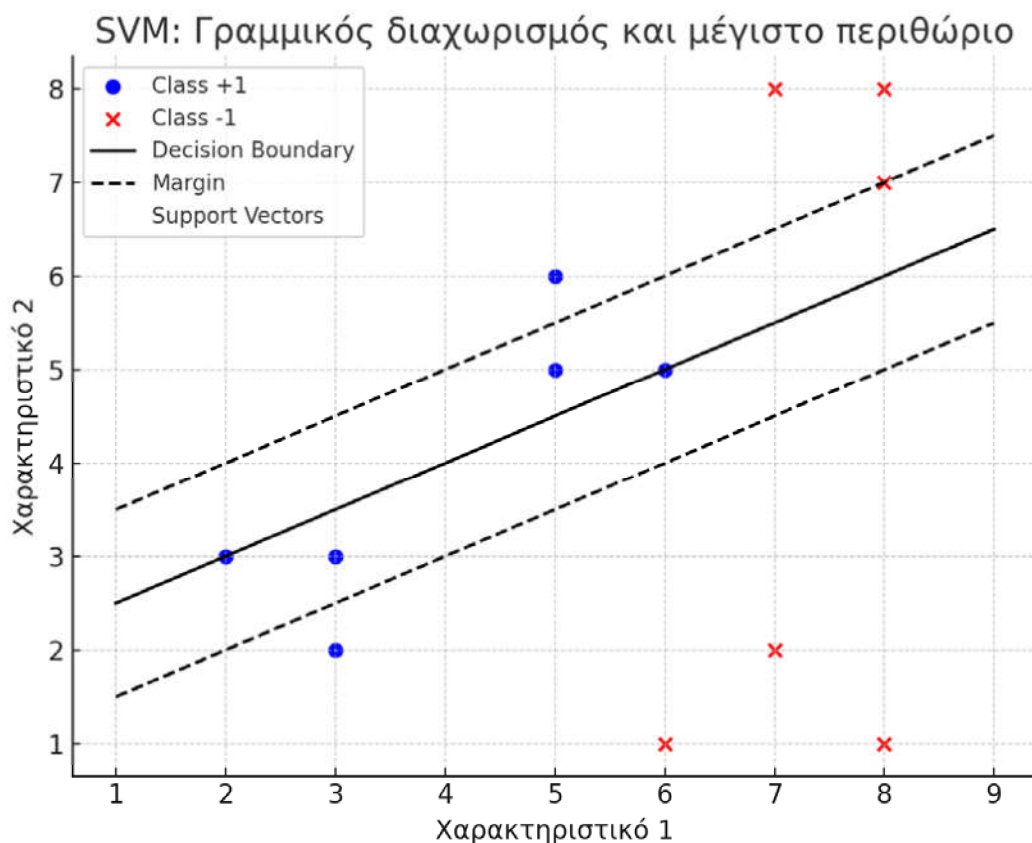
Ο αλγόριθμος SMO περιλαμβάνει τα παρακάτω βήματα:

1. Επιλογή ενός ζεύγους πολλαπλασιαστών ( $\alpha_i, \alpha_j$ ).
2. Επίλυση αναλυτικού υποπροβλήματος QP για αυτές τις δυο μεταβλητές.
3. Ενημέρωση των μεταβλητών και επανάληψη της διαδικασίας μέχρι να ικανοποιηθούν οι συνθήκες σύγκλισης (Platt, 1998).

Η επιλογή του ζεύγους μεταβλητών γίνεται συνήθως με βάση ευρετικές μεθόδους που διασφαλίζουν γρήγορη σύγκλιση (Keerthi et al., 2001).

Μια οπτικοποίηση της διαδικασίας βελτιστοποίησης, ειδικά του SMO, μπορεί να βοηθήσει στην καλύτερη κατανόηση των διαδικασιών επίλυσης. Σε τέτοια διαγράμματα, συνήθως φαίνεται η εξέλιξη των πολλαπλασιαστών Lagrange και η σταδιακή προσέγγιση της βέλτιστης λύσης.

Η επίλυση του προβλήματος βελτιστοποίησης που προκύπτει στις SVM είναι ένα κομβικό ζήτημα για την αποτελεσματική λειτουργία τους. Τεχνικές όπως ο τετραγωνικός προγραμματισμός και αλγόριθμοι όπως ο SMO έχουν συνεισφέρει σημαντικά στην ανάπτυξη και διάδοση των SVM στη μηχανική μάθηση.



Παραπάνω παρουσιάζεται το διάγραμμα που αναφέρεται στον γραμμικό διαχωρισμό και τη μεγιστοποίηση του περιθωρίου σε έναν SVM αλγόριθμο. Το διάγραμμα δείχνει καθαρά τις δύο διαφορετικές κατηγορίες (με μπλε κύκλους και κόκκινους σταυρούς), το βέλτιστο

υπερεπίπεδο διαχωρισμού (μαύρη συνεχής γραμμή), καθώς και τα δύο περιθώρια (μαύρες διακεκομμένες γραμμές). Οι υποστηρικτικοί φορείς (support vectors) σημειώνονται ειδικά με μεγαλύτερα και εντονότερα σημεία, καθώς αποτελούν τα πλησιέστερα στο υπερεπίπεδο σημεία, καθορίζοντας έτσι το περιθώριο του διαχωρισμού.

### Κριτήρια επιλογής συνάρτησης πυρήνα και υπερπαραμέτρων

Η επιτυχία ενός αλγορίθμου SVM εξαρτάται σε μεγάλο βαθμό από δύο βασικές σχεδιαστικές αποφάσεις:

(α) την επιλογή της συνάρτησης πυρήνα (kernel function) και  
(β) τη ρύθμιση των υπερπαραμέτρων (hyperparameters) που σχετίζονται τόσο με τον ίδιο τον SVM όσο και με τη συνάρτηση πυρήνα. Η διαδικασία αυτή δεν είναι απλώς τεχνική· εμπεριέχει τη στρατηγική αξιολόγηση της φύσης του προβλήματος, της κατανομής των δεδομένων και των διαθέσιμων υπολογιστικών πόρων (Schölkopf & Smola, 2002).

Η επιλογή της συνάρτησης πυρήνα οφείλει να λαμβάνει υπόψη το είδος της σχέσης που υπάρχει ανάμεσα στα χαρακτηριστικά και στις ετικέτες (labels):

- **Γραμμικός Πυρήνας (Linear Kernel):** Κατάλληλος όταν τα δεδομένα είναι σχεδόν γραμμικά διαχωρίσιμα. Ιδανικός για εφαρμογές όπου τα χαρακτηριστικά είναι ήδη κατάλληλα επεξεργασμένα και παρουσιάζουν απλή σχέση με την έξοδο, όπως σε προβλήματα ταξινόμησης κειμένου ή γονιδιακής έκφρασης (Joachims, 2002). Ο γραμμικός πυρήνας επιτρέπει και άμεση ερμηνεία των συντελεστών βαρών.
- **Πολυωνυμικός Πυρήνας (Polynomial Kernel):** Χρησιμοποιείται σε προβλήματα όπου υπάρχουν πολυωνυμικές σχέσεις ή αλληλεπιδράσεις μεταξύ μεταβλητών. Για παράδειγμα, σε οικονομικά ή φυσικά συστήματα όπου οι επιδράσεις δεν είναι γραμμικές αλλά διατηρούν πολυωνυμική δομή. Ο βαθμός του πολυωνύμου ddd καθορίζει το επίπεδο της μη γραμμικότητας.
- **Πυρήνας RBF (Radial Basis Function):** Ιδιαίτερα χρήσιμος σε προβλήματα με μη γραμμικές και περίπλοκες σχέσεις μεταξύ μεταβλητών. Ο RBF πυρήνας προβάλλει τα δεδομένα σε άπειρης διάστασης χώρο, δίνοντας εξαιρετική ικανότητα προσαρμογής. Είναι η πιο συχνή επιλογή όταν δεν έχουμε εις των προτέρων πληροφορία για τη φύση των σχέσεων (Cortes & Vapnik, 1995).

Το πλήθος και η διαστατικότητα των δεδομένων επηρεάζουν καθοριστικά την επιλογή πυρήνα:

- **Μεγάλα σύνολα δεδομένων** (πολλές χιλιάδες ή εκατομμύρια παρατηρήσεις): Ο γραμμικός πυρήνας είναι προτιμότερος, γιατί έχει χαμηλότερη υπολογιστική πολυπλοκότητα. Το γραμμικό SVM μπορεί επίσης να υλοποιηθεί αποδοτικά με αλγόριθμους όπως SGD (Stochastic Gradient Descent).
- **Μικρά έως μεσαία σύνολα:** Επιτρέπουν τη χρήση πιο πολύπλοκων πυρήνων, όπως ο RBF ή ο πολυωνυμικός, οι οποίοι έχουν μεγαλύτερη υπολογιστική απαίτηση αλλά και

μεγαλύτερη ισχύ έκφρασης. Στα μικρά δεδομένα, η υπερπροσαρμογή (overfitting) είναι πιο πιθανή, συνεπώς πρέπει να ρυθμίζονται με προσοχή οι υπερπαραμέτροι.

Οι πιο σημαντικές υπερπαραμέτροι είναι οι εξής:

- **Η παράμετρος C**

Ορίζει το trade-off ανάμεσα στην πολυπλοκότητα του μοντέλου και την ανοχή σε σφάλματα. Η αντικειμενική συνάρτηση γίνεται:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i$$

$$\text{υπό τον περιορισμό } y_i(\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

Όπου  $\xi_i$  είναι οι slack variables για την αποδοχή σφαλμάτων.

- Μεγάλο C: αυστηρή τιμωρία για τα σφάλματα  $\rightarrow$  πιθανό **overfitting**.
- Μικρό C: αποδοχή σφαλμάτων  $\rightarrow$  πιθανό **underfitting**.

- **Η παράμετρος  $\gamma$  στον RBF πυρήνα**

Καθορίζει τη "διάχυση" της επίδρασης κάθε δείγματος:

- Μεγάλο  $\gamma$ : μικρή ακτίνα επιρροής, πολύ λεπτομερές μοντέλο  $\rightarrow$  **overfitting**.
- Μικρό  $\gamma$ : μεγάλη ακτίνα, πιο απλοποιημένο μοντέλο  $\rightarrow$  **underfitting**.

Η συνάρτηση έχει τη μορφή:

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)$$

- **Ο βαθμός d στον Polynomial Kernel**

Ο βαθμός καθορίζει το επίπεδο πολυπλοκότητας:

- $d=1$ : γραμμικός διαχωρισμός.
- $d>2$ : μεγαλύτερη πολυπλοκότητα, αλλά και κίνδυνος υπερπροσαρμογής.

Η διασταυρούμενη επικύρωση (cross-validation) είναι η πιο διαδεδομένη μέθοδος για την πειραματική αξιολόγηση συνδυασμών πυρήνα-υπερπαραμέτρων. Διαδικασία:

1. Κατανομή του συνόλου εκπαίδευσης σε  $k$  υποσύνολα.
2. Εκπαίδευση σε  $k-1$  και επικύρωση στο υπόλοιπο 1.
3. Επανάληψη για όλες τις τιμές των υπερπαραμέτρων και των πυρήνων.

4. Επιλογή συνδυασμού με βάση τον μέσο όρο επιδόσεων.

### Χρησιμοποιούμενες Μετρικές Αξιολόγησης:

- **Ακρίβεια (Accuracy):** ποσοστό σωστά ταξινομημένων δειγμάτων.
- **F1-score:** συνδυάζει ακρίβεια (precision) και ανάκληση (recall).
- **ROC-AUC:** εμβαδόν κάτω από την καμπύλη ROC (για δυαδικά προβλήματα).

Το πλέγμα παραμέτρων (Grid Search) ή εξελεγμένες τεχνικές όπως Bayesian Optimization χρησιμοποιούνται για να επιλεγούν οι βέλτιστες τιμές C, γ, d κ.λπ.

Η επιλογή της συνάρτησης πυρήνα και η ρύθμιση των υπερπαραμέτρων αποτελούν κρίσιμους παράγοντες για την επίδοση των SVM. Η βέλτιστη στρατηγική προκύπτει μέσα από:

- θεωρητική κατανόηση της φύσης των δεδομένων,
- πειραματική δοκιμή εναλλακτικών πυρήνων και
- αυστηρό έλεγχο μέσω επικύρωσης και μετρικών απόδοσης.

Η προσεκτική ρύθμιση των παραμέτρων διασφαλίζει ότι ο SVM επιτυγχάνει μέγιστη γενίκευση με αποδεκτό ποσοστό σφαλμάτων.

### Μέθοδοι Αξιολόγησης και Μετρικές απόδοσης SVM

Η αξιολόγηση της απόδοσης ενός μοντέλου μηχανικής μάθησης αποτελεί κρίσιμο στάδιο στη διαδικασία ανάπτυξης και υλοποίησής του. Στην περίπτωση των Υποστηρικτικών Διανυσματικών Μηχανών (Support Vector Machines - SVM), οι μετρικές απόδοσης παρέχουν ουσιώδεις πληροφορίες για τη συμπεριφορά του μοντέλου τόσο στην ταξινόμηση όσο και στην παλινδρόμηση. Η ακρίβεια, η ανάκληση, η ακρίβεια πρόβλεψης (precision), η F1-score και η τεχνική διασταυρούμενης επικύρωσης (cross-validation) είναι μερικά από τα βασικότερα εργαλεία αξιολόγησης που χρησιμοποιούνται (Geron, 2019; James et al., 2021).

Η διασταυρούμενη επικύρωση είναι μια διαδικασία στατιστικής επαλήθευσης που στοχεύει να εκτιμήσει την επίδοση ενός αλγορίθμου πρόβλεψης, χρησιμοποιώντας διαφορετικά υποσύνολα του dataset (Hastie, Tibshirani & Friedman, 2009). Η πιο κοινή μορφή είναι η k-fold cross-validation, όπου το σύνολο δεδομένων διαχωρίζεται σε k ισομεγέθη τμήματα. Το μοντέλο εκπαιδεύεται σε k-1 από αυτά και δοκιμάζεται στο απομένον fold. Η διαδικασία επαναλαμβάνεται k φορές, έτσι ώστε κάθε τμήμα να χρησιμοποιηθεί μία φορά για επικύρωση.

Η επιλογή του αριθμού k εξαρτάται από το μέγεθος και τη φύση του dataset. Ένα συχνά χρησιμοποιούμενο όριο είναι το k=10 (10-fold cross-validation), το οποίο προσφέρει καλό συμβιβασμό μεταξύ bias και variance (Kohavi, 1995).

## Μετρικές Απόδοσης στην Ταξινόμηση

- **Accuracy (Ακρίβεια)**

Ορίζεται ως το ποσοστό των σωστά ταξινομημένων παραδειγμάτων προς το συνολικό αριθμό παραδειγμάτων:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Όπου:

- TP (True Positives): Πραγματικά θετικές και σωστά προβλεπόμενες τιμές.
- TN (True Negatives): Πραγματικά αρνητικές και σωστά προβλεπόμενες τιμές.
- FP (False Positives): Λανθασμένα προβλεπόμενες θετικές τιμές.
- FN (False Negatives): Λανθασμένα προβλεπόμενες αρνητικές τιμές.

Αν και εύκολη στην ερμηνεία, η ακρίβεια μπορεί να παραπλανήσει όταν τα δεδομένα είναι ανισόρροπα (imbalanced datasets) (Saito & Rehmsmeier, 2015).

- **Precision (Ακρίβεια Πρόβλεψης)**

Η ακρίβεια πρόβλεψης δείχνει το ποσοστό των θετικών προβλέψεων που είναι πράγματι σωστές:

$$Precision = \frac{TP}{TP + FP}$$

Είναι ιδιαίτερα χρήσιμη σε εφαρμογές όπου το κόστος των false positives είναι υψηλό, όπως στην ιατρική διάγνωση ή στην ανίχνευση απάτης (Aggarwal, 2015).

- **Recall (Ανάκληση)**

Η ανάκληση μετρά την ικανότητα του μοντέλου να αναγνωρίζει σωστά τις θετικές περιπτώσεις:

$$Recall = \frac{TP}{TP + FN}$$

Χρησιμοποιείται όταν η μη αναγνώριση μιας θετικής περίπτωσης είναι περισσότερο επιζήμια από ένα false positive.

- **F1-score**

Ο F1-score είναι ο αρμονικός μέσος της precision και της recall, δίνοντας μία ισορροπημένη θεώρηση:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Ο F1-score είναι χρήσιμος σε περιβάλλοντα με ανισόρροπες κλάσεις και απώλειες που δεν είναι συμμετρικές (Powers, 2011).

Οι SVM τείνουν να αποδίδουν καλύτερα όταν ο χώρος χαρακτηριστικών έχει υψηλή διαχωρισιμότητα. Επομένως, μετρικές όπως η ακρίβεια και ο F1-score μπορούν να δώσουν καλύτερη εικόνα για την απόδοση του μοντέλου συγκριτικά με το απλό accuracy, ειδικά σε περιπτώσεις με skewed δεδομένα (Schölkopf & Smola, 2002).

Η χρήση διασταυρούμενης επικύρωσης, σε συνδυασμό με βελτιστοποίηση υπερπαραμέτρων (όπως το C και το γ), επιτρέπει τη συστηματική εκπαίδευση και γενίκευση του μοντέλου σε νέα δεδομένα (Burges, 1998).

Η confusion matrix (ή πίνακας σύγχυσης) είναι ένα από τα πιο σημαντικά εργαλεία στη στατιστική ανάλυση και την εποπτευόμενη μάθηση, ειδικά όταν πρόκειται για προβλήματα δυαδικής ή πολυκατηγορικής ταξινόμησης. Ουσιαστικά, προσφέρει μια λεπτομερή ανάλυση της απόδοσης ενός ταξινομητή, παρουσιάζοντας όχι μόνο την ακρίβεια, αλλά και τις συγκεκριμένες κατηγορίες σφαλμάτων που κάνει το μοντέλο. Η confusion matrix είναι ένας τετραγωνικός πίνακας με διαστάσεις ίσες με τον αριθμό των κατηγοριών (classes) που προβλέπει το μοντέλο. Για ένα δυαδικό ταξινομητή, η τυπική μορφή της είναι 2x2 και έχει την εξής μορφή:

	Πραγματική Θετική (Positive)	Πραγματική Αρνητική (Negative)
Προβλεπόμενη Θετική	True Positive (TP)	False Positive (FP)
Προβλεπόμενη Αρνητική	False Negative (FN)	True Negative (TN)

Κάθε κελί του πίνακα αντιπροσωπεύει τον αριθμό των παραδειγμάτων που ανήκουν σε μία κατηγορία αλλά ταξινομήθηκαν (σωστά ή λανθασμένα) σε μια άλλη από το μοντέλο.

**Τι μετράει η Confusion Matrix;**

**1. True Positives (TP):** Περιπτώσεις όπου το μοντέλο **πρόβλεψε σωστά** την **θετική κλάση**. Παράδειγμα: το SVM προβλέπει ότι υπάρχει spam και πράγματι υπάρχει.

**2. True Negatives (TN):** Περιπτώσεις όπου το μοντέλο **πρόβλεψε σωστά** την **αρνητική κλάση**. Παράδειγμα: το SVM προβλέπει ότι δεν υπάρχει spam και πράγματι δεν υπάρχει.

**3. False Positives (FP):** Περιπτώσεις όπου το μοντέλο **πρόβλεψε λάθος** ως **θετική κλάση**. Παράδειγμα: το μοντέλο λέει ότι υπάρχει spam, αλλά στην πραγματικότητα δεν υπάρχει.

**4. False Negatives (FN):** Περιπτώσεις όπου το μοντέλο **πρόβλεψε λάθος** ως **αρνητική κλάση**. Παράδειγμα: το μοντέλο λέει ότι δεν υπάρχει spam, αλλά υπάρχει.

## Πλεονεκτήματα της Confusion Matrix

1. **Ακρίβεια κατηγοριοποίησης:** Δεν περιορίζεται σε μια συνολική μέτρηση (όπως η accuracy), αλλά προσφέρει ανάλυση για κάθε είδος σφάλματος.
2. **Ισορροπία μετρικών:** Χρήσιμη σε datasets με ανισόρροπες κλάσεις, καθώς επισημαίνει το κόστος των σφαλμάτων.
3. **Σύγκριση μοντέλων:** Διευκολύνει τη σύγκριση διαφορετικών μοντέλων ή παραμέτρων, π.χ., μεταξύ SVM με διαφορετικούς πυρήνες ή τιμές του C.

Οι Support Vector Machines είναι ακριβείς ταξινομητές, αλλά όπως κάθε μοντέλο, μπορούν να παρουν σφάλματα. Η confusion matrix χρησιμοποιείται για:

- **Βελτίωση υπερπαραμέτρων (π.χ., C, γ):** Χρησιμοποιώντας cross-validation σε συνδυασμό με την ανάλυση της confusion matrix.
- **Ανίχνευση bias του μοντέλου:** Αν ένα SVM προβλέπει συνεχώς την αρνητική κλάση σε ανισόρροπα δεδομένα, η confusion matrix θα δείξει αυξημένα FN.
- **Σύγκριση SVM με άλλους αλγορίθμους:** π.χ., Naive Bayes, Logistic Regression, k-NN.

## Παράδειγμα Confusion Matrix με Python (scikit-learn)

```
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
from sklearn.svm import SVC
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
# Φόρτωση δεδομένων και δημιουργία μοντέλου SVM
X, y = load_iris(return_X_y=True)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
svm_model = SVC(kernel='linear').fit(X_train, y_train)

# Πρόβλεψη και εμφάνιση confusion matrix
y_pred = svm_model.predict(X_test)
cm = confusion_matrix(y_test, y_pred)
disp = ConfusionMatrixDisplay(confusion_matrix=cm)
disp.plot()
```

Το confusion matrix δεν είναι απλώς ένα εργαλείο αποτίμησης απόδοσης, αλλά και ένα διαγνωστικό εργαλείο που αποκαλύπτει τις αδυναμίες και τις προκαταλήψεις ενός SVM

μοντέλου. Η κατανόησή της επιτρέπει τη βελτιστοποίηση παραμέτρων, την έγκαιρη αναγνώριση υποεκπροσώπησης και την ενίσχυση της γενίκευσης.

Η ολοκληρωμένη αξιολόγηση ενός SVM μοντέλου απαιτεί τον συνδυασμό ποσοτικών μετρικών (precision, recall, F1) και μεθοδολογιών (cross-validation), προκειμένου να διασφαλιστεί η γενίκευση και αξιοπιστία του σε νέα δεδομένα. Οι επιλογές αυτές εξαρτώνται από την φύση του προβλήματος, την κατανομή των κλάσεων και τους επιχειρησιακούς στόχους.

### 4.3.2 Πλεονεκτήματα και Μειονεκτήματα των SVM

Οι Υποστηρικτικές Διανυσματικές Μηχανές (*Support Vector Machines - SVM*) αποτελούν έναν από τους πιο ισχυρούς εποπτευόμενους αλγόριθμους μηχανικής μάθησης, ιδιαίτερα σε προβλήματα δυαδικής ταξινόμησης. Εισήχθησαν από τον Vladimir Vapnik και τους συνεργάτες του (Cortes & Vapnik, 1995) και έχουν εφαρμοστεί με επιτυχία σε διάφορους τομείς, όπως αναγνώριση προτύπων, ανάλυση συναισθήματος, ανίχνευση απάτης, ιατρική διάγνωση και ταξινόμηση περιεχομένου στα social media.

Ωστόσο, όπως κάθε αλγόριθμος, οι SVM δεν είναι απαλλαγμένες από περιορισμούς. Η επιλογή τους πρέπει να βασίζεται σε μια εμπεριστατωμένη αξιολόγηση του προβλήματος, του μεγέθους των δεδομένων, της γραμμικότητας των χαρακτηριστικών και της ερμηνευσιμότητας του αποτελέσματος.

#### Πλεονεκτήματα των SVM

- 1. Αποτελεσματική Γενίκευση:** Οι SVM παρουσιάζουν υψηλή ικανότητα γενίκευσης σε προβλήματα με χαμηλή έως μεσαία διάσταση, ακόμη και όταν το μέγεθος του συνόλου εκπαίδευσης είναι περιορισμένο (Cristianini & Shawe-Taylor, 2000). Αυτό οφείλεται στη μεγιστοποίηση του περιθωρίου μεταξύ των κλάσεων.
- 2. Λειτουργία σε Μη Γραμμικά Προβλήματα:** Η χρήση συναρτήσεων πυρήνα (kernel functions) επιτρέπει στους SVM να μετασχηματίζουν τα δεδομένα σε έναν υψηλότερων διαστάσεων χώρο, όπου η ταξινόμηση γίνεται γραμμικά (Schölkopf & Smola, 2002). Αυτή η προσέγγιση επιτρέπει την αντιμετώπιση σύνθετων, μη γραμμικών προβλημάτων.
- 3. Ανθεκτικότητα στον Υπερπροσδιορισμό (Overfitting):** Με τη σωστή επιλογή των υπερπαραμέτρων ( $C$ ,  $\gamma$ ), οι SVM τείνουν να είναι ανθεκτικοί στην υπερπροσαρμογή, σε αντίθεση με άλλους αλγόριθμους που εξαρτώνται από τη μορφή της κατανομής των δεδομένων (Hastie et al., 2009).

**4. Καλή Απόδοση σε Υψηλής Διάστασης Χώρους:** Οι SVM λειτουργούν καλά ακόμη και όταν ο αριθμός των χαρακτηριστικών είναι μεγαλύτερος από τον αριθμό των παραδειγμάτων, π.χ. σε αναλύσεις γονιδιακής έκφρασης (Guyon et al., 2002).

**5. Επεκτασιμότητα σε Πολυκατηγορικά Προβλήματα:** Αν και σχεδιάστηκαν αρχικά για δυαδική ταξινόμηση, μπορούν να προσαρμοστούν σε πολυκατηγορικά προβλήματα μέσω τεχνικών όπως *one-vs-all* και *one-vs-one* (Hsu & Lin, 2002).

**6. Μικρότερη Εξάρτηση από την Κατανομή των Δεδομένων|:** Σε αντίθεση με τους γενετικούς αλγορίθμους ή τα νευρωνικά δίκτυα, οι SVM δεν απαιτούν προϋποθέσεις ως προς την κατανομή των δεδομένων (Boser et al., 1992).

### Μειονεκτήματα των SVM

**1. Υψηλή Υπολογιστική Πολυπλοκότητα:** Η εκπαίδευση ενός SVM, ειδικά με μη γραμμικούς πυρήνες, μπορεί να είναι **χρονοβόρα**, κυρίως για μεγάλα σύνολα δεδομένων (Chang & Lin, 2011). Η υπολογιστική πολυπλοκότητα αυξάνεται σημαντικά λόγω του τετραγωνικού προγραμματισμού που απαιτείται.

**2. Δυσκολία στην Επιλογή Σωστής Συνάρτησης Πυρήνα:** Η απόδοση των SVM εξαρτάται έντονα από την επιλογή της συνάρτησης πυρήνα και των υπερπαραμέτρων της. Η λανθασμένη επιλογή μπορεί να μειώσει σημαντικά την ακρίβεια (Chapelle et al., 2002).

**3. Μειωμένη Ερμηνευσιμότητα:** Σε αντίθεση με τα **αποφασιστικά δέντρα**, οι SVM παρέχουν περιορισμένη **ερμηνευσιμότητα**, καθιστώντας δύσκολη την κατανόηση του τρόπου με τον οποίο λαμβάνεται η απόφαση (Doshi-Velez & Kim, 2017).

**4. Απόδοση σε Ανισόρροπα Δεδομένα:** Σε περιπτώσεις όπου οι κλάσεις είναι **ανισόρροπες**, οι SVM μπορεί να παρουσιάσουν μεροληψία υπέρ της πλειοψηφικής κλάσης, εκτός αν εφαρμοστούν ειδικές τεχνικές εξισορρόπησης (He & Garcia, 2009).

**5. Ευαισθησία στον Θόρυβο και στις Εκτοπισμένες Τιμές (Outliers):** Δεδομένα που βρίσκονται **κοντά στο υπερ-επίπεδο διαχωρισμού** μπορούν να επηρεάσουν σημαντικά το αποτέλεσμα, καθώς τα SVM βασίζονται στους *support vectors*, τα σημεία που καθορίζουν το περιθώριο (Schölkopf & Smola, 2002).

**6. Δυσκολία σε Online Learning:** Σε αντίθεση με άλλους αλγορίθμους, οι SVM δεν υποστηρίζουν καλά την **σταδιακή μάθηση (incremental learning)**, γεγονός που περιορίζει τη χρήση τους σε εφαρμογές streaming δεδομένων (Cauwenberghs & Poggio, 2001).

## Συνοπτικός Πίνακας Πλεονεκτημάτων - Μειονεκτημάτων

Πλεονεκτήματα	Μειονεκτήματα
Υψηλή Γενίκευση	Υψηλή Υπολογιστική Πολυπλοκότητα
Μη Γραμμική Ικανότητα μέσω Kernel	Δυσκολία στην Επιλογή Πυρήνα και Υπερπαραμέτρων
Ανθεκτικότητα στο Overfitting	Χαμηλή Ερμηνευσιμότητα
Ικανοποιητική Απόδοση σε Υψηλές Διαστάσεις	Ευαισθησία σε Ανισόρροπα Δεδομένα και Εκτοπισμένες Τιμές
Προσαρμογή σε Πολυκατηγορική Ταξινόμηση	Δυσκολία σε Incremental/Online Learning

### Σύγκριση με άλλους αλγορίθμους ταξινόμησης

Οι Support Vector Machines (SVM), όπως αναλύθηκε σε προηγούμενα κεφάλαια, αποτελούν μία από τις πιο αξιόπιστες μεθόδους εποπτευόμενης μάθησης για προβλήματα ταξινόμησης. Ωστόσο, δεν είναι η μοναδική. Υπάρχουν άλλοι διαδεδομένοι αλγόριθμοι ταξινόμησης που χρησιμοποιούνται ανάλογα με το είδος των δεδομένων, τις απαιτήσεις του προβλήματος και τους υπολογιστικούς πόρους. Στην παρούσα ενότητα, εξετάζουμε τη σύγκριση των SVM με:

- **Logistic Regression (LR)**
- **Random Forest (RF)**
- **k-Nearest Neighbors (k-NN)**
- **Νευρωνικά Δίκτυα (Neural Networks - NN)**

### **1. SVM vs Logistic Regression (LR)**

Χαρακτηριστικό	Support Vector Machines	Logistic Regression
Γενίκευση	Πολύ καλή σε υψηλές διαστάσεις	Καλή, αλλά όχι σε μη γραμμικά προβλήματα
Γραμμικός Διαχωρισμός	Υποστηρίζεται και βελτιώνεται μέσω μέγιστου περιθωρίου	Γραμμικός μόνο
Μη Γραμμικά Προβλήματα	Ναι, μέσω Kernel Trick	Όχι
Ερμηνευσιμότητα	Περιορισμένη	Υψηλή
Υπολογιστική Πολυπλοκότητα	Υψηλή	Χαμηλή

Η λογιστική παλινδρόμηση προτιμάται σε προβλήματα όπου η ερμηνευσιμότητα είναι κρίσιμη (π.χ. ιατρική), ενώ οι SVM υπερτερούν όταν το σύνολο δεδομένων είναι πιο σύνθετο ή μη γραμμικό (Hosmer et al., 2013; Cortes & Vapnik, 1995).

## 2. SVM vs Random Forest (RF)

Χαρακτηριστικό	Support Vector Machines	Random Forest
Ανθεκτικότητα σε Outliers	Ευαίσθητη	Πολύ ανθεκτική
Γραμμικότητα	Ισχυρή σε μη γραμμικά	Χωρίς προϋποθέσεις γραμμικότητας
Ακρίβεια	Πολύ υψηλή σε μικρά dataset	Υψηλή σε μεγάλα dataset
Χρόνος Εκπαίδευσης	Μεγάλος	Ταχύτερος σε μεγάλα δεδομένα
Ερμηνευσιμότητα	Περιορισμένη	Μέτρια

Οι Random Forests, ως ensemble μέθοδοι, είναι πιο ανθεκτικές στον θόρυβο και τις ακραίες τιμές, ενώ οι SVM παρέχουν καλύτερα αποτελέσματα σε προβλήματα υψηλής διάστασης με σαφή όρια διαχωρισμού (Breiman, 2001).

## 3. SVM vs k-Nearest Neighbors (k-NN)

Χαρακτηριστικό	Support Vector Machines	k-NN
Μάθηση	Batch learning	Lazy learning
Απόδοση σε υψηλές διαστάσεις	Καλή	Πολύ φτωχή (curse of dimensionality)
Ταχύτητα πρόβλεψης	Γρήγορη	Αργή
Κατανόηση ορίων απόφασης	Ξεκάθαρα όρια	Ασαφή όρια

Ο k-NN είναι εύκολος στην υλοποίηση και δεν απαιτεί εκπαίδευση, αλλά υστερεί σημαντικά σε υψηλής διάστασης χώρους, όπου οι SVM υπερτερούν (Cover & Hart, 1967; Duda et al., 2000).

## 4. SVM vs Νευρωνικά Δίκτυα (Neural Networks)

Χαρακτηριστικό	Support Vector Machines	Neural Networks
Απαιτήσεις σε δεδομένα	Λιγότερες	Μεγάλες
Μη Γραμμικά Προβλήματα	Ναι, μέσω Kernel	Ναι, φυσικά
Εκπαίδευση	Ακριβής αλλά βαριά	Χρονοβόρα και ευαίσθητη σε υπερπροσαρμογή
Ερμηνευσιμότητα	Μέτρια	Πολύ χαμηλή (black-box)
Κλιμάκωση	Περιορισμένη	Πολύ καλή (deep learning)

Τα νευρωνικά δίκτυα είναι κατάλληλα για προβλήματα με τεράστιες ποσότητες δεδομένων, όπως η αναγνώριση εικόνων ή φυσική γλώσσα. Οι SVM παραμένουν ισχυροί όταν υπάρχει

ανάγκη για προβλεψιμότητα και μικρότερο υπολογιστικό κόστος σε μέτρια dataset (LeCun et al., 2015).

Η χρήση SVM είναι ιδανική στις παρακάτω περιπτώσεις:

- Μικρά ή μεσαία σύνολα δεδομένων, όπου άλλες μέθοδοι (π.χ. NN) υπερπροσαρμόζουν.
- Υψηλής διάστασης προβλήματα (π.χ. γονιδιωματική, αναγνώριση κειμένου).
- Μη γραμμικά προβλήματα με σαφή διαχωρισμό των τάξεων.
- Όταν απαιτείται υψηλή ακρίβεια και όχι απαραίτητα ερμηνευσιμότητα.
- Όταν τα δεδομένα έχουν περιορισμένο θόρυβο και ισορροπία μεταξύ τάξεων.

Οι SVM βρίσκουν εφαρμογή σε μεγάλο εύρος προβλημάτων, κυρίως εκεί που άλλες μέθοδοι αποτυγχάνουν να διαχωρίσουν κλάσεις σε υψηλής διάστασης χώρους ή όταν δεν υπάρχει μεγάλη ποσότητα δεδομένων διαθέσιμη. Ωστόσο, δεν αντικαθιστούν πλήρως τις άλλες τεχνικές. Η τελική επιλογή εξαρτάται από τη φύση του προβλήματος, τις απαιτήσεις της εφαρμογής και τους διαθέσιμους πόρους.

### Αντιμετώπιση προβλημάτων και προκλήσεων

#### **Αντιμετώπιση Ανισόρροπων Κλάσεων (Imbalanced Datasets)**

Ένα από τα βασικά προβλήματα που αντιμετωπίζουν οι SVM είναι η ανισορροπία των κλάσεων, δηλαδή όταν μία κλάση (συνήθως η θετική) αντιπροσωπεύεται από πολύ λιγότερα παραδείγματα σε σχέση με την άλλη (αρνητική). Αυτό το φαινόμενο είναι συχνό σε εφαρμογές όπως:

- Ανίχνευση απάτης (fraud detection),
- Αναγνώριση σπάνιων ασθενειών,
- Φιλτράρισμα ανεπιθύμητης αλληλογραφίας (spam filtering),
- Αναγνώριση συναισθήματος σε social media.

#### **Προβλήματα που προκαλούνται:**

- Το μοντέλο τείνει να προβλέπει υπερβολικά την πλειοψηφική κλάση.
- Η μέγιστη ακρίβεια δεν αντανακλά την πραγματική απόδοση, αφού το μοντέλο μπορεί να αγνοεί πλήρως την ελάσσονα κλάση και παρ' όλα αυτά να φαίνεται «σωστό» τις περισσότερες φορές (He & Garcia, 2009).

## Τεχνικές Αντιμετώπισης:

**1. Επανασταθμισμένη συνάρτηση κόστους (Cost-sensitive SVM):** Η συνάρτηση κόστους του SVM προσαρμόζεται ώστε να επιβάλλει μεγαλύτερη ποινή στην εσφαλμένη ταξινόμηση παραδειγμάτων από την ελάσσονα κλάση. Η παραλλαγή αυτή τροποποιεί την αντικειμενική συνάρτηση ως εξής:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C_+ \sum_{i \in \text{positive}} \xi_i + C_- \sum_{j \in \text{negative}} \xi_j$$

όπου  $C_+ > C_-$ , ώστε η ποινή για τις θετικές (σπάνιες) παρατηρήσεις να είναι υψηλότερη.

## 2. Oversampling/Undersampling:

- Το SMOTE (Synthetic Minority Oversampling Technique) είναι μια κλασική μέθοδος που δημιουργεί συνθετικά δεδομένα για την ελάσσονα κλάση (Chawla et al., 2002).
- Το random undersampling αφαιρεί δείγματα από την πλειοψηφική κλάση για ισορροπία, αν και μπορεί να οδηγήσει σε απώλεια πληροφορίας.

**3. Χρήση ειδικών μετρικών απόδοσης:** Αποφεύγεται η χρήση της ακρίβειας και προτιμώνται F1-score, AUC-ROC και recall, οι οποίες αντανακλούν καλύτερα την απόδοση στα ανισόρροπα σύνολα (Saito & Rehmsmeier, 2015).

## Μείωση Πολυπλοκότητας (Regularization)

Η τακτοποίηση (regularization) αποτελεί βασικό εργαλείο για την αποτροπή της υπερπροσαρμογής (overfitting) και τη βελτίωση της γενίκευσης ενός SVM.

**1. Ο ρόλος της παραμέτρου C:** Η βασική υπερπαραμέτρος στον SVM είναι η παράμετρος regularization C, η οποία ρυθμίζει το trade-off μεταξύ μέγιστου περιθωρίου και ελάχιστου σφάλματος ταξινόμησης.

Η αντικειμενική συνάρτηση γίνεται:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad \text{με περιορισμό } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i$$

Όπου:

- **Μικρό C** → μεγάλο περιθώριο, λιγότερη έμφαση στο λάθος → **καλύτερη γενίκευση** αλλά περισσότερο bias.

- **Μεγάλο  $C$**   $\rightarrow$  ποινή για κάθε λάθος  $\rightarrow$  **χαμηλότερο bias** αλλά ενδεχόμενο overfitting (Cristianini & Shawe-Taylor, 2000).

**2. Kernel Regularization:** Στον μη γραμμικό διαχωρισμό με RBF kernel, η παράμετρος  $\gamma$  ελέγχει τη συμπεριφορά της καμπύλης του διαχωρισμού:

- **Υψηλή  $\gamma$**   $\rightarrow$  περιορισμένη επιρροή κάθε δείγματος  $\rightarrow$  overfitting.
- **Χαμηλή  $\gamma$**   $\rightarrow$  μεγάλη επιρροή  $\rightarrow$  underfitting.

**3. Επιλογή Regularization μέσω Cross-Validation:** Η διασταυρούμενη επικύρωση χρησιμοποιείται για την επιλογή του κατάλληλου  $C$  και  $\gamma$  μέσω πλέγματος τιμών (grid search). Η χρήση τεχνικών όπως k-fold cross-validation εξασφαλίζει ότι η τελική επιλογή παραμέτρων έχει βέλτιστη γενίκευση (Bergstra & Bengio, 2012).

Οι SVM, αν και ισχυροί, απαιτούν ιδιαίτερη προσοχή όταν εφαρμόζονται σε ανισόρροπα δεδομένα ή σύνθετα προβλήματα. Η αποτελεσματική αντιμετώπιση των προκλήσεων μέσω:

- προσαρμοσμένων συναρτήσεων κόστους,
- υπερπαραμετρικής βελτιστοποίησης ( $C, \gamma$ ),
- και εναλλακτικών μετρικών αξιολόγησης,

μπορεί να αυξήσει σημαντικά την αξιοπιστία και την πρακτική απόδοση του μοντέλου.

#### 4.3.3 Case study: Εφαρμογή των SVM σε πραγματικά δεδομένα Social Media

Η εκρηκτική αύξηση των δεδομένων στα μέσα κοινωνικής δικτύωσης καθιστά επιτακτική την ανάγκη για τεχνικές επεξεργασίας, κατανόησης και εξαγωγής πληροφορίας. Η Υποστήριξη Διανυσματικών Μηχανών (Support Vector Machines - SVM) έχει αποδειχθεί εξαιρετικά αποτελεσματική στην κατηγοριοποίηση αυτών των δεδομένων, ιδίως σε εργασίες όπως:

- Ανάλυση συναισθήματος (sentiment analysis),
- Ταξινόμηση τοξικού λόγου (toxic comment classification),
- Ανίχνευση παραπληροφόρησης ή ψευδών ειδήσεων (fake news detection),
- Κατηγοριοποίηση θεμάτων (topic classification),
- Ανίχνευση bots και μη αυθεντικής δραστηριότητας.

Στη συνέχεια, περιγράφεται αναλυτικά η διαδικασία εφαρμογής ενός SVM σε ένα πλήρες παράδειγμα ανάλυσης συναισθήματος στο Twitter.

## **Βήματα Εφαρμογής SVM σε Δεδομένα Social Media**

1. **Συλλογή Δεδομένων** Τα δεδομένα μπορεί να προέρχονται από Tweets, Facebook posts, σχόλια σε βίντεο ή reviews σε πλατφόρμες όπως YouTube, TripAdvisor ή Amazon. Τυπικά περιλαμβάνουν ένα κείμενο (post, comment) και μια ετικέτα (θετικό, αρνητικό, ουδέτερο συναίσθημα).
2. **Προεπεξεργασία Κειμένου** Περιλαμβάνει:
  - ο Καθαρισμό (αφαίρεση stopwords, συμβόλων, URLs),
  - ο Λεξικολογική ανάλυση (lemmatization, stemming),
  - ο Μετατροπή σε κατάλληλη αριθμητική μορφή μέσω τεχνικών όπως Bag-of-Words, TF-IDF ή ενσωμάτωση λέξεων (word embeddings).
3. **Διαχωρισμός Εκπαίδευσης και Δοκιμής** Χρησιμοποιείται συνήθως 70-80% για εκπαίδευση και 20-30% για δοκιμή, με τεχνικές διαστρωματικής δειγματοληψίας για ισορροπημένες κλάσεις.
4. **Εκπαίδευση SVM Μοντέλου** Ορίζεται το kernel (συνήθως RBF) και γίνεται επιλογή υπερπαραμέτρων C και γ μέσω διασταυρούμενης επικύρωσης.
5. **Αξιολόγηση Απόδοσης** Εκτός από τον υπολογισμό ακρίβειας, recall, precision και F1-score, μια Confusion Matrix δίνει οπτικά την εικόνα των TP (True Positives), TN (True Negatives), FP (False Positives) και FN (False Negatives). Το παραπάνω γράφημα δείχνει ξεκάθαρα το ποσοστό σωστών και λανθασμένων ταξινομήσεων του μοντέλου.
6. **Βελτιστοποίηση & Τελική Ανάπτυξη** Ανάλογα με τις απαιτήσεις εφαρμογής, γίνεται fine-tuning των παραμέτρων ή και σύγκριση με άλλους ταξινομητές όπως Random Forest ή Naive Bayes.

## 4.4 *k*-Nearest Neighbors (*k*-NN)

Ο αλγόριθμος *k*-Nearest Neighbors (*k*-NN) είναι μία από τις απλούστερες και πιο ευρέως γνωστές μεθόδους επιβλεπόμενης μάθησης, με εφαρμογές τόσο στην ταξινόμηση (classification) όσο και στην παλινδρόμηση (regression). Πρόκειται για έναν μη παραμετρικό αλγόριθμο “τεμπέλικης” μάθησης (*lazy learning*), καθώς δεν πραγματοποιεί ρητό στάδιο εκπαίδευσης αλλά αποθηκεύει απλώς τα δεδομένα εκπαίδευσης και στη συνέχεια λαμβάνει αποφάσεις κατά το χρόνο πρόβλεψης. Η θεμελιώδης υπόθεση πίσω από τον *k*-NN είναι ότι τα δείγματα με παρόμοια χαρακτηριστικά (δηλ. κοντά μεταξύ τους στον χώρο των χαρακτηριστικών) έχουν και παρόμοιες ετικέτες ή τιμές στόχου. Με άλλα λόγια, «κοντινά» σημεία στο χώρο πιθανότατα ανήκουν στην ίδια κατηγορία ή έχουν παρόμοια τιμή εξόδου. Ο αλγόριθμος στηρίζεται στην έννοια της εγγύτητας: για να προβλέψει την κατηγορία ή την τιμή ενός νέου σημείου, εξετάζει τις ετικέτες των *k* πιο κοντινών του σημείων στο σύνολο εκπαίδευσης και αποφασίζει με βάση αυτές.

Ιστορικά, οι πρώτες ιδέες για τον *k*-NN ανάγονται στις εργασίες των Fix & Hodges (1951), ενώ η θεωρητική θεμελίωση δόθηκε από τους Cover & Hart (1967). Ένα σημαντικό θεωρητικό αποτέλεσμα είναι ότι καθώς το πλήθος των δειγμάτων τείνει στο άπειρο, ο ταξινομητής 1-NN (με  $k=1$ ) έχει σφάλμα το πολύ διπλάσιο του βέλτιστου (Bayes) ταξινομητή. Γενικότερα, για κατάλληλη επιλογή του *k*, ο αλγόριθμος *k*-NN είναι συνεπής ταξινομητής, δηλαδή το ποσοστό σφάλματος του πλησιάζει ασυμπτωτικά το βέλτιστο Bayes σφάλμα καθώς το μέγεθος των δεδομένων αυξάνει.

Στην εποχή των Big Data και των κοινωνικών δικτύων (*social media*), ο *k*-NN συνεχίζει να βρίσκει εφαρμογές παρά την απλότητά του. Χρησιμοποιείται ως βασική μέθοδος αναφοράς σε πολλά προβλήματα, καθώς και σε πρακτικά συστήματα όπως απλά συστήματα συστάσεων και αναγνώριση προτύπων. Στα μέσα κοινωνικής δικτύωσης, ο *k*-NN έχει αξιοποιηθεί για ταξινόμηση χρηστών, ανάλυση συναισθήματος κειμένων, καθώς και για συστάσεις περιεχομένου, λόγω της ικανότητάς του να εκμεταλλεύεται την ομοιότητα μεταξύ χρηστών ή αντικειμένων. Στη συνέχεια θα παρουσιαστεί μια πλήρης και λεπτομερής ανάλυση του αλγορίθμου, εστιάζοντας στη θεωρητική/μαθηματική του βάση, καθώς και σε πρακτικές πτυχές και εφαρμογές του σε δεδομένα Social Media, με παράθεση σχετικής βιβλιογραφίας.

### 4.4.1 Αρχιτεκτονική και Βασική Λογική του *k*-NN

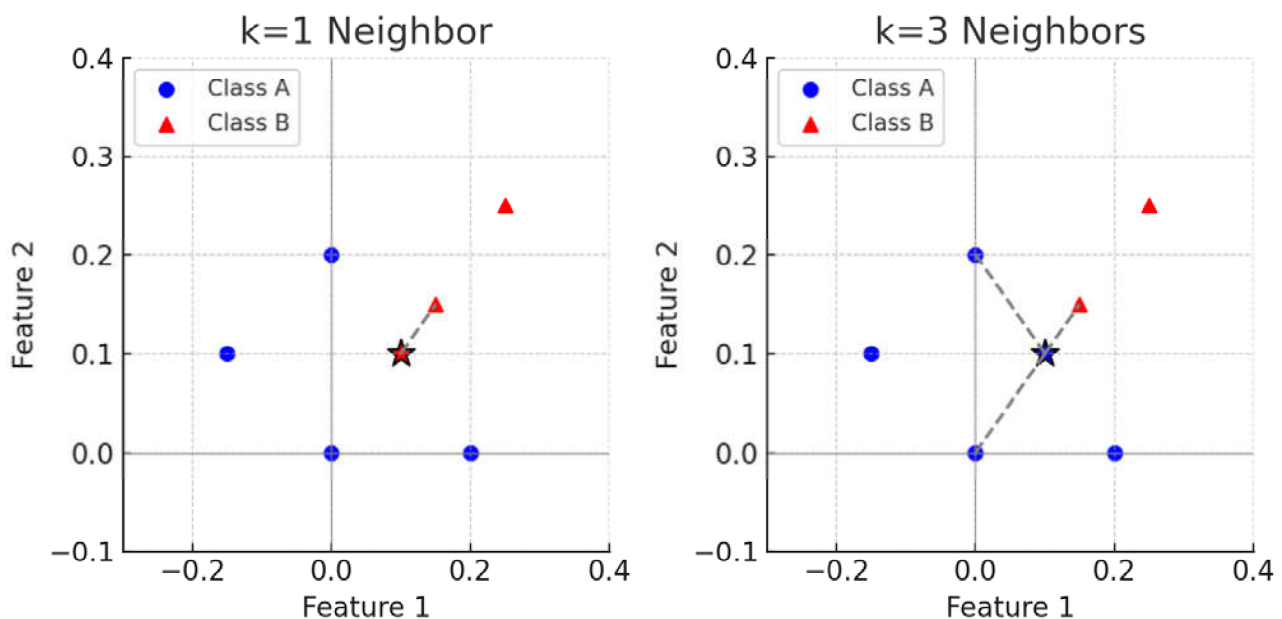
Ο αλγόριθμος *k*-NN ακολουθεί μια ιδιαίτερα απλή αρχιτεκτονική. Στάδιο εκπαίδευσης: δεν υπάρχει ρητή διαδικασία εκμάθησης μοντέλου· ο αλγόριθμος απλώς αποθηκεύει όλα τα πρότυπα εκπαίδευσης (feature vectors) μαζί με τις ετικέτες τους. Για τον λόγο αυτό

χαρκτηρίζεται ως instance-based ή memory-based μέθοδος, σε αντίθεση με αλγορίθμους “ενεργούς” μάθησης που μαθαίνουν ένα συνοπτικό μοντέλο.

Στάδιο πρόβλεψης (ταξινόμηση): Δεδομένου ενός νέου, μη επισημασμένου δείγματος (διάνυσμα χαρακτηριστικών) ως είσοδο, ο k-NN υπολογίζει την απόσταση μεταξύ του δείγματος αυτού και κάθε δείγματος του συνόλου εκπαίδευσης. Στη συνέχεια εντοπίζει τα k πλησιέστερα γειτονικά σημεία και εκχωρεί στο νέο δείγμα την ετικέτα που αποτελεί την πλειοψηφία μεταξύ των k αυτών γειτόνων. Δηλαδή, η προβλεπόμενη κλάση  $\hat{y}$  προκύπτει με πλειοψηφική ψηφοφορία (majority vote) των k κοντινότερων σημείων:

$$\hat{y} = \text{mode}\{y_i : x_i \text{ είναι μεταξύ των } k \text{ πλησιέστερων γειτόνων του } x\}$$

Σε περίπτωση ισοψηφίας (π.χ. δύο τάξεις εμφανίζονται εξίσου στους k γείτονες), συχνά επιλέγεται είτε τυχαία μία από τις ισοψηφήσασες τάξεις, είτε χρησιμοποιείται περιττός αριθμός k για να αποφεύγονται οι ισοψηφίες. Το σκεπτικό είναι ότι τα περισσότερα από τα k πλησιέστερα παραδείγματα, όντας “γειτονικά” στο χώρο χαρακτηριστικών, πιθανώς μοιράζονται την ίδια ετικέτα με το ζητούμενο σημείο. Στο παρακάτω διάγραμμα φαίνεται ένα παράδειγμα: για  $k=1$  ο νέος σημειωμένος με ★ ταξινομείται ως κόκκινος (Class B) βάσει του εγγύτερου γείτονα, ενώ για  $k=3$  ταξινομείται ως μπλε (Class A) λόγω πλειοψηφίας δύο προς ένα.



**Παράδειγμα ταξινόμησης με k-NN:** Το νέο σημείο (★) εξετάζει τους k πλησιέστερους γείτονες. **Αριστερά:** με  $k=1$ , ο πλησιέστερος γείτονας είναι κόκκινος και το νέο σημείο ταξινομείται ως κόκκινο (Class B). **Δεξιά:** με  $k=3$ , οι τρεις πλησιέστεροι γείτονες περιλαμβάνουν 2 μπλε και 1 κόκκινο, οπότε το σημείο ταξινομείται ως μπλε (Class A). Οι διακεκομμένες γραμμές δείχνουν τους γείτονες που λαμβάνονται υπόψη.

**Παλινδρόμηση:** Ο ίδιος αλγόριθμος προσαρμόζεται εύκολα για πρόβλεψη συνεχών τιμών (regression) αντί για διακριτές κλάσεις. Στην περίπτωση παλινδρόμησης, η πρόβλεψη για ένα

νέο δείγμα υπολογίζεται συνήθως ως ο μέσος όρος (ή γενικότερα ένα σταθμισμένο μέσο) των τιμών-στόχων των  $k$  πλησιέστερων γειτόνων. Για παράδειγμα, αν  $y_i$  είναι η πραγματική αριθμητική τιμή του  $i$ -οστού γείτονα, η πρόβλεψη δίνεται από:

$$\hat{y} = \frac{1}{k} \sum_{i \in N_k(x)} y_i$$

όπου  $N_k(x)$  είναι το σύνολο των  $k$  πλησιέστερων γειτονικών δειγμάτων του  $x$ . Έτσι, το  $k$ -NN υλοποιεί μια μορφή τοπικής παλινδρόμησης λαμβάνοντας υπόψη μόνο τα κοντινά ιστορικά δεδομένα. Είναι αξιοσημείωτο ότι ο ίδιος αλγόριθμος δύναται να εφαρμοστεί σε ποικιλία προβλημάτων χωρίς αλλαγές, αρκεί να οριστεί κατάλληλα ένα μέτρο απόστασης/ομοιότητας και να υπάρχει διαθέσιμο ένα επαρκές σύνολο παραδειγμάτων.

Βασικό στοιχείο του  $k$ -NN είναι η έννοια της απόστασης ή ομοιότητας μεταξύ δειγμάτων. Ο αλγόριθμος τυπικά λειτουργεί σε έναν  $d$ -διάστατο χώρο χαρακτηριστικών  $\mathbb{R}^d$  όπου κάθε δείγμα είναι ένα διάνυσμα  $x=(x_1, x_2, \dots, x_d)$ . Η εγγύτητα μεταξύ δύο δειγμάτων  $x$  και  $z$  μετρείται μέσω μιας μετρικής απόστασης  $d(x, z)$ .

Η πιο κοινή επιλογή είναι η Ευκλείδεια απόσταση, που ορίζεται ως:

$$d_E(x, z) = \sqrt{\sum_{i=1}^d (x_i - z_i)^2}$$

Πρόκειται για την απόσταση ευθείας γραμμής μεταξύ των σημείων  $x$  και  $z$  στο χώρο (η προέκταση του Πυθαγόρειου θεωρήματος σε περισσότερες διαστάσεις). Μια πιο γενική οικογένεια αποστάσεων είναι οι αποστάσεις Minkowski, οι οποίες ορίζονται ως:

$$d_{(p)}(x, z) = \left( \sum_{i=1}^d |x_i - z_i|^p \right)^{1/p}$$

Ανάλογα με την τιμή του εκθέτη  $p$ , προκύπτουν διαφορετικές γνωστές μετρικές ως ειδικές περιπτώσεις: όταν  $p=2$  έχουμε την Ευκλείδεια απόσταση, ενώ για  $p=1$  προκύπτει η απόσταση Manhattan (ή απόσταση πόλης), που ισούται με το άθροισμα των απόλυτων διαφορών:

$$d_{\text{Manhattan}}(x, z) = \sum_{i=1}^d |x_i - z_i|$$

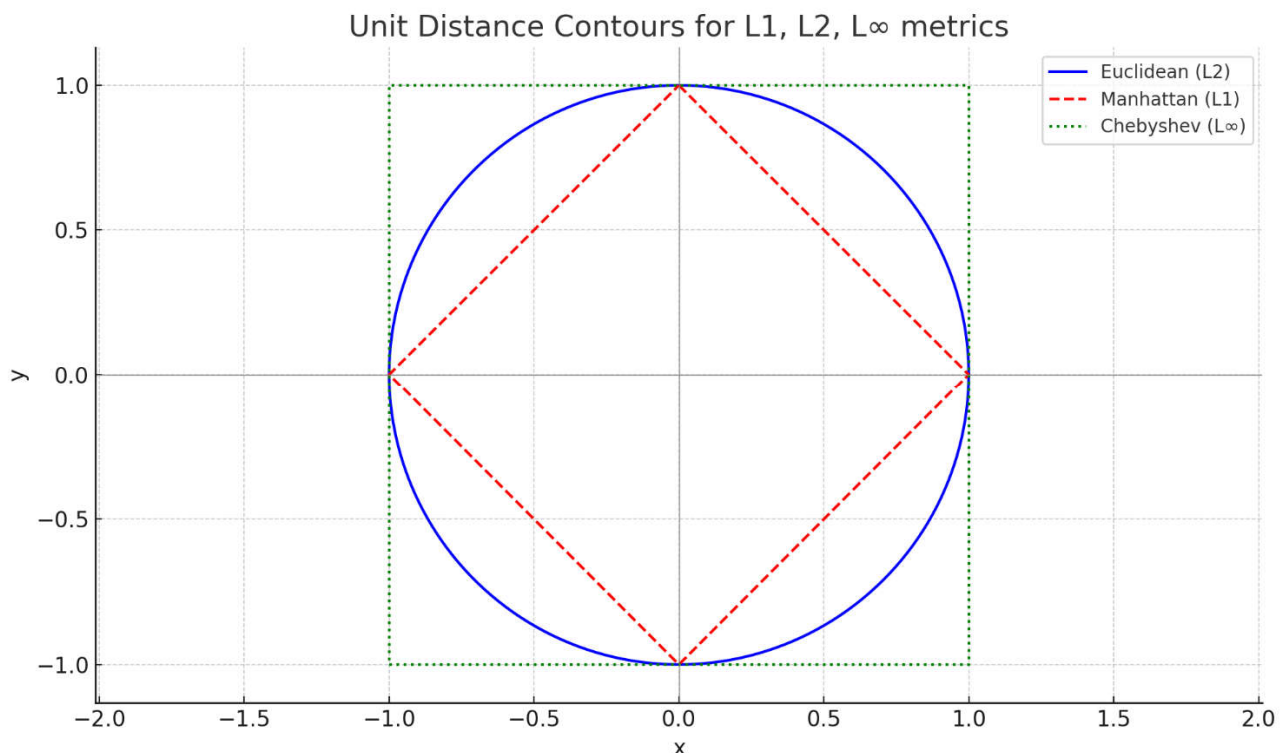
όπου:

- $x=(x_1,x_2,\dots,x_d)$  και  $z=(z_1,z_2,\dots,z_d)$  είναι δύο διανύσματα στον  $d$ -διάστατο χώρο χαρακτηριστικών.
- $|x_i-z_i|$  αντιστοιχεί στην απόλυτη τιμή της διαφοράς μεταξύ των αντίστοιχων στοιχείων των δύο διανυσμάτων.

Η ονομασία “Manhattan” (πόλη-μπλοκ) προκύπτει από την εικόνα της απόστασης σε ένα ορθογώνιο δίκτυο δρόμων – η απόσταση μεταξύ δύο σημείων είναι το άθροισμα των οριζόντιων και κάθετων αποστάσεών τους, όπως θα διένυε ένα ταξί σε μια πόλη με τετράγωνα οικοδομικά μπλοκ. Ένα άλλο ιδιαίτερο μέτρο είναι η απόσταση Chebyshev (ή  $L_\infty$ ), που αντιστοιχεί στην περίπτωση  $p \rightarrow \infty$  της απόστασης Minkowski. Σε αυτήν, η απόσταση δίνεται από το μέγιστο των απόλυτων διαφορών:

$$d^\infty(x,z)=\max_i |x_i-z_i|$$

Στο ακόλουθο σχήμα συγκρίνονται γεωμετρικά οι ισαπέχουσες καμπύλες (μονοσύνολα αποστάσεων) για τις αποστάσεις  $L_1$  (Manhattan),  $L_2$  (Ευκλείδεια) και  $L_\infty$  (Chebyshev). Παρατηρούμε ότι η  $L_2$  παράγει τον κλασικό κύκλο, η  $L_1$  έναν ρόμβο (διαμάντι), ενώ η  $L_\infty$  έναν τετράγωνο σχηματισμό.



Γεωμετρική απεικόνιση ισαπέχουσων καμπυλών: Οι μονοσύνοδες αποστάσεων (απόσταση = 1) για διαφορετικές μετρικές σε 2 διαστάσεις. Με μπλε συνεχές περιγράφεται η απόσταση Ευκλείδειας  $L_2$ , με κόκκινο διακεκομμένο η Manhattan  $L_1$  και με πράσινο σημειακό η Chebyshev  $L_\infty$ . Η επιλογή της μετρικής επηρεάζει τον τρόπο μέτρησης “γειτνιάσης” στον χώρο χαρακτηριστικών.

Άλλες μετρικές απόστασης μπορεί να χρησιμοποιηθούν ανάλογα με τη φύση των δεδομένων. Για παράδειγμα, σε δυαδικά ή κατηγορικά χαρακτηριστικά συνηθίζεται η απόσταση Hamming, που μετρά τον αριθμό θέσεων στις οποίες διαφέρουν δύο δυαδικά διανύσματα (ισοδύναμα, για δυαδικούς κώδικες είναι ο αριθμός bit που διαφέρουν). Για σύγκριση κειμένων ή συνόλων μπορεί να χρησιμοποιηθεί η απόσταση Jaccard, ενώ σε προβλήματα σύστασης και προτιμήσεων αξιοποιούνται συχνά συντελεστές συσχέτισης (όπως η συσχέτιση Pearson) ή συνελίξεις (π.χ. συνημίτονο γωνίας) ως μέτρα ομοιότητας. Η επιλογή του κατάλληλου μέτρου απόστασης είναι κρίσιμη: ο k-NN υποθέτει ότι τα σημεία που είναι κοντά ως προς αυτήν τη μετρική έχουν παρόμοια συμπεριφορά. Αν η μετρική δεν αντανακλά την πραγματική συσχέτιση με την έξοδο (π.χ. άσχετα χαρακτηριστικά), η επίδοση του αλγορίθμου θα υποβαθμιστεί.

Η παράμετρο  $k$ , δηλαδή ο αριθμός των γειτόνων που εξετάζει ο αλγόριθμος, παίζει καθοριστικό ρόλο στην απόδοση του k-NN. Ένα πολύ μικρό  $k$  κάνει τον ταξινομητή ευαίσθητο στον θόρυβο και στις τυχαίες διακυμάνσεις των δεδομένων (υψηλή διασπορά), οδηγώντας σε υπερ-προσαρμογή (overfitting) – π.χ. για  $k=1$  το μοντέλο απλά αναθέτει σε κάθε σημείο την ετικέτα του πλησιέστερου εκπαιδευτικού δείγματος, γεγονός που μπορεί να συλλάβει και τον θόρυβο στα δεδομένα. Αντίθετα, ένα πολύ μεγάλο  $k$  τον καθιστά πιο αμβλύ: η απόφαση λαμβάνεται από μια μεγάλη γειτονιά, ίσως συμπεριλαμβάνοντας σημεία από άλλες τάξεις, άρα το μοντέλο γίνεται υψηλής μεροληψίας (underfitting) και οι αποφάσεις του πλησιάζουν την απλή πρόβλεψη της πιο συχνής κλάσης στο σύνολο (για  $k = n$  όλες οι αποφάσεις είναι η πλειοψηφούσα κλάση). Συνεπώς, υπάρχει μία *συμβιβαστική επιλογή* (trade-off) μεταξύ μεροληψίας και διασποράς καθώς μεταβάλλεται το  $k$ .

Συνήθως, ο βέλτιστος αριθμός γειτόνων επιλέγεται εμπειρικά μέσω διαδικασίας διασταυρούμενης επικύρωσης (cross-validation) στο σύνολο εκπαίδευσης. Δοκιμάζονται διάφορες τιμές του  $k$  και επιλέγεται εκείνη που μεγιστοποιεί την απόδοση (π.χ. την ακρίβεια ταξινόμησης) σε δεδομένα επικύρωσης. Μερικές *εμπειρικές κατευθυντήριες γραμμές*: (α) χρησιμοποιείται συχνά περιττός αριθμός για  $k$  σε δυαδικά προβλήματα ώστε να αποφεύγονται οι ισοψηφίες, (β) σε δεδομένα με θόρυβο ή ανωμαλίες (outliers), ένα μεγαλύτερο  $k$  τείνει να μετριάσει την επίδραση των ακραίων περιπτώσεων, καθώς η απόφαση λαμβάνει υπόψη περισσότερους γείτονες, (γ) ένα τυπικό εύρος δοκιμής είναι από 1 μέχρι κάποια  $\sqrt{n}$  ή άλλη υπο-γραμμική συνάρτηση του αριθμού εκπαίδευσης  $n$ , ανάλογα με το πρόβλημα. Δεν υπάρχει μια καθολικά καλύτερη τιμή — το βέλτιστο  $k$  εξαρτάται από τα δεδομένα.

Εκτός από την απλή πλειοψηφική ψήφο, υπάρχουν παραλλαγές που επιχειρούν να μετριάσουν την επίδραση ενός ακατάλληλου  $k$ . Μία κοινή τεχνική είναι η σταθμισμένη ψηφοφορία: οι ψήφοι των γειτόνων μπορούν να σταθμιστούν ανάλογα με την απόστασή τους από το νέο σημείο, έτσι ώστε οι κοντινότεροι γείτονες να έχουν μεγαλύτερη συνεισφορά στην απόφαση. Συχνά χρησιμοποιείται συντελεστής βάρους ανάλογος του αντιστρόφου της απόστασης (π.χ. βάρος  $w_i=1/d(x_i,x)$ ), ώστε ένας πολύ κοντινός γείτονας να υπερσχύει έναντι

ενός οριακά μέσα στους  $k$  γείτονες. Αυτό μπορεί να βελτιώσει την απόδοση ειδικά όταν η πυκνότητα των σημείων διαφέρει μεταξύ τάξεων ή υπάρχουν ανισοκατανομή (skew) στις κλάσεις.

Στην πράξη, η επιλογή του  $k$  είναι μέρος της διαδικασίας βελτιστοποίησης του μοντέλου. Για κάθε νέο πρόβλημα, ο πειραματισμός με διάφορες τιμές και η αξιολόγηση μέσω κατάλληλων μετρικών (accuracy, F1, MSE κ.λπ.) είναι ουσιώδους σημασίας για την απόκτηση ενός αποτελεσματικού k-NN μοντέλου.

#### 4.4.2 Κανονικοποίηση Δεδομένων (Feature Scaling)

Η κλίμακα των χαρακτηριστικών επηρεάζει δραματικά τη λειτουργία του k-NN. Δεδομένου ότι η απόσταση υπολογίζεται πάνω στις πρώτες τιμές των χαρακτηριστικών, χαρακτηριστικά με μεγαλύτερες αριθμητικές τιμές ή ευρύτερη διασπορά μπορούν να κυριαρχήσουν στον υπολογισμό της απόστασης σε σχέση με χαρακτηριστικά με μικρότερη κλίμακα. Για παράδειγμα, αν ένα χαρακτηριστικό έχει τιμές από 0 έως 1 και ένα άλλο από 0 έως 1.000, η Ευκλείδεια απόσταση ουσιαστικά θα καθορίζεται από το δεύτερο χαρακτηριστικό, καθώς οι διαφορές σε αυτό θα είναι κατά πολύ μεγαλύτερες. Έτσι, το αποτέλεσμα του k-NN θα προκαθορίζεται από το χαρακτηριστικό μεγάλης κλίμακας, ακόμη κι αν τα μικρότερης κλίμακας χαρακτηριστικά είναι πιο σημαντικά για την ταξινόμηση.

Για τον λόγο αυτό, είναι αναγκαία η κανονικοποίηση ή τυποποίηση (normalization/standardization) των χαρακτηριστικών πριν από την εφαρμογή του k-NN. Με τον όρο κανονικοποίηση συνήθως εννοούμε τον ανασχεδιασμό των δεδομένων ώστε όλα τα χαρακτηριστικά να κυμαίνονται σε παρόμοιο εύρος, π.χ. μετασχηματίζοντας κάθε χαρακτηριστικό ώστε οι τιμές του να βρίσκονται στο  $[0,1]$  ή  $[-1,1]$ . Εναλλακτικά, η τυποποίηση (standardization) μετασχηματίζει τα χαρακτηριστικά ώστε να έχουν μηδενική μέση τιμή και μοναδιαία τυπική απόκλιση. Και οι δύο προσεγγίσεις αποσκοπούν στο να γίνουν οι διάφορες διαστάσεις συγκρίσιμες. Όπως χαρακτηριστικά αναφέρεται: «η κανονικοποίηση των μεταβλητών τις καθιστά συγκρίσιμες – ο υπολογισμός αποστάσεων σε μη κανονικοποιημένα δεδομένα αντιστοιχεί ουσιαστικά σε απονομή υψηλότερων βαρών στα χαρακτηριστικά με ευρύτερο εύρος τιμών».

Συνεπώς, σε προβλήματα όπου τα χαρακτηριστικά έχουν ετερογενείς μονάδες μέτρησης ή τάξεις μεγέθους, επιβάλλεται η προ-επεξεργασία των δεδομένων με κατάλληλη κλιμάκωση. Εξαιρέσεις μπορεί να υπάρξουν όταν τα ίδια τα δεδομένα περιλαμβάνουν έννοια βαρύτητας μέσω του μεγέθους (π.χ. σε ορισμένα προβλήματα βαρύνουμε σκόπιμα ένα χαρακτηριστικό). Όμως γενικά, η επιτυχία του k-NN προϋποθέτει *καλά ορισμένο χώρο απόστασης*, πράγμα που σημαίνει τόσο σωστή επιλογή μετρικής όσο και σωστή προεπεξεργασία των χαρακτηριστικών (κλιμάκωση, αντιμετώπιση κατηγορηματικών χαρακτηριστικών μέσω κατάλληλων κωδικοποιήσεων κ.λπ.).

Ένα κύριο μειονέκτημα του k-NN είναι το υπολογιστικό του κόστος κατά την φάση πρόβλεψης. Για να ταξινομηθεί ένα νέο δείγμα, ο αλγόριθμος πρέπει να υπολογίσει αποστάσεις μεταξύ του δείγματος και όλων των σημείων του συνόλου εκπαίδευσης και στη συνέχεια να εντοπίσει τους  $k$  πλησιέστερους. Αν το πλήθος των εκπαιδευτικών δειγμάτων είναι  $n$  και το πλήθος των χαρακτηριστικών  $d$ , η χρονική πολυπλοκότητα μιας απλής υλοποίησης είναι  $O(n \cdot d)$  ανά πρόβλεψη (για κάθε νέο σημείο). Συνεπώς, αν θέλουμε να ταξινομήσουμε  $m$  νέα σημεία, η πολυπλοκότητα είναι  $O(m \cdot n \cdot d)$ .

Αυτό σημαίνει ότι ο k-NN είναι σε γενικές γραμμές αργός για μεγάλα σύνολα δεδομένων, σε αντίθεση με το μηδενικό κόστος εκπαίδευσης. Όπως αναφέρει η IBM: «καθώς το σύνολο δεδομένων μεγαλώνει, ο k-NN γίνεται εξαιρετικά αναποτελεσματικός, θέτοντας σε κίνδυνο τη συνολική απόδοση». Για παράδειγμα, σε μια βάση με εκατομμύρια δείγματα, η κάθε αναζήτηση  $k$  γειτόνων απαιτεί δυνητικά εκατομμύρια υπολογισμούς αποστάσεων. Επιπλέον, η πολυπλοκότητα χώρου είναι επίσης μεγάλη, καθώς πρέπει να διατηρούνται όλοι οι εκπαιδευτικοί δεδομένοι στη μνήμη.

Υπάρχουν διάφορες προσεγγίσεις για την επιτάχυνση των αναζητήσεων εγγύτητας. Μια κατηγορία λύσεων είναι η χρήση δομών ευρετηρίασης (indexing structures) και αλγορίθμων εγγύς γειτόνων για αποφυγή του εξαντλητικού ελέγχου κάθε σημείου. Παραδείγματα τέτοιων δομών είναι τα δέντρα kd (kd-trees), οι μπάλες δέντρων (ball trees) και οι δέντρα κάλυψης (cover trees), τα οποία οργανώνουν τα δεδομένα με τρόπο ώστε να απορρίπτεται μεγάλο μέρος του χώρου κατά την αναζήτηση γειτόνων. Ωστόσο, η αποτελεσματικότητα αυτών μειώνεται σε υψηλές διαστάσεις (curse of dimensionality), όπου τα δεδομένα αραιώνονται και οι αποστάσεις μεταξύ σημείων τείνουν να εξισώνονται.

Μια άλλη προσέγγιση είναι η χρήση προσεγγιστικών μεθόδων (Approximate Nearest Neighbors - ANN). Αντί να βρίσκουμε ακριβώς τους  $k$  πλησιέστερους, μπορούμε να βρίσκουμε μια προσέγγιση αυτών με πολύ μικρότερο κόστος, επιτρέποντας έναν μικρό βαθμό σφάλματος στην εγγύτητα. Τεχνικές όπως οι Locality-Sensitive Hashing (LSH) και Hierarchical Navigable Small World (HNSW) γράφουν τις υψηλοδιάστατες αναπαραστάσεις σε δομές που επιτρέπουν αναζήτηση σε υπογραμμικό χρόνο. Για παράδειγμα, με μια κατάλληλη ANN δομή, η αναζήτηση μπορεί να μειωθεί από  $O(n)$  σε περίπου  $O(\log n)$  κατά μέσο όρο, με μια μικρή απώλεια στην ακρίβεια ανάκτησης γειτόνων. Αυτό είναι κρίσιμο σε εφαρμογές πραγματικού χρόνου, όπου η ταχύτητα είναι προτεραιότητα.

Τέλος, αξίζει να σημειωθεί ότι υπάρχουν και τεχνικές μείωσης δεδομένων (data condensation) ειδικά σχεδιασμένες για τον k-NN. Αυτές αφαιρούν ή συνοψίζουν σημεία εκπαίδευσης χωρίς σημαντική απώλεια στην απόδοση ταξινόμησης. Για παράδειγμα, ο αλγόριθμος Condensed Nearest Neighbor (CNN) του Hart (1968) κρατά ένα μικρό υποσύνολο των σημείων ως πρωτότυπα που αρκούν για να ταξινομηθεί σωστά το σύνολο εκπαίδευσης. Με τέτοιες μεθόδους μπορεί να μειωθεί δραστικά το πλήθος των σημείων αναφοράς και έτσι το κόστος ταξινόμησης, εις βάρος όμως πρόσθετου κόστους προεπεξεργασίας.

Συμπερασματικά, ο k-NN παρέχει μηδενικό κόστος εκπαίδευσης αλλά έχει υψηλό κόστος πρόβλεψης. Σε προβλήματα όπου το πλήθος των προβλέψεων είναι τεράστιο ή απαιτείται απόκριση σε πραγματικό χρόνο, συχνά ο k-NN θεωρείται μη πρακτικός εκτός αν συνδυαστεί με τις παραπάνω τεχνικές επιτάχυνσης ή περιοριστεί σε μικρότερα σύνολα.

Ο k-NN παρουσιάζει αρκετά πλεονεκτήματα που εξηγούν τη συνεχιζόμενη χρήση του:

- **Απλότητα και Ευκολία Κατανόησης:** Είναι ένας από τους πιο κατανοητούς αλγόριθμους – η λογική του (γειτονικά σημεία -> παρόμοιες αποφάσεις) είναι διαισθητική και η υλοποίησή του απλή. Δεν απαιτεί περίπλοκη εκπαίδευση ή ρύθμιση παραμέτρων.
- **Καθολικότητα / Ευελιξία:** Μπορεί να εφαρμοστεί σε προβλήματα ταξινόμησης και παλινδρόμησης, υποστηρίζοντας πολλαπλές κλάσεις χωρίς προσαρμογές, αλλά και διαφορετικούς τύπους δεδομένων και μετρικών απόστασης. Έτσι έχει χρησιμοποιηθεί σε πεδία από αναγνώριση προτύπων μέχρι συστήματα σύστασης.
- **Χωρίς ανάγκη εκπαίδευσης (Lazy):** Δεν απαιτείται χρόνος εκπαίδευσης για προσαρμογή παραμέτρων. Ο χρόνος προετοιμασίας δαπανάται κυρίως στην επιλογή χαρακτηριστικών και τον προκαταρκτικό μετασχηματισμό δεδομένων, όχι σε εκπαίδευση μοντέλου.
- **Προσαρμοστικότητα σε νέα δεδομένα:** Επειδή δεν υπάρχει “μοντέλο” με την παραδοσιακή έννοια, νέα δεδομένα μπορούν να προστεθούν εύκολα στη βάση γνώσης χωρίς ανάγκη επανεκπαίδευσης από την αρχή. Αυτό είναι χρήσιμο σε δυναμικά περιβάλλοντα όπου έρχονται συνεχώς νέες πληροφορίες.
- **Μη γραμμικά σύνορα αποφάσεων:** Ο k-NN μπορεί να προσεγγίσει πολύπλοκα σύνορα απόφασης καθώς είναι μη παραμετρικός. Με αρκετά δεδομένα, θεωρητικά μπορεί να προσεγγίσει οποιαδήποτε συνάρτηση, πράγμα που τον καθιστά ισχυρό στον περιορισμό του άπειρου δείγματος (είναι ασυμπτωτικά βέλτιστος όπως αναφέρθηκε).

Παρά τα πλεονεκτήματα, ο k-NN έχει και σημαντικές αδυναμίες:

- **Υπολογιστική επιβάρυνση:** Όπως αναλύθηκε, απαιτείται μεγάλος υπολογιστικός φόρτος για την πρόβλεψη όταν το σύνολο δεδομένων είναι μεγάλο, καθώς ο αλγόριθμος είναι ουσιαστικά *εξαντλητική αναζήτηση* στον χώρο. Αυτό μπορεί να καταστήσει τον k-NN αργό ή ανεφάρμοστο σε εφαρμογές με αυστηρά χρονικά περιθώρια ή τεράστια δεδομένα, αν δεν χρησιμοποιηθούν πρόσθετες τεχνικές.
- **Ευαισθησία σε άχρηστα χαρακτηριστικά:** Εάν το διάνυσμα χαρακτηριστικών περιλαμβάνει πολλές διαστάσεις που δεν σχετίζονται με την έξοδο, αυτές μπορεί να εισάγουν θόρυβο στον υπολογισμό αποστάσεων και να παραπλανήσουν τον αλγόριθμο. Απαιτείται συχνά προσεκτική επιλογή χαρακτηριστικών ή εκ των προτέρων μείωση διαστατικότητας (π.χ. PCA) για καλή απόδοση.
- **Ευαισθησία στη διαφορετική κλίμακα:** Όπως συζητήθηκε, αν τα χαρακτηριστικά δεν κλιμακωθούν κατάλληλα, ορισμένα θα κυριαρχήσουν στην απόφαση λόγω

μεγαλύτερου εύρους τιμών. Ο k-NN απαιτεί κανονικοποιημένα δεδομένα· διαφορετικά, το αποτέλεσμα μπορεί να είναι παραπλανητικό.

- **Ευαισθησία σε θόρυβο και ανωμαλίες:** Ένα μεμονωμένο θορυβώδες δείγμα ή outlier στη βάση εκπαίδευσης μπορεί να επηρεάσει δυσανάλογα τις αποφάσεις για κοντινά σημεία (ειδικά αν  $k$  είναι μικρό). Ο k-NN δεν έχει μηχανισμό από μόνος του να παραγνωρίζει τα outliers, σε αντίθεση π.χ. με τα δέντρα απόφασης που μπορούν να περιορίσουν την επίδρασή τους.
- **Απαίτηση μνήμης:** Πρέπει να διατηρηθούν όλα τα δείγματα εκπαίδευσης, κάτι που σε μεγάλα δεδομένα συνεπάγεται υψηλές απαιτήσεις μνήμης αποθήκευσης. Αυτό τον καθιστά λιγότερο πρακτικό σε περιβάλλοντα με μνήμη περιορισμένη.
- **Curse of Dimensionality:** Σε πολύ υψηλές διαστάσεις, η έννοια του “κοντινού” και “μακρινού” γίνεται θολή – όλα τα σημεία τείνουν να απέχουν σχεδόν εξίσου (οι αποστάσεις συγκλίνουν). Αυτό μειώνει την αποτελεσματικότητα του k-NN, καθώς η αναζήτηση εγγύτερων γειτόνων γίνεται λιγότερο αξιόπιστη. Επιπλέον, οι δομές ευρετηρίασης δεν λειτουργούν καλά όταν η διάσταση είναι μεγάλη. Συνεπώς, ο k-NN κλιμακώνεται άσχημα σε προβλήματα με πάρα πολλές διαστάσεις εκτός αν μειωθούν.
- **Πρόβλημα με ανισοκατανομή κλάσεων:** Εάν μια κλάση είναι πολύ πιο συχνή στο εκπαιδευτικό σύνολο (class imbalance), τότε οι γείτονες ενός νέου σημείου θα τείνουν να ανήκουν δυσανάλογα συχνά στη πολυπληθέστερη κλάση, απλώς λόγω πληθυσμιακών αναλογιών. Ο βασικός k-NN με πλειοψηφική ψήφο θα προκαταιληφθεί προς την πολυπληθέστερη κλάση. Η χρήση σταθμισμένης απόστασης ή τεχνικών oversampling/undersampling μπορεί να μετριάσει εν μέρει το ζήτημα.

Συνολικά, ο k-NN είναι ελκυστικός για την απλότητά του αλλά απαιτεί προσοχή στην προεπεξεργασία, επιλογή χαρακτηριστικών και αντιμετώπιση ζητημάτων κλίμακας για να αποδώσει βέλτιστα.

Με τα χρόνια έχουν προταθεί διάφορες παραλλαγές του αλγορίθμου k-NN που βελτιώνουν πτυχές του ή τον προσαρμόζουν σε ειδικές περιπτώσεις. Παρακάτω αναφέρονται μερικές σημαντικές:

- **Μεγάλο Περιθώριο k-πλησιέστερων Γειτόνων (LMNN):** Πρόκειται για αλγόριθμο εκμάθησης μετρικής απόστασης που πρότειναν οι Weinberger et al. (2006) με στόχο να βελτιστοποιηθεί η απόδοση του k-NN. Η βασική ιδέα του Large Margin Nearest Neighbor (LMNN) είναι να μάθει μέσω εποπτευόμενης μάθησης μια μετασχηματισμένη χωρική απόσταση (συνήθως μια Μαχαλανοβική απόσταση) έτσι ώστε: (α) κάθε δείγμα να έχει τους  $k$  πλησιέστερους γείτονές του να ανήκουν στην ίδια κλάση (target neighbors) και (β) δείγματα διαφορετικών κλάσεων να βρίσκονται χωρισμένα από έναν μεγάλο περιθώριο (margin). Αυτό διατυπώνεται ως ένα πρόβλημα βελτιστοποίησης (η ελαχιστοποίηση ενός κόστους που ενθαρρύνει μικρές αποστάσεις σε ομοταγή σημεία και μεγάλες αποστάσεις σε ετερότατα σημεία) και μπορεί να λυθεί με μεθόδους ημι-οριστικού προγραμματισμού. Το αποτέλεσμα είναι

μια προσαρμοσμένη μετρική όπου ο  $k$ -NN δίνει πολύ καλύτερη ακρίβεια. Σε πειράματα, η LMNN έφερε σημαντικές βελτιώσεις στην απόδοση του  $k$ -NN – π.χ. σφάλμα μόλις 1,3% στο σύνολο MNIST, συγκρίσιμο με πιο πολύπλοκους ταξινομητές. Ουσιαστικά, η LMNN επιχειρεί να αξιοποιήσει τις ιδέες των SVM (μεγάλο περιθώριο) στο πλαίσιο των γειτονικών σημείων, μαθαίνοντας τον βέλτιστο “χώρο” όπου η απλή ψηφοφορία γειτόνων δουλεύει άριστα. Σημειώνεται ότι παρόμοιες προσεγγίσεις metric learning έχουν αναπτυχθεί, όπως η Neighbourhood Components Analysis (NCA) και άλλες, με στόχο όλες να βρουν μετασχηματισμούς των δεδομένων που μειώνουν την διάσταση ή αναδιατάσσουν τα σημεία ώστε οι σωστοί γείτονες να έρχονται κοντά. Στις μέρες μας, η φιλοσοφία αυτή ενσωματώνεται και σε βαθιά μοντέλα, π.χ. στο FaceNet (εξαγωγή εμβαδών προσώπου με triplet loss ώστε οι εικόνες του ίδιου ατόμου να είναι γειτονικές).

- **Σταθμισμένος  $k$ -NN:** Όπως προαναφέρθηκε, μια απλή αλλά αποτελεσματική παραλλαγή είναι να γίνεται χρήση βάρων κατά την απόφαση. Ο σταθμισμένος  $k$ -NN δίνει διαφορετική βαρύτητα στον κάθε ένα από τους  $k$  γείτονες ανάλογα με την απόστασή του: συνήθως, όσο πιο κοντινός ο γείτονας, τόσο μεγαλύτερο βάρος. Στην ταξινόμηση, αυτό σημαίνει ότι η ψήφος ενός κοντινού σημείου μετρά περισσότερο από ενός οριακά μέσα στους  $k$  γείτονες, ενώ στην παλινδρόμηση αντί για απλό μέσο όρο παίρνουμε σταθμισμένο μέσο. Η τεχνική αυτή μπορεί να αυξήσει την ακρίβεια, ιδίως όταν η πυκνότητα των δεδομένων δεν είναι ομοιόμορφη.
- **Fuzzy  $k$ -NN:** Αντί να παράγει μια *κρίσιμη* απόφαση, ο αλγόριθμος fuzzy  $k$ -NN αποδίδει στο νέο δείγμα βαθμούς συμμετοχής (membership grades) σε κάθε κλάση, με βάση τους γείτονες. Ουσιαστικά αντί για οριστική ετικέτα δίνει πιθανότητες ή βαθμούς πεποίθησης. Αυτό μπορεί να είναι χρήσιμο σε εφαρμογές όπου η αβεβαιότητα του μοντέλου είναι σημαντική.
- **Παραλλαγές για Πολυ-ετικετικές κλάσεις (Multi-label kNN):** Σε ορισμένα προβλήματα, κάθε δείγμα μπορεί να ανήκει ταυτόχρονα σε πολλές κατηγορίες (multi-label classification). Για τέτοια σενάρια έχει προταθεί η επέκταση ML-kNN (Multi-Label kNN). Στον ML-kNN (Zhang & Zhou, 2007), ο αλγόριθμος συλλέγει πληροφορία από τους  $k$  γείτονες για κάθε πιθανή ετικέτα και χρησιμοποιεί μια Bayesian προσέγγιση για να αποφασίσει ποιες ετικέτες θα αποδοθούν στο νέο δείγμα. Παραλλαγές του ML-kNN και βελτιώσεις του (π.χ. Ιεραρχικός kNN για πολυ-επίπεδες ετικέτες, ή παραμετρικός kNN όπου το  $k$  επιλέγεται δυναμικά) έχουν εμφανιστεί στη βιβλιογραφία για να αντιμετωπίσουν ειδικές ανάγκες πολυ-ετικετικού χαρακτήρισμού.
- **Μείωση δεδομένων (Edited/Condensed kNN):** Μια κατηγορία παραλλαγών επικεντρώνεται στη βελτίωση της αποδοτικότητας μέσω μείωσης του μεγέθους του συνόλου αναφοράς. Ο αλγόριθμος Condensed NN που αναφέρθηκε παραπάνω αφαιρεί περιττά δείγματα, ενώ άλλοι όπως ο Edited kNN αφαιρούν θορυβώδη ή λανθασμένα ταξινομημένα δείγματα από το σύνολο για βελτίωση της ακρίβειας. Οι τεχνικές αυτές μπορούν να ιδωθούν ως διαδικασίες φιλτραρίσματος του συνόλου εκπαίδευσης πριν την εφαρμογή του  $k$ -NN.

Εν ολίγοις, ο βασικός αλγόριθμος k-NN έχει αποτελέσει βάση για πολλές επεκτάσεις. Ανάλογα με το πρόβλημα, μπορούν να ενσωματωθούν μαθημένες μετρικές, σταθμίσεις, ή εξειδικευμένοι χειρισμοί δεδομένων ώστε να αντιμετωπιστούν οι αδυναμίες του και να αξιοποιηθούν τα πλεονεκτήματά του.

#### **4.4.3 Εφαρμογές του k-NN σε Δεδομένα Social Media**

Λόγω της απλότητας και της ευελιξίας του, ο k-NN έχει εφαρμοστεί σε πληθώρα προβλημάτων που σχετίζονται με δεδομένα από μέσα κοινωνικής δικτύωσης. Παρακάτω εξετάζονται τρία χαρακτηριστικά παραδείγματα: η ταξινόμηση χρηστών, η ανάλυση συναισθήματος σε αναρτήσεις και η σύσταση περιεχομένου.

##### **Ταξινόμηση Χρηστών σε Social Media**

Η ταξινόμηση χρηστών αναφέρεται στον διαχωρισμό των χρηστών μιας πλατφόρμας σε συγκεκριμένες κατηγορίες βάσει των δεδομένων τους. Για παράδειγμα, οι χρήστες μπορεί να χαρακτηριστούν ως “spammers” vs “μη spammers”, ως διαφορετικά *segments* αγοράς (π.χ. ενδιαφέροντα: τεχνολογία, αθλητισμός, μόδα), ή να τους εκχωρηθούν πολλαπλές ετικέτες ενδιαφερόντων. Ο k-NN έχει χρησιμοποιηθεί ως απλή αλλά αποτελεσματική μέθοδος για τέτοιες ταξινομήσεις.

Σε ένα σενάριο, διαθέτουμε για κάθε χρήστη χαρακτηριστικά όπως αριθμό φίλων, συχνότητα δημοσιεύσεων, περιγραφή προφίλ, συμμετοχή σε ομάδες κλπ, μαζί με κάποιες ετικέτες (labels) που γνωρίζουμε για ένα υποσύνολο χρηστών (π.χ. ποιοι χρήστες είναι bots ή ποιοι ενδιαφέρονται για θεματολογία Α). Ο k-NN μπορεί να ταξινομήσει έναν νέο χρήστη βρίσκοντας τους  $k$  πιο παρόμοιους υπάρχοντες χρήστες (π.χ. σε όρους συμπεριφοράς ή προφίλ) και λαμβάνοντας ψηφοφορία των ετικετών τους. Αυτή η προσέγγιση είναι απλή και δεν απαιτεί εκτενή εκπαίδευση· συχνά χρησιμοποιείται ως baseline σύγκρισης με πιο πολύπλοκα μοντέλα.

Για παράδειγμα, σε πρόσφατη έρευνα των Huang et al. (2023), προτάθηκε ένα μοντέλο πολυ-ετικετικής ταξινόμησης χρηστών με βάση έναν εξελιγμένο k-NN αλγόριθμο (ML-KNN) σε δεδομένα από social media. Οι συγγραφείς ανέλυσαν σύνθετα κοινωνικά δίκτυα και εφάρμοσαν τον ML-KNN για να ταξινομήσουν χρήστες σε πολλαπλές θεματικές ετικέτες ταυτόχρονα, επιτυγχάνοντας καλύτερα αποτελέσματα από υπάρχουσες μεθόδους. Συγκεκριμένα, το μοντέλο τους μπόρεσε να αποδώσει πολλαπλές ετικέτες ενδιαφερόντων σε χρήστες του Twitter, αξιοποιώντας τόσο τα περιεχόμενα των δημοσιεύσεων όσο και τη δομή του δικτύου (κοινότητες). Αυτό το παράδειγμα δείχνει πως ο k-NN (και οι παραλλαγές του) μπορούν να επεκταθούν για απαιτητικές εργασίες στα social media, όπως η ταυτόχρονη ταξινόμηση σε πολλές κατηγορίες (multi-label classification).

Άλλες εφαρμογές ταξινόμησης χρηστών με k-NN περιλαμβάνουν: ανίχνευση κακόβουλων χρηστών ή bots (π.χ. ένας λογαριασμός χαρακτηρίζεται ως ύποπτος αν οι  $k$  πιο όμοιοί του γνωστοί λογαριασμοί είναι bots), ταξινόμηση ανά δημογραφικές ή γεωγραφικές ομάδες και πρόβλεψη ενδιαφερόντων βάσει γειτονικών χρηστών. Σε όλες αυτές, η ιδέα είναι ότι ένας χρήστης περιβάλλεται από “γειτονικούς” χρήστες με παρόμοια χαρακτηριστικά ή συμπεριφορές, οπότε οι ετικέτες των γειτόνων αποτελούν καλή ένδειξη και για τον ίδιο.

### **Ανάλυση Συναισθήματος σε Social Media**

Η ανάλυση συναισθήματος (sentiment analysis) αφορά την αυτοματοποιημένη κατηγοριοποίηση κειμένων (όπως tweets, σχόλια, αναρτήσεις) σε κατηγορίες συναισθήματος, συνήθως *θετικό*, *αρνητικό*, (ή *ουδέτερο*). Ο k-NN έχει αξιοποιηθεί σε αυτό το πρόβλημα ως ένας απλός ταξινομητής κειμένου: πρώτα μετατρέπει τα κείμενα σε διανύσματα χαρακτηριστικών (π.χ. με χρήση tf-idf για λέξεις, ή ενσωματώσεις λέξεων/προτάσεων) και έπειτα, για μια νέα ανάρτηση, βρίσκει τις  $k$  πιο κοντινές (σε όρους ομοιότητας κειμένου) αναρτήσεις με γνωστό συναίσθημα, αναθέτοντάς της την επικρατούσα πολικότητα.

Παρά το ότι πιο σύνθετες μέθοδοι (π.χ. νευρωνικά δίκτυα) είναι σήμερα δημοφιλείς στην ανάλυση συναισθήματος, ο k-NN έχει χρησιμοποιηθεί είτε ως baseline είτε σε συνδυασμό με άλλες τεχνικές. Για παράδειγμα, ο Isnain et al. (2021) εφάρμοσαν k-NN σε σύνολο 1825 tweets στα Ινδονησιακά σχετικά με την τηλεπαιδείωση, αφού πρώτα μετέτρεψαν τα tweets σε διανύσματα TF-IDF. Στα πειράματά τους, δοκιμάστηκε διάφορο  $k$  και βρέθηκε ότι  $k=10$  έδωσε την υψηλότερη ακρίβεια (~84.65% ακρίβεια ταξινόμησης σε θετικό/αρνητικό). Αυτό δείχνει ότι με κατάλληλη προεπεξεργασία κειμένου, ο k-NN μπορεί να είναι ανταγωνιστικός σε τέτοια προβλήματα. Άλλα αποτελέσματα της μελέτης έδειξαν precision ~87% και recall ~86%, γεγονός που υπογραμμίζει ότι ένας τόσο απλός αλγόριθμος κατάφερε αξιόλογη απόδοση στην ανάλυση πραγματικών δεδομένων Twitter.

Σε πρακτικές υλοποιήσεις, η χρήση του k-NN για κείμενα συχνά συνεπάγεται την επιλογή κατάλληλης μετρικής ομοιότητας – αντί της απλής Ευκλείδειας απόστασης σε TF-IDF, μπορεί να χρησιμοποιηθεί κοσινική ομοιότητα (cosine similarity), η οποία είναι ιδιαίτερα δημοφιλής για συγκρίσεις κειμένων. Η κοσινική ομοιότητα ουσιαστικά μετρά την γωνία μεταξύ των διανυσμάτων χαρακτηριστικών (tf-idf), η οποία είναι ανεξάρτητη του απόλυτου μεγέθους τους και λειτουργεί καλά σε υψηλοδιάστατα αραιά δεδομένα όπως είναι τα διανύσματα λέξεων. Ένας k-NN που χρησιμοποιεί την κοσινική ομοιότητα μπορεί να θεωρηθεί “k-NN με βάση την ομοιότητα” παρά την απόσταση, αλλά μαθηματικά ισοδυναμεί με χρήση μιας κατάλληλης μετρικής απόστασης (της συμπληρωματικής ως προς το 1 – cosine distance).

Άλλες εφαρμογές στα social media όπου συναντάται ο k-NN είναι η κατηγοριοποίηση θεματολογίας κειμένων (π.χ. ταξινόμηση ενός tweet ως σχετικό με αθλητισμό, πολιτική, κ.ο.κ.), όπου επίσης λειτουργεί με γείτονες στο χώρο των λέξεων. Σε γενικές γραμμές, ο k-NN είναι μια ανταγωνιστική απλή μέθοδος ταξινόμησης κειμένου, ειδικά όταν το μέγεθος δεδομένων δεν δικαιολογεί εκπαίδευση βαρύτερων μοντέλων ή όταν απαιτείται ένα γρήγορο σύστημα αναφοράς.

### Συστάσεις Περιεχομένου (Content Recommendation)

Τα κοινωνικά δίκτυα και πλατφόρμες συχνά επιδιώκουν να προτείνουν περιεχόμενο στους χρήστες τους – όπως φίλους που ίσως γνωρίζουν, ομάδες για να ενταχθούν, ή δημοσιεύσεις/άρθρα που μπορεί να τους ενδιαφέρουν. Πολλά από αυτά τα προβλήματα μπορούν να ιδωθούν μέσα από τον φακό της εύρεσης γειτονικών ομοιοτήτων και υλοποιούνται με παραλλαγές του αλγορίθμου k-NN.

Μία κλασική προσέγγιση στις συστάσεις είναι το συνεργατικό φιλτράρισμα (collaborative filtering). Υπάρχουν δύο κύριες εκδοχές: το *user-based* και το *item-based* φιλτράρισμα. Στο *user-based*, για να προτείνουμε περιεχόμενο σε έναν χρήστη, βρίσκουμε χρήστες παρόμοιους με αυτόν (γειτονικούς σε προτιμήσεις) και προτείνουμε αντικείμενα που εκείνοι προτιμούν. Στο *item-based*, βρίσκουμε αντικείμενα (π.χ. αναρτήσεις, βίντεο, προϊόντα) παρόμοια με αυτά που ο χρήστης έχει ήδη δείξει ενδιαφέρον και προτείνουμε τα “γειτονικά” τους. Ουσιαστικά και οι δύο προσεγγίσεις είναι μια εφαρμογή του k-NN: στο πρώτο οι k πλησιέστεροι γείτονες είναι άλλοι χρήστες, στο δεύτερο είναι άλλα αντικείμενα/περιεχόμενο.

Για παράδειγμα, το σύστημα προτάσεων φίλων του Facebook (“People You May Know”) μπορεί να χρησιμοποιήσει έναν αλγόριθμο όπου για έναν χρήστη A, βρίσκει τους k πιο “κοντινούς” χρήστες με βάση κοινές συνδέσεις και αλληλεπιδράσεις και προτείνει αυτούς τους ανθρώπους ως πιθανούς φίλους. Αντίστοιχα, μια υπηρεσία όπως το YouTube ή το TikTok μπορεί για ένα βίντεο X που παρακολουθεί ένας χρήστης, να βρει τα k πιο παρόμοια βίντεο (π.χ. σε όρους ετικετών, θεματολογίας ή ακόμα και οπτικού περιεχομένου) και να τα προτείνει δίπλα ως “σχετικά βίντεο”.

Στο χώρο των συστάσεων ταινιών/προϊόντων, έχουν υπάρξει γνωστά συστήματα που βασίζονται σε k-NN. Για παράδειγμα, μια κλασική μέθοδος *item-based* συνεργατικού φιλτραρίσματος (Sarwar et al., 2001) υπολογίζει την ομοιότητα μεταξύ όλων των ζευγών ταινιών με βάση τις αξιολογήσεις των χρηστών. Κατόπιν, για έναν χρήστη, βρίσκει τις πιο όμοιες ταινίες με αυτές που έχει ήδη βαθμολογήσει υψηλά και συστήνει εκείνες που δεν έχει δει. Ουσιαστικά, αυτό είναι k-NN στο χώρο των αντικειμένων χρησιμοποιώντας ως χαρακτηριστικά το “προφίλ αξιολογήσεων” κάθε αντικειμένου.

Η απλότητα του k-NN καθιστά εύκολη την υλοποίηση τέτοιων συστημάτων. Επιπλέον, μπορεί να αξιοποιήσει οποιουδήποτε είδους χαρακτηριστικά για την ομοιότητα: στα social media, μπορεί να είναι δημογραφικά στοιχεία, το κείμενο των δημοσιεύσεων, στατιστικά αλληλεπίδρασης (likes, shares), κ.λπ. Για παράδειγμα, μια μελέτη προτείνει ένα σύστημα συστάσεων συνδυάζοντας τόσο τη δομή φιλίας όσο και το περιεχόμενο: ένας χρήστης θεωρείται γείτονας αν έχει παρόμοιο κοινωνικό “εγδίκτυο” και δημοσιεύει παρόμοιο περιεχόμενο. Αυτό οδηγεί σε πιο προσωποποιημένες προτάσεις.

Επειδή όμως ο αριθμός των χρηστών και του περιεχομένου στα social media είναι τεράστιος, οι υλοποιήσεις των συστημάτων σύστασης συνήθως ενσωματώνουν

βελτιστοποιήσεις του k-NN ως προς την ταχύτητα, χρησιμοποιώντας τεχνικές προσεγγιστικής γειτνίασης που αναφέρθηκαν. Για παράδειγμα, η αναζήτηση σε ενσωματωμένους χώρους (vector search) είναι μια τάση όπου πρώτα τα αντικείμενα (π.χ. αναρτήσεις) μετατρέπονται σε διανύσματα σε έναν πολυδιάστατο χώρο μέσω μοντέλων (π.χ. μετασχηματιστές για μετατροπή κειμένου σε embedding) και μετά χρησιμοποιείται μια δομή όπως HNSW ή Annoy για να βρίσκονται γρήγορα τα κοντινότερα διανύσματα σε αυτόν τον χώρο. Ουσιαστικά πρόκειται για k-NN αναζητήσεις σε υψηλή κλίμακα με ανεικτική προσεγγιστική ακρίβεια. Αυτές οι τεχνικές έχουν καταστήσει δυνατή την εφαρμογή της λογικής των k πλησιέστερων γειτόνων ακόμα και σε προβλήματα σύστασης με δισεκατομμύρια αντικείμενα.

Συμπερασματικά, στις εφαρμογές social media ο k-NN παρέχει έναν απλό αλλά ισχυρό μηχανισμό εύρεσης ομοιοτήτων, που μπορεί να χρησιμοποιηθεί ευρέως: από την ανάλυση συμπεριφοράς χρηστών και την κατανόηση κειμένου, μέχρι την προσωποποίηση του περιεχομένου που βλέπει ο χρήστης. Παρότι συχνά αντικαθίσταται από πιο σύνθετα μοντέλα μάθησης, διατηρείται ως σημείο αναφοράς και ως τμήμα υβριδικών λύσεων (π.χ. φίλτρο για υποψήφιες προτάσεις πριν την τελική κατάταξη).

Ο αλγόριθμος k-Nearest Neighbors αποτελεί θεμελιώδη μέθοδο μηχανικής μάθησης, συνδυάζοντας απλότητα στην υλοποίηση, διαισθητικότητα και ευελιξία στην εφαρμογή. Στην παρούσα ανάλυση εξετάστηκε η θεωρητική και μαθηματική βάση του k-NN: από τη βασική του λειτουργία ως lazy classifier που βασίζεται σε αποστάσεις, μέχρι τις διάφορες μετρικές απόστασης (Ευκλείδεια, Manhattan, Minkowski κ.ά.), τη σημασία της σωστής επιλογής του υπερπαραμέτρου  $k$  και της κανονικοποίησης των χαρακτηριστικών, καθώς και την πολυπλοκότητα των υπολογισμών του. Επισημάνθηκαν τα πλεονεκτήματα του αλγορίθμου, όπως η μηδενική εκπαίδευση και η προσαρμοστικότητα, αλλά και τα μειονεκτήματα όπως το υψηλό κόστος πρόβλεψης, η ευαισθησία σε θόρυβο και το “curse of dimensionality”.

Επιπλέον, παρουσιάστηκαν σημαντικές επεκτάσεις του k-NN, με κυριότερη το LMNN που μαθαίνει μια κατάλληλη μετασχηματισμένη μετρική για βελτίωση των επιδόσεων, καθώς και άλλες παραλλαγές (σταθμισμένες ψήφοι, multi-label, fuzzy, condensed datasets) που προσαρμόζουν τον αλγόριθμο σε διάφορες απαιτήσεις. Τέλος, συζητήθηκαν συγκεκριμένες εφαρμογές στα Social Media: είδαμε πώς ο k-NN μπορεί να χρησιμοποιηθεί για να ταξινομήσει χρήστες κοινωνικών δικτύων σε κατηγορίες, να αναλύσει το συναίσθημα σε δημοσιεύσεις χρηστών και να συστήσει περιεχόμενο, αξιοποιώντας την εγγύτητα είτε μεταξύ χρηστών είτε μεταξύ αντικειμένων. Σε κάθε περίπτωση, ο k-NN λειτουργεί ως ένας “έξυπνος γείτονας” που λαμβάνει αποφάσεις βασιζόμενος στα παραδείγματα που βρίσκονται πιο κοντά του.

Παρά την εμφάνιση πιο εξελιγμένων αλγορίθμων, ο k-NN διατηρεί τη θέση του ως ένα βασικό εργαλείο ανάλυσης δεδομένων. Αποτελεί συχνά σημείο εκκίνησης για την επίλυση ενός προβλήματος μάθησης, αλλά και σημείο σύγκρισης (baseline) για την αξιολόγηση πιο πολύπλοκων μοντέλων. Η θεωρητική κατανόηση του αλγορίθμου – από τις συνθήκες συνεκτικότητάς του μέχρι τις βέλτιστες πρακτικές στον υπολογισμό αποστάσεων – είναι

κρίσιμη για κάθε επιστήμονα δεδομένων. Στο πεδίο των social media, όπου τα δεδομένα είναι άφθονα και πολυδιάστατα, η απλότητα του k-NN σε συνδυασμό με τις σύγχρονες βελτιώσεις (π.χ. metric learning, approximate neighbors) εξασφαλίζουν ότι θα συνεχίσει να προσφέρει μια αξιόπιστη προσέγγιση βασισμένη στην έννοια του «δείξε μου τους φίλους σου, να σου πω τι είσαι».

## 4.5 Clustering (*k*-means - DBSCAN)

### 4.5.1 Αλγόριθμος K-Means

Ο K-means είναι ένας από τους πιο δημοφιλείς αλγορίθμους μη-επιβλεπόμενης ομαδοποίησης (clustering) στην ανάλυση δεδομένων. Η βασική ιδέα του αλγορίθμου είναι να χωρίσει ένα σύνολο δεδομένων σε (K) ομάδες (clusters) έτσι ώστε τα σημεία κάθε ομάδας να είναι “κοντά” μεταξύ τους στον χώρο, ενώ σημεία διαφορετικών ομάδων να είναι πιο “μακριά” μεταξύ τους. Κάθε ομάδα χαρακτηρίζεται από το κέντρο της, το οποίο είναι η μέση τιμή (centroid) των σημείων που ανήκουν σε αυτήν την ομάδα. Γεωμετρικά, το κέντρο μιας ομάδας είναι το βαρυκεντρικό σημείο (centroid) των σημείων της ομάδας στον (d)-διάστατο χώρο (υπολογίζεται ως ο μέσος όρος των συντεταγμένων τους). Ο αλγόριθμος K-means προσπαθεί να τοποθετήσει αυτά τα κέντρα με τέτοιο τρόπο ώστε κάθε σημείο δεδομένων να αντιστοιχιστεί στο εγγύτερο κέντρο ομάδας.

Η ομαδοποίηση με βάση τον K-means μπορεί να θεωρηθεί ως μια μορφή κβαντοποίησης χώρου (vector quantization). Η διαίρεση του χώρου σε περιοχές επιρροής γύρω από κάθε κέντρο είναι γνωστή ως διάμεριση Voronoi: κάθε κέντρο ορίζει μία κυψέλη Voronoi που περιλαμβάνει όλα τα σημεία που είναι πιο κοντά σε αυτό το κέντρο παρά σε οποιοδήποτε άλλο. Αυτό σημαίνει ότι το σύνολο των σημείων κάθε ομάδας ( $C_i$ ) ισούται με την αντίστοιχη κυψέλη Voronoi που αντιστοιχεί στο κέντρο ( $\mu_i$ ). Τα όρια μεταξύ των ομάδων είναι υπερεπίπεδα (hyperplanes) που βρίσκονται στους μέσους καθέτους μεταξύ των κέντρων – στην επίπεδη περίπτωση (2 διαστάσεις) είναι ευθείες γραμμές που χωρίζουν το επίπεδο σε περιοχές, όπως φαίνεται στην Εικόνα 1. Έτσι, ο K-means έχει μια σαφή γεωμετρική ερμηνεία: παράγει μια διάσπαση του χώρου σε περιοχές Voronoi με κέντρα τα centroids των ομάδων.

Ιστορικά, η ιδέα της ομαδοποίησης βάσει μέσων τιμών εντοπίζεται ήδη από τον Hugo Steinhaus το 1957. Ο όρος “k-means” εισήχθη από τον James MacQueen (1967) στο πλαίσιο στατιστικής ανάλυσης πολλαπλών μεταβλητών. Παράλληλα, μια ισοδύναμη διαδικασία προτάθηκε ανεξάρτητα από τον Stuart Lloyd το 1957 σε τεχνική έκθεση της Bell Labs (δημοσιεύθηκε το 1982) ως μέθοδος κβαντοποίησης σήματος (Pulse-Code Modulation). Επίσης, ο Edward Forgy (1965) παρουσίασε ουσιαστικά τον ίδιο αλγόριθμο, γι’ αυτό συχνά αναφέρεται και ως αλγόριθμος Lloyd–Forgy. Σήμερα ο K-means αποτελεί τον απλούστερο και ευρύτερα χρησιμοποιούμενο εκπρόσωπο των διαμεριστικών μεθόδων ομαδοποίησης (partitional clustering) σε πλήθος επιστημονικών πεδίων, λόγω της απλότητας, της αποτελεσματικότητας και της πρακτικής του επιτυχίας (Jain, 2010).

Ο αλγόριθμος  $\text{\textbf{K-means}}$  διατυπώνεται ως ένα πρόβλημα ελαχιστοποίησης ενός κόστους εντός των ομάδων. Δεδομένου ενός συνόλου παρατηρήσεων

$$X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \subset \mathbb{R}^d$$

ο στόχος είναι να βρεθεί μια κατάτμηση σε  $K$  ομάδες

$$C = \{C_1, C_2, \dots, C_K\}$$

που ελαχιστοποιεί το άθροισμα των τετραγώνων των αποστάσεων κάθε σημείου από το κέντρο της ομάδας του. Η συνάρτηση κόστους (επίσης γνωστή ως within-cluster sum of squares ή WCSS) ορίζεται ως:

$$J(C) = \sum_{i=1}^K \sum_{\mathbf{x} \in C_i} \|\mathbf{x} - \mu_i\|^2$$

όπου το  $\mu_i$  είναι το κεντροειδές (ή μέσος όρος) της ομάδας  $C_i$ , δηλαδή:

$$\mu_i = \frac{1}{|C_i|} \sum_{\mathbf{x} \in C_i} \mathbf{x}$$

Δηλαδή, για κάθε ομάδα  $C_i$ , υπολογίζεται το άθροισμα των τετραγωνικών αποστάσεων των σημείων από το κέντρο της και το συνολικό κόστος  $J(C)$  είναι το άθροισμα αυτών των ποσοτήτων για όλες τις ομάδες.

Ο στόχος του K-means είναι να βρει την κατάτμηση  $C$  που ελαχιστοποιεί το  $J(C)$ , δηλαδή να ελαχιστοποιήσει τη διασπορά εντός των ομάδων — και ισοδύναμα να μεγιστοποιήσει τη διάκριση μεταξύ των ομάδων, αφού η συνολική διασπορά των δεδομένων θεωρείται σταθερή.

Η παραπάνω βελτιστοποίηση είναι διακριτή (καθώς αφορά την ανάθεση των σημείων σε ομάδες) και υπολογιστικά δύσκολη: έχει αποδειχθεί ότι το πρόβλημα εύρεσης του απόλυτου παγκόσμιου ελαχίστου του  $J(C)$  είναι NP-πλήρες ακόμη και για  $K=2$  σε γενικές συνθήκες.

Στην πράξη, ο αλγόριθμος K-means χρησιμοποιεί μια ευρετική επαναληπτική διαδικασία με στόχο την προσέγγιση ενός (τοπικά βέλτιστου) ελαχίστου του  $J$ . Η διαδικασία αυτή (που περιγράφεται στην επόμενη ενότητα) εξασφαλίζει ότι σε κάθε επανάληψη η τιμή της συνάρτησης κόστους  $J$  μειώνεται μονοτονικά, έως ότου ο αλγόριθμος συγκλίνει σε μια διαμέριση των δεδομένων. Ωστόσο, η τελική αυτή διαμέριση αντιστοιχεί συνήθως σε ένα τοπικό ελάχιστο της  $J$  και όχι απαραίτητα στο ολικό ελάχιστο, λόγω της εξάρτησης από την αρχική επιλογή των κέντρων.

Εναλλακτικά, η συνάρτηση κόστους  $J$  μπορεί να εκφραστεί και ως το άθροισμα των ενδο-ομαδικών διασπορών πολλαπλασιασμένων με το μέγεθος κάθε ομάδας:

$$J(C) = \sum_{i=1}^K |C_i| \cdot \text{Var}(C_i)$$

όπου

$$\text{Var}(C_i) = \frac{1}{|C_i|} \sum_{\mathbf{x} \in C_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2$$

είναι η διακύμανση της ομάδας  $C_i$ . Δηλαδή, το συνολικό κόστος ισοδυναμεί με τη σταθμισμένη άθροιση των διασπορών των επιμέρους ομάδων.

Όπως φαίνεται από τους παραπάνω τύπους, ο K-means βασίζεται στην  $\text{Euclidean distance}$  ως μέτρο ομοιότητας μεταξύ σημείων και κέντρων, συγκεκριμένα χρησιμοποιεί το  $\text{tetragono}$  της Ευκλείδειας απόστασης, δηλαδή:

$$\|\mathbf{x} - \mathbf{y}\|^2 = \sum_{j=1}^d (x_j - y_j)^2$$

Η χρήση διαφορετικής μετρικής, όπως η  $\text{Manhattan distance}$  (L1 norm):

$$\|\mathbf{x} - \mathbf{y}\|_1 = \sum_{j=1}^d |x_j - y_j|$$

αλλάζει το πρόβλημα και η ενημέρωση του μέσου όρου μπορεί να μην οδηγεί σε σύγκλιση. Γι' αυτό έχουν αναπτυχθεί παραλλαγές του αλγορίθμου όπως:

- Ο  $\text{K-medians}$ , που χρησιμοποιεί τη διάμεσο αντί για τη μέση τιμή.
- Ο  $\text{K-medoids}$ , που επιλέγει ως κέντρο το πιο «αντιπροσωπευτικό» υπαρκτό σημείο της ομάδας.
- Ο  $\text{Fuzzy K-means}$ , που επιτρέπει σε κάθε σημείο να ανήκει με βαθμούς συμμετοχής σε πολλές ομάδες.

Ωστόσο, στην κλασική του μορφή, ο K-means είναι άρρηκτα συνδεδεμένος με την Ευκλείδεια γεωμετρία και την αριθμητική μέση τιμή ως το βέλτιστο κέντρο κάθε ομάδας.

Ο αλγόριθμος K-means ακολουθεί μια επαναληπτική διαδικασία δύο βασικών βημάτων: (α) **Βήμα Ανάθεσης (Assignment step)** και (β) **Βήμα Ενημέρωσης (Update step)**. Πριν από αυτά, απαιτείται μια αρχικοποίηση των κέντρων των ομάδων. Συνοπτικά, η διαδικασία έχει ως εξής:

1. **Αρχικοποίηση κέντρων:** Επιλέγουμε αρχικές θέσεις για τα (K) κέντρα (centroids) των ομάδων. Αυτό μπορεί να γίνει τυχαία (π.χ. επιλέγοντας τυχαία (K) σημεία από το σύνολο δεδομένων – γνωστή ως αρχικοποίηση Forgy) ή με πιο εξελιγμένες μεθόδους. Μια σύγχρονη βελτιωμένη μέθοδος αρχικοποίησης είναι ο αλγόριθμος **k-means++** (Arthur & Vassilvitskii, 2007), όπου τα αρχικά κέντρα επιλέγονται έξυπνα ώστε να είναι διάσπαρτα, βελτιώνοντας έτσι τις πιθανότητες σύγκλισης σε καλύτερη λύση. Η επιλογή αρχικών κέντρων είναι κρίσιμη, καθώς διαφορετικές αρχικές τιμές μπορεί να οδηγήσουν τον αλγόριθμο σε διαφορετικά τοπικά ελάχιστα.
2. **Βήμα 1 - Ανάθεση σημείων σε ομάδες:** Δίνοντας τα τρέχοντα κέντρα  $\{\mu_1, \mu_2, \dots, \mu_K\}$ , κάθε σημείο δεδομένων  $x_p$  ανατίθεται στην ομάδα της οποίας το κέντρο βρίσκεται πλησιέστερα με βάση την Ευκλείδεια απόσταση. Δηλαδή, για κάθε  $p$ , βρίσκουμε το  $i$  που ελαχιστοποιεί:

$$i = \arg \min_{j \in \{1, \dots, K\}} \|\mathbf{x}_p - \mu_j\|$$

και αποδίδουμε το  $x_p$  στην ομάδα  $C_i$ .

Αυτό ισοδυναμεί με το να χωρίσουμε τον χώρο σύμφωνα με το διάγραμμα  $\text{Voronoi}$  των τρεχόντων κέντρων: κάθε σημείο πέφτει στην κυψέλη Voronoi του κοντινότερου κέντρου. Μετά το βήμα αυτό, έχουμε μια (ενδεχομένως νέα) κατάτμηση των δεδομένων σε ομάδες, βάσει των επικαιροποιημένων αναθέσεων.

3. **Βήμα 2 - Ενημέρωση των κέντρων:**

Για κάθε ομάδα, υπολογίζουμε εκ νέου το κέντρο της ομάδας ως τον μέσο όρο όλων των σημείων που της έχουν ανατεθεί στο προηγούμενο βήμα. Δηλαδή, αν  $C_i^{(t)}$  είναι το σύνολο σημείων που ανήκουν στην ομάδα  $i$  μετά το βήμα ανάθεσης της επανάληψης  $t$ , τότε το νέο κέντρο δίνεται από:

$$\mu_i^{(t+1)} = \frac{1}{|C_i^{(t)}|} \sum_{\mathbf{x}_j \in C_i^{(t)}} \mathbf{x}_j$$

Αυτό το βήμα επικαιροποιεί τα κέντρα (centroids) ώστε να ευθυγραμμιστούν με τον τρέχοντα διαχωρισμό των ομάδων.

4. **Έλεγχος σύγκλισης:** Εάν μετά τα βήματα 1 και 2 καμία παρατήρηση δεν αλλάξει ομάδα (ή ισοδύναμα, αν τα κέντρα δεν μεταβάλλονται σημαντικά), τότε ο αλγόριθμος έχει συγκλίνει και τερματίζει. Διαφορετικά, επαναλαμβάνουμε τα βήματα Ανάθεσης και Ενημέρωσης (βήματα 2 και 3) διαδοχικά, αποτελώντας μια επαναληπτική διαδικασία βελτίωσης.

Το παραπάνω σχήμα είναι γνωστό και ως αλγόριθμος του Lloyd (Lloyd's algorithm), ειδικά στη βιβλιογραφία της πληροφορικής. Ο Lloyd πρότεινε αυτόν τον επαναληπτικό τρόπο το 1957, ενώ ο MacQueen (1967) περιέγραψε μια παρόμοια διαδικασία σε στατιστικό πλαίσιο. Μια διαφορά είναι ότι ο MacQueen εισήγαγε και μια στοχαστική (online) εκδοχή: αντί να επαναυπολογίζει όλους τους μέσους περιοδικά, ενημέρωνε κάθε μέσο (mu<sub>i</sub>) κάθε φορά που ένα νέο σημείο εισερχόταν στην ομάδα (δηλαδή αναθεώρηση στη ροή των δεδομένων). Παρόλα αυτά, η πιο διαδεδομένη μορφή είναι ο batch αλγόριθμος που περιγράφηκε παραπάνω.

Σε κάθε επανάληψη, τόσο το Βήμα Ανάθεσης όσο και το Βήμα Ενημέρωσης μειώνουν (ή αφήνουν αμετάβλητο) την τιμή της συνάρτησης κόστους (J). Το Βήμα Ανάθεσης, δεδομένων των κέντρων, τοποθετεί κάθε σημείο στην πλησιέστερη ομάδα, που είναι προφανώς η βέλτιστη επιλογή για τη μείωση της συνεισφοράς του στο κόστος. Το Βήμα Ενημέρωσης, δεδομένης της ανάθεσης σημείων, υπολογίζει τον βέλτιστο δυνατό νέο centroid (τον μέσο όρο) που ελαχιστοποιεί το άθροισμα αποστάσεων για τα σημεία της ομάδας. Έτσι, κάθε πλήρης επανάληψη δεν μπορεί να αυξήσει το κόστος και συνήθως το μειώνει αυστηρά. Αυτό είναι κρίσιμο για τις ιδιότητες σύγκλισης του αλγορίθμου, οι οποίες συζητούνται αμέσως μετά.

Ο αλγόριθμος K-means διαθέτει ορισμένες σημαντικές θεωρητικές ιδιότητες:

- **Σύγκλιση:** Ο K-means πάντα συγκλίνει μετά από πεπερασμένο αριθμό επαναλήψεων. Αυτό συμβαίνει διότι υπάρχουν πεπερασμένοι δυνατοί τρόποι να χωριστούν (n) παρατηρήσεις σε (K) ομάδες και σε κάθε επανάληψη το κόστος (J) είτε μειώνεται είτε παραμένει το ίδιο. Αφού το (J) δεν μπορεί να μειώνεται επ' άπειρον (έχει κάτω φράγμα το 0), σε πεπερασμένα βήματα ο αλγόριθμος θα βρεθεί σε σημείο όπου το κόστος δεν μειώνεται περαιτέρω – εκεί επιτυγχάνεται σύγκλιση σε μια σταθερή κατάρτηση των δεδομένων. Συνήθως ο αλγόριθμος συγκλίνει σχετικά γρήγορα (σε λίγες δεκάδες επαναλήψεις ή λιγότερο) στην πράξη.
- **Μονοτονία (μη-αύξηση του κόστους):** Όπως αναφέρθηκε, η τιμή του (J) δεν αυξάνεται ποτέ από επανάληψη σε επανάληψη – αντιθέτως, μειώνεται μονοτονικά ή παραμένει σταθερή όταν φτάσουμε σε βέλτιστη (τοπικά) λύση. Αυτό σημαίνει ότι κάθε βήμα του αλγορίθμου βελτιώνει ή διατηρεί την ποιότητα της ομαδοποίησης. Η μονοτονία εξασφαλίζει ότι δεν θα υπάρξουν κύκλοι όπου ο αλγόριθμος επιστρέφει σε προηγούμενη διάταξη ομάδων με υψηλότερο κόστος.

- **Τοπικά ελάχιστα και εξάρτηση από αρχικοποίηση:** Παρά τη σύγκλιση, το αποτέλεσμα στο οποίο καταλήγει ο K-means μπορεί να είναι ένα τοπικό ελάχιστο του κόστους και όχι το παγκόσμιο βέλτιστο. Ο αλγόριθμος είναι ευαίσθητος στα αρχικά κέντρα: διαφορετικές αρχικές τιμές μπορεί να οδηγήσουν σε διαφορετικές τελικές ομαδοποιήσεις. Για να μετριαστεί αυτό, στην πράξη συχνά τρέχουμε τον K-means πολλαπλές φορές με διαφορετικές τυχαίες αρχικοποιήσεις και κρατούμε το καλύτερο αποτέλεσμα (με το μικρότερο κόστος). Η εισαγωγή της μεθόδου k-means++ (Arthur & Vassilvitskii, 2007) έχει στόχο ακριβώς να βελτιώσει την αρχικοποίηση, έτσι ώστε ο αλγόριθμος να ξεκινά από μια “καλή” κατάσταση και να αποφεύγει με μεγαλύτερη πιθανότητα κακά τοπικά ελάχιστα. Ωστόσο, ακόμη και με έξυπνη αρχικοποίηση, το πρόβλημα παραμένει μη κυρτό και ο αλγόριθμος μπορεί να παγιδευτεί σε ένα τοπικά βέλτιστο διαχωρισμό.
- **Διαμέριση Voronoi & γραμμικά όρια:** Σε κάθε φάση, η ανάθεση των σημείων σε ομάδες ισοδυναμεί με διαίρεση του χώρου σύμφωνα με το διάγραμμα Voronoi των τρεχόντων κέντρων. Στο τελικό στάδιο σύγκλισης, κάθε σημείο βρίσκεται στην κυψέλη Voronoi του πλησιέστερου κέντρου και τα όρια μεταξύ των ομάδων είναι υπερεπίπεδα (γραμμικά όρια). Επομένως, ο K-means παράγει πολυεδρικές (convex polytope) ομάδες στον χώρο: κάθε ομάδα είναι ένα κυρτό πολυτόπιο που ορίζεται από τομή ημίχωρων (τους μέσους καθέτους μεταξύ των τελικών κέντρων). Αυτή η ιδιότητα συνεπάγεται ότι ο K-means λειτουργεί καλύτερα όταν οι «φυσικές» ομάδες στα δεδομένα έχουν περίπου σφαιρικό ή κυρτό σχήμα και παρόμοια διάσταση. Αν τα πραγματικά clusters έχουν πολύ ανισομεγέθεις διαστάσεις ή μη-κυρτά σχήματα, τα γραμμικά όρια του K-means μπορεί να μην μπορέσουν να τα διαχωρίσουν ικανοποιητικά.
- **Υπολογιστική πολυπλοκότητα:** Ο αλγόριθμος Lloyd έχει πολυπλοκότητα περίπου  $\mathcal{O}(n \cdot K \cdot T)$ , όπου (n) ο αριθμός σημείων και (T) ο αριθμός επαναλήψεων μέχρι τη σύγκλιση. Παρότι στη χειρότερη περίπτωση το (T) μπορεί να είναι μεγάλο (θεωρητικά εκθετικό σε κάποιες παθολογικές περιπτώσεις), στην πράξη το (T) είναι συνήθως μικρό. Μεγάλες τιμές του (n) ή/και του (K) μπορεί να κάνουν τον αλγόριθμο αργό, ωστόσο υπάρχουν βελτιστοποιήσεις (π.χ. χρήση δενδροειδών δομών ή τεχνικών μείωσης υπολογισμών) που επιταχύνουν την εκτέλεσή του χωρίς να αλλάζουν το αποτέλεσμα. Για παράδειγμα, υπάρχουν μέθοδοι επιτάχυνσης αξιοποιώντας γεωμετρικές ιδιότητες (π.χ. Elkan’s algorithm που χρησιμοποιεί τρίγωνη ανισότητα, ή μεθόδους συνόλων υποψήφιων centroids)

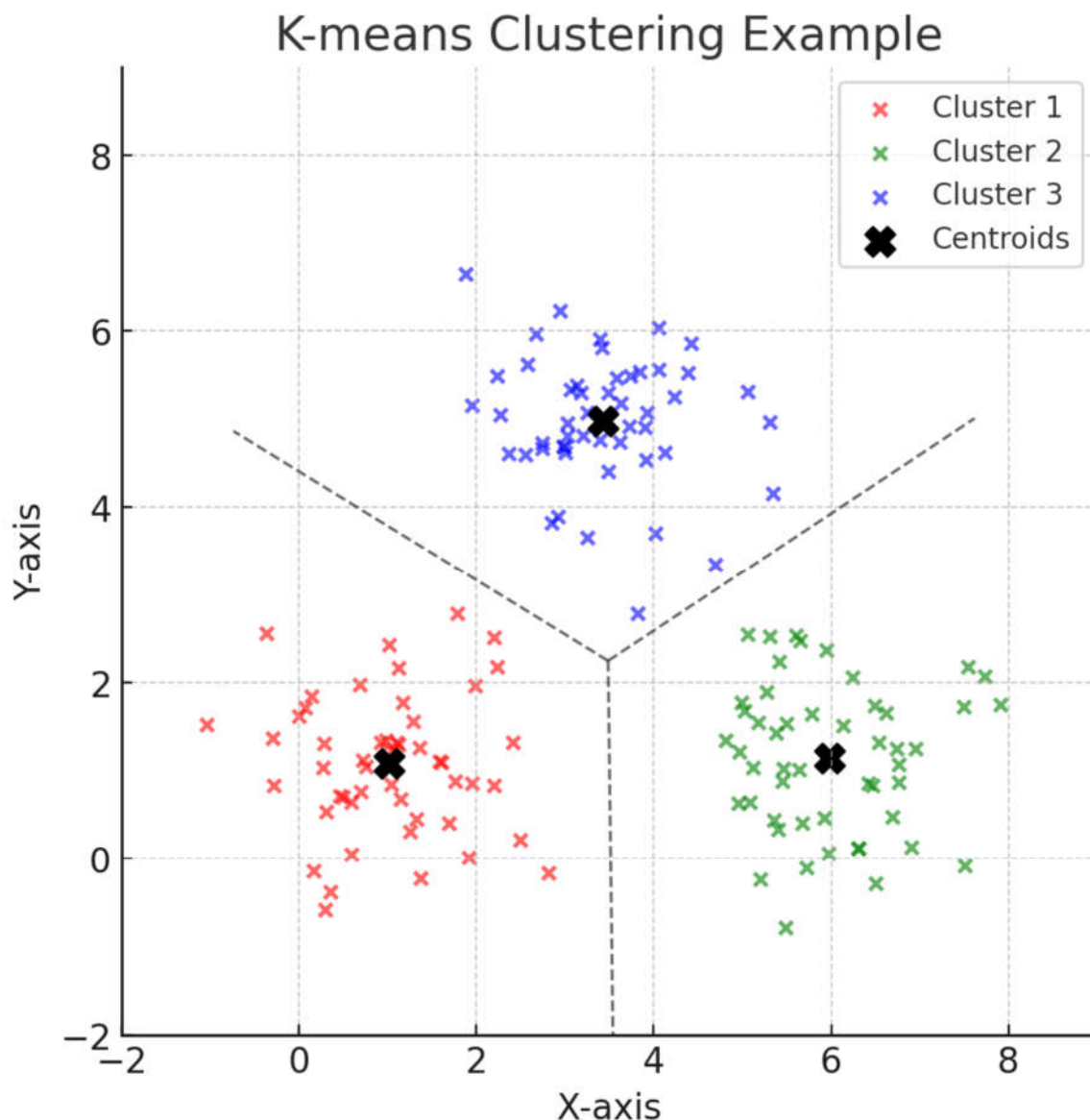
Συνοψίζοντας, ο K-means είναι ένας αλγόριθμος που βελτιστοποιεί εναλλάξ ανάμεσα σε ανάθεση και επανυπολογισμό κέντρων, μειώνοντας μονοτονικά ένα συγκεκριμένο κριτήριο κόστους. Εγγυάται σύγκλιση σε μια λύση (τοπικό ελάχιστο), η οποία χαρακτηρίζεται από γραμμικά όρια Voronoi μεταξύ των παραγόμενων ομάδων. Αυτά τα χαρακτηριστικά –μαζί με την απλότητά του– εξηγούν γιατί ο K-means παραμένει βασικό εργαλείο για την ομαδοποίηση δεδομένων, παρά τους περιορισμούς του.

Ο αλγόριθμος K-means έχει στενή σχέση με την οικογένεια των αλγορίθμων Expectation-Maximization (EM) που χρησιμοποιούνται για την προσαρμογή μοντέλων μειγμάτων κατανομών (mixture models). Συγκεκριμένα, μπορεί να θεωρηθεί ως μια ειδική περίπτωση του EM για ένα μοντέλο μίγματος κανονικών (Gaussian Mixture Model, GMM) υπό αυστηρούς περιορισμούς. Ας εξετάσουμε αυτή τη συσχέτιση:

Στο γενικό μοντέλο μειγμάτων Gauss, θεωρούμε ότι τα δεδομένα προέρχονται από  $\setminus(K\setminus)$  διαφορετικές κανονικές κατανομές. Για την απλότητα, έστω ένα μείγμα ισοτροπικών κανονικών κατανομών με μέσους ( $\mu_k$ ), ομοιοσκεδασίες (ίδια διασπορά) ( $\sigma^2$ ) για κάθε συνιστώσα και ίσα βάρη ανάμειξης  $\pi_k = \frac{1}{K}$ . Ο αλγόριθμος EM για την προσαρμογή αυτού του μοντέλου περιλαμβάνει: (E) βήμα υπολογισμού των πιθανοτήτων συμμετοχής κάθε σημείου σε κάθε κατανομή (soft ανάθεση) και (M) βήμα ενημέρωσης των παραμέτρων (υπολογισμός νέων μέσων και διασποράς) μεγιστοποιώντας τη πιθανοφάνεια. Αν τώρα λάβουμε το όριο όπου η διασπορά ( $\sigma^2$ ) τείνει σε πολύ μικρή τιμή, η πιθανότητα ενός σημείου να ανήκει σε μια συγκεκριμένη κατανομή θα συγκεντρώνεται σχεδόν εξ ολοκλήρου στη κοντινότερη κατανομή (κοντινότερο μέσο). Σε αυτό το όριο, οι *a posteriori* πιθανότητες του E-βήματος γίνονται δυαδικές: κάθε σημείο “ανήκει” πλήρως στο πιο πιθανό cluster. Έτσι, το E-βήμα μετατρέπεται σε ένα σκληρό βήμα ανάθεσης ακριβώς όπως στον K-means (αντί για πιθανότητες συμμετοχής, κάθε σημείο ανατίθεται στο cluster με το μεγαλύτερο βάρος – ή ισοδύναμα με το μικρότερο τετραγωνικό σφάλμα). Στη συνέχεια, το M-βήμα που ενημερώνει τους μέσους υπολογίζοντας το αναμενόμενο μέσο όρο υπό αυτές τις “μονοκατοικημένες” κατανομές, ουσιαστικά υπολογίζει τον μέσο των σημείων κάθε cluster – ακριβώς δηλαδή το βήμα ενημέρωσης centroid του K-means. Οι διασπορές δε χρειάζεται να ενημερωθούν διότι θεωρούνται κοινές και σταθερές (ή στην οριακή περίπτωση, τείνουν στο μηδέν). Με αυτόν τον τρόπο, βλέπουμε ότι ο K-means προκύπτει ως ειδική περίπτωση του EM για ένα συμμετρικό μείγμα Gauss: όλες οι κατανομές έχουν ίδια σχήματα (σφαιρικές), ίσες διασπορές και ίσα *a priori* βάρη. Το μόνο που παραμένει είναι η εναλλαγή ανάμεσα σε πλήρεις αναθέσεις (E-step) και υπολογισμό μέσων (M-step). Αυτή η ισοδυναμία κάνει πιο σαφές το γιατί ο K-means συγκλίνει: είναι ουσιαστικά μια περίπτωση *εναλλακτικής μεγιστοποίησης* (alternating maximization) πάνω σε μια συνάρτηση πιθανότητας (ή ισοδύναμα ελαχιστοποίησης ενός κόστους).

Επιπλέον, η σύγκριση K-means και GMM αναδεικνύει τις υποθέσεις του K-means. Ο K-means αναζητά ομάδες περίπου ίδιου “μεγέθους” και “σχήματος” στο χώρο, διότι επιβάλλει σφαιρική συμμετρία και δεν εκτιμά παραμέτρους διασποράς ή συσχετίσεις μεταξύ μεταβλητών. Ένας πλήρης αλγόριθμος EM για GMM μπορεί να προσαρμόσει ελλειψοειδή clusters με διαφορετικές διαστάσεις, περιστροφές και μεγέθη και παρέχει *πιθανότητες* συμμετοχής (soft clustering), αντί για τη μονοσήμαντη ανάθεση του K-means. Από την άλλη, ο K-means, επειδή ακριβώς είναι απλούστερος, απαιτεί λιγότερες παραμέτρους και είναι φθηνότερος υπολογιστικά, πράγμα που εξηγεί γιατί χρησιμοποιείται συχνά όταν υποθέτουμε ότι οι ομάδες μας είναι περίπου “σφαιρικές” και διακριτές. Συνοψίζοντας, μπορούμε να θεωρήσουμε τον K-means ως μια απλοποιημένη έκδοση του EM για σφαιρικά Gaussian mixtures με ίσες διασπορές και ίσες *a priori* πιθανότητες.

Ο K-means βρίσκει εφαρμογή σε πληθώρα προβλημάτων όπου απαιτείται ομαδοποίηση δεδομένων. Κλασικά παραδείγματα περιλαμβάνουν την ομαδοποίηση πελατών στο μάρκετινγκ (segmentation), την τμηματοποίηση εικόνων (π.χ. ομαδοποίηση pixel με παρόμοιο χρώμα ή ένταση), την ανάκτηση πληροφοριών (clustering εγγράφων ή όρων), την γενετική ανάλυση (ομαδοποίηση γονιδιακών εκφράσεων) και γενικά οποιοδήποτε πρόβλημα όπου θέλουμε να βρούμε φυσικές ομάδες σε ένα σύνολο δεδομένων. Η δημοτικότητα του K-means οφείλεται στο ότι είναι εύκολος στην κατανόηση και υλοποίηση, καθώς και αρκετά αποδοτικός ακόμα και για μεγάλα δεδομένα. Σε πολλές περιπτώσεις, ο K-means χρησιμοποιείται ως αρχικό βήμα (π.χ. για αρχική ομαδοποίηση που θα βελτιωθεί με άλλες μεθόδους, ή για συμπίεση δεδομένων μέσω vector quantization).



Παράδειγμα οπτικής απεικόνισης του αλγορίθμου K-means σε διδιάστατο σύνολο δεδομένων. Τα σημεία διακρίνονται σε τρεις ομάδες (με διαφορετικό χρώμα για κάθε cluster) και τα μαύρα σύμβολα "X" υποδεικνύουν τα κέντρα (centroids)

των ομάδων μετά τη σύγκλιση. Οι διακεκομμένες γραμμές είναι τα σύνορα *Voronoi* που χωρίζουν τον χώρο ανάμεσα στα κέντρα – κάθε τέτοια γραμμή ισαπέχει από δύο κοντινά *centroids* και ορίζει το όριο μεταξύ των δύο αντίστοιχων ομάδων.

Στην παραπάνω εικόνα παρουσιάζεται ένα παράδειγμα εφαρμογής του K-means σε ένα απλό διδιάστατο σύνολο δεδομένων. Τα δεδομένα δημιουργούν φυσικά τρεις πυκνές συστάδες. Ο K-means (με  $K=3$ ) τοποθετεί τρία κέντρα (μαύρα X) και ορίζει τις ομάδες με βάση την εγγύτητα σε αυτά τα κέντρα. Παρατηρούμε ότι κάθε ομάδα καλύπτει μια “περιοχή” του επιπέδου και τα όρια μεταξύ των περιοχών είναι ευθείες γραμμές (*Voronoi* όρια). Τα σημεία κάθε ομάδας συγκεντρώνονται γύρω από το αντίστοιχο *centroid*. Αυτό συμφωνεί με τη θεωρητική πρόβλεψη ότι οι ομάδες του K-means έχουν περιπου σφαιρικό/κυρτό σχήμα.

Παρά τη χρησιμότητά του, είναι επίσης σημαντικό να αναγνωρίζουμε τις περιπτώσεις όπου ο K-means δεν αποδίδει καλά. Αν τα δεδομένα έχουν ομάδες με πολύ ανισοσκελή μεγέθη ή μη συμμετρικές κατανομές, ή αν υπάρχουν outliers, ο K-means μπορεί να δώσει μη ικανοποιητικά αποτελέσματα. Για παράδειγμα, αν δύο clusters έχουν επικαλυπτόμενες ελλειπτικές μορφές, ο K-means (με βάση μόνο αποστάσεις από μέσα) ίσως αποτύχει να τα ξεχωρίσει, σε αντίθεση με πιο περίπλοκα μοντέλα όπως τα Gaussian mixtures που μπορούν να εκτιμήσουν διαφορετικές διασπορές. Επίσης, ο K-means πάντα “διαμερίζει” τον χώρο σε (K) ομάδες ακόμη και αν δεν υπάρχουν τόσες φυσικές ομάδες – έτσι, ο σωστός καθορισμός του (K) είναι ένα ξεχωριστό πρόβλημα (συχνά αντιμετωπίζεται με κριτήρια όπως το Elbow method, ο Silhouette coefficient κ.ά. ή με μεθόδους αυτομάτου προσδιορισμού όπως το X-means).

Συνολικά, ο K-means λόγω της απλότητάς του συχνά λειτουργεί ως πρώτη προσέγγιση για ομαδοποίηση. Τα αποτελέσματά του μπορούν να ερμηνευτούν εύκολα γεωμετρικά (βλέποντας τις θέσεις των *centroids* και τις περιοχές επιρροής τους) και μπορούν να χρησιμεύσουν για περαιτέρω αναλύσεις. Αν το πρόβλημα απαιτεί πιο σύνθετα σχήματα clusters, μπορεί να επεκταθεί ή να αντικατασταθεί από άλλες μεθόδους (π.χ. Gaussian mixture models με EM, ιεραρχική ομαδοποίηση, DBSCAN για αυθαίρετα σχήματα clusters κλπ.). Παρόλα αυτά, η κατανόηση του K-means αποτελεί θεμέλιο λίθο για την κατανόηση των πιο πολύπλοκων τεχνικών ομαδοποίησης.

#### 4.5.2 Αλγόριθμος DBSCAN

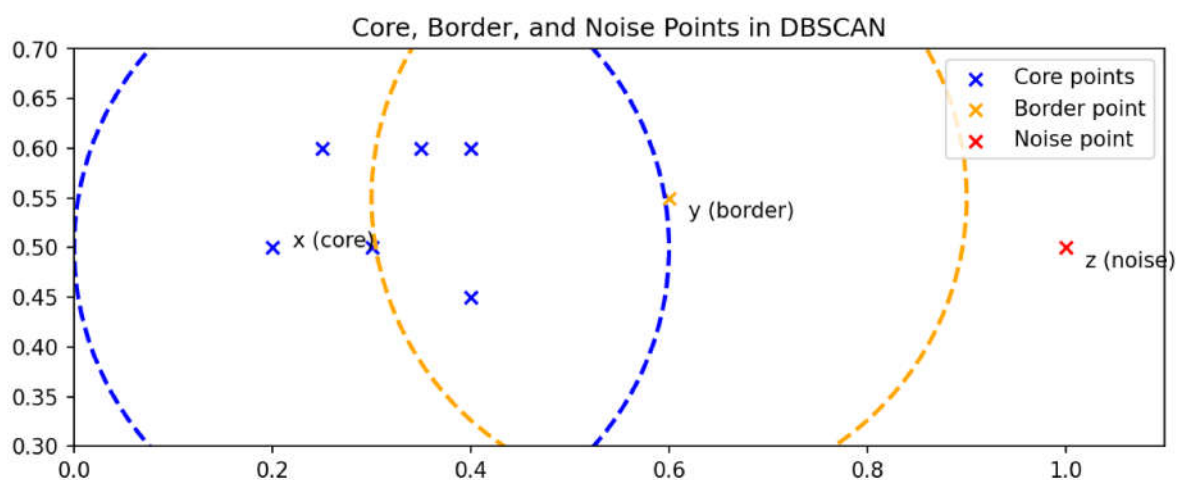
Ο DBSCAN (Density-Based Spatial Clustering of Applications with Noise) είναι ένας από τους πιο διαδεδομένους αλγόριθμους συσταδοποίησης, ο οποίος βασίζεται στην πυκνότητα των σημείων στον χώρο. Σε αντίθεση με τον K-means και άλλες ιεραρχικές μεθόδους που προϋποθέτουν την εκ των προτέρων επιλογή αριθμού συστάδων και την ύπαρξη σφαιρικών ομάδων με παρόμοιο μέγεθος, ο DBSCAN μπορεί να εντοπίσει συστάδες αυθαίρετου σχήματος, χρησιμοποιώντας αποκλειστικά κριτήρια πυκνότητας (Ester et al., 1996).

Η βασική ιδέα του DBSCAN είναι ότι οι συστάδες μπορούν να θεωρηθούν ως περιοχές με υψηλή συγκέντρωση σημείων, οι οποίες διαχωρίζονται μεταξύ τους από περιοχές με χαμηλότερη πυκνότητα. Η μέθοδος αυτή είναι εξαιρετικά αποτελεσματική στην αναγνώριση ανωμαλιών και θορύβου, καθώς τα σημεία που βρίσκονται σε αραιές περιοχές δεν εντάσσονται σε καμία συστάδα αλλά χαρακτηρίζονται ως θόρυβος (noise) (Hex.Tech, 2023).

Για να υλοποιήσει τη στρατηγική του, ο DBSCAN εισάγει τις εξής βασικές κατηγορίες σημείων:

- **Πυρήνες (core points):** Σημεία που περιβάλλονται από τουλάχιστον  $MinPts$  άλλα σημεία εντός μιας απόστασης  $\epsilon$  (epsilon). Αυτά αποτελούν τα κεντρικά σημεία μιας συστάδας.
- **Οριακά σημεία (border points):** Δεν πληρούν το κριτήριο των  $MinPts$ , αλλά βρίσκονται εντός της ακτίνας  $\epsilon$  ενός πυρήνα και γι' αυτό εντάσσονται στη συστάδα του.
- **Σημεία θορύβου (noise points):** Δεν πληρούν καμία από τις παραπάνω προϋποθέσεις και θεωρούνται ως μη ανήκοντα σε οποιαδήποτε συστάδα.

Η χρήση του DBSCAN είναι ιδιαίτερα ευεργετική σε περιβάλλοντα υψηλής διαστασιμότητας ή δεδομένων με μη γραμμικές συσχετίσεις, όπου παραδοσιακές τεχνικές αποτυγχάνουν να εντοπίσουν πολυπλοκότητες στη δομή των δεδομένων. Σε τέτοια πλαίσια, η μηχανική μάθηση –και ειδικά μέθοδοι όπως ο DBSCAN– επιτρέπουν την εξαγωγή σημαντικών συμπερασμάτων, διευκολύνοντας τη λήψη αποφάσεων και την πρόβλεψη μέσω της αποκάλυψης κρυμμένων προτύπων (James et. al., 2021). Στο παρακάτω σχήμα απεικονίζεται η ταξινόμηση σημείων σε αυτές τις κατηγορίες. Το σημείο x (μπλε) είναι πυρήνας, καθώς στη γειτονιά ακτίνας  $\epsilon$  (μπλε διακεκομμένος κύκλος) έχει επαρκείς γείτονες. Το σημείο y (πορτοκαλί) είναι σημείο ορίου: βρίσκεται εντός της  $\epsilon$ -γειτονιάς του x (πορτοκαλί κύκλος που επικαλύπτεται με τον μπλε), αλλά από μόνο του δεν έχει αρκετούς γείτονες ώστε να είναι πυρήνας. Το σημείο z (κόκκινο) είναι θόρυβος, καθώς δεν βρίσκεται στην πυκνή περιοχή κανενός πυρήνα (βρίσκεται εκτός των κύκλων).



Εικονογράφηση των τύπων σημείων στον DBSCAN. Το σημείο  $x$  (μπλε) είναι πυρήνας (core), με γείτονες εντός ακτίνας  $\varepsilon$  (μπλε διακεκομμένη)  $\geq MinPts$ . Το σημείο  $y$  (πορτοκαλί) είναι οριακό (border), εντός της  $\varepsilon$ -γειτονιάς του  $x$ , αλλά με λιγότερους γείτονες από  $MinPts$ . Το σημείο  $z$  (κόκκινο) είναι θόρυβος (noise), εκτός πυκνών περιοχών.

Μια ακόμη θεμελιώδης έννοια είναι η πυκνότητα-συνδεσιμότητα των σημείων. Ο DBSCAN ορίζει ότι μια συστάδα είναι το σύνολο των σημείων που είναι πυκνωτικά συνδεδεμένα μεταξύ τους μέσω μιας αλυσίδας σημείων-πυρήνων. Πιο τυπικά:

- Ένα σημείο  $q$  είναι άμεσα πυκνωτικά προσεγγίσιμο (directly density-reachable) από ένα σημείο  $p$  αν  $q$  βρίσκεται εντός της  $\varepsilon$ -γειτονιάς του  $p$  και το  $p$  είναι πυρήνας.
- Ένα σημείο  $q$  είναι πυκνωτικά προσεγγίσιμο (density-reachable) από ένα σημείο  $p$  αν υπάρχει αλληλουχία σημείων  $p = p_1, p_2, \dots, p_n = q$ , όπου κάθε  $p_{i+1}$  είναι άμεσα προσεγγίσιμο από το  $p_i$ . Αυτό σημαίνει πρακτικά ότι μπορούμε να «περπατήσουμε» από το  $p$  στο  $q$  μετακινώντας από πυρήνα σε πυρήνα, με βήματα που δεν ξεπερνούν την απόσταση  $\varepsilon$ .
- Δύο σημεία  $p$  και  $q$  είναι πυκνωτικά συνδεδεμένα (density-connected) αν υπάρχει κάποιο σημείο  $o$  τέτοιο ώστε τόσο το  $p$  όσο και το  $q$  να είναι πυκνωτικά προσεγγίσιμα από το  $o$ . Δηλαδή, υπάρχει τουλάχιστον ένας πυρήνας από τον οποίο και τα δύο σημεία είναι προσβάσιμα μέσω πυκνών διαδρομών.

Με βάση αυτά, ορίζεται ότι μια συστάδα DBSCAN είναι κάθε μέγιστο σύνολο σημείων που είναι όλα πυκνωτικά συνδεδεμένα μεταξύ τους. Κάθε συστάδα περιλαμβάνει τουλάχιστον ένα σημείο-πυρήνα και πιθανά οριακά σημεία συνδεδεμένα με αυτόν. Τα σημεία θορύβου είναι ακριβώς εκείνα που δεν είναι πυκνωτικά συνδεδεμένα με κανένα άλλο σημείο – με άλλα λόγια, είναι απομονωμένα.

Βάσει των παραπάνω εννοιών, ο αλγόριθμος DBSCAN λειτουργεί αναζητώντας περιοχές υψηλής πυκνότητας και επεκτείνοντας τις συστάδες από σημεία-πυρήνες προς τα έξω. Τα βασικά βήματα του αλγορίθμου είναι τα εξής:

1. **Έναρξη με ένα μη επισκεφθέν σημείο:** Ο αλγόριθμος ξεκινά επιλέγοντας ένα αυθαίρετο σημείο από το σύνολο δεδομένων που δεν έχει ακόμα επισκεφθεί (δηλαδή δεν έχει ταξινομηθεί σε κάποια συστάδα ή χαρακτηριστεί ως θόρυβος).
2. **Εύρεση γειτόνων:** Υπολογίζονται όλοι οι γείτονες του επιλεγμένου σημείου που βρίσκονται σε απόσταση  $\leq \varepsilon$  (ακτίνα). Αυτή είναι η  $\varepsilon$ -γειτονιά του σημείου.
  - ο Αν ο αριθμός των σημείων στη γειτονιά (συμπεριλαμβανομένου και του ίδιου) είναι μικρότερος από  $MinPts$ , τότε το αρχικό σημείο χαρακτηρίζεται προσωρινά ως θόρυβος (δεν μπορεί να ξεκινήσει συστάδα). Σημειώνουμε ότι ένα σημείο χαρακτηρισμένο αρχικά ως θόρυβος μπορεί αργότερα να γίνει μέρος κάποιας συστάδας αν βρεθεί κοντά σε κάποιον άλλο πυρήνα.
  - ο Αν βρεθούν τουλάχιστον  $MinPts$  σημεία στη γειτονιά, τότε το αρχικό σημείο είναι πυρήνας. Δημιουργείται μια νέα συστάδα που περιλαμβάνει αυτό το σημείο και όλους τους γείτονές του (όλα τα σημεία εντός ακτίνας  $\varepsilon$ ).
3. **Επέκταση της συστάδας:** Για κάθε σημείο που προστέθηκε στη συστάδα, αν αυτό το σημείο είναι πυρήνας (έχει και αυτό αρκετούς δικούς του γείτονες), τότε όλοι οι

γείτονες του προστίθενται επίσης στη συστάδα. Η διαδικασία επεκτείνεται επαναληπτικά: κάθε φορά που προστίθενται νέα σημεία-πυρήνες, εξετάζονται και οι δικές τους  $\varepsilon$ -γειτονιές, προσθέτοντας επιπλέον σημεία στη συστάδα. Έτσι, η συστάδα «μεγαλώνει» συνδέοντας μεταξύ τους γειτονικές περιοχές υψηλής πυκνότητας.

4. **Οριστικοποίηση συστάδας:** Η επέκταση συνεχίζεται έως ότου δεν μπορούν να προστεθούν άλλα σημεία στη συστάδα (δηλαδή όταν για όλα τα σημεία της συστάδας είτε δεν υπάρχουν μη επισκεφθέντες γείτονες, είτε οι γείτονές τους είναι ήδη στη συστάδα ή δεν πληρούν τα κριτήρια πυκνότητας). Τότε η τρέχουσα συστάδα θεωρείται ολοκληρωμένη.
5. **Συνέχιση με επόμενο σημείο:** Επιλέγεται ένα νέο μη επισκεφθέν σημείο εκτός συστάδων και επαναλαμβάνεται η διαδικασία (βήματα 2–4) για ενδεχόμενη δημιουργία νέας συστάδας. Αυτό επαναλαμβάνεται μέχρι να επισκεφθούν όλα τα σημεία του συνόλου δεδομένων.
6. **Σημεία θορύβου:** Όσα σημεία παραμένουν χωρίς να ενταχθούν σε κάποια συστάδα μετά την ολοκλήρωση των παραπάνω βημάτων επισημαίνονται τελικά ως θόρυβος. Αυτά είναι τα σημεία που είτε χαρακτηρίστηκαν θόρυβος εξ αρχής είτε δεν βρέθηκαν ποτέ εντός  $\varepsilon$ -γειτονιάς κάποιου πυρήνα κατά τη διάρκεια της διαδικασίας.

Με αυτή τη διαδικασία, ο DBSCAN κατορθώνει να ανακαλύψει συστάδες αυθαίρετων σχημάτων και μεγέθους, αρκεί να είναι επαρκώς πυκνές περιοχές στον χώρο. Επιπλέον, απομονώνει φυσικά τα διάσπαρτα σημεία (αραιές περιοχές) ως θόρυβο. Σημειώνεται ότι η ποιότητα των αποτελεσμάτων εξαρτάται σε μεγάλο βαθμό από την ορθή επιλογή των δύο βασικών παραμέτρων του αλγορίθμου ( $\varepsilon$  και  $MinPts$ ), οι οποίες συζητώνται στη συνέχεια.

Ο DBSCAN έχει δύο κύριες παραμέτρους που καθορίζουν τη συμπεριφορά του: την ακτίνα  $\varepsilon$  (έψιλον) και τον ελάχιστο αριθμό σημείων  $MinPts$ . Η σωστή ρύθμιση αυτών των παραμέτρων είναι κρίσιμη για την επιτυχία του αλγορίθμου, καθώς προσδιορίζει τι θεωρείται «πυκνή» περιοχή.

- **Ακτίνα  $\varepsilon$ :** Καθορίζει την εμβέλεια της “γειτονιάς” ενός σημείου. Δύο σημεία θεωρούνται γείτονες (δηλ. συνδέονται άμεσα) εάν η απόσταση μεταξύ τους είναι μικρότερη ή ίση από  $\varepsilon$ . Μια μικρή τιμή  $\varepsilon$  σημαίνει ότι απαιτούμε τα σημεία να είναι πολύ κοντά για να θεωρηθούν μέρος της ίδιας πυκνής περιοχής, ενώ μια μεγάλη τιμή  $\varepsilon$  επιτρέπει ευρύτερες αραιές περιοχές να συγχωνεύονται σε μία συστάδα. Η επιλογή του  $\varepsilon$  είναι κρίσιμη:
  - ο Αν  $\varepsilon$  είναι υπερβολικά μικρό, τότε μόνο ελάχιστα σημεία θα θεωρούνται γείτονες και πολλά πραγματικά συσταδοποιημένα σημεία μπορεί να μην συνδεθούν – με αποτέλεσμα πολλές μικρές συστάδες ή πολλά σημεία να χαρακτηριστούν ως θόρυβος.
  - ο Αν  $\varepsilon$  είναι υπερβολικά μεγάλο, τότε περιοχές που θα έπρεπε να είναι διακριτές συστάδες μπορεί να συγχωνευθούν σε μία και ακόμη χειρότερα, μπορεί ουσιαστικά ολόκληρο το σύνολο να θεωρηθεί μια ενιαία συστάδα. Η

υπερβολική αύξηση του  $\varepsilon$  ουσιαστικά μειώνει την ευαισθησία του αλγορίθμου στην τοπική πυκνότητα, θολώνοντας τη «δομή» των δεδομένων.

- ο Ένας πρακτικός τρόπος εύρεσης κατάλληλου  $\varepsilon$  είναι η χρήση διαγράμματος απόστασης  $k$ -πλησιέστερου γείτονα ( $k$ -distance graph). Συγκεκριμένα, υπολογίζεται για κάθε σημείο η απόσταση από τον  $k$ -οστό κοντινότερο γείτονα του (όπου  $k = \text{MinPts}$ ). Οι αποστάσεις αυτές ταξινομούνται και σχεδιάζονται. Το σημείο στο οποίο η καμπύλη του διαγράμματος παρουσιάζει “γόνατο” (elbow) – δηλαδή μια απότομη μεταβολή της κλίσης – θεωρείται συχνά καλή επιλογή για  $\varepsilon$ . Αυτό ανταποκρίνεται στην ιδέα ότι μέχρι ένα συγκεκριμένο όριο απόστασης τα σημεία βρίσκουν αρκετούς γείτονες (εντός πυκνών περιοχών), ενώ πέρα από αυτό το όριο οι αποστάσεις αυξάνονται απότομα (αραιές περιοχές). Το “γόνατο” αυτής της καμπύλης δίνει μια κατά προσέγγιση τιμή για το  $\varepsilon$  που διαχωρίζει τις πυκνές από τις αραιές περιοχές.
- ο Τέλος, η επιλογή του  $\varepsilon$  μπορεί να καθοδηγηθεί και από γνώση του πεδίου (domain knowledge). Αν γνωρίζουμε, για παράδειγμα, ότι στη συγκεκριμένη εφαρμογή μια “φυσική” ακτίνα πυκνότητας είναι μια συγκεκριμένη τιμή (π.χ. 100 μέτρα σε γεωγραφικά δεδομένα), μπορούμε να την χρησιμοποιήσουμε ως αφετηρία.
- **Ελάχιστος αριθμός σημείων  $\text{MinPts}$ :** Καθορίζει πόσα σημεία απαιτούνται τουλάχιστον εντός της ακτίνας  $\varepsilon$  για να θεωρηθεί μια περιοχή αρκετά πυκνή ώστε να αποτελεί συστάδα. Με άλλα λόγια,  $\text{MinPts}$  είναι το κατώφλι για να χαρακτηριστεί ένα σημείο ως πυρήνας. Συνήθως  $\text{MinPts} \geq 3$  (καθώς με 1 ή 2 σημεία δεν έχει νόημα να οριστεί συστάδα). Ο αντίκτυπος του  $\text{MinPts}$  είναι ο εξής:
  - ο Μικρό  $\text{MinPts}$  (π.χ. 2 ή 3) κάνει τον αλγόριθμο πιο ευαίσθητο στο να σχηματίζει συστάδες. Ακόμη και λίγα σημεία κοντά θα δημιουργήσουν συστάδα, κάτι που μπορεί να οδηγήσει σε υπερβολικό κατακερματισμό του συνόλου (μικρές συστάδες ή θόρυβο να γίνεται συστάδα).
  - ο Μεγάλο  $\text{MinPts}$  κάνει τον αλγόριθμο πιο αυστηρό – απαιτεί πιο πυκνές συγκεντρώσεις για να αναγνωρίσει συστάδα. Αυτό μπορεί να αγνοήσει μικρότερες αλλά σημαντικές συστάδες και να τις θεωρήσει θόρυβο.
  - ο Μια ευρέως χρησιμοποιούμενη εμπειρική υπόδειξη είναι να ρυθμίζεται το  $\text{MinPts}$  ανάλογα με τις διαστάσεις (features) του δεδομένου: π.χ.  $\text{MinPts}$  περίπου ίσο με  $2 * D$ , όπου  $D$  ο αριθμός διαστάσεων, μπορεί να είναι ένα σημείο εκκίνησης. Σε διδιάστατα δεδομένα ( $D=2$ ) συχνά  $\text{MinPts} = 4$  ή  $5$  είναι μια λογική αρχική τιμή.
  - ο Εάν τα δεδομένα περιέχουν πολύ θόρυβο ή εάν θέλουμε να εντοπίσουμε και πολύ μικρές συστάδες, μπορεί να χρησιμοποιηθεί μικρότερο  $\text{MinPts}$ . Αντίθετα, σε πολύ μεγάλα σύνολα δεδομένων, ένα μεγαλύτερο  $\text{MinPts}$  βοηθά να αποφύγουμε τη δημιουργία πολυάριθμων μικρών συστάδων και να έχουμε πιο ουσιαστικές ομάδες.

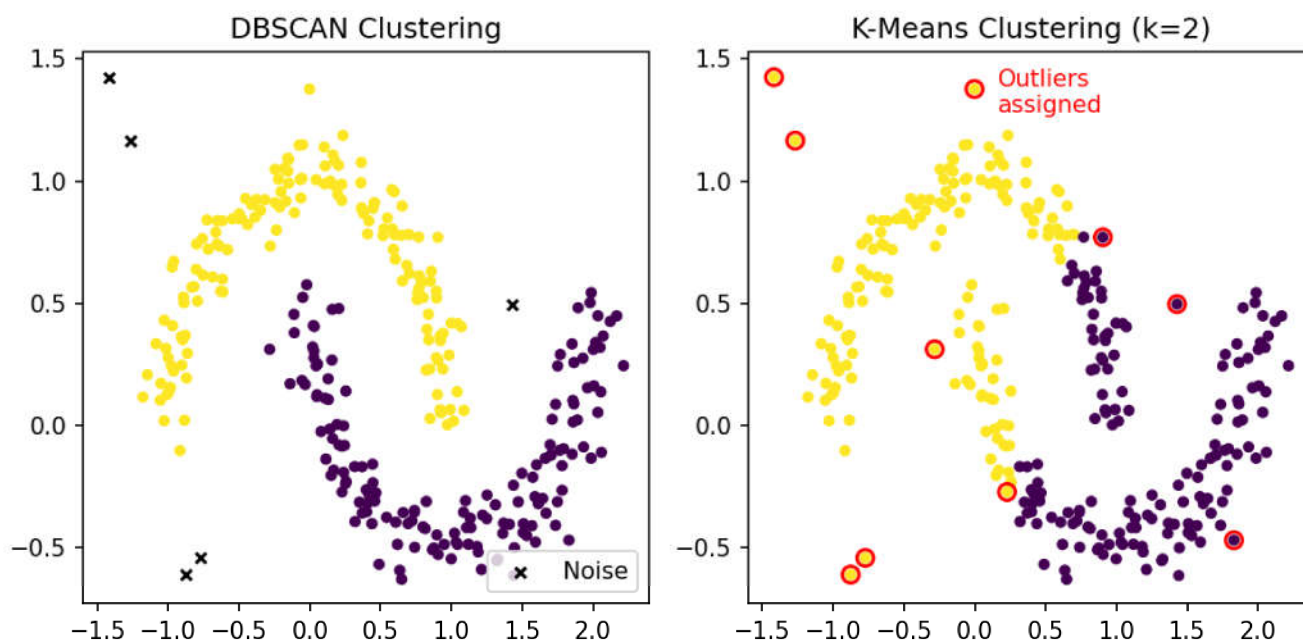
Επειδή η επιλογή των  $\varepsilon$  και  $\text{MinPts}$  είναι αλληλεξαρτώμενη, συχνά απαιτείται πειραματισμός: δοκιμή διαφόρων συνδυασμών και αξιολόγηση των αποτελεσμάτων (π.χ. μέσω μετρικών όπως ο δείκτης σιλουέτας ή οπτική επιθεώρηση των συστάδων). Συνοψίζοντας, ένα λογικό

workflow είναι: αρχική εκτίμηση  $\varepsilon$  με το διάγραμμα  $k$ -απόστασης, επιλογή  $MinPts$  με βάση τις διαστάσεις (ή εμπειρία) και στη συνέχεια δοκιμή/διόρθωση αυτών των τιμών λαμβάνοντας υπόψη τη γνώση του προβλήματος και την ποιότητα των συστάδων.

Στην πράξη, ο DBSCAN ξεχωρίζει διότι μπορεί να εντοπίσει συστάδες αυθαίρετων σχημάτων και να αφήσει «εκτός» τα σημεία θορύβου (Ester et al., 1996). Παρακάτω παρουσιάζεται ένα παράδειγμα σύγκρισης του DBSCAN με τον K-means σε ένα δισδιάστατο σύνολο δεδομένων. Το σύνολο δεδομένων αποτελείται από δύο ημικυκλικές συστάδες (το κλασικό σετ “two moons”) με προσθήκη λίγων σημείων θορύβου διάσπαρτων τριγύρω.

Στο παρακάτω σχήμα φαίνεται πως ο DBSCAN ταξινομεί σωστά τις δύο ημικυκλικές συγκεντρώσεις σημείων ως ξεχωριστές συστάδες (κίτρινα και μωβ σημεία αριστερά). Επιπλέον, τα λίγα σημεία που είναι διάσπαρτα μακριά από αυτές τις συστάδες δεν εντάσσονται σε καμία από τις δύο – σημειώνονται ως θόρυβος (μαύρα σημάδια x). Αντίθετα, ο αλγόριθμος K-means με  $k=2$  (δεξιά) προσπάθησε να χωρίσει τα δεδομένα σε δύο ομάδες, αλλά λόγω της υπόθεσης του για σφαιρικές συστάδες καταλήγει να *κόψει* τα ημικύκλια σε μη φυσικούς διαχωρισμούς (Jain, 2010; Tan et. al., 2018).

Επιπρόσθετα, ο K-means δεν έχει έννοια θορύβου: όλα τα σημεία πρέπει να ανήκουν σε κάποια ομάδα. Έτσι, τα απομονωμένα outlier σημεία (κόκκινοι κύκλοι στο δεξί διάγραμμα) αναγκαστικά αποδίδονται σε μία από τις δύο ομάδες, παρότι δεν ταιριάζουν καλά σε καμία. Αυτό έχει ως αποτέλεσμα την αλλοίωση των κεντροειδών και εν τέλει λιγότερο ικανοποιητική ομαδοποίηση για τα κύρια clusters (Xu & Tian, 2015).



Σύγκριση αποτελέσματος συσταδοποίησης με DBSCAN (αριστερά) και K-means (δεξιά) στο ίδιο σύνολο δεδομένων. Ο DBSCAN εντοπίζει σωστά τις δύο καμπύλες συστάδες (με χρώματα μωβ και κίτρινο) και επισημαίνει τα διάσπαρτα

σημεία ως θόρυβο (μαύρα  $x$ ). Ο *K-means* με  $k=2$  χωρίζει τα δεδομένα σε δύο ομάδες, όμως αναγκάζεται να εντάξει και τα απομονωμένα *outliers* σε αυτές (τα περιγεγραμμένα με κόκκινο κύκλο σημεία δεξιά), διαστρεβλώνοντας έτσι τα κέντρα των δύο ομάδων.

Το παράδειγμα αυτό καταδεικνύει δύο βασικές αρετές του DBSCAN:

- (α) την ικανότητα να εντοπίζει συστάδες με περίπλοκο σχήμα (όπως οι καμπύλες “μισοφέγγαρα” που αποτυγχάνουν να αναγνωριστούν σωστά από τον *K-means*) και
- (β) την εγγενή αντιμετώπιση των θορυβωδών σημείων, αφήνοντάς τα εκτός συστάδων αντί να αλλοιωθούν οι συστάδες από αυτά.

Αυτές οι ιδιότητες είναι πολύτιμες σε δεδομένα πραγματικών εφαρμογών όπου οι συστάδες δεν είναι συμμετρικές ή όπου υπάρχουν μεμονωμένες ανωμαλίες (Scikit-learn Developers, 2025; Aggarwal, 2015).

Για να αξιολογηθεί ολοκληρωμένα ο DBSCAN, είναι σημαντικό να τον συγκρίνουμε με άλλες δημοφιλείς μεθόδους συσταδοποίησης, όπως ο *K-means* και ο PAM (Partitioning Around Medoids), αναδεικνύοντας τα πλεονεκτήματα και τα μειονεκτήματά του σε σχέση με αυτές (Jain, 2010).

### Σε σχέση με τον K-means

Ο αλγόριθμος *K-means* είναι ίσως η πιο γνωστή μέθοδος clustering. Λειτουργεί τοποθετώντας  $k$  κέντρα (*centroids*) και αναθέτοντας κάθε σημείο στο κοντινότερο κέντρο, επαναλαμβάνοντας μέχρι τα κέντρα να σταθεροποιηθούν (MacQueen, 1967; Lloyd, 1982). Ο *K-means* έχει μερικές σημαντικές διαφορές από τον DBSCAN:

- **Απαίτηση για αριθμό συστάδων:** Ο *K-means* απαιτεί από τον χρήστη να ορίσει εκ των προτέρων τον αριθμό συστάδων  $k$ . Αυτή η ανάγκη μπορεί να είναι περιοριστική όταν δεν υπάρχει γνώση του πόσες ομάδες υπάρχουν στα δεδομένα (Tan, Steinbach, & Kumar, 2018). Αντίθετα, ο DBSCAN δεν χρειάζεται αριθμό συστάδων εκ των προτέρων – καθορίζει από μόνος του πόσες συστάδες υπάρχουν με βάση τα δεδομένα και τις παραμέτρους πυκνότητας (Ester et al., 1996). Αυτό καθιστά τον DBSCAN ιδιαίτερα κατάλληλο για εξερευνητική ανάλυση χωρίς εκ των προτέρων υποθέσεις για τη δομή των δεδομένων.
- **Σχήμα συστάδων:** Ο *K-means* βασίζεται στην έννοια του κεντροειδούς και της απόστασης (συνήθως ευκλείδειας) και τείνει να δημιουργεί σφαιρικές (*globular*) και περίπου ίσου μεγέθους συστάδες (Xu & Tian, 2015). Εάν τα πραγματικά δεδομένα έχουν συστάδες επιμηκυμένες, καμπύλες ή γενικά αυθαίρετου σχήματος, ο *K-means* ενδέχεται να αποτύχει να τις συλλάβει σωστά. Αντίθετα, ο DBSCAN μπορεί να εντοπίσει συστάδες οποιουδήποτε σχήματος, αρκεί αυτές να είναι πυκνές (Schubert et al., 2017). Όπως φάνηκε και στο προηγούμενο σχήμα, ο DBSCAN εντοπίζει τα “μισοφέγγαρα” ενώ ο *K-means* τα παραμορφώνει.

- **Ευαισθησία σε θόρυβο και ανωμαλίες:** Ο K-means δεν διαθέτει μηχανισμό χαρακτηρισμού θορύβου. Κάθε σημείο, ακόμα και απομονωμένο, θα ανατεθεί σε κάποιο cluster, γεγονός που τον καθιστά ευαίσθητο σε outliers (Aggarwal, 2015). Αντίθετα, ο DBSCAN διαθέτει ενσωματωμένο μηχανισμό απομόνωσης των outliers – οι θορυβώδεις παρατηρήσεις δεν εντάσσονται σε καμία ομάδα, με αποτέλεσμα οι συστάδες να μην αλλοιώνονται από μεμονωμένα σημεία εκτός πυκνότητας.
- **Υπολογιστική απόδοση:** Ο K-means έχει συνήθως γραμμική πολυπλοκότητα ως προς τον αριθμό σημείων  $n$ , δηλαδή  $O(n \cdot k \cdot T)$ , όπου  $T$  είναι ο αριθμός επαναλήψεων (Arthur & Vassilvitskii, 2007). Αυτό τον καθιστά αποδοτικό σε μεγάλα σύνολα δεδομένων. Ο DBSCAN έχει θεωρητικά χειρότερη πολυπλοκότητα  $O(n \log n)$  με ευρετήριο γειτονιάς (όπως kd-tree), ή και  $O(n^2)$  στη χειρότερη περίπτωση χωρίς ευρετήριο (Schubert et al., 2017). Συνεπώς, για πολύ μεγάλα δεδομένα, ο K-means ενδέχεται να είναι ταχύτερος.
- **Ποιότητα συστάδων:** Σε δεδομένα όπου οι συστάδες είναι πράγματι σφαιρικές και χωρίς θόρυβο, ο K-means δίνει καλά αποτελέσματα, με καθαρά διαχωρισμένες ομάδες (Jain, 2010). Μάλιστα, σε τέτοια περιβάλλοντα είναι προτιμητέος λόγω απλότητας και ταχύτητας. Ωστόσο, σε δεδομένα με πολύπλοκη δομή ή άγνωστο αριθμό συστάδων, ο DBSCAN αποδίδει καλύτερα, αναδεικνύοντας πρότυπα που ο K-means μπορεί να αγνοήσει.

Συνοψίζοντας, ο K-means υπερέχει σε απλότητα και ταχύτητα όταν οι συστάδες είναι συμμετρικές και καθαρά διαχωρισμένες, αλλά υστερεί σε περιβάλλοντα με θόρυβο ή αυθαίρετες γεωμετρικές δομές. Ο DBSCAN είναι πιο ευέλικτος και ανθεκτικός σε θόρυβο και ενδείκνυται όταν δεν έχουμε καθορισμένες παραδοχές για τη μορφή των δεδομένων (Ester et al., 1996; Scikit-learn Developers, 2025).

### Σε σχέση με τον PAM (K-medoids)

Ο αλγόριθμος PAM (Partitioning Around Medoids), γνωστός και ως k-medoids, αποτελεί μια παραλλαγή του K-means με βασική διαφορά ότι χρησιμοποιεί medoids αντί για centroids. Ένας medoid είναι ένα πραγματικό σημείο από το σύνολο των δεδομένων που ελαχιστοποιεί τη συνολική απόσταση προς τα υπόλοιπα σημεία της ομάδας του (Kaufman & Rousseeuw, 1990). Αυτό τον καθιστά λιγότερο ευαίσθητο σε ακραίες τιμές από τον K-means.

- **Αριθμός συστάδων και σχήμα:** Όπως και ο K-means, ο PAM απαιτεί την προκαθορισμένη εισαγωγή του αριθμού των συστάδων  $k$ , πράγμα που περιορίζει την ικανότητα προσαρμογής του σε άγνωστες δομές δεδομένων (Kassambara, 2017). Επιπλέον, λόγω της βασισμένης σε medoids προσέγγισης και της απόδοσης κάθε σημείου στο κοντινότερο medoid, ο PAM είναι περισσότερο κατάλληλος για εύρεση συμπαγών και σχετικά ισομεγέθων συστάδων (Park & Jun, 2009). Αν και υποστηρίζει διαφορετικά μέτρα απόστασης πέρα από την ευκλείδεια, δεν αποδίδει καλά όταν οι

συστάδες έχουν μη σφαιρικά ή περίπλοκα σχήματα (Spath, 1980). Αντίθετα, ο DBSCAN μπορεί να ανιχνεύσει αυθαίρετων σχημάτων συστάδες, όπως καμπύλες ή δακτυλιοειδή σχήματα, αρκεί να ικανοποιείται η συνθήκη της τοπικής πυκνότητας (Ester et al., 1996).

- **Αντοχή σε θόρυβο:** Ο PAM παρουσιάζει μεγαλύτερη ανθεκτικότητα στον θόρυβο από τον K-means, λόγω της χρήσης πραγματικών σημείων (medoids) που δεν επηρεάζονται έντονα από μεμονωμένα outliers (Arora et al., 2016). Παρόλα αυτά, ο PAM δεν απορρίπτει κανένα σημείο – όλα εντάσσονται σε κάποια συστάδα. Έτσι, outliers ενδέχεται να επηρεάσουν τις συστάδες, αν και λιγότερο από ό,τι στον K-means. Σε αντίθεση, ο DBSCAN χαρακτηρίζει ρητά τα σημεία εκτός περιοχών υψηλής πυκνότητας ως θόρυβο και δεν τα εντάσσει σε καμία συστάδα, καθιστώντας τον ιδιαίτερα αποτελεσματικό σε εφαρμογές με ανωμαλίες ή απομονωμένες παρατηρήσεις (Schubert et al., 2017).
- **Υπολογιστική πολυπλοκότητα:** Η υπολογιστική πολυπλοκότητα του PAM είναι υψηλή:  $O(k(n-k)^2)$  ανά επανάληψη, γεγονός που τον καθιστά ανεπαρκή για πολύ μεγάλα σύνολα δεδομένων (Kaufman & Rousseeuw, 1990). Παρόλο που έχουν προταθεί βελτιωμένες εκδοχές όπως οι CLARA και CLARANS για μείωση της πολυπλοκότητας (Ng & Han, 2002), ο DBSCAN τείνει να είναι υπολογιστικά πιο αποδοτικός σε μεγάλα σύνολα, ιδιαίτερα όταν χρησιμοποιούνται ευρετήρια για τα ερωτήματα γειτνίασης, με πολυπλοκότητα κοντά στο  $O(n \log n)$  (Schubert et al., 2017).
- **Δυνατότητα επιλογής μέτρου απόστασης:** Ο PAM υποστηρίζει οποιοδήποτε μέτρο απόστασης ή ομοιότητας, π.χ. Manhattan, Minkowski, ή ακόμη και κατηγορικές αποστάσεις με το Gower ή Hamming, καθιστώντας τον ευέλικτο σε δεδομένα με κατηγορικές ή μικτές μεταβλητές (Kassambara, 2017). Ο DBSCAN, αν και χρησιμοποιείται κυρίως με ευκλείδεια απόσταση, μπορεί και αυτός να λειτουργήσει με προσαρμοσμένες μετρικές, όπως η Haversine απόσταση για γεωχωρικά δεδομένα ή cosine για διανύσματα λέξεων (Pedregosa et al., 2011). Επομένως και οι δύο αλγόριθμοι προσφέρουν προσαρμοστικότητα ως προς τη μετρική ομοιότητας.

Συμπερασματικά, συγκριτικά με τους αλγορίθμους διαμέρισης όπως ο K-means και ο PAM, ο DBSCAN υπερέρχει στα εξής:

- Δεν απαιτεί τον καθορισμό του αριθμού των συστάδων εξ αρχής.
- Μπορεί να εντοπίσει συστάδες οποιουδήποτε σχήματος.
- Διαχειρίζεται αποτελεσματικά θόρυβο και απομονωμένα σημεία.

Ωστόσο, υστερεί ενδεχομένως σε περιβάλλοντα υψηλών διαστάσεων (curse of dimensionality) και απαιτεί προσεκτική παραμετροποίηση των τιμών  $\epsilon$  και minPts. Παρ' όλα αυτά, η ικανότητά του να ανταποκρίνεται σε πολύπλοκες δομές χωρίς σφαιρικές υποθέσεις τον καθιστά ιδιαίτερα κατάλληλο για εφαρμογές σε πραγματικά δεδομένα όπου η υπόθεση συμμετρίας δεν ισχύει (Tan et al., 2018; Jain, 2010). Ο παρακάτω πίνακας συνοψίζει μια συγκριτική θεώρηση:

## Σύγκριση DBSCAN με K-means και PAM

Κριτήριο	DBSCAN	K-means	PAM (K-medoids)
Προκαθορισμένος αριθμός συστάδων;	Όχι	Ναι (πρέπει να οριστεί k)	Ναι (πρέπει να οριστεί k)
Σχήμα συστάδων	Αυθαίρετο (οποιοδήποτε σχήμα)	Σφαιρικό/ισομεγέθες	Συμπαγές γύρω από medoids (τυπικά κοντά σε σφαιρικό) Μέτρα – λιγότερο επηρεαζόμενο από outliers σε σχέση με K-means medium.com αλλά δεν αγνοεί κανένα σημείο
Ευαισθησία σε θόρυβο	Ανθεκτικό – αγνοεί θόρυβο (ετικέτα -1)	Υψηλή – όλα τα σημεία ταξινομούνται, επηρεάζεται από outliers	
Απαιτήσεις παραμέτρων	$\epsilon$ , MinPts (δύσκολος προσδιορισμός)	k (μπορεί να εκτιμηθεί με Elbow method)	k (παρόμοια με K-means)
Πολυπλοκότητα	$O(n^2)$ worst-case, βελτίωση με ευρετήρια σε $\sim O(n \log n)$	$\approx O(n \cdot k \cdot I)$ (συνήθως γραμμικό σε n)	$O(k(n-k)^2)$ (πολύ βαρύ για μεγάλα n)
Κατάλληλο για υψηλές διαστάσεις;	Περιορισμένα (curse of dimensionality)	Μέτρα (εξαρτάται από μέτρο απόστασης, παρόμοιο πρόβλημα)	Μέτρα (υποφέρει και αυτός από υψηλές διαστάσεις)
Δυνατότητα διαφορετικών αποστάσεων;	Ναι (υποστηρίζει custom metric)	Συνήθως ευκλείδειο (υλοποιήσεις μπορεί να επιτρέπουν άλλες αποστάσεις)	Ναι (φυσικά, μπορεί οποιοδήποτε γιατί βασίζεται σε αποστάσεις)
Ανιχνεύει μικρές συστάδες;	Ναι, αν ρυθμιστεί κατάλληλα (μικρό $\epsilon$ , μικρό MinPts)	Όχι αν k είναι μικρό; (θα τις συγχωνεύσει)	Όχι, πρέπει να αυξήσεις k επίτηδες

Η αποδοτική υλοποίηση του DBSCAN και η κλιμάκωσή του σε μεγάλα δεδομένα αποτελεί ένα πρακτικό ζήτημα. Αρχικά, ας εξετάσουμε την πολυπλοκότητα: Στην απλούστερη υλοποίηση, για κάθε σημείο ο αλγόριθμος πρέπει να βρει όλους τους γείτονές του εντός απόστασης  $\epsilon$ . Μια αφελής προσέγγιση θα υπολόγιζε αποστάσεις από αυτό το σημείο προς όλα τα υπόλοιπα σημεία – αυτό είναι  $O(n)$  για ένα σημείο και επαναλαμβανόμενο για  $n$  σημεία δίνει συνολικά πολυπλοκότητα  $O(n^2)$ . Πράγματι, αναφέρεται στη βιβλιογραφία

ότι η χειρότερη περίπτωση για τον DBSCAN είναι  $O(n^2)$  (π.χ. αν όλα τα σημεία καταλήγουν σε μία τεράστια συστάδα, ο αλγόριθμος μπορεί να χρειαστεί να εξετάσει πολλούς γείτονες επαναληπτικά).

Ωστόσο, μια από τις δυνάμεις του DBSCAN είναι ότι μπορεί να υλοποιηθεί πιο αποδοτικά χρησιμοποιώντας δομές ευρετηρίασης χωρικών ερωτημάτων (spatial index), όπως δέντρα  $k$ - $d$ , R-trees ή άλλες δομές γειτνίασης. Οι δομές αυτές επιτρέπουν ερωτήματα εύρεσης γειτόνων σε υπογραμμική πολυπλοκότητα. Για παράδειγμα, ένα  $k$ - $d$  tree μπορεί να βρει όλα τα σημεία εντός μιας δεδομένης ακτίνας  $\epsilon$  συνήθως σε χρόνο  $O(\sqrt{\log n})$  για κάθε σημείο (αυτή είναι μια μέση περίπτωση – σε χειρότερη περίπτωση μπορεί να είναι  $O(n)$ , αλλά συνήθως πολύ καλύτερα). Εάν χρησιμοποιηθεί τέτοια δομή, η εμπειρική πολυπλοκότητα του DBSCAN βελτιώνεται σε περίπου  $O(n \sqrt{\log n})$ , καθιστώντας τον πολύ πιο διαχειρίσιμο σε μεγάλα σύνολα. Πολλές σύγχρονες υλοποιήσεις (όπως της scikit-learn) αξιοποιούν τέτοιες δομές για βελτίωση ταχύτητας.

### Συνθήκες αποδοτικής χρήσης:

- Ο DBSCAN τείνει να είναι αποδοτικός σε δεδομένα με μέτρια διαστασιμότητα (π.χ. μέχρι δεκάδες διαστάσεις). Σε πολύ υψηλές διαστάσεις, το φαινόμενο γνωστό ως “curse of dimensionality” επηρεάζει αρνητικά οποιονδήποτε αλγόριθμο βασισμένο σε αποστάσεις: όλες οι αποστάσεις αρχίζουν να μοιάζουν μεταξύ τους και οι δομές ευρετηρίασης χάνουν την αποτελεσματικότητά τους. Έτσι, σε εξαιρετικά πολυδιάστατα δεδομένα, ο DBSCAN μπορεί να μην είναι αποδοτικός ούτε και αποτελεσματικός στην εύρεση clusters.
- Σε δεδομένα όπου οι πυκνότητες διαφέρουν πολύ (π.χ. μια μικρή αλλά πολύ πυκνή συστάδα και μια μεγάλη αραιότερη συστάδα), ο DBSCAN μπορεί να χρειαστεί διαφορετικές ρυθμίσεις  $\epsilon$  σε διαφορετικές περιοχές – κάτι που δεν υποστηρίζει εγγενώς. Υπάρχει η βελτιωμένη εκδοχή HDBSCAN που αντιμετωπίζει σε έναν βαθμό αυτό το πρόβλημα παράγοντας ιεραρχία συστάδων με μεταβαλλόμενη πυκνότητα, αλλά ο βασικός αλγόριθμος DBSCAN μπορεί σε τέτοιες περιπτώσεις να θεωρηθεί λιγότερο “αποδοτικός” εννοιολογικά, γιατί δεν θα αποδώσει ιδανικά αποτελέσματα με ένα μόνο  $\epsilon$ . Δηλαδή, μπορεί ή να σπάσει τη μεγάλη αραιή συστάδα σε κομμάτια ή να συγχωνεύσει την πυκνή μικρή συστάδα με γειτονικό θόρυβο, αναλόγως της ρύθμισης. Αυτό δεν είναι ακριβώς ζήτημα υπολογιστικής πολυπλοκότητας, αλλά “αποδοτικής χρήσης” με την έννοια της επιτυχημένης εφαρμογής.
- Από πλευράς μνήμης, ο DBSCAN χρειάζεται να αποθηκεύει την γειτονική λίστα ή να μπορεί να επισημάνει/επισκέπτεται τα σημεία. Η πολυπλοκότητα μνήμης είναι  $O(n)$ , που είναι αναμενόμενο για αλγορίθμους που πρέπει να διατηρούν μια ετικέτα για κάθε σημείο.
- Σε πολύ μεγάλα datasets (εκατομμύρια σημεία), έχουν προταθεί παραλλαγές (π.χ. Parallel DBSCAN, MR-DBSCAN για MapReduce) που κατατείνουν στη διαμοιρασμένη εκτέλεση του αλγορίθμου, ώστε να εκμεταλλευτούν πολλαπλούς πυρήνες ή καταναμεμημένα συστήματα. Ο πυρήνας όμως παραμένει ο ίδιος: εύρεση γειτόνων και επέκταση συστάδων.

- Σημειώνεται ότι το πρόβλημα της γειτονικής εύρεσης (range query) είναι το κύριο φορτίο του DBSCAN. Αν αυτό βελτιστοποιηθεί, ο αλγόριθμος είναι αρκετά αποδοτικός. Η εσωτερική δομή της συστάδας δεν απαιτεί βαριές πράξεις (σε αντίθεση, π.χ., με υπολογισμό κόστους ανάκτησης συνεχώς στον K-means).

Συνοψίζοντας, ο DBSCAN σε έναν μέσο όρο περίπτωσης εκτελείται αρκετά γρήγορα, εκμεταλλευόμενος δομές αναζήτησης γειτόνων. Η θεωρητική πολυπλοκότητα  $O(n^2)$  συνήθως μειώνεται πρακτικά κοντά στο  $O(n \log n)$  με σωστή χρήση ευρετηρίων. Φυσικά, ακραίες περιπτώσεις (όλα τα σημεία πολύ κοντά μεταξύ τους ή πολύ υψηλές διαστάσεις) μπορεί να τον φέρουν στο τετράγωνο, αλλά αυτές είναι σπάνιες ή μη ρεαλιστικές για χρήσιμα datasets.

Η ικανότητα του DBSCAN να εντοπίζει συστάδες χωρίς προκαθορισμένο σχήμα και να αντιμετωπίζει τον θόρυβο τού δίνει ένα πλεονέκτημα σε πολλά πεδία. Παρακάτω παρουσιάζονται ορισμένες χαρακτηριστικές εφαρμογές:

**1. Γεωχωρικά Δεδομένα και Αναλύσεις Τοποθεσίας:** Σε προβλήματα γεωγραφικών πληροφοριακών συστημάτων (GIS) και χωροταξικών δεδομένων, ο DBSCAN είναι ιδιαίτερα δημοφιλής. Για παράδειγμα, σκεφτείτε δεδομένα με συντεταγμένες (γεωγραφικό μήκος/πλάτος) συμβάντων όπως σεισμοί, εγκληματικά περιστατικά ή θέσεις καταναλωτών σε μια πόλη. Ο DBSCAN μπορεί να εντοπίσει πυκνές περιοχές σημείων οι οποίες αντιστοιχούν σε *hotspots* στο χώρο, χωρίς να απαιτεί προκαθορισμένα γεωγραφικά όρια (Ester et al., 1996).

- Στον αστικό σχεδιασμό, μπορεί να βρει πυκνοκατοικημένες ή πολυσύχναστες περιοχές (π.χ. εμπορικά κέντρα) αναλύοντας δεδομένα GPS από κινητές συσκευές πολιτών. Αυτό βοηθά σε αποφάσεις όπως η τοποθέτηση υποδομών: ομαδοποιώντας σημεία όπου συγκεντρώνεται ο κόσμος μπορούμε να προτείνουμε σημεία για νέες στάσεις μετρό ή λεωφορείου.
- Στη χαρτογράφηση εγκληματικότητας, δεδομένα θέσεων εγκλημάτων σε μια πόλη μπορούν να συσταδοποιηθούν ώστε να αποκαλυφθούν περιοχές με ιδιαίτερα υψηλή συχνότητα περιστατικών (crime hotspots). Οι αρχές μπορούν να αξιοποιήσουν αυτή την πληροφορία για στοχευμένες περιπολίες ή παρεμβάσεις.
- Σε περιβαλλοντικά δεδομένα, ο DBSCAN μπορεί να βρει συστάδες ρύπανσης ή αισθητήρων που καταγράφουν ακραίες μετρήσεις. Για παράδειγμα, σημεία με υψηλή μόλυνση σε έναν χάρτη μπορούν να εντοπιστούν ως συστάδες σε ένα νέφος σημείων χαμηλότερης μόλυνσης.

Το πλεονέκτημα είναι ότι δεν χρειάζεται να γνωρίζουμε πόσες περιοχές ενδιαφέροντος υπάρχουν εκ των προτέρων – ο DBSCAN αποκαλύπτει εγγενώς αυτές τις πυκνές περιοχές και μας δίνει και τα σημεία που είναι εκτός ως θόρυβο (τα οποία μπορεί να αγνοηθούν ως μεμονωμένα περιστατικά).

**2. Αναγνώριση Προτύπων σε Δεδομένα & Υπολογιστική Όραση:** Η αναγνώριση προτύπων περιλαμβάνει την ανίχνευση υποκείμενων δομών ή ομάδων σε δεδομένα, συχνά υψηλών διαστάσεων (Abiodun et. al., 2019). Ο DBSCAN έχει βρει θέση σε διάφορες εφαρμογές:

- Στην υπολογιστική όραση, π.χ. επεξεργασία εικόνων, μπορεί να χρησιμοποιηθεί για τμηματοποίηση εικόνας (image segmentation) βάσει πυκνότητας pixel ή χαρακτηριστικών. Αν θεωρήσουμε μια εικόνα ως σύνολο σημείων σε χώρο χρώματος ή σε χώρο χαρακτηριστικών, ο DBSCAN μπορεί να ομαδοποιήσει pixel που σχηματίζουν ένα αντικείμενο. Σε ιατρικές εικόνες, για παράδειγμα, έχει χρησιμοποιηθεί για να απομονώσει όγκους ή ανωμαλίες ως πυκνές περιοχές pixel που ξεχωρίζουν από το γύρω υγιές υπόβαθρο. Επειδή οι όγκοι μπορεί να έχουν ακανόνιστα περιγράμματα, μια μέθοδος σαν τον K-means θα ήταν αδέξια, ενώ ο DBSCAN συλλαμβάνει το ακριβές σχήμα της περιοχής ενδιαφέροντος και αφήνει έξω τα διάσπαρτα pixel θορύβου.
- Στην ανάλυση δεδομένων αισθητήρων (π.χ. IoT), υποθέστε πλήθος αισθητήρων που παράγουν δεδομένα (θερμοκρασία, πίεση, κ.λπ.). Ο DBSCAN μπορεί να αναγνωρίσει πρότυπα όπως ομάδες αισθητήρων με παρόμοιες αναγνώσεις που αντανακλούν μια κοινή κατάσταση. Επίσης, σε χρονοσειρές (αν και όχι απευθείας κατάλληλος για χρονική συσταδοποίηση χωρίς μετασχηματισμό), έχουν υπάρξει επεκτάσεις του DBSCAN για εύρεση προτύπων σε πολυδιάστατες χρονοσειρές μετατρέποντάς τις σε σημεία σε χώρο χαρακτηριστικών (feature space) και βρίσκοντας συστάδες (π.χ. ομαδοποίηση περιόδων με παρόμοια συμπεριφορά).
- Σε σύνολα δεδομένων χαρακτηριστικών για αναγνώριση προτύπων, όπως η ομαδοποίηση πελατών με βάση συμπεριφορές ή χαρακτηριστικά, ο DBSCAN μπορεί να αποκαλύψει ομάδες πελατών με παρόμοια μοτίβα, ενώ αφήνει στην άκρη περιπτώσεις πελατών που δεν ανήκουν σε καμία ομάδα (δηλ. “έξωτικοί” πελάτες) χωρίς να τους εντάξει με το ζόρι σε μια κατηγορία. Αυτό σχετίζεται βέβαια και με την παρακάτω κατηγορία (ανίχνευση ανωμαλιών), καθώς ο απομονωμένος πελάτης θα μπορούσε να θεωρηθεί ανωμαλία.

**3. Ανίχνευση Ανωμαλιών (Anomaly/Outlier Detection):** Ένας τομέας όπου ο DBSCAN χρησιμοποιείται κατά κόρον είναι η ανίχνευση ανωμαλιών. Ουσιαστικά, η ιδιότητά του να χαρακτηρίζει κάποια σημεία ως θόρυβο είναι απευθείας μια ένδειξη ότι αυτά τα σημεία είναι ανωμαλικά σε σχέση με τα υπόλοιπα (Schubert et al., 2017). Μερικές εφαρμογές:

- **Ασφάλεια Δικτύων:** Στην ανίχνευση εισβολών, κάθε καταγραφή δικτυακής κίνησης μπορεί να ειπωθεί ως ένα σημείο σε χώρο χαρακτηριστικών (π.χ. αριθμός πακέτων, bytes, πόρτες επικοινωνίας, κ.λπ.). Ο DBSCAN μπορεί να βρει συστάδες “φυσιολογικής” κίνησης, ενώ τυχόν ύποπτα μοτίβα που δεν ταιριάζουν σε καμία συστάδα θα εμφανιστούν ως θόρυβος. Έτσι, οι πιθανές επιθέσεις ή ανωμαλίες στην κίνηση του δικτύου απομονώνονται για περαιτέρω διερεύνηση. Αυτό λειτουργεί και σε συνδυασμό με άλλες τεχνικές, αλλά ο DBSCAN αποτελεί ένα απλό φίλτρο outlier.

- **Ανίχνευση Απάτης στα Οικονομικά:** Σε δεδομένα συναλλαγών, ο DBSCAN μπορεί να ομαδοποιήσει τις συνηθισμένες συμπεριφορές πελατών (π.χ. αγορές συγκεκριμένης συχνότητας και ποσών) και να αναδείξει ως θόρυβο συναλλαγές που δεν ανήκουν σε καμία κανονική ομάδα. Αυτές οι μεμονωμένες συναλλαγές – για παράδειγμα μια ασυνήθιστα μεγάλη αγορά ή μια αγορά σε διαφορετική χώρα – θα μπορούσαν να αποτελούν περιπτώσεις απάτης και να αξίζουν διερεύνησης. Ο DBSCAN βοηθά επειδή δεν αναγκάζει αυτές τις συναλλαγές να μπουν σε ένα “κανονικό” cluster (όπως θα έκανε π.χ. ο K-means), αλλά τις εμφανίζει ξεκάθαρα ως outliers.
- **Βιομηχανική παρακολούθηση και IoT:** Σε ένα εργοστάσιο με πολλούς αισθητήρες, οι φυσιολογικές λειτουργίες θα παράγουν συστάδες σημείων (συνδυασμοί μετρήσεων αισθητήρων που συμβαίνουν συχνά). Αν κάποιο μηχάνημα αρχίσει να δυσλειτουργεί, οι μετρήσεις του μπορεί να βγουν εκτός αυτών των συστάδων και να εμφανιστούν ως μεμονωμένα σημεία. Έτσι, ο DBSCAN μπορεί να χρησιμεύσει στην προγνωστική συντήρηση: να ειδοποιήσει ότι “υπάρχουν σημεία μετρήσεων εκτός των γνωστών clusters λειτουργίας”, που ισοδυναμεί με ανωμαλία εξοπλισμού.

Σε όλες αυτές τις περιπτώσεις, το ισχυρό στοιχείο είναι ότι ο DBSCAN δεν χρειάζεται ρητά να εκπαιδευτεί τι είναι “ανωμαλία” – μαθαίνει τα κανονικά μοτίβα (clusters) και ό,τι δεν ανήκει σε κάποιο από αυτά θεωρείται ανωμαλικό. Φυσικά, απαιτείται σωστός χειρισμός των παραμέτρων  $\epsilon$  και  $MinPts$  ώστε τα “κανονικά” δεδομένα να σχηματίσουν συστάδες και τα “μη κανονικά” να μείνουν εκτός.

**4. Άλλες Εφαρμογές:** Πέραν των παραπάνω, ο DBSCAN έχει χρησιμοποιηθεί σε πάρα πολλούς τομείς, όπως:

- **Βιοπληροφορική:** π.χ. συσταδοποίηση γονιδιακής έκφρασης. Γονίδια με παρόμοια επίπεδα έκφρασης σε διάφορες συνθήκες μπορεί να ομαδοποιηθούν. Ο DBSCAN βοηθά διότι αγνοεί γονίδια που δεν ανήκουν σε καμία ομάδα (θόρυβος) και μπορεί να βρει ομάδες γονιδίων που δρουν από κοινού ακόμη κι αν το μέγεθός τους ποικίλει (Kaur & Wasan, 2006).
- **Ανάλυση Αγοράς/Marketing:** Όπως αναφέρθηκε, η συσταδοποίηση πελατών ή προϊόντων. Ο DBSCAN έχει χρησιμοποιηθεί για τμηματοποίηση αγοράς όπου δεν είναι σαφής ο αριθμός τμημάτων εκ των προτέρων. Μπορεί να προκύψουν π.χ. 3 μεγάλα clusters πελατών και μερικοί πελάτες να μείνουν εκτός (θόρυβος) διότι η συμπεριφορά τους δεν μοιάζει με των υπολοίπων – αυτοί μπορεί να αντιμετωπιστούν ξεχωριστά σε στοχευμένες καμπάνιες (Ester et al., 199);
- **Αστρονομία:** Στην ανάλυση κατανομών άστρων ή γαλαξιών, η κατανομή ύλης στο σύμπαν παρουσιάζει πυκνότητες και κενά. Ο DBSCAN έχει εφαρμοστεί για να βρει σμήνη γαλαξιών ή άστρων ως πυκνές συγκεντρώσεις στο διάστημα (Schubert et al., 2012). Επίσης στην ανακάλυψη εξωπλανητών, μπορεί να εντοπίσει ομάδες δεδομένων

από τηλεσκοπία που υποδηλώνουν παρουσία πλανητών, αφήνοντας έξω μεμονωμένα ψευδοστίγματα.

Κάθε φορά, η ικανότητα του DBSCAN να μορφοποιεί το σχήμα της συστάδας πάνω στα δεδομένα (αντί να επιβάλλει σχήμα) και να παραμερίζει τον θόρυβο, είναι αυτό που προσφέρει την ιδιαίτερη αξία του. Φυσικά, αν τα δεδομένα δεν έχουν εμφανή δομή πυκνότητας (π.χ. είναι περίπου ομοιόμορφα κατανεμημένα), ο DBSCAN δεν θα βρει κάτι ουσιαστικό – στην ακραία περίπτωση θα θεωρήσει τα πάντα μια συστάδα ή όλα θόρυβο αναλόγως τις παραμέτρους. Γι' αυτό απαιτείται προσοχή: το ότι ο DBSCAN βρίσκει “κάτι” δεν σημαίνει πως αυτό είναι πάντα σημασιολογικά χρήσιμο· πρέπει να υπάρχει πραγματική πυκνότητα/δομή στα δεδομένα.

## Παράδειγμα Υλοποίησης σε Python (scikit-learn)

Για να δείξουμε πώς χρησιμοποιείται πρακτικά ο DBSCAN, δίνουμε ένα παράδειγμα κώδικα σε Python κάνοντας χρήση της βιβλιοθήκης **scikit-learn**. Το παράδειγμα περιλαμβάνει: δημιουργία ενός απλού συνόλου δεδομένων, εφαρμογή του DBSCAN και οπτικοποίηση των αποτελεσμάτων.

```
# Εισαγωγή αναγκαίων βιβλιοθηκών
import numpy as np
from sklearn.cluster import DBSCAN
from sklearn.datasets import make_blobs
import matplotlib.pyplot as plt

# 1. Δημιουργία συνθετικού συνόλου δεδομένων
# Θα δημιουργήσουμε 3 clusters σημείων σε 2 διαστάσεις και θα προσθέσουμε και
μερικά σημεία θορύβου.
X, y_true = make_blobs(n_samples=300, centers=[(0,0), (5,5), (0,5)],
                        cluster_std=0.5, random_state=0)
# Προσθήκη 5 τυχαίων "θορυβωδών" σημείων εκτός των clusters
np.random.seed(1)
noise_points = np.random.uniform(low=-2, high=7, size=(5, 2))
X = np.vstack([X, noise_points])

# 2. Εφαρμογή DBSCAN
eps_value = 0.8          # απόσταση ακτίνας
minpts_value = 5         # ελάχιστοι γείτονες
dbscan = DBSCAN(eps=eps_value, min_samples=minpts_value)
labels = dbscan.fit_predict(X) # υπολογίζει τα clusters και επιστρέφει τις
ετικέτες (-1 για θόρυβο)

# 3. Έλεγχος των αποτελεσμάτων
num_clusters = len(set(labels) - {-1}) # πλήθος clusters (εξαιρώντας το -1)
print(f"Βρέθηκαν συνολικά {num_clusters} συστάδες.")
print(f"Ετικέτες συστάδων για κάθε σημείο: {labels}")

# 4. Οπτικοποίηση των συστάδων
plt.figure(figsize=(6, 5))
unique_labels = set(labels)
colors = ['red', 'green', 'blue', 'purple', 'orange'] # χρώματα για ως 5
clusters
for label in unique_labels:
    if label == -1:
```

```

        # Σημεία θορύβου με μαύρο 'x'
        plt.scatter(X[labels == label, 0], X[labels == label, 1], c='k',
marker='x', label='θόρυβος')
    else:
        plt.scatter(X[labels == label, 0], X[labels == label, 1],
                    color=colors[label % len(colors)], label=f'Cluster {label}')
plt.title(f"DBSCAN αποτέλεσμα: {num_clusters} clusters")
plt.legend()
plt.show()

```

### Επεξήγηση κώδικα:

- Χρησιμοποιούμε τη `make_blobs` για να δημιουργήσουμε 3 ομάδες σημείων (blobs) σε διαφορετικές θέσεις, προσθέτοντας λίγη τυχειότητα (θόρυβο) μέσα σε καθεμία. Έπειτα προσθέτουμε χειροκίνητα 5 σημεία-θόρυβο εκτός των blob περιοχών.
- Αρχικοποιούμε το DBSCAN με `eps=0.8` και `min_samples=5`. Αυτές οι τιμές επιλέχθηκαν εμπειρικά: ξέρουμε ότι μέσα σε κάθε blob τα σημεία έχουν τυπική απόκλιση 0.5, οπότε μια ακτίνα 0.8 λογικά θα συμπεριλάβει γειτονικά σημεία εντός κάθε πυκνής ομάδας. Το `min_samples=5` ορίζει ότι χρειαζόμαστε τουλάχιστον 5 σημεία για να θεωρήσουμε μια περιοχή πυκνή (σε 2 διαστάσεις, 5 είναι μια συνηθισμένη τιμή εκκίνησης).
- Καλούμε τη `fit_predict` που υπολογίζει τις συστάδες και επιστρέφει έναν πίνακα `labels` με τον αριθμό συστάδας για κάθε σημείο (ή -1 αν το σημείο χαρακτηρίστηκε ως θόρυβος). Εκτυπώνουμε κάποιες βασικές πληροφορίες: τον αριθμό συστάδων που βρέθηκαν και τις ετικέτες για κάθε σημείο (μπορεί να μη τις δίνουμε όλες σε πραγματικά μεγάλο dataset, εδώ είναι επίδειξη).
- Τέλος, σχεδιάζουμε τα σημεία με διαφορετικά χρώματα ανά cluster. Όσα έχουν `label = -1` (θόρυβος) τα σχεδιάζουμε με μαύρο χρώμα και το σύμβολο 'x' για να ξεχωρίζουν. Αναμένουμε ότι ο DBSCAN θα πρέπει να βρει περίπου 3 clusters που αντιστοιχούν στα 3 blobs που φτιάξαμε και να αφήσει εκτός (ως θόρυβο) τα 5 διάσπαρτα σημεία.

Αν εκτελεστεί ο παραπάνω κώδικας, ένα πιθανό αποτέλεσμα είναι:

```

less
ΑντιγραφήΕπεξεργασία
Βρέθηκαν συνολικά 3 συστάδες.
Ετικέτες συστάδων για κάθε σημείο: [ 0  1  2  0  1  2 ... -1 ...]

```

Δηλαδή, τρεις συστάδες με ετικέτες 0, 1, 2 και μερικά σημεία -1. Το διάγραμμα θα δείξει τρεις καθαρές ομάδες σημείων χρώματος κόκκινου, πράσινου, μπλε (για παράδειγμα) και τα σημεία θορύβου ως μαύρα x διάσπαρτα μακριά τους. Αυτό επαληθεύει ότι ο DBSCAN κατάφερε να ξαναβρεί τα clusters που δημιουργήσαμε και να απομονώσει τα outliers, με τις κατάλληλες ρυθμίσεις παραμέτρων.

*Σημείωση:* Στον πραγματικό κόσμο, θα δοκιμάζαμε διάφορες τιμές  $\epsilon$ ps ή θα χρησιμοποιούσαμε το διάγραμμα απόστασης όπως περιγράφηκε, προκειμένου να βελτιστοποιήσουμε αυτές τις παραμέτρους. Επίσης, ίσως χρειαζόταν να δοκιμάσουμε διαφορετικά κανονικοποιημένα δεδομένα ή μετρικές απόστασης, ανάλογα με το πρόβλημα.

## Συμπεράσματα

Ο αλγόριθμος DBSCAN αποτελεί ένα ισχυρό εργαλείο συσταδοποίησης που πλεονεκτεί σε περιπτώσεις όπου οι συστάδες δεν είναι καλά διαχωρισμένες, έχουν ακανόνιστα σχήματα ή περιέχουν θορυβώδη σημεία. Στην παρούσα ανάλυση, εξετάσαμε λεπτομερώς τη θεωρητική του βάση: ορισμοί όπως η πυκνωτική προσέγγιση και συνδεσιμότητα και είδαμε πώς αυτά υλοποιούνται στα βήματα του αλγορίθμου. Οι παράμετροι  $\epsilon$  και  $MinPts$  αναδείχθηκαν ως κρίσιμες – αν και η επιλογή τους δεν είναι πάντα προφανής, υπάρχουν πρακτικές μέθοδοι (όπως τα διαγράμματα  $k$ -απόστασης) για να καθοδηγήσουν την ρύθμισή τους.

Σε σύγκριση με κλασικούς αλγορίθμους όπως ο K-means, ο DBSCAN προσφέρει αυτονομία ως προς τον αριθμό συστάδων και ανθεκτικότητα σε θόρυβο, ενώ υπερέρχει στον εντοπισμό συστάδων πολύπλοκων μορφών. Από την άλλη, έχει και αυτός τις αδυναμίες του: η απόδοσή του μπορεί να μειωθεί σε υψηλοδιάστατα δεδομένα και για να δουλέψει σωστά προϋποθέτει ότι οι συστάδες έχουν σημαντική διαφορά πυκνότητας από το περιβάλλον τους (σε περιπτώσεις μεταβαλλόμενης πυκνότητας, ίσως απαιτούνται πιο εξελιγμένες παραλλαγές όπως ο HDBSCAN). Υπολογιστικά, με τη βοήθεια χωρικών ευρετηρίων ο DBSCAN μπορεί να κλιμακωθεί σχετικά καλά, αν και στη χειρότερη θεωρητική περίπτωση είναι τετραγωνικός.

Οι εφαρμογές του DBSCAN που παρουσιάστηκαν – από γεωχωρική ανάλυση μέχρι ανίχνευση απάτων – δείχνουν την ευελιξία του: λειτουργεί ως “ανιχνευτής” δομής σε δεδομένα όπου άλλα μέτρα αποτυγχάνουν ή απαιτούν περισσότερη επίβλεψη. Συγκεκριμένα, η ικανότητα να ξεχωρίζει τις ανωμαλίες αντί να τις κρύβει μέσα σε συστάδες είναι καθοριστική για πολλά πρακτικά προβλήματα.

Συμπερασματικά, ο DBSCAN είναι ένα απαραίτητο εργαλείο στη φαρέτρα του ερευνητή δεδομένων και του επιστήμονα μηχανικής μάθησης όταν πρόκειται για ανεπιτήρητη μάθηση. Καμία μέθοδος clustering δεν είναι πανάκεια – ο DBSCAN δεν αποτελεί εξαίρεση – αλλά όταν οι προϋποθέσεις ευνοούν τη χρήση του (διακριτές πυκνές περιοχές, ύπαρξη θορύβου, άγνωστος αριθμός clusters), τότε μπορεί να αποκαλύψει γνώσεις από τα δεδομένα που δύσκολα θα αναδύονταν με άλλους τρόπους. Η κατανόηση του τρόπου λειτουργίας του, όπως επιχειρήθηκε σε αυτό το κείμενο, βοηθά στη σωστή εφαρμογή του και στη μέγιστη αξιοποίηση των δυνατοτήτων του.

## 4.6 Νευρωνικά Δίκτυα και Βαθιά Μάθηση (Deep Learning)

### 4.6.1 Τεχνητά Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα (ΤΝΔ – Artificial Neural Networks, ANN), αποτελούν λογισμικά υπολογιστικά μοντέλα που είναι εμπνευσμένα από τη λειτουργία του ανθρώπινου εγκεφάλου και χρησιμοποιούνται ευρέως στο πεδίο της μηχανικής μάθησης και της τεχνητής νοημοσύνης. Αν και δεν αναπαριστούν πιστά τη βιολογική πολυπλοκότητα του ανθρώπινου εγκεφάλου, έχουν σχεδιαστεί με βάση την αρχή ότι οι νευρώνες και οι μεταξύ τους συνδέσεις αποτελούν τα θεμέλια της νοητικής επεξεργασίας (Haykin, 2009). Στόχος των ΤΝΔ είναι να προσεγγίσουν λειτουργικά τον τρόπο με τον οποίο τα βιολογικά νευρωνικά δίκτυα επεξεργάζονται πληροφορίες και λαμβάνουν αποφάσεις (McCulloch & Pitts, 1943).

Ο ανθρώπινος εγκέφαλος επιδεικνύει εξαιρετικές δυνατότητες επεξεργασίας σύνθετων πληροφοριών σε πραγματικό χρόνο. Δραστηριότητες όπως η αναγνώριση φωνής, η κατανόηση γλώσσας και η αναγνώριση προσώπων εκτελούνται με μεγάλη ακρίβεια και ταχύτητα, συχνά χωρίς συνειδητή προσπάθεια. Το ερώτημα που τίθεται είναι κατά πόσο οι υπολογιστές μπορούν να αναπαράγουν αυτές τις λειτουργίες με την ίδια αποδοτικότητα. Οι διαφορές στη φυσιολογική και λειτουργική δομή μεταξύ υπολογιστών και ανθρώπινου εγκεφάλου είναι αξιοσημείωτες, καθιστώντας δύσκολη την άμεση αντιστοίχιση των ανθρώπινων γνωστικών ικανοτήτων με τις υπολογιστικές δυνατότητες (Churchland & Sejnowski, 1992).

Παρά τη ραγδαία εξέλιξη των υπολογιστικών συστημάτων, η αντιμετώπιση "απλών" για τον άνθρωπο εργασιών, όπως η αναγνώριση ομιλίας ή χειρονομιών, εξακολουθεί να αποτελεί πρόκληση για τους υπολογιστές. Το γεγονός αυτό οδήγησε στην ιδέα της μίμησης της νευρωνικής αρχιτεκτονικής του εγκεφάλου μέσω τεχνητών μοντέλων. Έτσι, τα ANN σχεδιάζονται ως προσεγγιστικά μοντέλα που επιχειρούν να ενσωματώσουν τον παραλληλισμό, τη συνδεσιμότητα και την προσαρμοστικότητα, βασικά χαρακτηριστικά των βιολογικών νευρωνικών συστημάτων (Russell & Norvig, 2020).

Η κατασκευή ενός τεχνητού νευρωνικού δικτύου περιλαμβάνει την ανάπτυξη ενός συστήματος από τεχνητούς νευρώνες (τεχνητές μονάδες υπολογισμού) που οργανώνονται σε στρώματα. Αυτοί οι νευρώνες είναι συνδεδεμένοι μεταξύ τους με βαθμονομημένα βάρη (weights) και εκτελούν απλές υπολογιστικές λειτουργίες. Με τη διαδοχική επεξεργασία της πληροφορίας μέσω πολλών στρωμάτων, τα ΤΝΔ καταφέρνουν να "μάθουν" και να γενικεύσουν μοτίβα που υπάρχουν στα δεδομένα εισόδου (Goodfellow et. al., 2016). Η εκπαίδευση αυτών των μοντέλων βασίζεται σε αλγορίθμους βελτιστοποίησης, όπως η οπισθοδιάδοση (backpropagation) και τεχνικές βελτιστοποίησης, όπως ο καθορισμός βαθμίδωσης (gradient descent).

Μία από τις θεμελιώδεις διαφορές μεταξύ των βιολογικών και των τεχνητών νευρωνικών δικτύων είναι η μέθοδος μάθησης. Ενώ τα ANN βασίζονται σε μαθηματικές προσεγγίσεις και προγραμματισμένους αλγόριθμους, τα βιολογικά συστήματα διαθέτουν πλαστικότητα, δηλαδή ικανότητα αναδιοργάνωσης και επανασύνδεσης των νευρωνικών κυκλωμάτων, φαινόμενο που δύσκολα προσομοιώνεται τεχνητά (LeCun, Bengio, & Hinton, 2015). Παράλληλα, τα ΤΝΔ χρειάζονται μεγάλα σύνολα δεδομένων και επαναληπτική εκπαίδευση για να επιτύχουν παρόμοια αποτελέσματα.

Η πρόοδος στην έρευνα των ΤΝΔ τα τελευταία τριάντα χρόνια είναι εντυπωσιακή, με σταδιακή μετάβαση από τα απλά μονοστρωματικά δίκτυα σε βαθιά νευρωνικά δίκτυα (deep neural networks – DNNs). Τα δίκτυα αυτά, χάρη στην αύξηση της υπολογιστικής ισχύος και τη διαθεσιμότητα μεγάλων συνόλων δεδομένων, έχουν σημειώσει σημαντικές επιτυχίες σε ποικίλα πεδία, όπως η αναγνώριση εικόνων, η αυτόματη μετάφραση και η ιατρική διάγνωση (Schmidhuber, 2015). Ωστόσο, η διαφορά στην πολυπλοκότητα μεταξύ των τεχνητών και των βιολογικών νευρωνικών συστημάτων παραμένει χαώδης: ο ανθρώπινος εγκέφαλος περιλαμβάνει περίπου 86 δισεκατομμύρια νευρώνες, με τρισεκατομμύρια συνάψεις, ενώ ακόμη και τα πιο προχωρημένα ANN αριθμούν λίγα εκατομμύρια τεχνητούς νευρώνες.

Παρά τους περιορισμούς, τα ANN εξακολουθούν να αντιπροσωπεύουν μια υποσχόμενη τεχνολογική κατεύθυνση, κυρίως για την ανάπτυξη ευφύων συστημάτων που μπορούν να προσαρμόζονται, να μαθαίνουν από δεδομένα και να αυτοβελτιώνονται με την πάροδο του χρόνου. Το ενδιαφέρον της επιστημονικής κοινότητας παραμένει έντονο, με έρευνες να επικεντρώνονται σε νέες αρχιτεκτονικές, όπως τα convolutional και recurrent networks, αλλά και στην ανάπτυξη νευρωνικών μοντέλων με βιολογικά ρεαλιστικά χαρακτηριστικά (Hassabis et al., 2017).

Ο βασικός στόχος της λειτουργίας ενός Τεχνητού Νευρωνικού Δικτύου (ΤΝΔ) είναι η ικανότητά του να εκτελεί αυτόματα συγκεκριμένες εργασίες, όπως για παράδειγμα η αναγνώριση εικόνων, μέσω της διαδικασίας της μάθησης από δεδομένα. Τα ΤΝΔ βασίζονται στην προσαρμογή της εσωτερικής τους δομής ώστε να ανταποκρίνονται σε εισόδους με τρόπο όμοιο με τον ανθρώπινο εγκέφαλο (Haykin, 2009). Ουσιαστικά, πρόκειται για προγράμματα που μαθαίνουν να «προβλέπουν» ή να «αναγνωρίζουν» πρότυπα, χωρίς να απαιτείται η διατύπωση προκαθορισμένων κανόνων ή μαθηματικών μοντέλων.

Κατά τη φάση εκπαίδευσης, το δίκτυο μαθαίνει μέσα από παραδείγματα. Κάθε παράδειγμα αποτελείται από ένα ζεύγος εισόδου-επιθυμητής εξόδου. Το ΤΝΔ προσαρμόζει τις συνδέσεις (συντελεστές ή βάρη) μεταξύ των τεχνητών νευρώνων έτσι ώστε να ελαχιστοποιήσει τη διαφορά μεταξύ της προβλεπόμενης και της πραγματικής εξόδου. Η εκπαίδευση αυτή βασίζεται σε αλγόριθμους όπως η οπισθοδιάδοση σφάλματος (backpropagation) και ολοκληρώνεται όταν το συνολικό σφάλμα μειωθεί κάτω από ένα αποδεκτό όριο (Goodfellow, Bengio, & Courville, 2016). Το δίκτυο, με αυτόν τον τρόπο, αποκτά την ικανότητα γενίκευσης, δηλαδή να προβλέπει σωστά ακόμη και για είσοδοι που δεν έχει "δει" ακριβώς.

Μετά την εκπαίδευση, το ΤΝΔ περνά στη φάση ελέγχου, κατά την οποία αξιολογείται η αποδοτικότητα του με τη χρήση νέων, άγνωστων παραδειγμάτων. Τα δεδομένα ελέγχου πρέπει να έχουν χαρακτηριστικά παρόμοια με εκείνα των εκπαιδευτικών παραδειγμάτων ώστε να μπορεί να εκτιμηθεί με ακρίβεια η ικανότητα του δικτύου να γενικεύει. Η επιτυχία σε αυτή τη φάση δείχνει την ικανότητα του ΤΝΔ να εφαρμόζει τις γνώσεις του σε καταστάσεις πραγματικού κόσμου, που δεν είναι ίδιες με εκείνες που χρησιμοποιήθηκαν για την εκπαίδευση (Russell & Norvig, 2020).

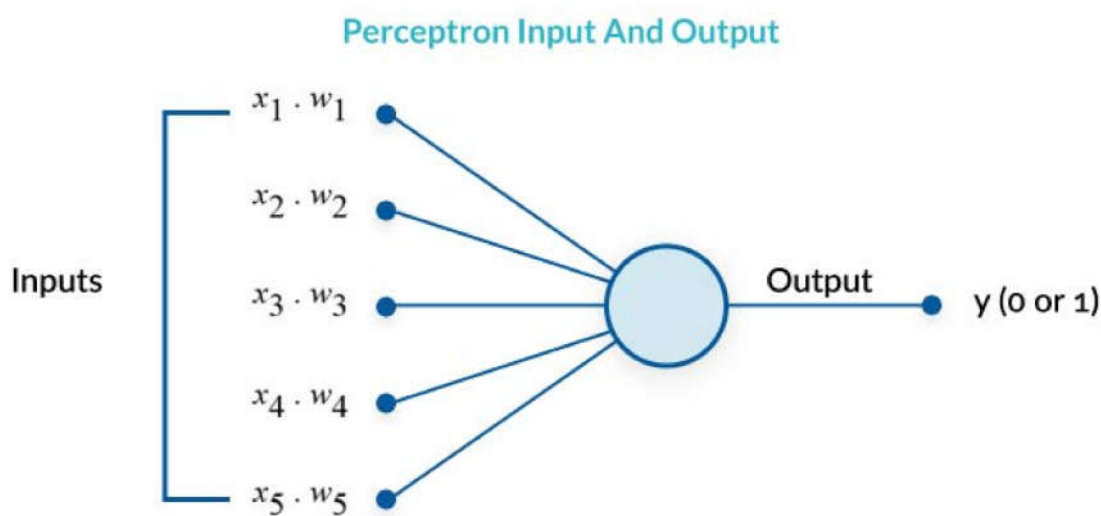
Ένα βασικό χαρακτηριστικό των ΤΝΔ είναι ότι δεν απαιτείται ανθρώπινη γνώση για τα ειδικά χαρακτηριστικά των κατηγοριών που αναγνωρίζουν. Αντί να βασίζονται σε χειροκίνητα ορισμένα χαρακτηριστικά ή κανόνες, τα ΤΝΔ «μαθαίνουν» τα χαρακτηριστικά αυτά από τα ίδια τα δεδομένα, ενδογενώς, μέσα από τα βάρη των συνδέσεων τους. Για παράδειγμα, για να αναγνωρίσει αν μια εικόνα περιέχει έναν σκύλο, το ΤΝΔ δεν χρειάζεται να γνωρίζει προκαθορισμένα κριτήρια (όπως μήκος αυτιών ή σχήμα σώματος). Αντίθετα, μέσα από την εκπαίδευση, δημιουργεί εσωτερικές αναπαραστάσεις των χαρακτηριστικών που είναι πιο σχετικές με την έννοια «σκύλος» (LeCun, Bengio, & Hinton, 2015).

Αυτή η ικανότητα εκμάθησης χωρίς εξωτερικό προγραμματισμό καθιστά τα ΤΝΔ ιδιαίτερα ισχυρά εργαλεία για εφαρμογές όπως η ταξινόμηση εικόνων, η αναγνώριση προτύπων, η πρόβλεψη τιμών, η μετάφραση φυσικής γλώσσας και πολλά ακόμη. Ωστόσο, η «μαύρη φύση» των μοντέλων αυτών – δηλαδή το γεγονός ότι δεν μπορούμε εύκολα να εξηγήσουμε πώς έφτασαν στο τελικό τους αποτέλεσμα – παραμένει ένα σημαντικό ζήτημα (Doshi-Velez & Kim, 2017).

Η ανάπτυξη των ΤΝΔ ξεκίνησε κατά το δεύτερο μισό του 20ού αιώνα, αρχικά με απλά μοντέλα όπως το Perceptron και αργότερα με πιο σύνθετες αρχιτεκτονικές όπως τα πολυστρωματικά perceptrons (MLP). Από τη δεκαετία του 1980 και εξής, η αυξανόμενη υπολογιστική ισχύς και η ανάπτυξη μεγάλων συνόλων δεδομένων επέτρεψαν τη ραγδαία πρόοδο του πεδίου. Τα τελευταία χρόνια, τα λεγόμενα βαθιά νευρωνικά δίκτυα (deep neural networks) έχουν φέρει επανάσταση στην τεχνητή νοημοσύνη (Schmidhuber, 2015). Ωστόσο, τα ΤΝΔ δεν είναι κατάλληλα για όλους τους τύπους προβλημάτων. Παρά την υψηλή τους απόδοση σε προβλήματα γενίκευσης, όπως η κατηγοριοποίηση, η αναγνώριση προτύπων και η πρόβλεψη τάσεων, παρουσιάζουν αδυναμίες σε προβλήματα που απαιτούν συμβολική λογική, εξηγησιμότητα ή ακριβή αριθμητική επεξεργασία. Ειδικά όταν το μέγεθος των δικτύων μεγαλώνει, αυξάνονται οι απαιτήσεις σε υπολογιστική ισχύ και δεδομένα, ενώ καθίσταται πιο δύσκολη η παρακολούθηση της συμπεριφοράς του μοντέλου (Alpaydin, 2020).

### 4.6.2 Κυριότερες κατηγορίες ANN

**Perceptron:** Το Perceptron αναπτύχθηκε από τον Frank Rosenblatt στο Πανεπιστήμιο Cornell το 1958 ως ένας απλός αλλά ισχυρός αλγόριθμος δυαδικής ταξινόμησης. Σκοπός του ήταν η ταξινόμηση εισόδων σε δύο κατηγορίες, «ναι» ή «όχι», με βάση ένα σύνολο γραμμικά διαχωρίσιμων δεδομένων. Η βασική ιδέα πίσω από το perceptron ήταν να μιμηθεί τη λειτουργία ενός βιολογικού νευρώνα, ενσωματώνοντας την ικανότητα προσαρμογής μέσω εκπαίδευσης. Το perceptron αποτέλεσε ένα από τα πρώτα μοντέλα τεχνητής νοημοσύνης που προσέγγισαν το πρόβλημα της μάθησης με συστηματικό τρόπο (Rosenblatt, 1958).



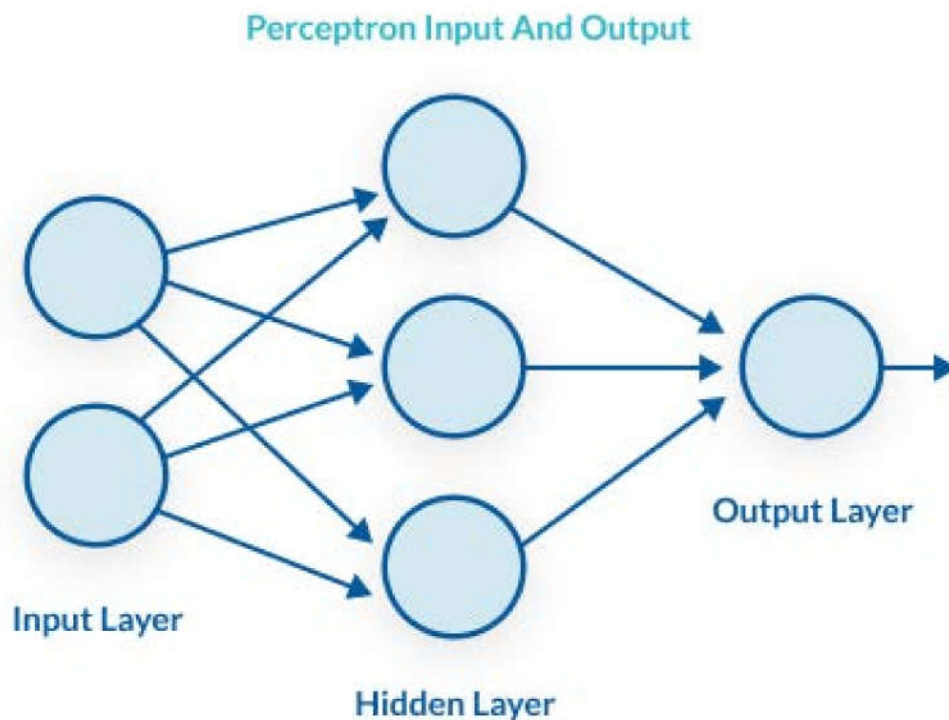
*Απλό παράδειγμα ΤΝΔ τύπου Perceptron*

Ένα απλό perceptron λαμβάνει ως είσοδο ένα διάνυσμα εισόδων  $x_1, x_2, \dots, x_n$  και αντιστοιχίζει σε κάθε είσοδο ένα βάρος  $w_1, w_2, \dots, w_n$ . Υπολογίζει το σταθμισμένο άθροισμα των εισόδων, εφαρμόζει μια συνάρτηση ενεργοποίησης (συνήθως βηματική) και παράγει μια δυαδική έξοδο – 0 ή 1. Η έξοδος αυτή αντιστοιχεί στην πρόβλεψη του μοντέλου για την κατηγορία στην οποία ανήκει το δείγμα. Παρότι απλό, το perceptron έχει περιορισμούς – δεν μπορεί να λύσει προβλήματα που δεν είναι γραμμικά διαχωρίσιμα, όπως το πρόβλημα XOR (Minsky & Papert, 1969).

Η σύνθεση πολλών perceptrons οδηγεί στη δημιουργία ενός τεχνητού νευρωνικού δικτύου (ΤΝΔ). Όταν τα perceptrons οργανώνονται σε στρώματα και συνδέονται μεταξύ τους, επιτρέπουν στο δίκτυο να επιλύει πολύπλοκα προβλήματα, αρκεί να διαθέτει επαρκή υπολογιστική ισχύ και να εκπαιδευτεί κατάλληλα (Goodfellow et al., 2016). Αυτή η πολυεπίπεδη αρχιτεκτονική επιτρέπει την αναπαράσταση μη γραμμικών συναρτήσεων και επιτρέπει στο μοντέλο να μαθαίνει περίπλοκα μοτίβα.

**Multilayer Perceptron:** Το Multilayer Perceptron (MLP) είναι μια επέκταση του κλασικού perceptron και αποτελείται από πολλαπλά στρώματα: ένα εισαγωγικό (input layer), ένα ή περισσότερα κρυφά (hidden layers) και ένα εξαγωγικό στρώμα (output layer). Κάθε

perceptron σε ένα επίπεδο συνδέεται με όλους τους perceptrons στο επόμενο επίπεδο (πλήρως συνδεδεμένο δίκτυο). Αυτή η διάταξη επιτρέπει την ανάλυση πολυδιάστατων σχέσεων στα δεδομένα και χρησιμοποιείται σε εφαρμογές όπως η αναγνώριση εικόνας, η πρόβλεψη χρηματιστηριακών τιμών και η ανάλυση συναισθήματος (Haykin, 2009).

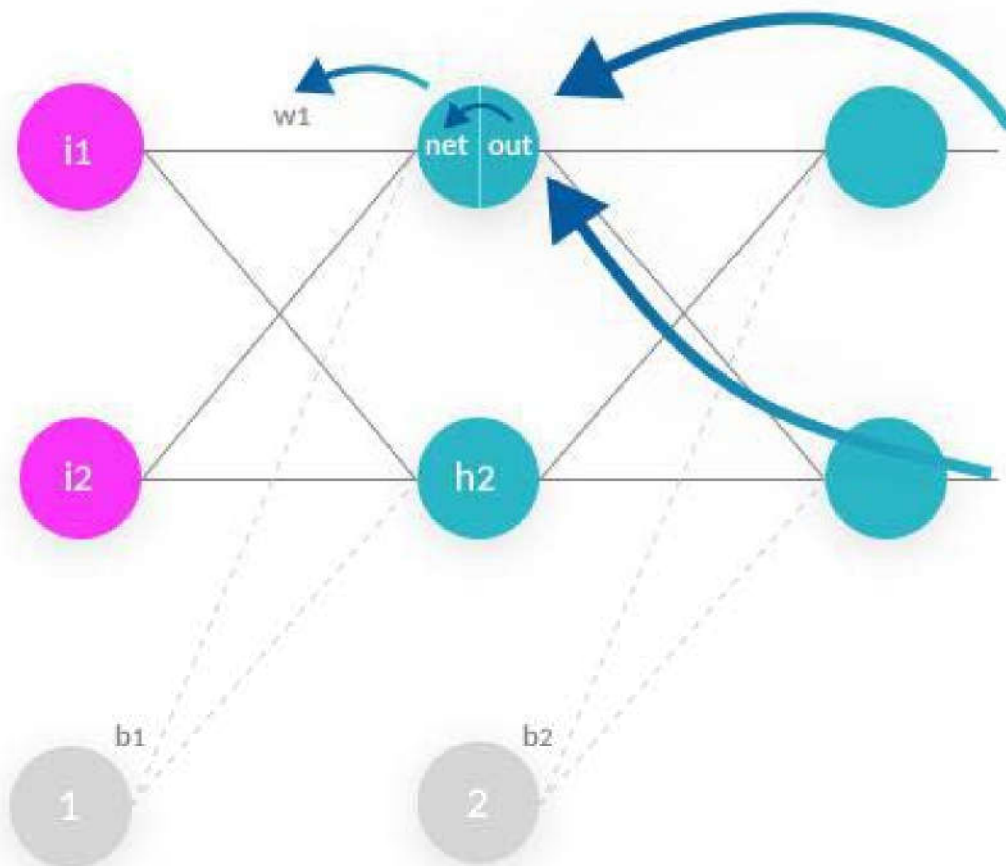


*Απλό παράδειγμα TNΔ τύπου Multilayer Perceptron με ένα κρυφό στρώμα.*

Σε ένα απλό MLP με ένα μόνο κρυφό στρώμα, οι νευρώνες του πρώτου στρώματος λαμβάνουν εισόδους και υπολογίζουν σταθμισμένα αθροίσματα που διέρχονται από μια μη γραμμική συνάρτηση ενεργοποίησης (π.χ., sigmoid, tanh ή ReLU). Το αποτέλεσμα μεταδίδεται στους νευρώνες του επόμενου επιπέδου. Η χρήση μη γραμμικών ενεργοποιήσεων επιτρέπει στα MLP να επιλύουν προβλήματα ταξινόμησης με πολλαπλές εξόδους ή συνεχείς τιμές (Nielsen, 2015). Όταν το MLP έχει τρία ή περισσότερα επίπεδα, χαρακτηρίζεται ως "βαθύ νευρωνικό δίκτυο" (Deep Neural Network – DNN), καθιστώντας το κατάλληλο για προηγμένες εφαρμογές.

Η τεχνική της οπισθοδρομικής διάδοσης (backpropagation) χρησιμοποιείται κατά την εκπαίδευση ενός MLP. Αρχικά, τα βάρη του δικτύου έχουν τυχαίες ή μηδενικές τιμές. Καθώς το δίκτυο δέχεται δεδομένα, υπολογίζει εξόδους και συγκρίνει τα αποτελέσματα με τις επιθυμητές εξόδους (στόχους). Το σφάλμα που προκύπτει μεταφέρεται προς τα πίσω (από την έξοδο στην είσοδο), τροποποιώντας τα βάρη των συνδέσεων μέσω του αλγορίθμου gradient descent, ώστε να μειωθεί το σφάλμα. Η διαδικασία επαναλαμβάνεται μέχρι να επιτευχθεί αποδεκτό επίπεδο ακρίβειας (Rumelhart et. al., 1986).

**Back Propagation NN:** Η οπισθοδρομική μάθηση (backpropagation) αποτελεί έναν από τους πιο θεμελιώδεις και διαδεδομένους αλγορίθμους εκπαίδευσης τεχνητών νευρωνικών δικτύων (ΤΝΔ). Χρησιμοποιείται για τη ρύθμιση των βαρών των συνδέσεων μεταξύ των νευρώνων ώστε να ελαχιστοποιηθεί το σφάλμα πρόβλεψης του δικτύου. Όταν δημιουργείται ένα ΤΝΔ, τα βάρη κάθε νευρώνα αρχικοποιούνται με τυχαίες ή μηδενικές τιμές. Καθώς το δίκτυο δέχεται εισόδους, παράγει εξόδους με βάση τα τρέχοντα βάρη.



*Απλό παράδειγμα ANN τύπου Back Propagation.*

Η πραγματική έξοδος συγκρίνεται με την επιθυμητή (γνωστή) έξοδο και το σφάλμα (error) που προκύπτει χρησιμοποιείται για την αναπροσαρμογή των βαρών. Το σφάλμα μεταδίδεται προς τα πίσω από την έξοδο προς την είσοδο του δικτύου, σύμφωνα με τον αλγόριθμο backpropagation, ο οποίος εφαρμόζει τον κανόνα της αλυσίδας παραγώγων (chain rule) για τον υπολογισμό των βαθμίδων του σφάλματος. Η διαδικασία αυτή επαναλαμβάνεται για πολλές εποχές (epochs) μέχρι το σφάλμα να μειωθεί κάτω από ένα προκαθορισμένο όριο. Ο αλγόριθμος αυτός επέτρεψε για πρώτη φορά την αποτελεσματική εκπαίδευση πολυεπίπεδων perceptrons (MLPs) και αποτέλεσε ορόσημο για την πρόοδο της τεχνητής νοημοσύνης (Rumelhart et. al., 1986).

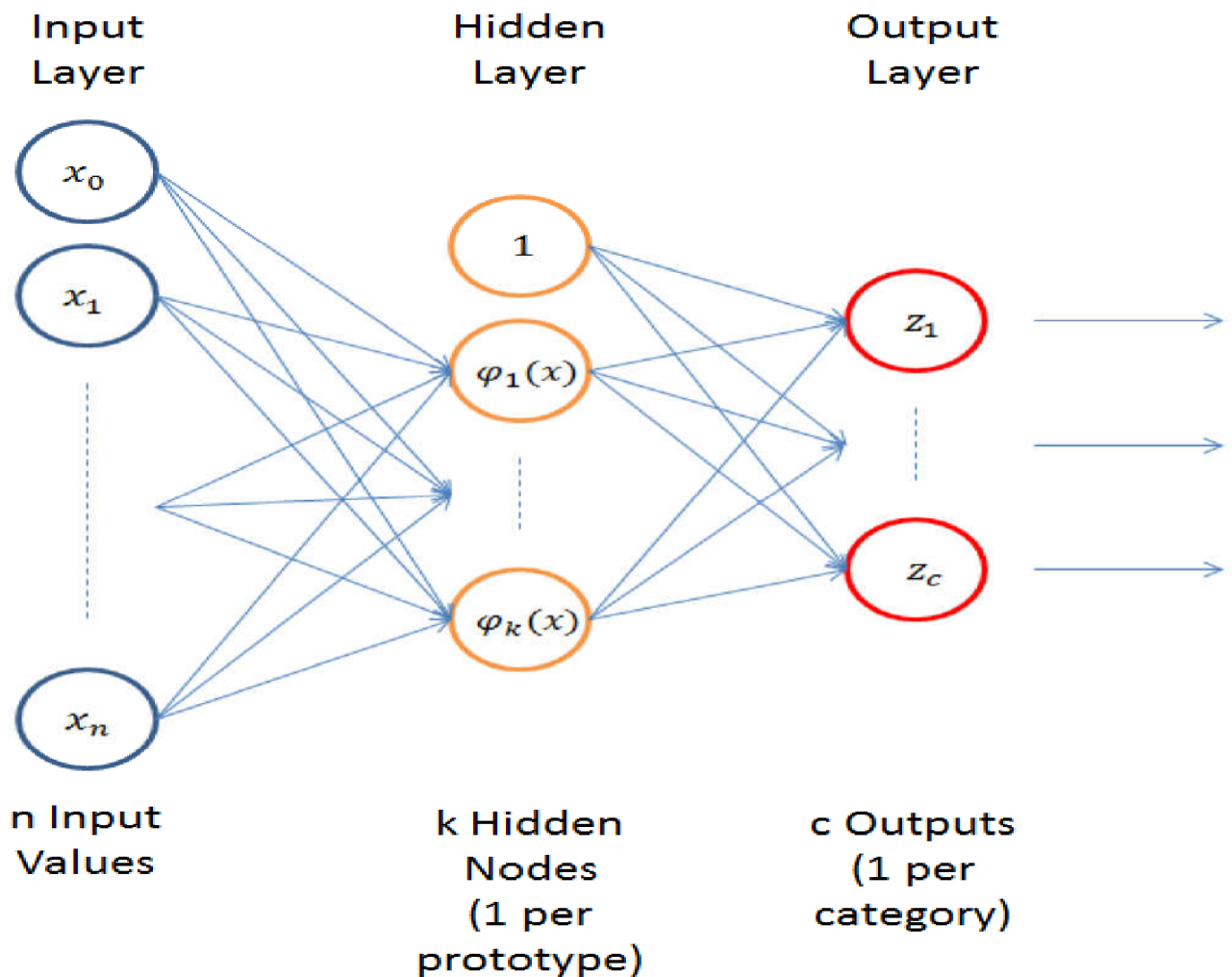
**Radial Basis Function NN (RBFNN):** Ένα Δίκτυο Λειτουργιών Ακτινικής Βάσης (Radial Basis Function Network – RBFN) αποτελεί έναν ειδικό τύπο τεχνητού νευρωνικού

δικτύου (ΓΝΔ), ο οποίος διαφέρει σημαντικά από τα παραδοσιακά πολυεπίπεδα perceptrons (MLPs). Ενώ τα MLP χρησιμοποιούν συνήθως γραμμικούς συνδυασμούς εισόδων και μη γραμμικές συναρτήσεις ενεργοποίησης (όπως ReLU ή sigmoid), το RBFN βασίζεται σε συναρτήσεις ομοιότητας και χρησιμοποιεί μη γραμμικές συναρτήσεις μετασχηματισμού (radial basis functions) όπως η Gaussian (Bishop, 2006).

Στα MLP, κάθε νευρώνας λειτουργεί με βάση το σταθμισμένο άθροισμα των εισόδων του και την εφαρμογή μιας συνάρτησης ενεργοποίησης. Αυτά τα δίκτυα είναι ικανά να λειτουργούν τόσο ως γραμμικοί όσο και ως μη γραμμικοί ταξινομητές, ανάλογα με την πολυπλοκότητα της αρχιτεκτονικής (Goodfellow, Bengio, & Courville, 2016). Η επεξεργασία στα MLPs βασίζεται κυρίως στη σταδιακή μάθηση των βαρών των συνδέσεων μέσω οπισθοδρόμησης (backpropagation).

Αντίθετα, τα RBFN εκτελούν την ταξινόμηση μετρώντας την ομοιότητα της εισόδου με συγκεκριμένα πρότυπα ή «πρωτότυπα» που έχουν αποθηκευτεί από το σύνολο εκπαίδευσης. Κάθε νευρώνας RBF στο κρυφό στρώμα αντιστοιχεί σε ένα συγκεκριμένο παράδειγμα εκπαίδευσης και υπολογίζει την Ευκλείδεια απόσταση μεταξύ της νέας εισόδου και του κέντρου του (prototype), εφαρμόζοντας συνήθως μια Gaussian συνάρτηση (Haykin, 2009). Η έξοδος αυτής της συνάρτησης αντικατοπτρίζει την ομοιότητα: όσο μικρότερη η απόσταση, τόσο μεγαλύτερη η ενεργοποίηση του νευρώνα.

Για παράδειγμα, σε ένα πρόβλημα δυαδικής ταξινόμησης με δύο κλάσεις (Α και Β), αν η είσοδος είναι πιο κοντά στα «πρωτότυπα» της κλάσης Α, το δίκτυο θα την ταξινομήσει σε αυτή την κατηγορία. Η διαδικασία αυτή μοιάζει με μέθοδο πλησιέστερου γείτονα, αλλά ενσωματώνεται σε ένα συνεχές και παραμετροποιήσιμο νευρωνικό πλαίσιο (Moody & Darken, 1989).



*Απλό παράδειγμα ANN τύπου Radial Basis Function*

Το RBFN αποτελείται γενικά από τρία στρώματα:

1. Το εισαγωγικό στρώμα, όπου το διάνυσμα εισόδου μεταφέρεται στο δίκτυο.
2. Το κρυφό στρώμα, στο οποίο εφαρμόζονται οι συναρτήσεις ακτινικής βάσης – εδώ μετρώνται οι αποστάσεις μεταξύ εισόδων και πρωτοτύπων.
3. Το εξαγωγικό στρώμα, στο οποίο γίνεται ένας γραμμικός συνδυασμός των εξόδων του κρυφού στρώματος ώστε να παραχθεί η τελική έξοδος του δικτύου.

Η λειτουργία αυτή παρέχει στα RBFN τη δυνατότητα να ανταποκρίνονται αποτελεσματικά σε προβλήματα παλινδρόμησης και ταξινόμησης, ειδικά όταν ο όγκος των δεδομένων είναι περιορισμένος ή τα πρότυπα είναι καλά διαχωρίσιμα (Zhang et al., 1992). Σε αντίθεση με τα MLP, τα RBFN μπορούν να εκπαιδευτούν γρήγορα, δεδομένου ότι η εκπαίδευση του κρυφού στρώματος βασίζεται κυρίως στην τοποθέτηση των πρωτοτύπων, ενώ το εξαγωγικό στρώμα μπορεί να προσαρμοστεί με τεχνικές παλινδρόμησης..

**Generalized Regression Neural Network (GRNN):** Το Generalized Regression Neural Network (GRNN), δηλαδή το Γενικευμένο Νευρωνικό Δίκτυο Παλινδρόμησης, ανήκει στην κατηγορία των τεχνητών νευρωνικών δικτύων ακτινικής βάσης (Radial Basis Function Networks - RBFN) και προτάθηκε από τον Donald F. Specht το 1991. Το GRNN αποτελεί μια μη παραμετρική προσέγγιση παλινδρόμησης που έχει σχεδιαστεί για προβλήματα πρόβλεψης, παλινδρόμησης και ταξινόμησης, με ιδιαίτερη επιτυχία σε συστήματα με υψηλή αβεβαιότητα ή μεταβαλλόμενα δεδομένα.

Το βασικό χαρακτηριστικό που το διαχωρίζει από άλλα δίκτυα RBF είναι η δυνατότητά του να προσεγγίζει οποιαδήποτε συνάρτηση παλινδρόμησης χωρίς να απαιτεί επαναληπτική εκπαίδευση ή ρύθμιση βαρών μέσω οπισθοδρόμησης (Haykin, 2009). Σε αντίθεση με τα MLP ή ακόμα και τα βασικά RBFN, το GRNN δεν απαιτεί παραδοσιακή διαδικασία εκπαίδευσης, καθώς βασίζεται σε στατιστικές αρχές και συγκεκριμένα στον εκτιμητή πυκνότητας πιθανότητας τύπου Parzen.

Η δομή του GRNN αποτελείται από τέσσερα στρώματα:

1. Το εισαγωγικό στρώμα (input layer),
2. Το στρώμα μοτίβου (pattern layer), όπου κάθε νευρώνας αντιστοιχεί σε ένα σημείο του συνόλου εκπαίδευσης,
3. Το στρώμα συνάθροισης (summation layer), το οποίο υπολογίζει τις σταθμισμένες συναρτήσεις ομοιότητας και
4. Το στρώμα εξόδου (output layer), όπου υπολογίζεται η τελική πρόβλεψη βάσει της σχέσης σταθμισμένου μέσου (Parzen window estimator).

Η κεντρική αρχή του GRNN είναι ότι κάθε σημείο εκπαίδευσης συμβάλλει στη συνολική πρόβλεψη, ανάλογα με το πόσο κοντά βρίσκεται στην υπό εξέταση είσοδο, με βάση την Ευκλείδεια απόσταση και τη χρήση μιας Gaussian συνάρτησης μεταφοράς. Αυτή η δομή καθιστά το GRNN ιδιαίτερα κατάλληλο για προβλήματα όπου απαιτείται γρήγορη μάθηση από μικρό αριθμό παραδειγμάτων, όπως σε εφαρμογές ελέγχου ρομπότ, πρόβλεψης και διαχείρισης πολύπλοκων συστημάτων (Specht, 1991; Yegnanarayana, 2009).

Επιπλέον, λόγω της δυναμικής φύσης του – δηλαδή της δυνατότητας να ενσωματώνει νέα δεδομένα χωρίς την ανάγκη επανεκπαίδευσης ολόκληρου του δικτύου – το GRNN χρησιμοποιείται ευρέως σε online learning συστήματα, καθώς και σε χρηματοοικονομική πρόβλεψη, πρόγνωση σειρών χρόνου και ιατρικές διαγνώσεις (Zhang et. al., 1998).

**Self Organizing Maps (SOM):** Οι Αυτοοργανούμενοι Χάρτες (Self-Organizing Maps – SOM), γνωστοί και ως Χάρτες Kohonen, αποτελούν ένα είδος τεχνητού νευρωνικού δικτύου μη επιβλεπόμενης μάθησης, το οποίο εισήχθη από τον Φινλανδό επιστήμονα Τεουνο Kohonen στις αρχές της δεκαετίας του 1980 (Kohonen, 1982). Η βασική λειτουργία των SOM είναι η προβολή δεδομένων υψηλής διάστασης σε έναν δισδιάστατο χάρτη, επιτρέποντας τη μείωση διαστάσεων και τη διαισθητική οπτική κατανόηση των σχέσεων και των μοτίβων στα δεδομένα (Kohonen, 2001).

Σε αντίθεση με άλλα τεχνητά νευρωνικά δίκτυα, όπως τα MLP που εκπαιδεύονται με επιβλεπόμενη μάθηση μέσω διόρθωσης σφάλματος, οι SOM βασίζονται σε ανταγωνιστική

μάθηση (competitive learning), όπου οι νευρώνες "ανταγωνίζονται" για να ενεργοποιηθούν από μια δεδομένη είσοδο. Ο νευρώνας που παρουσιάζει τη μεγαλύτερη ομοιότητα με την είσοδο ανακηρύσσεται "νικητής" (winner-takes-all) και προσαρμόζει τα βάρη του, καθώς και οι νευρώνες της γειτονιάς του, ενισχύοντας έτσι τη διατήρηση της τοπολογικής δομής του αρχικού χώρου εισόδου (Haykin, 2009).

Το δίκτυο των SOM αποτελείται από έναν πίνακα νευρώνων – συνήθως δισδιάστατο – όπου κάθε κόμβος αντιπροσωπεύει ένα διάνυσμα αναφοράς (weight vector). Η εκπαίδευση του SOM βασίζεται στη σταδιακή μετακίνηση των διανυσμάτων αναφοράς προς τις εισερχόμενες τιμές, δημιουργώντας μια χωρική αναπαράσταση του συνόλου δεδομένων που ομαδοποιεί παρόμοια πρότυπα (Vesanto & Alhoniemi, 2000). Αυτό επιτρέπει την εφαρμογή των SOM σε εξερεύνηση δεδομένων, ομαδοποίηση (clustering) και οπτικοποίηση σύνθετων πολυδιάστατων συνόλων δεδομένων, χωρίς την ανάγκη για προκαθορισμένες ετικέτες.

Επιπλέον, σημαντική καινοτομία των SOM αποτελεί η δυνατότητά τους να προβάλλουν τη μη γραμμική συσχέτιση μεταξύ μεταβλητών με τέτοιο τρόπο ώστε οι γειτονικοί νευρώνες στο χάρτη να αντιστοιχούν σε παρόμοια πρότυπα εισόδου. Αυτή η διατήρηση τοπολογίας καθιστά τους SOM ιδανικούς για ανάλυση φασματογραφικών δεδομένων, ταξινόμηση εικόνων, χρηματοοικονομικά δεδομένα, αιόμη και βιοϊατρικές εφαρμογές (Deboeck & Kohonen, 1998).

**Learning Vector Quantization (LVQ):** Η Εκπαιδευτική Ανίχνευση Διανύσματος (Learning Vector Quantization – LVQ) είναι μια τεχνική εποπτευόμενης μάθησης που ανήκει στην κατηγορία των ανταγωνιστικών τεχνητών νευρωνικών δικτύων (competitive neural networks) και αναπτύχθηκε από τον Teuvo Kohonen στα τέλη της δεκαετίας του 1980 (Kohonen, 1990). Σε αντίθεση με τα Self-Organizing Maps (SOM), τα οποία χρησιμοποιούν μη επιβλεπόμενη μάθηση, το LVQ ενσωματώνει εποπτευόμενη εκπαίδευση, προσφέροντας σαφή ερμηνευσιμότητα και αυξημένη ακρίβεια στην κατηγοριοποίηση προτύπων.

Η βασική λειτουργία του LVQ στηρίζεται στην αρχή της ταξινόμησης μοτίβων, όπου κάθε μονάδα εξόδου αντιστοιχεί σε μία συγκεκριμένη κλάση. Κατά τη διάρκεια της εκπαίδευσης, παρουσιάζεται στο δίκτυο ένα σύνολο δεδομένων με γνωστές ετικέτες (labels) και το δίκτυο προσαρμόζει τα διανύσματα αναφοράς (reference vectors ή prototypes) έτσι ώστε να πλησιάζουν τα πρότυπα της αντίστοιχης κλάσης (Haykin, 2009).

Η διαδικασία περιλαμβάνει υπολογισμό της Ευκλείδειας απόστασης μεταξύ ενός εισερχόμενου διανύσματος και όλων των διανυσμάτων αναφοράς. Αν η ετικέτα της νικήτριας μονάδας (best matching unit) ταιριάζει με την ετικέτα της εισόδου, τότε το διάνυσμα αναφοράς κινείται προς την είσοδο. Διαφορετικά, απομακρύνεται από αυτή. Με τον τρόπο αυτό, το δίκτυο μαθαίνει να αυξάνει την ομοιότητα μεταξύ παραδειγμάτων της ίδιας κλάσης και να ενισχύει τη διάκριση μεταξύ διαφορετικών κλάσεων (Kohonen, 2001).

Μετά την ολοκλήρωση της εκπαίδευσης, το LVQ μπορεί να χρησιμοποιηθεί για την ταξινόμηση νέων, άγνωστων διανυσμάτων εισόδου, αντιστοιχίζοντάς τα με την πλησιέστερη μονάδα εξόδου. Αυτή η διαδικασία καθιστά το LVQ κατάλληλο για εφαρμογές όπως η αναγνώριση ομιλίας, η ιατρική διάγνωση, η επεξεργασία σήματος και η κατηγοριοποίηση βιομετρικών δεδομένων.

Επιπλέον, σε αντίθεση με πιο περίπλοκα μοντέλα, το LVQ είναι απλό στην υλοποίηση, δεν απαιτεί μεγάλη υπολογιστική ισχύ και προσφέρει μεγαλύτερη διαφάνεια στην απόφαση ταξινόμησης, αφού η απόσταση ενός εισερχόμενου διανύσματος από το πρότυπο ταξινόμησης είναι εύκολα υπολογίσιμη και ερμηνεύσιμη (Kohonen, 2001).

#### 4.6.3 Ιστορικά στοιχεία για τα ΤΝΔ

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) έχουν μια πλούσια ιστορία, παρουσιάζοντας σημαντικές εξελίξεις στον τομέα της μηχανικής μάθησης. Ας δούμε μερικά κλειδιά ιστορικά στοιχεία:

- 1943 - Τα Μοντέλα Νευρώνων McCulloch-Pitts: Η έννοια των τεχνητών νευρωνικών δικτύων έχει τις ρίζες της στη δεκαετία του 1940 με το έργο του Warren McCulloch και του Walter Pitts. Δημιούργησαν ένα μαθηματικό μοντέλο νευρώνων που έπαιξε σημαντικό ρόλο στην ανάπτυξη των ΤΝΔ.,
- 1949, ο Donald Hebb ενίσχυσε την έννοια των νευρώνων στο βιβλίο του «Ο Οργανισμός Συμπεριφοράς». Επισημαίνει ότι οι νευρικές οδοί ενισχύονται κάθε φορά που χρησιμοποιούνται.
- Στη δεκαετία του 1950, ο Nathaniel Rochester από τα ερευνητικά εργαστήρια της IBM έκανε πρώτος προσπάθεια να προσομοιώσει ένα ΤΝΔ.
- 1956, το Dartmouth Summer Research Project για την Τεχνητή Νοημοσύνη (Artificial Intelligence, AI) έδωσε ώθηση τόσο στην τεχνητή νοημοσύνη όσο και στα ΤΝΔ. Έτσι ξεκίνησε η έρευνα στην AI σε συσχέτιση με τις χαμηλότερου επιπέδου λειτουργίες του νευρικού συστήματος του εγκεφάλου.
- 1957, ο John von Neumann πρότεινε να γίνει απομίμηση απλών λειτουργιών νευρώνων χρησιμοποιώντας το τηλεγράφημα ή σωλήνες κενού.
- 1958 - Το Πρώτο Νευρωνικό Δίκτυο του Rosenblatt: Ο Frank Rosenblatt παρουσίασε το Perceptron, ένα απλό νευρωνικό δίκτυο, το οποίο ήταν ικανό να μάθει και να εκτελεί δυαδική ταξινόμηση.
- 1959, οι Bernard Widrow και Marcian Hoff από το Πανεπιστήμιο Stanford των ΗΠΑ παρουσίασαν μοντέλα που ονόμασαν ADALINE και MADALINE, χρησιμοποιώντας πολλαπλά ADaptive LINEar Elements. Το MADALINE, το οποίο είναι ένα προσαρμοστικό φίλτρο, εφαρμόστηκε για πρώτη φορά σε πραγματικό πρόβλημα. Είναι ένα προσαρμοστικό φίλτρο (adaptive filter) που εξαλείφει την ηχώ στις τηλεφωνικές γραμμές (echo cancellation). Εξαιτίας των επιτυχημένων εφαρμογών του, το MADALINE εξακολουθεί να χρησιμοποιείται εμπορικά σήμερα.
- 1969 - Περίοδος Πτώσης: Ένα άρθρο από τον Marvin Minsky και τον Seymour Papert ανέδειξε τις περιορισμένες δυνατότητες του Perceptron, οδηγώντας σε έναν εκτενή χρονικό διάστημα αμφισβήτησης για τη χρησιμότητα των ΤΝΔ.

- 1982, ο John Hopfield παρουσίασε την προσέγγισή του για τη δημιουργία χρήσιμων διατάξεων ΤΝΔ σε ένα άρθρο που δημοσιεύτηκε στην αμερικανική Εθνική Ακαδημία Επιστημών. Την ίδια χρονιά, κατά τη διάρκεια της κοινής διάσκεψης ΗΠΑ-Ιαπωνίας για τα Συνεργατικά / Ανταγωνιστικά Νευρωνικά Δίκτυα, η Ιαπωνία παρουσίασε την 5η γενιά των προσπαθειών της στον τομέα αυτόν. Αυτό έκανε τις ΗΠΑ να ανησυχούν για την πιθανή καθυστέρηση των ερευνητικών τους προσπαθειών. Ωστόσο, σύντομα η χρηματοδότηση για την έρευνα στα ΤΝΔ ανακτήθηκε και επανήλθε στα επίπεδα προηγούμενων ετών.
- 1985, το Αμερικανικό Ινστιτούτο Φυσικής (AIP) εκκίνησε την ετήσια συνάντηση που πλέον γνωρίζουμε ως "Νευρωνικά Δίκτυα για την Πληροφορική". Έως το 1987, το πρώτο Διεθνές Συνέδριο για τα ΤΝΔ του Ινστιτούτου Ηλεκτρολόγων Ηλεκτρονικών Μηχανικών (IEEE) συγκέντρωσε περισσότερους από 1.800 συμμετέχοντες.
- 1986 - Η Επανεκκίνηση με τα Πολλαπλά Στρώματα: Η ανακάλυψη της πίσω διάδοσης (backpropagation) από τους Rumelhart, Hinton και Williams επέτρεψε στα ΤΝΔ να εκπαιδεύονται με πολλαπλά στρώματα, ανοίγοντας τον δρόμο για πιο πολύπλοκες λειτουργίες.
- Δεκαετία του '90 - Άνθηση των Νευρωνικών Δικτύων: Οι εξελίξεις στη θεωρία και την υπολογιστική ισχύ οδήγησαν σε μεγάλη αύξηση του ενδιαφέροντος για τα ΤΝΔ, με εφαρμογές σε ευρύ φάσμα πεδίων.
- Δεκαετία του 2000 - Deep Learning και Επανεκκίνηση: Η ανάπτυξη της υπολογιστικής ισχύος και η ανακάλυψη νέων αρχιτεκτονικών έφεραν την αναγέννηση των ΤΝΔ, οδηγώντας στην εικρητική ανάπτυξη του Deep Learning.
- Σήμερα - Εφαρμογές και Εξελίξεις: Τα ΤΝΔ χρησιμοποιούνται ευρέως σε εφαρμογές όπως η αναγνώριση εικόνας, η φωνητική αναγνώριση, η αυτόματη μετάφραση, τα αυτόνομα οχήματα, κ.ά. Η έρευνα συνεχίζεται για να βελτιωθούν οι αρχιτεκτονικές και οι αλγόριθμοι των ΤΝΔ.

Το παραπάνω αποτελεί μια συνοπτική ανασκόπηση, καθώς η προκοπή στον τομέα των ΤΝΔ είναι συνεχής και διαρκώς εξελίσσεται.

#### 4.6.4 Εφαρμογές των ΤΝΔ

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) έχουν ευρεία εφαρμογή σε πολλούς τομείς λόγω της ικανότητάς τους να εξάγουν πολύπλοκα μοντέλα από δεδομένα και να προβλέπουν, ταξινομούν, ή εξάγουν σημαντικά χαρακτηριστικά. Αναλύοντας τις εφαρμογές των ΤΝΔ:

- Αναγνώριση Προτύπων: Χρησιμοποιούνται για αναγνώριση προτύπων σε εικόνες, ήχο και βίντεο. Εφαρμόζονται στον τομέα της γραφολογίας και της αναγνώρισης χειρόγραφων κειμένων.
- Επεξεργασία Σημάτων: Χρησιμοποιούνται για τη συμπίεση δεδομένων χωρίς απώλειες. Χρησιμοποιούνται για εξαγωγή χαρακτηριστικών και καταστολή θορύβου σε σήματα.

- **Πρόβλεψη και Εκτίμηση:** Χρησιμοποιούνται για πρόβλεψη οικονομικών δεικτών και τάσεων. Εφαρμόζονται στην ασφαλιστική βιομηχανία για την εκτίμηση κινδύνων.
- **Αυτοματοποιημένο Συστήματα:** Χρησιμοποιούνται για την εκπαίδευση ρομπότ και την αυτοματοποίηση της λήψης αποφάσεων.
- **Συστήματα Ελέγχου:** Εφαρμόζονται για τον έλεγχο και την παρακολούθηση διαφόρων διεργασιών.
- **Συστήματα Συστάσεων:** Χρησιμοποιούνται για τη δημιουργία προσωποποιημένων συστάσεων σε διάφορες πλατφόρμες, όπως σε ιστοσελίδες και υπηρεσίες streaming.
- **Ιατρική:** Εφαρμόζονται για διάφορες ιατρικές εφαρμογές, συμπεριλαμβανομένης της διάγνωσης νοσημάτων και της πρόβλεψης αποτελεσμάτων θεραπείας.
- **Χρηματοοικονομικά:** Χρησιμοποιούνται για τον αναλυτικό έλεγχο αγορών, τη διαχείριση κινδύνων και τον προσδιορισμό ευκαιριών επενδύσεων.
- **Διαχείριση Ενέργειας:** Εφαρμόζονται για τον έλεγχο και τη βελτιστοποίηση συστημάτων διανομής ενέργειας και για την πρόβλεψη της κατανάλωσης ενέργειας.
- **Άμυνα:** Αντιτρομοκρατία, αναγνώριση προσώπου, εξαγωγή χαρακτηριστικών, καταστολή θορύβου, διάκριση αντικειμένων, αισθητήρες, σόναρ, επεξεργασία ραντάρ και σήματος εικόνας, αναγνώριση σημάτων / εικόνων, παρακολούθηση στόχων και διαχείριση όπλων.
- **Εκπαίδευση:** Προσαρμοστικό λογισμικό μάθησης, δυναμική πρόβλεψη, ανάλυση και πρόβλεψη του εκπαιδευτικού συστήματος, μοντελοποίηση των επιδόσεων των σπουδαστών και διαμόρφωση προφίλ προσωπικότητας.

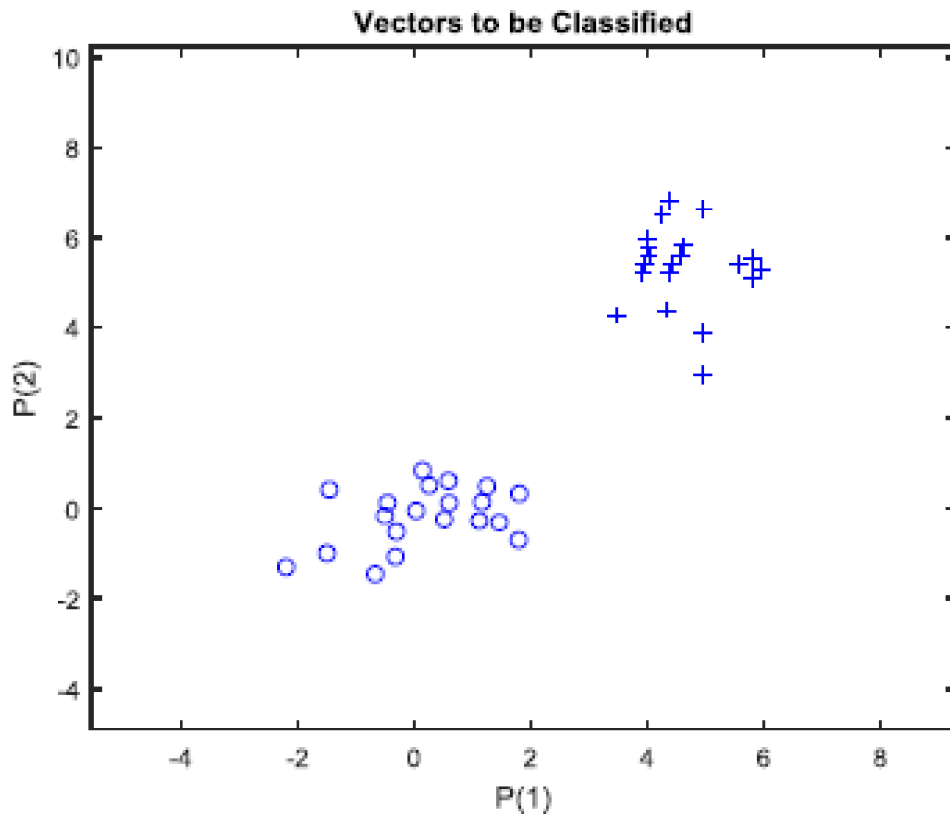
Αυτές είναι μόνο μερικές από τις πολλές εφαρμογές των Τεχνητών Νευρωνικών Δικτύων, οι οποίες συνεχώς επεκτείνονται καθώς η τεχνολογία εξελίσσεται.

#### **4.6.5 Παραδείγματα επίλυσης προβλημάτων με ΤΝΔ**

##### **Ταξινόμηση γραμμικά διαχωρίσιμων δεδομένων με Perceptron**

Δύο ομάδες δεδομένων, που ανήκουν σε δύο κατηγορίες, ορίζονται σε έναν διδιάστατο χώρο εισόδων. Υποτίθεται ότι οι 2 τάξεις είναι γραμμικά διαχωρίσιμες. Ο στόχος είναι να κατασκευαστεί και να εκπαιδευτεί ένα Perceptron για την ορθή ταξινόμηση άγνωστων δεδομένων εισόδων.

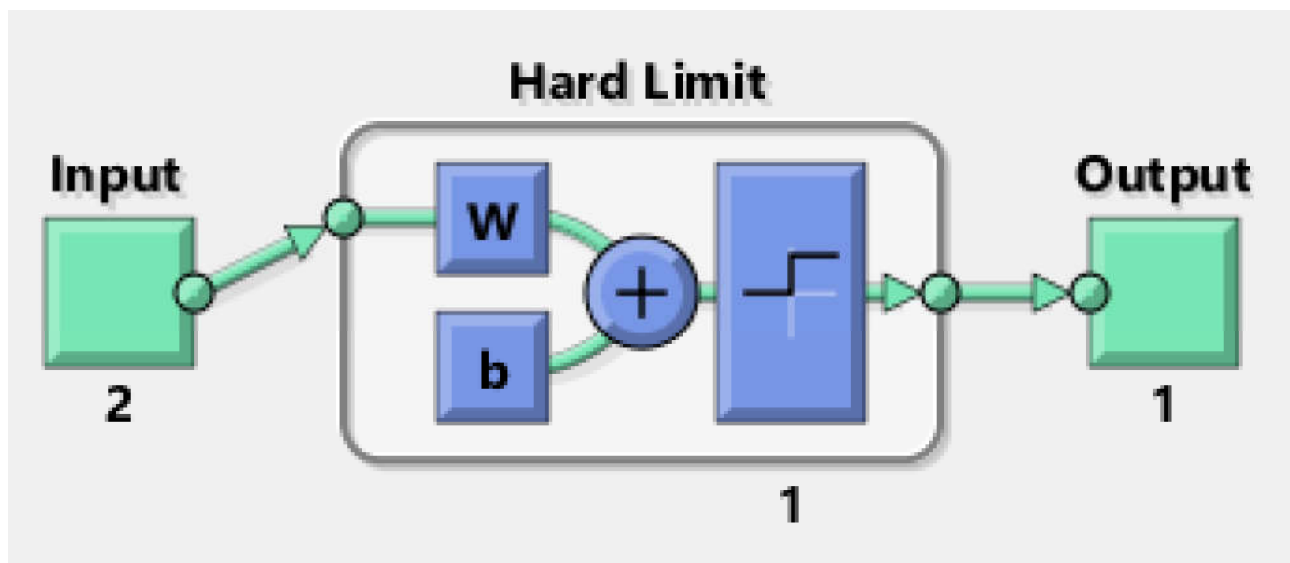
**(α) Ορίζονται τα δεδομένα εισόδου και εξόδου για την εκπαίδευση του δικτύου.**



*Δημιουργία δύο ομάδων δεδομένων για διαχωρισμό αυτών.*

Το Perceptron είναι συνήθως χρησιμοποιούμενο για να διαχωρίζει γραμμικά δεδομένα εισόδου σε δύο κατηγορίες. Επομένως, το χαρακτηρίζουμε και ως γραμμικό δυαδικό ταξινομητή. Κάθε Perceptron με έναν νευρώνα αναλαμβάνει τον ρόλο της ταξινόμησης προτύπων που ανήκουν σε δύο διακριτικές κλάσεις. Συνεπώς, ο στόχος του είναι η σωστή ομαδοποίηση όλων των εισόδων σε μία από τις δύο αυτές κλάσεις.

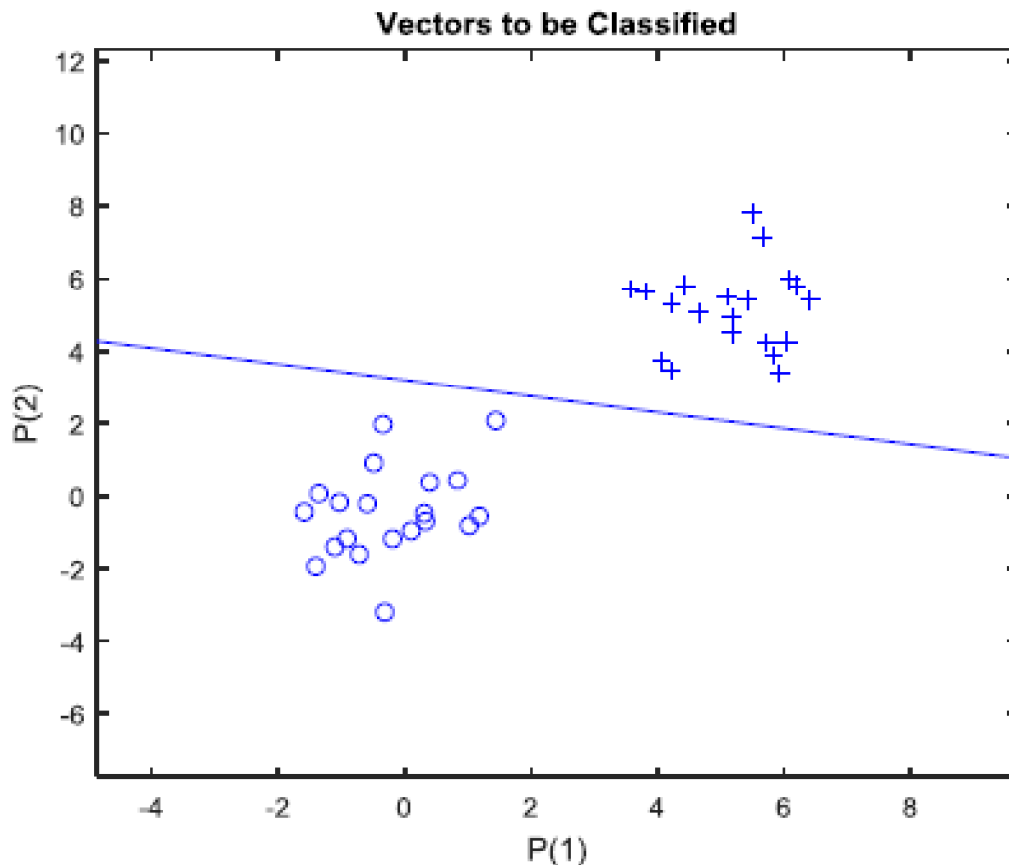
**(β) Δημιουργείται και εκπαιδεύεται το Perceptron:**



*Δημιουργία Perceptron για δυαδική ταξινόμηση*

Οι είσοδοι (inputs) πολλαπλασιάζονται με τα βάρη (weights) τους  $w$  και στο γινόμενο αθροίζεται και μία σταθερά (offset ή bias). Πολλαπλασιάζει όλες τις τιμές στον αθροιστή και δημιουργεί ένα σταθμισμένο άθροισμα. Έπειτα αυτό το άθροισμα το περνάει από μία συνάρτηση ενεργοποίησης (π.χ. βηματική συνάρτηση στην περίπτωση της παραπάνω εικόνας) και παράγει την έξοδο.

(γ) Σχεδιάζεται το όριο απόφασης (γραμμή διαχωρισμού των 2 τάξεων).



*Δημιουργία διαχωριστικής γραμμής για επιβεβαίωση σωστής ταξινόμησης*

#### (δ) Έλεγχος

Μετά το τέλος της εκπαίδευσης, χρησιμοποιείται η γραμμή διαχωρισμού των 2 τάξεων για τη σωστή ταξινόμηση νέων, άγνωστων δεδομένων εισόδων.

#### Επίλυση του προβλήματος XOR με Perceptron πολλαπλών στρώσεων

Το βασικό μειονέκτημα του Perceptron είναι ότι αντιμετωπίζει δυσκολία στην επίλυση προβλημάτων που δεν μπορούν να διαχωριστούν γραμμικά. Στην πραγματική ζωή, τα περισσότερα προβλήματα είναι μη γραμμικά διαχωρίσιμα, ενώ τα γραμμικά προβλήματα μπορούν συχνά να λυθούν με πιο απλούς τρόπους από τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) και το Perceptron.

Ένα παράδειγμα μη γραμμικά διαχωρίσιμου προβλήματος είναι η λογική συνάρτηση XOR, το οποίο δεν μπορεί να αντιπροσωπευθεί από ένα μόνο νευρώνα. Αυτό το πρόβλημα

οδήγησε στην ανάγκη να εξεταστεί το πολυεπίπεδο Perceptron, ένα πιο περίπλοκο δίκτυο που περιλαμβάνει πολλούς νευρώνες.

Στα πολυεπίπεδα Perceptron, το δίκτυο αποτελείται από περισσότερα από ένα γραμμικά στρώματα, γνωστά και ως συνδυασμοί νευρώνων. Σε ένα απλό παράδειγμα με τρία στρώματα, το πρώτο αποτελεί το στρώμα εισόδου, το τελευταίο το στρώμα εξόδου, ενώ το ενδιάμεσο ονομάζεται κρυμμένο στρώμα. Τα δεδομένα εισόδου τροφοδοτούνται στο στρώμα εισόδου, ενώ ο χρήστης λαμβάνει την τελική έξοδο από το στρώμα εξόδου. Ο αριθμός των κρυμμένων στρωμάτων μπορεί να αυξηθεί όσο χρειάζεται, καθιστώντας την αρχιτεκτονική πιο πολύπλοκη, ανάλογα με την πολυπλοκότητα του προβλήματος που επιλύεται.

Αυτή η πολυεπίπεδη αρχιτεκτονική είναι πιο εξελιγμένη από τον κλασικό Perceptron, επιτρέποντάς του να αντιμετωπίζει μη γραμμικά διαχωρίσιμα προβλήματα. Με τη σωστή συνδυασμό βαρών και μετά από εκπαίδευση, το δίκτυο μπορεί να επιτύχει σωστή ταξινόμηση, ακόμη και σε δύσκολες περιπτώσεις όπως η συνάρτηση XOR.

Το πρόβλημα XOR είναι ένα κλασικό πρόβλημα στην έρευνα των ΤΝΔ. Είναι το πρόβλημα της χρήσης ενός ΤΝΔ για την πρόβλεψη των εξόδων των λογικών πυλών XOR που δέχονται δύο δυαδικές εισόδους. Μια πύλη XOR θα πρέπει να επιστρέφει «TRUE» (π.χ. 1) εάν οι δύο εισοδοί δεν είναι ίσες μεταξύ τους και «FALSE» (π.χ. «0» ή «-1») εάν είναι ίσες μεταξύ τους. Όλες οι δυνατές εισοδοί και οι προβλεπόμενες έξοδοι εμφανίζονται στον Πίνακα Αληθείας της πύλης XOR παρακάτω.

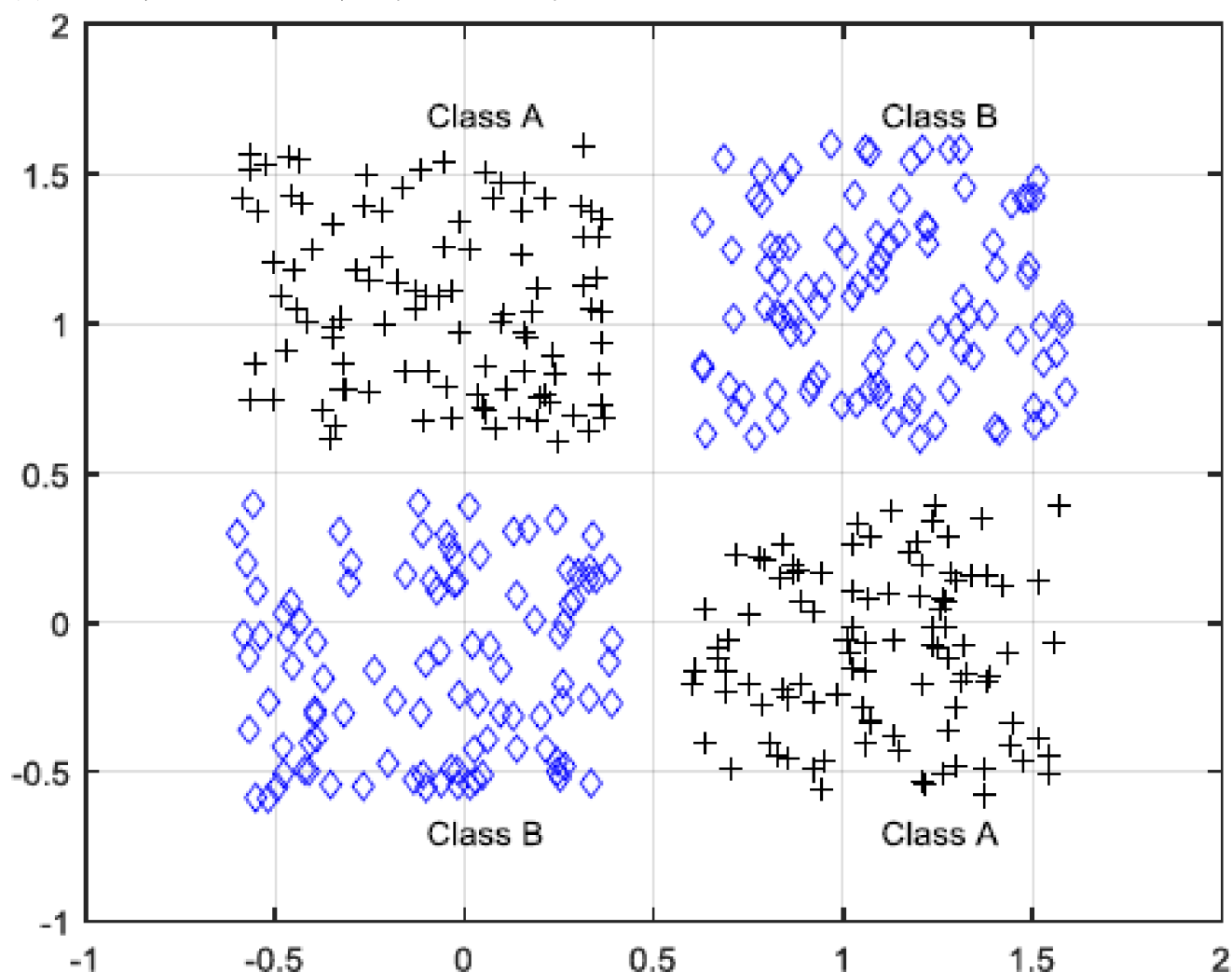
<b>X<sub>1</sub></b>	<b>X<sub>2</sub></b>	<b>X<sub>1</sub> XOR X<sub>2</sub></b>
-1	-1	-1
-1	1	1
1	-1	1
1	1	-1

Πίνακας Αληθείας της λογικής πράξης Exclusive OR (XOR). True = 1, False = -1

Το XOR είναι ένα γνωστό πρόβλημα ταξινόμησης για το οποίο οι αναμενόμενες έξοδοι είναι γνωστές εκ των προτέρων. Επομένως, είναι σκόπιμο να χρησιμοποιηθεί μια προσέγγιση εκπαίδευσης με επίβλεψη (supervised learning) για την εκπαίδευση του ΤΝΔ.

Τέσσερις ομάδες δεδομένων (A, B, C, D) ορίζονται σε έναν 2-διαστάσεων χώρο εισόδου. Τα ζεύγη (A, C) και (B, D) αντιπροσωπεύουν πρόβλημα ταξινόμησης XOR. Ο στόχος είναι να σχεδιαστεί ένα ΤΝΔ για την επίλυση του προβλήματος XOR.

(α) Καθορίζονται τέσσερις ομάδες δεδομένων εισόδου.



*Δημιουργία δύο κλάσεων με δύο πρότυπα η καθεμία, για δυαδική κατηγοριοποίηση*

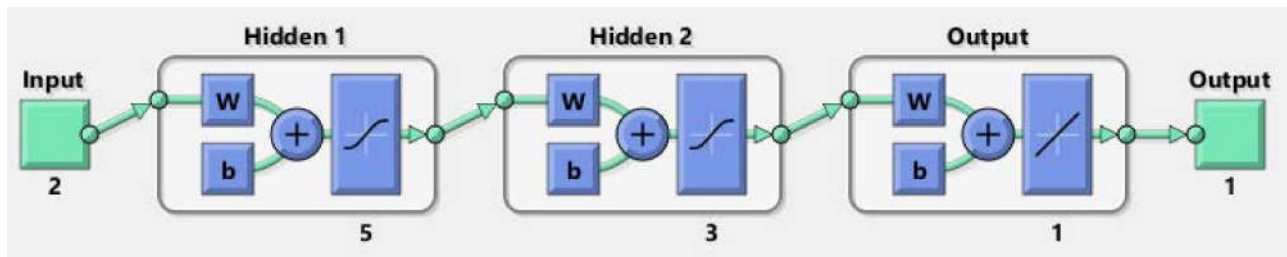
Στο διάγραμμα του XOR, παρατηρούμε δύο κλάσεις με δύο πρότυπα η καθεμία. Η ενδεχόμενη διαχωριστική γραμμή που θα μπορούσε να χωρίσει τις δύο κλάσεις είναι γραμμικά αδύνατη, επομένως απαιτείται η χρήση περισσότερων από μιας γραμμών. Στη συγκεκριμένη περίπτωση, χρησιμοποιούνται δύο γραμμές, μία οριζόντια και μία κατακόρυφη, οι οποίες διασταυρώνονται στο κέντρο του διαγράμματος.

Το νευρωνικό δίκτυο που παρουσιάζεται χρησιμοποιεί πολλαπλούς κρυμμένους κόμβους για να περιγράψει αυτό τον μη γραμμικό διαχωρισμό. Κάθε κρυμμένος κόμβος αντιστοιχεί σε μια από τις γραμμές που αναφέραμε, επιτρέποντας έτσι στο δίκτυο να αντιμετωπίσει μη γραμμικά διαχωρίσιμα προβλήματα, όπως αυτό του XOR.

(β) Ορίζεται η κωδικοποίηση εξόδου για το πρόβλημα XOR.

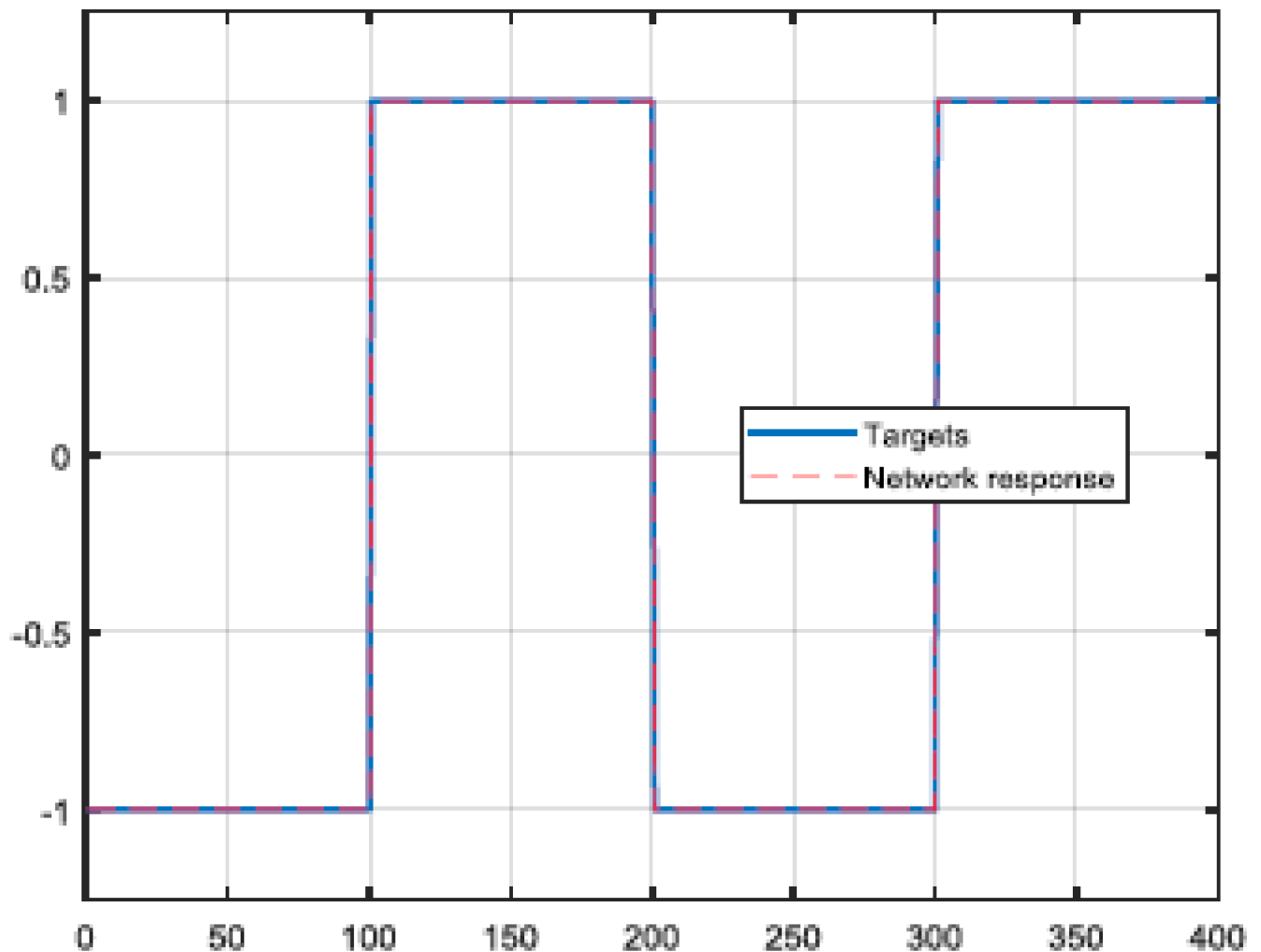
Προετοιμασία εισόδων και εξόδων για εκπαίδευση του δικτύου.

(γ) Δημιουργία και εκπαίδευση ενός Multilayer Perceptron.



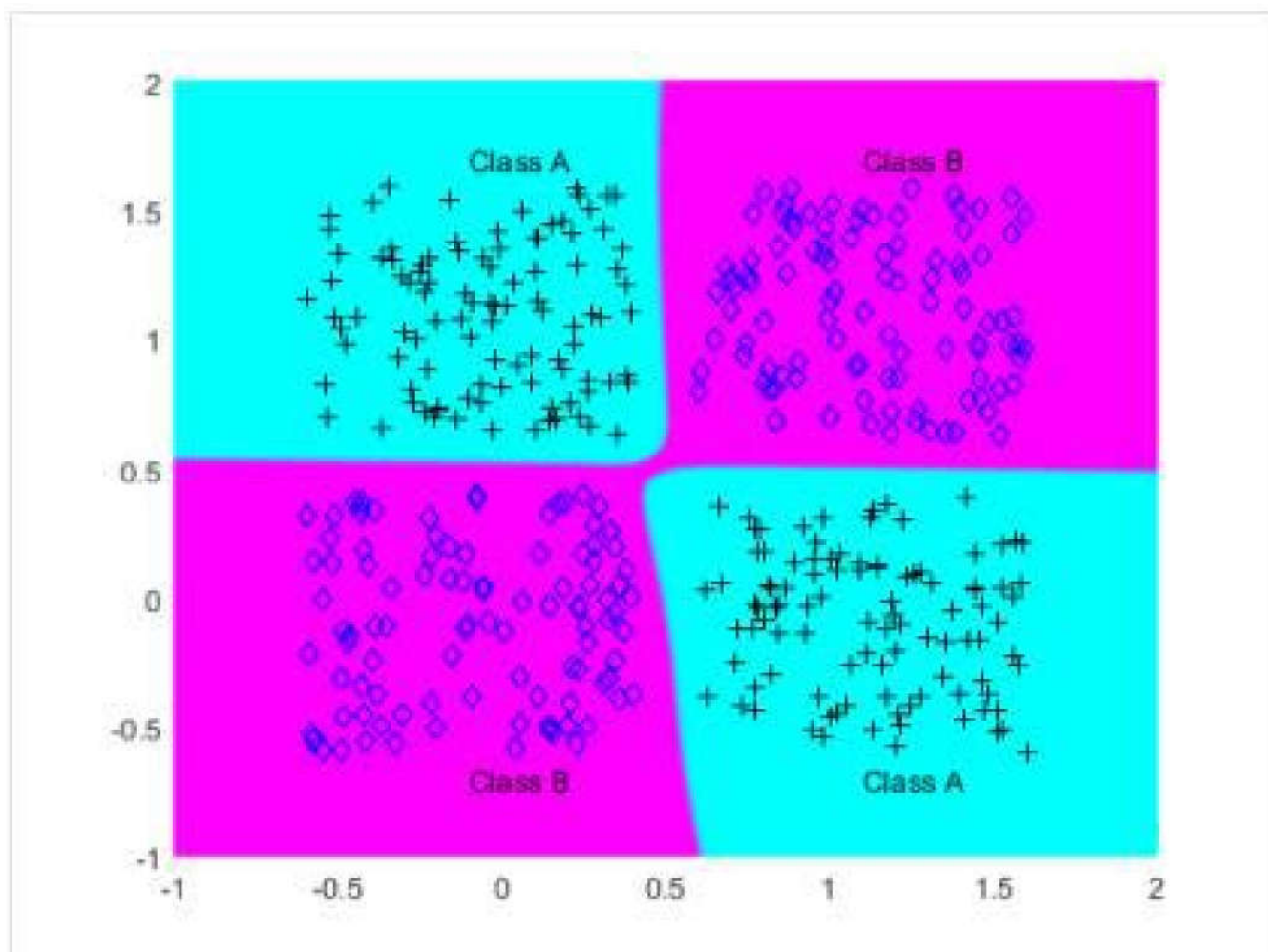
Δημιουργία Multilayer Perceptron για την ταξινόμηση

Σχεδιάζονται μία σειρά από εισόδους και η αντίστοιχη απόκριση (έξοδος) του ΤΝΔ για ναδειχθεί πόσο καλά το ΤΝΔ «μαθαίνει» τα δεδομένα εισόδου.



Δημιουργία Targets και Network response για έλεγχο του δίκτυο πάνω στα δεδομένα

(γ) Χάραξη των διαχωριστικών γραμμών:



Δημιουργία διαχωριστικών πεδίων με την βοήθεια των δύο χρωμάτων για την σωστή ταξινόμηση των δύο κλάσεων.

(δ) Έλεγχος

Σωστή ταξινόμηση άγνωστων δεδομένων εισόδου στις δύο κλάσεις, ανάλογα με τις διαχωριστικές γραμμές που χαράχθηκαν κατά τη φάση εκπαίδευσης.

#### 4.6.6 Βαθιά Μάθηση και Νευρωνικά Δίκτυα Βαθιάς Μάθησης

##### 1 Μηχανική μάθηση - Βαθιά μάθηση

Η Μηχανική Μάθηση (Machine Learning, ML) αναδύθηκε ως κλάδος της Επιστήμης των Υπολογιστών και της Τεχνητής Νοημοσύνης, με στόχο την ανάπτυξη μεθόδων που επιτρέπουν στους υπολογιστές να "μαθαίνουν" από εμπειρία χωρίς να είναι ρητά προγραμματισμένοι. Εστιάζει στην αναγνώριση προτύπων, την εξαγωγή γνώσης από

δεδομένα και την οικοδόμηση αλγορίθμων που επιτρέπουν στα συστήματα να προβλέπουν αποτελέσματα ή να λαμβάνουν αποφάσεις με βάση την εμπειρική ανάλυση δεδομένων (Alpaydin, 2020). Η θεμελιώδης ιδέα είναι ότι ένα σύστημα μπορεί να βελτιώνει τη συμπεριφορά του καθώς εκτίθεται σε περισσότερα δεδομένα, μαθαίνοντας από προηγούμενα παραδείγματα.

Οι αλγόριθμοι μηχανικής μάθησης δημιουργούν μοντέλα πρόβλεψης ή απόφασης μέσω εκπαίδευσης σε ιστορικά ή πειραματικά δεδομένα. Ένας από τους πιο ισχυρούς μηχανισμούς για την υλοποίηση αυτών των αλγορίθμων είναι τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ). Εμπνευσμένα από τη λειτουργία του ανθρώπινου εγκεφάλου, τα ΤΝΔ αποτελούνται από πλήρως διασυνδεδεμένες υπολογιστικές μονάδες – νευρώνες – που προσαρμόζουν τα βάρη των συνδέσεων τους κατά την εκπαίδευση. Η ικανότητά τους να μοντελοποιούν πολύπλοκες, μη γραμμικές σχέσεις τα καθιστά ιδιαίτερα κατάλληλα για προβλήματα όπως η αναγνώριση εικόνας, η αναγνώριση φωνής και η πρόβλεψη συμπεριφοράς (Goodfellow et.al., 2016).

Ο όρος "fuzzy" αναφέρεται στην έννοια της ασάφειας, δηλαδή της αδυναμίας να καθοριστεί με ακρίβεια αν μια δήλωση είναι απολύτως αληθής ή απολύτως ψευδής. Για παράδειγμα, η δήλωση "αυτός ο μαθητής είναι ψηλός" ενέχει υποκειμενικότητα και σχετικότητα και δεν μπορεί να προσδιοριστεί με αυστηρά δυαδικά κριτήρια. Η ανάγκη για χειρισμό τέτοιων ασαφών πληροφοριών, που είναι κοινές στην καθημερινότητα και στις ανθρώπινες αποφάσεις, οδήγησε στην ανάπτυξη της θεωρίας Ασαφούς Λογικής (Fuzzy Logic) (Zadeh, 1965).

Η Ασαφής Λογική (Fuzzy Logic) εισήχθη για να μοντελοποιήσει την ανθρώπινη αντίληψη και τη διαδικασία λήψης αποφάσεων υπό συνθήκες αβεβαιότητας. Η Fuzzy Logic δεν περιορίζεται σε δυαδικές τιμές (0 ή 1), αλλά επιτρέπει ενδιάμεσες τιμές της αλήθειας, ενισχύοντας την ευελιξία στη μοντελοποίηση προβλημάτων που περιλαμβάνουν γλωσσικούς όρους ή ποιοτικές κρίσεις. Ο αλγόριθμος βασίζεται στην αξιολόγηση βαθμών συμμετοχής ενός στοιχείου σε μία ή περισσότερες ασαφείς ομάδες και έχει εφαρμοστεί επιτυχώς σε συστήματα ελέγχου, αναγνώριση προτύπων, οικονομικά μοντέλα και ιατρικές διαγνώσεις (Ross, 2010).

Οι Support Vector Machines (SVM) συγκαταλέγονται στους αλγορίθμους μηχανικής μάθησης με επίβλεψη (supervised learning) και χρησιμοποιούνται κυρίως για προβλήματα ταξινόμησης (classification), αν και μπορούν να εφαρμοστούν και σε προβλήματα παλινδρόμησης (regression). Ο βασικός στόχος ενός SVM είναι να βρει το υπερεπίπεδο (hyperplane) που διαχωρίζει τα δεδομένα σε δύο ή περισσότερες κατηγορίες με το μέγιστο δυνατό περιθώριο (maximum margin), δηλαδή με τη μεγαλύτερη δυνατή απόσταση μεταξύ των σημείων των διαφορετικών κλάσεων και της γραμμής διαχωρισμού. Η ιδιαιτερότητα των SVM έγκειται στη χρήση kernel functions, που επιτρέπουν την αποτελεσματική χαρτογράφηση των δεδομένων σε πολυδιάστατους χώρους, επιλύοντας προβλήματα που δεν είναι γραμμικά διαχωρίσιμα στον αρχικό χώρο (Cortes & Vapnik, 1995).

Οι Γενετικοί Αλγόριθμοι (Genetic Algorithms – GA) ανήκουν στους αλγόριθμους βελτιστοποίησης εμπνευσμένους από τη θεωρία της εξέλιξης του Δαρβίνου. Βασίζονται στις έννοιες της επιλογής, της μετάλλαξης και της διασταύρωσης, που εφαρμόζονται πάνω σε πληθυσμούς λύσεων ώστε να αναπαραχθούν νέες, βελτιωμένες λύσεις μέσα από επαναλαμβανόμενους κύκλους (Mitchell, 1997). Οι γενετικοί αλγόριθμοι είναι ιδιαίτερα χρήσιμοι για προβλήματα αναζήτησης και βελτιστοποίησης σε περιβάλλοντα υψηλής πολυπλοκότητας και μη γραμμικότητας. Χάρη στη φύση τους, δεν απαιτούν εις των προτέρων γνώση των μαθηματικών ιδιοτήτων της λειτουργίας κόστους και μπορούν να αποφύγουν την παγίδευση σε τοπικά ελάχιστα, κάτι που αποτελεί πλεονέκτημα έναντι παραδοσιακών τεχνικών.

Η Μηχανική Μάθηση (Machine Learning – ML) και η Βαθιά Μάθηση (Deep Learning – DL) είναι δύο στενά συνδεδεμένοι τομείς που εντάσσονται κάτω από την ευρύτερη ομπρέλα της Τεχνητής Νοημοσύνης (Artificial Intelligence – AI). Η μηχανική μάθηση περιλαμβάνει αλγόριθμους που βασίζονται σε στατιστικές μεθόδους και απαιτούν από τους προγραμματιστές να παρέχουν δομημένα δεδομένα και συγκεκριμένα χαρακτηριστικά (features), πάνω στα οποία εκπαιδεύονται τα μοντέλα ώστε να πραγματοποιούν ταξινομήσεις ή προβλέψεις (Alpaydin, 2020). Οι αποφάσεις που λαμβάνονται βασίζονται σε κανόνες που παράγονται κατά τη διάρκεια της εκπαίδευσης μέσω ιστορικών δεδομένων.

Αντίθετα, η βαθιά μάθηση (DL) χρησιμοποιεί πολυεπίπεδα τεχνητά νευρωνικά δίκτυα (Deep Neural Networks – DNN), που έχουν την ικανότητα να εξάγουν αυτόματα πολύπλοκα χαρακτηριστικά από μη δομημένα δεδομένα όπως εικόνες, ήχο ή φυσική γλώσσα. Η βαθιά μάθηση επιτυγχάνει εκμάθηση αναπαραστάσεων (representation learning), επιτρέποντας στα μοντέλα να ανακαλύπτουν ιεραρχικές σχέσεις μεταξύ των δεδομένων χωρίς την ανάγκη ανθρώπινης παρέμβασης στον σχεδιασμό των χαρακτηριστικών (LeCun et. al., 2015). Αυτή η ιδιότητα καθιστά τα συστήματα βαθιάς μάθησης ιδιαίτερα ισχυρά σε εφαρμογές όπως η αυτόνομη οδήγηση, η αναγνώριση φωνής και η ιατρική διάγνωση.

Η βαθιά μάθηση εκπαιδεύει τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) με μια προσέγγιση που μοιάζει με τον ανθρώπινο εγκέφαλο, καθώς βασίζεται στη διαδοχική επεξεργασία των πληροφοριών μέσα από πολυεπίπεδες δομές. Σε αντίθεση με τη μηχανική μάθηση, όπου οι προγραμματιστές καθορίζουν ρητούς κανόνες ή επιλέγουν με το χέρι τα χαρακτηριστικά (features) που θα χρησιμοποιηθούν, η βαθιά μάθηση επιτρέπει στις μηχανές να αναλύουν ακατέργαστα δεδομένα και να μαθαίνουν μέσω αναπαραστάσεων που παράγονται αυτόματα. Έτσι, παρακάμπτεται η ανάγκη προκαθορισμένων κανόνων, κάτι που κάνει τα μοντέλα βαθιάς μάθησης εξαιρετικά ευέλικτα και ικανά να αντιμετωπίσουν περίπλοκα προβλήματα (Goodfellow et. al., 2016).

Τα μοντέλα βαθιάς μάθησης εκπαιδεύονται απευθείας μέσω μεγάλου όγκου και ποικιλίας δεδομένων που συλλέγουν και επεξεργάζονται – διαδικασία που απαιτεί υψηλούς

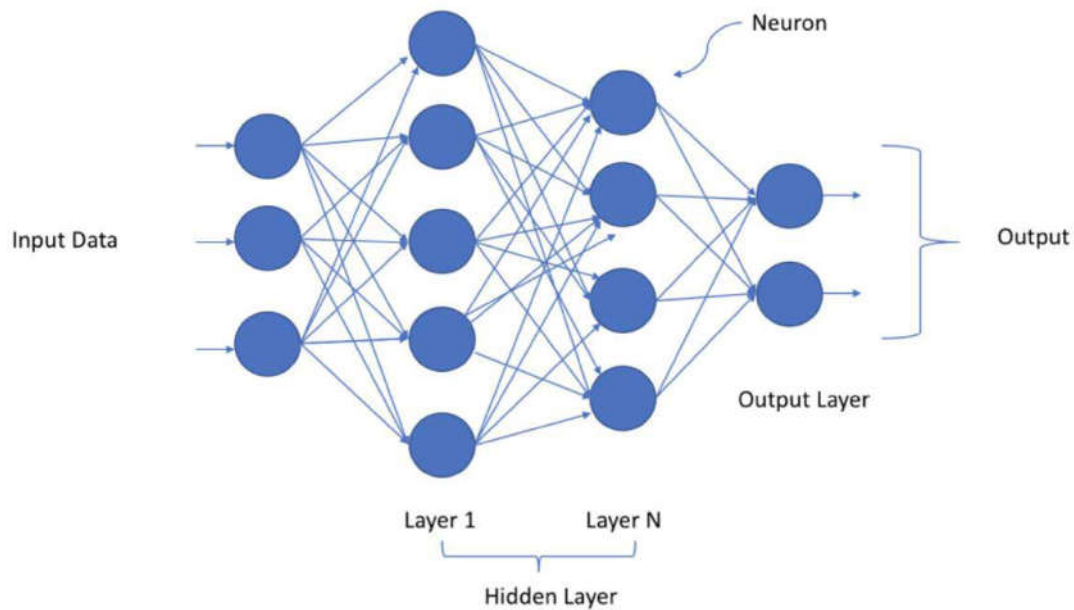
υπολογιστικούς πόρους αλλά οδηγεί σε εντυπωσιακά αποτελέσματα. Τα ίδια τα δεδομένα λειτουργούν ως «καύσιμο» για την εκπαίδευση αυτών των συστημάτων, τα οποία γίνονται ικανά να γενικεύουν και να προβλέπουν με υψηλή ακρίβεια, χωρίς ανθρώπινη επίβλεψη σε κάθε βήμα. Συνολικά, η βαθιά μάθηση αποτελεί την πιο προηγμένη πτυχή της τεχνητής νοημοσύνης, επιτρέποντας στις μηχανές να προσαρμόζονται σε πολύπλοκα περιβάλλοντα και να εξελίσσονται με τρόπο σχεδόν αυτόνομο (LeCun et. al., 2015).

Όλες αυτές οι εξελιγμένες δυνατότητες γίνονται εφικτές χάρη στη χρήση πολλαπλών στρωμάτων (layers) στα βαθιά νευρωνικά δίκτυα (Deep Neural Networks – DNN). Ένα δίκτυο που αποτελείται από τρία ή περισσότερα επίπεδα – εισόδου, πολλαπλά κρυφά στρώματα (hidden layers) και εξόδου – ονομάζεται «βαθύ». Αυτά τα κρυφά στρώματα επιτρέπουν στο δίκτυο να συλλαμβάνει πολυεπίπεδες και αφηρημένες αναπαραστάσεις των δεδομένων, με κάθε στρώμα να εξάγει ολοένα και πιο σύνθετα χαρακτηριστικά. Η δομή αυτή αποτελεί τον πυρήνα του μηχανισμού μάθησης και αποδίδει εξαιρετικά αποτελέσματα σε τομείς όπως η αναγνώριση εικόνας, η φυσική γλώσσα και η πρόβλεψη χρονοσειρών (Schmidhuber, 2015).

Συνολικά, οι τεχνικές της βαθιάς μάθησης και της μηχανικής μάθησης είναι στενά αλληλένδετες. Οι νευρωνικοί αλγόριθμοι αποτελούν τη βασική αρχιτεκτονική υποστήριξη της βαθιάς μάθησης, ενώ η εξέλιξη των μοντέλων αυτών μέσω της χρήσης πολλαπλών στρωμάτων επιτρέπει την αποτελεσματική ανάλυση και κατανόηση τεράστιων και περίπλοκων όγκων δεδομένων. Με τη συνεχή πρόοδο στον υπολογιστικό εξοπλισμό και τη διαθεσιμότητα δεδομένων, η βαθιά μάθηση αναμένεται να διαδραματίσει πρωταγωνιστικό ρόλο σε όλους τους τομείς της τεχνητής νοημοσύνης (Zhang et al., 2021).

## **2. Νευρωνικά Δίκτυα Βαθιάς Μάθησης**

Ένα βαθύ νευρωνικό δίκτυο (Deep Neural Network – DNN) αποτελεί μια απλοποιημένη αλλά ισχυρή αναπαράσταση της λειτουργίας του ανθρώπινου εγκεφάλου, βασισμένη σε μια ιεραρχική (στρωματοποιημένη) αρχιτεκτονική. Οι νευρώνες στα DNN είναι λογικές υπολογιστικές μονάδες που συνδέονται μεταξύ τους σχηματίζοντας ένα σύνθετο δίκτυο. Αυτοί οι νευρώνες επικοινωνούν μεταξύ τους μέσω συνδέσεων με αναλογικά βάρη, μεταδίδοντας σήματα ανάλογα με το ερέθισμα που λαμβάνουν, σε μια διαδικασία παρόμοια με την ανάδραση που συμβαίνει στο βιολογικό νευρικό σύστημα (Goodfellow et. al., 2016).



*Ένα βαθύ νευρωνικό δίκτυο με  $N$  κρυφά στρώματα.*

Η αρχική είσοδος (input) μεταφέρεται στο πρώτο στρώμα του δικτύου, που ονομάζεται στρώμα εισόδου. Από εκεί, τα δεδομένα περνούν διαδοχικά μέσω ενός ή περισσότερων κρυφών στρωμάτων (hidden layers), όπου κάθε στρώμα εξάγει μια πιο αφηρημένη αναπαράσταση των δεδομένων και τελικά φτάνουν στο στρώμα εξόδου (output layer). Η έξοδος του δικτύου μπορεί να πάρει διαφορετικές μορφές, όπως μια δυαδική πρόβλεψη (π.χ., "Ναι" ή "Όχι"), μια πιθανότητα, μια ταξινομητική κατηγορία ή μια αριθμητική τιμή, ανάλογα με το πρόβλημα που καλείται να επιλύσει το δίκτυο (LeCun et. al., 2015).

Κάθε νευρώνας εκτελεί μια βασική μαθηματική πράξη, η οποία περιλαμβάνει το άθροισμα των εισερχόμενων σημάτων πολλαπλασιασμένων με τα αντίστοιχα βάρη τους. Το αποτέλεσμα αυτής της πράξης περνά στη συνέχεια μέσα από μια λειτουργία ενεργοποίησης (activation function), όπως η sigmoid, η ReLU ή η tanh, που προσδιορίζει αν το σήμα θα μεταδοθεί περαιτέρω ή όχι. Οι λειτουργίες ενεργοποίησης είναι απαραίτητες για την εισαγωγή μη γραμμικότητας στο δίκτυο και καθιστούν δυνατή την επίλυση πολύπλοκων προβλημάτων (Nielsen, 2015).

Η ισχύς των DNN έγκειται στο γεγονός ότι τα βάρη μεταξύ των νευρώνων προσαρμόζονται κατά τη διάρκεια της φάσης της εκπαίδευσης, χρησιμοποιώντας αλγορίθμους όπως η οπισθοδρομική διάδοση (backpropagation) και η καθοδική βαθμίδωση (gradient descent). Μέσω αυτών, το δίκτυο ελαχιστοποιεί το σφάλμα πρόβλεψης συγκρίνοντας τις εξόδους με τις επιθυμητές τιμές και αναπροσαρμόζει τα βάρη του σταδιακά ώστε να μάθει το βέλτιστο μοντέλο πρόβλεψης (Schmidhuber, 2015).

Η διαφορά μεταξύ "ρηχών" και "βαθιών" νευρωνικών δικτύων βασίζεται στον αριθμό των κρυφών στρωμάτων. Τα ρηχά δίκτυα (Shallow Neural Networks – SNN) διαθέτουν μόνο ένα κρυφό στρώμα και είναι ικανά να επιλύσουν σχετικά απλά προβλήματα. Αντίθετα, τα βαθιά νευρωνικά δίκτυα (DNN), με περισσότερα από τρία στρώματα (εισόδου, εξόδου και τουλάχιστον δύο κρυφά), μπορούν να χειριστούν σύνθετα προβλήματα μη γραμμικής φύσης, εξαγοντας αυτόματα αφηρημένα χαρακτηριστικά από τα δεδομένα (Zhang et al., 2021).

Η έρευνα έχει δείξει ότι η αυξημένη βάθος (depth) του δικτύου οδηγεί σε βελτιωμένη ικανότητα γενίκευσης, μεγαλύτερη ακρίβεια και καλύτερη απόδοση σε σύνθετες εφαρμογές, όπως η αναγνώριση φωνής, εικόνas και φυσικής γλώσσας. Ωστόσο, αυτό συνεπάγεται αυξημένες υπολογιστικές απαιτήσεις και κινδύνους υπερεικπαίδευσης, καθιστώντας τη ρύθμιση των παραμέτρων (hyperparameter tuning) κρίσιμης σημασίας για την επιτυχία του μοντέλου (Aggarwal, 2018).

### **3. Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks)**

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNNs) αποτελούν έναν από τους πιο ισχυρούς και δημοφιλείς αλγορίθμους βαθιάς μάθησης, ιδιαίτερα κατάλληλους για εφαρμογές σε προβλήματα υπολογιστικής όρασης, όπως αναγνώριση εικόνas, εντοπισμός αντικειμένων και κατηγοριοποίηση. Η δομή τους είναι εμπνευσμένη από την οργάνωση του πρωτεύοντος οπτικού φλοιού στον ανθρώπινο εγκέφαλο, που είναι υπεύθυνος για την επεξεργασία οπτικών ερεθισμάτων (LeCun et al., 2015).

#### **• Στρώμα Συνέλιξης (Convolutional Layer)**

Το στρώμα συνέλιξης αποτελεί τον θεμέλιο λίθο των CNNs. Εκτελεί συνέλιξη (convolution) της εισόδου με φίλτρα (kernels), με σκοπό την εξαγωγή τοπικών χαρακτηριστικών από τα δεδομένα εικόνas, όπως ακμές, γωνίες ή μοτίβα υφής. Το πλεονέκτημα της συνέλιξης είναι ότι ελαττώνει τον αριθμό των παραμέτρων του δικτύου σε σύγκριση με πλήρως συνδεδεμένα δίκτυα, καθιστώντας το πιο αποτελεσματικό και ανθεκτικό σε υπερπροσαρμογή (Goodfellow et al., 2016).

#### **• Στρώμα Υποδειγματοληψίας (Pooling Layer)**

Το pooling layer μειώνει τις διαστάσεις των εξαγόμενων χαρακτηριστικών, περιορίζοντας ταυτόχρονα το υπολογιστικό κόστος και τον αριθμό των παραμέτρων. Η πιο κοινή τεχνική είναι το max pooling, που επιλέγει τη μέγιστη τιμή από κάθε περιοχή και ακολουθεί το average pooling, που υπολογίζει τον μέσο όρο. Η χρήση αυτών των στρωμάτων οδηγεί σε χωρική συμπίεση και αύξηση της γενικευσιμότητας του δικτύου (Zhang et al., 2021).

#### **• Στρώμα Πλήρως Διασυνδεδεμένων (Fully Connected Layer)**

Τα εξαγόμενα χαρακτηριστικά των προηγούμενων στρωμάτων προωθούνται σε ένα ή περισσότερα fully connected layers, στα οποία κάθε νευρώνας συνδέεται με όλους τους νευρώνες του επόμενου στρώματος. Στο τελικό στάδιο της επεξεργασίας, το δίκτυο λαμβάνει

μια τελική απόφαση, π.χ. ποια κλάση ανήκει η εικόνα. Τα βάρη που αντιστοιχούν στις συνδέσεις προσαρμόζονται μέσω διαδικασιών μάθησης, όπως η οπισθοδρομική διάδοση (backpropagation) (Aggarwal, 2018).

### • Συναρτήσεις Ενεργοποίησης

Για την εισαγωγή **μη γραμμικότητας** στο δίκτυο, χρησιμοποιούνται activation functions. Η ReLU (Rectified Linear Unit) είναι η πιο διαδεδομένη, καθώς παρέχει υπολογιστική αποδοτικότητα και καταπολεμά το πρόβλημα της εξαφάνισης του βαθμωτού σφάλματος (vanishing gradient). Εναλλακτικά, χρησιμοποιούνται η sigmoid, η tanh, ή νεότερες όπως η Leaky ReLU, ειδικά σε βαθύτερα δίκτυα (Nielsen, 2015).

### • Εκπαίδευση

Κατά τη φάση της εκπαίδευσης, τα CNNs μαθαίνουν προσαρμόζοντας τα βάρη τους έτσι ώστε να ελαχιστοποιούν το σφάλμα μεταξύ προβλεπόμενης και πραγματικής εξόδου. Χρησιμοποιούνται αλγόριθμοι βελτιστοποίησης όπως ο SGD (Stochastic Gradient Descent), το Adam και το RMSprop, οι οποίοι προσαρμόζουν τα βάρη με βάση τον βαθμό σφάλματος σε κάθε βήμα (Kingma & Ba, 2014).

### • Μεταφορά Μάθησης (Transfer Learning)

Τα CNNs είναι ιδιαίτερα κατάλληλα για μεταφορά μάθησης, μια τεχνική κατά την οποία προεκπαιδευμένα μοντέλα (όπως τα VGG16, ResNet, Inception) επαναχρησιμοποιούνται και προσαρμόζονται για νέες εργασίες. Αυτή η προσέγγιση μειώνει τον χρόνο εκπαίδευσης και απαιτεί μικρότερη ποσότητα δεδομένων, καθιστώντας τα CNNs εξαιρετικά αποδοτικά για πρακτικές εφαρμογές (Yosinski et al., 2014).

Τα Convolutional Neural Networks (CNNs) αποτελούν βασική υποκατηγορία των βαθιών νευρωνικών δικτύων (Deep Neural Networks - DNNs) και έχουν φέρει επανάσταση στον τομέα της υπολογιστικής όρασης (computer vision), χάρη στη δυνατότητά τους να αναγνωρίζουν μοτίβα σε δεδομένα εικόνες με ελάχιστη ανάγκη για προεπεξεργασία. Η αρχιτεκτονική τους, η οποία βασίζεται στη λειτουργία του ανθρώπινου οπτικού φλοιού, επιτρέπει την εξαγωγή χαρακτηριστικών από εικόνες μέσω στρωμάτων συνέλιξης και υποδειγματοληψίας (LeCun et al., 2015).

Ένα από τα πλέον γνωστά πεδία εφαρμογής των CNNs είναι η αναγνώριση αντικειμένων και προσώπων σε εικόνες και βίντεο. Οι αλγόριθμοι CNN έχουν χρησιμοποιηθεί σε λογισμικά ταυτοποίησης προσώπου (όπως το FaceNet της Google), σε συστήματα ασφαλείας και σε έξυπνες κάμερες επιτήρησης, αποδεικνύοντας εξαιρετικά υψηλά ποσοστά ακρίβειας (Parkhi et al., 2015).

Στον τομέα της αυτόνομης οδήγησης, τα CNNs είναι κρίσιμα για τη λειτουργία συστημάτων αντίληψης. Αναλύοντας ροές βίντεο από κάμερες τοποθετημένες στο όχημα, τα CNNs εντοπίζουν πεζούς, πινακίδες, οχήματα και εμπόδια, καθιστώντας δυνατή την αντίδραση του οχήματος σε πραγματικό χρόνο (Bojarski et al., 2016). Αυτή η ικανότητα καθιστά τα CNNs απαραίτητα για τις εταιρείες που αναπτύσσουν αυτόνομα ή υποβοηθούμενα συστήματα οδήγησης, όπως η Tesla, η Waymo και η Nvidia.

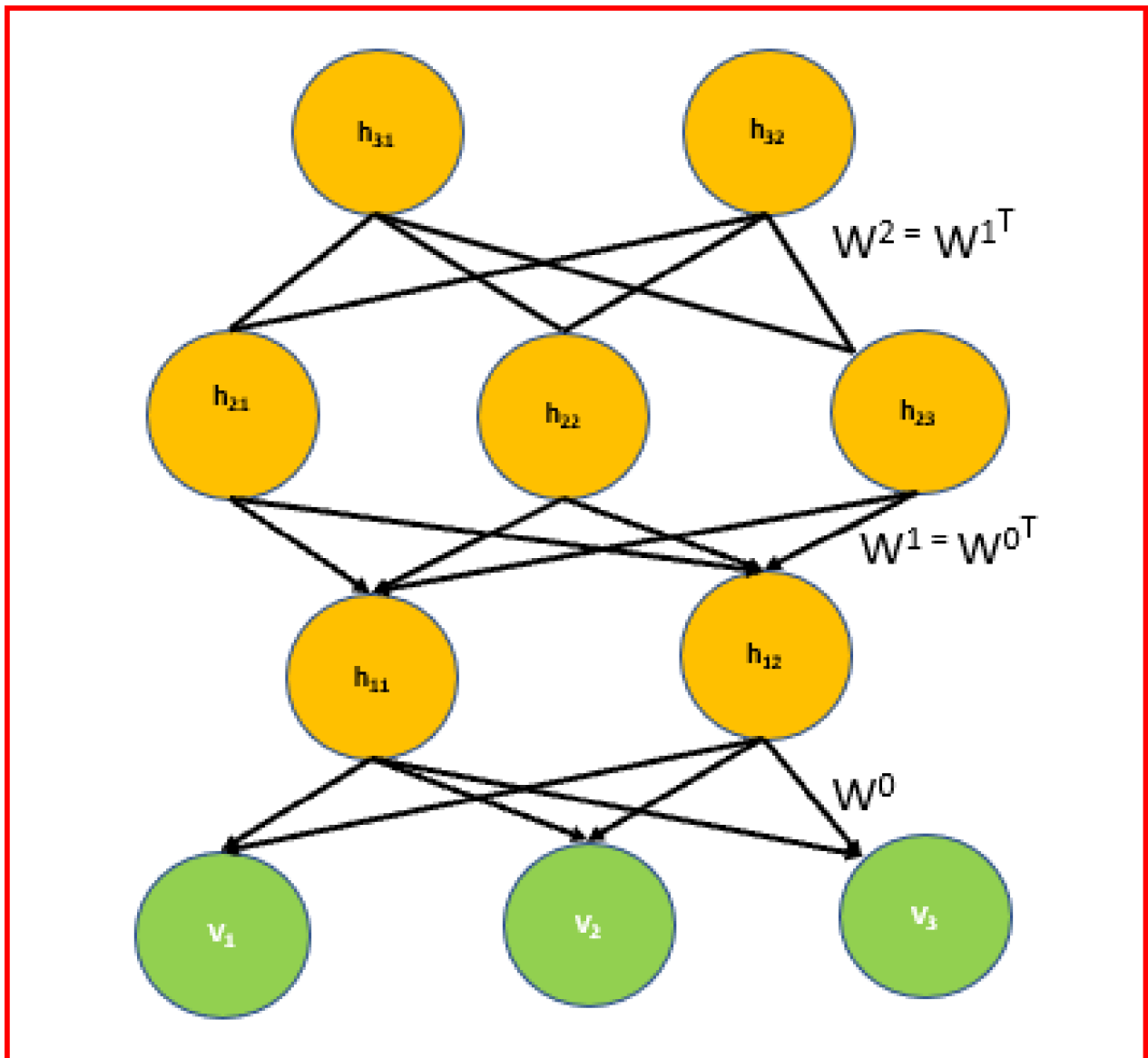
Εξίσου εντυπωσιακές είναι οι εφαρμογές των CNNs στην υγειονομική περιθαλψη, ειδικά στην ιατρική απεικόνιση. Τα CNNs χρησιμοποιούνται για την ανάλυση ακτινογραφιών, μαγνητικών τομογραφιών (MRI), αξονικών τομογραφιών (CT scans) και άλλων ειδών ιατρικής εικόνας. Έχουν καταγράψει σημαντικές επιτυχίες στην έγκαιρη ανίχνευση ασθενειών, όπως ο καρκίνος, η διαβητική αμφιβληστροειδοπάθεια και τα εγκεφαλικά επεισόδια (Esteva et al., 2017). Τα CNNs προσφέρουν συγκρίσιμη ή και ανώτερη ακρίβεια σε σχέση με τους γιατρούς, ειδικά σε συγκεκριμένες διαγνωστικές περιπτώσεις.

Επιπλέον, τα CNNs βρίσκουν εφαρμογή και σε συστήματα αυτόματης μετάφρασης εικόνων σε κείμενο, όπως στη μετάφραση χειρόγραφων ή έντυπων εγγράφων (OCR), καθώς και σε συστήματα αναγνώρισης συναισθημάτων από εκφράσεις προσώπου ή χειρονομίες. Επίσης, έχουν αξιοποιηθεί στην κατάτμηση εικόνων (image segmentation) και στην ανακατασκευή 3D μοντέλων από 2D εικόνες.

Κατά την επεξεργασία μιας εικόνας, τα CNNs αναλύουν την είσοδο σε μικρές περιοχές (patches) και αναγνωρίζουν τοπικά χαρακτηριστικά όπως ακμές, γωνίες ή υφές, τα οποία συνδυάζουν προοδευτικά για να κατανοήσουν σύνθετες δομές. Με αυτόν τον τρόπο, η αναπαράσταση της εικόνας γίνεται ιεραρχικά, γεγονός που καθιστά το δίκτυο ικανό να εντοπίσει με ακρίβεια αντικείμενα και σχέσεις μεταξύ τους (Goodfellow et al., 2016).

#### **4. Deep Belief Neural Networks**

Τα Δυσδικά Νευρωνικά Δίκτυα (Deep Belief Neural Networks – DBNNs) ανήκουν στην ευρύτερη κατηγορία των βαθιών νευρωνικών δικτύων και έχουν διαμορφωθεί με βάση την αρχιτεκτονική και λειτουργικότητα του ανθρώπινου εγκεφάλου. Το κύριο χαρακτηριστικό τους είναι η χρήση πιθανολογικών μοντέλων **και** εκπαίδευσης χωρίς επίβλεψη, με σκοπό τη μάθηση λανθάνουσών αναπαραστάσεων (latent representations) των δεδομένων. Τα DBNNs βασίζονται σε δυσδικές λανθάνουσες μεταβλητές και αποτελούνται από στοιβάδες τόσο κατευθυνόμενων όσο και μη κατευθυνόμενων συνδέσεων, καθιστώντας τα κατάλληλα για σύνθετες εργασίες αναγνώρισης προτύπων και γενίκευσης πληροφορίας (Hinton et al., 2006).



Αρχιτεκτονική DBNN

Σε αντίθεση με άλλες μορφές βαθιάς μάθησης, όπως τα Συνελικτικά Νευρωνικά Δίκτυα (CNNs), τα DBNN έχουν την ιδιαιτερότητα ότι κάθε επίπεδο του δικτύου μπορεί να μάθει ολιστικά χαρακτηριστικά της εισόδου. Ενώ στα CNNs τα πρώτα στρώματα εντοπίζουν απλά χαρακτηριστικά όπως ακμές και τα επόμενα στρώματα ενώνουν αυτά τα χαρακτηριστικά σε πιο πολύπλοκες μορφές, στα DBNN κάθε επίπεδο αντιμετωπίζει την είσοδο ως ενιαίο σύνολο, επαναλαμβάνοντας τη διαδικασία μάθησης με τρόπο πιο γενικευμένο και αφηρημένο (Bengio, 2009).

Η αρχιτεκτονική ενός DBNN αποτελείται από μια στοίβα Μειωμένων Μηχανών Boltzmann (Restricted Boltzmann Machines – RBMs). Κάθε RBM μαθαίνει ένα απλοποιημένο μοντέλο των εισόδων και περνάει τα αποτελέσματά του ως είσοδο στο επόμενο επίπεδο. Οι κόμβοι σε διαφορετικά στρώματα είναι πλήρως συνδεδεμένοι, αλλά δεν υπάρχουν πλευρικές

συνδέσεις στο ίδιο επίπεδο. Οι συνδέσεις στα υψηλότερα επίπεδα είναι μη κατευθυντικές, συμβάλλοντας στη δημιουργία συσχετιστικών αναπαραστάσεων, ενώ στα κατώτερα επίπεδα οι συνδέσεις είναι κατευθυντικές, επιτρέποντας τη μεταφορά πληροφορίας από τα δεδομένα προς τα χαρακτηριστικά (Hinton et al., 2006).

Το DBNN βασίζεται στην τεχνική greedy layer-wise pretraining, η οποία περιλαμβάνει την ανεξάρτητη εκπαίδευση κάθε επιπέδου (ή RBM) με στόχο τη βελτιστοποίηση των πιθανολογικών κατανομών των δεδομένων. Αυτή η τεχνική επιτρέπει σε κάθε επίπεδο να μάθει σημαντικά χαρακτηριστικά χωρίς την ανάγκη για εποπτευόμενη μάθηση, μειώνοντας παράλληλα τον κίνδυνο υπερπροσαρμογής. Μετά την προ-εκπαίδευση, το δίκτυο μπορεί να "βελτιστοποιηθεί" με χρήση backpropagation, εφόσον διατίθενται επισημασμένα δεδομένα (Erhan et al., 2010).

Οι greedy αλγόριθμοι είναι καθοριστικοί για την επιτυχία των DBNN, καθώς επιτρέπουν την ιεραρχική εκμάθηση με μειωμένες υπολογιστικές απαιτήσεις. Η διαδικασία ξεκινά από το κατώτερο επίπεδο και προχωρά προς τα ανώτερα, χρησιμοποιώντας την έξοδο κάθε στρώματος ως είσοδο για το επόμενο. Αυτό καθιστά το μοντέλο πολύ ευέλικτο και ικανό να εντοπίζει πολύπλοκα πρότυπα και συσχετίσεις στα δεδομένα, γεγονός που ενισχύει την απόδοσή του σε εφαρμογές όπως η αναγνώριση εικόνων, η ανάλυση κειμένου και τα συστήματα σύστασης (Larochelle et al., 2009).

## **5. Εφαρμογές των DBNN**

Τα Δίκτυα Βαθιάς Νευρωνικής Διαδικασίας (Deep Belief Neural Networks – DBNN) έχουν εφαρμογές σε πληθώρα τομέων λόγω της ικανότητάς τους να εξάγουν σύνθετα χαρακτηριστικά και να διαχειρίζονται μεγάλα και ετερογενή δεδομένα. Η ικανότητά τους να μαθαίνουν ιεραρχικές αναπαραστάσεις τους καθιστά εξαιρετικά χρήσιμα σε πεδία που απαιτούν υψηλή ακρίβεια, προσαρμοστικότητα και δυνατότητα γενίκευσης.

### **Επεξεργασία Εικόνων και Βίντεο:**

Τα DBNN χρησιμοποιούνται ευρέως στην αναγνώριση αντικειμένων και στην κατηγοριοποίηση εικόνων. Μέσω της στοίβαξης πολλαπλών Restricted Boltzmann Machines (RBMs), μπορούν να αναγνωρίσουν αφηρημένα μοτίβα και να δημιουργήσουν ισχυρές αναπαραστάσεις που χρησιμοποιούνται στην ταξινόμηση εικόνων και στον εντοπισμό αντικειμένων. Ενδείκνυνται ιδιαίτερα για εφαρμογές υπολογιστικής όρασης, ακόμα και σε δυναμικά περιβάλλοντα, όπως είναι η επεξεργασία βίντεο (Lee et al., 2009).

### **Επεξεργασία Φυσικής Γλώσσας (NLP)**

Στον τομέα της φυσικής γλώσσας, τα DBNN εφαρμόζονται στην αναγνώριση και κατανόηση κειμένου, όπως στην ανάλυση συναισθήματος, στην παραγωγή γλώσσας και στην αυτόματη μετάφραση. Η ικανότητά τους να εντοπίζουν συσχετίσεις και αφηρημένες έννοιες χωρίς να απαιτούν εποπτεία τα καθιστά πολύτιμα για πολυγλωσσικές και πολυσήμαντες εφαρμογές (Larochelle et al., 2012).

## Ιατρικές Εφαρμογές

Στην ιατρική, τα DBNN έχουν αναδειχθεί ως ισχυρά εργαλεία στην ανάλυση ιατρικών εικόνων (όπως ακτινογραφίες, μαγνητικές τομογραφίες και υπερηχογραφήματα), επιτρέποντας τον εντοπισμό παθολογικών περιοχών και την υποστήριξη της διάγνωσης ασθενειών βάσει κλινικών δεδομένων. Ειδικά στη διάγνωση καρκίνων ή καρδιοπαθειών, η εφαρμογή βαθιάς εκμάθησης έχει μειώσει την ανάγκη για ανθρώπινη επισκόπηση σε πρώιμα στάδια διάγνωσης (Shen et al., 2017).

## Αυτόνομη Οδήγηση και Ρομποτική

Στον χώρο της ρομποτικής και της αυτόνομης οδήγησης, τα DBNN αξιοποιούνται για την ανίχνευση αντικειμένων, την κατηγοριοποίηση σκηνών και την πλοήγηση σε απρόβλεπτα περιβάλλοντα. Μέσω της δυναμικής τους μάθησης, τα δίκτυα προσαρμόζονται σε νέες καταστάσεις και επιτρέπουν στα συστήματα να λαμβάνουν βέλτιστες αποφάσεις με ελάχιστη ανθρώπινη παρέμβαση (Tian et al., 2022).

## Χρηματοπιστωτικές Εφαρμογές

Στις χρηματοπιστωτικές αγορές, τα DBNN χρησιμοποιούνται στην πρόβλεψη τιμών μετοχών, στη βελτιστοποίηση επενδυτικών στρατηγικών, αλλά και στην ανίχνευση απάτης. Η ικανότητά τους να ανακαλύπτουν πολύπλοκα μοτίβα σε χρονοσειρές και να μαθαίνουν από μεγάλες ιστορικές βάσεις δεδομένων τους καθιστά εξαιρετικά πολύτιμους συνεργάτες για αλγοριθμικούς επενδυτές και τράπεζες (Heaton et al., 2016).

Τα Δίκτυα Βαθιάς Νευρωνικής Διαδικασίας, χάρη στην ευελιξία, την επεκτασιμότητα και την ικανότητα εκμάθησης αφηρημένων εννοιών, βρίσκουν εφαρμογή σε ένα ευρύ φάσμα επιστημονικών, τεχνολογικών, οικονομικών και κοινωνικών τομέων. Η ικανότητά τους να συνδυάζουν χωρικά και χρονικά χαρακτηριστικά, να μαθαίνουν από τα δεδομένα και να γενικεύουν αποτελέσματα σε νέες περιπτώσεις καθιστά την τεχνολογία αυτή ακρογωνιαίο λίθο της σύγχρονης τεχνητής νοημοσύνης.

### 4.6.7 Συμπεράσματα

Τα νευρωνικά δίκτυα αποτελούν μια συναρπαστική πτυχή της εξέλιξης της τεχνητής νοημοσύνης (AI), ένα διεπιστημονικό πεδίο που συνδυάζει τη βιολογία, τη στατιστική και τη μηχανική μάθηση για τη δημιουργία ευφύων και ευέλικτων συστημάτων (Goodfellow et al., 2016). Μετασχηματίζοντας τον τρόπο με τον οποίο επεξεργαζόμαστε τις πληροφορίες, τα νευρωνικά δίκτυα ανοίγουν νέους ορίζοντες στην τεχνολογία και επαναπροσδιορίζουν την ανθρώπινη αλληλεπίδραση με τα συστήματα πληροφορικής (LeCun et. al., 2015).

Στη ρίζα των νευρωνικών δικτύων βρίσκεται η προσπάθεια αναπαράστασης της λειτουργίας του ανθρώπινου εγκεφάλου. Αντί να χρησιμοποιούμε συμβατικούς κανόνες λογικής και ακολουθιακούς αλγόριθμους, τα τεχνητά νευρωνικά δίκτυα βασίζονται στη σύνδεση μεταξύ τεχνητών νευρώνων, προσομοιώνοντας τις συνάψεις των βιολογικών νευρώνων, με στόχο την αναγνώριση προτύπων και την πρόβλεψη αποτελεσμάτων (Haykin, 2009). Αυτή η δομή παρέχει μια εντελώς νέα προσέγγιση στην υπολογιστική, που βασίζεται στην μάθηση από τα δεδομένα αντί στην προγραμματισμένη λογική.

Η ανάπτυξη των νευρωνικών δικτύων ξεπέρασε τη θεωρητική μελέτη, εδραιώνοντας την πρακτική τους χρήση ως βασικό υπόβαθρο της σύγχρονης τεχνητής νοημοσύνης. Με την ικανότητά τους να μαθαίνουν από τα δεδομένα, να προσαρμόζονται σε νέες πληροφορίες και να αυτοβελτιώνονται, τα Τεχνητά Νευρωνικά Δίκτυα (ANNs) προσφέρουν ευελιξία και ισχύ σε πεδία όπως η αναγνώριση εικόνων, η ανάλυση φυσικής γλώσσας και η πρόβλεψη ακολουθιών (Schmidhuber, 2015).

Η τεχνητή νοημοσύνη δεν αποτελεί πλέον μια επιστημονική φαντασία, αλλά μια τεχνολογική πραγματικότητα που αναδιαμορφώνει την καθημερινότητα. Τα νευρωνικά δίκτυα χρησιμοποιούνται σε εφαρμογές όπως η αυτόνομη οδήγηση, τα συστήματα σύστασης περιεχομένου και η ανάλυση φωνής, αποτελώντας τον ακρογωνιαίο λίθο της εξατομικευμένης εμπειρίας χρήστη και της αυτοματοποιημένης λήψης αποφάσεων (Russell & Norvig, 2020). Χάρη στην ικανότητά τους να αναγνωρίζουν μοτίβα και να γενικεύουν από τα δεδομένα, καθίστανται ισχυρά εργαλεία κατανόησης και πρόβλεψης της ανθρώπινης συμπεριφοράς.

Ωστόσο, η επιτυχία των νευρωνικών δικτύων συνοδεύεται και από προκλήσεις. Η ερμηνευσιμότητα των αποτελεσμάτων τους, η ανάγκη για διαφάνεια στους αλγόριθμους, καθώς και η διασφάλιση της ιδιωτικότητας και της ηθικής χρήσης των δεδομένων, αποτελούν κρίσιμα ζητήματα στον δημόσιο διάλογο (Doshi-Velez & Kim, 2017). Παράλληλα, η ραγδαία αύξηση της υπολογιστικής ισχύος και η πρόοδος στην αρχιτεκτονική των Βαθιών Νευρωνικών Δικτύων (DNN) δημιουργούν το υπόβαθρο για τη συνεχή εξέλιξη του πεδίου.

Σήμερα, τα νευρωνικά δίκτυα βρίσκονται στην αιχμή της τεχνολογικής επανάστασης, καθοδηγώντας την καινοτομία στην υπολογιστική όραση, τη ρομποτική, τη βιοϊατρική μηχανική και πολλούς άλλους τομείς (Aggarwal, 2018). Η προσαρμοστικότητα τους επιτρέπει τη διαχείριση δυναμικών, πολυδιάστατων και ατελών δεδομένων, επιτυγχάνοντας αποδόσεις που ξεπερνούν τις κλασικές μεθόδους μηχανικής μάθησης.

Τελικά, τα νευρωνικά δίκτυα αποτελούν μια γέφυρα μεταξύ της βιολογίας και της τεχνολογίας – ένα βήμα πιο κοντά στην ψηφιακή αναπαραγωγή της ανθρώπινης ευφυΐας. Καθώς η επιστήμη και η τεχνολογία προχωρούν, τα δίκτυα αυτά ανοίγουν τον δρόμο για νέες δυνατότητες στην εξερεύνηση της τεχνητής νοημοσύνης και την εφαρμογή της σε ζητήματα παγκόσμιας σημασίας – από την υγεία και την ενέργεια μέχρι την εκπαίδευση και το περιβάλλον.

## *Κεφάλαιο 5: Μέσα Κοινωνικής Δικτύωσης και Ανάλυση Δεδομένων*

### *5.1 Δομή και Ιδιαιτερότητες των Δεδομένων των Κοινωνικών Δικτύων*

Τα μέσα κοινωνικής δικτύωσης έχουν μετασχηματίσει ριζικά τον τρόπο με τον οποίο οι άνθρωποι επικοινωνούν, αλληλεπιδρούν και μοιράζονται πληροφορίες. Στο πλαίσιο αυτό, έχει παρατηρηθεί μία εκρηκτική αύξηση του όγκου, της ποικιλίας και της ταχύτητας με την οποία παράγονται δεδομένα. Αυτά τα δεδομένα δεν είναι απλώς μεγάλου όγκου· φέρουν χαρακτηριστικά που καθιστούν τη διαχείριση και ανάλυσή τους ιδιαίτερα απαιτητική. Χαρακτηρίζονται από έντονη ετερογένεια, δυναμικότητα, πολυπλοκότητα, αλλά και υψηλό βαθμό αβεβαιότητας (González-Bailón, 2013).

Στον πυρήνα της δομής αυτών των δεδομένων βρίσκεται η θεμελιώδης αντίληψη του γράφου. Τα κοινωνικά δίκτυα μοντελοποιούνται συχνά ως γράφοι, στους οποίους οι κόμβοι αναπαριστούν χρήστες ή αντικείμενα (όπως αναρτήσεις, εικόνες ή hashtags), ενώ οι ακμές αποτυπώνουν τις μεταξύ τους σχέσεις και αλληλεπιδράσεις. Οι σχέσεις αυτές μπορεί να είναι κατευθυνόμενες, όπως στην περίπτωση ενός χρήστη που ακολουθεί έναν άλλον στο Twitter, ή αμφίδρομες, όπως η αμοιβαία «φιλία» στο Facebook. Επιπλέον, οι ακμές μπορεί να έχουν βάρη που αντιπροσωπεύουν την ένταση ή τη συχνότητα των αλληλεπιδράσεων (Wasserman & Faust, 1994).

Η ετερογένεια των δεδομένων είναι ένα από τα πλέον χαρακτηριστικά και ταυτόχρονα απαιτητικά στοιχεία των κοινωνικών δικτύων. Το περιεχόμενο που αναρτούν ή διαμοιράζονται οι χρήστες περιλαμβάνει κείμενα, εικόνες, βίντεο, γεωγραφικές συντεταγμένες, hashtags και άλλες ετικέτες μεταδεδομένων. Για παράδειγμα, μία ανάρτηση στο Instagram μπορεί να περιλαμβάνει φωτογραφία, κείμενο, hashtags, σχόλια και αντιδράσεις· όλα αυτά αποτελούν διαφορετικούς τύπους δεδομένων, που απαιτούν διαφορετικές τεχνικές επεξεργασίας (Cambria & White, 2014). Η πολυτροπική φύση των δεδομένων αυτών προϋποθέτει τη χρήση εξειδικευμένων τεχνικών συνδυασμού πληροφοριών από πολλαπλά κανάλια (multimodal data fusion), γεγονός που αυξάνει την πολυπλοκότητα του υπολογιστικού προβλήματος.

Ένα εξίσου σημαντικό χαρακτηριστικό των κοινωνικών δικτύων είναι η δυναμικότητα της δομής τους. Σε αντίθεση με στατικά δίκτυα, τα κοινωνικά δίκτυα μεταβάλλονται διαρκώς: νέοι χρήστες εισέρχονται, άλλοι αποχωρούν, ενώ οι μεταξύ τους σχέσεις επαναπροσδιορίζονται συνεχώς. Οι αλληλεπιδράσεις που λαμβάνουν χώρα είναι χρονικά ευαίσθητες και το πλαίσιο της πληροφορίας ενδέχεται να αλλάξει ριζικά μέσα σε μικρά χρονικά διαστήματα. Το χαρακτηριστικό αυτό επιβάλλει την ανάγκη για ανάλυση δεδομένων σε πραγματικό χρόνο και για την αξιοποίηση μοντέλων ανίχνευσης ροών δεδομένων (data

stream mining) που μπορούν να διαχειριστούν εισερχόμενη πληροφορία χωρίς να απαιτείται η πλήρης αποθήκευση του παρελθόντος (Aggarwal, 2011).

Η αβεβαιότητα αποτελεί μια ακόμη παράμετρο που διαφοροποιεί τα δεδομένα των κοινωνικών μέσων από άλλες παραδοσιακές μορφές πληροφορίας. Εξαιτίας της ανθρώπινης φύσης των χρηστών, τα δεδομένα συχνά περιέχουν θόρυβο, ασάφεια και υποκειμενικότητα. Οι εκφράσεις συναισθήματος, η χρήση ειρωνείας ή η υιοθέτηση νεολογισμών δυσχεραίνουν την αυτοματοποιημένη κατανόηση και κατηγοριοποίηση των αναρτήσεων. Επιπλέον, οι ίδιες λέξεις ή hashtags μπορεί να φέρουν διαφορετικό νοηματικό φορτίο, ανάλογα με το κοινωνικό ή πολιτισμικό υπόβαθρο των χρηστών και το χρονικό πλαίσιο χρήσης τους (Pak & Paroubek, 2010).

Η θεωρία γράφων παρέχει τη βασική γλώσσα μοντελοποίησης και ανάλυσης των δικτύων αυτών. Ένας γράφος μπορεί να αναπαρασταθεί με διάφορες μορφές, όπως ο πίνακας γειτνίασης (adjacency matrix), όπου κάθε στοιχείο δείχνει την ύπαρξη ή μη σύνδεσης μεταξύ δύο κόμβων· ο πίνακας συσχέτισης ακμών (incidence matrix), ο οποίος συνδέει κόμβους και ακμές· και η απλή λίστα ακμών (edge list). Η επιλογή της αναπαράστασης εξαρτάται από τη φύση του προβλήματος και τις απαιτήσεις της ανάλυσης.

Προχωρώντας σε πιο προχωρημένες τεχνικές, αναπτύσσονται μέτρα κεντρικότητας για να αξιολογηθεί η σημασία ενός κόμβου μέσα στο δίκτυο. Η «κεντρικότητα βαθμού» (degree centrality) βασίζεται στον αριθμό των συνδέσεων ενός κόμβου, υποδεικνύοντας τη δημοτικότητά του. Αντίθετα, η «μεταξύτητα» (betweenness centrality) εστιάζει στον ρόλο του κόμβου ως ενδιάμεσος στη ροή πληροφορίας· ένα υψηλό επίπεδο μεταξύτητας υποδηλώνει στρατηγική θέση μέσα στο δίκτυο. Το «συντελεστής συσσωμάτωσης» (clustering coefficient) αναφέρεται στην τάση δημιουργίας τριγώνων εντός του δικτύου, δηλαδή στο πόσο πιθανό είναι οι φίλοι ενός χρήστη να είναι επίσης φίλοι μεταξύ τους. Αυτά τα μέτρα επιτρέπουν την αναγνώριση χρηστών με επιρροή, τον εντοπισμό κοινοτήτων, αλλά και την κατανόηση των μοτίβων διάδοσης πληροφορίας (Wasserman & Faust, 1994).

Ενδιαφέρον παρουσιάζει και η ανάλυση της ίδιας της πλατφόρμας προέλευσης των δεδομένων. Διαφορετικά κοινωνικά δίκτυα προσφέρουν διαφορετικές δομές και μορφές πληροφορίας. Για παράδειγμα, το Twitter διαθέτει κατευθυνόμενους γράφους μέσω της έννοιας του «ακολουθήσε», ενώ το Facebook δημιουργεί αμφίδρομους δεσμούς μέσω της έννοιας της φιλίας. Το Instagram στηρίζεται κυρίως σε οπτικά δεδομένα, ενώ το TikTok δίνει έμφαση στη διάδοση μέσω ήχου και χρήσης βίντεο, δημιουργώντας πολυδιάστατες μορφές γράφων. Η μελέτη αυτών των δομών είναι καθοριστική για την επιλογή των κατάλληλων τεχνικών επεξεργασίας, ανάλυσης και μοντελοποίησης.

Οι τεχνικές προκλήσεις που απορρέουν από τα παραπάνω είναι πολυάριθμες. Η αποθήκευση και επεξεργασία δεδομένων σε πραγματικό χρόνο είναι συχνά ανέφικτη χωρίς προηγμένες τεχνικές κατανεμημένου υπολογισμού και αποδοτικών αλγορίθμων φιλτραρίσματος. Η προστασία της ιδιωτικότητας και των προσωπικών δεδομένων αποτελεί έναν διαρκή ηθικό και νομικό προβληματισμό. Επιπλέον, η ανίχνευση και απομόνωση ψευδούς πληροφορίας,

bots και spammers καθίσταται επιτακτική, καθώς επηρεάζει άμεσα την ποιότητα της πληροφορίας και την αξιοπιστία των αναλύσεων.

Συνολικά, η δομή και οι ιδιαιτερότητες των δεδομένων των κοινωνικών δικτύων συγκροτούν ένα σύνθετο πλαίσιο που απαιτεί τη διαθεματική χρήση τεχνικών από τη θεωρία γράφων, τη μηχανική μάθηση, την επεξεργασία φυσικής γλώσσας, αλλά και τη στατιστική και κοινωνιολογική ανάλυση. Η επιτυχής κατανόηση και αξιοποίηση αυτών των δεδομένων είναι δυνατή μόνο μέσα από ένα ολιστικό και διεπιστημονικό πλαίσιο προσέγγισης.

## 5.2 Πηγές και Τύποι Δεδομένων (κείμενο, εικόνα, βίντεο, hashtags, χρονικά δεδομένα)

Η ανάπτυξη των κοινωνικών μέσων δικτύωσης έχει συνοδευτεί από μία ριζική μεταβολή στη φύση, το περιεχόμενο και τη μορφή των δεδομένων που παράγονται και διακινούνται σε καθημερινή βάση. Οι πλατφόρμες όπως το Twitter, το Facebook, το Instagram, το TikTok και το YouTube αποτελούν ταυτόχρονα πηγές κειμένου, εικόνας, ήχου, βίντεο και διαφόρων μεταδεδομένων, δημιουργώντας έτσι μια εξαιρετικά πολύμορφη και δυναμική πληροφοριακή πραγματικότητα. Η ετερογένεια αυτή των δεδομένων απαιτεί προηγμένες τεχνικές ανάλυσης και προσαρμοσμένες μεθοδολογίες ανάλογα με το είδος του περιεχομένου (Zhao et al., 2011).

Το κείμενο αποτελεί τον πιο συχνό τύπο δεδομένων στα μέσα κοινωνικής δικτύωσης, περιλαμβάνοντας αναρτήσεις, σχόλια, άμεσες συνομιλίες, τίτλους και περιγραφές. Κείμενα όπως tweets, posts και status updates είναι συνήθως σύντομα, ανεπίσημα και γεμάτα με γλωσσικές παραλλαγές, emoticons, hashtags και συντομογραφίες. Αυτά τα στοιχεία δημιουργούν προκλήσεις για τη φυσική γλωσσική επεξεργασία (NLP), όπως η αναγνώριση οντοτήτων, η ανάλυση συναισθήματος και η κατανόηση του συμφραζομένου (Liu, 2012).

Η συλλογή τέτοιων δεδομένων πραγματοποιείται μέσω των APIs των πλατφορμών, ενώ η ανάλυσή τους συχνά προϋποθέτει προκαταρκτικά στάδια όπως αφαίρεση θορύβου, κανονικοποίηση λέξεων, αφαίρεση stop-words και lemmatization. Επιπλέον, στις αναλύσεις συχνά χρησιμοποιούνται τεχνικές εξαγωγής χαρακτηριστικών, όπως TF-IDF, word embeddings (Word2Vec, GloVe), ή πιο πρόσφατα, γλωσσικά μοντέλα όπως το BERT (Devlin et al., 2019).

Οι εικόνες καταλαμβάνουν κεντρική θέση σε πλατφόρμες όπως το Instagram, το Facebook και το Snapchat. Παράγονται κυρίως από χρήστες και χρησιμοποιούνται για προσωπική έκφραση, επαγγελματική προβολή, ή σχολιασμό γεγονότων. Σε πολλές περιπτώσεις, συνδυάζονται με λεζάντες, hashtags και σχόλια, επιτρέποντας την πολυτροπική ανάλυση του περιεχομένου (Βασιλειάδης, 2025).

Η ανάλυση εικόνας στα social media περιλαμβάνει τεχνικές υπολογιστικής όρασης, όπως αναγνώριση αντικειμένων, κατηγοριοποίηση σκηνών, ανίχνευση προσώπων και ανάλυση συναισθήματος. Τα πιο διαδεδομένα εργαλεία βασίζονται σε συνελικτικά νευρωνικά δίκτυα (CNNs), τα οποία μαθαίνουν να εξάγουν μορφολογικά και εννοιολογικά χαρακτηριστικά από τις εικόνες. Η πληροφορία από την εικόνα μπορεί να συνδυαστεί με κείμενο ή μεταδεδομένα για ενισχυμένη ερμηνεία (Vinyals et al., 2015).

Το βίντεο είναι κυρίαρχο μέσο έκφρασης σε πλατφόρμες όπως το TikTok, το YouTube και το Instagram Reels. Η ανάλυση του βίντεο περιλαμβάνει την επεξεργασία χρονικών ακολουθιών εικόνων (frames) καθώς και ηχητικής πληροφορίας, γεγονός που καθιστά την ανάλυση του πολύ πιο απαιτητική από αυτή της εικόνας ή του κειμένου.

Η οπτικοακουστική επεξεργασία περιλαμβάνει την απομόνωση σημαντικών στιγμών (keyframe selection), την αναγνώριση προσώπων, χειρονομιών, αντικειμένων και την ανάλυση συναισθήματος βάσει φωνής. Πλαίσια όπως τα 3D CNNs ή τα Recurrent Neural Networks (RNNs) έχουν προταθεί για την ανάλυση διαχρονικών δεδομένων, συνδυάζοντας πληροφορίες από πολλαπλά επίπεδα (Βασιλειάδης, 2025b).

Τα hashtags αποτελούν μορφή ετικέτας (metadata) που χρησιμοποιούνται για την κατηγοριοποίηση του περιεχομένου και τη σύνδεση χρηστών γύρω από κοινά ενδιαφέροντα ή θέματα. Η γλωσσική και συμβολική τους δομή επιτρέπει την εξαγωγή θεμάτων, τάσεων και την ανάλυση συναισθηματικών ή ιδεολογικών τοποθετήσεων (Tsur & Rappoport, 2012).

Οι αναφορές σε άλλους χρήστες (mentions) επιτρέπουν την ανασύσταση της κοινωνικής δομής, τον εντοπισμό επιρροών (influencers) και την ανάλυση δικτύου. Με τη βοήθεια αυτών των στοιχείων κατασκευάζονται γράφοι που περιγράφουν την πληροφοριακή ροή και τις αλληλεπιδράσεις.

Το χρονικό στίγμα (timestamp) κάθε ανάρτησης ή αλληλεπίδρασης αποτελεί κρίσιμο δεδομένο για την κατανόηση της εξέλιξης της πληροφορίας στο χρόνο. Τα δεδομένα των social media είναι χρονικά σειριοποιημένα, επιτρέποντας τη μελέτη της δυναμικής διάδοσης, την πρόβλεψη γεγονότων και την εντοπισμό περιόδων αυξημένης δραστηριότητας (event bursts).

Η ανάλυση χρονικών δεδομένων περιλαμβάνει τεχνικές όπως time series modeling, moving averages, change point detection και autoregressive models (ARIMA, LSTM). Χρησιμοποιούνται κυρίως για την κατανόηση μοτίβων συμπεριφοράς, την ανίχνευση κρίσεων ή την πρόβλεψη τάσεων στο κοινωνικό σύνολο (Syed et al., 2021).

Η ενσωμάτωση όλων των παραπάνω τύπων δεδομένων σε ενιαίο αναλυτικό πλαίσιο αποτελεί ένα από τα σημαντικότερα ερευνητικά ζητήματα στη σύγχρονη επιστήμη των δεδομένων. Η πολυτροπική ανάλυση (multimodal analytics) επιδιώκει να συνδυάσει κειμενική, οπτική, ηχητική και χρονική πληροφορία ώστε να ενισχύσει την ακρίβεια, το βάθος και τη σημασιολογική πληρότητα των προβλέψεων ή ταξινομήσεων.

Η σύνθεση αυτή επιτυγχάνεται με τεχνικές όπως το early fusion (συνδυασμός δεδομένων πριν την ανάλυση) και το late fusion (συνδυασμός αποτελεσμάτων ξεχωριστών μοντέλων), ενώ πιο σύγχρονες προσεγγίσεις χρησιμοποιούν attention-based networks και transformer models για τον ταυτόχρονο χειρισμό πολλαπλών τύπων πληροφορίας (Tsai et al., 2019).

### ***5.3 Προ-επεξεργασία Δεδομένων από Κοινωνικά Δίκτυα (Data Cleaning & Feature Engineering)***

Η προεπεξεργασία δεδομένων αποτελεί κρίσιμη προϋπόθεση για την επιτυχημένη εφαρμογή μεθόδων μηχανικής μάθησης και εξαγωγής γνώσης από τα μέσα κοινωνικής δικτύωσης. Τα δεδομένα που προέρχονται από πλατφόρμες όπως το Twitter, το Facebook, το Instagram και το TikTok εμφανίζουν ιδιαίτερη ετερογένεια, ακαταστασία και ασάφεια, καθιστώντας τη διαδικασία του καθαρισμού και της μετατροπής τους σε χρήσιμα χαρακτηριστικά απαραίτητη για την επίτευξη ακρίβειας και αξιοπιστίας στις αναλύσεις (Gandomi & Haider, 2015).

Η φύση των κοινωνικών δεδομένων τα καθιστά ακατέργαστα: οι αναρτήσεις συχνά περιλαμβάνουν συντομογραφίες, λάθη πληκτρολόγησης, emojis, hashtags, συνδέσμους, ακόμη και σύμβολα ή γλώσσες που δεν είναι απαραίτητα συμβατές με τυπικές προσεγγίσεις στατιστικής επεξεργασίας. Η προεπεξεργασία περιλαμβάνει βήματα όπως η κανονικοποίηση της γλώσσας (π.χ. μετατροπή όλων των χαρακτήρων σε πεζά), η αφαίρεση των σημείων στίξης, η εξάλειψη των κοινών λέξεων που δεν φέρουν σημαντική πληροφορία (stopwords), καθώς και η αντιμετώπιση των emojis και των hashtags. Στο πλαίσιο αυτό, τεχνικές λεματοποίησης (lemmatization) ή αποκοπής (stemming) εφαρμόζονται για να επαναφέρουν τις λέξεις στη βασική τους μορφή, μειώνοντας έτσι τη διασπορά των δεδομένων και βελτιώνοντας την κανονικότητα της πληροφορίας (Manning et. al., 2008).

Ένα από τα σημαντικότερα στάδια της προεπεξεργασίας είναι η αφαίρεση του θορύβου. Οι χρήστες των κοινωνικών δικτύων εκφράζονται συχνά με ελεύθερο, απρόβλεπτο και ασυνεπή τρόπο, κάνοντας χρήση προσωπικών συντομογραφιών (όπως “gt8” για “great”) ή παραβλέποντας βασικούς κανόνες σύνταξης. Αυτά τα φαινόμενα καθιστούν απαραίτητη την κανονικοποίηση των όρων, η οποία επιτυγχάνεται με εργαλεία αναγνώρισης και μετασχηματισμού slang (Han et. al., 2013). Επιπλέον, η παρουσία διπλών αναρτήσεων ή spam περιεχομένου οδηγεί στη διαστρέβλωση των στατιστικών κατανομών. Η απομάκρυνση τέτοιων στοιχείων είναι κρίσιμη, ιδίως όταν αναλύονται τάσεις ή πρότυπα σε πραγματικό χρόνο.

Η μηχανική χαρακτηριστικών (feature engineering) έρχεται να καλύψει το επόμενο στάδιο της διαδικασίας, μετατρέποντας τα προεπεξεργασμένα δεδομένα σε αναπαραστάσεις κατάλληλες για αλγοριθμική ανάλυση. Στην περίπτωση των κοινωνικών δικτύων, όπου το κείμενο αποτελεί τον πιο κοινό τύπο πληροφορίας, τεχνικές όπως το Bag of Words (BoW) και το TF-IDF χρησιμοποιούνται εκτενώς για την αναπαράσταση του λεξιλογίου και της πληροφορίας σε διανυσματικό χώρο. Το BoW αποτελεί μία από τις παλαιότερες αλλά και πιο αποτελεσματικές προσεγγίσεις, όπου κάθε έγγραφο μετατρέπεται σε ένα διάνυσμα που περιέχει τη συχνότητα εμφάνισης κάθε όρου. Η προσέγγιση TF-IDF, από την άλλη, σταθμίζει την παρουσία των όρων ανάλογα με τη σπανιότητά τους στο σύνολο του σώματος κειμένων, δίνοντας έμφαση σε όρους που είναι διακριτικοί και όχι υπερβολικά κοινοί (Rajaraman & Ullman, 2011).

Πέραν των βασικών αναπαραστάσεων, τα κοινωνικά δεδομένα συχνά απαιτούν πιο εξελιγμένες τεχνικές, όπως η ανάλυση n-grams, η οποία επιτρέπει τη σύλληψη συμφραζομένων και την αναγνώριση κοινών εκφραστικών σχημάτων, όπως bigrams ή trigrams. Αυτές οι τεχνικές συνδυάζονται με αναλύσεις συναισθήματος (sentiment analysis), όπου κάθε ανάρτηση αποδίδεται σε μία ποσοτική κλίμακα συναισθηματικής έκφρασης (αρνητική, θετική, ουδέτερη), στηριζόμενη σε λεξικά συναισθήματος ή προεκπαιδευμένα μοντέλα.

Στην περίπτωση των πολυτροπικών δεδομένων (multimodal data), όπως οι εικόνες και τα βίντεο, η εξαγωγή χαρακτηριστικών γίνεται με τη βοήθεια υπολογιστικής όρασης. Εδώ, οι τεχνικές όπως τα Histogram of Oriented Gradients (HOG) ή οι περιγραφείς SURF και SIFT επιτρέπουν τη συστηματική αναγνώριση μορφολογικών δομών. Τα τελευταία χρόνια, συνελκτικά νευρωνικά δίκτυα (CNNs) έχουν κυριαρχήσει στην εξαγωγή υψηλού επιπέδου χαρακτηριστικών από πολυδιάστατα δεδομένα, όπως στην περίπτωση του ResNet ή του VGG, προσφέροντας ενσωματωμένες αναπαραστάσεις (embeddings) οι οποίες ενσωματώνουν την οπτική πολυπλοκότητα των κοινωνικών εικόνων (Du et. al., 2025).

Παράλληλα με τα δομικά και περιεχομενικά δεδομένα, σημαντικό ρόλο παίζουν και τα μεταδεδομένα (metadata), όπως η ώρα και ημέρα ανάρτησης, ο αριθμός των αλληλεπιδράσεων, τα hashtags, οι συνδέσεις του χρήστη και ο αριθμός ακολούθων. Αυτά τα χαρακτηριστικά προσφέρουν πολύτιμες πληροφορίες σχετικά με τη δυναμική και τη χρονική εξάπλωση του περιεχομένου.

Για τη μείωση της διάστασης του συνόλου χαρακτηριστικών και την εστίαση στα πιο πληροφοριακά χαρακτηριστικά, τεχνικές όπως η Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis - PCA) και η t-SNE εφαρμόζονται για την συμπύκνωση του χώρου και την οπτικοποίηση των clusters. Ιδιαίτερα στις αναλύσεις μεγάλων δεδομένων, η χρήση embeddings όπως τα Word2Vec, GloVe και BERT επιτρέπει την απόδοση σημασιολογικά πλούσιων αναπαραστάσεων, βασισμένων στην κατανομή και στα συμφραζόμενα των λέξεων σε τεράστια κειμενικά σύνολα.

Αναμφίβολα, η διαδικασία της προεπεξεργασίας δεν στερείται προκλήσεων. Η γλωσσική ποικιλομορφία, η χρήση διαλέκτων και διαφορετικών αλφαβήτων, η ειρωνεία και η αμφισημία της γλώσσας αποτελούν εμπόδια που πρέπει να αντιμετωπιστούν με εξελιγμένες τεχνικές και εξειδικευμένα μοντέλα φυσικής γλώσσας. Επιπλέον, οι δυναμικές φύσεις των χρηστών και η χρονική ρευστότητα των κοινωνικών φαινομένων απαιτούν συστήματα ικανά να προσαρμόζονται σε real-time ροές δεδομένων και να εντοπίζουν συσχετισμούς και μοτίβα σε εξελισσόμενο περιεχόμενο.

Συμπερασματικά, η επιτυχής εφαρμογή της μηχανικής μάθησης στα δεδομένα των social media εξαρτάται ουσιαδώς από την ποιότητα της προεπεξεργασίας και την επιλογή των κατάλληλων χαρακτηριστικών. Η συστηματική και στρατηγική προσέγγιση αυτών των

σταδίων παρέχει τη βάση για αξιόπιστη κατηγοριοποίηση, ανάλυση συναισθήματος, εντοπισμό θεμάτων και άλλα σημαντικά καθήκοντα στον χώρο της κοινωνικής πληροφορικής.

## 5.4 Ανάλυση Συναισθήματος (Sentiment Analysis)

Η ανάλυση συναισθήματος, γνωστή και ως εξαγωγή συναισθηματικού περιεχομένου, αποτελεί έναν από τους πιο καθοριστικούς και ενεργούς κλάδους της επεξεργασίας φυσικής γλώσσας (NLP), ιδίως μετά την ραγδαία άνοδο των κοινωνικών μέσων δικτύωσης ως πεδίο έρευνας και επιχειρηματικής αξιοποίησης. Πρόκειται για τη διαδικασία αυτόματης ανίχνευσης και κατηγοριοποίησης της συναισθηματικής φόρτισης που φέρει ένα κείμενο, κατά κανόνα ταξινομώντας το σε θετικό, αρνητικό ή ουδέτερο, ενώ πλέον έχουν αρχίσει να εξετάζονται και πιο εκλεπτυσμένες συναισθηματικές αποχρώσεις όπως ειρωνεία, θυμός, φόβος ή χαρά (Liu, 2012). Το ενδιαφέρον για την ανάλυση συναισθήματος εντείνεται λόγω της αυξανόμενης επιθυμίας επιχειρήσεων, οργανισμών και κυβερνήσεων να παρακολουθούν σε πραγματικό χρόνο τη δημόσια διάθεση, τις αντιδράσεις του κοινού και τα κοινωνικά αισθήματα.

Από θεωρητικής πλευράς, η ανάλυση βασίζεται στην αρχή ότι οι λέξεις και οι φράσεις φέρουν πολικότητα, δηλαδή είναι ικανές να εκφράσουν συναισθηματική διάθεση. Το κλασικό πολικό μοντέλο ταξινομεί την πληροφορία σε θετική, αρνητική και ουδέτερη, αλλά τα σύγχρονα μοντέλα επεκτείνονται και σε περισσότερες διαστάσεις συναισθήματος, είτε μέσω θεωριών όπως το Circumplex Model του Russell, είτε μέσω παρατηρησιακών δεδομένων. Η πληροφορία αυτή μπορεί να ανιχνευθεί σε λέξεις, φράσεις ή σε συνδυασμούς προτάσεων, ενώ καθοριστικό ρόλο παίζει το συμφραζόμενο, το οποίο συχνά αλλοιώνει την ερμηνεία μιας μεμονωμένης λέξης. Για παράδειγμα, η λέξη «υπέροχα» σε κυριολεκτικό συμφραζόμενο είναι θετική, ενώ σε ειρωνικό περιεχόμενο μπορεί να φέρει αρνητική σημασία (Pang & Lee, 2008).

Η προσέγγιση της ανάλυσης συναισθήματος μπορεί να είναι λεξικοκεντρική, στατιστική ή να βασίζεται σε μηχανική μάθηση. Οι λεξικοκεντρικές μέθοδοι χρησιμοποιούν έτοιμες βάσεις λέξεων και συναισθηματικής πολικότητας όπως τα SentiWordNet, LIWC και NRC Emotion Lexicon. Σε αυτή την περίπτωση, ο αλγόριθμος σαρώνει το κείμενο, εντοπίζει τις λέξεις και τις βαθμολογεί σύμφωνα με τις τιμές που φέρουν στα λεξικά. Το τελικό συναίσθημα προκύπτει από το άθροισμα των πολικότητων. Η ακρίβεια αυτής της μεθόδου εξαρτάται από την ποιότητα και την πληρότητα του λεξικού, όμως συχνά αποτυγχάνει όταν οι λέξεις χρησιμοποιούνται σε διαφορετικό συμφραζόμενο, ή όταν υπάρχουν στοιχεία μεταφορικής γλώσσας ή ειρωνείας (Taboada et al., 2011).

Οι στατιστικές τεχνικές και οι μέθοδοι μηχανικής μάθησης βασίζονται σε επισημασμένα σύνολα δεδομένων (corpora) που περιλαμβάνουν αναρτήσεις, σχόλια ή κριτικές οι οποίες έχουν ήδη κριθεί από ανθρώπους ως θετικές, αρνητικές ή ουδέτερες. Από αυτά τα δεδομένα, ο αλγόριθμος εκπαιδεύεται ώστε να αναγνωρίζει μοτίβα και να ταξινομεί νέα, αταξινόμητα κείμενα. Τυπικές μέθοδοι που χρησιμοποιούνται είναι οι Naïve Bayes, Logistic Regression και SVM. Τα δεδομένα μετατρέπονται σε διανύσματα μέσω τεχνικών όπως TF-IDF ή Word Embeddings (π.χ. Word2Vec, GloVe), που επιτρέπουν την αριθμητική αναπαράσταση

λέξεων και φράσεων (Medhat et. al., 2014). Η εξέλιξη των αλγορίθμων βαθιάς μάθησης και η ανάπτυξη προεκπαιδευμένων μοντέλων όπως τα LSTM, BERT και GPT έφεραν ριζικές αλλαγές στην ανάλυση συναισθήματος. Τα μοντέλα αυτά βασίζονται σε αρχιτεκτονικές νευρωνικών δικτύων που μπορούν να μάθουν από τα συμφραζόμενα, να εντοπίζουν την πολυσημία, να διαχειρίζονται μακροχρόνιες εξαρτήσεις και να ερμηνεύουν ακόμη και μεταφορικές εκφράσεις. Για παράδειγμα, το μοντέλο BERT επιτρέπει τη διδιάστατη κατανόηση του κειμένου – εξετάζει δηλαδή τη λέξη τόσο με βάση το προηγούμενο όσο και το επόμενο συμφραζόμενο – γεγονός που ενισχύει τη ρεαλιστικότητα και την ευαισθησία στην ερμηνεία (Devlin et al., 2019).

Ωστόσο, η εφαρμογή της ανάλυσης συναισθήματος στα κοινωνικά δίκτυα δεν είναι απλή υπόθεση. Τα δεδομένα είναι συχνά θορυβώδη, ανορθόγραφα, γεμάτα emoticons, hashtags και συντομογραφίες. Επιπλέον, οι χρήστες συχνά χρησιμοποιούν ειρωνεία, σαρκασμό ή σύγχρονη γλώσσα του διαδικτύου που αποκλίνει από τα κλασικά λεξικά ή τα γλωσσικά μοντέλα που έχουν εκπαιδευτεί σε πιο «καθαρό» λόγο (Riloff et al., 2013). Επίσης, τα δεδομένα μπορεί να είναι ανισόρροπα, δηλαδή να παρουσιάζουν υπερεκπροσώπηση μιας κλάσης – για παράδειγμα, οι χρήστες να εκφράζουν συχνότερα θετικά σχόλια σε σχέση με αρνητικά. Αυτό δημιουργεί στρεβλώσεις στην εκπαίδευση των ταξινομητών και απαιτεί τεχνικές εξισορρόπησης όπως oversampling, undersampling, χρήση σταθμισμένων απωλειών και προσεκτική επιλογή μετρικών αξιολόγησης (π.χ. F1-score αντί ακρίβειας).

Οι εφαρμογές της ανάλυσης συναισθήματος είναι πολυάριθμες. Εταιρείες χρησιμοποιούν τεχνικές παρακολούθησης φήμης (reputation monitoring) για να αντιληφθούν έγκαιρα τις αλλαγές στο κοινό αίσθημα απέναντι στα προϊόντα ή τις υπηρεσίες τους. Πολιτικά κόμματα αναλύουν την κοινή γνώμη για να ρυθμίσουν τον λόγο τους ή να εντοπίσουν θέματα αιχμής. Δημόσιοι φορείς και οργανισμοί υγείας μπορούν να αντλήσουν κρίσιμες πληροφορίες για την αντίδραση του πληθυσμού κατά τη διάρκεια κρίσεων, όπως φυσικές καταστροφές ή πανδημίες (Greaves et al., 2013).

Στο ερευνητικό πεδίο, η τάση οδηγεί σε συνδυασμένες προσεγγίσεις που ενοποιούν δεδομένα από πολλαπλούς τρόπους – κείμενο, ήχο, εικόνα και βίντεο – για μια πολυτροπική ανάλυση συναισθήματος. Αυτή η προσέγγιση καθίσταται απαραίτητη σε περιβάλλοντα όπως το TikTok ή το YouTube, όπου η πληροφορία δεν εντοπίζεται αποκλειστικά στο κείμενο, αλλά κατανέμεται σε εικόνες, εκφράσεις προσώπου, φωνητικό τόνο και μουσική υπόκρουση. Οι σύγχρονες προσεγγίσεις εξετάζουν επίσης την ανάγκη για διαφάνεια (explainability) και ερμηνευσιμότητα, κυρίως λόγω της ευρείας χρήσης των αναλύσεων σε τομείς όπου η ανθρώπινη κρίση επηρεάζεται από αλγοριθμικές αποφάσεις (Dang et. al., 2020).

Η επιτυχής εφαρμογή της ανάλυσης συναισθήματος στα κοινωνικά δίκτυα απαιτεί την ενοποίηση τεχνικών από το NLP, τη μηχανική μάθηση, την ηθική τεχνητής νοημοσύνης και τη θεωρία κοινωνικής συμπεριφοράς. Σε ένα συνεχώς μεταβαλλόμενο και δυναμικό περιβάλλον, η ακρίβεια, η προσαρμοστικότητα και η ευαισθησία των μοντέλων καθορίζουν τη χρηστικότητα και την αποτελεσματικότητα της εξαγόμενης πληροφορίας.

## 5.5 Ανίχνευση Θεμάτων και Τάσεων (Topic Modeling & Trend Detection)

Η ανάλυση μεγάλου όγκου κοινωνικών δεδομένων προϋποθέτει μεθοδολογίες που να αποδίδουν όχι μόνο τη θεματική πολυπλοκότητα του περιεχομένου, αλλά και τη χρονική δυναμική της πληροφορίας. Δύο τέτοιες μεθοδολογίες είναι η ανίχνευση θεμάτων (topic modeling) και η ανίχνευση τάσεων (trend detection), οι οποίες εφαρμόζονται εκτενώς στα μέσα κοινωνικής δικτύωσης για την κατανόηση των λαθανουσών αφηγήσεων και την έγκαιρη καταγραφή των μεταβολών στην κοινή γνώμη ή στη δημόσια διάθεση (Liu., 2009).

Η ανίχνευση θεμάτων αποτελεί διαδικασία ανάδειξης κρυφών δομών μέσα σε σώματα κειμένων, με σκοπό την ομαδοποίηση λέξεων σε σημασιολογικά συνεκτικές ενότητες. Τα θέματα αυτά, αν και δεν επισημαίνονται ρητά, προκύπτουν μέσω συστηματικής ανάλυσης της συν-εμφάνισης λέξεων. Η σημαντικότερη τεχνική σε αυτό το πλαίσιο είναι το Latent Dirichlet Allocation (LDA), μια στατιστική μέθοδος που μοντελοποιεί κάθε έγγραφο ως ένα μείγμα θεμάτων και κάθε θέμα ως μια κατανομή λέξεων, αξιοποιώντας την πιθανοθεωρητική προσέγγιση για να αναπαραστήσει τη λαθάνουσα θεματολογία ενός συνόλου δεδομένων (Blei et. al., 2003). Ο αλγόριθμος LDA λειτουργεί υπό το παραδοσιακό Bayesian paradigm, εκτιμώντας τις πιθανότητες συνύπαρξης λέξεων και αποδίδοντάς τους σε «topics» που αναδύονται μέσα από τη συλλογική σημασιολογική δομή των κειμένων.

Η εφαρμογή του LDA στα μέσα κοινωνικής δικτύωσης παρουσιάζει αρκετές προκλήσεις. Η συντομία των κειμένων, όπως τα tweets, περιορίζει την πληροφοριακή πυκνότητα κάθε παρατήρησης, ενώ η πολυμορφία της γλώσσας – με παρουσία emoticons, hashtags, abbreviations και memes – προσθέτει σημαντικό θόρυβο στην ανάλυση. Για την αντιμετώπιση αυτών των περιορισμών έχουν αναπτυχθεί παραλλαγές του βασικού αλγορίθμου, όπως το Twitter-LDA, που λαμβάνει υπόψη τη φύση του συγκεκριμένου μέσου και εισάγει συμπληρωματικές μεταβλητές, όπως το user context ή τις θεματικές των hashtags, για να βελτιώσει την απόδοση της θεματικής ομαδοποίησης (Zhao et al., 2011).

Πέραν του LDA και άλλες τεχνικές όπως η Non-negative Matrix Factorization (NMF) έχουν χρησιμοποιηθεί επιτυχώς για την ανίχνευση θεμάτων. Η NMF επιτρέπει τη διάσπαση του πίνακα όρων-εγγράφων σε δύο μη αρνητικούς πίνακες, ενισχύοντας την ερμηνευσιμότητα των παραγόμενων θεμάτων, ενώ ταυτόχρονα προσφέρει μεγαλύτερη υπολογιστική αποδοτικότητα, ιδίως σε περιβάλλοντα με πολύ μεγάλες ποσότητες δεδομένων (Lee & Seung, 2000).

Από την άλλη πλευρά, η ανίχνευση τάσεων συνιστά διαδικασία χρονικά προσδιορισμένης ανάλυσης, με έμφαση στην καταγραφή της μεταβολής της συχνότητας εμφάνισης όρων, θεμάτων ή γεγονότων μέσα στον χρόνο. Βασίζεται στη συλλογή και επεξεργασία χρονοσειρών και εφαρμόζεται για την αναγνώριση «εκρήξεων» πληροφορίας (bursts),

εντοπίζοντας ξαφνικές αλλαγές στο δημόσιο ενδιαφέρον. Ένα από τα πιο κλασικά εργαλεία σε αυτή την κατηγορία είναι το μοντέλο του Kleinberg (2003), το οποίο αξιοποιεί κατανομές Poisson για την ανίχνευση ασυνήθιστων αυξήσεων στη συχνότητα λέξεων ή όρων μέσα από «χρονικά παράθυρα» μεταβλητού μήκους.

Οι μέθοδοι ανίχνευσης τάσεων εμπλουτίζονται επιπλέον με τεχνικές ανάλυσης χρονοσειρών όπως τα autoregressive models και το Holt-Winters exponential smoothing, τα οποία επιτρέπουν την εξομάλυνση και πρόβλεψη της συμπεριφοράς των τάσεων σε μελλοντικούς χρονικούς ορίζοντες. Οι αλγόριθμοι αυτοί είναι ιδιαίτερα χρήσιμοι σε επιχειρησιακές ή πολιτικές εφαρμογές, όπου η παρακολούθηση της δημόσιας συζήτησης σε (quasi) real-time παρέχει συγκριτικό πλεονέκτημα στην ανταπόκριση και τον στρατηγικό σχεδιασμό (Imran et al., 2015). Η δυναμική των hashtags είναι κρίσιμη τόσο για την ανίχνευση θεμάτων όσο και για την παρακολούθηση τάσεων. Λειτουργούν ως σημασιολογικά σημεία αναφοράς, προσφέροντας μεταδεδομένα που ενισχύουν τη σημασιολογική οργάνωση του περιεχομένου. Ειδικά όταν χρησιμοποιούνται σε συνδυασμό με γεωγραφικά ή χρονικά δεδομένα (π.χ. geo-tagged tweets), η ανάλυση γίνεται πολυεπίπεδη, επιτρέποντας την ταυτόχρονη διερεύνηση περιεχομένου, χωρικών συσχετίσεων και χρονικών μεταβολών.

Η οπτικοποίηση των θεμάτων και των τάσεων αποτελεί αναπόσπαστο κομμάτι της ερμηνείας των αποτελεσμάτων. Τεχνικές όπως τα word clouds προσφέρουν απλουστευμένες, αλλά ιδιαίτερα χρήσιμες απεικονίσεις των κυρίαρχων λέξεων ενός θέματος, ενώ τα heatmaps και τα timeline graphs επιτρέπουν την αναγνώριση αλλαγών σε θεματολογική ένταση, καθώς και τη χαρτογράφηση της κοινωνικής αντήχησης (resonance) ενός γεγονότος στο πέρασμα του χρόνου. Στην αξιολόγηση της θεματικής συνοχής χρησιμοποιούνται μετρικές όπως το topic coherence score (Newman et al., 2010), ενώ για τη σταθερότητα μιας τάσης εφαρμόζονται δείκτες διακρίμανσης και προβλεψιμότητας.

Η συνδυαστική εφαρμογή τεχνικών topic modeling και trend detection έχει βρει εφαρμογή σε ποικίλα πεδία. Στην πολιτική επικοινωνία, οι τεχνικές αυτές χρησιμοποιούνται για τον εντοπισμό κεντρικών θεμάτων προεκλογικών καμπανιών ή δημόσιων αντιπαράθεσεων, επιτρέποντας στα κόμματα να επαναπροσδιορίσουν τη στρατηγική τους βάσει των κυρίαρχων ανησυχιών των πολιτών (Wang et al., 2012). Στο πεδίο της διαχείρισης κρίσεων, προσφέρουν εργαλεία για την έγκαιρη αναγνώριση αναγκών και συναισθημάτων κατά τη διάρκεια φυσικών καταστροφών, συμβάλλοντας στην προσαρμογή του δημόσιου λόγου και της κρατικής παρέμβασης. Τέλος, στον επιχειρηματικό τομέα, χρησιμοποιούνται για την παρακολούθηση θεμάτων και τάσεων γύρω από προϊόντα ή υπηρεσίες, αναδεικνύοντας με ακρίβεια ευκαιρίες και αδυναμίες στις αγορές στόχου (Tuatob & Tucker, 2015).

Εν κατακλείδι, η ενσωμάτωση της ανίχνευσης θεμάτων και τάσεων στο πλαίσιο της ανάλυσης κοινωνικών δεδομένων παρέχει ένα σύνθετο αλλά ισχυρό εργαλείο ερμηνείας της ψηφιακής κοινωνικής ζωής. Με την αξιοποίηση μεθόδων της στατιστικής, της μηχανικής μάθησης και της υπολογιστικής γλωσσολογίας, επιτρέπεται όχι μόνο η κατανόηση του παρόντος αλλά και η πρόβλεψη του μέλλοντος, ενισχύοντας τη δυνατότητα παρέμβασης σε ποικίλους τομείς της ανθρώπινης δραστηριότητας.

## 5.6 Ανάλυση Δικτύων και Κοινοτήτων (*Network Analysis & Community Detection*)

Η ανάλυση δικτύων (network analysis) και η ανίχνευση κοινοτήτων (community detection) αποτελούν θεμελιώδη πεδία της υπολογιστικής κοινωνικής επιστήμης και της ανάλυσης δεδομένων από μέσα κοινωνικής δικτύωσης. Οι κοινωνικές πλατφόρμες, όπως το Facebook, το Twitter και το Instagram, λειτουργούν εγγενώς ως δίκτυα, στα οποία οι χρήστες (κόμβοι) συνδέονται μεταξύ τους μέσω διαφόρων τύπων σχέσεων (ακμές), όπως φιλία, ακολούθηση, κοινοποιήσεις, απαντήσεις και αναφορές (Wasserman & Faust, 1994). Η μελέτη αυτών των δομών παρέχει πολύτιμες πληροφορίες για τη δυναμική της διάδοσης της πληροφορίας, την επιρροή, την πολυπλοκότητα της επικοινωνίας και τη διαμόρφωση κοινωνικών ομάδων.

Η θεωρητική βάση της ανάλυσης δικτύων στηρίζεται στη θεωρία γράφων, όπου ένα κοινωνικό δίκτυο αναπαρίσταται ως μη κατευθυνόμενος ή κατευθυνόμενος γράφος  $G = (V, E)$ , με  $V$  το σύνολο των κόμβων (π.χ. χρηστών) και  $E$  το σύνολο των ακμών (π.χ. σχέσεων). Ανάλογα με τον τύπο των σχέσεων, ο γράφος μπορεί να είναι σταθμισμένος (weighted) ή μη σταθμισμένος (unweighted), κατευθυνόμενος (directed) ή μη (undirected), ενώ η πολυπλοκότητά του αυξάνεται με την προσθήκη χρονικών (temporal) ή πολυτροπικών (multimodal) διαστάσεων (Newman, 2010).

Μία από τις βασικότερες εφαρμογές της ανάλυσης δικτύων είναι η εντόπιση σημαντικών κόμβων, μέσω μετρικών όπως η βαθμοκεντρικότητα (degree centrality), η μεσοκεντρικότητα (betweenness centrality) και η ιδιοτιμοκεντρικότητα (eigenvector centrality). Η βαθμοκεντρικότητα αποτυπώνει τον αριθμό των συνδέσεων που έχει ένας κόμβος και αντανακλά την άμεση κοινωνική του δραστηριότητα. Η μεσοκεντρικότητα δείχνει τον αριθμό των μικρότερων διαδρομών που περνούν από έναν κόμβο, καθιστώντας τον στρατηγικό σημείο ελέγχου της πληροφορίας. Αντίστοιχα, η ιδιοτιμοκεντρικότητα μετρά τη σημασία ενός κόμβου βάσει της επιρροής των κόμβων με τους οποίους συνδέεται – μια έννοια που βρίσκεται στον πυρήνα του αλγορίθμου PageRank της Google (Bonacich, 1987; Page et al., 1998).

Η ανίχνευση κοινοτήτων, ή αλλιώς η αναγνώριση ομάδων κόμβων με πυκνή μεταξύ τους διασύνδεση και αραιές συνδέσεις με τον υπόλοιπο γράφο, είναι εξαιρετικά σημαντική για την κατανόηση της κοινωνικής συνοχής και της διάχυσης πληροφορίας. Οι κοινότητες σε κοινωνικά δίκτυα μπορεί να αντιστοιχούν σε ομάδες χρηστών με κοινά ενδιαφέροντα, γεωγραφική εγγύτητα, πολιτική προτίμηση ή κοινούς στόχους (Fortunato & Hric, 2016).

Υπάρχουν πολλές μέθοδοι ανίχνευσης κοινοτήτων. Η πιο διαδεδομένη είναι η μεγιστοποίηση της διαμόρφωσης (modularity maximization), όπου ο στόχος είναι η εύρεση τέτοιας διαίρεσης των κόμβων ώστε η ενδο-ομαδική συνδεσιμότητα να είναι σημαντικά υψηλότερη από την αναμενόμενη σε τυχαίο γράφο. Ο αλγόριθμος Louvain (Blondel et al., 2008) αποτελεί μία από τις πιο αποδοτικές προσεγγίσεις για μεγάλης κλίμακας δίκτυα.

Εναλλακτικά, ιεραρχικές τεχνικές (π.χ. dendrograms), φασματικές μέθοδοι που χρησιμοποιούν τις ιδιοτιμές του λαπλασιανού πίνακα του γράφου και μέθοδοι βασισμένες σε πυκνότητα όπως ο αλγόριθμος Label Propagation, προσφέρουν διαφορετικά εργαλεία αναπαράστασης και ερμηνείας των κοινοτήτων (Newman, 2006).

Η εφαρμογή της ανάλυσης δικτύων και κοινοτήτων στα δεδομένα κοινωνικών μέσων είναι πολλαπλή. Στο Twitter, για παράδειγμα, οι κόμβοι μπορεί να είναι χρήστες και οι ακμές να εκπροσωπούν retweets ή αναφορές. Μέσω αυτής της αναπαράστασης, μπορεί να αναγνωριστεί πώς μια είδηση ή ένα hashtag εξαπλώνεται, ποιοι χρήστες λειτουργούν ως «κόμβοι διάχυσης» (influencers) και ποιες κοινότητες παρουσιάζουν αντίστοιχη θεματολογία ή στάση. Η συνδυασμένη ανάλυση κοινοτήτων και συναισθήματος μπορεί να αποκαλύψει ομάδες με ομοιογενή συναισθηματική απήχηση, όπως θετικά ή αρνητικά συντονισμένες ομάδες γύρω από πολιτικά ζητήματα (Conover et al., 2011).

Επιπλέον, η χρονική ανάλυση των γραφικών δομών – δηλαδή η εξέταση της εξέλιξης του κοινωνικού γράφου μέσα στο χρόνο – επιτρέπει την ανίχνευση αλλαγών στην κοινωνική δυναμική. Για παράδειγμα, η εμφάνιση νέων συνδέσεων σε κρίσιμες χρονικές στιγμές μπορεί να υποδεικνύει συγκέντρωση προσοχής γύρω από ένα γεγονός, όπως ένα κοινωνικό κίνημα ή μια κρίση. Οι δυναμικοί γράφοι (dynamic graphs) και η ανάλυση ροής πληροφορίας (information flow analysis) συνεισφέρουν στη μοντελοποίηση τέτοιων φαινομένων.

Σημαντικό εργαλείο αποτελεί και η οπτικοποίηση των δικτύων, η οποία συμβάλλει στην κατανόηση της δομής και της πυκνότητας των συνδέσεων. Εργαλεία όπως το Gephi, το Cytoscape και το NetworkX επιτρέπουν τη χαρτογράφηση σύνθετων δικτύων, την αποτύπωση κοινοτήτων με διαφορετικά χρώματα και την ανάλυση των μετρικών κόμβων και υποδικτύων. Η οπτικοποίηση είναι καθοριστική σε περιπτώσεις πολιτικής ανάλυσης, ανίχνευσης bot activity και εντοπισμού clusters παραπληροφόρησης.

Προκλήσεις που εντοπίζονται στην ανάλυση δικτύων στα social media περιλαμβάνουν την αντιμετώπιση ψευδών συνδέσεων (π.χ. bots, αυτοματοποιημένα likes), την επεξεργασία τεράστιων δεδομένων σε πραγματικό χρόνο (streaming graphs) και την πολυτροπική φύση των δεδομένων, όπου χρήστες συνδέονται ταυτόχρονα μέσω κειμένου, εικόνας, βίντεο και αλληλεπιδράσεων.

Παρόλα αυτά, η ανάλυση δικτύων και κοινοτήτων στα κοινωνικά μέσα παραμένει μία από τις πλέον αποδοτικές μεθόδους κατανόησης της κοινωνικής δομής, της ροής της πληροφορίας και των μηχανισμών επιρροής, με εφαρμογές από την πολιτική και το marketing έως τη διαχείριση κρίσεων και τη δημόσια ασφάλεια.

## 5.7 Προβλέψεις και Κατηγοριοποιήσεις (Classification & Prediction)

Η κατηγοριοποίηση και η πρόβλεψη αποτελούν δύο από τους πιο κρίσιμους άξονες εφαρμογής της μηχανικής μάθησης στην επεξεργασία δεδομένων από τα μέσα κοινωνικής δικτύωσης. Καθώς οι πλατφόρμες κοινωνικής δικτύωσης παράγουν έναν τεράστιο όγκο πολυτροπικών δεδομένων (κείμενα, εικόνες, βίντεο, μεταδεδομένα), η ανάγκη για αυτόματη οργάνωση, ερμηνεία και πρόβλεψη της συμπεριφοράς των χρηστών καθίσταται ιδιαίτερα επιτακτική. Η ταξινόμηση (classification) αναφέρεται σε εκείνες τις περιπτώσεις όπου ο αλγόριθμος καλείται να αναγνωρίσει σε ποια κατηγορία ανήκει ένα νέο, άγνωστο δείγμα, βάσει ενός συνόλου εκπαίδευσης με επισημασμένα δεδομένα (Tan et al., 2018). Αντίστοιχα, η πρόβλεψη (prediction) στοχεύει στην εκτίμηση μιας συνεχιζόμενης ή μελλοντικής συμπεριφοράς, είτε πρόκειται για ποσοτικές ειβάσεις (π.χ. αριθμός αλληλεπιδράσεων, viral δυναμική), είτε για χρονικά εξαρτώμενα φαινόμενα, όπως η εξάπλωση ενός θέματος στον κοινωνικό γράφο.

Η κατηγοριοποίηση στα κοινωνικά δίκτυα μπορεί να λάβει πολλές μορφές. Κλασικές εφαρμογές περιλαμβάνουν την ανάλυση συναισθήματος, την ανίχνευση ψευδών ειδήσεων, την αναγνώριση spam ή την κατηγοριοποίηση του τύπου περιεχομένου (π.χ., διαφημιστικό, ενημερωτικό, κοινωνικό, ψυχαγωγικό). Οι αλγόριθμοι που χρησιμοποιούνται ποικίλλουν από απλά μοντέλα όπως το Naïve Bayes, μέχρι πιο σύνθετους όπως τα Support Vector Machines (SVM), τα νευρωνικά δίκτυα και τα μοντέλα τυχαίων δασών (Random Forests). Η επιλογή του κατάλληλου αλγορίθμου εξαρτάται όχι μόνο από το μέγεθος και τη φύση των δεδομένων αλλά και από το είδος της ετικέτας στόχου, την πολυπλοκότητα των χαρακτηριστικών και την ανάγκη για ερμηνευσιμότητα (Aggarwal, 2011).

Η αναπαράσταση των δεδομένων είναι ένα κρίσιμο ζήτημα. Τα κοινωνικά δεδομένα, ιδιαίτερα όταν πρόκειται για σύντομα κείμενα (όπως tweets), απαιτούν τεχνικές που αναδεικνύουν τη σημασιολογική πληροφορία και διατηρούν τη συντακτική συνοχή. Κλασικές τεχνικές όπως το Bag-of-Words και το TF-IDF βασίζονται στη συχνότητα των όρων, όμως αποτυγχάνουν να αποδώσουν το πλαίσιο χρήσης και τις συναισθηματικές αποχρώσεις. Για τον λόγο αυτό, προτιμώνται ολοένα και περισσότερο τεχνικές word embeddings, όπως τα Word2Vec, GloVe και πιο πρόσφατα τα BERT και RoBERTa, που επιτρέπουν την εκπαίδευση των μοντέλων με βάση τη σημασιολογική εγγύτητα μεταξύ λέξεων (Devlin et al., 2019).

Από την άλλη πλευρά, η πρόβλεψη αποτελεί πεδίο που συνδέεται άμεσα με την ανάλυση χρονικών σειρών και τη μελέτη της συμπεριφοράς στο χρόνο. Οι προβλέψεις περιλαμβάνουν την εκτίμηση του αριθμού αλληλεπιδράσεων που θα δεχτεί μία ανάρτηση, την πιθανότητα μετάδοσης ενός hashtag, ή ακόμη και τη δυναμική ενός ψηφιακού γεγονότος. Οι μέθοδοι ARIMA και Prophet προσφέρουν στατιστική προσέγγιση στις χρονοσειρές, ενώ τα Recurrent Neural Networks (RNNs) και τα Long Short-Term Memory Networks

(LSTMs) παρέχουν εξελιγμένες δυνατότητες μάθησης χρονικών εξαρτήσεων και αναπαράστασης μακροχρόνιων σχέσεων. Τα μοντέλα αυτά έχουν χρησιμοποιηθεί για την πρόβλεψη της δημοτικότητας περιεχομένου, της εξάπλωσης τάσεων, ακόμα και της εγκληματικής συμπεριφοράς (Tumasjan et al., 2010; Zhao et al., 2015).

Στις σύγχρονες εφαρμογές, παρατηρείται μετατόπιση από μονοτροπικά σε πολυτροπικά μοντέλα, που ενοποιούν δεδομένα από διαφορετικές πηγές – κείμενα, εικόνες, metadata – με στόχο τη βελτιστοποίηση των προβλέψεων. Για παράδειγμα, η δημοτικότητα μιας ανάρτησης στο Instagram μπορεί να επηρεάζεται από το ύψος της εικόνας, την περιγραφή της λεζάντας, την ώρα ανάρτησης και τα χαρακτηριστικά του χρήστη. Τα multimodal deep learning μοντέλα χρησιμοποιούν συνελικτικά νευρωνικά δίκτυα (CNNs) για τις εικόνες, επαναλαμβανόμενα δίκτυα (LSTMs) για τα κείμενα και πλήρως συνδεδεμένα επίπεδα (dense layers) για τα μεταδεδομένα, συνθέτοντας μια ενιαία εκπροσώπηση που τροφοδοτείται σε έναν ταξινομητή ή παλινδρομητή.

Η αποτελεσματικότητα των ταξινομητών εξαρτάται επίσης από την ισορροπία των δεδομένων. Σε πολλές περιπτώσεις, οι κλάσεις είναι άνισες, με υπερεκπροσώπηση θετικών ή ουδέτερων σχολίων, κάτι που οδηγεί σε bias των μοντέλων. Η διαχείριση του προβλήματος αυτού μπορεί να γίνει είτε με τεχνικές εξισορρόπησης όπως το SMOTE ή το random undersampling, είτε με χρήση σταθμισμένων συναρτήσεων κόστους, είτε ακόμη και με ενσωμάτωση της πληροφορίας της ασυμμετρίας κατά την επιλογή των μετρικών αξιολόγησης.

Η αξιολόγηση των αποτελεσμάτων είναι κρίσιμη. Οι πιο συχνά χρησιμοποιούμενες μετρικές για την ταξινόμηση περιλαμβάνουν την ακρίβεια (accuracy), την ακρίβεια πρόβλεψης (precision), την ανάκληση (recall), το F1-score και την AUC-ROC καμπύλη. Στα προβλήματα παλινδρόμησης, συνηθίζονται οι μετρικές MAE και RMSE. Επιπλέον, η εφαρμογή διασταυρούμενης επικύρωσης (k-fold cross-validation) είναι απαραίτητη για την αξιολόγηση της γενικευσιμότητας και της σταθερότητας του μοντέλου, καθώς περιορίζει τον κίνδυνο υπερεκπαίδευσης και βελτιστοποιεί την επιλογή υπερπαραμέτρων.

Ένα ζήτημα που κερδίζει έδαφος είναι η ανάγκη για διαφάνεια και ερμηνευσιμότητα στα αποτελέσματα των προβλέψεων, ιδιαίτερα σε ευαίσθητες εφαρμογές που αφορούν την υγεία, την ασφάλεια ή την πολιτική. Τα black-box μοντέλα βαθιάς μάθησης, αν και αποδίδουν εξαιρετικά στις ταξινομήσεις, δεν προσφέρουν εύκολη ερμηνεία της απόφασής τους. Η ανάπτυξη interpretable models, όπως τα attention-based δίκτυα και τα decision trees, προσπαθεί να εξισορροπήσει την απόδοση με την κατανόηση. Σε ορισμένα πεδία, όπως η ανάλυση δημόσιου λόγου, η διαφάνεια των μοντέλων καθίσταται ακόμη πιο κρίσιμη, προκειμένου να αποφευχθεί η αλλοίωση ή παρερμηνεία των κοινωνικών τάσεων (Shu et. al., 2017).

Συνοψίζοντας, η κατηγοριοποίηση και η πρόβλεψη δεδομένων στα κοινωνικά μέσα αποτελούν ακρογωνιαίο λίθο στην αξιοποίηση της κοινωνικής πληροφορίας. Καθώς οι πλατφόρμες συνεχίζουν να εξελίσσονται, η ανάγκη για μοντέλα που είναι προσαρμόσιμα,

ευφυή, διαφανή και ικανά να χειριστούν τη γλωσσική, χρονική και πολυτροπική πολυπλοκότητα των κοινωνικών δεδομένων αυξάνεται ραγδαία. Οι τεχνολογικές εξελίξεις σε deep learning, NLP και computational social science συνδυάζονται για να προσφέρουν νέους τρόπους ανάλυσης, προβλεπτικότητας και δράσης, καθιστώντας τη συγκεκριμένη ενότητα έναν από τους πιο σύγχρονους και παραγωγικούς τομείς της επιστήμης δεδομένων.

# Κεφάλαιο 6: Εφαρμογές Μηχανικής Μάθησης στα Κοινωνικά Δίκτυα

## 6.1 Ανίχνευση *Fake News* και Παραπληροφόρησης

Η εξάπλωση ψευδών ειδήσεων (fake news) και γενικότερα παραπλανητικού περιεχομένου (misinformation και disinformation) έχει αναδειχθεί ως ένα από τα σημαντικότερα προβλήματα της ψηφιακής εποχής, ιδίως στο πεδίο των κοινωνικών δικτύων. Η εύκολη παραγωγή, αναπαραγωγή και διάχυση περιεχομένου σε πλατφόρμες όπως το Facebook, το Twitter και το TikTok έχει δημιουργήσει ένα περιβάλλον όπου η πληροφορία δεν ελέγχεται απαραίτητα από επίσημους ή έγκυρους φορείς. Αυτό οδηγεί συχνά σε παραμορφωμένες αντιλήψεις της πραγματικότητας, ιδίως σε περιόδους κρίσης όπως εκλογικές διαδικασίες, πανδημίες, ή κοινωνικές αναταραχές (Lazer et al., 2018).

Η ανάγκη για τεχνολογικές λύσεις εντοπισμού της παραπληροφόρησης έχει οδηγήσει στην εφαρμογή τεχνικών μηχανικής μάθησης και επεξεργασίας φυσικής γλώσσας (Natural Language Processing – NLP). Σε θεωρητικό επίπεδο, το πρόβλημα της ανίχνευσης ψευδών ειδήσεων προσεγγίζεται ως πρόβλημα ταξινόμησης: δίνεται ένα κείμενο ή ανάρτηση και ο αλγόριθμος καλείται να την κατατάξει ως αληθή ή ψευδή. Το έργο αυτό είναι εξαιρετικά πολύπλοκο, δεδομένης της συντομίας, της γλωσσικής ποικιλομορφίας και του context-dependence των κοινωνικών μηνυμάτων. Οι τεχνικές που εφαρμόζονται περιλαμβάνουν τόσο επιβλεπόμενες (supervised) όσο και μη επιβλεπόμενες (unsupervised) μεθόδους, αλλά και μοντέλα βαθιάς μάθησης, ιδίως στις πιο πρόσφατες προσεγγίσεις (Shu et al., 2017).

Η ανίχνευση ψευδών ειδήσεων στηρίζεται σε δύο βασικούς άξονες: την ανάλυση του περιεχομένου (content-based detection) και την ανάλυση της διάδοσης στο δίκτυο (propagation-based detection). Στην πρώτη περίπτωση, ο αλγόριθμος επιχειρεί να εντοπίσει γλωσσικά, λεκτικά ή σημασιολογικά μοτίβα που χαρακτηρίζουν τις ψευδείς ειδήσεις. Συχνά, τα fake news εμφανίζουν υπερβολικά συναισθηματική γλώσσα, χρήση επιθέτων, επιφωνημάτων ή κεφαλαίων γραμμάτων, καθώς και ιδιαίτερες λεξιλογικές επιλογές που αποσκοπούν στην πρόκληση εντυπώσεων και την αύξηση της viral διάδοσης (Rubin et al., 2015). Για τον σκοπό αυτό, χρησιμοποιούνται τεχνικές όπως TF-IDF, n-grams, ανάλυση συναισθήματος, lexical features και συχνά συνδυάζονται με word embeddings ή προεμπαιδευμένα γλωσσικά μοντέλα όπως το BERT, που επιτρέπουν την καλύτερη κατανόηση των συμφραζομένων και την αναγνώριση των υπονοουμένων και της ειρωνείας (Devlin et al., 2019).

Η δεύτερη προσέγγιση, που αφορά την ανάλυση του μοτίβου διάδοσης μιας πληροφορίας, βασίζεται στην υπόθεση ότι οι ψευδείς ειδήσεις διαχέονται διαφορετικά σε σχέση με τις έγκυρες. Για παράδειγμα, έχει βρεθεί ότι τα fake news συχνά εξαπλώνονται ταχύτερα, διαμέσου χρηστών που είναι λιγότερο διασυνδεδεμένοι με επίσημους φορείς και έχουν μικρότερη αξιοπιστία ή ιστορικό spam (Vosoughi et al., 2018). Τέτοιες αναλύσεις αξιοποιούν τη θεωρία γράφων και τεχνικές ανάλυσης κοινωνικών δικτύων, εντοπίζοντας

κοινότητες χρηστών που αλληλεπιδρούν συστηματικά με ύποπτο περιεχόμενο ή παρακολουθώντας την "πορεία" της πληροφορίας μέσα από τον κοινωνικό ιστό. Αλγόριθμοι όπως το Propagation Tree Kernel και τα Graph Convolutional Networks έχουν εφαρμοστεί με επιτυχία για την αναπαράσταση των σχέσεων μεταξύ των χρηστών και των αναρτήσεων.

Ένα κρίσιμο πρόβλημα στην ανίχνευση παραπληροφόρησης είναι η ετικετοποίηση των δεδομένων. Η δημιουργία επισημασμένων συνόλων δεδομένων (labeled datasets) απαιτεί συχνά ανθρώπινη επιμέλεια, ή εξάρτηση από fact-checking οργανισμούς, γεγονός που περιορίζει την κλίμακα των δεδομένων και την επικαιρότητα των μοντέλων. Παράλληλα, η έννοια της «ψευδούς είδησης» δεν είναι πάντα απόλυτη, αλλά περιέχει και γκριζες ζώνες, όπως οι παρερμηνείες, οι ανακρίβειες και οι ημι-αληθείς πληροφορίες, καθιστώντας την ταξινόμηση πιο υποκειμενική.

Αξιοσημείωτο είναι ότι τα μοντέλα μηχανικής μάθησης μπορούν να χρησιμοποιηθούν όχι μόνο για την ανίχνευση, αλλά και για την εξήγηση (explanation) των λόγων που ένα κείμενο θεωρείται ψευδές. Η χρήση attention mechanisms ή explainable AI τεχνικών προσφέρει τη δυνατότητα κατανόησης της «λογικής» του μοντέλου, ενισχύοντας τη διαφάνεια και την αποδοχή από το κοινό ή τις αρχές. Σε εφαρμογές δημόσιας πολιτικής και ασφάλειας, η ερμηνευσιμότητα είναι κρίσιμη, καθώς η απόφαση ενός αλγορίθμου μπορεί να οδηγήσει σε απόκρυψη, επισήμανση ή αφαίρεση περιεχομένου.

Οι σύγχρονες προσεγγίσεις ανιχνεύουν fake news και μέσω πολυτροπικής πληροφορίας: δηλαδή συνδυάζοντας κείμενο, εικόνα, βίντεο και μεταδεδομένα. Για παράδειγμα, η ίδια είδηση συνοδεύεται πολλές φορές από εικόνες που μπορεί να είναι τροποποιημένες ή να προέρχονται από άσχετα γεγονότα. Η χρήση CNNs για ανάλυση εικόνων και μοντέλων OCR για εντοπισμό κειμένου σε εικόνες, προσφέρει έναν επιπλέον άξονα για την ανίχνευση απάτης και χειραγώγησης. Τα fake news detection pipelines πλέον ενσωματώνουν και metadata όπως το ιστορικό του χρήστη, τον αριθμό των αλληλεπιδράσεων και τη χρονική ακολουθία της διάδοσης, προσφέροντας έναν πιο σφαιρικό και ακριβή μηχανισμό αξιολόγησης (Rubin et. al., 2016).

Είναι σημαντικό να σημειωθεί ότι η ανίχνευση ψευδών ειδήσεων δεν μπορεί να επιτευχθεί αποκλειστικά με τεχνικά μέσα. Η ανθρώπινη κρίση, η δημοσιογραφική επιμέλεια και η διασταύρωση πηγών παραμένουν κρίσιμα εργαλεία. Παρ' όλα αυτά, η μηχανική μάθηση και τα συστήματα τεχνητής νοημοσύνης προσφέρουν απαραίτητες υποδομές για μαζική ανίχνευση και φιλτράρισμα περιεχομένου, σε κλίμακες που ξεπερνούν κατά πολύ τις δυνατότητες της χειροκίνητης παρακολούθησης.

Η αντιμετώπιση του φαινομένου της παραπληροφόρησης απαιτεί έναν πολυπαραγοντικό και συνεργατικό μηχανισμό, στον οποίο η μηχανική μάθηση, η ανάλυση κοινωνικών δικτύων, η

επιστήμη της πληροφορίας και η δεοντολογία οφείλουν να λειτουργούν συμπληρωματικά. Ο στόχος δεν είναι η απλή απομάκρυνση λανθασμένων αναρτήσεων, αλλά η ενίσχυση της κριτικής σκέψης, η προώθηση της διαφάνειας και η αποκατάσταση της εμπιστοσύνης στην πληροφορία.

## 6.2 Στοχευμένη Διαφήμιση και Προβλέψεις Συμπεριφοράς Χρηστών

Η στοχευμένη διαφήμιση (targeted advertising) και η πρόβλεψη συμπεριφοράς χρηστών (user behavior prediction) αποτελούν δύο από τις πλέον κερδοφόρες και τεχνολογικά προηγμένες εφαρμογές της μηχανικής μάθησης στα κοινωνικά δίκτυα. Ο πυρήνας τους στηρίζεται στη συλλογή, ανάλυση και αξιοποίηση μεγάλων ποσοτήτων προσωπικών δεδομένων για τη βελτιστοποίηση της εμπορικής απήχησης και τη μεγιστοποίηση της απόδοσης των διαφημιστικών καμπανιών (Kosinski et al., 2013).

Η στοχευμένη διαφήμιση βασίζεται στην ιδέα ότι τα μηνύματα marketing είναι πιο αποτελεσματικά όταν απευθύνονται σε χρήστες με συγκεκριμένα ενδιαφέροντα, δημογραφικά χαρακτηριστικά ή πρότυπα συμπεριφοράς. Οι αλγόριθμοι επιτρέπουν την αναγνώριση τέτοιων προτύπων μέσω της καταγραφής δεδομένων όπως οι προτιμήσεις (likes), οι αναζητήσεις, οι αλληλεπιδράσεις με περιεχόμενο, οι επισκέψεις σε ιστοσελίδες, ακόμη και το είδος των emojis που χρησιμοποιεί κάποιος (Chaffey & Ellis-Chadwick, 2019). Ο συνδυασμός τέτοιων δεδομένων δημιουργεί ένα ψυχολογικό προφίλ του χρήστη, βάσει του οποίου μπορεί να προβλεφθεί η ανταπόκρισή του σε μια συγκεκριμένη διαφήμιση ή προσφορά.

Οι τεχνικές που χρησιμοποιούνται περιλαμβάνουν supervised learning για την πρόβλεψη της πιθανότητας ενός click (CTR prediction), unsupervised learning για την τμηματοποίηση χρηστών (clustering), καθώς και reinforcement learning για την βελτιστοποίηση διαφημιστικών στρατηγικών σε πραγματικό χρόνο. Μοντέλα όπως τα Logistic Regression, Random Forests και Gradient Boosting χρησιμοποιούνται εκτενώς για την πρόβλεψη της αλληλεπίδρασης του χρήστη με μια διαφήμιση, ενώ τα deep learning μοντέλα επιτρέπουν τη σύνθεση δεδομένων από πολλαπλές πηγές, όπως εικόνες και κείμενα (Yu et al., 2022).

Η πρόβλεψη συμπεριφοράς χρηστών εστιάζει στην πρόβλεψη μελλοντικών ενεργειών ή προτιμήσεων βάσει ιστορικών δεδομένων. Αυτό μπορεί να αφορά προβλέψεις όπως: ποιος χρήστης είναι πιθανότερο να κάνει like ή share ένα post, ποιος πρόκειται να εγκαταλείψει την πλατφόρμα (churn prediction), ή ποιος θα ανταποκριθεί σε μια προσφορά. Η εφαρμογή RNNs και LSTMs είναι ιδιαίτερα αποτελεσματική όταν τα δεδομένα έχουν χρονική διάσταση, όπως για παράδειγμα η αλληλουχία των clicks ή των views σε μια συνεδρία χρήστη.

Ένα από τα πιο επιτυχημένα παραδείγματα στοχευμένης διαφήμισης είναι η λειτουργία των lookalike audiences του Facebook, όπου ο διαφημιζόμενος μπορεί να στοχεύσει χρήστες που μοιάζουν προφίλ-ια με εκείνους που ήδη έχουν αλληλεπιδράσει με το brand. Τα lookalike μοντέλα βασίζονται σε high-dimensional vector representations του χρήστη, οι οποίες εκπαιδεύονται με deep learning τεχνικές. Επιπλέον, η χρήση τεχνικών collaborative

filtering – εμπνευσμένων από τα συστήματα συστάσεων (recommendation systems) – επιτρέπει την πρόβλεψη προϊόντων που πιθανώς ενδιαφέρουν τον χρήστη, βάσει της συμπεριφοράς άλλων με παρόμοια χαρακτηριστικά (Koren et. al., 2009).

Η στόχευση γίνεται όχι μόνο με βάση το περιεχόμενο, αλλά και με βάση τη χρονική συγκυρία (temporal targeting). Οι πλατφόρμες μπορούν να καθορίσουν τη βέλτιστη στιγμή για την εμφάνιση μιας διαφήμισης ή να προβλέψουν πότε είναι πιο πιθανό ο χρήστης να απαντήσει θετικά. Χρησιμοποιώντας μοντέλα χρονικών σειρών, όπως τα Prophet, ARIMA, ή Temporal Fusion Transformers, επιτυγχάνεται προσωποποιημένη βελτιστοποίηση του marketing.

Ωστόσο, η χρήση τέτοιων τεχνικών συνοδεύεται από σοβαρά ηθικά και κοινωνικά ζητήματα. Η υπερβολική στόχευση (microtargeting) μπορεί να οδηγήσει σε φαινόμενα «φούσικας πληροφορίας» (filter bubble), περιορίζοντας την έκθεση του χρήστη σε διαφορετικές απόψεις και περιεχόμενο. Επίσης, η χρήση ευαίσθητων προσωπικών δεδομένων χωρίς συναίνεση εγείρει ζητήματα ιδιωτικότητας, ειδικά όταν οι προβλέψεις αγγίζουν πεδία όπως η πολιτική προτίμηση, η ψυχική κατάσταση ή η σεξουαλικότητα (Zuboff, 2019).

Η Ευρωπαϊκή Νομοθεσία και συγκεκριμένα ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR), έχει θέσει αυστηρούς περιορισμούς στη χρήση προσωπικών δεδομένων για αυτοματοποιημένες αποφάσεις, όπως η στοχευμένη διαφήμιση. Αυτό ενισχύει την ανάγκη για explainable AI, ώστε οι χρήστες να κατανοούν τον τρόπο με τον οποίο επιλέγονται ή αποκλείονται από μια διαφημιστική εκστρατεία.

Η τάση πλέον κινείται προς τα hybrid models, τα οποία συνδυάζουν δεδομένα περιεχομένου, δομής (social graph), χρονικής ροής και metadata, επιτρέποντας την καλύτερη αποτύπωση του πλαισίου. Παράλληλα, η εισαγωγή των Generative AI τεχνολογιών ανοίγει νέες προοπτικές για τη δημιουργία προσωποποιημένων μηνυμάτων διαφήμισης, καθώς τα μοντέλα μπορούν να προσαρμόσουν το διαφημιστικό περιεχόμενο στη γλώσσα, το ύφος και τις προτιμήσεις του κάθε χρήστη (Brown et al., 2020).

Εν κατακλείδι, η στοχευμένη διαφήμιση και οι προβλέψεις συμπεριφοράς είναι κομβικές λειτουργίες της μηχανικής μάθησης στα κοινωνικά δίκτυα, προσφέροντας εντυπωσιακά αποτελέσματα σε επίπεδο εμπορικής αξιοποίησης. Ωστόσο, απαιτείται σταθερή ισορροπία μεταξύ αποδοτικότητας, διαφάνειας και σεβασμού στην ιδιωτικότητα, ώστε οι τεχνολογίες αυτές να υπηρετούν όχι μόνο τους επιχειρηματικούς σκοπούς, αλλά και το δημόσιο συμφέρον.

### 6.3 Ανίχνευση Ασυνήθιστης Συμπεριφοράς και Spam

Η ταχεία ανάπτυξη των κοινωνικών μέσων δικτύωσης, σε συνδυασμό με την ελευθερία ανάρτησης και αλληλεπίδρασης που προσφέρουν, έχει οδηγήσει στη μαζική παραγωγή και διάδοση περιεχομένου από χρήστες διαφορετικών προθέσεων και κινήτρων. Παρότι η πλειοψηφία των δραστηριοτήτων είναι γνήσια και οργανική, ένα ολοένα και μεγαλύτερο ποσοστό της διαδικτυακής δραστηριότητας προέρχεται από λογαριασμούς ή μηχανισμούς που σκοπό έχουν είτε την παραπλάνηση, είτε τη χειραγώγηση του δημόσιου διαλόγου, είτε την αποστολή ανεπιθύμητων ή αυτοματοποιημένων μηνυμάτων (spam). Το πρόβλημα της ανίχνευσης τέτοιων συμπεριφορών είναι ιδιαίτερα κρίσιμο, καθώς συνδέεται τόσο με την ακεραιότητα της πλατφόρμας όσο και με ευρύτερες κοινωνικές, πολιτικές και οικονομικές επιπτώσεις (Ferrara et al., 2016).

Η έννοια της ασυνήθιστης συμπεριφοράς στις κοινωνικές πλατφόρμες αναφέρεται σε αποκλίσεις από τα στατιστικά πρότυπα που εμφανίζει το σύνολο των χρηστών. Αυτές οι αποκλίσεις μπορεί να εκδηλώνονται με εξαιρετικά υψηλή ή μη φυσιολογική δραστηριότητα (π.χ., μαζικά posts σε σύντομο χρονικό διάστημα), με περιορισμένη θεματική ποικιλία (π.χ., συνεχείς αναρτήσεις για το ίδιο θέμα ή hashtag), ή ακόμη και με ασυνήθιστα πρότυπα δικτύωσης (π.χ., αλληλεπίδραση με μικρό αριθμό επαναλαμβανόμενων λογαριασμών). Το πρόβλημα καθίσταται πιο περίπλοκο όταν τέτοιες συμπεριφορές είναι συγκαλυμμένες ή «καμουφλαρισμένες» ώστε να μιμούνται οργανικές αλληλεπιδράσεις. Στο πλαίσιο αυτό, οι τεχνικές ανίχνευσης ανωμαλιών και spam καθίστανται θεμελιώδεις, αξιοποιώντας αλγόριθμους που προσεγγίζουν το πρόβλημα είτε με επιβλεπόμενη είτε με μη επιβλεπόμενη μάθηση (Chandola et al., 2009).

Η επιβλεπόμενη μάθηση χρησιμοποιείται κυρίως στην περίπτωση που υπάρχει επαρκής αριθμός ετικετών (labels) για να διαχωριστεί η κανονική από την ανεπιθύμητη συμπεριφορά. Αλγόριθμοι όπως οι Support Vector Machines, τα Random Forests και τα k-Nearest Neighbors εκπαιδεύονται ώστε να διακρίνουν, βάσει χαρακτηριστικών όπως η χρονικότητα αναρτήσεων, ο λόγος αρχικού προς αναδημοσιευμένο περιεχόμενο και η συχνότητα εμφάνισης συνδέσμων, εάν ένας λογαριασμός παρουσιάζει χαρακτηριστικά κανονικής ή ύποπτης χρήσης. Οι ίδιες μέθοδοι μπορούν να εφαρμοστούν και για την κατηγοριοποίηση δημοσιεύσεων που παραπέμπουν σε spam, με βάση το λεξιλόγιο, την επαναληψιμότητα και το engagement.

Σε περιπτώσεις όπου δεν υπάρχουν labels – δηλαδή δεν γνωρίζουμε εκ των προτέρων ποια είναι τα ύποπτα παραδείγματα – χρησιμοποιούνται τεχνικές μη επιβλεπόμενης μάθησης, όπως οι autoencoders, το One-Class SVM και το Isolation Forest. Αυτοί οι αλγόριθμοι προσπαθούν να μάθουν τα κανονικά πρότυπα των δεδομένων και να εντοπίσουν τις εξαιρέσεις, δηλαδή τα σημεία που αποκλίνουν σημαντικά από το υπόλοιπο σύνολο. Ιδιαίτερη έμφαση δίνεται στα autoencoders, που επιχειρούν να συμπιέσουν τα δεδομένα σε έναν

συμπαγή, ενδιάμεσο χώρο αναπαράστασης και να τα αναδομήσουν με όσο το δυνατόν μικρότερη απώλεια πληροφορίας. Οι παρατηρήσεις που δεν μπορούν να αναδομηθούν αποτελεσματικά θεωρούνται ανωμαλίες.

Τα τελευταία χρόνια, το πεδίο έχει στραφεί σε αλγορίθμους βαθιάς μάθησης και ειδικά σε αρχιτεκτονικές όπως τα Graph Neural Networks (GNNs), οι οποίες αξιοποιούν την εγγενή γραφική δομή των κοινωνικών δικτύων. Με αυτά τα μοντέλα είναι δυνατή η αναπαράσταση όχι μόνο των χαρακτηριστικών των κόμβων (δηλαδή των χρηστών ή των posts), αλλά και των μεταξύ τους συνδέσεων. Αυτό καθιστά τα GNNs ιδιαίτερα ισχυρά για την ανίχνευση συντονισμένων και συγκαλυμμένων καμπανιών, στις οποίες πολλαπλοί λογαριασμοί λειτουργούν με παρόμοιο τρόπο για να παρακάμψουν τους κανόνες της πλατφόρμας (Pacheco et al., 2021).

Η περίπτωση του Twitter είναι χαρακτηριστική: έχουν παρατηρηθεί botnets τα οποία λειτουργούν με ακρίβεια δευτερολέπτου για να προωθήσουν hashtags ή να διαδώσουν παραπλανητικό περιεχόμενο. Στην ανάλυσή τους χρησιμοποιούνται τόσο χρονικά δεδομένα όσο και δομικά χαρακτηριστικά του κοινωνικού γράφου, με σκοπό την αναγνώριση συγχρονισμένων προτύπων. Αντίστοιχα, στο Facebook έχουν παρατηρηθεί οργανωμένα δίκτυα spam με λογαριασμούς που λειτουργούν «υπογείως» για διαφημιστικούς ή πολιτικούς σκοπούς. Εδώ, η αξιοποίηση metadata, όπως η IP, η συχνότητα εισόδου/εξόδου και το interaction με άλλους λογαριασμούς, προσφέρουν κρίσιμες ενδείξεις για την ανίχνευση τέτοιων περιπτώσεων (Subrahmanian et al., 2016).

Ένα ιδιαίτερο πρόβλημα αφορά τη συντονισμένη μη αυθεντική δραστηριότητα (coordinated inauthentic behavior), η οποία εκτελείται από ομάδες λογαριασμών με κοινή στρατηγική. Στις περιπτώσεις αυτές, η απλή μοντελοποίηση της δραστηριότητας σε ατομικό επίπεδο είναι ανεπαρκής. Αντίθετα, απαιτείται συστημική προσέγγιση, με ταυτόχρονη ανάλυση χρονικών, θεματικών και δομικών συσχετίσεων, που οδηγεί στην κατασκευή similarity graphs και dynamic subgraph analysis. Σε πολλές περιπτώσεις, η ύπαρξη εξαιρετικά παρόμοιων προτύπων δημοσιεύσεων, η χρήση ίδιων hashtags σε μικρά χρονικά παράθυρα, ή οι κοινές πηγές περιεχομένου, αποτελούν ισχυρές ενδείξεις οργανωμένης καμπάνιας (Ferrara et al., 2016).

Ωστόσο, η προσπάθεια αυτοματοποιημένης ανίχνευσης δεν είναι χωρίς κινδύνους. Ο σημαντικότερος είναι η εσφαλμένη αναγνώριση κανονικής συμπεριφοράς ως spam ή κακόβουλης (false positive), κάτι που μπορεί να έχει επιπτώσεις στην ελευθερία του λόγου και στην αξιοπιστία των μηχανισμών moderation. Για τον λόγο αυτό, προτείνεται συχνά ο συνδυασμός αλγοριθμικών και ανθρώπινων συστημάτων επιβεβαίωσης, καθώς και η ανάπτυξη ερμηνεύσιμων μοντέλων που μπορούν να δικαιολογούν τις αποφάσεις τους (explainable AI).

Συνοψίζοντας, η ανίχνευση ασυνήθιστης συμπεριφοράς και spam στα κοινωνικά δίκτυα συνιστά έναν πολυδιάστατο τομέα που συνδυάζει στοιχεία επεξεργασίας φυσικής γλώσσας, ανάλυσης κοινωνικών γράφων, στατιστικής και βαθιάς μάθησης. Η συνεχιζόμενη εξέλιξη των απειλών επιβάλλει τη διαρκή προσαρμογή των μοντέλων, ενώ οι ηθικές και νομικές

διαστάσεις του θέματος απαιτούν απόλυτη διαφάνεια, υπευθυνότητα και συμμόρφωση με τους κανονισμούς προστασίας προσωπικών δεδομένων.

## 6.4 Ανάλυση Επιρροής (*Influence Analysis*)

Η ανάλυση επιρροής (influence analysis) στα κοινωνικά δίκτυα αποτελεί έναν από τους πιο στρατηγικούς τομείς έρευνας και εφαρμογής της μηχανικής μάθησης και της ανάλυσης κοινωνικών γραφημάτων, με εφαρμογές που εκτείνονται από το μάρκετινγκ και την πολιτική επικοινωνία, μέχρι τη δημόσια ασφάλεια και την πληροφορική κοινωνιολογία. Στο επίκεντρο αυτής της μεθοδολογίας βρίσκεται η επιδίωξη εντοπισμού των ατόμων, ομάδων ή κόμβων ενός κοινωνικού δικτύου που διαθέτουν τη μεγαλύτερη ικανότητα να επηρεάσουν την πληροφοριακή ή συμπεριφορική ροή εντός του συστήματος (Katz & Lazarsfeld, 1955).

Η επιρροή δεν ταυτίζεται απαραίτητα με την αριθμητική δημοτικότητα. Ένας χρήστης με πολλούς ακολούθους (followers) δεν είναι κατ' ανάγκη επιδραστικός, εάν το περιεχόμενό του δεν ενεργοποιεί αλληλεπιδράσεις (engagement) ή εάν το δίκτυό του αποτελείται από κόμβους χαμηλής συνδεσιμότητας. Αντιθέτως, ένας χρήστης με μικρό αριθμό συνδέσεων αλλά υψηλό ποσοστό αναμεταδόσεων ή επιρροής στους κύριους κόμβους μπορεί να έχει σημαντικά υψηλότερο impact. Επομένως, η επιρροή είναι μία πολύπλοκη έννοια, πολυδιάστατη και σχεσιακή, που μελετάται μέσω ποσοτικών και ποιοτικών αναλύσεων (Freeman, 1979).

Οι μεθοδολογίες ανάλυσης επιρροής βασίζονται κυρίως στη θεωρία γράφων και στην εξαγωγή μετริกών από την τοπολογία του κοινωνικού δικτύου. Οι πιο διαδεδομένες μετρικές είναι η degree centrality, που μετρά τον αριθμό των άμεσων συνδέσεων ενός κόμβου, η betweenness centrality, που μετρά πόσο συχνά ένας κόμβος λειτουργεί ως «γέφυρα» σε συντομότερα μονοπάτια άλλων κόμβων και η eigenvector centrality, που προσδιορίζει την επιρροή ενός κόμβου με βάση την επιρροή των γειτόνων του (Bonacich, 1987).

Η PageRank, αρχικά ανεπτυγμένη από την Google για την κατάταξη ιστοσελίδων, αποτελεί επίσης μία από τις πιο αποδοτικές μεθόδους αξιολόγησης επιρροής. Στο πλαίσιο των κοινωνικών δικτύων, το PageRank αποδίδει υψηλότερη επιρροή σε κόμβους που έχουν ισχυρούς γείτονες, δίνοντας έμφαση στην ποιότητα των συνδέσεων και όχι απλώς στην ποσότητα (Brin & Page, 1998). Ενδεικτικά, ένας χρήστης στο Twitter που αλληλεπιδρά τακτικά με ηγέτες κοινής γνώμης, μπορεί να κατατάσσεται ψηλά σε PageRank ακόμη και με περιορισμένο αριθμό followers.

Η επιρροή μπορεί να μελετηθεί και σε χρονικό βάθος. Ορισμένα μοντέλα, όπως το Influence Cascade Model, αναλύουν τη διάδοση πληροφοριών (information diffusion) και εξετάζουν την ικανότητα ενός κόμβου να προκαλεί δευτερογενείς και τριτογενείς αναμεταδόσεις (Kempe et al., 2003). Οι Independent Cascade και Linear Threshold Models αποτελούν βασικά εργαλεία για την προσομοίωση της μετάδοσης ενός μηνύματος ή ιδέας μέσα στο δίκτυο, υπολογίζοντας τη «δύναμη ενεργοποίησης» των επιδραστικών κόμβων.

Η επιρροή δεν είναι πάντοτε θετική ή ουδέτερη. Στην πράξη, παρατηρείται ότι λογαριασμοί με έντονη τοξικότητα ή ψευδοπληροφόρηση ενδέχεται να έχουν υψηλό επίπεδο διάδοσης

λόγω των ακραίων θέσεων που υιοθετούν. Η αρνητική επιρροή, όπως και η επικίνδυνη χειραγώγηση της κοινής γνώμης από αυτόκλητους influencers, καθιστά την ανάλυση επιρροής απαραίτητο εργαλείο για τη στρατηγική διαχείριση του δημόσιου διαλόγου αλλά και για την προστασία του κοινωνικού ιστού (Ferrara, 2017).

Σε επίπεδο εφαρμογής, η ανάλυση επιρροής χρησιμοποιείται στο influencer marketing για την επιλογή κατάλληλων προσώπων προς προώθηση προϊόντων, με βάση την πραγματική τους απήχηση και όχι τα επιφανειακά metrics. Επίσης, χρησιμοποιείται από πολιτικά επιτελεία, οργανώσεις και μέσα ενημέρωσης για τον εντοπισμό χρηστών που καθορίζουν την ατζέντα ή μετατοπίζουν την κατεύθυνση του διαλόγου (Bovet & Makse, 2019). Επιπλέον, σε περιόδους κρίσεων, όπως κατά τη διάρκεια φυσικών καταστροφών ή πολιτικών εξεγέρσεων, η ταχεία ταυτοποίηση των πιο επιδραστικών κόμβων επιτρέπει την αποτελεσματικότερη διάχυση κρίσιμων πληροφοριών.

Τα μοντέλα επιρροής εξελίσσονται συνεχώς ώστε να ενσωματώνουν πολυτροπικά δεδομένα, δηλαδή στοιχεία που προέρχονται ταυτόχρονα από κείμενα, εικόνες, βίντεο και αλληλεπιδράσεις. Η χρήση deep learning και ειδικά των Graph Convolutional Networks (GCNs) έχει ενισχύσει σημαντικά τη δυνατότητα εξαγωγής δυναμικών αναπαραστάσεων επιρροής, αξιοποιώντας ταυτόχρονα το περιεχόμενο και τη δομή του γράφου (Kipf & Welling, 2016). Παράλληλα, αναπτύσσονται metrics όπως το temporal influence, που λαμβάνει υπόψη όχι μόνο το ποιος επηρεάζει, αλλά και το πότε και για πόσο διάστημα.

Ωστόσο, η ανάλυση επιρροής θέτει και σημαντικά ηθικά και νομικά ζητήματα. Η στόχευση και η χειραγώγηση εύλωτων ομάδων, η ιδιωτικότητα και η ακούσια αποκάλυψη κοινωνικών σχέσεων, αποτελούν προκλήσεις που πρέπει να αντιμετωπίζονται με υπευθυνότητα. Το ζήτημα της ερμηνευσιμότητας είναι επίσης κρίσιμο: καθώς τα μοντέλα γίνονται πιο περίπλοκα, η ανάγκη κατανόησης του γιατί ένα άτομο θεωρείται επιδραστικό καθίσταται θεμελιώδης, τόσο για διαφανείς διαδικασίες όσο και για λογοδοσία.

Η ανάλυση επιρροής αποτελεί σήμερα έναν από τους βασικούς πυλώνες της αναλυτικής επιστήμης των κοινωνικών δικτύων. Η συνεχής εξέλιξη των εργαλείων και τεχνικών της, η διεπιστημονική της φύση και η στρατηγική της σημασία την καθιστούν πεδίο αυξημένου ερευνητικού και επιχειρηματικού ενδιαφέροντος. Όσο το τοπίο της διαδικτυακής κοινωνικής δυναμικής μεταβάλλεται, η κατανόηση της επιρροής θα συνεχίσει να αποτελεί κρίσιμο πλεονέκτημα για την επιβίωση και την επιτυχία ατόμων, οργανισμών και θεσμών στο ψηφιακό οικοσύστημα.

## 6.5 Ανάλυση Καμπανιών και Social Media Marketing

Η ανάλυση καμπανιών στα μέσα κοινωνικής δικτύωσης (Social Media Campaign Analysis) αποτελεί έναν κρίσιμο μηχανισμό αξιολόγησης της αποτελεσματικότητας των στρατηγικών ψηφιακού μάρκετινγκ. Η διάδοση περιεχομένου, η αλληλεπίδραση των χρηστών, η δημιουργία κοινοτήτων γύρω από ένα brand ή μήνυμα και η καταγραφή των επιδόσεων κάθε καμπάνιας συγκροτούν ένα σύνθετο αλλά μετρήσιμο οικοσύστημα, στο οποίο εμπλέκονται τεχνικές μηχανικής μάθησης, ανάλυσης συναισθήματος, δικτύωσης και οικονομικών δεδομένων (Tuten & Solomon, 2018).

Το social media marketing βασίζεται στην προσαρμοσμένη, συστηματική χρήση των κοινωνικών δικτύων για την προβολή προϊόντων, υπηρεσιών ή κοινωνικών πρωτοβουλιών, αξιοποιώντας τεχνικές targeting, οργανικό και πληρωμένο περιεχόμενο, influence marketing και αυτοματισμούς διαχείρισης. Η αναλυτική του διάσταση είναι αυτή που επιτρέπει τη βελτιστοποίηση των ενεργειών, με στόχο τη μέγιστη απόδοση (ROI), την κατανόηση του καταναλωτή και τη στρατηγική ανατροφοδότηση (Chaffey & Ellis-Chadwick, 2019).

Η ανάλυση μιας καμπάνιας ξεκινά από τον καθορισμό στόχων: επίτευξη προβολών (impressions), αύξηση αλληλεπιδράσεων (engagement), αύξηση επισκεψιμότητας (traffic), καταγραφές αγορών (conversions) ή ενίσχυση της φήμης και πιστότητας του brand. Οι στόχοι αυτοί επιδρούν άμεσα στον σχεδιασμό των μετρικών αξιολόγησης (KPIs), όπως οι CTR (Click-Through Rate), CPL (Cost per Lead), CPA (Cost per Acquisition) και CPM (Cost per Mille) (Kingsnorth, 2019).

Σε λειτουργικό επίπεδο, η ανάλυση καμπανιών βασίζεται σε πολυδιάστατη ανάλυση δεδομένων που προέρχονται από διαφορετικά κοινωνικά κανάλια (π.χ. Facebook, Instagram, TikTok, Twitter), τα οποία μπορεί να διαφέρουν σε μορφή, αλγόριθμους και χρήσεις. Η ενοποίηση των δεδομένων (data integration) και η εξαγωγή συγκρίσιμων δεικτών από ετερογενείς πηγές αποτελούν τεχνική και μεθοδολογική πρόκληση. Επιπλέον, η αλληλεπίδραση περιεχομένου (likes, shares, retweets, comments) δεν έχει πάντα την ίδια βαρύτητα. Ένα «share», για παράδειγμα, έχει μεγαλύτερη αξία διάδοσης από ένα «like», ενώ ένα αρνητικό σχόλιο μπορεί να έχει δυσανάλογο αντίκτυπο στην αντίληψη του κοινού (Heath & Heath, 2007).

Η ανάλυση αλγορίθμων engagement είναι ουσιώδης για την κατανόηση του τρόπου με τον οποίο το περιεχόμενο γίνεται viral ή όχι. Χρησιμοποιούνται μοντέλα πρόβλεψης διάδοσης, όπως τα Hawkes Processes και metrics όπως το virality coefficient ή το engagement velocity. Αυτά βοηθούν στον εντοπισμό των πιο αποτελεσματικών στιγμών και κατηγοριών περιεχομένου (Goel et al., 2015).

Η κατηγοριοποίηση περιεχομένου αποτελεί άλλη μια σημαντική πτυχή της ανάλυσης. Μέσα από τεχνικές NLP, γίνεται ανάλυση του είδους των αναρτήσεων (π.χ., ενημερωτικές, προωθητικές, ψυχαγωγικές) και η συσχέτισή τους με τα αποτελέσματα των καμπανιών. Ορισμένες κατηγορίες περιεχομένου, όπως τα συναισθηματικά φορτισμένα μηνύματα,

φαίνεται να προκαλούν υψηλότερα ποσοστά διαμοιρασμού και αλληλεπίδρασης, ιδιαίτερα όταν συνοδεύονται από πολυμεσιά στοιχεία (π.χ. εικόνες, βίντεο) (Berger & Milkman, 2012).

Η χρήση συνελκτικών νευρωνικών δικτύων (CNNs) για την ανάλυση οπτικού περιεχομένου, καθώς και recurrent neural networks (RNNs) για την ανάλυση χρονικών προτύπων συμπεριφοράς, συνιστούν πλέον καθιερωμένες πρακτικές. Επιπλέον, τα transformer-based μοντέλα, όπως το BERT και το GPT, επιτρέπουν την κατανόηση του περιεχομένου σε βάθος, προσφέροντας insight για τη συναισθηματική αντίδραση του κοινού (Devlin et al., 2019).

Σε επίπεδο προγνωστικής ανάλυσης, χρησιμοποιούνται μοντέλα μηχανικής μάθησης που προβλέπουν το performance μιας καμπάνιας, ανάλογα με τις λέξεις-κλειδιά, τα hashtags, το οπτικό στυλ και τη χρονική στιγμή δημοσίευσης. Η διαδικασία αυτή είναι ιδιαίτερα χρήσιμη για τη βελτιστοποίηση της στρατηγικής σε πραγματικό χρόνο (real-time optimization).

Η γεωγραφική και δημογραφική ανάλυση επιτρέπει την εξατομικευμένη στόχευση (micro-targeting), προσδιορίζοντας ποιοι τύποι περιεχομένου λειτουργούν καλύτερα ανά ηλικιακή ομάδα, φύλο, τοποθεσία ή ακόμα και ψυχογραφικό προφίλ. Ορισμένες εταιρείες προχωρούν σε συμπεριφορική μοντελοποίηση, συνδυάζοντας τα δεδομένα από social media με τρίτες βάσεις δεδομένων (third-party data), ώστε να προσεγγίσουν το κοινό με ακριβέστερα προσωποποιημένα μηνύματα (Zuboff, 2019).

Ωστόσο, η ανάλυση καμπανιών εγείρει σοβαρά ζητήματα δεοντολογίας και προστασίας προσωπικών δεδομένων. Η αμφισβήτηση της διαφάνειας στις μεθόδους στοχοποίησης και η ανάδυση φαινομένων όπως το Cambridge Analytica αναδεικνύουν την ανάγκη για ρυθμιστικό έλεγχο, ενημερωμένη συγκατάθεση και διατήρηση της εμπιστοσύνης του χρήστη (Cadwalladr & Graham-Harrison, 2018).

Στο πρακτικό επίπεδο, η ανάλυση καμπανιών ολοκληρώνεται με αναφορές και dashboards που επιτρέπουν τη γρήγορη αποτίμηση της απόδοσης. Χρησιμοποιούνται εργαλεία όπως Google Analytics, Meta Business Suite, HubSpot, Hootsuite, Sprout Social και Tableau για την οπτικοποίηση των KPIs και την ανάλυση της αποτελεσματικότητας ανά κανάλι, περιεχόμενο και χρονική περίοδο.

Η μελλοντική κατεύθυνση του social media marketing βασίζεται στην αυτοματοποιημένη βελτιστοποίηση (AI-driven optimization), όπου τα ίδια τα συστήματα μαθαίνουν από τα δεδομένα των καμπανιών και προσαρμόζουν δυναμικά το περιεχόμενο, το κοινό και τη χρονική στιγμή. Παράλληλα, η εξηγήσιμη τεχνητή νοημοσύνη (Explainable AI) επιτρέπει στους marketers να κατανοούν το «γιατί» πίσω από την επιτυχία ή την αποτυχία ενός μηνύματος, ενισχύοντας τη λογοδοσία και την εμπιστοσύνη.

Η ανάλυση καμπανιών και το social media marketing αποτελούν πυρήνα της σύγχρονης ψηφιακής στρατηγικής. Όσο τα κοινωνικά δίκτυα εξελίσσονται σε πλατφόρμες σύνθετης ανθρώπινης έκφρασης και εμπορικής δραστηριότητας, η συστηματική τους μελέτη και αξιοποίηση καθίστανται όχι απλώς επιθυμητές, αλλά αναγκαίες για κάθε οργανισμό που επιδιώκει να επιβιώσει και να αναπτυχθεί στο δυναμικό ψηφιακό περιβάλλον.

## 6.6 Deep Learning και Ανάλυση Πολυμεσικών Περιεχομένων (εικόνες, βίντεο)

Η ανάλυση πολυμεσικών δεδομένων – κυρίως εικόνων και βίντεο – αποτελεί αναπόσπαστο τμήμα της σύγχρονης μελέτης των κοινωνικών μέσων, δεδομένου ότι μεγάλο ποσοστό της επικοινωνίας στις πλατφόρμες αυτές γίνεται πλέον μέσω οπτικοακουστικού περιεχομένου. Το Instagram, το TikTok, το Snapchat και το YouTube, για παράδειγμα, βασίζονται πρωτίστως στην εικόνα, το βίντεο και τον ήχο, απαιτώντας νέες υπολογιστικές μεθόδους για την ανάλυση και κατανόηση των πολυμεσικών πληροφοριών (Βλαχοπούλου, 2020).

Η ραγδαία ανάπτυξη του Deep Learning και η ανάδειξη αρχιτεκτονικών όπως τα συνελκτικά νευρωνικά δίκτυα (Convolutional Neural Networks – CNNs) και τα επαναλαμβανόμενα νευρωνικά δίκτυα (Recurrent Neural Networks – RNNs) έχουν οδηγήσει σε εντυπωσιακή πρόοδο στην αυτόματη κατανόηση εικόνων και βίντεο. Τα CNNs, ιδιαίτερα, έχουν καθιερωθεί ως η κυρίαρχη τεχνική για την εξαγωγή οπτικών χαρακτηριστικών, την ταξινόμηση εικόνων και την αναγνώριση αντικειμένων (Krizhevsky et al., 2012). Μέσω της μάθησης ιεραρχικών χαρακτηριστικών – από απλές ακμές και υφές έως αφηρημένες έννοιες – τα CNNs επιτρέπουν στα υπολογιστικά μοντέλα να κατανοούν το περιεχόμενο μιας εικόνας με τρόπο παρόμοιο με τον ανθρώπινο εγκέφαλο.

Στα κοινωνικά μέσα, οι τεχνικές αυτές εφαρμόζονται σε πολλαπλά επίπεδα. Η αναγνώριση προσώπων (face detection and recognition), η ταυτοποίηση αντικειμένων και σκηνών, η ανάλυση έκφρασης προσώπου, ακόμη και η αισθητική ανάλυση εικόνων (image aesthetics analysis) χρησιμοποιούνται για την κατανόηση των συναισθημάτων, των ενδιαφερόντων και της εμπλοκής των χρηστών με συγκεκριμένο περιεχόμενο (Vinyals et al., 2015). Για παράδειγμα, οι εικόνες με φωτεινά χρώματα, ανθρώπινα πρόσωπα και υψηλή συμμετρία συσχετίζονται συχνά με αυξημένα επίπεδα engagement, κάτι που έχει αποδειχθεί εμπειρικά με deep learning-based μοντέλα πρόβλεψης (Khosla et al., 2014).

Η ανάλυση βίντεο απαιτεί ακόμα μεγαλύτερη υπολογιστική πολυπλοκότητα, καθώς περιλαμβάνει τόσο την κατανόηση των επιμέρους πλάνων όσο και τη χρονική αλληλουχία αυτών. Τα RNNs, ιδίως τα Long Short-Term Memory (LSTM) δίκτυα, χρησιμοποιούνται για την κατανόηση αυτών των χρονικών εξαρτήσεων, επιτρέποντας την ανάλυση εννοιών όπως κίνηση, συνέπεια δράσης και διάλογος. Επιπλέον, η συνδυασμένη χρήση CNNs για spatial features και LSTMs για temporal patterns (π.χ., στο μοντέλο CNN-LSTM) επιτρέπει την ταυτόχρονη ανάλυση εικόνας και κίνησης, γεγονός απαραίτητο για τη συναισθηματική ανάλυση βίντεο ή την κατηγοριοποίηση τύπων περιεχομένου (Donahue et al., 2015).

Ένα ιδιαίτερο ενδιαφέρον παρουσιάζουν τα multimodal deep learning συστήματα, τα οποία συνδυάζουν δεδομένα από διαφορετικούς τύπους – π.χ., εικόνα, κείμενο και ήχο – για την

κατανόηση του συνολικού νοήματος ενός μηνύματος. Η λεγόμενη πολυτροπική ανάλυση (multimodal analysis) επιτρέπει για παράδειγμα σε ένα μοντέλο να συνδέσει μια συγκεκριμένη έκφραση προσώπου με μια φράση σε ένα σχόλιο, για να κατανοήσει εάν η ανάρτηση είναι ειρωνική, σαρκαστική ή ενθουσιώδης. Τέτοιες προσεγγίσεις έχουν χρησιμοποιηθεί ευρέως στην πολυτροπική ανάλυση συναισθήματος (multimodal sentiment analysis) (Poria et al., 2016).

Η τεχνολογία μεταφοράς στυλ (style transfer) και η γενετική δημιουργία περιεχομένου (generative modeling), μέσω GANs (Generative Adversarial Networks), έχουν επίσης αξιοποιηθεί στα social media τόσο για δημιουργικούς σκοπούς (π.χ., φίλτρα προσώπου, avatar generation) όσο και για πιο σκοτεινές χρήσεις όπως τα deepfakes, γεγονός που καθιστά επιτακτική την ανάγκη για τεχνικές ανίχνευσης παραποιημένου περιεχομένου (Nguyen et al., 2019).

Επιπλέον, ένα αυξανόμενο πεδίο ενδιαφέροντος είναι η οπτική περιγραφή εικόνων (image captioning), δηλαδή η αυτόματη παραγωγή κειμένου που περιγράφει τι απεικονίζεται σε μια εικόνα. Τα πιο σύγχρονα μοντέλα (όπως το Show and Tell και το CLIP) βασίζονται σε συνδυασμούς CNN και transformer-based αρχιτεκτονικών, επιτρέποντας την αποδοτική σύνδεση γλώσσας και εικόνας (Radford et al., 2021).

Στον τομέα της διαφήμισης και του marketing, οι τεχνικές βαθιάς μάθησης χρησιμοποιούνται για την πρόβλεψη απόδοσης οπτικού περιεχομένου, τη δημιουργία visual embeddings που συσχετίζουν οπτικά χαρακτηριστικά με αποτελέσματα (π.χ. clicks, likes, shares) και τη στόχευση περιεχομένου σε συγκεκριμένα κοινά ανάλογα με τις προτιμήσεις τους. Στο TikTok, για παράδειγμα, ο αλγόριθμος αξιοποιεί πολυμεσικά δεδομένα για την ταξινόμηση των βίντεο και τη ροή προτάσεων με βάση την πιθανότητα engagement (Kaye et al., 2020).

Ωστόσο, η χρήση deep learning σε πολυμεσικά δεδομένα δεν είναι απαλλαγμένη από προκλήσεις. Πρώτον, απαιτείται μεγάλος υπολογιστικός πόρος, ειδικά για training σε raw video frames. Δεύτερον, υπάρχει υψηλή πολυπλοκότητα ως προς την ερμηνεία των μοντέλων – ένα CNN μπορεί να εντοπίσει ότι μια εικόνα συσχετίζεται με υψηλό engagement, αλλά είναι δύσκολο να εξηγηθεί *γιατί*. Αυτό έχει εντείνει τη συζήτηση γύρω από την ερμηνευσιμότητα (explainability) και την ανάγκη για διάφανες, δίκαιες και μη προικατελημμένες (unbiased) τεχνολογίες (Doshi-Velez & Kim, 2017).

Ακόμα, οι ηθικές προεκτάσεις της χρήσης οπτικών δεδομένων σε κοινωνικά δίκτυα αφορούν ζητήματα ιδιωτικότητας, συναίνεσης και παραπληροφόρησης. Οι πολιτικές πλατφορμών όπως το Meta και το YouTube περιλαμβάνουν πλέον συστήματα flagging και αυτοματοποιημένες ανιχνεύσεις περιεχομένου βασισμένες σε deep learning, ενώ ταυτόχρονα αναπτύσσονται πλαίσια όπως το Fairness, Accountability, and Transparency in Machine Learning (FAT/ML) για τη ρύθμιση τέτοιων εφαρμογών (Barocas et al., 2019).

Συνολικά, η ανάλυση πολυμεσικού περιεχομένου μέσω deep learning μετασχηματίζει τον τρόπο με τον οποίο αντιλαμβανόμαστε, αξιολογούμε και προβλέπουμε τη συμπεριφορά χρηστών στα κοινωνικά δίκτυα. Η συνδυαστική χρήση εικόνας, βίντεο, ήχου και κειμένου δημιουργεί νέες δυνατότητες για προβλέψεις, ταξινομήσεις, ανίχνευση συναισθημάτων και δημιουργία προσωποποιημένων εμπειριών, ενισχύοντας τόσο την εμπορική αξία των social media όσο και τις αναλυτικές δυνατότητες της κοινωνικής επιστήμης.

## *Κεφάλαιο 7: Πλεονεκτήματα και Προκλήσεις*

### *7.1 Οφέλη της Χρήσης Μηχανικής Μάθησης στα Μέσα Κοινωνικής Δικτύωσης*

Η ενσωμάτωση τεχνικών μηχανικής μάθησης (machine learning – ML) στα μέσα κοινωνικής δικτύωσης (social media) έχει οδηγήσει σε μια ριζική μεταμόρφωση του τρόπου με τον οποίο συλλέγεται, επεξεργάζεται και αξιοποιείται η πληροφορία στις ψηφιακές πλατφόρμες. Η ταχύτητα, η αυτοματοποίηση και η ικανότητα προσαρμογής των ML συστημάτων προσδίδουν πολυδιάστατα οφέλη που επηρεάζουν τόσο τις επιχειρήσεις όσο και τους απλούς χρήστες, τους ερευνητές και τους διαχειριστές πλατφορμών (Gandomi & Haider, 2015).

Ένα από τα βασικά πλεονεκτήματα αφορά τη βελτιστοποίηση της εμπειρίας χρήστη (user experience). Μέσω αλγορίθμων συστάσεων (recommendation systems), οι χρήστες εκτίθενται σε περιεχόμενο που είναι εξατομικευμένο με βάση τις προτιμήσεις, τις συμπεριφορές και τα ιστορικά δεδομένα τους. Πλατφόρμες όπως το YouTube, το Instagram και το TikTok χρησιμοποιούν deep learning για να προβλέψουν ποιο περιεχόμενο είναι πιθανότερο να τραβήξει την προσοχή ενός χρήστη, αυξάνοντας το engagement και το χρόνο παραμονής στην εφαρμογή (Covington, Adams, & Sargin, 2016). Η δυνατότητα αυτή δεν θα ήταν εφικτή χωρίς την αυτοματοποιημένη μάθηση και προσαρμογή των ML μοντέλων στις συνεχώς μεταβαλλόμενες συνήθειες των χρηστών.

Παράλληλα, η μηχανική μάθηση ενισχύει τη βελτίωση της αποδοτικότητας στον εντοπισμό και την κατηγοριοποίηση περιεχομένου. Οι αλγόριθμοι μπορούν να φιλτράρουν spam, να ανιχνεύσουν προσβλητικό ή επιβλαβές περιεχόμενο, να εντοπίσουν παραβιάσεις των όρων χρήσης και να ενισχύσουν τους μηχανισμούς moderation με ελάχιστη ανθρώπινη παρέμβαση (Schmidt & Wiegand, 2017). Αυτό είναι ιδιαίτερα κρίσιμο σε περιβάλλοντα όπου αναρτώνται εκατομμύρια μηνύματα ημερησίως, καθιστώντας την ανθρώπινη επιτήρηση ανεπαρκή και καθυστερημένη.

Ένα ακόμα καθοριστικό όφελος είναι η ανάλυση συναισθήματος και κοινωνικού κλίματος, επιτρέποντας την κατανόηση της δημόσιας γνώμης σε πραγματικό χρόνο. Με χρήση τεχνικών NLP (Natural Language Processing), η μηχανική μάθηση καθιστά δυνατή την αυτόματη ερμηνεία των συναισθημάτων και των απόψεων που εκφράζονται στα κοινωνικά μέσα, προσφέροντας πολύτιμα insights για πολιτική επικοινωνία, στρατηγικό marketing, ή ακόμη και για τη διαχείριση κρίσεων (Liu, 2012). Επιχειρήσεις και οργανισμοί μπορούν έτσι να παρακολουθούν τη φήμη τους, να αντιλαμβάνονται προληπτικά την εμφάνιση αρνητικών τάσεων και να ανταποκρίνονται ταχύτερα σε κοινωνικές αντιδράσεις.

Επιπροσθέτως, η ML υποστηρίζει την πρόβλεψη συμπεριφοράς και την προσωποποίηση υπηρεσιών. Τα predictive models, είτε βασισμένα σε supervised learning είτε σε time-series forecasting, μπορούν να εκτιμήσουν τη μελλοντική δημοτικότητα ενός περιεχομένου, να προβλέψουν την πιθανότητα απάντησης ή κοινοποίησης, αλλά και να προσδιορίσουν την

πιθανότητα αποχώρησης ενός χρήστη από την πλατφόρμα (user churn) (Li et al., 2021). Η πρόβλεψη αυτή ενισχύει τη διατήρηση πελατών και την αποτελεσματικότητα των διαφημιστικών εκστρατειών, καθώς προσφέρει τακτικές προσωποποιημένες παρεμβάσεις.

Στο πεδίο της διαφήμισης και της εμπορικής αξιοποίησης, η ML παρέχει εργαλεία για στοχευμένη διαφήμιση (targeted advertising), όπου οι καταναλωτές λαμβάνουν προωθητικά μηνύματα που ανταποκρίνονται στις ανάγκες και τις επιθυμίες τους. Μέσω της επεξεργασίας δεδομένων προτιμήσεων, clicks, αγορών και πλοήγησης, οι διαφημιστικοί αλγόριθμοι προσαρμόζουν δυναμικά τα διαφημιστικά banners και βίντεο σε πραγματικό χρόνο (Viktoratos & Tsadiras, 2021). Αυτή η διαδικασία οδηγεί σε αυξημένα ποσοστά μετατροπής και μειωμένο διαφημιστικό κόστος.

Από πλευράς επιστημονικής έρευνας, τα εργαλεία ML προσφέρουν δυνατότητες για ανακάλυψη κοινωνικών μοτίβων, δομής κοινότητας, επιρροής και διάδοσης ιδεών. Για παράδειγμα, με χρήση τεχνικών community detection και graph analysis, μπορούν να αναγνωριστούν ομάδες με κοινά ενδιαφέροντα, influencers, ή ακόμη και συντονισμένες καμπάνιες (Ferrara et al., 2016). Αυτές οι δυνατότητες συμβάλλουν σημαντικά στη μελέτη της σύγχρονης κοινωνικής συμπεριφοράς, της πολιτικής επικοινωνίας και της πληροφορικής κοινωνιολογίας.

Τέλος, η ML μπορεί να συμβάλει στη βελτίωση της ασφάλειας και της εμπιστοσύνης στα κοινωνικά δίκτυα. Η έγκαιρη ανίχνευση ύποπτων λογαριασμών, η παρακολούθηση παραπλανητικής πληροφόρησης και η διασφάλιση της γνησιότητας των χρηστών (π.χ., face verification μέσω CNNs), αποτελούν παραδείγματα κρίσιμων εφαρμογών που στηρίζονται σε τεχνικές deep learning και anomaly detection. Η αυτοματοποίηση αυτών των λειτουργιών μειώνει τους κινδύνους για τους χρήστες και προστατεύει τη δημόσια σφαίρα από κακόβουλες παρεμβάσεις (Subrahmanian et al., 2016).

Συμπερασματικά, η χρήση μηχανικής μάθησης στα μέσα κοινωνικής δικτύωσης παρέχει έναν εξαιρετικά πλούσιο μηχανισμό ανάλυσης, πρόβλεψης, κατανόησης και παρεμβατικής δράσης. Η ταχύτητα, η ευελιξία, η κλίμακα και η δυναμική προσαρμοστικότητα των αλγορίθμων ML ενισχύουν την αποτελεσματικότητα των κοινωνικών πλατφορμών, αναδιαμορφώνουν την εμπορική στρατηγική, διευκολύνουν την επιστημονική έρευνα και εν τέλει, μετασχηματίζουν την ίδια τη φύση της κοινωνικής επικοινωνίας στον ψηφιακό χώρο.

## 7.2 Τεχνικές και Θεωρητικές Προκλήσεις

Η εφαρμογή της μηχανικής μάθησης στα μέσα κοινωνικής δικτύωσης συνοδεύεται από ένα σύνολο πολυεπίπεδων προκλήσεων – τόσο τεχνικών όσο και θεωρητικών – που αναδεικνύουν τη δυναμικότητα αλλά και την πολυπλοκότητα του πεδίου. Οι προκλήσεις αυτές δεν αφορούν μόνο την επεξεργασία τεράστιων ποσοτήτων δεδομένων, αλλά επεκτείνονται σε ζητήματα γενίκευσης, μοντελοποίησης της κοινωνικής συμπεριφοράς, αποδοτικής αναπαράστασης πληροφορίας και αλγοριθμικής αποδοτικότητας (Zhang et al., 2018).

Μια από τις θεμελιώδεις τεχνικές προκλήσεις είναι η ετερογένεια των δεδομένων. Τα κοινωνικά δίκτυα περιλαμβάνουν δεδομένα διαφορετικών τύπων – κείμενα, εικόνες, βίντεο, ήχο, μεταδεδομένα, πληροφορίες από γραφήματα (π.χ. ποιος συνδέεται με ποιον) – τα οποία δεν μπορούν να αναλυθούν αποτελεσματικά με έναν ενιαίο αλγόριθμο ή εκπαιδευτικό μοντέλο. Η ανάγκη για πολυτροπική μοντελοποίηση (multimodal modeling) είναι καθοριστική, καθώς κάθε μορφή δεδομένων απαιτεί διαφορετικό τρόπο αναπαράστασης, εξαγωγής χαρακτηριστικών και συνδυασμού με άλλα (Dimitri, 2022).

Επιπλέον, τα δεδομένα είναι δυναμικά και εξελισσόμενα. Οι συνδέσεις, οι προτιμήσεις, οι αλληλεπιδράσεις και τα θέματα αλλάζουν διαρκώς, γεγονός που καθιστά τα στατικά μοντέλα ανεπαρκή. Τα μοντέλα μηχανικής μάθησης καλούνται να ενσωματώσουν τη χρονική διάσταση και να προσαρμόζονται σε online learning περιβάλλοντα, όπου η εισερχόμενη πληροφορία επηρεάζει άμεσα τις προβλέψεις και τις κατηγοριοποιήσεις (Aggarwal, 2011).

Η σημασιολογική ασάφεια είναι μία ακόμη βασική πρόκληση, ιδίως στα δεδομένα κειμένου. Η χρήση ειρωνείας, μεταφορών, πολυσημίας, καθώς και οι διαφορές ανάμεσα σε πολιτισμικά και γλωσσικά συμφραζόμενα, καθιστούν δύσκολη την αυτοματοποιημένη κατανόηση (Riloff et al., 2013). Παρότι τα μοντέλα βαθιάς μάθησης έχουν επιτύχει σημαντικές βελτιώσεις μέσω contextual embeddings, εξακολουθούν να υπάρχουν περιπτώσεις όπου το μοντέλο αποτυγχάνει να κατανοήσει την πραγματική πρόθεση πίσω από μια φράση.

Η ερμηνευσιμότητα (interpretability) των μοντέλων αποτελεί επίσης θεωρητικό και πρακτικό εμπόδιο, ειδικά όταν πρόκειται για βαθιά νευρωνικά δίκτυα τα οποία λειτουργούν ως «μαύρα κουτιά». Η κατανόηση του *γιατί* ένα σύστημα πήρε μια συγκεκριμένη απόφαση είναι κρίσιμη σε εφαρμογές όπως η ανίχνευση ψευδών ειδήσεων, η ανάλυση συναισθήματος ή η απομάκρυνση περιεχομένου, όπου εμπλέκονται ζητήματα ηθικής και λογοδοσίας (Doshi-Velez & Kim, 2017).

Παρά την πληθώρα των δεδομένων στα κοινωνικά δίκτυα, το ζήτημα των επισημασμένων παραδειγμάτων παραμένει κρίσιμο. Η διαδικασία του labeling απαιτεί ανθρώπινη εργασία, γνώση του πλαισίου και είναι συχνά ακριβή και χρονοβόρα. Ειδικά σε προβλήματα όπως η ανάλυση συναισθήματος, η ανίχνευση παραπληροφόρησης ή η ταξινόμηση ειδικών τύπων περιεχομένου, η ερμηνεία μπορεί να είναι υποκειμενική, αυξάνοντας την αβεβαιότητα στην εκπαίδευση των μοντέλων (Pan et al., 2020).

Ακόμα, η ανισορροπία μεταξύ κλάσεων (class imbalance) είναι έντονη: σε πολλές περιπτώσεις, τα θετικά παραδείγματα (π.χ., spam, fake news, τοξικά σχόλια) αποτελούν μικρό ποσοστό του συνόλου. Αυτή η ασυμμετρία επηρεάζει αρνητικά τους αλγορίθμους ταξινόμησης, που συχνά μαθαίνουν να προβλέπουν τη «κανονική» συμπεριφορά αγνοώντας τις σπάνιες και σημαντικές εξαιρέσεις. Η χρήση τεχνικών όπως oversampling, SMOTE, cost-sensitive learning και weighted loss functions είναι απαραίτητη, αλλά δεν λύνει πλήρως το πρόβλημα.

Ένα άλλο τεχνικό εμπόδιο είναι η υπολογιστική πολυπλοκότητα των σύγχρονων μοντέλων. Η επεξεργασία δεδομένων από πλατφόρμες όπως το Twitter ή το Facebook σε πραγματικό χρόνο απαιτεί επεξεργασία υψηλού όγκου, ταχύτητα και αποδοτική χρήση πόρων. Τα μεγάλα προεκπαιδευμένα μοντέλα (όπως τα BERT, GPT) είναι εξαιρετικά αποδοτικά στην κατανόηση γλώσσας, αλλά έχουν σημαντικό κόστος υπολογισμού και απαιτούν εξειδικευμένο hardware (π.χ., GPUs ή TPUs). Το trade-off μεταξύ ακρίβειας και αποδοτικότητας παραμένει σημαντικό ερώτημα, ιδίως για εφαρμογές σε edge συσκευές ή μικρότερες επιχειρήσεις (Sanh et al., 2019).

Ένα μεγάλο θεωρητικό ζήτημα είναι η ικανότητα γενίκευσης των μοντέλων. Τα περισσότερα μοντέλα μαθαίνουν από δεδομένα συγκεκριμένου χρονικού πλαισίου ή πολιτισμικού πλαισίου και δυσκολεύονται να γενικεύσουν σε νέα δεδομένα (π.χ., νέα trends, hashtags, memes). Η πρόκληση έγκειται στη δημιουργία ευέλικτων μοντέλων που να μπορούν να ανανεώνονται ή να επαναεκπαιδεύονται (model retraining) χωρίς να χάνουν την προηγούμενη γνώση (catastrophic forgetting), αλλά και χωρίς να απαιτούν εξ αρχής πλήρη επανεκπαίδευση.

Τέλος, η μηχανική μάθηση στα social media φέρνει μαζί της ένα σύνθετο πλέγμα ηθικών και κοινωνικών προκλήσεων, που σχετίζονται με την ιδιωτικότητα, τον αλγοριθμικό αποκλεισμό, τη μεροληψία (bias), αλλά και την ψυχολογική επίδραση των μοντέλων συστάσεων ή της στοχοποίησης χρηστών. Αν δεν ελεγχθούν, οι τεχνολογίες αυτές ενδέχεται να ενισχύσουν κοινωνικές διακρίσεις ή να παραπλανήσουν το κοινό (Zou & Schiebinger, 2018). Η ανάγκη για διαφάνεια, λογοδοσία και τεχνολογική ηθική είναι πιο επίκαιρη από ποτέ.

Η εφαρμογή της μηχανικής μάθησης στα κοινωνικά δίκτυα ανοίγει τεράστιες προοπτικές, αλλά ταυτόχρονα φέρνει στην επιφάνεια ένα πλέγμα τεχνικών και θεωρητικών προκλήσεων. Από την πολυπλοκότητα των δεδομένων και την ανάγκη για ερμηνευσιμότητα, μέχρι την εξισορρόπηση των δεδομένων και τα ηθικά διλήμματα, οι ερευνητές και οι μηχανικοί καλούνται να αναπτύξουν λύσεις που είναι όχι μόνο αποτελεσματικές, αλλά και δίκαιες, εξηγήσιμες και υπεύθυνες. Το μέλλον της μηχανικής μάθησης στα social media δεν θα καθοριστεί μόνο από τις τεχνικές καινοτομίες, αλλά και από την ικανότητα των ανθρώπων να τις εντάξουν με ορθό και ηθικό τρόπο στο κοινωνικό γίγνεσθαι.

### 7.3 Ζητήματα Ιδιωτικότητας και Ηθικής

Η ενσωμάτωση τεχνολογιών μηχανικής μάθησης στα μέσα κοινωνικής δικτύωσης εντείνει τα ηθικά και νομικά διλήμματα που σχετίζονται με την ιδιωτικότητα των χρηστών, τη χρήση και την επεξεργασία προσωπικών δεδομένων, τη διαφάνεια στις αποφάσεις των αλγορίθμων και τη δυνατότητα παρεμβάσεων σε κοινωνική κλίμακα χωρίς συναίνεση ή επίγνωση των εμπλεκομένων (Zuboff, 2019). Η αποθήκευση, η ανάλυση και η αξιοποίηση δεδομένων που προέρχονται από πλατφόρμες όπως το Facebook, το Twitter ή το TikTok φέρνουν στο προσκήνιο το ερώτημα: ποια είναι τα όρια μεταξύ τεχνολογικής καινοτομίας και σεβασμού της ανθρώπινης ελευθερίας και αξιοπρέπειας;

Η συναίνεση του χρήστη για τη συλλογή δεδομένων στα κοινωνικά δίκτυα θεωρείται, σε πολλές περιπτώσεις, ρητή μόνο στα χαρτιά. Οι περισσότεροι χρήστες αποδέχονται τους όρους χρήσης χωρίς να τους διαβάζουν, γεγονός που καθιστά την έννοια της συγκατάθεσης προβληματική. Επιπλέον, η πολυπλοκότητα των τεχνικών ανάλυσης – όπως η πρόβλεψη προσωπικότητας, η ανάλυση συναισθήματος ή η στόχευση βάσει ψυχομετρικών προφίλ – δημιουργούν έναν ψηφιακό καθρέφτη στον οποίο οι χρήστες εκτίθενται χωρίς να κατανοούν ούτε το εύρος ούτε τις επιπτώσεις αυτής της έκθεσης (Acquisti et. al., 2015).

Η αδιαφάνεια των αλγορίθμων (algorithmic opacity) ενισχύει την έλλειψη ελέγχου από τον χρήστη. Οι τεχνικές deep learning, αν και αποτελεσματικές, συχνά λειτουργούν ως «μαύρα κουτιά», αδυνατώντας να εξηγήσουν πώς καταλήγουν σε μία απόφαση (Burrell, 2016). Αυτό έχει σοβαρές συνέπειες σε περιπτώσεις λογοκρισίας περιεχομένου, αποκλεισμού λογαριασμών ή αποσιώπησης συγκεκριμένων φωνών – ιδίως όταν τέτοιες αποφάσεις λαμβάνονται χωρίς ανθρώπινη εποπτεία.

Τα δεδομένα των χρηστών στα social media δεν περιλαμβάνουν μόνο αυτά που δηλώνονται ρητά (π.χ. ονοματεπώνυμο, ηλικία), αλλά και προσωπικά συμπεράσματα που εξάγονται με αλγορίθμους, όπως πολιτικές προτιμήσεις, σεξουαλικός προσανατολισμός ή ψυχολογική κατάσταση (Kosinski et. al., 2013). Η χρήση τέτοιων προφίλ χωρίς ενημέρωση και συναίνεση αντιβαίνει στην αρχή της ιδιωτικότητας ως αυτονομίας, δηλαδή του δικαιώματος κάθε ατόμου να διαχειρίζεται την προσωπική του πληροφορία.

Το φαινόμενο αυτό αναδεικνύεται έντονα στην περίπτωση του Cambridge Analytica, όπου δεδομένα εκατομμυρίων χρηστών του Facebook χρησιμοποιήθηκαν για την πολιτική στόχευση και χειραγώγηση της κοινής γνώμης στις προεδρικές εκλογές των ΗΠΑ το 2016 και στο δημοψήφισμα για το Brexit (Isaak & Hanna, 2018). Παρότι το περιστατικό θεωρήθηκε εξαιρετικό, η δυνατότητα επανάληψής του παραμένει υπαρκτή, δεδομένης της διαθεσιμότητας των APIs και της ευκολίας πρόσβασης σε δημόσια δεδομένα.

Ένα άλλο σημαντικό ηθικό ζήτημα είναι η μεροληψία (bias) των αλγορίθμων. Τα μοντέλα μηχανικής μάθησης μαθαίνουν από ιστορικά δεδομένα, τα οποία συχνά αντικατοπτρίζουν κοινωνικές ανισότητες και προκαταλήψεις. Έτσι, η χρήση τους μπορεί να οδηγήσει στην ενίσχυση φυλετικών, φύλων ή κοινωνικών διακρίσεων, ακόμη και χωρίς πρόθεση (Barocas et.

al., 2019). Για παράδειγμα, η ανάλυση συναισθήματος μπορεί να ερμηνεύσει αρνητικά εκφράσεις που χρησιμοποιούνται συχνότερα σε συγκεκριμένες γλωσσικές κοινότητες, ή να στοχοποιήσει χρήστες που συμμετέχουν σε πολιτικό λόγο ως ακραίους ή επικίνδυνους.

Η ανίχνευση και διόρθωση του bias είναι μία από τις μεγαλύτερες ηθικές προκλήσεις στον τομέα και απαιτεί τόσο τεχνική προσέγγιση (fairness metrics, adversarial training) όσο και νομικό/θεσμικό έλεγχο.

Οι πλατφόρμες κοινωνικής δικτύωσης χρησιμοποιούν πλέον αυτοματοποιημένα συστήματα για τον εντοπισμό και την απομάκρυνση «επιβλαβούς» περιεχομένου. Όμως, η μεταβίβαση της ευθύνης από τον άνθρωπο στη μηχανή προκαλεί ανησυχίες σχετικά με το ποιος έχει τον έλεγχο του δημόσιου λόγου. Ο αλγόριθμος δεν έχει την ικανότητα να εντοπίσει τη λεπτότητα της ειρωνείας, της κριτικής ή του σαρκασμού, γεγονός που μπορεί να οδηγήσει σε αυθαίρετη λογοκρισία και περιορισμό της ελευθερίας έκφρασης (Gillespie, 2018).

Ταυτόχρονα, η ενίσχυση του φαινομένου echo chamber – δηλαδή η ενίσχυση των απόψεων που ήδη συμφωνούν με αυτές του χρήστη μέσω συστημάτων συστάσεων – μπορεί να οδηγήσει σε πολωμένο δημόσιο διάλογο, εξαλείφοντας την πολυφωνία και ενισχύοντας την ριζοσπαστικοποίηση.

Η προστασία των προσωπικών δεδομένων ρυθμίζεται από νομικά πλαίσια όπως ο Γενικός Κανονισμός για την Προστασία Δεδομένων (GDPR) στην Ευρωπαϊκή Ένωση, ο οποίος θέτει αυστηρές προδιαγραφές για τη συλλογή, την αποθήκευση και την επεξεργασία των δεδομένων. Ωστόσο, η ταχύτητα της τεχνολογικής εξέλιξης συχνά υπερβαίνει τις ρυθμιστικές δυνατότητες, καθιστώντας την τεχνολογική δεοντολογία ακόμη πιο απαραίτητη.

Οι ερευνητές και οι προγραμματιστές καλούνται να ενσωματώνουν αρχές ηθικής by design, δηλαδή από το στάδιο του σχεδιασμού των αλγορίθμων: διαφάνεια, δικαιοσύνη, προστασία ιδιωτικότητας και δυνατότητα ελέγχου από τον χρήστη.

Τα ζητήματα ιδιωτικότητας και ηθικής στη χρήση μηχανικής μάθησης στα μέσα κοινωνικής δικτύωσης αποτελούν μια πολυπαραγοντική πρόκληση, η οποία συνδυάζει την τεχνολογική, νομική, κοινωνική και ηθική διάσταση. Η προοπτική ενός ψηφιακού οικοσυστήματος που σέβεται την αυτονομία του χρήστη, διασφαλίζει τη δικαιοσύνη και λειτουργεί με διαφάνεια και λογοδοσία, δεν αποτελεί μόνο τεχνικό επίτευγμα, αλλά βαθιά ανθρωπιστική απαίτηση. Η αναγκαία ισορροπία μεταξύ καινοτομίας και προστασίας αποτελεί το μεγάλο στοίχημα της εποχής μας.

## 7.4 Ερμηνευσιμότητα και Διαφάνεια των Μοντέλων

Η ταχύτατη ενσωμάτωση αλγορίθμων μηχανικής μάθησης στα μέσα κοινωνικής δικτύωσης, ειδικά για εφαρμογές όπως ταξινόμηση περιεχομένου, στόχευση διαφημίσεων, ανάλυση συναισθήματος και ανίχνευση ψευδών ειδήσεων, έχει δημιουργήσει επείγουσα ανάγκη για ερμηνευσιμότητα (interpretability) και διαφάνεια (transparency) των μοντέλων. Η ερμηνευσιμότητα δεν είναι πολυτέλεια, αλλά απαραίτητη προϋπόθεση για την εμπιστοσύνη, την υπευθυνότητα και τη συμμόρφωση με νομικά και ηθικά πρότυπα (Doshi-Velez & Kim, 2017).

Ένα σύστημα μηχανικής μάθησης που λαμβάνει αποφάσεις χωρίς να μπορεί να εξηγή «πώς» και «γιατί» έφτασε σε αυτές – γνωστό και ως “black-box model” – δημιουργεί σημαντικά προβλήματα αξιοπιστίας, ελέγχου και αποδοχής. Αυτό είναι ιδιαίτερα κρίσιμο όταν οι αποφάσεις αφορούν ευαίσθητα ζητήματα, όπως αποκλεισμό λογαριασμών, ανάλυση πολιτικών διαθέσεων ή αυτόματη λογοκρισία (Lipton, 2018). Οι χρήστες, οι ελεγκτικές αρχές αλλά και οι ίδιοι οι σχεδιαστές των αλγορίθμων οφείλουν να γνωρίζουν τη λογική πίσω από τις προβλέψεις ενός συστήματος.

Η ερμηνευσιμότητα επιτρέπει:

- Την κατανόηση της λειτουργίας του μοντέλου και των βασικών χαρακτηριστικών που επηρεάζουν την απόφαση.
- Την ανίχνευση μεροληψίας, αδικιών ή λαθών στο εκπαιδευτικό σύνολο.
- Την εμπιστοσύνη του τελικού χρήστη στη χρήση των αποτελεσμάτων.
- Τη νομική λογοδοσία σε περιβάλλοντα που ρυθμίζονται από κανονισμούς, όπως ο GDPR, ο οποίος κατοχυρώνει το δικαίωμα του πολίτη σε εξήγηση (right to explanation) (Goodman & Flaxman, 2017).

Τα μοντέλα μπορούν να είναι ενδογενώς ερμηνεύσιμα (intrinsically interpretable) ή να αποκτούν ερμηνευσιμότητα μετά την εκπαίδευση (post-hoc interpretability). Τα πρώτα περιλαμβάνουν μοντέλα όπως decision trees, logistic regression ή generalized additive models (GAMs), όπου η διαδικασία λήψης απόφασης είναι από τη φύση της διαφανής. Αντίθετα, μοντέλα όπως τα deep neural networks, τα ensemble methods (π.χ., random forests, gradient boosting) και τα transformer-based models (π.χ., BERT) θεωρούνται black-box και απαιτούν ειδικές τεχνικές για την εξαγωγή εξηγήσεων (Molnar, 2022).

Στη σύγχρονη πρακτική, πολλές μέθοδοι έχουν αναπτυχθεί για την ερμηνεία black-box μοντέλων:

- LIME (Local Interpretable Model-Agnostic Explanations): Δημιουργεί τοπικά απλά μοντέλα γύρω από μια πρόβλεψη για να εξηγήσει την απόφαση (Ribeiro et al., 2016).
- SHAP (SHapley Additive exPlanations): Βασίζεται στη θεωρία παιγνίων για να κατανοήσει τη συνεισφορά κάθε χαρακτηριστικού στην τελική πρόβλεψη (Lundberg & Lee, 2017).

- Attention Mechanisms σε νευρωνικά δίκτυα, ειδικά transformers, αποδίδουν βάρη σε τμήματα του κειμένου ή της εικόνας που επηρεάζουν περισσότερο την απόφαση.
- Feature importance από ensemble models όπως τα Random Forests, που δείχνουν ποια χαρακτηριστικά συνεισφέρουν περισσότερο.

Η διαφάνεια δεν περιορίζεται στην τεχνική επεξήγηση των μοντέλων, αλλά αφορά και τον σχεδιασμό των συστημάτων: ποια δεδομένα χρησιμοποιούνται, ποιος τα συλλέγει, πώς προεπεξεργάζονται, ποιος τα αξιολογεί και ποιος αποφασίζει. Η έλλειψη διαφάνειας ευνοεί την αθέατη επιρροή των αλγορίθμων πάνω στις επιλογές, τις συμπεριφορές και τις απόψεις των χρηστών – ένα φαινόμενο που ονομάζεται αυλός του φίλτρου (filter bubble) ή αλγοριθμικός αποικισμός (Pariser, 2011).

Η Ευρωπαϊκή Ένωση, μέσα από το AI Act και τον Digital Services Act, επιχειρεί να θέσει ελάχιστα πρότυπα διαφάνειας για τα αλγοριθμικά συστήματα που επηρεάζουν τη δημόσια ζωή. Στο ίδιο πνεύμα, η επιστημονική κοινότητα προτείνει την τεκμηρίωση μοντέλων (model cards), τα οποία περιγράφουν τις παραμέτρους, τα δεδομένα, τις παραδοχές και τους περιορισμούς κάθε αλγορίθμου (Mitchell et al., 2019).

Η εφαρμογή αυτών των αρχών στα κοινωνικά δίκτυα είναι εξαιρετικά πολύπλοκη, καθώς τα συστήματα βασίζονται σε πολυτροπικά δεδομένα (κείμενο, εικόνα, βίντεο), συνεχώς εξελισσόμενες συμπεριφορές και ασταθή συμφραζόμενα. Η δημιουργία εξηγούμενων συστημάτων συστάσεων, moderation και ανάλυσης πρέπει να γίνεται σε πραγματικό χρόνο και να είναι ανθεκτική σε κακόβουλες επεμβάσεις.

Ταυτόχρονα, η ερμηνευσιμότητα αποτελεί εργαλείο κοινωνικού ελέγχου και λογοδοσίας. Η δυνατότητα ενός χρήστη να γνωρίζει γιατί αποκλείστηκε, γιατί δεν εμφανίστηκε η ανάρτησή του ή γιατί είδε μία διαφήμιση είναι κρίσιμο ζήτημα δημοκρατίας και ατομικής αυτοδιάθεσης.

Η ερμηνευσιμότητα και η διαφάνεια δεν είναι απλώς τεχνικές απαιτήσεις, αλλά θεμέλια εμπιστοσύνης, δικαιοσύνης και κοινωνικής αποδοχής για κάθε εφαρμογή τεχνητής νοημοσύνης στα κοινωνικά μέσα. Σε ένα οικοσύστημα όπου οι αποφάσεις των αλγορίθμων επηρεάζουν μαζικά την ανθρώπινη επικοινωνία, τις πεποιθήσεις και τη δημοκρατική συμμετοχή, η δυνατότητα να κατανοούμε και να αμφισβητούμε αυτές τις αποφάσεις δεν είναι απλώς επιθυμητή – είναι ηθικά επιβεβλημένη.

## 7.5 Δεδομένα Εκτός Πλαισίου και Σφάλματα Πρόβλεψης

Η πρόκληση των δεδομένων εκτός πλαισίου (Out-of-Distribution, OOD) αναδεικνύεται σε θεμελιώδες ζήτημα της μηχανικής μάθησης, ιδιαίτερα σε εφαρμογές που αφορούν τα κοινωνικά δίκτυα, όπου το περιβάλλον μεταβάλλεται ταχύτατα. Η πλειονότητα των αλγορίθμων επιβλεπόμενης μάθησης βασίζεται στην υπόθεση ότι τα δεδομένα πρόβλεψης ανήκουν στην ίδια στατιστική κατανομή με αυτά της εκπαίδευσης. Η υπόθεση αυτή, γνωστή ως i.i.d. (independent and identically distributed), είναι ωστόσο ανεπαρκής όταν εφαρμοστεί σε πλατφόρμες κοινωνικών μέσων, όπου παρατηρούνται έντονες χρονικές, γλωσσικές, πολιτισμικές και θεματικές αποκλίσεις (Recht et al., 2019).

Η έννοια των OOD δεδομένων αναφέρεται σε παρατηρήσεις που αποκλίνουν ουσιαστικά από τα γνωστά πρότυπα του εκπαιδευτικού συνόλου. Αυτή η απόκλιση δεν οφείλεται απαραίτητα σε σφάλμα ή θόρυβο, αλλά σε μια βαθύτερη στατιστική διαφοροποίηση — είτε στη δομή των χαρακτηριστικών, είτε στο περιεχόμενο των δεδομένων. Χαρακτηριστικά παραδείγματα αποτελούν η χρήση νέας ορολογίας, η εισαγωγή νεολογισμών, οι μεταφορικές εκφράσεις ή η παρουσία συναισθηματικών αποχρώσεων που δεν υπήρχαν στο αρχικό dataset (Kong et al., 2024).

Ένα μοντέλο μηχανικής μάθησης που έχει εκπαιδευτεί, για παράδειγμα, να αναγνωρίζει το συναίσθημα σε tweets του 2018, μπορεί να αποτύχει το 2023 όταν αναλύει αναρτήσεις που περιέχουν πολιτισμικά φορτισμένες λέξεις, νέα hashtags ή ειρωνικά σχήματα λόγου. Αυτή η αποτυχία οφείλεται σε μεταβολή της στατιστικής κατανομής των δεδομένων — φαινόμενο γνωστό ως concept drift (Lu et al., 2019).

Οι συνέπειες της εφαρμογής ενός μοντέλου σε OOD δεδομένα είναι πολλαπλές. Καταρχάς, μειώνεται η ακρίβεια και η αξιοπιστία της πρόβλεψης, ενώ ταυτόχρονα αυξάνεται η πιθανότητα υπερβολικά βέβαιων αλλά λανθασμένων αποτελεσμάτων. Αυτή η υπεραυτοπεποίθηση (overconfident misclassification) θεωρείται ιδιαίτερα επικίνδυνη, ιδίως όταν το σύστημα χρησιμοποιείται για ευαίσθητες εφαρμογές, όπως είναι η ταξινόμηση λογαριασμών ως spam, η ανάλυση συναισθήματος σε δημόσιο διάλογο ή η εντολή αποκλεισμού αναρτήσεων (Nguyen et al., 2015).

Η φύση των social media εντείνει το πρόβλημα. Πρόκειται για περιβάλλοντα με εξαιρετική δυναμικότητα, καθώς κάθε δευτερόλεπτο αναρτώνται νέες μορφές έκφρασης, memes, hashtags, λεκτικά σχήματα ή visual culture που ξεφεύγουν από κάθε προκαθορισμένο λεξιλόγιο ή πρότυπο. Ταυτόχρονα, η πολυγλωσσικότητα, η χρήση slang, η κωδικοποιημένη γλώσσα και η ανάδυση παγκόσμιων ή τοπικών τάσεων (trends) δημιουργούν ένα πολυσύνθετο σημασιολογικό πεδίο που αμφισβητεί την προσαρμοστικότητα των στατικών μοντέλων.

Αυτές οι διαφορές επιφέρουν τρία βασικά είδη σφαλμάτων:

- Σφάλματα προκατάληψης (bias errors): Παρουσιάζονται όταν το μοντέλο έχει μάθει λανθασμένες συσχετίσεις λόγω μεροληπτικού dataset (π.χ., λέξεις που σχετίζονται με συγκεκριμένες κοινωνικές ομάδες ελλοιπώνονται ως "αρνητικές").
- Σφάλματα διασποράς (variance errors): Εκδηλώνονται όταν το μοντέλο είναι υπερευαίσθητο σε μικρές αλλαγές των δεδομένων.
- Σφάλματα OOD: Είναι πιο επικίνδυνα, καθώς το μοντέλο συμπεριφέρεται "με αυτοπεποίθηση" σε δεδομένα τα οποία δεν έχει ξανασυναντήσει, χωρίς να αναγνωρίζει την αβεβαιότητά του.

Η ανίχνευση και η διαχείριση τέτοιων δεδομένων αποτελεί αντικείμενο εντατικής ερευνητικής δραστηριότητας. Ορισμένες από τις πιο διαδεδομένες προσεγγίσεις περιλαμβάνουν τη χρήση Bayesian neural networks και μεθόδων όπως το Monte Carlo dropout, οι οποίες επιτρέπουν την εκτίμηση της αβεβαιότητας κατά την πρόβλεψη (Gal & Ghahramani, 2016). Παράλληλα, τα ensemble models και τα outlier detection techniques (όπως Isolation Forest ή One-Class SVM) συμβάλλουν στη βελτίωση της γενικευσιμότητας.

Σημαντική εξέλιξη αποτελεί και η εφαρμογή τεχνικών continual learning και meta-learning, οι οποίες επιτρέπουν στο μοντέλο να προσαρμόζεται σε νέα δεδομένα μέσω επαναλαμβανόμενης εκπαίδευσης ή προσαρμογής των παραμέτρων του. Ειδικά στο περιβάλλον των κοινωνικών δικτύων, όπου η πληροφορία εξελίσσεται γλωσσικά και θεματικά, τα στατικά μοντέλα καθίστανται γρήγορα ξεπερασμένα (Hendrycks et al., 2021).

Ωστόσο, ακόμα και με προηγμένα τεχνικά μέσα, το πρόβλημα της "τυφλής εμπιστοσύνης" παραμένει. Ένα μοντέλο που δεν μπορεί να αναγνωρίσει ότι βρίσκεται εκτός πλαισίου είναι ένα επικίνδυνο μοντέλο. Η λύση περιλαμβάνει την ενίσχυση της ερμηνευσιμότητας (explainability), την ανθρώπινη επίβλεψη (human-in-the-loop) και τη διαφάνεια στον σχεδιασμό και την εφαρμογή. Οι χρήστες πρέπει να γνωρίζουν πώς και γιατί ελήφθη μια απόφαση από ένα σύστημα, ειδικά όταν αυτή έχει επιπτώσεις στην παρουσία τους στα social media ή στη δημόσια εικόνα τους.

Η ενσωμάτωση τεχνικών ελέγχου ποιότητας (π.χ., auditing algorithms), η διαμόρφωση σαφών πολιτικών αλγοριθμικής ευθύνης και η συμμόρφωση με ρυθμιστικά πλαίσια όπως ο GDPR και η AI Act αποτελούν απαραίτητα βήματα για τη βιώσιμη και ηθικά αποδεκτή χρήση της μηχανικής μάθησης σε περιβάλλοντα με υψηλό ρίσκο όπως τα κοινωνικά δίκτυα.

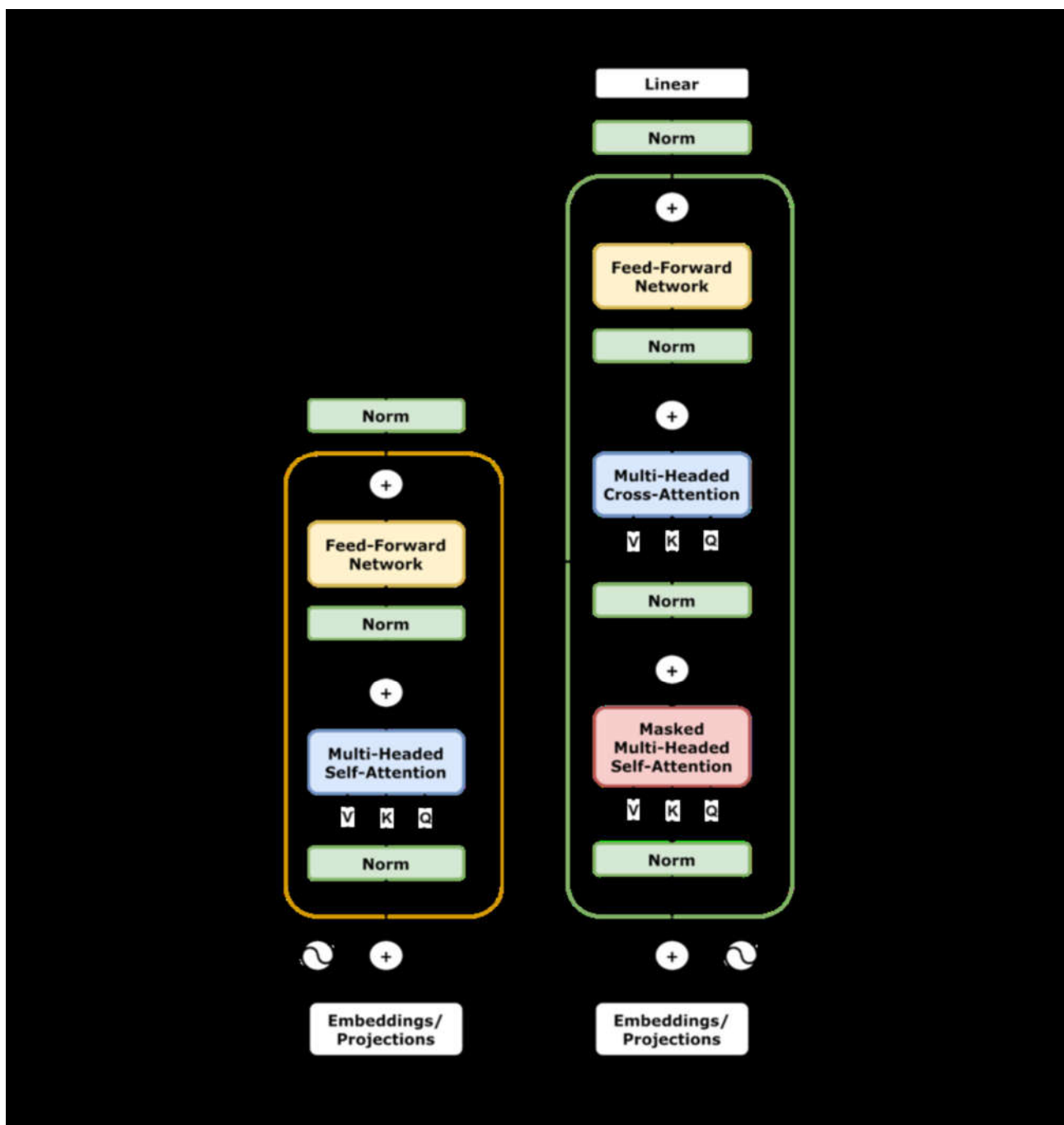
Συνολικά, η ανίχνευση και η διαχείριση δεδομένων εκτός πλαισίου δεν είναι απλώς ένα τεχνικό πρόβλημα, αλλά αφορά θεμελιώδη ζητήματα γενικευσιμότητας, υπευθυνότητας και εμπιστοσύνης. Χωρίς κατάλληλους μηχανισμούς ελέγχου, καμία πρόβλεψη, όσο τεχνικά άρτια και αν φαίνεται, δεν μπορεί να θεωρείται αξιόπιστη.

## Κεφάλαιο 8: Ανοιχτά Ζητήματα και Κατευθύνσεις για Μελλοντική Έρευνα

### 8.1 Αλγόριθμοι νέας γενιάς (π.χ. *Transformers*, *BERT*, *LLMs*)

Η ραγδαία εξέλιξη της μηχανικής μάθησης και ειδικότερα της επεξεργασίας φυσικής γλώσσας (Natural Language Processing - NLP), έχει οδηγήσει σε μια νέα γενιά αλγορίθμων που βασίζονται σε αρχιτεκτονικές βαθιάς μάθησης με επίκεντρο τους *Transformers*. Οι μοντέλα αυτοί, όπως το *BERT* (Bidirectional Encoder Representations from *Transformers*) και τα *LLMs* (Large Language Models), αποτελούν τομή στον τρόπο με τον οποίο τα υπολογιστικά συστήματα αντιλαμβάνονται, κατανοούν και παράγουν φυσική γλώσσα, προσφέροντας πρωτοφανή ακρίβεια και ευελιξία σε πλήθος εφαρμογών στα κοινωνικά δίκτυα (Devlin et al., 2019).

Οι *Transformers* έχουν επιφέρει επανάσταση στη μηχανική μάθηση, ιδιαίτερα στην επεξεργασία φυσικής γλώσσας, αποτελώντας τη βάση για τα σύγχρονα Large Language Models (*LLMs*). Σε αντίθεση με παλαιότερες αρχιτεκτονικές ακολουθιακών δικτύων (*RNN/LSTM*), τα *Transformers* επιτρέπουν την παράλληλη επεξεργασία των δεδομένων ακολουθίας και μπορούν να συλλάβουν μακροπρόθεσμες εξαρτήσεις μέσω του μηχανισμού *attention* (Ouyang et. al., 2022). Μεγάλα προεκπαιδευμένα γλωσσικά μοντέλα όπως το *BERT* (της Google), η σειρά *GPT* (του OpenAI), το *T5* (της Google), το *RoBERTa* (του Facebook), καθώς και πρόσφατα συστήματα διαλόγου όπως το *ChatGPT* και το *Claude*, βασίζονται στην αρχιτεκτονική *Transformer*. Αυτά τα μοντέλα έχουν δισεκατομμύρια παραμέτρους και παρουσιάζουν *emergent* ικανότητες (συμπεριφορές που δεν υπήρχαν σε μικρότερα μοντέλα), επιτυγχάνοντας πρωτοφανή αποτελέσματα σε πληθώρα γλωσσικών εργασιών (Bai et. al., 2022). Στην παρούσα ανάλυση εξετάζεται το θεωρητικό υπόβαθρο των *Transformers* και του μηχανισμού προσοχής (*attention*), παρουσιάζεται η αρχιτεκτονική του *BERT* και των Large Language Models και συζητώνται εφαρμογές σε κοινωνικά δεδομένα (όπως ανάλυση συναισθήματος και ανίχνευση ψευδών ειδήσεων). Επιπλέον, αναλύονται τα πλεονεκτήματα και οι προκλήσεις αυτών των μοντέλων (συμπεριλαμβανομένου του *bias*, της ερμηνευσιμότητας και των *hallucinations*), η εξέλιξή τους σε πολυτροπικά μοντέλα που συνδυάζουν κείμενο με εικόνες/βίντεο, καθώς και η υλοποίησή τους σε πραγματικό χρόνο και σε συσκευές αιχμής (π.χ. μέσω εκδόσεων *TinyBERT* και *DistilBERT*).



Σχηματική απεικόνιση μιας τυπικής αρχιτεκτονικής Transformer, με τον κωδικοποιητή (encoder, αριστερά) και τον αποκωδικοποιητή (decoder, δεξιά). Κάθε στρώμα encoder αποτελείται από πολυκέφαλο μηχανισμό αυτο-προσοχής (multi-head self-attention) και ένα πλήρως συνδεδεμένο νευρωνικό δίκτυο (feed-forward), ενώ κάθε στρώμα decoder περιλαμβάνει επιπλέον έναν μηχανισμό προσοχής διασταύρωσης προς τις εξόδους του encoder (cross-attention).

Η καρδιά του Transformer είναι ο μηχανισμός αυτο-προσοχής (self-attention), ο οποίος επιτρέπει στο μοντέλο να αξιολογεί δυναμικά τη σημαντικότητα κάθε λέξης μιας ακολουθίας σε σχέση με τις υπόλοιπες. Θεωρητικά, η προσοχή υλοποιείται μέσω των διανυσμάτων *Query* (*Q*), *Key* (*K*) και *Value* (*V*) που παράγονται για κάθε συμβολική είσοδο. Ο μηχανισμός attention χαρτογραφεί ένα *query* μαζί με ένα σύνολο από ζεύγη *key-value* σε ένα αποτέλεσμα

– ουσιαστικά υπολογίζει ένα βάρος προσοχής για κάθε value αναλόγως της «συμβατότητας» του με το query και το αντίστοιχο key (Vaswani et. al., 2017) . Συγκεκριμένα, στην ευρέως χρησιμοποιούμενη εκδοχή scaled dot-product attention, το αποτέλεσμα δίνεται από τη σχέση:

$$\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V$$

Όπου:

- $Q \in \mathbb{R}^{n \times d_k}$ : το ερώτημα (query) matrix
- $K \in \mathbb{R}^{m \times d_k}$ : το κλειδί (key) matrix
- $V \in \mathbb{R}^{m \times d_v}$ : το τιμή (value) matrix
- $d_k$ : η διάσταση των διανυσμάτων Q και K
- $\sqrt{d_k}$ : κλίμακα κανονικοποίησης (scaling factor)

Η κλίμακα  $\frac{1}{\sqrt{d_k}}$  χρησιμοποιείται για να περιορίσει τις τιμές του εσωτερικού γινομένου  $QK^T$  ώστε να μη γίνουν υπερβολικά μεγάλες, ειδικά όταν η διάσταση  $d_k$  είναι μεγάλη. Αν δεν χρησιμοποιηθεί, οι μεγάλες τιμές μπορεί να οδηγήσουν σε πολύ απότομες τιμές της softmax και κατά συνέπεια σε προβλήματα στην οπισθοδιάδοση (gradient vanishing/exploding). Η έξοδος της attention είναι ένας σταθμισμένος μέσος όρος των διανυσμάτων V, με τα βάρη να προέρχονται από τις softmax-normalized ομοιότητες μεταξύ Q και K.

Μία ακόμα καινοτομία των Transformers είναι ο πολυκέφαλος μηχανισμός προσοχής (*multi-head attention*). Αντί να υπολογίζεται ένα μοναδικό attention, το μοντέλο μαθαίνει πολλαπλές “κεφαλές” προσοχής ταυτόχρονα: το  $QK^T$ ,  $KK^T$ ,  $VV^T$  προβάλλονται γραμμικά σε χαμηλότερες διαστάσεις και υπολογίζονται παράλληλα πολλαπλές attention, των οποίων τα αποτελέσματα τελικά συνενώνονται (concatenate). Αυτή η τεχνική επιτρέπει στο δίκτυο να εστιάζει σε διαφορετικές υπο-χωρικότητες της αναπαράστασης (π.χ. διαφορετικές σχέσεις λέξεων) ταυτόχρονα, βελτιώνοντας την ικανότητα του μοντέλου να συλλαμβάνει πολύπλοκα μοτίβα.

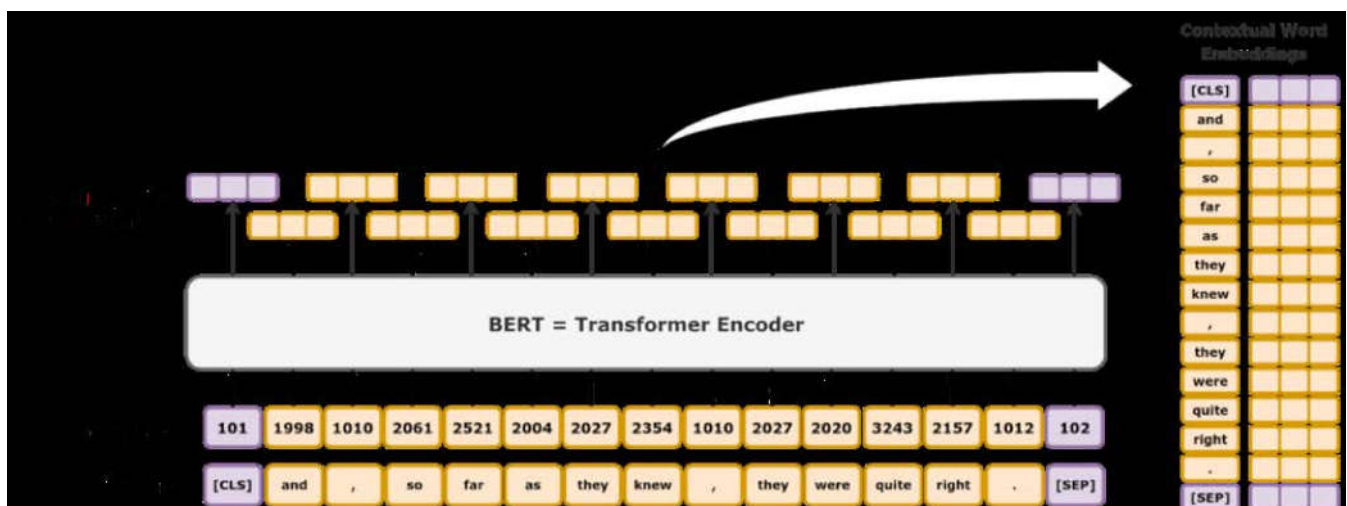
Ένα Transformer αποτελείται από κωδικοποιητές και αποκωδικοποιητές (συνήθως πολλαπλά επάλληλα στρώματα). Ο αρχικός Transformer (Vaswani et al., 2017) για μετάφραση χρησιμοποιούσε 6 στρώματα encoder και 6 decoder. Κάθε encoder layer περιλαμβάνει έναν πολυκέφαλο self-attention και ένα πλήρως συνδεδεμένο feed-forward υποστρώμα, με υπολειμματικές συνδέσεις και κανονικοποίηση (layer norm) σε κάθε υποστρώμα. Ο κάθε decoder layer περιλαμβάνει επιπλέον ένα sub-layer προσοχής που κοιτάει στις εξόδους του encoder (γνωστό ως *encoder-decoder attention* ή *cross-attention*). Για να

μπορέσει το δίκτυο να εκμεταλλευτεί τη θέση των λέξεων (εφόσον δεν υπάρχει πια κάποια εγγενής ακολουθιακή δομή όπως στα RNN), προστίθενται διανύσματα θέσης (*positional encodings*) στις εισόδους, τα οποία κωδικοποιούν τη σειρά των tokens.

Μια κρίσιμη ιδιότητα των Transformers είναι ότι δεν βασίζονται σε αναδρομικές επεξεργασίες· έτσι, μπορούν να επεξεργάζονται όλα τα tokens παράλληλα, επιτυγχάνοντας σημαντική επιτάχυνση στην εκπαίδευση σε σχέση με τα RNNs. Αυτό κατέστησε εφικτή την προεκπαίδευση πολύ μεγάλων μοντέλων σε τεράστια σώματα κειμένου. Ωστόσο, το μειονέκτημα είναι ότι η πολυπλοκότητα μνήμης/υπολογισμού κλιμακώνεται τετραγωνικά ως προς το μήκος της ακολουθίας (λόγω του  $\$QK^T\$$ ), καθιστώντας δύσκολη την διαχείριση πολύ μεγάλων κειμένων. Εντούτοις, διάφορες βελτιώσεις έχουν προταθεί (π.χ. *efficient transformers* με μείωση της πολυπλοκότητας) για να αντιμετωπιστεί αυτό το ζήτημα, ενώ συνεχίζεται η έρευνα για νέες αρχιτεκτονικές μετά την προσοχή (post-attention paradigms) (Tay et. al., 2020).

Με βάση το πλαίσιο των Transformers, έχουν αναπτυχθεί διαφορετικές κατηγορίες μοντέλων ανάλογα με το αν χρησιμοποιούν μόνο το μέρος του encoder, μόνο το decoder ή και τα δύο. Το BERT (Bidirectional Encoder Representations from Transformers) είναι ένα χαρακτηριστικό *encoder-only* μοντέλο – δηλαδή αποτελείται αποκλειστικά από στρώματα κωδικοποιητών του Transformer. Αντίθετα, τα μοντέλα τύπου GPT (Generative Pre-trained Transformer) είναι *decoder-only*, χρησιμοποιώντας μόνο αποκωδικοποιητές και παράγοντας κείμενο αυτοπαλινδρόμησης (ένα token τη φορά). Υπάρχουν και υβριδικά *encoder-decoder* μοντέλα (seq2seq), όπως το T5, που αξιοποιούν και τα δύο μέρη (π.χ. για μετατροπή εισόδου σε έξοδο άλλης μορφής, όπως στη μετάφραση).

BERT – Bidirectional Encoder Representations from Transformers: Το BERT εισήχθη από ερευνητές της Google (Devlin et al., 2018) και σηματοδότησε μια νέα γενιά *προ-εκπαιδευμένων* μοντέλων γλώσσας. Ο στόχος του BERT είναι να μάθει πλούσιες *διδιάστατες* (bidirectional) αναπαραστάσεις – δηλαδή να λαμβάνει υπόψη συμφραζόμενα και από αριστερά και από δεξιά κάθε λέξης ταυτόχρονα σε όλα τα επίπεδα. Το BERT αποτελείται από μια στρώμα  $\$N\$$  επάλληλων Transformer encoder layers (για το BERT\_Base,  $\$N=12\$$  στρώματα με κρυφή διάσταση 768, ενώ το BERT\_Large έχει  $\$N=24\$$  και διάσταση 1024). Με συνολικά ~110 εκατομμύρια παραμέτρους (BERT\_Base) ή 340M (BERT\_Large), προεκπαιδεύτηκε σε ένα τεράστιο σώμα ~3,3 δισεκατομμυρίων λέξεων (BookCorpus και Wikipedia). Στο παρακάτω σχήμα φαίνεται σχηματικά η ροή δεδομένων στο BERT: τα κείμενα εισόδου μετατρέπονται σε tokens (υπο-λέξεις μέσω WordPiece tokenizer) και προστίθενται ειδικά διακριτικά ([CLS], [SEP]) πριν τροφοδοτηθούν στον Transformer encoder, του οποίου οι τελικές ενεργοποιήσεις παράγουν *συμφραζόμενες ενσωματώσεις* λέξεων (contextual embeddings) (Jiao et. al., 2020).



Υψηλού επιπέδου διάγραμμα του BERT. Ένα εισαγωγικό token [CLS] προστίθεται στην αρχή και ένα διακριτικό [SEP] στο τέλος του κειμένου. Το κείμενο "and so far as they knew, they were quite right." διασπάται σε υπο-λέξεις (WordPiece tokens) (π.χ. "so", "##far") που αντιστοιχούν σε μοναδικούς αριθμητικούς δείκτες. Αυτοί οι δείκτες εισέρχονται στον μετασχηματιστή (BERT = Transformer Encoder) και παράγουν, μετά από πολλαπλά επίπεδα, τις τελικές ενσωματωμένες αναπαραστάσεις κάθε λέξης λαμβάνοντας υπόψη το πλήρες συμφραζόμενο.

Το BERT χρησιμοποιεί δύο προσομοιωτικές προ-εκπαιδευτικές εργασίες: (1) Masked Language Modeling (MLM) και (2) Next Sentence Prediction (NSP). Στο MLM, το μοντέλο μαθαίνει να προβλέπει τυχαία καλυμμένες λέξεις μέσα σε μια πρόταση. Συγκεκριμένα, περίπου 15% των tokens μιας πρότασης αντικαθίστανται είτε με το ειδικό σύμβολο [MASK] είτε με κάποια άλλη τυχαία λέξη (ή αφήνονται ως έχουν με μικρή πιθανότητα) και το BERT καλείται να αναγνωρίσει τις αρχικές λέξεις. Αυτή η διαδικασία το αναγκάζει να μάθει διπλής κατεύθυνσης συμφραζόμενα – δηλαδή να κατανοεί μια λέξη λαμβάνοντας υπόψη τόσο τα προηγούμενα όσο και τα επόμενα tokens της. Στο NSP, στοχεύει να μάθει τις σχέσεις μεταξύ προτάσεων: του δίνεται ένα ζεύγος προτάσεων και καλείται να προβλέψει αν η δεύτερη ακολουθεί λογικά μετά την πρώτη στο αρχικό κείμενο ή όχι. Αυτό βοηθά το μοντέλο να συλλάβει τη συνοχή μεταξύ προτάσεων (χρήσιμο για κατανόηση κειμένου σε ανώτερο επίπεδο, π.χ. ερωτήσεις απάντησης σε παραγράφους). Μετά την προεκπαίδευση, το BERT μπορεί να fine-tune-αριστεί εύκολα: προσθέτοντας απλώς ένα μικρό head κατάλληλο για μια συγκεκριμένη εργασία (π.χ. ταξινόμηση με μια πυκνή έξοδο πάνω από το [CLS] token για sentiment analysis) και εκπαιδεύοντας το μοντέλο για λίγες εποχές, επιτυγχάνονται κορυφαίες επιδόσεις σε πολλές εργασίες NLP (Devlin et. al., 2019). Πράγματι, το BERT πέτυχε νέα state-of-the-art αποτελέσματα σε 11 διαφορετικές εργασίες NLP, όπως το σύνολο GLUE, το SQuAD (ερωτήσεις/απαντήσεις) κ.ά., όταν πρωτοπαρουσιάστηκε.

**GPT και μεγάλα γλωσσικά μοντέλα:** Η σειρά μοντέλων GPT (Generative Pre-trained Transformer), αναπτυγμένη από την OpenAI, ακολούθησε μια διαφορετική φιλοσοφία. Τα GPT είναι *αυτοπαλινδρομα* μοντέλα γλώσσας (autoregressive), τα οποία εκπαιδεύονται να προβλέπουν το επόμενο token με δεδομένο το ιστορικό προηγούμενων tokens. Ως εκ τούτου, είναι μοντέλα *decoder-only* (χρησιμοποιούν μόνο αποκωδικοποιητές Transformer) με μονοκατευθυντικό χαρακτήρα – κάθε πρόβλεψη βασίζεται μόνο στο ήδη παραχθέν τμήμα

της ακολουθίας και όχι σε μελλοντικές λέξεις. Το πρώτο GPT (2018) είχε 117 εκατ. παραμέτρους, ενώ το GPT-2 (2019) μεγάλωσε στα 1,5 δισ. παραμέτρους, προκαλώντας αίσθηση για την ικανότητά του να παράγει συνεκτικό κείμενο συνεχίζοντας μια δοθείσα φράση. Το άλμα ήρθε με το GPT-3 (2020), με 175 δισεκατομμύρια παραμέτρους, εκπαιδευμένο σε 300 δισ. tokens από ένα μίγμα δεδομένων (Common Crawl, βιβλία, Wikipedia κ.α.). Το GPT-3 επέδειξε την αξιοσημείωτη ικανότητα few-shot learning: χωρίς περαιτέρω fine-tuning, μπορούσε, δίνοντάς του λίγα παραδείγματα (prompts) μιας νέας εργασίας στη φυσική γλώσσα, να προσαρμοστεί και να παράγει σωστές απαντήσεις. Αυτό ανέδειξε ότι τα μεγάλα LLMs μπορούν να γενικεύσουν ευρύτερα απ' ό,τι αναμενόταν, παρουσιάζοντας *emergent* συμπεριφορές καθώς αυξάνεται το μέγεθός τους (Brown et. al., 2020).

### Model Size Comparison of GPT Models

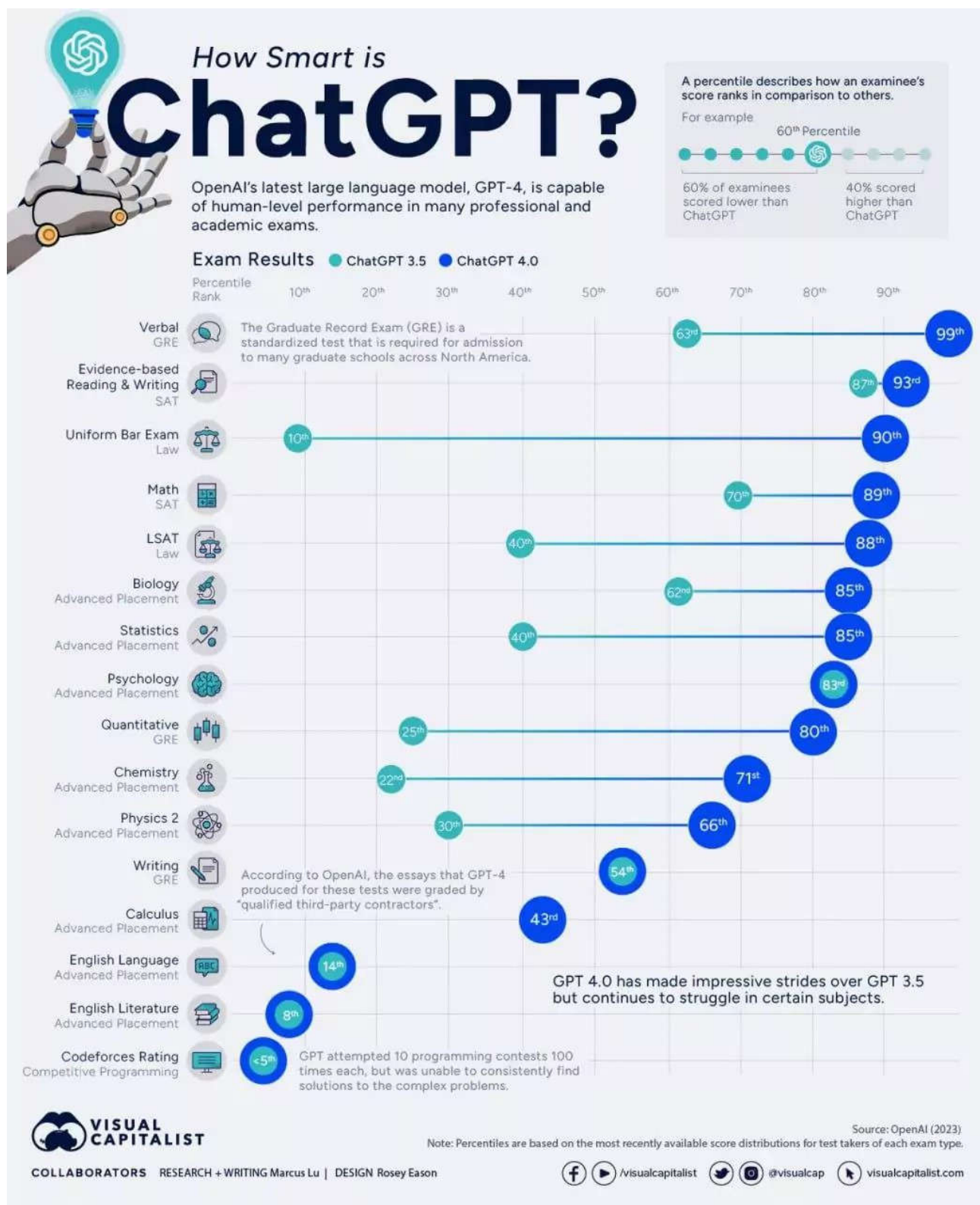
This slide showcases comparison of model sizes prevalent in multiple GPT models. It provides details about parameters, decoder layers, context length, hidden layer size, etc.

	Parameters	Decoder Layers	Context Length	Hidden Layer Size
<b>GPT-1</b>	117million	12	512	768
<b>GPT-2</b>	1.5 billion	48	1024	1600
<b>GPT-3</b>	175 billion	96	2048	12288
<b>GPT-4</b>	1.76 trillion	120	8000*	20k*

This slide is 100% editable. Adapt it to your needs and capture your audience's attention.

Πηγή: <https://www.slideteam.net/introduction-to-gpt-4-powerpoint-ppt-template-bundles-chatgpt-mm.html>

Συνολικά, τα Large Language Models (LLMs) αναφέρονται σε γιγαντιαία μοντέλα τύπου Transformer εκπαιδευμένα σε τεράστιες κειμενικές συλλογές. Εκτός του GPT-3, άλλα σημαντικά LLMs είναι το PaLM (της Google, με 540 δισ. παραμέτρους), το LLaMA (της Meta AI, με μοντέλα 7-70 δισ. παραμέτρων ανοιχτής πρόσβασης) και το GPT-4 (2023, διάδοχος του GPT-3 με ~1,7 τρισ. παραμέτρους σύμφωνα με αναφορές). Τα μοντέλα αυτά ξεχωρίζουν όχι μόνο για το μέγεθός τους αλλά και για τις ικανότητες γενικής γλωσσικής κατανόησης και παραγωγής που επιδεικνύουν. Για παράδειγμα, το GPT-4 και το ChatGPT μπορούν να εκτελέσουν πολύπλοκες οδηγίες πολλαπλών βημάτων, να λύσουν προβλήματα λογικής και να προσαρμοστούν σε νέα καθήκοντα απλά μέσω της διατύπωσης κατάλληλων προτροπών (prompts) από τον χρήστη (Bubeck et. al., 2023).



Πηγή: <https://academichelp.net/blog/academic-writing-tips/gpt-3-5-vs-gpt-4-what-is-the-difference-ai-text-generator-showdown.html>

**RoBERTa και άλλες παραλλαγές:** Μετά το BERT, προτάθηκαν βελτιώσεις στην προεκπαίδευσή του. Το RoBERTa (Liu et al., 2020) ήταν μια «βελτιστοποιημένη» έκδοση του BERT: καταργήθηκε ο στόχος NSP, αυξήθηκε το batch size και το μήκος εκπαίδευσης

και χρησιμοποιήθηκε πολύ μεγαλύτερο corpus (συμπεριλαμβανομένου ~2 τρισεκατομμυρίων tokens από ειδησεογραφικά άρθρα και το Common Crawl). Αυτό οδήγησε σε περαιτέρω βελτίωση της απόδοσης σε πολλές εργασίες. Άλλες παραλλαγές encoder-only μοντέλων περιλαμβάνουν το ALBERT (lighter BERT με διαμοιρασμό παραμέτρων), το XLNet (αυτοπαλινδρομο διδιάστατο μοντέλο), το DeBERTa κ.ά. – επιβεβαιώνοντας ότι η βάση των Transformers μπορεί να βελτιστοποιηθεί για διάφορες ανάγκες.

ChatGPT και μοντέλα διαλόγου (GPT-3.5/4 με ανθρώπινη ανάδραση): Μια σημαντική εξέλιξη στα LLMs ήταν η προσαρμογή τους ώστε να ακολουθούν οδηγίες χρήστη και να παράγουν ωφέλιμες, μη τοξικές αποκρίσεις. Το ChatGPT (κυκλοφόρησε τέλη 2022) είναι ουσιαστικά μια έκδοση του GPT-3.5 προσαρμοσμένη μέσω *fine-tuning* με ανθρώπινη ανάδραση (τεχνική RLHF, Reinforcement Learning from Human Feedback). Συγκεκριμένα, οι ερευνητές της OpenAI (Ouyang et al., 2022) πρώτα fine-tunάρισαν το GPT-3 σε ένα σύνολο συνομιλιών με σωστά παραδείγματα συμπεριφοράς (επιβλεπόμενη μάθηση) και έπειτα βελτίωσαν περαιτέρω το μοντέλο με ενισχυτική μάθηση, χρησιμοποιώντας ανθρώπινους αξιολογητές που βαθμολόγησαν εξόδους του μοντέλου. Το αποτέλεσμα, γνωστό ως InstructGPT, παρουσίασε θεαματικά αποτελέσματα: οι άνθρωποι αξιολογητές προτίμησαν τις απαντήσεις ενός μικρότερου InstructGPT (μόλις 1.3B παραμέτρων) έναντι των απαντήσεων του αρχικού GPT-3 175B, ενώ παράλληλα μειώθηκαν η τοξικότητα και οι ανακρίβειες των απαντήσεων. Το ChatGPT, βασισμένο σε αυτή την προσέγγιση, έχει τη δυνατότητα διαλογικής αλληλεπίδρασης: θυμάται το ιστορικό της συζήτησης, ζητά διευκρινίσεις και παράγει συνεπείς απαντήσεις σε φυσικό ύφος. Παρόμοια, η Anthropic παρουσίασε το μοντέλο Claude, ένα chatbot που δίνει έμφαση στην ασφάλη και ηθική συμπεριφορά. Το Claude αναπτύχθηκε με μεθοδολογίες όπως η *Constitutional AI*, όπου αντί για άμεσο RLHF χρησιμοποιείται ένα σύνολο αρχών (μια “συνταγματική” λίστα κανόνων) για να καθοδηγήσει το μοντέλο να βελτιώνει τις απαντήσεις του αυτόνομα, περιορίζοντας μισαλλόδοξες ή επιβλαβείς εκφράσεις. Τόσο το ChatGPT όσο και το Claude αντιπροσωπεύουν μεγάλα γλωσσικά μοντέλα προσαρμοσμένα για διαλογική χρήση, τα οποία μπορούν να εφαρμοστούν σε πληθώρα πρακτικών εφαρμογών, από εικονικούς βοηθούς μέχρι εργασία συγγραφής κειμένου.

**LLaMA και ανοιχτά LLMs:** Ένα ζήτημα με πολλά προηγμένα LLMs είναι ότι είναι κλειστά (ιδιοκτησιακά). Το 2023, η Meta AI κυκλοφόρησε το LLaMA (Large Language Model Meta AI), μια σειρά μοντέλων με 7, 13, 33 και 65 δισεκατομμύρια παραμέτρους, που προσφέρθηκαν στην ερευνητική κοινότητα. Παρά το μικρότερο μέγεθος σε σχέση με το GPT-3, τα LLaMA είχαν συγκρίσιμη απόδοση μέσω αποτελεσματικής προπόνησης (1-2 τρισεκατομμύρια tokens υψηλής ποιότητας κείμενο). Αυτό επέτρεψε σε τρίτους να δημιουργήσουν παράγωγα μοντέλα (π.χ. Alpaca, Vicuna – fine-tuned εκδοχές του LLaMA για συνομιλίες). Η διαθεσιμότητα ενός ισχυρού LLM ανοικτού κώδικα ενίσχυσε την έρευνα και την καινοτομία, επιτρέποντας ακόμη και την εκτέλεση LLMs σε μικρότερης κλίμακας

υπολογιστές (με βελτιστοποιήσεις). Το LLaMA 2 (2023) ακολούθησε, διαθέσιμο ελεύθερα για εμπορική χρήση, υποστηρίζοντας περαιτέρω το οικοσύστημα ανοικτών LLMs.

Η ισχύς των Transformers και των μεγάλων γλωσσικών μοντέλων έχει αξιοποιηθεί σε ποικίλα προβλήματα ανάλυσης κοινωνικού περιεχομένου. Στην περίπτωση της ανάλυσης συναισθήματος (sentiment analysis), μοντέλα όπως το BERT έχουν εκπαιδευτεί σε σύνολα δεδομένων από μέσα κοινωνικής δικτύωσης, όπως το Twitter και κριτικές προϊόντων, με στόχο την αυτόματη κατηγοριοποίηση του συναισθήματος ενός κειμένου σε θετικό, αρνητικό ή ουδέτερο (Devlin et al., 2019). Χάρη στις contextual embeddings που παράγει, το BERT υπερέχει στην κατανόηση του πολυδιάστατου νοήματος λέξεων εντός πρότασης, καθώς μπορεί να εντοπίσει, για παράδειγμα, τον σαρκασμό βασιζόμενο στο συμφραζόμενο (Sun et al., 2019). Επιπλέον, τα μοντέλα αυτά μπορούν να προσαρμοστούν (fine-tuned) εύκολα σε ειδικά γλωσσικά ιδιώματα κάθε πλατφόρμας, όπως η slang του Twitter, διατηρώντας παράλληλα υψηλή απόδοση.

**Ανίχνευση ψευδών ειδήσεων (fake news detection):** Όσον αφορά την ανίχνευση ψευδών ειδήσεων (fake news detection), η ικανότητα των Transformers να συλλαμβάνουν μακροχρόνιες σχέσεις και λεπτές γλωσσικές ενδείξεις τα καθιστά ιδιαίτερα αποτελεσματικά στην αντιμετώπιση του φαινομένου της παραπληροφόρησης (Kaliyar et al., 2021). Για παράδειγμα, μοντέλα τύπου BERT έχουν χρησιμοποιηθεί με επιτυχία για τον εντοπισμό ψευδών ειδήσεων, όχι μόνο με βάση λέξεις-κλειδιά, αλλά αξιοποιώντας το ύφος γραφής, εντοπίζοντας ασυνέπειες ή συγκρίνοντας το περιεχόμενο με άλλες πηγές (Zhou et al., 2020). Με κατάλληλη εκπαίδευση σε δεδομένα που περιλαμβάνουν επισημασμένες ψευδείς και αληθείς ειδήσεις, ένα μοντέλο όπως το BERT μπορεί να επιτύχει ακρίβεια άνω του 90% στην κατηγοριοποίηση των αναρτήσεων (Yang et al., 2025).

Ενδιαφέρον παρουσιάζουν επίσης μελέτες που συγκρίνουν τις επιδόσεις fine-tuned BERT μοντέλων με εκείνες μεγάλων generative LLMs όπως το GPT-3 σε εργασίες ανίχνευσης παραπληροφόρησης. Συγκεκριμένα, ενώ τα BERT μοντέλα εμφάνισαν υψηλότερη ακρίβεια σε καθαρά ταξινομητικές εργασίες, το GPT-3 παρουσίασε μεγαλύτερη ανθεκτικότητα σε παραλλαγές του κειμένου, όπως αλλαγές στη σύνταξη ή τροποποιήσεις του ύφους, γεγονός που υποδεικνύει υψηλότερη ικανότητα γενίκευσης (Brown et al., 2020; Bubeck et al., 2023). Αυτό αναδεικνύει μια σημαντική διάκριση: τα μεγάλα LLMs ενδέχεται να είναι πιο ευέλικτα σε μη προβλέψιμα inputs, αλλά τα στοχευμένα, προσαρμοσμένα transformers, όπως το BERT, παραμένουν πιο αποτελεσματικά όταν η εργασία είναι σαφώς καθορισμένη και τα δεδομένα συνεπή.

**Θεματική ταξινόμηση και ανάλυση περιεχομένου:** Τα LLMs χρησιμοποιούνται ευρέως για την *κατηγοριοποίηση θεματολογίας* σε σύνολα κοινωνικών δεδομένων, όπως η ομαδοποίηση αναρτήσεων ανά θέμα (π.χ. πολιτική, αθλητισμός, υγεία). Προηγουμένως, τεχνικές όπως Latent Dirichlet Allocation (LDA) παρείχαν *topic modeling*, όμως τα σύγχρονα μοντέλα Transformer μπορούν να κάνουν *topic classification* με επιβλεπόμενη μάθηση, επιτυγχάνοντας καλύτερη ακρίβεια και συλλαμβάνοντας λεπτές θεματικές διαφορές. Για παράδειγμα, ένα fine-tuned RoBERTa μπορεί να διακρίνει αν ένα tweet αφορά «πολιτική διαφωνία» έναντι

«ρητορική μίσους», λαμβάνοντας υπόψη συμφραζόμενα και υπαινιγμούς που θα ξέφευγαν από απλές μεθόδους. Επιπλέον, τα LLMs μπορούν να παράγουν και *περίληψη θεματολογίου* – π.χ. δίνοντας τους ένα σύνολο άρθρων, μπορούν να παραγάγουν σύντομη των κυρίαρχων θεμάτων. Αυτό αξιοποιείται σε πλατφόρμες που παρακολουθούν τάσεις στα social media (trend analysis), όπου ένα μοντέλο σαν το T5 (Text-to-Text Transfer Transformer) μπορεί να μετατρέψει τις αναρτήσεις σε συνοπτικές περιγραφές ανά θέμα (Brown et al., 2020).

Σημαντικό είναι ότι τα μεγάλα προεκπαιδευμένα μοντέλα παρέχουν *γενίκευση και μεταφορά γνώσης*: ακόμη κι αν τα δεδομένα κοινωνικών δικτύων έχουν ιδιωτισμούς, εmojis, ορθογραφικά λάθη κ.λπ., το μοντέλο έχει ήδη «δει» παρόμοια γλωσσικά φαινόμενα στην προεκπαίδευση (π.χ. στο τεράστιο corpus του GPT-3 συμπεριλαμβάνονται και συνομιλίες Reddit). Έτσι, με λίγη προσαρμογή, μπορούν να αποδώσουν εντυπωσιακά σε αυτές τις πραγματικές πηγές δεδομένων.

Τα LLMs όπως το GPT-3/4, το BERT κ.ά. προσφέρουν πλεονεκτήματα που έχουν μετασχηματίσει το πεδίο της TN: (α) Υψηλές επιδόσεις: πετυχαίνουν κορυφαία ακρίβεια σε μεγάλο εύρος εργασιών, από μετάφραση μέχρι ανάλυση συναισθήματος, συχνά υπερβαίνοντας ανθρώπινη επίδοση σε συγκεκριμένα benchmarks. (β) Γενική ικανότητα και προσαρμοστικότητα: ένα ενιαίο προμοτέλο μπορεί, με κατάλληλο prompt ή ελάχιστη περαιτέρω εκπαίδευση, να λύσει εργασίες για τις οποίες δεν είχε εκπαιδευτεί ρητά (π.χ. επίλυση απλών μαθηματικών προβλημάτων, σύνταξη κώδικα). (γ) Αποφυγή ειδικής σχεδίασης χαρακτηριστικών: σε αντίθεση με παλαιότερες προσεγγίσεις που απαιτούσαν *feature engineering*, τα Transformers μαθαίνουν αυτόματα πλούσιες αναπαραστάσεις. (δ) Μεταφορά μάθησης: η γνώση που αποκτούν από την προεκπαίδευση σε τεράστιο κείμενο μπορεί να μεταφερθεί σε εξειδικευμένες εφαρμογές με λίγα δεδομένα, μέσω fine-tuning ή prompt-tuning, λύνοντας έτσι προβλήματα έλλειψης δεδομένων (Brown et al., 2020).

Παρόλα αυτά, υπάρχουν σημαντικές προκλήσεις και περιορισμοί:

- **Bias (Μεροληψίες):** Τα μεγάλα μοντέλα κληρονομούν τις προκαταλήψεις των δεδομένων πάνω στα οποία εκπαιδεύονται. Εάν το κείμενο περιέχει στερεότυπα, ρατσιστικό ή σεξιστικό λόγο, το μοντέλο πιθανώς θα τα αναπαράγει. Έχουν καταγραφεί παραδείγματα όπου μοντέλα απάντησαν με φυλετικά ή έμφυλα μεροληπτικές προτάσεις ή διάκριση έναντι μειονοτήτων. Η αντιμετώπιση αυτών των μεροληπιών είναι δύσκολη: απαιτεί είτε φιλτράρισμα/εξισορρόπηση των δεδομένων εκπαίδευσης, είτε μετα-επεξεργασία των εξόδων, είτε εξειδικευμένη *fine-tuning* ως προς ευαίσθητες διαστάσεις. Οι ερευνητές αναπτύσσουν benchmarks αξιολόγησης της δικαιοσύνης και αμεροληψίας των LLMs και τεχνικές για τον μετριασμό τους. Ωστόσο, η πλήρης εξάλειψη του bias παραμένει ανοικτό πρόβλημα, καθώς τα μοντέλα είναι τόσο μεγάλα που συχνά περιέχουν λεπτές μορφές bias δύσκολο να εντοπιστούν (Bender et al., 2021).

- Explainability (Ερμηνευσιμότητα):** Τα LLMs λειτουργούν ουσιαστικά ως *μαύρα κουτιά*. Με δισεκατομμύρια παραμέτρους, είναι εξαιρετικά δύσκολο να εξηγηθεί γιατί δίνουν μια συγκεκριμένη έξοδο ή ποιες εσωτερικές αναπαραστάσεις οδηγούν σε μια απόφαση. Αυτό είναι πρόβλημα σε εφαρμογές όπου η διαφάνεια είναι κρίσιμη (π.χ. ιατρικές ή νομικές συμβουλές από ένα μοντέλο – πρέπει να αιτιολογείται). Παρότι ο μηχανισμός attention αρχικά θεωρήθηκε ότι θα μπορούσε να προσφέρει κάποιου είδους ερμηνευσιμότητα (βλέποντας σε ποιες λέξεις «κοιτάει» το μοντέλο), στην πράξη αυτό δεν επαρκεί ως εξήγηση. Υπάρχει ενεργή ερευνητική προσπάθεια για μεθόδους ερμηνείας των Transformers: π.χ. ανάλυση των attention heads για να δούμε ποιες σχέσεις κωδικοποιούν, χρήση *input perturbation* (ελέγχοντας πώς αλλάζουν οι έξοδοι αν τροποποιηθούν οι είσοδοι), εφαρμογή τεχνικών όπως SHAP/LIME σε embedding χώρους, ή ακόμα και *μηχανιστική ερμηνευτική* (mechanistic interpretability) όπου επιχειρείται η χαρτογράφηση νευρώνων του μοντέλου σε συγκεκριμένες σημασιολογικές λειτουργίες. Παρότι μερικές ενδιαφέρουσες παρατηρήσεις έχουν γίνει (π.χ. συγκεκριμένοι neurons στο GPT που ανιχνεύουν γράμματα ή λέξεις), δεν έχουμε ακόμη έναν αξιόπιστο τρόπο να εξηγήσουμε πλήρως τις αποφάσεις των LLMs (Belinkov & Glass, 2019). Αυτό δημιουργεί ζητήματα εμπιστοσύνης: π.χ. σε ένα σύστημα αυτόματης λήψης αποφάσεων, πώς μπορούμε να είμαστε βέβαιοι ότι δεν θα κάνει κάποιο σοβαρό λάθος αν δεν μπορούμε να καταλάβουμε το εσωτερικό του σκεπτικού; Η έλλειψη ερμηνευσιμότητας συνδέεται και με τις μεροληψίες: δύσκολα διορθώνεις κάτι που δεν μπορείς εύκολα να εντοπίσεις μέσα στο μοντέλο.
- Hallucinations (Ψευδαισθήσεις):** Ένα από τα πλέον συζητημένα προβλήματα των LLMs, ιδιαίτερα σε γενετικά μοντέλα όπως το GPT-3/4, είναι η τάση τους να παράγουν *πειστικά αλλά αναληθή ή ανύπαρκτα* στοιχεία – αυτό που ονομάζεται “hallucination” ή τεχνητή ψευδαίσθηση. Για παράδειγμα, ένα μοντέλο μπορεί να επινοήσει μια βιβλιογραφική αναφορά που φαίνεται πραγματική αλλά δεν υπάρχει, ή να δώσει μια απάντηση σε μια ερώτηση γνώσης που ακούγεται λογική αλλά είναι λανθασμένη. Οι έρευνες δείχνουν ότι τα chatbots μπορεί να παράγουν τέτοια λανθασμένα στοιχεία σε ποσοστό ~20-30% των περιπτώσεων. Οι hallucinations οφείλονται στο ότι το μοντέλο βάζει την *ευφράδεια* πάνω από την *αλήθεια*: έχει μάθει να παράγει πιθανόλογο κείμενο που ταιριάζει στο prompt, ακόμα κι αν αυτό σημαίνει να «μαντέψει» πληροφορίες όταν δεν είναι βέβαιο. Επειδή δεν έχει πρόσβαση σε μια βάση πραγματικών γνώσεων την ώρα της χρήσης (εκτός αν συνδυαστεί με αναζήτηση), απαντά με το πιο πιθανό συμπληρωματικό κείμενο σύμφωνα με τα πρότυπα γλώσσας – το οποίο όμως μπορεί να είναι λάθος. Αυτή η συμπεριφορά αποτελεί σοβαρό εμπόδιο για υιοθέτηση LLMs σε ευαίσθητα πεδία (π.χ. ιατρική διάγνωση), καθώς το μοντέλο μπορεί να δώσει σφάλματα με υπερβολική αυτοπεποίθηση. Διάφορες προσεγγίσεις εξετάζονται: ενσωμάτωση ελεγκτών αλήθειας (fact-checkers) που επαληθεύουν τις δηλώσεις του μοντέλου, χρήση *prompt engineering* για να ζητείται από το μοντέλο να είναι σίγουρο μόνο όταν γνωρίζει (π.χ. να απαντά «δεν γνωρίζω» αλλιώς), ή συνδυασμός LLM με εργαλεία ανάκτησης πληροφορίας (βλ. π.χ. αναδυόμενο πεδίο retrieval-augmented generation, όπου το μοντέλο ψάχνει σε μια βάση γνώσεων κατά την παραγωγή). Παρά τη μείωση τέτοιων φαινομένων στα

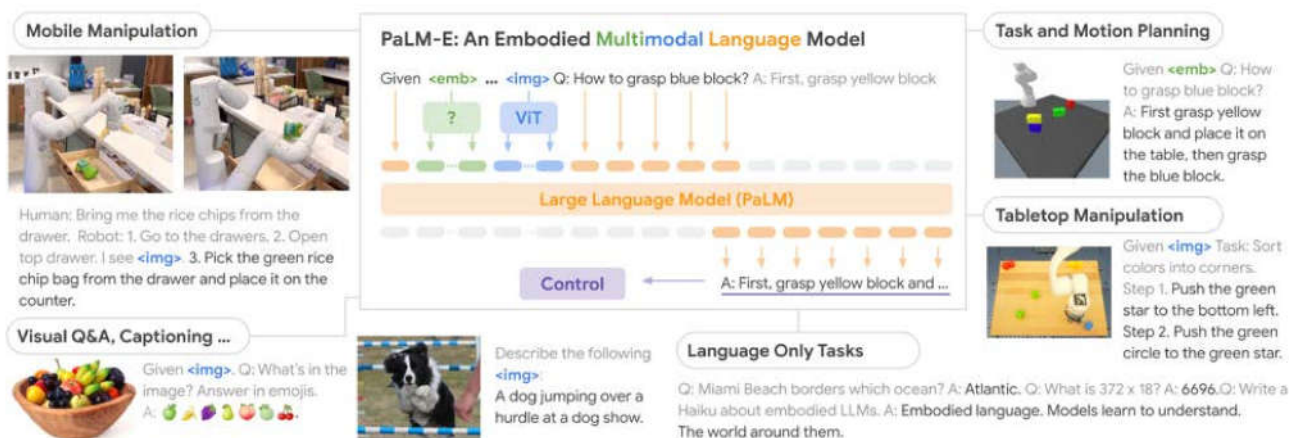
νεότερα μοντέλα (το GPT-4 είναι πιο “έγκυρο” από το GPT-3.5), οι hallucinations δεν έχουν εξαλειφθεί και παραμένουν ενεργό πεδίο έρευνας (Ji et. al., 2023).

- **Άλλες προκλήσεις:** Ένα πρακτικό ζήτημα είναι το υπολογιστικό κόστος. Η εκπαίδευση LLMs απαιτεί τεράστιους πόρους (για παράδειγμα, το GPT-3 εκπαιδεύτηκε σε δεκάδες χιλιάδες GPU ώρες). Αυτό εγείρει ζητήματα ενεργειακής κατανάλωσης και περιβαλλοντικού αποτυπώματος. Επιπλέον, το μέγεθος τους καθιστά δύσκολη την εξαγωγή τους σε συσκευές με περιορισμένη μνήμη. Επίσης, υπάρχει το θέμα της ευθυγράμμισης με την ανθρώπινη πρόθεση (*alignment*): παρότι τις προσαρμογές (όπως το RLHF), τα μοντέλα μπορεί ακόμα να παράγουν ανάρμοστο περιεχόμενο ή να χρησιμοποιηθούν κακόβουλα (π.χ. για μαζική παραγωγή παραπληροφόρησης). Η ασφαλής ανάπτυξή τους απαιτεί συνδυασμό τεχνικών φραγμών και ρυθμιστικών πλαισίων (Bubeck et al., 2023). Τέλος, η επικαιρότητα γνώσεων: τα μοντέλα εκπαιδεύονται σε δεδομένα μέχρι μια χρονική στιγμή και αγνοούν νεότερα γεγονότα. Χρειάζονται μηχανισμοί ενημέρωσης (χωρίς πλήρη επανεκπαίδευση) για να διατηρούνται up-to-date.

Καθώς τα LLMs απέδειξαν την αξία τους στο κείμενο, η επόμενη πρόκληση ήταν η επέκτασή τους σε πολυτροπικά μοντέλα που επεξεργάζονται πολλαπλές μορφές δεδομένων (κειμένου, εικόνων, βίντεο, ήχου). Η ιδέα είναι ένα ενιαίο μοντέλο να μπορεί να «βλέπει» και να «διαβάζει», συνδυάζοντας οπτικές πληροφορίες με γλωσσικές.

Μια πρωτοποριακή δουλειά σε αυτό το πεδίο ήταν το Flamingo (DeepMind, 2022). Το Flamingo είναι ένα Visual Language Model (VLM) το οποίο μπορεί να λάβει ως είσοδο ακολουθίες εναλλασσόμενων εικόνων και κειμένου και να παράγει κείμενο (απαντήσεις, λεζάντες) σχετικά με αυτές. Το σημαντικό είναι ότι το Flamingo επιδεικνύει few-shot learning σε ανοιχτού τύπου πολυτροπικές εργασίες: δηλαδή, με λίγα παραδείγματα εικόνων και περιγραφών ως prompt, μπορεί να επιλύσει νέες εργασίες χωρίς περαιτέρω εξειδικευμένη εκπαίδευση. Για παράδειγμα, μπορεί να δει 2-3 ζεύγη εικόνων με ερωτήσεις και απαντήσεις και στη συνέχεια να απαντήσει σε μια ερώτηση για μια νέα εικόνα. Σε πειράματα, ένα μοντέλο Flamingo (~80B παράμετροι) πέτυχε νέο state-of-the-art σε 16 διαφορετικά πολυμεσικά benchmark με λίγα δείγματα ανά εργασία, ξεπερνώντας μάλιστα μοντέλα που ήταν ειδικά fine-tuned σε κάθε ξεχωριστό task. *Πώς λειτουργεί;* Αρχιτεκτονικά, το Flamingo γεφυρώνει ένα προεκπαιδευμένο vision model (π.χ. ένα μοντέλο ανίχνευσης εικόνων) με ένα προεκπαιδευμένο language model (Chinchilla 70B) μέσω πρόσθετων «ενδιάμεσων» στρωμάτων προσοχής που επιτρέπουν την αλληλεπίδραση μεταξύ των οπτικών και γλωσσικών χαρακτηριστικών. Τα οπτικά δεδομένα (εικόνες ή ακόμα και βίντεο καρέ) μετατρέπονται σε μια ακολουθία εσωτερικών αναπαραστάσεων που εισάγονται στο γλωσσικό μοντέλο σαν να ήταν επιπλέον tokens. Έτσι, το μοντέλο μπορεί να συνεχίσει την ακολουθία παράγοντας κείμενο, ενώ έχει «δει» το τι περιέχεται στην εικόνα. Το αποτέλεσμα είναι εντυπωσιακές δυνατότητες όπως: οπτική ερώτηση-απάντηση (Visual QA), δημιουργία λεζάντας σε εικόνες, περιγραφή βίντεο, όλα αυτά με ελάχιστη επίδειξη παραδειγμάτων, κάτι

που προηγουμένως απαιτούσε ξεχωριστά μοντέλα ή εκτενή fine-tuning (Alayrac et al., 2022).



Παράδειγμα πολυτροπικού μοντέλου (PaLM-E της Google). Το PaLM-E συνδυάζει ένα μεγάλο γλωσσικό μοντέλο (PaLM 540B) με ένα οπτικό σύστημα (ViT-22B) ώστε να χειρίζεται ταυτόχρονα εισόδους εικόνας και κειμένου. Στο διάγραμμα φαίνεται η αρχιτεκτονική αρχή: αισθητηριακές παρατηρήσεις (εικόνες, κατάσταση ρομπότ) κωδικοποιούνται σε διανύσματα που εισχέονται στον χώρο ενσωμάτωσης λέξεων του γλωσσικού μοντέλου, επιτρέποντας στο σύστημα να παράγει αποκρίσεις ή βήματα δράσης με βάση τόσο οπτική όσο και λεκτική αντίληψη. Το PaLM-E επιδεικνύει ικανότητες στην εκτέλεση εντολών ρομποτικής (task planning) και ταυτόχρονα είναι ισχυρό μοντέλο όρασης-γλώσσας (π.χ. πέτυχε SOTA στην απάντηση οπτικών ερωτήσεων OK-VQA χωρίς ειδικό fine-tuning).

Ένα άλλο σημαντικό βήμα είναι το PaLM-E που παρουσιάστηκε από την Google το 2023 και προτείνεται ως το πρώτο embodied multimodal μεγάλο μοντέλο. Το PaLM-E “ενσωματώνει” (Embodied) ένα τεράστιο γλωσσικό μοντέλο (PaLM) με αισθητηριακά δεδομένα από ρομποτικές πλατφόρμες – ουσιαστικά, τροφοδοτεί εικόνες από την κάμερα ενός ρομπότ και μετρήσεις κατάστασης (state estimates) στο γλωσσικό μοντέλο μαζί με κείμενο οδηγιών. Με αυτόν τον τρόπο, το μοντέλο μπορεί να δημιουργεί εξόδους που είναι είτε περιγραφικές (κείμενο που εξηγεί τι βλέπει) είτε εντολές προς το ρομπότ για δράση. Στις δοκιμές, ένα ενιαίο PaLM-E 562B παραμέτρων μπορούσε να εκτελεί εργασίες όπως: σχεδιασμός αλληλουχίας ενεργειών για έναν ρομποτικό βραχίονα (π.χ. «φέρε μου τις πατάτες από το συρτάρι» – το μοντέλο παράγει βήμα-βήμα οδηγίες προς το ρομπότ) και ταυτόχρονα να απαντά σε οπτικές ερωτήσεις ή να περιγράφει σκηνές με state-of-the-art ακρίβεια. Σημαντικό είναι ότι η προσθήκη οπτικών δεδομένων δεν μείωσε τις γλωσσικές ικανότητες: το PaLM-E διατήρησε τις επιδόσεις του PaLM στο καθαρά γλωσσικό κομμάτι, ενώ απέκτησε και νέες ικανότητες όρασης. Αυτό υποδεικνύει ότι το να “εφοδιάσουμε” τα LLMs με επιπλέον τροπικότητες είναι εφικτό και μπορεί να τα μετατρέψει σε γενικούς πράκτορες τεχνητής νοημοσύνης που αντιλαμβάνονται και δρουν στον φυσικό κόσμο (Google, 2023). Παρόμοια, το GPT-4 φημολογείται ότι είναι πολυτροπικό, αποδεχόμενο και εικόνες ως είσοδο (π.χ. αν του δώσεις μια φωτογραφία, μπορεί να την περιγράψει ή να απαντήσει σε ερωτήσεις επί αυτής), συνδυάζοντας έτσι όραση και γλώσσα.

Εκτός από εικόνες, γίνονται προσπάθειες ενσωμάτωσης και άλλων modality: π.χ. βίντεο (όπου ηχρονοσειρά καρέ πρέπει να αντιμετωπιστεί, συχνά με επέκταση της αυτο-προσοχής σε χωροχρονικό επίπεδο) και ήχος/ομιλία (συνδυάζοντας μετατροπείς ομιλίας σε κείμενο με

LLM για κατανόηση προφορικών εντολών). Τα πολυτροπικά μοντέλα ανοίγουν τον δρόμο για εφαρμογές όπως: έξυπνοι βοηθοί που αντιλαμβάνονται το περιβάλλον τους (π.χ. ένα ρομπότ που βλέπει ένα αντικείμενο και ακούει μια εντολή και απαντά ή πράττει ανάλογα), αυτόματη λεζάντα και *content moderation* σε βίντεο/εικόνες για πλατφόρμες κοινωνικής δικτύωσης, ή εργασία για ΑμεΑ (π.χ. περιγραφή περιβάλλοντος σε τυφλούς χρήστες σε πραγματικό χρόνο).

Τα περισσότερα από τα προαναφερθέντα μοντέλα είναι εξαιρετικά μεγάλα και απαιτητικά, γεγονός που δυσχεραίνει την άμεση αξιοποίησή τους σε περιβάλλοντα περιορισμένων πόρων ή εφαρμογές πραγματικού χρόνου όπου η ταχύτητα είναι κρίσιμη (π.χ. σε ένα κινητό τηλέφωνο ή σε μια εφαρμογή που χρειάζεται απόκριση σε κλάσματα δευτερολέπτου). Ωστόσο, η έρευνα έχει αναπτύξει τεχνικές για συμπίεση και επιτάχυνση των Transformers, ώστε να διατηρούν σημαντικό μέρος της απόδοσής τους με πολύ μικρότερο αποτύπωμα.

Μία δημοφιλής προσέγγιση είναι η απόσταξη γνώσης (*knowledge distillation*). Σε αυτήν, ένα μεγάλο εκπαιδευμένο “teacher” μοντέλο χρησιμοποιείται για να παράξει *pseudo-outputs* σε ένα σύνολο δεδομένων και ένα μικρότερο “student” μοντέλο εκπαιδεύεται να μιμηθεί τις προβλέψεις του teacher. Με αυτό τον τρόπο, συμπυκνώνεται η «γνώση» του μεγάλου μοντέλου στο μικρό. Το DistilBERT (Sanh et al., 2019) ήταν ένα από τα πρώτα τέτοια μοντέλα: κατάφερε να μειώσει ~40% τον αριθμό παραμέτρων του BERT (αφαιρώντας περίπου κάθε δεύτερο layer, μεταξύ άλλων βελτιώσεων) και να τρέχει ~60% γρηγορότερα, ενώ η πτώση στην ακρίβεια σε εργασίες όπως το SQuAD ήταν μόνο ~1-3%. Συγκεκριμένα, στο σύνολο ερωτήσεων-απαντήσεων SQuAD1.1 το DistilBERT διατήρησε ~97% των επιδόσεων του BERT με \$1.7\times\$ ταχύτερη ταχύτητα.

Ένα ακόμη πιο συμπίεσμένο μοντέλο είναι το TinyBERT (Jiao et al., 2020). Το TinyBERT εφάρμοσε απόσταξη τόσο στο στάδιο της προ-εκπαίδευσης όσο και στο στάδιο της fine-tuning, με ένα ειδικό πολυ-σταδιακό εκπαιδευτικό σχήμα: πρώτα μαθαίνει να μιμείται τους ενδιάμεσους αναπαραστατικούς χώρους του teacher (BERT) και μετά τις τελικές του εξόδους. Το αποτέλεσμα ήταν ένας “tiny” BERT με 4 ή 6 στρώματα αντί για 12, που τρέχει περίπου \$2\times\$ γρηγορότερα από τον BERT-Base και επιτυγχάνει χάσμα ακρίβειας λιγότερο από 1% σε GLUE/SQuAD. Με άλλα λόγια, 2x speedup με πρακτικά αμελητέα απώλεια επίδοσης. Μια μεταγενέστερη εργασία εισήγαγε το Dynamic TinyBERT, χρησιμοποιώντας και μείωση μήκους ακολουθιών και βέλτιστη ρύθμιση υπερ-παραμέτρων, επιτυγχάνοντας ακόμη μεγαλύτερη επιτάχυνση (~3.3x) με <1% απώλεια.

Η σημασία αυτών των μοντέλων είναι τεράστια για υλοποίηση σε edge devices (όπως smartphones, IoT συσκευές) και σε περιβάλλοντα όπου δεν υπάρχει πανίσχυροι server GPUs. Για παράδειγμα, το DistilBERT έχει ενσωματωθεί σε εφαρμογές κινητών για *offline* ανάλυση κειμένου (χωρίς ανάγκη δικτύου), όπως η άμεση ταξινόμηση εισερχόμενων email στη συσκευή. Ομοίως, εκδοχές TinyBERT θα μπορούσαν να «τρέχουν» σε έναν browser για

να φιλτράρουν τοξικά σχόλια σε πραγματικό χρόνο από τα κοινωνικά μέσα, παρέχοντας ταχύτητα και ιδιωτικότητα (επεξεργασία τοπικά).

Άλλες τεχνικές συμπίεσης περιλαμβάνουν την ποσοτικοποίηση (quantization) των βάσεων – π.χ. μείωση ακρίβειας από 32-bit floating point σε 8-bit integers, που μπορεί να επιφέρει σημαντική μείωση μνήμης/υπολογισμών με μικρή επίπτωση στην απόδοση. Σήμερα υπάρχουν βιβλιοθήκες που επιτρέπουν τη φόρτωση π.χ. του GPT-2 μοντέλου σε 8-bit και εκτέλεση σε συμβατικές CPU με αξιοπρεπή ταχύτητα. Επιπλέον, τεχνικές όπως η περικοπή μοντέλου (pruning) απομακρύνουν νευρώνες ή κεφάλες attention που θεωρούνται λιγότερο σημαντικές. Με συνδυασμό απόσταξης και ποσοτικοποίησης, μοντέλα όπως το MiniLM ή το ALBERT έχουν επίσης προσφέρει επιλογές για ταχύτερα αλλά ισχυρά language models.

Συγκεκριμένα για real-time εφαρμογές, έχει σημασία και η ολιστική καθυστέρηση (latency). Τα μοντέλα decoder-only (GPT) όταν χρησιμοποιούνται για παραγωγή κειμένου, παράγουν λέξη-λέξη, κάτι που μπορεί να καθυστερεί σε μεγάλες απαντήσεις. Υπάρχει έρευνα για βελτιστοποίηση της διαδικασίας δειγματοληψίας (π.χ. *parallel decoding* τεχνικές για συγκεκριμένες περιπτώσεις). Επίσης, υλοποιήσεις σε γλώσσες συστήματος και η χρήση GPU/TPU κάνει εφικτή την χρήση LLMs σε διαδραστικά περιβάλλοντα (όπως το ChatGPT στο διαδίκτυο, που παρότι τρέχει σε servers, δίνει απαντήσεις σχεδόν αμέσως για μεσαίου μεγέθους prompt).

Τέλος, η διαθεσιμότητα ελαφριών μοντέλων επιτρέπει και το edge deployment στην παραγωγή: για παράδειγμα, το BERT-Base αρχικά ήταν πολύ βαρύ για παραγωγικές υπηρεσίες με χιλιάδες αιτήματα ανά δευτερόλεπτο, αλλά με το DistilBERT ή το TinyBERT είναι πλέον εφικτό να έχουμε ένα BERT-like μοντέλο σε κλίμακα (όπως στην κατάταξη αναζητήσεων του Bing, όπου χρησιμοποιήθηκε ένα μοντέλο TinyBERT για real-time ranking). Ακόμα και για φωνητικούς βοηθούς (Siri, Alexa), έχουν ενσωματωθεί μικρότερα Transformers για φυσικότερο χειρισμό εντολών, τρέχοντας πάνω στα chips των συσκευών.

Οι αλγόριθμοι νέας γενιάς στη μηχανική μάθηση και ειδικότερα οι αρχιτεκτονικές Transformers και τα μεγάλα γλωσσικά μοντέλα, έχουν αναδιαμορφώσει το τοπίο της Τεχνητής Νοημοσύνης. Με τον μηχανισμό προσοχής να επιτρέπει την αποτελεσματική εκμάθηση μακροχρόνιων εξαρτήσεων, τα Transformers πέτυχαν πρωτοφανή αποτελέσματα σε προβλήματα γλώσσας, όρασης και πέραν αυτών. Μοντέλα όπως το BERT έθεσαν τις βάσεις για μεταφορά γνώσης και ευρεία υιοθέτηση προεκπαιδευμένων δικτύων, ενώ τα GPT-3/4 και τα συναφή LLMs έδειξαν πως η κλιμάκωση του μεγέθους μπορεί να επιφέρει γενικότερη νοημοσύνη και ικανότητα προσαρμογής. Ταυτόχρονα, αναδείχθηκαν προκλήσεις – από τις εγγενείς μεροληψίες και την έλλειψη διαφάνειας, μέχρι την τάση των μοντέλων να «φαντάζονται» ανακρίβειες. Η αντιμετώπιση αυτών των θεμάτων αποτελεί πεδίο εντατικής έρευνας, με στόχο πιο αξιόπιστα και δίκαια συστήματα. Η επένταση σε πολυτροπικά μοντέλα είναι ήδη πραγματικότητα, ανοίγοντας τον δρόμο για AI που κατανοεί πολυδιάστατα τον κόσμο (βλέπει και ακούει, πέρα από το να διαβάζει). Τέλος, χάρι σε μεθόδους συμπίεσης, οι προηγμένες δυνατότητες των Transformers γίνονται διαθέσιμες ακόμη και σε συνθήκες πραγματικού χρόνου και σε φθηνό υλικό, διευρύνοντας την πρακτική

τους επίδραση. Οι εξελίξεις αυτές σηματοδοτούν ένα βήμα πιο κοντά σε γενικούς, ευέλικτους και αξιόπιστους «πράκτορες τεχνητής νοημοσύνης» που θα αποτελέσουν βασικά δομικά στοιχεία μελλοντικών εφαρμογών.

## 8.2 Συνδυασμός Μηχανικής Μάθησης με Μετα-Ανάλυση και Big Data

Η ανάλυση δεδομένων από τα κοινωνικά δίκτυα αποτελεί ένα πεδίο που παράγει τεράστιο όγκο, ποικιλία και ταχύτητα πληροφορίας – χαρακτηριστικά γνωστά ως οι «5V» του Big Data. Σε αυτό το πλαίσιο, η Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP) και ειδικότερα η Ανάλυση Συναισθήματος αποτελούν κεντρικές εφαρμογές, αξιοποιώντας αλγορίθμους Μηχανικής Μάθησης (Machine Learning) για την εξαγωγή γνώσης από τα μαζικά κείμενα των χρηστών. Τα τελευταία χρόνια παρατηρείται σημαντική αύξηση του ενδιαφέροντος για την ανάλυση συναισθήματος, τόσο από την επιστημονική κοινότητα όσο και από επιχειρήσεις και κυβερνητικούς φορείς. Η εξάπλωση των μέσων κοινωνικής δικτύωσης και του περιεχομένου που δημιουργείται από χρήστες (user-generated content) έχει οδηγήσει στη συλλογή πρωτοφανούς μεγέθους δεδομένων κειμένου· η ανάλυση αυτού του τεράστιου όγκου δεδομένων μπορεί να αποκαλύψει τάσεις της κοινής γνώμης και να στηρίξει τη λήψη αποφάσεων σε πραγματικό χρόνο.

Στο υπόβαθρο αυτό, ο συνδυασμός της μηχανικής μάθησης με μεθόδους Μετα-Ανάλυσης έχει αναδειχθεί ως μια πολλά υποσχόμενη προσέγγιση για τη βελτίωση των αποτελεσμάτων σε περιβάλλοντα Big Data. Ο όρος *μετα-ανάλυση* αναφέρεται στη στατιστική συνδυαστική ανάλυση αποτελεσμάτων από πολλές ανεξάρτητες μελέτες, μέσω μιας συστηματικής ανασκόπησης όλων των διαθέσιμων δεδομένων. Με τον τρόπο αυτό, εξάγονται γενικευμένα συμπεράσματα με αυξημένη αξιοπιστία, καθώς η μετα-ανάλυση αξιοποιεί συλλογικά την πληροφορία από το σύνολο των πηγών μειώνοντας την επίδραση μεμονωμένων αποκλίσεων. Η ενσωμάτωση τέτοιων τεχνικών στον χώρο της μηχανικής μάθησης μπορεί να προσδώσει μια “συμβιωτική” ενίσχυση: η μετα-ανάλυση συνθέτει την υπάρχουσα γνώση, ενώ οι αλγόριθμοι ML ανακαλύπτουν κρυμμένα πρότυπα μέσα στα συγκεντρωτικά δεδομένα. Αυτή η συνέργεια εκτιμάται ότι βελτιώνει την ακρίβεια και την προβλεπτική ικανότητα των μοντέλων, ιδίως όταν το διαθέσιμο δεδομένο είναι πλούσιο αλλά και πολύπλευρο (όπως συμβαίνει στα social media) (Borenstein et al., 2009). Σημαντική προϋπόθεση, βέβαια, είναι η ορθή ευθυγράμμιση των δεδομένων και των μελετών που συνδυάζονται – το αξίωμα του “να μη συγχέονται ανόμοια στοιχεία” στη μετα-ανάλυση υποδεικνύει ότι δεν πρέπει να συνδυάζονται αποτελέσματα με ουσιαδώς διαφορετικά χαρακτηριστικά. Η αρχή αυτή αντικατοπτρίζεται και στη Μηχανική Μάθηση, όπου η ποιότητα και η ομοιογένεια των δεδομένων εκπαίδευσης είναι κρίσιμες για αξιόπιστα αποτελέσματα. Επομένως, η σύζευξη ML και μετα-ανάλυσης απαιτεί μεθοδολογική αυστηρότητα: συστηματική συλλογή και τεκμηρίωση των δεδομένων (π.χ. με διαδικασίες τύπου PRISMA για ανασκοπήσεις) και τυποποίηση στον τρόπο παρουσίασης των αποτελεσμάτων, ώστε ο συνδυασμός να είναι έγκυρος και αναπαραγωγίμος.

**Εφαρμογές σε NLP και Ανάλυση Συναισθήματος:** Στην πράξη, πολλοί από τους σύγχρονους αλγορίθμους ανάλυσης συναισθήματος σε δεδομένα κοινωνικών δικτύων αξιοποιούν τεχνικές μηχανικής μάθησης. Αρχικά χρησιμοποιήθηκαν κλασικοί ταξινομητές (όπως Naïve Bayes, SVM, δέντρα απόφασης), αλλά η ανάγκη για επεξεργασία μεγαλύτερων και πιο πολύπλοκων συνόλων δεδομένων οδήγησε στην ευρεία υιοθέτηση μεθόδων Βαθιάς

Μάθησης (Deep Learning). Πράγματι, επεκτάθηκε η χρήση βαθιών νευρωνικών δικτύων για τη διαχείριση “μεγαλύτερων δεδομένων” και τη βελτίωση της ακρίβειας ταξινόμησης σε σχέση με τα κλασικά μοντέλα ML. Οι τεχνικές βαθιάς μάθησης έχουν το πλεονέκτημα ότι μπορούν να εξαγάγουν αυτόματα χαρακτηριστικά από κείμενα μεγάλης κλίμακας και να αντιμετωπίσουν πολύπλοκα μοτίβα, κάτι που τις κάνει ιδιαίτερα κατάλληλες για Big Data. Ωστόσο, ένα μεμονωμένο μοντέλο βαθιάς μάθησης μπορεί να εμφανίζει περιορισμούς, όπως υψηλή πολυπλοκότητα ως “μαύρο κουτί” και ευαισθησία στις παραμετροποιήσεις. Σε αυτό το σημείο εισέρχεται η μετα-αναλυτική προσέγγιση στο επίπεδο των μοντέλων: ο συνδυασμός πολλαπλών μοντέλων σε ένα ενιαίο σύστημα μπορεί να ενισχύσει την απόδοση και τη γενίκευση.

Μια χαρακτηριστική τεχνική συνδυαστικής μάθησης είναι η δημιουργία συνόλων μοντέλων (ensemble learning), όπου τα αποτελέσματα πολλών ταξινομητών συγχωνεύονται. Απλές μορφές ensemble (όπως η ψηφοφορία πλειοψηφίας ή το bagging/boosting) έχουν εφαρμοστεί στην ανάλυση συναισθήματος, όμως πρόσφατες έρευνες προχωρούν πέρα από αυτά, σε πολυεπίπεδες προσεγγίσεις μετα-μάθησης. Για παράδειγμα, οι Kora και Mohammed (2023) πρότειναν ένα «meta-ensemble» βαθιάς μάθησης για ταξινόμηση συναισθήματος, όπου συνδυάζονται πολλαπλά νευρωνικά δίκτυα σε τρία διαδοχικά επίπεδα μετα-ταξινομητών. Κάθε επίπεδο μαθαίνει από τις προβλέψεις του προηγούμενου, συνθέτοντας έτσι την πληροφορία σε διαρκώς υψηλότερο επίπεδο αφαίρεσης. Τα πειράματα σε μεγάλο πλήθος δεδομένων (περιλαμβανομένου ενός επεκταμένου corpus με 50.000 tweets) έδειξαν ότι η μετα-ensemble προσέγγιση υπερνικά σε ακρίβεια κάθε μεμονωμένο μοντέλο βαθιάς μάθησης, επιτυγχάνοντας σημαντικά καλύτερη απόδοση στην ανάλυση συναισθήματος. Μάλιστα, παρατηρήθηκε περαιτέρω βελτίωση όταν ως είσοδος στους meta-learners χρησιμοποιήθηκαν οι κατανομές πιθανοτήτων των επιμέρους προβλέψεων αντί για απλές κατηγορίες, αξιοποιώντας έτσι περισσότερη πληροφορία αβεβαιότητας από κάθε μοντέλο. Αυτό το παράδειγμα αναδεικνύει πώς η μετα-ανάλυση στη πράξη (μέσω συνδυασμού μοντέλων) ενισχύει τη μηχανική μάθηση: συγκεντρώνοντας γνώση από πολλαπλές πηγές (μοντέλα, γλώσσες ή σύνολα δεδομένων) και αξιοποιώντας την υπολογιστική ισχύ για να εξαχθεί μια τελική απόφαση με μεγαλύτερη αξιοπιστία.

**Υποστήριξη από Τεχνολογίες Big Data:** Η υλοποίηση τέτοιων συνδυαστικών προσεγγίσεων σε πραγματικά προβλήματα καθίσταται εφικτή χάρη στις σύγχρονες τεχνολογίες Big Data. Πλατφόρμες όπως το Apache Hadoop και το Apache Spark επιτρέπουν την αποθήκευση και κατανεμημένη επεξεργασία τεράστιων όγκων δεδομένων κοινωνικών μέσων, προσφέροντας τη δυνατότητα ανάλυσης σε κλίμακα και σε πραγματικό χρόνο. Κατανεμημένα υπολογιστικά πλαίσια σαν το Spark είναι καίρια για εφαρμογές όπως η ανάλυση συναισθήματος streaming δεδομένων (π.χ. ζωντανής ροής tweets), διότι πρόκειται για από τις πιο απαιτητικές περιπτώσεις ανάλυσης όπου απαιτούνται ισχυρά εργαλεία Big Data (Gandomi & Haider, 2015). Πράγματι, έχει διαπιστωθεί ότι η ανάλυση συναισθήματος σε πραγματικό χρόνο “απαιτεί ισχυρές πλατφόρμες ανάλυσης Big Data όπως

το Spark” για να διαχειριστεί τον όγκο και την ταχύτητα των δεδομένων. Για παράδειγμα, με χρήση του οικοσυστήματος Hadoop, μια επιχείρηση μπορεί να εξορύξει δεδομένα από Twitter, Facebook και άλλα κοινωνικά δίκτυα και να αναλύσει το συναίσθημα των χρηστών, παρέχοντας άμεσα αξιοποιήσιμα αποτελέσματα για τη λήψη αποφάσεων. Οι τεχνολογίες Big Data καθιστούν επίσης δυνατή την ενοποίηση πολλαπλών πηγών δεδομένων: μπορούν να συγκεντρωθούν και να αναλυθούν από κοινού δεδομένα από διαφορετικές πλατφόρμες (π.χ. συνδυασμός σχολίων από Twitter, YouTube, forums), επιτρέποντας μια μετα-αναλυτική θεώρηση που εντοπίζει ευρύτερα πρότυπα πέρα από τα όρια μιας μόνο πηγής. Η παράλληλη επεξεργασία και η αρχιτεκτονική δεδομένων αυτών των συστημάτων διασφαλίζουν ότι ακόμη και πολύπλοκες μέθοδοι όπως τα προαναφερθέντα ensembles μπορούν να εκπαιδευτούν και να εφαρμοστούν σε χρονικά πλαίσια αποδεκτά για πρακτική χρήση.

**Πλεονεκτήματα.** Ο συνδυασμός μηχανικής μάθησης, μετα-ανάλυσης και Big Data προσφέρει σημαντικά οφέλη:

- **Βελτιωμένη ακρίβεια και απόδοση μοντέλων:** Με την ενσωμάτωση αποτελεσμάτων από πολλαπλά υπομοντέλα ή μελέτες, μειώνεται η τυχαία σφάλιση ενός μεμονωμένου μοντέλου. Μετα-συνδυαστικές τεχνικές όπως τα ensembles αποδεδειγμένα επιτυγχάνουν υψηλότερη ακρίβεια σε έργα ανάλυσης συναισθήματος, σε σύγκριση με μονολιθικές προσεγγίσεις. Η συλλογική αξιοποίηση των προβλέψεων πολλών μοντέλων λειτουργεί διορθωτικά ως προς τα λάθη τους, ενισχύοντας την γενικευσιμότητα των αποτελεσμάτων. Επιπλέον, η σύνθεση ευρημάτων από πολλές μελέτες μέσω μετα-ανάλυσης αυξάνει τη στατιστική ισχύ, επιτρέποντας την ανίχνευση πιο λεπτών αποτελεσμάτων που μπορεί να διέφευγαν σε μικρότερα δείγματα (Dietterich, 2000).
- **Αξιοποίηση μεγάλου όγκου δεδομένων και αυτόματη εξαγωγή γνώσης:** Σε περιβάλλον Big Data, οι μέθοδοι βαθιάς μάθησης μπορούν να εκπαιδευτούν σε τεράστια σύνολα δεδομένων και να εξάγουν αυτόματα χαρακτηριστικά, αντί να απαιτείται χειροκίνητη επιλογή γνωρισμάτων. Αυτό σημαίνει ότι όσο περισσότερα δεδομένα διατίθενται, τόσο πλουσιότερο γίνεται το μοντέλο μάθησης. Ο συνδυασμός δεδομένων από πολλαπλές πηγές (π.χ. διαφορετικά κοινωνικά δίκτυα ή χρονικές περιόδους) μέσω τεχνικών μετα-ανάλυσης διευρύνει το πληροφοριακό πεδίο του συστήματος, οδηγώντας σε πιο σφαιρική κατανόηση των κοινωνικών τάσεων και συναισθημάτων (Lecun et. al., 2015).
- **Ενίσχυση αξιοπιστίας και εγκυρότητας συμπερασμάτων:** Η μετα-ανάλυση, ως μεθοδολογία, στηρίζεται στη συνεκτική αξιολόγηση όλων των διαθέσιμων δεδομένων και στην τεκμηριωμένη διαδικασία συλλογής τους. Εφαρμοζόμενη στην ανάλυση κοινωνικών μέσων, εξασφαλίζει ότι τα συμπεράσματα δεν βασίζονται σε μεμονωμένες περιπτώσεις ή σε ειδικές συνθήκες ενός dataset, αλλά έχουν ελεγχθεί διαχρονικά και δια-πηγιακά. Αυτό προσδίδει μεγαλύτερη σταθερότητα στα ευρήματα. Παράλληλα, η χρήση Big Data τεχνολογιών σημαίνει ότι ακόμη και καθώς τα δεδομένα ρέουν συνεχώς (streaming data), τα μοντέλα μπορούν να επικαιροποιούνται και τα συμπεράσματα να παραμένουν επίκαιρα, αυξάνοντας την πρακτική αξιοπιστία τους (Borenstein et al., 2009).

- **Υποστήριξη λήψης αποφάσεων σε πραγματικό χρόνο:** Οι συνδυαστικές τεχνικές σε big data περιβάλλον επιτρέπουν την ταχεία επεξεργασία και ανάλυση, κάτι που είναι κρίσιμο όταν πρόκειται για social media όπου οι τάσεις εξελίσσονται γρήγορα. Για παράδειγμα, η δυνατότητα εξαγωγής συγκεντρωτικών δεικτών συναισθήματος από χιλιάδες tweets ανά λεπτό και η αυτόματη σύγκρισή τους (π.χ. με μετα-ανάλυση χρονικών στιγμών) μπορεί να δώσει σε οργανισμούς ένα έγκαιρο εργαλείο αντίδρασης (όπως στην ανίχνευση κρίσεων ή στην αξιολόγηση μιας καμπάνιας) (Shu et al., 2017). Η ενσωμάτωση διαφορετικών πηγών επίσης μειώνει την πιθανότητα να βασίζεται η απόφαση σε *μερική εικόνα* της κοινής γνώμης – αντίθετα, αντλεί από ένα ευρύτερο φάσμα δεδομένων.

**Περιορισμοί.** Παρά τα σημαντικά πλεονεκτήματα, η συνδυαστική αυτή προσέγγιση συνοδεύεται και από προκλήσεις και περιορισμούς που πρέπει να αντιμετωπιστούν:

- **Ετερογένεια δεδομένων και μοντέλων:** Η συγχώνευση πληροφοριών από διαφορετικές πηγές ή η συνδυαστική χρήση πολλών μοντέλων θέτει ζητήματα ετερογένειας. Στη μετα-ανάλυση είναι γνωστό ότι δεν πρέπει να συγχέονται αποτελέσματα από μελέτες που διαφέρουν δραστηρικά (π.χ. ως προς τον πληθυσμό ή τις μεθόδους). Ομοίως, στον χώρο των κοινωνικών μέσων, τα δεδομένα μπορεί να προέρχονται από διαφορετικές πλατφόρμες (με διαφορετική γλώσσα, ύφος ή δημογραφικά χρήστη), γεγονός που δυσχεραίνει τον άμεσο συνδυασμό τους. Απαιτείται λοιπόν προσεκτική προ-επεξεργασία και ομογενοποίηση των δεδομένων (π.χ. μετατροπή όλων σε κοινούς μορφότυπους, γλωσσική κανονικοποίηση) πριν την ανάλυση. Επιπλέον, τα διαφορετικά μοντέλα ML ενδέχεται να αποδίδουν διαφορετικά σε διαφορετικά υποσύνολα δεδομένων· η επιλογή ή στάθμιση των μοντέλων στο ensemble πρέπει να γίνει με τρόπο που να λαμβάνει υπόψη αυτές τις διαφοροποιήσεις (Borenstein et al., 2009).
- **Υψηλό υπολογιστικό κόστος και πολυπλοκότητα:** Η εφαρμογή μετα-ανάλυσης σε κλίμακα Big Data και η εκπαίδευση πολλαπλών μοντέλων παράλληλα μπορεί να απαιτήσει τεράστιους υπολογιστικούς πόρους. Η διαδικασία είναι χρονοβόρα και απαιτητική σε μνήμη/επεξεργαστική ισχύ, ειδικά όταν τα δεδομένα εισέρχονται με μεγάλη ταχύτητα (streaming). Παρά την ύπαρξη του Hadoop/Spark, το κόστος σε χρόνο πραγματικό για εκπαίδευση και διαρκή επικαιροποίηση ενός σύνθετου μοντέλου μπορεί να αποτελεί εμπόδιο. Επιπλέον, η πολυπλοκότητα του συστήματος αυξάνεται: αντί για ένα μοντέλο, έχουμε πολλά διαδοχικά στρώματα μοντέλων, κάτι που δυσχεραίνει τον εντοπισμό σφαλμάτων και τη συντήρηση του κώδικα. Συνεπώς, η αποσφαλμάτωση και βελτιστοποίηση τέτοιων συστημάτων είναι δυσκολότερη, ενώ και ο συντονισμός (tuning) των υπερπαραμέτρων πολλαπλών μοντέλων ταυτόχρονα γίνεται σύνθετο έργο (Zhang & Lu, 2021).
- **Ζήτημα ερμηνευσιμότητας και διαφάνειας:** Τα μοντέλα μηχανικής μάθησης – ιδιαίτερα τα βαθιά νευρωνικά δίκτυα – συχνά χαρακτηρίζονται ως «*μαύρα κουτιά*»,

όπου οι εσωτερικές τους διεργασίες δεν είναι εύκολα ερμηνεύσιμες. Όταν συνδυάζονται πολλά τέτοια μοντέλα, το πρόβλημα της ερμηνευσιμότητας μπορεί να ενταθεί. Η τελική πρόβλεψη προκύπτει από μια αλληλουχία μοντελοποίησης, καθιστώντας δύσκολο να αποδοθούν αιτιώδεις εξηγήσεις (π.χ. ποιο χαρακτηριστικό του κειμένου οδήγησε τελικά σε μια αρνητική ταξινόμηση συναισθήματος). Η έλλειψη διαφάνειας μπορεί να μειώσει την εμπιστοσύνη σε ευαίσθητες εφαρμογές (π.χ. στην πολιτική ανάλυση ή στη διάγνωση τάσεων δημόσιας υγείας μέσω social media) (Binns, 2018. Για τον λόγο αυτό, πρόσφατες έρευνες επιχειρούν να ενσωματώσουν μεθόδους ερμηνεύσιμης Τεχνητής Νοημοσύνης (Explainable AI) παράλληλα με τα μοντέλα, ώστε να παρέχουν ερμηνείες ακόμη και σε επίπεδο ensemble (π.χ. με τεχνικές ανάλυσης σημαντικότητας χαρακτηριστικών για το συνολικό μοντέλο).

- **Ποιότητα και αξιοπιστία δεδομένων:** Στα κοινωνικά δίκτυα, τα δεδομένα συχνά περιέχουν θόρυβο, ψευδείς πληροφορίες ή αυτοματοποιημένους λογαριασμούς (bots). Ένα μεροληπτικό ή χαμηλής ποιότητας σύνολο δεδομένων, αν χρησιμοποιηθεί απευθείας, θα οδηγήσει σε μεροληπτικά ή λανθασμένα αποτελέσματα ακόμη και μετά τον συνδυασμό. Για παράδειγμα, εάν πολλοί χρήστες-«φραντάσματα» (ψεύτικοι λογαριασμοί) εκφράζουν έντονα μια άποψη, το μοντέλο ανάλυσης συναισθήματος μπορεί να υπερεκτιμήσει την πραγματική τάση. Μεγάλης κλίμακας μελέτες έχουν δείξει ότι η ανίχνευση και φιλτράρισμα ψευδών ή κακόβουλων πηγών είναι κρίσιμο βήμα: ενσωματώνοντας μια υπηρεσία εντοπισμού ψεύτικων λογαριασμών σε ένα πλαίσιο ανάλυσης συναισθήματος βελτιώνεται η ακρίβεια των αποτελεσμάτων. Συνεπώς, η διαδικασία θα πρέπει να περιλαμβάνει στάδια καθαρισμού των δεδομένων (data cleaning) και αξιολόγησης της αξιοπιστίας τους. Επιπλέον, υπάρχει ο κίνδυνος της υπερεφαρμογής: αν συνδυαστούν δεδομένα με συστημικές μεροληψίες, η μετα-ανάλυση δεν μπορεί να διορθώσει το πρόβλημα στη ρίζα του – αντιθέτως, μπορεί να επιβεβαιώσει λανθασμένα μοτίβα (Shu et al., 2017). Η προσεκτική επιλογή των πηγών και η εξακρίβωση της ποιότητάς τους είναι απαραίτητη προϋπόθεση.

Συμπερασματικά, ο συνδυασμός μηχανικής μάθησης, μετα-ανάλυσης και τεχνολογιών Big Data ανοίγει νέους δρόμους στην ανάλυση των κοινωνικών μέσων. Προσφέρει τη δυνατότητα πιο αποτελεσματικής και έγκυρης εξαγωγής γνώσης από τον τεράστιο πληροφοριακό όγκο που παρέχουν τα social media, ενισχύοντας τόσο την επιστημονική κατανόηση των διαδικτυακών κοινωνικών φαινομένων όσο και την πρακτική αξιοποίηση της πληροφορίας σε εφαρμογές από το επιχειρηματικό μάρκετινγκ μέχρι τη διαμόρφωση δημόσιας πολιτικής. Παρά τις προκλήσεις, η συνέργεια μεταξύ μετα-ανάλυσης και μηχανικής μάθησης – υποστηριζόμενη από τις υποδομές των Big Data – αποτελεί ένα εξελισσόμενο πεδίο με μεγάλες προοπτικές, ικανό να αναβαθμίσει την ακρίβεια, τη βαθύτερη κατανόηση και την εμβέλεια των αναλύσεων στα κοινωνικά δίκτυα.

### *8.3 Χρήση των Social Media ως εργαλεία Κοινωνικής και Πολιτικής Έρευνας*

Τα μέσα κοινωνικής δικτύωσης (social media) έχουν μεταβάλει ριζικά τον τρόπο έκφρασης, επικοινωνίας και συμμετοχής μας στον δημόσιο λόγο, δημιουργώντας τεράστιους όγκους περιεχομένου που καλύπτουν κάθε πτυχή της κοινωνικής ζωής. Πάνω από το 30% του παγκόσμιου πληθυσμού χρησιμοποιεί πλέον τα social media, με αποτέλεσμα τη συνεχή παραγωγή μεγάλων ποσοτήτων δεδομένων σε πραγματικό χρόνο. Αυτά τα ψηφιακά ίχνη αποτελούν ολοένα και πιο σημαντικές πηγές δεδομένων για κοινωνικούς επιστήμονες, παρέχοντας πρωτοφανή πρόσβαση σε πληροφορίες σχετικά με ανθρώπινες συμπεριφορές, απόψεις και αλληλεπιδράσεις σε κοινωνικά και πολιτικά ζητήματα.

Πράγματι, έχει καταδειχθεί ότι τα δεδομένα από πλατφόρμες όπως το Twitter, το Facebook και το YouTube μπορούν να συμβάλουν σημαντικά στην κατανόηση κοινωνικών τάσεων και μακρο-επιπέδου διεργασιών. Οι ερευνητές αξιοποιούν ποικίλες μεθόδους για την ανάλυση αυτών των δεδομένων – από την εξόρυξη κειμένου και την ανάλυση συναισθήματος, μέχρι την ανάλυση κοινωνικών δικτύων – ώστε να αντλήσουν συμπεράσματα για φαινόμενα όπως οι πολιτικές συζητήσεις, οι ειστρατείες υγείας, ακόμη και οι εξελίξεις στα ΜΜΕ. Τα social media, λοιπόν, λειτουργούν ως εύφορο πεδίο για κοινωνική και πολιτική έρευνα, προσφέροντας πλούσιο υλικό για μελέτη σε τομείς από τη δημόσια υγεία και την οικονομία μέχρι την πολιτική συμπεριφορά και τις κοινωνικές σχέσεις.

Μία από τις κυριότερες εφαρμογές των δεδομένων social media στην πολιτική έρευνα είναι η ανάλυση συναισθήματος του κοινού ως προς πρόσωπα, κόμματα και γεγονότα. Η δυνατότητα συλλογής εκατομμυρίων δημοσίων αναρτήσεων (π.χ. tweets) επιτρέπει τη μέτρηση της διάθεσης και των τάσεων του εκλογικού σώματος σε σχεδόν πραγματικό χρόνο. Για παράδειγμα, η ανάλυση συναισθήματος στο Twitter έχει αξιοποιηθεί για την εκτίμηση της δημόσιας γνώμης και ακόμη και για την πρόβλεψη εκλογικών αποτελεσμάτων, αναδεικνυόμενη σε ένα ταχέως αναπτυσσόμενο πεδίο έρευνας (Tumasjan et al., 2010).

Με την επεξεργασία φυσικής γλώσσας και τεχνικές μηχανικής μάθησης, οι ερευνητές εξάγουν το συναισθηματικό πρόσημο (θετικό, αρνητικό, ουδέτερο) των πολιτικών αναρτήσεων, γεγονός που μπορεί να αποκαλύψει τη στάση των ψηφοφόρων απέναντι σε υποψήφιους ή πολιτικές προτάσεις σε πραγματικό χρόνο. Πολλές μελέτες, μάλιστα, έχουν επιχειρήσει να συσχετίσουν τη συλλογική διάθεση των social media με εκλογικές συμπεριφορές, καταγράφοντας περιπτώσεις όπου η αυξημένη θετικότητα ή αρνητικότητα στα κοινωνικά δίκτυα προηγήθηκε αντίστοιχων τάσεων στην κάλπη.

Πέρα από τις εκλογές, η ανάλυση περιεχομένου των social media χρησιμοποιείται για την ανίχνευση πολιτικών τάσεων και θεμάτων που απασχολούν την κοινή γνώμη. Ο δημόσιος διάλογος σε πλατφόρμες όπως το Facebook και τα blogs μπορεί να δείξει ποιες πολιτικές

πρωτοβουλίες συγκεντρώνουν υποστήριξη ή αντιδράσεις. Η θεματολογία που αναδύεται από τα hashtags και τις κοινοποιήσεις επιτρέπει στους ερευνητές να παρακολουθούν τις μεταβολές στην πολιτική ατζέντα και να εντοπίζουν έγκαιρα αναδυόμενα ζητήματα. Επιπλέον, σε περιόδους κρίσεων (π.χ. φυσικές καταστροφές, τρομοκρατικά χτυπήματα, υγειονομικές κρίσεις), τα social media λειτουργούν ως πεδίο άμεσης αντίδρασης του κοινού. Μελέτες έχουν δείξει ότι σε καταστάσεις έκτακτης ανάγκης οι πολίτες καταφεύγουν στο Twitter, το Facebook ή το YouTube για να εκφράσουν συναισθήματα, ανησυχίες ή και για να διαδώσουν πληροφορίες, παρέχοντας στους ερευνητές πλούσιο υλικό για ανάλυση της δημόσιας αντίδρασης σε πραγματικό χρόνο.

Χαρακτηριστικά, μέσω ανάλυσης συναισθήματος σε αναρτήσεις κατά την πανδημία COVID-19, αναδείχθηκαν οι δυναμικές της εμπιστοσύνης του κοινού: αναρτήσεις από επίσημους κυβερνητικούς φορείς προκάλεσαν θετικό αντίκτυπο στους χρήστες, ενώ ειδήσεις από αξιόπιστους ειδικούς κέρδισαν μεγαλύτερη αποδοχή κατά την κρίση. Τέτοιες εφαρμογές καταδεικνύουν πως η ανάλυση συναισθήματος στα social media μπορεί να λειτουργήσει ως ένας έγκαιρος δείκτης της κοινωνικής διάθεσης, βοηθώντας στην κατανόηση και ενδεχομένως την πρόβλεψη της κοινής γνώμης σε κρίσιμες συγκυρίες.

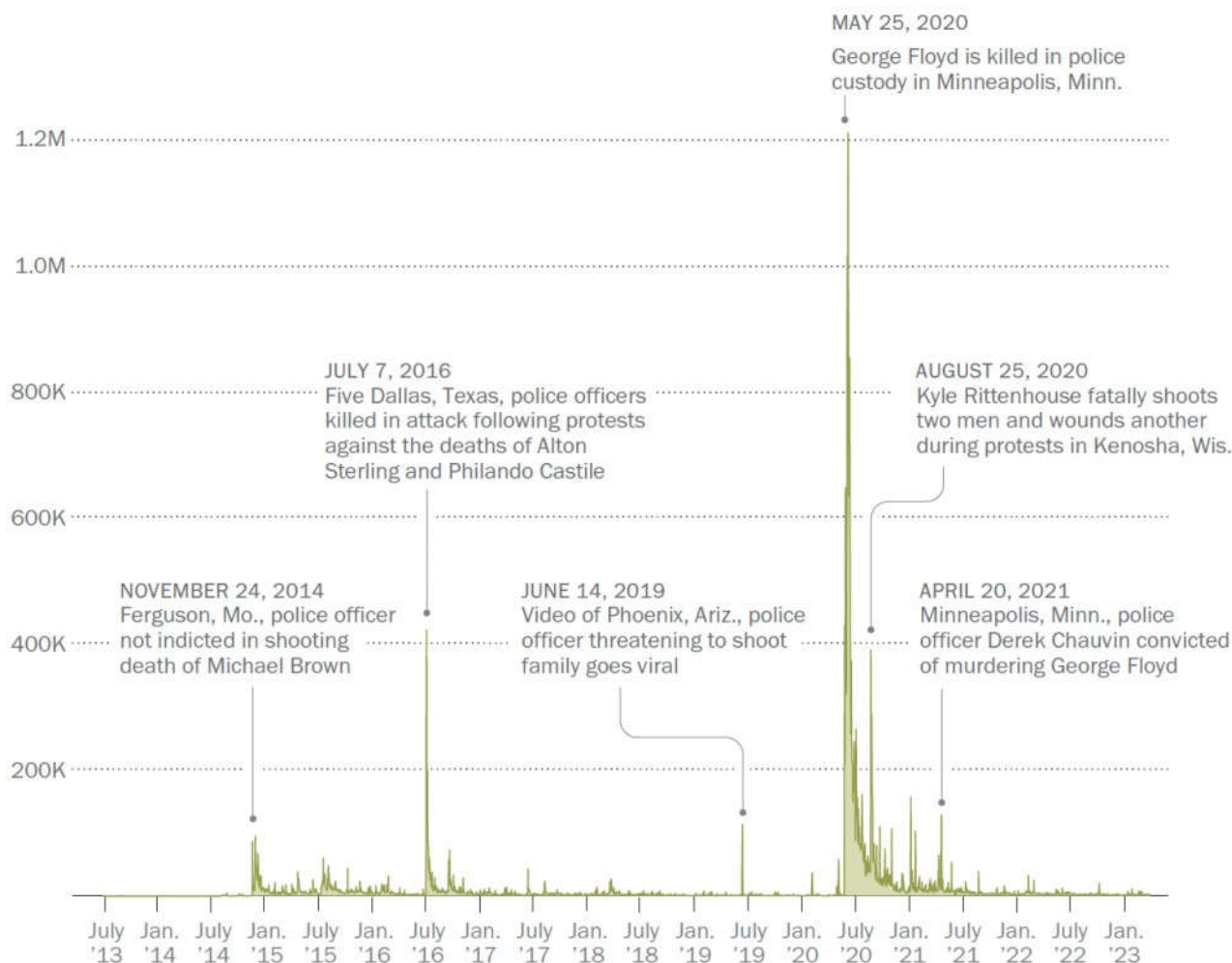
Τα social media έχουν επίσης αναδειχθεί σε κρίσιμο εργαλείο για την παρακολούθηση και μελέτη κοινωνικών κινημάτων και διαμαρτυριών. Μέσω πλατφορμών όπως το Twitter και το Instagram, τα κινήματα μπορούν να οργανωθούν, να διαδώσουν τα μηνύματά τους και να κινητοποιήσουν υποστηρικτές με ταχύτητα και έκταση που δεν ήταν δυνατή σε προγενέστερες εποχές. Για τους ερευνητές, αυτό μεταφράζεται σε μια νέα πηγή δεδομένων για την κατανόηση της δυναμικής των κινημάτων. Ψηφιακά δεδομένα (όπως hashtags, κοινοποιήσεις και σχόλια) επιτρέπουν τον εντοπισμό των θεμάτων που εγείρουν τα κινήματα, τον προσδιορισμό βασικών κόμβων επιρροής (influencers) μέσα σε αυτά, καθώς και την ανάλυση των δικτύων υποστήριξης.

Στην πράξη, το διαδίκτυο και ειδικά τα social media έχουν μετασχηματίσει τον τρόπο οργάνωσης των σύγχρονων διαμαρτυριών, διευκολύνοντας τη μαζική συμμετοχή και τη διάχυση πληροφοριών (π.χ. σε πραγματικό χρόνο ενημερώσεις για τόπους διαδήλωσης). Όπως επισημαίνεται στη βιβλιογραφία, τα ψηφιακά μέσα προσφέρουν σημαντικά νέα εργαλεία σε κοινωνικά κινήματα και ακτιβιστές, ενισχύοντας την ικανότητά τους να δικτυώνονται και να ασκούν κριτική στο κράτος.

Παρακάτω παρουσιάζεται η ημερήσια συχνότητα χρήσης του hashtag #BlackLivesMatter στο Twitter (Ιούλ. 2013–Μάρ. 2023). Παρατηρούνται απότομες αιχμές μετά από σημαντικά γεγονότα (π.χ. δολοφονία George Floyd τον Μάιο 2020), γεγονός που καταδεικνύει πώς τα δεδομένα των social media αποτυπώνουν την ένταση και την εξέλιξη ενός κοινωνικού κινήματος σε βάθος χρόνου.

## Use of the #BlackLivesMatter hashtag surged following the death of George Floyd

Daily number of public Twitter posts using the #BlackLivesMatter hashtag, July 2013–March 2023



Source: Pew Research Center analysis of publicly available tweets from July 2013 through March 2023. Data collected March 1–May 12, 2023, using Twitter API.

"#BlackLivesMatter Turns 10"

PEW RESEARCH CENTER

Πηγή: <https://www.pewresearch.org/internet/2023/06/29/ten-years-of-blacklivesmatter-on-twitter/>

Οι ερευνητές αξιοποιούν τέτοια δεδομένα για να χαρτογραφήσουν την πορεία των κινημάτων. Για παράδειγμα, σε μια δεκαετή ανάλυση του κινήματος Black Lives Matter συγκεντρώθηκαν πάνω από 44 εκατομμύρια tweets με το hashtag #BlackLivesMatter, επιτρέποντας τη μελέτη της εξέλιξης του διαλόγου και της συμμετοχής του κοινού στον χρόνο. Τα δεδομένα αυτά αποκάλυψαν περιόδους αιχμής στη δραστηριότητα (όπως το καλοκαίρι του 2020) όπου σημειώθηκαν εκατοντάδες χιλιάδες αναρτήσεις ημερησίως, σε ανταπόκριση μείζονων γεγονότων αστυνομικής βίας. Μέσω τέτοιων προσεγγίσεων, οι κοινωνικοί επιστήμονες μπορούν να παρακολουθούν σε πραγματικό χρόνο την ανάπτυξη

ενός κινήματος, να μετρούν τη δημόσια υποστήριξη ή αντίθεση και να αναλύουν τις αφηγήσεις που διαμορφώνονται γύρω από αυτό. Παράλληλα, τα ίδια τα social media παρέχουν στις αρχές ένα παράθυρο ψηφιακής επιτήρησης των διαμαρτυριών, καθώς κρατικοί φορείς δύνανται να παρακολουθούν τη δραστηριότητα και ενίοτε να την παρεμποδίζουν (π.χ. μέσω παρεμβάσεων στο διαδίκτυο). Έτσι, η μελέτη των κοινωνικών κινήμάτων στα social media συνδυάζει δύο αντικρουόμενες διαστάσεις: αφενός την ενδυνάμωση της συλλογικής δράσης μέσω δικτύωσης, αφετέρου την νέα μορφή καταστολής ή χειραγώγησης ψηφιακά (digital repression) μέσω παρακολούθησης, ψευδο-πληροφόρησης ή παρεμπόδισης της ροής πληροφοριών.

Η ανοιχτή φύση και η ταχύτητα διάδοσης των πληροφοριών στα social media τα καθιστούν γόνιμο έδαφος όχι μόνο για έγκυρη ενημέρωση αλλά και για παραπληροφόρηση. Τα τελευταία χρόνια παρατηρείται μια έξαρση ψευδών ειδήσεων (fake news), θεωριών συνωμοσίας και γενικότερα ανακριβούς περιεχομένου που εξαπλώνεται ταχύτατα μέσω Facebook, Twitter, YouTube και άλλων δικτύων. Η αντιμετώπιση αυτού του φαινομένου έχει αναδειχθεί σε μείζονα πρόκληση για την ακαδημαϊκή κοινότητα αλλά και για τις ίδιες τις πλατφόρμες. Έρευνες δείχνουν ότι, παρά τα οφέλη της άμεσης επικοινωνίας που προσφέρουν τα κοινωνικά δίκτυα, η ανεξέλεγκτη διάδοση ψευδούς περιεχομένου αποτελεί μία από τις μεγαλύτερες προκλήσεις τους (Lazer et al., 2018). Ο εντοπισμός των fake news αποδεικνύεται σύνθετο και ακόμη άλυτο πρόβλημα, καθώς οι ψευδείς πληροφορίες συχνά σχεδιάζονται να μιμούνται την αλήθεια και να παραπλανούν το κοινό.

Στο πεδίο αυτό, οι επιστήμονες υπολογιστών και οι κοινωνικοί επιστήμονες συνεργάζονται για την ανάπτυξη μεθόδων ανίχνευσης παραπληροφόρησης. Αξιοποιούν τεχνικές μηχανικής μάθησης που αναλύουν τόσο τα γλωσσικά χαρακτηριστικά μιας είδησης (π.χ. τίτλος, λεξιλόγιο, ύφος) όσο και τα μεταδεδομένα διάδοσής της (π.χ. πόσο γρήγορα και από ποιους λογαριασμούς κοινοποιείται). Ωστόσο, η εφαρμογή τεχνητής νοημοσύνης δεν έχει δώσει ακόμα οριστική λύση: τα ίδια τα εργαλεία ΑΙ μπορούν να χρησιμοποιηθούν για τη δημιουργία ψεύτικου περιεχομένου (π.χ. κείμενα ή βίντεο “deepfake”), δυσχεραίνοντας περαιτέρω την ανίχνευση. Μελέτες ανασκόπησης επιβεβαιώνουν ότι το ζήτημα παραμένει ανοιχτό, με πολυάριθμες προσεγγίσεις υπό διερεύνηση – από αλγορίθμους που φιλτράρουν ειδήσεις βάσει αξιοπιστίας πηγών, έως ανάλυση δικτύων bot που ενορχηστρώνουν ειστρατείες παραπληροφόρησης (Zhou & Zafarani, 2020). Τα social media, επομένως, δεν είναι μόνο αντικείμενο έρευνας για το πώς διαδίδεται η παραπληροφόρηση (ρυθμοί διάχυσης, χαρακτηριστικά “viral” περιεχομένου), αλλά και εργαλείο για την αντιμετώπισή της – π.χ. μέσω της έγκαιρης ανίχνευσης ύποπτων προτύπων διάδοσης και της σήμανσης ή απόσυρσης ψευδών ειδήσεων. Συνολικά, η μελέτη της παραπληροφόρησης στα μέσα κοινωνικής δικτύωσης αποτελεί κρίσιμο πεδίο όπου η κοινωνική έρευνα συναντά την υπολογιστική, με στόχο την θωράκιση της ενημέρωσης απέναντι στις σύγχρονες απειλές ψηφιακού περιβάλλοντος.

Ένα ιδιαίτερα καινοτόμο πεδίο που έχει αναδειχθεί με την έλευση των big data από τα social media είναι η ψηφιακή δημογραφία – δηλαδή η εξαγωγή δημογραφικών και κοινωνικών χαρακτηριστικών πληθυσμών βάσει των ψηφιακών τους ιχνών. Κάθε άτομο που

χρησιμοποιεί πλατφόρμες κοινωνικής δικτύωσης αφήνει πίσω του δεδομένα όπως προτιμήσεις (likes), κοινωνικές συνδέσεις, γεωγραφικές ετικέτες (geotags) και δραστηριότητες. Αυτά τα δεδομένα, σε συγκεντρωτικό επίπεδο, μπορούν να αξιοποιηθούν για την χαρτογράφηση τάσεων πληθυσμού και τη συμπλήρωση παραδοσιακών στατιστικών. Για παράδειγμα, αναλύοντας τα δημογραφικά στοιχεία των χρηστών μιας πλατφόρμας ή το περιεχόμενο που κοινοποιούν, οι ερευνητές μπορούν να εκτιμήσουν την κατανομή χαρακτηριστικών όπως ηλικία, φύλο, ή τόπο διαμονής των χρηστών και πώς αυτά συσχετίζονται με κοινωνικές συμπεριφορές. Σε ορισμένες περιπτώσεις, τα δεδομένα των social media λειτουργούν σαν ένας “ψηφιακός πληθυσμιακός κατάλογος”: χαρακτηριστικά είναι τα δεδομένα διαφημίσεων του Facebook, τα οποία θεωρούνται ένα είδος “ψηφιακής απογραφής” με πληροφορίες για πάνω από 2 δισεκατομμύρια χρήστες, επιτρέποντας την άντληση συγκεντρωτικών στοιχείων για δημογραφικά γνωρίσματα και ενδιαφέροντα κοινού (Ribeiro et al., 2020).

Η βιβλιογραφία αναφέρει πολλαπλά παραδείγματα όπου τα ψηφιακά ίχνη χρησιμοποιούνται για δημογραφική έρευνα. Έχουν γίνει μελέτες που συσχετίζουν τη συχνότητα και τρόπο χρήσης των social media με δείκτες γονιμότητας, θνησιμότητας ή μετανάστευσης, παρέχοντας νέα δεδομένα σε κλασικά δημογραφικά ερωτήματα. Για παράδειγμα, αναλύοντας τη γεωγραφική κατανομή των tweets, μπορούν να εξαχθούν συμπεράσματα για τα μοτίβα εσωτερικής μετανάστευσης ή τις μετακινήσεις πληθυσμού μετά από γεγονότα (π.χ. φυσικές καταστροφές). Επίσης, μελετώντας τις γλώσσες και το περιεχόμενο που χρησιμοποιούν οι χρήστες σε διαφορετικές περιοχές, είναι δυνατόν να χαρτογραφηθούν πολιτισμικά και κοινωνικά χαρακτηριστικά κοινοτήτων. Τα πλεονεκτήματα τέτοιων προσεγγίσεων είναι προφανή: αντί να βασίζεται κανείς αποκλειστικά σε χρονοβόρες και ακριβές μεθόδους (όπως απογραφές ή δειγματοληπτικές έρευνες), μπορεί να αξιοποιήσει άμεσα διαθέσιμα ψηφιακά δεδομένα για να αποτυπώσει την κοινωνική πραγματικότητα σε πραγματικό χρόνο. Ωστόσο, απαιτείται προσοχή στην αντιμετώπιση ζητημάτων αντιπροσωπευτικότητας και μεροληψίας των δεδομένων (bias): οι χρήστες των social media δεν αποτελούν απαραίτητα μικρογραφία του γενικού πληθυσμού και ομάδες που δεν χρησιμοποιούν την τεχνολογία αυτή μπορεί να υπο-εκπροσωπούνται. Επομένως, οι μέθοδοι αυτές συχνά συνδυάζονται με σταθμίσεις ή με παραδοσιακά δεδομένα (π.χ. δημοσκοπήσεις) ώστε να διασφαλιστεί η αξιοπιστία των συμπερασμάτων (Cesare et al., 2018).

Η αξιοποίηση δεδομένων από τα social media προσφέρει σημαντικά μεθοδολογικά οφέλη για την κοινωνική και πολιτική έρευνα, αλλά ταυτόχρονα ενέχει και προκλήσεις/περιορισμούς που πρέπει να λαμβάνονται υπόψη σε μια ακαδημαϊκή μελέτη. Ακολουθεί σύνοψη των κυριότερων οφελών και περιορισμών:

- **Μεγάλη κλίμακα και σε πραγματικό χρόνο:** Τα social media παράγουν τεράστιο όγκο δεδομένων συνεχώς. Αυτό δίνει τη δυνατότητα στους ερευνητές να αναλύσουν μεγάλα δείγματα πληθυσμού και να ανιχνεύσουν φαινόμενα τη στιγμή που

συμβαίνουν, κάτι που υπερβαίνει τις δυνατότητες των παραδοσιακών μεθόδων (π.χ. περιοδικές δημοσκοπήσεις). Η ανάλυση big data επιτρέπει την ανάδειξη λεπτών τάσεων ή μειονοτικών απόψεων που θα χάνονταν σε μικρότερα δείγματα (Boyd & Crawford, 2012).

- **Πλούτος περιεχομένου και τύπων δεδομένων:** Σε αντίθεση με τις τυποποιημένες απαντήσεις ερωτηματολογίων, τα δεδομένα social media είναι ανεπιτήδευτα και πολυδιάστατα – περιλαμβάνουν κείμενο, εικόνες, βίντεο, κοινωνικές συνδέσεις, αντιδράσεις (likes, shares) κ.ο.κ. Αυτό επιτρέπει μια πολυεπίπεδη ανάλυση των κοινωνικών φαινομένων. Για παράδειγμα, μπορεί ταυτόχρονα να μελετηθεί το περιεχόμενο ενός μηνύματος (γλωσσική ανάλυση), η διάδοσή του (ανάλυση δικτύου) και η αποδοχή του (αριθμός κοινοποιήσεων/likes) (Bruns & Stieglitz, 2013).
- **Φυσική (μη παρεμβατική) παρατήρηση:** Τα social media καταγράφουν τις αντιδράσεις και τις αλληλεπιδράσεις των ανθρώπων αυθόρμητα, χωρίς την παρουσία ερευνητή που θα μπορούσε να επηρεάσει τις απαντήσεις. Αυτή η μη αντιδραστική συλλογή δεδομένων (non-reactive data collection) επιτρέπει την παρατήρηση κοινωνικών συμπεριφορών σε φυσικό ψηφιακό περιβάλλον. Επιπλέον, οι πλατφόρμες αυτές προσφέρουν ποικιλία προοπτικών και φωνών, καθώς δίνουν βήμα έκφρασης σε άτομα που παραδοσιακά μπορεί να μην ακούγονταν στον δημόσιο λόγο (Murthy, 2008).
- **Νέες ερευνητικές δυνατότητες:** Τα δεδομένα κοινωνικών δικτύων μπορούν να συνδυαστούν με προηγμένους αλγορίθμους για την ανακάλυψη προτύπων που δεν είναι προφανή. Για παράδειγμα, οι τεχνικές μηχανικής μάθησης μπορούν να εντοπίσουν συσχετίσεις ανάμεσα σε διαδικτυακές συμπεριφορές και offline φαινόμενα (π.χ. πρόβλεψη έξαρσης γρίπης από αναρτήσεις σχετικά με συμπτώματα). Επίσης, οι ερευνητές μπορούν να πραγματοποιήσουν πειράματα πεδίου στα social media ή να συλλέξουν δεδομένα από επιμέρους ομάδες πληθυσμού με τρόπο πιο ευέλικτο από ό,τι με παραδοσιακές μεθόδους.

Παρά τα οφέλη αυτά, υπάρχουν σοβαροί περιορισμοί και προκλήσεις στη χρήση δεδομένων social media:

- **Θέματα αντιπροσωπευτικότητας και μεροληψίας:** Ο πληθυσμός των χρηστών social media δεν αντιστοιχεί απαραίτητα στο γενικό πληθυσμό. Ορισμένες κοινωνικές ομάδες (π.χ. ηλικιωμένοι, άτομα χωρίς πρόσβαση στο διαδίκτυο) υποεκπροσωπούνται, ενώ άλλες (π.χ. νεότερες ηλικίες) είναι υπερεκπροσωπημένες. Επιπλέον, οι απόψεις που εκφράζονται online μπορεί να είναι μεροληπτικές – συχνά ακούγονται πιο έντονα οι ακραίες φωνές ή οργανωμένες εκστρατείες (π.χ. από bots ή ομάδες πίεσης), με αποτέλεσμα τα δεδομένα να μην αντικατοπτρίζουν πιστά την offline πραγματικότητα. Όπως σημειώνεται, η διαδικτυακή συμπεριφορά των χρηστών δεν αντιστοιχεί πάντα στις πραγματικές τους πεποιθήσεις ή δράσεις εκτός διαδικτύου και αυτό πρέπει να λαμβάνεται υπόψη στην ερμηνεία (Tufekci, 2014).
- **Ποιότητα και θόρυβος δεδομένων:** Τα δεδομένα που συλλέγονται από τα social media είναι συχνά αδόμητα, ανομοιογενή και θορυβώδη. Περιλαμβάνουν συντομογραφίες, emojis, λανθασμένη ορθογραφία, πολλαπλές γλώσσες, καθώς και

sram ή αυτοματοποιημένο περιεχόμενο. Η επεξεργασία και καθαρισμός αυτών των δεδομένων απαιτεί σημαντική προσπάθεια και μπορεί να εισαγάγει σφάλματα. Επιπλέον, η πληρότητα των δεδομένων δεν είναι εγγυημένη – οι χρήστες μπορεί να διαγράφουν αναρτήσεις ή να ιδιωτικοποιούν τους λογαριασμούς τους, ενώ διαρκώς δημιουργούνται νέα δεδομένα που είναι αδύνατο να συλλεχθούν στο σύνολό τους. (Pang & Lee, 2008)

- **Πρόσβαση σε δεδομένα και νομικοί περιορισμοί:** Αν και ορισμένα δεδομένα είναι δημόσια διαθέσιμα, τα τελευταία χρόνια οι μεγάλες πλατφόρμες έχουν περιορίσει σημαντικά την πρόσβαση των ερευνητών σε δεδομένα, ιδιαίτερα μετά από σκάνδαλα όπως αυτό της Cambridge Analytica το 2018. Οι αλλαγές στις πολιτικές των API (Application Programming Interface) του Twitter, του Facebook/Instagram και άλλων έχουν επιβάλει όρια ή κόστη στην εξαγωγή δεδομένων. Για παράδειγμα, το Twitter πλέον χρεώνει για πρόσβαση σε μεγάλο όγκο ιστορικών tweets και το Facebook έχει κλείσει τα περισσότερα “παράθυρα” για συλλογή δεδομένων χωρίς εμπορική συμφωνία. Αυτό δυσχεραίνει την ανεξάρτητη ακαδημαϊκή έρευνα και μπορεί να οδηγήσει σε μεροληψία επιλογής δεδομένων (οι ερευνητές με πρόσβαση σε δεδομένα να είναι μόνο όσοι συνεργάζονται με τις εταιρείες ή έχουν μεγάλα κονδύλια). Συνολικά, το τοπίο έχει χαρακτηριστεί ως “μετα-API εποχή” όπου οι κοινωνικοί επιστήμονες βρίσκονται σε όλο και πιο αδύναμη θέση όσον αφορά την πρόσβαση σε δεδομένα social media (Bruns, 2019).
- **Ηθικά και ιδιωτικό απόρρητο:** Τα δεδομένα των social media, παρότι δημόσια σε κάποιο βαθμό, αφορούν προσωπικές πληροφορίες και απόψεις ατόμων. Η άντληση και ανάλυσή τους εγείρει ηθικά ζητήματα, όπως η ανάγκη σεβασμού της ιδιωτικότητας και της συναίνεσης των χρηστών. Οι χρήστες θεωρούν τα δεδομένα τους σε κάποιο βαθμό ευαίσθητα και αναμένουν να προστατεύονται συχνά δηλώνουν ότι η έρευνα θα πρέπει να γίνεται με σαφή οφέλη και με τη συγκατάθεσή τους όπου είναι δυνατόν. Έτσι, οι ερευνητές οφείλουν να συμμορφώνονται με κανόνες δεοντολογίας: ανωνυμοποίηση δεδομένων, αποφυγή δημοσίευσης προσωπικά αναγνωρίσιμων πληροφοριών και διαφάνεια ως προς τον σκοπό της έρευνας. Επιπλέον, υπάρχουν νομικά πλαίσια (π.χ. ο Γενικός Κανονισμός Προστασίας Δεδομένων - GDPR στην ΕΕ) που περιορίζουν τι επιτρέπεται να συλλεχθεί και πώς να αποθηκευτεί/χρησιμοποιηθεί. Η ισορροπία ανάμεσα στην επιστημονική πρόοδο και την προστασία των δικαιωμάτων των υποκειμένων είναι κεντρικής σημασίας στη σύγχρονη έρευνα των social media (Townsend & Wallace, 2016).

Συμπερασματικά, τα μέσα κοινωνικής δικτύωσης αποτελούν πανίσχυρα εργαλεία για κοινωνική και πολιτική έρευνα, συνδυάζοντας την ευκαιρία για νέες ανακαλύψεις με την πρόκληση της υπεύθυνης και μεθοδικά ορθής χρήσης τους. Η ενότητα αυτή ανέδειξε τις πολλαπλές δυνατότητες που προσφέρουν – από την ανάλυση της κοινής γνώμης και τη μελέτη κινημάτων, μέχρι την παρακολούθηση της παραπληροφόρησης και τη χαρτογράφηση πληθυσμών. Παράλληλα, υπογράμμισε την ανάγκη προσεκτικής

αντιμετώπισης των μεθοδολογικών περιορισμών. Με τη συνεπή ένταξη της πρόσφατης βιβλιογραφίας και την εξέταση των ορίων και των δυνατοτήτων, οι ερευνητές μπορούν να αξιοποιήσουν τα social media με τρόπο που προάγει την επιστημονική γνώση, συμβάλλοντας στην κατανόηση των σύγχρονων κοινωνικοπολιτικών φαινομένων με έναν τρόπο που έως πριν λίγα χρόνια θα φάνταζε αδύνατος.

## 8.4 Επικαιροποίηση Μοντέλων σε Δυναμικά Περιβάλλοντα

Οι σύγχρονες εφαρμογές μηχανικής μάθησης καλούνται συχνά να λειτουργούν σε δυναμικά περιβάλλοντα όπου τα δεδομένα και τα μοτίβα μεταβάλλονται διαρκώς – χαρακτηριστικό παράδειγμα αποτελούν τα μέσα κοινωνικής δικτύωσης (social media). Σε τέτοια πλαίσια, ένα μοντέλο που εκπαιδεύτηκε κάποτε με στατικά δεδομένα ενδέχεται να υποστεί υποβάθμιση απόδοσης με την πάροδο του χρόνου, λόγω αλλαγών στην κατανομή των δεδομένων (φαινόμενο *concept drift*). Το φαινόμενο του *concept drift*, όπου οι όροι και τα σήματα που σχετίζονται με ένα θέμα αλλάζουν με τον χρόνο, μπορεί να καταστήσει αναποτελεσματικά τα στατικά μοντέλα μηχανικής μάθησης (Gama et al., 2014). Για την αντιμετώπιση αυτών των προκλήσεων, έχουν αναπτυχθεί τεχνικές συνεχούς ενημέρωσης και προσαρμογής μοντέλων, όπως η *online learning*, η διαχείριση του *concept drift* και η *σταδιακή εκπαίδευση (incremental training)*, που επιτρέπουν στα μοντέλα να επικαιροποιούνται διαρκώς με νέα δεδομένα. Στη συνέχεια αναλύονται αυτές οι τεχνικές, οι προκλήσεις και τα πλεονεκτήματά τους, με έμφαση στις εφαρμογές στα social media, όπως η ανάλυση συναισθήματος, η ανίχνευση παραπληροφόρησης και η παρακολούθηση πολιτικών ή κοινωνικών δυναμικών.

Η *online learning* είναι μια προσέγγιση κατά την οποία το μοντέλο εκπαιδεύεται και ενημερώνεται διαρκώς, επεξεργαζόμενο τα δεδομένα ένα προς ένα ή σε μικρές διαδοχικές παρτίδες αντί για ολόκληρα σύνολα δεδομένων εκ των προτέρων. Σε αντίθεση με την παραδοσιακή *batch* εκπαίδευση (όπου το μοντέλο εκπαιδεύεται εφάπαξ σε ένα στατικό σύνολο δεδομένων και στη συνέχεια χρησιμοποιείται χωρίς ενημέρωση), η *online* εκπαίδευση επιτρέπει στο μοντέλο να προσαρμόζει παραμέτρους του σε πραγματικό χρόνο καθώς φτάνουν νέα δεδομένα (Hoi et al., 2018). Για παράδειγμα, ένα μοντέλο ανάλυσης συναισθήματος μπορεί να ενημερώνεται συνεχώς με κάθε νέο *tweet*, διατηρώντας έτσι την επαφή του με την τρέχουσα γλώσσα και τα θέματα συζήτησης.

Πλεονεκτήματα: Η *online* μάθηση καθιστά το σύστημα πιο ευέλικτο και ανθεκτικό σε αλλαγές. Το μοντέλο αποφεύγει το κόστος πλήρους επαναπροπόνησης από την αρχή κάθε φορά που εμφανίζονται νέα δεδομένα, κάτι που βελτιώνει την κλιμάκωση σε μεγάλα συνεχή ρεύματα δεδομένων. Επιπλέον, το μοντέλο μπορεί να αντιδρά άμεσα σε αλλαγές τάσεων ή θεμάτων – κρίσιμο σε περιβάλλοντα όπως τα social media όπου οι τάσεις εμφανίζονται και εξαφανίζονται γρήγορα (Ariga, 2025).

Προκλήσεις: Η συνεχής ενημέρωση ενός μοντέλου ενέχει τον κίνδυνο της *υπερβολικής προσαρμογής* στα πιο πρόσφατα δεδομένα και της λήθης (*catastrophic forgetting*) παλαιότερων γνώσεων. Αν το νέο εισερχόμενο δεδομένο είναι θορυβώδες ή μη αντιπροσωπευτικό, η *online* προπόνηση μπορεί να υποβαθμίσει προσωρινά την απόδοση. Απαιτείται μια προσεκτική ισορροπία μεταξύ πλαστικότητας και σταθερότητας: το μοντέλο πρέπει να προσαρμόζεται αρκετά ώστε να μαθαίνει νέες τάσεις, αλλά όχι τόσο γρήγορα ώστε να ξεχνά μοτίβα που εξακολουθούν να ισχύουν. Τεχνικές όπως η ρύθμιση χαμηλού ρυθμού

εικμάθησης, η χρήση παραθύρων χρονικού διαστήματος (sliding window) ή η βάρη στην ιστορική μνήμη μπορούν να βοηθήσουν, αλλά εισάγουν επιπλέον πολυπλοκότητα στον σχεδιασμό του αλγορίθμου.

Το concept drift αναφέρεται συγκεκριμένα στην αλλαγή της σχέσης μεταξύ των εισόδων ενός μοντέλου και της επιθυμητής εξόδου του με την πάροδο του χρόνου. Σε δυναμικά δεδομένα, όπως τα κοινωνικά δίκτυα, το concept drift εμφανίζεται συχνά με διάφορες μορφές: *σταδιακή* ή *αργή μετατόπιση* (π.χ. αργές αλλαγές στη γλώσσα ή στις προτιμήσεις των χρηστών), *ξαφνική μετατόπιση* (π.χ. απότομη αλλαγή θεματολογίας μετά από ένα έκτακτο γεγονός), *επαναλαμβανόμενη/εποχική* (όπου μοτίβα εμφανίζονται περιοδικά, όπως οι τάσεις κατά την προεικλογική περίοδο) κ.ά. Η γρήγορη δυναμική των social media μπορεί μεν να συλλάβει υποκείμενες τάσεις, αλλά δημιουργεί και το πρόβλημα ότι αλγόριθμοι εκπαιδευμένοι σε παλαιότερα δεδομένα υστερούν σε σύγχρονα δεδομένα λόγω drift (Müller & Salathé, 2019). Για παράδειγμα, μελέτες έδειξαν ότι τα μοντέλα που εκπαιδεύτηκαν σε δεδομένα πριν την πανδημία του COVID-19 δεν μπόρεσαν να ανιχνεύσουν σωστά τη μεταβολή του αισθήματος σχετικά με τα εμβόλια το 2020, εκτός αν προσαρμόστηκαν καταλλήλως.

**Τεχνικές διαχείρισης:** Η αντιμετώπιση του concept drift μπορεί να γίνει είτε *προληπτικά* είτε *αντιδραστικά*. Μια κοινή προσέγγιση είναι η συνεχής παρακολούθηση της απόδοσης του μοντέλου και η επανεκπαίδευση ή προσαρμογή του όταν εντοπιστεί σημαντική πτώση ακρίβειας. Υπάρχουν μέθοδοι *ανίχνευσης drift* που βασίζονται σε στατιστικούς ελέγχους (π.χ. έλεγχοι Kolmogorov-Smirnov για αλλαγές κατανομών) ή στην παρακολούθηση του σφάλματος του μοντέλου στον χρόνο. Άλλη τεχνική είναι η χρήση ενός κινητού παραθύρου πρόσφατων δεδομένων: το μοντέλο εκπαιδεύεται διαρκώς μόνο στα πιο πρόσφατα δείγματα, “ξεχνώντας” σταδιακά τα παλαιότερα. Έχειδειχθεί ότι η εκπαίδευση με μόνο τα πιο πρόσφατα δεδομένα είναι συχνά απαραίτητη για την αποφυγή υποβάθμισης λόγω concept drift. Επιπλέον, ειδικές αρχιτεκτονικές μπορούν να βοηθήσουν, όπως ensemble μοντέλα που διατηρούν πολλά ταξινομητικά υπο-μοντέλα εκπαιδευμένα σε διαφορετικές χρονικές περιόδους: όταν το νέο μοτίβο αποκλίνει, κάποια υπο-μοντέλα θα το καλύψουν καλύτερα και το σύστημα μπορεί να αυξομειώσει το βάρος τους ανάλογα. Τέτοιες προσεγγίσεις μειώνουν τον κίνδυνο να βασίζεται το σύστημα σε μία μόνο “ξεπερασμένη” υπόθεση για τα δεδομένα (Gama et al., 2014).

**Προκλήσεις:** Η βασική δυσκολία είναι να ξεχωρίσει κανείς μια πραγματική αλλαγή *concept* από προσωρινό θόρυβο. Η υπερβολικά συχνή προσαρμογή (π.χ. πολύ μικρό παράθυρο δεδομένων) μπορεί να οδηγήσει σε αστάθεια, ενώ η αργοπορημένη προσαρμογή σημαίνει ότι το μοντέλο θα έχει ήδη μειωμένη απόδοση για ένα διάστημα. Επιπλέον, σε ορισμένες περιπτώσεις οι παλιές γνώσεις παραμένουν σχετικές (π.χ. επαναφορά παλαιότερων θεμάτων ή εποχικές τάσεις): ένα μοντέλο που έχει τελείως ξεχάσει το παρελθόν του ίσως αποτύχει να αναγνωρίσει την επανεμφάνιση παλαιών μοτίβων. Συνεπώς, πολλές προσεγγίσεις drift προσπαθούν να διατηρούν ισορροπία, π.χ. διατηρώντας ένα μικρό υποσύνολο παλιών δεδομένων ως “μνήμη”. Τέλος, η υπολογιστική επιβάρυνση της συνεχούς παρακολούθησης και προσαρμογής δεν είναι αμελητέα: σε περιπτώσεις *big data* ροών (όπως το πλήθος

δεδομένων κοινωνικών δικτύων) πρέπει οι αλγόριθμοι ανίχνευσης και ενημέρωσης να είναι ιδιαίτερα αποδοτικοί (Suprem et al., 2019).

Η σταδιακή (προσαυξητική) εκπαίδευση αναφέρεται στην τεχνική όπου το μοντέλο εκπαιδεύεται επανειλημμένα με νέα δεδομένα που προστίθενται σταδιακά, αντί να εκπαιδεύεται μία φορά εξ αρχής. Ουσιαστικά, αποτελεί μια ενδιάμεση προσέγγιση μεταξύ batch και online learning: το μοντέλο μπορεί να ενημερώνεται ανά τακτά διαστήματα ή όταν συσσωρευτεί ένας επαρκής όγκος νέων δειγμάτων. Για παράδειγμα, ένα σύστημα ανίχνευσης παραπληροφόρησης μπορεί να επανεκπαιδεύεται κάθε βδομάδα ή μήνα, ενσωματώνοντας στο σύνολο εκπαίδευσης τις πρόσφατες αναρτήσεις που έχουν σημειωθεί ως ψευδείς ή αληθείς ειδήσεις (Ariga, 2025). Έτσι, η γνώση του μοντέλου επεκτείνεται προοδευτικά, χωρίς να ξεκινά από μηδενική βάση κάθε φορά.

**Πλεονεκτήματα:** Η σταδιακή εκπαίδευση εκμεταλλεύεται το γεγονός ότι το ήδη εκπαιδευμένο μοντέλο έχει μάθει μια βάση γνώσης που μπορεί να διατηρηθεί. Κάθε νέα εκπαίδευση *fine-tuning* πάνω στο υπάρχον μοντέλο απαιτεί συνήθως λιγότερο χρόνο από μια πλήρη εκπαίδευση εξ αρχής, ειδικά αν το μοντέλο είναι μεγάλης κλίμακας (π.χ. ένα βαθύ νευρωνικό δίκτυο). Επιτρέπει επίσης τη διατήρηση παλαιότερων δεδομένων στον βαθμό που αυτά παραμένουν στο σύνολο εκπαίδευσης (αν αθροίζουμε δεδομένα) ή τουλάχιστον στη μορφή των ήδη μαθημένων παραμέτρων. Συγκριτικά με την αμιγώς online ενημέρωση, η σταδιακή επανεκπαίδευση σε μπλοκ δεδομένων μπορεί να είναι πιο σταθερή, καθώς *ομαλοποιεί* την επίδραση μεμονωμένων ακραίων δειγμάτων (κάθε μπλοκ είναι αντιπροσωπευτικό μεγαλύτερου διαστήματος) (Hoi et al., 2018).

**Προκλήσεις:** Καθώς συσσωρεύονται δεδομένα, μια στρατηγική είναι να περιορίζεται το σύνολο εκπαίδευσης στα πιο πρόσφατα (όπως προαναφέρθηκε με τα παράθυρα). Αν όμως διατηρούμε ολόκληρο το ιστορικό και απλώς προσθέτουμε δεδομένα, το κόστος εκπαίδευσης μπορεί να αυξηθεί σημαντικά με τον χρόνο. Επιπλέον, υπάρχει κίνδυνος το μοντέλο να δομήσει την εκπαίδευσή του γύρω από ένα μεροληπτικό ή μη αντιπροσωπευτικό σύνολο αν το *concept drift* είναι σταθερό: αν π.χ. τα δεδομένα των τελευταίων μηνών είναι έντονα μονομερή θεματικά, το μοντέλο θα προσαρμοστεί σε αυτά εις βάρος παλαιότερων πιο ποικίλων δεδομένων. Μια λύση είναι η σταθμισμένη εκπαίδευση, όπου δίνεται μεγαλύτερο βάρος στα πρόσφατα δεδομένα αλλά δεν εγκαταλείπονται εντελώς τα παλιά. Άλλη πρόκληση είναι η συμβατότητα των αρχιτεκτονικών: όχι όλα τα μοντέλα δεν υποστηρίζουν εύκολα την επαυξητική εκπαίδευση (Gama et al., 2014). Για παράδειγμα, ορισμένοι αλγόριθμοι δέντρων αποφάσεων ή μοντέλα μεθόδων συνόλου μπορεί να χρειάζονται ειδικές τροποποιήσεις για να ενημερώνονται χωρίς επανεκκίνηση.

Η ανάγκη για συνεχή επικαιροποίηση των μοντέλων είναι ιδιαίτερα εμφανής στα social media, όπου το περιεχόμενο και η συμπεριφορά των χρηστών εξελίσσονται αιατάπαυστα.

Ακολουθούν μερικά χαρακτηριστικά παραδείγματα εφαρμογών και πώς οι παραπάνω τεχνικές παίζουν ρόλο:

- **Ανάλυση συναισθήματος:** Το λεξιλόγιο και οι εκφράσεις στα κοινωνικά δίκτυα μεταβάλλονται διαρκώς. Νέες λέξεις (συχνά αργκό ή memes) εμφανίζονται και αποκτούν συναισθηματική σημασία, ενώ παλαιότερες χάνουν τη σημασία τους. Ένα μοντέλο sentiment analysis που δεν ενημερώνεται θα αρχίσει να αποτυγχάνει να αναγνωρίσει σωστά το συναίσθημα πίσω από νέες φράσεις ή hashtags. Μέσω online learning ή συχνής σταδιακής επανεκπαίδευσης, το μοντέλο μπορεί να διατηρεί υψηλή ακρίβεια παρά τις γλωσσικές αλλαγές (Müller & Salathé, 2019). Για παράδειγμα, εάν ξεκινήσει μια νέα δημοφιλής φράση που εκφράζει ειρωνεία, ένα ευέλικτο μοντέλο θα την μάθει γρήγορα από τα συμφραζόμενα και τις αντιδράσεις των χρηστών.
- **Ανίχνευση παραπληροφόρησης (fake news):** Οι τακτικές και τα θέματα παραπληροφόρησης εξελίσσονται όσο οι διακινητές ψευδών ειδήσεων προσαρμόζονται και εκμεταλλεύονται επίκαιρα γεγονότα. Με την πάροδο του χρόνου, τα χαρακτηριστικά που διέκριναν τις ψευδείς ειδήσεις πριν από ένα έτος μπορεί να μην ισχύουν για τις καινούριες περιπτώσεις. Μελέτες έχουν δείξει ότι οι αλλαγές στις τάσεις και στο ύφος του περιεχομένου μπορούν να οδηγήσουν σε δραματική πτώση ακρίβειας ενός ανιχνευτή fake news, εκτός αν το σύστημα αντιμετωπίσει ενεργά το concept drift. Ως εκ τούτου, η συνεχής προσαρμογή είναι ζωτικής σημασίας: ένα σύστημα θα πρέπει να εντοπίζει εγκαίρως ότι το μοντέλο του παλιώνει και να το επανεκπαιδεύει με πρόσφατα δεδομένα. Η χρήση συναισθηματικών χαρακτηριστικών αντί απλώς λέξεων έχει προταθεί ως τρόπος μείωσης της ευαισθησίας ενός μοντέλου στο drift, καθώς π.χ. το ύφος του συναισθήματος σε ψευδείς ειδήσεις μπορεί να παραμένει πιο σταθερό από τις ίδιες τις λέξεις (Suprem et al., 2019). Παρ' όλ' αυτά και αυτά τα μοντέλα χρειάζονται περιοδική ενημέρωση για να παρακολουθούν νέες μορφές παραπληροφόρησης.
- **Παρακολούθηση πολιτικής και κοινωνικής δυναμικής:** Τα κοινωνικά δίκτυα χρησιμοποιούνται ως "βαρόμετρο" της κοινής γνώμης και της κοινωνικής τάσης. Συστήματα που παρακολουθούν αυτές τις δυναμικές – π.χ. για να εκτιμήσουν την στάση των πολιτών σε μια πολιτική απόφαση ή για να ανιχνεύσουν την άνοδο ενός κοινωνικού κινήματος – πρέπει να ανταποκρίνονται σε ξαφνικές μεταβολές. Ένα γεγονός όπως μια κρίση, μια εκλογική αναμέτρηση ή ένα viral βίντεο μπορεί να αλλάξει απότομα το θέμα και το συναίσθημα των συζητήσεων. Τα μοντέλα πρόγνωσης ή ταξινόμησης που τροφοδοτούν τέτοιου είδους monitoring χρειάζεται να εκπαιδεύονται εκ νέου ή να ενημερώνονται online ώστε να *πιάνουν τον παλμό* των νέων εξελίξεων. Για παράδειγμα, αν ξαφνικά εμφανιστεί ένα νέο hashtag που συγκεντρώνει μαζική προσοχή, το σύστημα θα πρέπει να το αντιληφθεί και να μάθει τη σημασία του μέσα στο αντίστοιχο συμφραζόμενο. Ερευνητικά ευρήματα υποστηρίζουν ότι τα συστήματα ανάλυσης κοινωνικών μέσων πρέπει να αντιμετωπίζουν το concept drift συνεχώς για να αποφεύγονται συστηματικές αστοχίες, ειδικά σε περιόδους κρίσεων όπου τα δεδομένα μεταβάλλονται ραγδαία (Müller & Salathé, 2019).

Συνολικά, η συνεχής προσαρμογή μοντέλων δεν είναι απλώς ένα τεχνικό πολυτέλεια αλλά ουσιώδης απαίτηση για διατήρηση της αξιοπιστίας και της ακρίβειας σε περιβάλλοντα με μη στάσιμα δεδομένα. Κάθε προσέγγιση – online learning, ειδικοί χειρισμοί του concept drift, incremental training – προσφέρει ένα σύνολο εργαλείων για την αντιμετώπιση της πρόκλησης αυτής. Το κλειδί βρίσκεται στην κατάλληλη συνδυαστική χρήση τους: για παράδειγμα, ένα σύστημα μπορεί να χρησιμοποιεί ανιχνευτές concept drift για να αποφασίσει πότε θα ενεργοποιήσει μια σταδιακή επανεκπαίδευση του μοντέλου ή να εκμεταλλεύεται την online ενημέρωση σε συνδυασμό με ένα παράθυρο πρόσφατων δεδομένων. Με αυτόν τον τρόπο, τα μοντέλα μπορούν να παραμένουν *επίκαιρα* και *αξιόπιστα*, παρέχοντας συνεπείς προβλέψεις παρά τις διαρκείς αλλαγές στα δεδομένα των κοινωνικών δικτύων.

## ***8.5 Ενσωμάτωση με Τεχνολογίες Metaverse και Augmented Social Media***

Το Metaverse έχει χαρακτηριστεί ως ένα «μετα-πραγματικό» σύμπαν – ένα διαρκές, πολυχρηστικό εικονικό περιβάλλον που συγχωνεύει τη φυσική πραγματικότητα με την ψηφιακή εικονικότητα (Mystakidis, 2022). Σε αυτό το πλαίσιο, ο όρος Augmented Social Media αναφέρεται στις εμπλουτισμένες εμπειρίες μέσω κοινωνικής δικτύωσης μέσω τεχνολογιών επαυξημένης και εικονικής πραγματικότητας, όπου οι χρήστες αλληλεπιδρούν με ψηφιακό περιεχόμενο ενσωματωμένο στον πραγματικό κόσμο. Η εξέλιξη αυτή θεωρείται το επόμενο στάδιο της κοινωνικής δικτύωσης, μετατρέποντας την παραδοσιακή 2D online επικοινωνία σε μια καθηλωτική εμπειρία 3D με συνεχείς αλληλεπιδράσεις και παρουσία σε πραγματικό χρόνο μέσα σε εικονικούς κόσμους.

**Πλατφόρμες και Υποδομές:** Η υλοποίηση του metaverse και των augmented social media βασίζεται σε ένα σύνθετο οικοσύστημα πλατφορμών και τεχνολογιών υποδομών. Ήδη έχουν αναπτυχθεί immersive κοινωνικά δίκτυα και εικονικές πλατφόρμες όπου πολλοί χρήστες συνυπάρχουν ως ψηφιακά avatars – ενδεικτικά παραδείγματα αποτελούν το VRChat, το Horizon Worlds της Meta και αποκεντρωμένα εικονικά περιβάλλοντα όπως το Decentraland. Αυτές οι πλατφόρμες αξιοποιούν προηγμένα VR/AR SDKs και μηχανές γραφικών (π.χ. Unity, Unreal Engine) για τη δημιουργία τρισδιάστατων κόσμων, καθώς και τεχνολογίες cloud για τη φιλοξενία μεγάλου αριθμού ταυτόχρονων χρηστών. Παράλληλα, η blockchain τεχνολογία ενσωματώνεται για τη διαχείριση ψηφιακής ταυτότητας και ιδιοκτησίας: μέσω αποκεντρωμένων μητρώων, οι χρήστες μπορούν να κατέχουν με ασφάλεια τα avatar τους, τα ψηφιακά αγαθά ή τα συλλεκτικά στοιχεία (NFTs) στον εικονικό χώρο (Gadekallu et al., 2023). Αυτό το πολυεπίπεδο τεχνολογικό υπόβαθρο διασφαλίζει ότι το metaverse λειτουργεί ως μια ενιαία, διασυνδεδεμένη εμπειρία, όπου διαφορετικές πλατφόρμες και συσκευές (από VR headsets μέχρι AR mobile εφαρμογές) συνεργάζονται απρόσκοπτα.

**Αλγόριθμοι για Περιεχόμενο και Αλληλεπίδραση:** Στα εικονικά και μεικτά περιβάλλοντα του metaverse, κρίσιμο ρόλο παίζουν οι εξελιγμένοι αλγόριθμοι μηχανικής μάθησης και γραφικών που επιτρέπουν τη δημιουργία, διαχείριση και δυναμική προσαρμογή του περιεχομένου. Αλγόριθμοι γεννητικής τεχνητής νοημοσύνης χρησιμοποιούνται για την αυτόματη δημιουργία ρεαλιστικών τρισδιάστατων αντικειμένων και σκηνών – για παράδειγμα, μοντέλα Deep Learning μπορούν να παράγουν διαδικαστικά εικονικά τοπία ή πιστές ψηφιακές ανθρώπινες μορφές. Επιπλέον, σε εφαρμογές Augmented Reality (AR) εφαρμόζονται μέθοδοι ταυτόχρονης εντοπισμού και χαρτογράφησης (SLAM algorithms) ώστε το σύστημα να αναγνωρίζει το φυσικό περιβάλλον και να τοποθετεί σταθερά τα ψηφιακά στοιχεία μέσα σε αυτό. Η διαχείριση του περιεχομένου σε πραγματικό χρόνο απαιτεί επίσης αποδοτικούς αλγόριθμους ροής δεδομένων και βελτιστοποίησης δικτύου, προκειμένου οι εικονικές εμπειρίες να παραμένουν ομαλές ακόμη και με χιλιάδες συμμετέχοντες. Ταυτόχρονα, οι ίδιες οι πλατφόρμες συλλέγουν τεράστιο όγκο δεδομένων (κίνηση χρηστών, προτιμήσεις, κοινωνικά γραφήματα), τα οποία αναλύονται με τεχνικές big data και analytics. Με αυτόν τον τρόπο, μοντέλα machine learning μπορούν να

εξατομικεύουν το περιεχόμενο για κάθε χρήστη – από την προσαρμογή του περιβάλλοντος και των προτάσεων κοινωνικής αλληλεπίδρασης, μέχρι τον εντοπισμό ανάρμοστης συμπεριφοράς και τον αυτόματο μετριασμό περιεχομένου. Χαρακτηριστικά, έχει υποστηριχθεί ότι η αξιοποίηση της ΤΝ στο metaverse επιτρέπει τη δημιουργία ιδιαίτερα προσωποποιημένων εμπειριών βάσει της συμπεριφοράς και των προτιμήσεων του κάθε χρήστη, αυτοματοποιώντας παράλληλα επαναλαμβανόμενες διαδικασίες διαχείρισης (Almeman et al., 2025).

**Avatars και Πράκτορες Τεχνητής Νοημοσύνης:** Στις καθελωτικές κοινωνικές εμπειρίες του metaverse, οι χρήστες εκπροσωπούνται από avatars, δηλαδή ψηφιακές ενσαρκώσεις με εμφάνιση που μπορεί να κυμαίνεται από ρεαλιστική ανθρώπινη μορφή μέχρι φανταστικούς χαρακτήρες. Η αλληλεπίδραση μεταξύ χρηστών μέσω των avatars τους υποβοηθάται από αλγόριθμους ανίχνευσης κίνησης και αναγνώρισης εκφράσεων: για παράδειγμα, συστήματα καταγραφής χειρονομιών και προσώπου επιτρέπουν στο avatar να αντικατοπτρίζει τις κινήσεις και τα συναισθήματα του χρήστη σε πραγματικό χρόνο. Πέρα όμως από τα avatars των πραγματικών χρηστών, στο metaverse συμμετέχουν και αυτόνομοι ψηφιακοί χαρακτήρες – οι λεγόμενοι agents ή πράκτορες ΤΝ. Αυτοί μπορεί να είναι AI avatars ή bots που κινούνται και συνομιλούν αυτόνομα, καθοδηγούμενοι από αλγόριθμους τεχνητής νοημοσύνης (π.χ. μοντέλα NLP για κατανόηση και παραγωγή φυσικής γλώσσας). Στόχος των agents είναι να εμπλουτίσουν το περιβάλλον με διαδραστικότητα, λειτουργώντας ως εικονικοί βοηθοί, οδηγοί ή ακόμη και σύντροφοι των χρηστών. Ερευνητικές μελέτες δείχνουν ότι οι άνθρωποι μπορούν να αντιδράσουν κοινωνικά προς αυτούς τους εικονικούς ανθρώπους – είτε πρόκειται για avatars άλλων χρηστών είτε για αυτόνομους agents – ιδίως όταν η συμπεριφορά και η εμφάνισή τους γίνονται αρκετά πειστικές (Kyrllitsias & Michael-Grigoriou, 2022). Αυτό υπογραμμίζει τη σημασία της ρεαλιστικής απεικόνισης και της πειθοποιητικής τεχνητής νοημοσύνης στην ενίσχυση του αισθήματος social presence, δηλαδή της αίσθησης ότι «συναντούμε» πραγματικά άλλους ανθρώπους μέσα στο εικονικό περιβάλλον.

**Εφαρμογές και Παραδείγματα:** Η σύζευξη του metaverse με τα κοινωνικά δίκτυα έχει ήδη οδηγήσει σε καινοφανείς εφαρμογές που επεκτείνουν τα όρια της κοινωνικής αλληλεπίδρασης. Ένα χαρακτηριστικό παράδειγμα αποτελούν οι virtual influencers – ψηφιακοί «διάσημοι» στα social media, που εμφανίζονται ως ανθρωπόμορφοι χαρακτήρες δημιουργημένοι με γραφικά υπολογιστών και τεχνητή νοημοσύνη. Πρόσωπα όπως η Lil Miquela στο Instagram συγκεντρώνουν εκατομμύρια ακολούθους, επηρεάζοντας την προώθηση προϊόντων και τη μόδα, παρόλο που δεν υπάρχουν φυσικά. Η ύπαρξή τους κατέστη δυνατή μέσω αλγορίθμων CGI για την απόδοση ρεαλιστικών εικόνων, σε συνδυασμό με στρατηγικές σεναρίων και μηχανικής μάθησης που καθοδηγούν τη «προσωπικότητά» τους στα μέσα (Gulan et al., 2025). Παρομοίως, πολλές εταιρείες εξερευνούν AI avatars ως εικονικούς εκπροσώπους για εξυπηρέτηση πελατών ή ως ψηφιακούς παρουσιαστές, αξιοποιώντας την ικανότητα των μοντέλων ΤΝ να παράγουν ανθρώπινη φωνή

και έκφραση. Επιπλέον, εμφανίζονται agents ευαίσθητοι στο συναίσθημα (sentiment-aware agents) σε πλατφόρμες κοινωνικής VR: αυτοί οι πράκτορες μπορούν να αναλύουν τον τόνο φωνής ή το κείμενο στις συνομιλίες και να ανιχνεύουν συναισθήματα, προσαρμόζοντας ανάλογα τις αντιδράσεις τους. Για παράδειγμα, ένας εικονικός βοηθός θα μπορούσε να αντιληφθεί ότι ένας χρήστης είναι θυμωμένος από τη χροιά της φωνής του και να τροποποιήσει την προσέγγισή του – τεχνική που αντλεί από τον κλάδο του affective computing στην ανάλυση κοινωνικών δεδομένων. Τέλος, η έννοια του spatial computing βρίσκει εφαρμογή στην κοινωνική δικτύωση μέσω επαυξημένης πραγματικότητας: οι χρήστες με συσκευές AR (π.χ. έξυπνα γυαλιά) μπορούν να βλέπουν ψηφιακές πληροφορίες υπερτιθέμενες στους γύρω ανθρώπους και χώρους, μετατρέποντας την φυσική συνάντηση σε μια εμπειρία επαυξημένου κοινωνικού δικτύου. Για παράδειγμα, σε ένα συνέδριο με AR γυαλιά, θα ήταν εφικτό να εμφανίζονται πάνω από κάθε άτομο τα στοιχεία του προφίλ του από το LinkedIn ή πρόσφατες αναρτήσεις του, παρέχοντας ένα νέο επίπεδο πληροφορίας στη διά ζώσης συζήτηση. Αυτός ο συγκεκριασμός φυσικού και ψηφιακού επιπέδου καθιστά τους χρήστες ενεργούς συμμετέχοντες σε μικτές κοινωνικές εμπειρίες, επεκτείνοντας την διάδραση πέρα από τα όρια της οθόνης (Annetta et. al., 2024).

Όλες αυτές οι εφαρμογές όχι μόνο θολώνουν τα όρια μεταξύ πραγματικότητας και ψηφιακού κόσμου, αλλά παράγουν και πρωτοφανείς όγκους δεδομένων αλληλεπίδρασης. Η ανάλυση αυτών των big data με σύγχρονες μεθόδους κοινωνικής δικτύωσης και μηχανικής μάθησης προσφέρει νέες δυνατότητες κατανόησης της ανθρώπινης συμπεριφοράς, ανοίγοντας δρόμους για καινοτόμες υπηρεσίες και έρευνα στην κοινωνική αναλυτική του metaverse.

## Κεφάλαιο 9: Συμπεράσματα

### 9.1 Συνοπτική Επισκόπηση

Η αλματώδης εξάπλωση των μέσων κοινωνικής δικτύωσης (ΜΚΔ) έχει δημιουργήσει τεράστιους όγκους δεδομένων, καθιστώντας απαραίτητη την αξιοποίηση προηγμένων μεθόδων ανάλυσης. Η Μηχανική Μάθηση (ΜΜ) αναδεικνύεται ως καίριο εργαλείο για την επεξεργασία αυτών των δεδομένων, επιτρέποντας την εξαγωγή γνώσης και την ανάπτυξη εφαρμογών σε ένα ευρύ φάσμα πεδίων – από το μάρκετινγκ μέχρι την ανίχνευση ψευδών ειδήσεων. Στο πλαίσιο αυτό, παρουσιάζεται το θεωρητικό υπόβαθρο των ΜΚΔ και των μεθόδων ΜΜ, με έμφαση στις βασικές κατηγορίες αλγορίθμων (επιβλεπόμενη, μη επιβλεπόμενη και ενισχυτική μάθηση) και στα νευρωνικά δίκτυα βαθιάς μάθησης που αποτελούν τη βάση των σύγχρονων προσεγγίσεων. Ο στόχος είναι να δοθεί μια ολοκληρωμένη εικόνα του τρόπου με τον οποίο οι αλγόριθμοι ΜΜ συμβάλλουν στην κατανόηση της κοινωνικής συμπεριφοράς και των τάσεων σε ένα διαρκώς εξελισσόμενο ψηφιακό οικοσύστημα (Kapur & Harikrishnan, 2022).

Ένα μεγάλο μέρος των δεδομένων στα ΜΚΔ αποτελείται από κείμενο γραμμένο σε φυσική γλώσσα (π.χ. αναρτήσεις, σχόλια). Η ανάλυση αυτού του περιεχομένου απαιτεί προηγμένες τεχνικές Natural Language Processing (NLP) για την εξαγωγή νοήματος, την αναγνώριση οντοτήτων και τη σύνοψη πληροφοριών από μη δομημένο κείμενο. Δεδομένης της ιδιομορφίας της γλώσσας στα social media (π.χ. slang, συντομογραφίες, emojis), οι παραδοσιακές μέθοδοι επεξεργασίας κειμένου επεκτείνονται με σύγχρονες προσεγγίσεις που συνδυάζουν σημασιολογική ανάλυση και μηχανική μάθηση. Για παράδειγμα, τεχνικές ανάλυσης θεματολογίας (topic modeling) και μοντέλα μετασχηματιστών εφαρμόζονται για την κατανόηση του περιεχομένου σε μεγάλη κλίμακα. Έχει τεκμηριωθεί ότι η χρήση εξελιγμένων μεθόδων NLP βελτιώνει αισθητά την ποιότητα της πληροφορίας που εξάγεται από τα χρήστη-παράγόμενα κείμενα, επιτρέποντας ακριβέστερη κατηγοριοποίηση και εξαγωγή συμπερασμάτων. Συνεπώς, οι τεχνικές επεξεργασίας φυσικής γλώσσας αποτελούν θεμέλιο για πολλές εφαρμογές ανάλυσης κοινωνικών δικτύων.

Η ταξινόμηση συναισθήματος (sentiment analysis) αποτελεί μία από τις πιο σημαντικές εφαρμογές NLP στα κοινωνικά μέσα. Στοχεύει στην αυτόματη ανίχνευση και κατηγοριοποίηση της συναισθηματικής φόρτισης που εκφράζουν οι χρήστες σε αναρτήσεις ή σχόλια, ταξινομώντας τα π.χ. ως θετικά, αρνητικά ή ουδέτερα. Μέσω της ανάλυσης συναισθήματος αποκαλύπτεται η στάση του κοινού απέναντι σε προϊόντα, υπηρεσίες ή γεγονότα σε πραγματικό χρόνο, παρέχοντας πολύτιμα insights σε οργανισμούς και επιχειρήσεις. Για παράδειγμα, μια εταιρεία μπορεί να παρακολουθεί το συναίσθημα των χρηστών σχετικά με το brand της στα tweets και τα σχόλια, ώστε να αντιληφθεί έγκαιρα πιθανές κρίσεις ή να μετρήσει την αποδοχή μιας νέας καμπάνιας (Rodríguez-Ibáñez et al.,

2023). Η αξιοπιστία των αποτελεσμάτων εξαρτάται από την ποιότητα των δεδομένων και των λεξιλογίων/μοντέλων που χρησιμοποιούνται, όμως τα σύγχρονα βαθιά νευρωνικά δίκτυα (π.χ. BERT, LSTM) έχουν βελτιώσει σημαντικά την ακρίβεια της ανάλυσης συναισθήματος.

Με την εξάπλωση των κοινωνικών μέσων, η παραπληροφόρηση και τα ψευδή νέα (fake news) έχουν καταστεί μείζον ζήτημα. Η ανεξέλεγκτη διάδοση αναληθών ή χειραγωγημένων πληροφοριών μέσω πλατφορμών όπως το Facebook και το Twitter μπορεί να προκαλέσει σοβαρές κοινωνικές και πολιτικές επιπτώσεις. Οι αλγόριθμοι MM χρησιμοποιούνται ενεργά για την ανίχνευση ψευδών ειδήσεων, φιλτράροντας το παραπλανητικό περιεχόμενο. Τεχνικές όπως η ανάλυση περιεχομένου και δικτύων διασποράς εφαρμόζονται ώστε να εντοπιστούν μοτίβα που χαρακτηρίζουν τα fake news (π.χ. υπερβολικά ομογενοποιημένη γλώσσα ή ασυνήθιστα σχήματα διάδοσης). Η έρευνα έχει δείξει ότι συνδυάζοντας χαρακτηριστικά κειμένου, μετα-δεδομένα (π.χ. πότε και από ποιον δημοσιεύθηκε) και δικτυακά γνωρίσματα, τα μοντέλα μπορούν να διακρίνουν με ολοένα αυξανόμενη ακρίβεια την παραπληροφόρηση file-6bot3qynkzj1bcd3rjbbtv. Παρά τις προόδους, παραμένουν προκλήσεις όπως η αντιμετώπιση ολοένα πιο εξελιγμένων deep fakes και η εξισορρόπηση ανάμεσα στη λογοκρισία και την ελεύθερη έκφραση (Shu et al., 2017).

Πέρα από το περιεχόμενο, οι αλγόριθμοι MM χρησιμοποιούνται για την πρόβλεψη συμπεριφοράς χρηστών και την στοχευμένη παροχή περιεχομένου. Αναλύοντας ιστορικά δεδομένα χρήσης, μπορούν να προβλέψουν μελλοντικές ενέργειες των χρηστών, όπως την πιθανότητα να αλληλεπιδράσουν με συγκεκριμένες αναρτήσεις ή να εγκαταλείψουν την πλατφόρμα. Τέτοια predictive analytics βοηθούν τις επιχειρήσεις να βελτιστοποιήσουν τις στρατηγικές τους: π.χ. ένα μοντέλο μπορεί να προβλέψει ποιοι χρήστες κινδυνεύουν να πάψουν να είναι ενεργοί και έτσι η πλατφόρμα να λάβει μέτρα για την διατήρησή τους (προσφέροντας κατάλληλο περιεχόμενο ή προσφορές). Επίσης, στην στοχευμένη διαφήμιση, τα μοντέλα μηχανικής μάθησης συνδυάζουν δημογραφικά και συμπεριφορικά δεδομένα για να προβάλλουν σε κάθε χρήστη διαφημίσεις που ταιριάζουν στα ενδιαφέροντά του. Αυτές οι προγνωστικές δυνατότητες έχουν σημαντική οικονομική αξία, καθώς μεγιστοποιούν το engagement και την ικανοποίηση χρήστη, ενισχύοντας παράλληλα την απόδοση επενδύσεων στο μάρκετινγκ (Lin, 2016). Παράλληλα, εγείρονται ζητήματα ιδιωτικότητας, αφού η εκτεταμένη συλλογή και ανάλυση προσωπικών δεδομένων απαιτεί τηρητές ασφαλείας και κανονιστικές ρυθμίσεις (π.χ. συμμόρφωση με GDPR).

Η δομή των κοινωνικών δικτύων και οι σχέσεις μεταξύ των χρηστών αναλύονται μέσω τεχνικών ανάλυσης κοινοτήτων και ανάλυσης επιρροής. Με την ανάλυση κοινοτήτων (community detection) εντοπίζονται ομάδες χρηστών με κοινά ενδιαφέροντα ή πρότυπα συμπεριφοράς. Αυτές οι κοινότητες μπορούν να αποκαλύψουν υπο-δικτυώματα μέσα σε πλατφόρμες όπως το Twitter (π.χ. ομάδες γύρω από ένα θέμα ή hashtag), βοηθώντας στην κατανόηση της κοινωνικής δομής και της δυναμικής διάδοσης της πληροφορίας. Η αναγνώριση των κοινοτήτων επιτρέπει πιο στοχευμένες παρεμβάσεις – για παράδειγμα, μια καμπάνια δημόσιας υγείας μπορεί να εντοπίσει μια κοινότητα σκεπτικιστών και να προσαρμόσει ανάλογα το μήνυμά της. Σχετική με την κοινοτική δομή είναι και η έννοια της επιρροής: σε κάθε δίκτυο, ορισμένοι κόμβοι (χρήστες) λειτουργούν ως “influencers” ή

κόμβοι-κλειδιά διάχυσης της πληροφορίας. Η ανάλυση επιρροής (influence analysis) μελετά ποιοι χρήστες έχουν τη μεγαλύτερη απήχηση ή ικανότητα να επηρεάζουν άλλους (π.χ. μέσω αναδημοσιεύσεων, likes κ.λπ.). Αυτή η ανάλυση έχει άμεσες πρακτικές εφαρμογές στο influencer marketing, όπου οι εταιρείες επιλέγουν συνεργασίες με βάση το πραγματικό δίκτυο επιρροής ενός χρήστη και όχι μόνο τον αριθμό ακολούθων του. Επιπλέον, η χαρτογράφηση των «ηγετών γνώμης» και των κοινοτήτων βοηθά στην κατανόηση του πώς διαχέονται οι πληροφορίες και οι τάσεις στο κοινωνικό σύνολο. Σε ένα πιο στρατηγικό επίπεδο, η ανίχνευση των βασικών κόμβων επιρροής είναι κρίσιμη και για την προστασία του δημόσιου διαλόγου: η ύπαρξη αυτόκλητων επιδραστών ή ακόμη και αυτοματοποιημένων bots που χειραγωγούν τη συζήτηση (π.χ. διάδοση προπαγάνδας) καθιστά αναγκαία την παρακολούθηση και τον έλεγχο της επιρροής στο διαδίκτυο. Σύγχρονες μελέτες υπογραμμίζουν ότι η κατανόηση των μηχανισμών επιρροής μπορεί να συμβάλει στην αντιμετώπιση συντονισμένων εκστρατειών παραπληροφόρησης και στην ενίσχυση της ανθεκτικότητας των κοινωνικών δικτύων απέναντι σε κακόβουλες παρεμβάσεις (Ferrara, 2017).

Η ενσωμάτωση αλγορίθμων μηχανικής μάθησης στα social media έχει επιφέρει σημαντικά οφέλη ως προς τον τρόπο συλλογής, επεξεργασίας και αξιοποίησης της πληροφορίας. Ταχύτητα και αυτοματοποίηση: Οι μηχανές μαθαίνουν να επεξεργάζονται τεράστιες ροές δεδομένων σε πραγματικό χρόνο, προσφέροντας αποτελέσματα και προβλέψεις ασύγκριτα ταχύτερα από ό,τι θα ήταν δυνατό με χειροκίνητες μεθόδους. Κλίμακα και προσαρμοστικότητα: Τα ML συστήματα κλιμακώνονται σε πλήθος χρηστών και δεδομένων, προσαρμόζοντας τα μοντέλα τους σε νέες πληροφορίες. Αυτά τα χαρακτηριστικά έχουν πολυεπίπεδη επίδραση – ωφελούν τις επιχειρήσεις (π.χ. μέσω αυτοματοποιημένης ανάλυσης αγοραστικής συμπεριφοράς), τους ερευνητές (π.χ. με εξαγωγή κοινωνιολογικών τάσεων από μεγάλα δεδομένα) και τους ίδιους τους χρήστες (π.χ. μέσω προσωποποιημένου περιεχομένου που βελτιώνει την εμπειρία χρήστη). Ενδεικτικά, συστήματα συστάσεων όπως αυτά του YouTube ή του TikTok χρησιμοποιούν MM για να προβλέψουν και να προτείνουν περιεχόμενο που πιθανότατα θα ενδιαφέρει τον εκάστοτε χρήστη, αυξάνοντας την ικανοποίηση και την αλληλεπίδρασή του με την πλατφόρμα. Επιπρόσθετα, η αυτοματοποίηση συμβάλλει στην αποτελεσματικότερη διαχείριση περιεχομένου: εργασίες όπως ο εντοπισμός spam, η φιλτραρισμός προσβλητικού υλικού ή η ανίχνευση παραβίασης πολιτικών κοινότητας μπορούν να διεκπεραιωθούν σε μεγάλο βαθμό αυτόματα, κάτι ιδιαίτερα κρίσιμο όταν δημοσιεύονται εκατομμύρια μηνύματα ημερησίως. Συνολικά, η MM βελτιστοποιεί την απόδοση και αποτελεσματικότητα των πλατφορμών κοινωνικής δικτύωσης (Gandomi & Haider, 2015).

Παρά τα οφέλη, προκύπτουν σημαντικές προκλήσεις από την εφαρμογή της MM στα κοινωνικά δίκτυα. Ποιότητα και προκατάληψη δεδομένων: Τα μοντέλα είναι τόσο καλά όσο τα δεδομένα με τα οποία εκπαιδεύονται. Τα social media δεδομένα συχνά περιέχουν θόρυβο, ανακρίβειες ή δεν είναι αντιπροσωπευτικά του πληθυσμού, οδηγώντας σε

μεροληπτικά ή ανακριβή αποτελέσματα. Διαφάνεια και ερμηνευσιμότητα: Πολλοί προηγμένοι αλγόριθμοι (π.χ. deep learning) λειτουργούν ως «μαύρα κουτιά», καθιστώντας δύσκολο να εξηγηθεί γιατί πάρθηκε μια απόφαση. Αυτό έχει τονίσει την ανάγκη για ανάπτυξη πιο ερμηνεύσιμων και δίκαιων μοντέλων, καθώς και πλαισίων αξιολόγησης της ευθυδικίας τους. Ιδιωτικότητα και ηθική: Η συλλογή και ανάλυση προσωπικών δεδομένων χρηστών εγείρει ζητήματα ιδιωτικότητας και συναίνεσης. Υπάρχει λεπτή ισορροπία μεταξύ της αξιοποίησης των δεδομένων και της προστασίας των δικαιωμάτων των χρηστών – χαρακτηριστικό είναι το παράδειγμα του σκανδάλου Cambridge Analytica, που ανέδειξε το πώς τα δεδομένα μπορούν να χρησιμοποιηθούν καταχρηστικά. Σε απάντηση, έχουν αρχίσει να εμφανίζονται κανονιστικά πλαίσια και πρωτοβουλίες, όπως το κίνημα Fairness, Accountability and Transparency in ML (FAT/ML), που στοχεύουν στη ρύθμιση και καθοδήγηση της ανάπτυξης ηθικών αλγορίθμων. Τέλος, ζητήματα ασφαλείας (security) και ανθεκτικότητας των μοντέλων (robustness) επίσης προκύπτουν – επιθέσεις adversarial μπορούν να παραπλανήσουν τα μοντέλα, ενώ η συνεχής εξέλιξη της πλατφόρμας απαιτεί διαρκή επανεκτίμηση των υποθέσεων των αλγορίθμων. Η αντιμετώπιση αυτών των προκλήσεων αποτελεί ενεργό πεδίο έρευνας, συνδυάζοντας τεχνικές λύσεις με νομικά και κοινωνικά μέτρα (Barocas et al., 2019).

Η φύση των δεδομένων στα ΜΚΔ – τεράστιος όγκος, ταχύτητα ροής και ποικιλία μορφών – τα κατατάσσει στα Big Data, απαιτώντας εξειδικευμένες τεχνολογίες για την αποθήκευση και επεξεργασία τους. Σύγχρονες πλατφόρμες κατανεμημένης επεξεργασίας όπως το Apache Hadoop και το Spark επιτρέπουν την ανάλυση κοινωνικών δεδομένων σε μεγάλη κλίμακα και σε πραγματικό χρόνο. Για παράδειγμα, η ανάλυση συναισθήματος σε ροές ζωντανών tweets είναι εφικτή μόνο με τη βοήθεια τέτοιων εργαλείων, καθώς “απαιτεί ισχυρές πλατφόρμες ανάλυσης Big Data όπως το Spark” για να διαχειριστεί τον όγκο και την ταχύτητα των δεδομένων. Επιπλέον, τα big data συστήματα επιτρέπουν την ενοποίηση πολλαπλών πηγών δεδομένων: μπορούν να συλλέξουν και να συνεπεξεργαστούν δεδομένα από διάφορες πλατφόρμες (π.χ. Twitter, Instagram, YouTube) ταυτόχρονα. Αυτή η δυνατότητα οδηγεί στην εφαρμογή τεχνικών μετα-ανάλυσης στην ανάλυση κοινωνικών δικτύων – ουσιαστικά, συνδυάζοντας ευρήματα από πολλές διαφορετικές πηγές ή μελέτες για να εξαχθούν πιο έγκυρα και γενικεύσιμα συμπεράσματα. Ο συνδυασμός της ΜΜ με τη μετα-ανάλυση έχει βρεθεί ότι βελτιώνει την αξιοπιστία και την ακρίβεια των αποτελεσμάτων: ενοποιώντας προβλέψεις από πολλαπλά μοντέλα ή σύνολα δεδομένων, μειώνεται η τυχαία σφάλση και μετριάζεται η επίδραση τυχόν μεροληψιών ενός μεμονωμένου δείγματος. Πρακτικά, αυτό πραγματοποιείται μέσω ensemble μεθόδων και συλλογής των αποτελεσμάτων τους – μια προσέγγιση όπου διάφορα υπο-μοντέλα συμβάλλουν από κοινού στην τελική απόφαση. Έχει αποδειχθεί ότι τέτοιες μετα-συνδυαστικές τεχνικές (ensemble learning) συχνά υπερτερούν των μεμονωμένων προσεγγίσεων, ιδίως σε προβλήματα όπως η ανάλυση συναισθήματος, εντοπίζοντας λεπτές τάσεις που μπορεί να διέφευγαν υπό άλλες συνθήκες (Dietterich, 2000). Συνολικά, η αξιοποίηση εργαλείων Big Data σε συνδυασμό με μετα-αναλυτικές μεθόδους ενισχύει τη γενικευσιμότητα των μοντέλων και προσφέρει μια πιο σφαιρική θεώρηση των κοινωνικών φαινομένων πέρα από τα όρια μίας μόνο πλατφόρμας ή μελέτης.

Τα δεδομένα των social media δεν είναι στατικά – αντιθέτως, το περιεχόμενο και η συμπεριφορά των χρηστών μεταβάλλονται διαρκώς. Ένα μοντέλο που λειτουργεί άριστα σήμερα μπορεί να υποστεί μείωση απόδοσης σε λίγους μήνες καθώς αλλάζουν οι τάσεις, η γλώσσα (νέες αργκό, memes) ή το ενδιαφέρον του κοινού. Το φαινόμενο αυτό περιγράφεται ως concept drift, δηλαδή η σταδιακή ή απότομη μετατόπιση της στατιστικής κατανομής των δεδομένων εισόδου και της σχέσης τους με τις εξόδους με την πάροδο του χρόνου. Στα κοινωνικά δίκτυα το drift είναι συχνό – π.χ. μια ξαφνική αλλαγή στη θεματολογία των συζητήσεων μετά από ένα έκτακτο γεγονός (όπως μια πανδημία) ή οι περιοδικές διακυμάνσεις ενδιαφέροντος κατά την προεκλογική περίοδο. Τα μοντέλα MM χρειάζεται να προσαρμόζονται σε αυτές τις αλλαγές, διαφορετικά οι προβλέψεις τους θα παλιώσουν και θα χάσουν ακρίβεια. Για την αντιμετώπιση του concept drift έχουν αναπτυχθεί μέθοδοι προσαρμοστικής μάθησης. Μία προσέγγιση είναι η διαδικτυακή εκμάθηση (online learning) ή αλλιώς σταδιακή/προσαυξητική εκπαίδευση: το μοντέλο ενημερώνεται συνεχώς με νέα δεδομένα καθώς αυτά καταφθάνουν, αντί να εκπαιδεύεται άπαξ σε ένα στατικό σύνολο. Έτσι, διατηρείται “φρέσκο” ως προς τις τρέχουσες τάσεις. Ωστόσο, η συνεχής ενημέρωση ενέχει κινδύνους – αν το μοντέλο προσαρμόζεται υπερβολικά στα πιο πρόσφατα δεδομένα, μπορεί να ξεχάσει χρήσιμα μοτίβα του παρελθόντος (catastrophic forgetting). Για τον λόγο αυτό, συχνά χρησιμοποιούνται τεχνικές όπως sliding windows (το μοντέλο εκπαιδεύεται κάθε φορά μόνο στα πιο πρόσφατα δείγματα, ξεχνώντας σταδιακά τα παλιότερα) ή μέθοδοι που διατηρούν μια μικρή “μνήμη” παλαιών δεδομένων για σταθερότητα. Μια άλλη αποτελεσματική στρατηγική είναι η χρήση συνόλων ταξινομητών (ensemble models) ειδικά σχεδιασμένων για δυναμικά περιβάλλοντα: διατηρούνται πολλαπλά υπο-μοντέλα, το καθένα εκπαιδευμένο σε διαφορετικές χρονικές περιόδους ή χαρακτηριστικά και το σύστημα συνδυάζει τις προβλέψεις τους ανάλογα με το ποιο είναι πιο κατάλληλο τη δεδομένη στιγμή. Τέτοιες προσεγγίσεις μειώνουν τον κίνδυνο να βασιζεται το σύστημα σε μία και μόνη “ξεπερασμένη” υπόθεση για τα δεδομένα. Φυσικά, υπάρχει μια λεπτή ισορροπία: η πολύ συχνή προσαρμογή μπορεί να κάνει το μοντέλο ασταθές (κυνηγώντας θόρυβο), ενώ η πολύ αργή προσαρμογή αφήνει το μοντέλο να λειτουργεί με απαρχαιωμένες παραδοχές. Ως εκ τούτου, η ανίχνευση του πότε συμβαίνει σημαντικό drift είναι κρίσιμη – στατιστικοί έλεγχοι κατανομών ή η παρακολούθηση της απόδοσης του μοντέλου σε νέο δεδομένο μπορούν να σηματοδοτήσουν την ανάγκη προσαρμογής. Η διαχείριση του concept drift εξακολουθεί να αποτελεί πρόκληση για τα συστήματα MM, αλλά η πρόοδός της είναι απαραίτητη ώστε τα μοντέλα να παραμένουν έγκυρα και αξιόπιστα σε ένα πεδίο τόσο δυναμικό όσο τα κοινωνικά δίκτυα (Gama et al., 2014).

Οι πρόσφατες εξελίξεις στη MM, ιδιαίτερα στον τομέα της επεξεργασίας γλώσσας, έχουν φέρει στο προσκήνιο μια νέα γενιά αλγορίθμων. Οι Transformers και τα παράγωγά τους – συμπεριλαμβανομένων προεκπαιδευμένων μοντέλων όπως το BERT και οι Large Language Models (LLMs) τύπου GPT – έχουν φέρει επανάσταση στον τρόπο που αντιμετωπίζουμε δεδομένα κειμένου. Σε αντίθεση με τα παλιότερα επαναληπτικά δίκτυα (RNN/LSTM), οι transformers επιτρέπουν παράλληλη επεξεργασία ακολουθιακών δεδομένων και μπορούν να συλλάβουν μακροπρόθεσμες εξαρτήσεις μέσω του μηχανισμού προσοχής (attention)file-

6bot3qynkzj1bcd3rjbbtv. Αυτό σημαίνει ότι μπορούν να κατανοήσουν καλύτερα το συμφραζόμενο ενός post ή tweet, ακόμα κι αν οι κρίσιμες πληροφορίες απέχουν πολλές λέξεις μεταξύ τους. Η εισαγωγή των μοντέλων αυτών έχει ανεβάσει τον πήχη της απόδοσης σε ποικίλες εργασίες: από την ανάλυση συναισθήματος με λεπτές διαβαθμίσεις έως τη δημιουργία περιεχομένου (όπως αυτόματες περιλήψεις ή chatbot απαντήσεις) σε πραγματικό χρόνο. Σε έρευνες έχει φανεί ότι τα μεγάλα γλωσσικά μοντέλα μπορούν να γενικεύσουν γνώσεις από τεράστια corpora (π.χ. όλο το διαδίκτυο) και να απαντήσουν με ανθρώπινη-φαινόμενη συνοχή σε σύνθετα ερωτήματα. Για τα social media, αυτό ανοίγει δρόμους τόσο για την προηγμένη κατανόηση των συζητήσεων (π.χ. ανάλυση συναισθηματικού τόνου με βάση συμφραζόμενα) όσο και για νέες εφαρμογές όπως η αυτόματη παραγωγή περιεχομένου ή η εξατομικευμένη αλληλεπίδραση με χρήστες. Ωστόσο, παρά τα οφέλη, αυτοί οι αλγόριθμοι νέας γενιάς φέρνουν και νέες προκλήσεις: είναι υπολογιστικά απαιτητικοί, δυσκολότερο να ερμηνευθούν και παρουσιάζουν ενίοτε ανεπιθύμητες συμπεριφορές (π.χ. αναπαραγωγή κοινωνικών προκαταλήψεων που υπάρχουν στα δεδομένα εκπαίδευσής τους). Εντούτοις, αποτελούν κορυφαία κατεύθυνση της έρευνας και της πρακτικής – η ενσωμάτωσή τους στην ανάλυση ΜΚΔ βρίσκεται ήδη σε εξέλιξη, με στόχο την επίτευξη πιο έξυπνων και ευέλικτων συστημάτων (Ouyang et al., 2022)

Η σύζευξη των κοινωνικών δικτύων με τεχνολογίες Εικονικής και Επαυξημένης Πραγματικότητας (VR/AR) διαμορφώνει το όραμα του Metaverse, ενός επίμονου ψηφιακού σύμπαντος όπου οι χρήστες αλληλεπιδρούν μέσω avatars σε καθηλωτικά περιβάλλοντα. Στο πλαίσιο αυτό, έχει εισαχθεί ο όρος Augmented Social Media, που αναφέρεται σε εμπλουτισμένες κοινωνικές εμπειρίες με τη βοήθεια VR/AR: η παραδοσιακή δισδιάστατη online επικοινωνία μετατρέπεται σε μια βυθιστική εμπειρία 3D, με συνεχή, σε πραγματικό χρόνο αλληλεπίδραση σε εικονικούς κόσμους. Ήδη έχουν εμφανιστεί πλατφόρμες που υλοποιούν αυτές τις ιδέες – από το VRChat και το Horizon Worlds της Meta, μέχρι αποικεντρωμένα εικονικά περιβάλλοντα όπως το Decentraland – αξιοποιώντας προηγμένες μηχανές γραφικών και υποδομές cloud για να φιλοξενήσουν χιλιάδες σύγχρονους χρήστες. Η Μηχανική Μάθηση διαδραματίζει κρίσιμο ρόλο σε αυτό το νέο πλαίσιο: από τους αλγορίθμους γραφικών που δημιουργούν ρεαλιστικούς τρισδιάστατους κόσμους, μέχρι τα συστήματα που διαχειρίζονται δυναμικά τη ροή πληροφοριών και αλληλεπίδρασης μεταξύ των χρηστών. Επιπλέον, νέες προκλήσεις εγείρονται σχετικά με την ανάλυση κοινωνικών δεδομένων σε μικτά περιβάλλοντα – π.χ. η καταγραφή και ερμηνεία της συμπεριφοράς ενός avatar σε ένα εικονικό meeting ή η αξιοποίηση δεδομένων επαυξημένης πραγματικότητας (όπου το ψηφιακό περιεχόμενο υπερτίθεται στον φυσικό κόσμο). Στόχος των ερευνών είναι η δημιουργία έξυπνων πράκτορων και εργαλείων που θα μπορούν να βοηθούν τους χρήστες σε αυτές τις περιβάλλοντες, παρέχοντας πραγματικού χρόνου ανάλυση και υποστήριξη μέσα στο μετασύμπαν.

Μια αξιοσημείωτη τάση στο πλαίσιο του metaverse και των εξελιγμένων κοινωνικών δικτύων είναι η εμφάνιση εικονικών παραγόντων – από έξυπνους βοηθούς έως ψηφιακές περσόνες που δρουν ως αυτόνομα κοινωνικά υποκείμενα. Για παράδειγμα, agents ευαίσθητοι στο συναισθηματικό έχουν αρχίσει να δοκιμάζονται σε πλατφόρμες κοινωνικής VR: τέτοιοι πράκτορες μπορούν να αναλύουν τον τόνο της φωνής ή το κείμενο μιας συνομιλίας και να

ανιχνεύουν το συναίσθημα του χρήστη, προσαρμόζοντας ανάλογα τη συμπεριφορά τους (τεχνικές του πεδίου Affective Computing). Ένας εικονικός βοηθός σε ένα VR περιβάλλον μπορεί έτσι να αντιληφθεί ότι ο συνομιλητής του είναι θυμωμένος ή στενοχωρημένος και να μεταβάλει τον τρόπο αλληλεπίδρασης (π.χ. πιο συμπαθητικές απαντήσεις), βελτιώνοντας την εμπειρία χρήστη. Παράλληλα, στα ίδια τα παραδοσιακά ΜΚΔ έχει αναδειχθεί το φαινόμενο των virtual influencers – ψηφιακών «διασήμων» που εμφανίζονται ως ανθρωπόμορφοι χαρακτήρες δημιουργημένοι με γραφικά και Τεχνητή Νοημοσύνη. Περιπτώσεις όπως η Lil Miquela στο Instagram, ένα εικονικό μοντέλο με εκατομμύρια ακολούθους, καταδεικνύουν πως μια ψηφιακή περσόνα μπορεί να επηρεάζει τάσεις μόδας και να διαφημίζει προϊόντα χωρίς να υπάρχει φυσικά. Η ύπαρξη τέτοιων influencers κατέστη δυνατή μέσω συνδυασμού CGI τεχνολογίας και αλγορίθμων ML που καθοδηγούν την «προσωπικότητά» τους και το περιεχόμενο που δημοσιεύουν. Μελέτες μάλιστα αρχίζουν να εξετάζουν την αποτελεσματικότητα των εικονικών influencers στο μάρκετινγκ – π.χ. πώς η αντιλαμβανόμενη “πρωτοτυπία” ή ανθρωπομορφία τους επηρεάζει τη δέσμευση των καταναλωτών (Gulan et al., 2025). Οι εξελίξεις αυτές θολώνουν τα όρια μεταξύ πραγματικών και τεχνητών κοινωνικών οντοτήτων, δημιουργώντας νέα ερωτήματα σχετικά με την αυθεντικότητα, την εμπιστοσύνη του κοινού και τα ηθικά όρια στη χρήση τέτοιων τεχνικών.

Συμπερασματικά, η ανάλυση των μέσων κοινωνικής δικτύωσης με αλγορίθμους μηχανικής μάθησης αποτελεί ένα πολυδιάστατο και συνεχώς εξελισσόμενο πεδίο. Οι τρέχουσες τεχνικές έχουν ήδη καταδείξει τη δύναμή τους στο να αποκαλύπτουν γνώση από μεγάλα και πολύπλοκα δεδομένα – από την εξαγωγή συναισθήματος και γνώμης, μέχρι την ανίχνευση κοινοτήτων και επιδραστικών προσώπων. Ταυτόχρονα, αντιμετωπίζουν προκλήσεις που σχετίζονται με την ποιότητα των δεδομένων, το δυναμικό περιβάλλον και τις κοινωνικές προεκτάσεις της τεχνολογίας. Η ενσωμάτωση νέων τεχνολογιών (Big Data υποδομές, μετα-ανάλυση, VR/AR) και νέων αλγορίθμων (LLMs, transformers) διαμορφώνει το μέλλον της περιοχής, προσφέροντας ευκαιρίες για ακόμη βαθύτερη κατανόηση της διαδικτυακής κοινωνικής συμπεριφοράς. Η διεπιστημονική προσέγγιση – συνδυάζοντας την υπολογιστική δύναμη με την θεωρητική κατανόηση της ανθρώπινης συμπεριφοράς και τη θεσμική μέριμνα για ηθική – θα είναι καθοριστική για να αξιοποιηθεί πλήρως το δυναμικό της μηχανικής μάθησης στα social media την επόμενη δεκαετία.

## 9.2 Απαντήσεις στα Ερευνητικά Ερωτήματα

Η παρούσα ενότητα αποσκοπεί στην αναλυτική απάντηση των ερευνητικών ερωτημάτων που τέθηκαν στο Κεφάλαιο 1.2 της εργασίας, τα οποία διαμόρφωσαν το θεωρητικό και μεθοδολογικό πλαίσιο της μελέτης. Τα ερωτήματα αυτά δεν διατυπώθηκαν απλώς ως στόχοι κατανόησης της σχέσης μεταξύ μηχανικής μάθησης και κοινωνικών δικτύων, αλλά και ως μοχλοί διερεύνησης για την αποσαφήνιση των αλγοριθμικών τεχνικών, των εφαρμογών, των προκλήσεων και των μελλοντικών προοπτικών ενός ραγδαία εξελισσόμενου πεδίου. Μέσα από τη συνδυασμένη ανάλυση θεωρητικών πηγών, τεχνικών μοντέλων και εμπειρικών παραδειγμάτων, επιχειρείται η τεκμηριωμένη και συγκροτημένη απάντηση στα τέσσερα βασικά ερωτήματα της έρευνας. Οι απαντήσεις που ακολουθούν αντικατοπτρίζουν τόσο τα συμπεράσματα που αντλήθηκαν από τη βιβλιογραφική επισκόπηση όσο και τα ευρήματα των σύγχρονων εξελίξεων στην τεχνητή νοημοσύνη, υπογραμμίζοντας την πολυπλοκότητα, τη δυναμική και τη διεπιστημονικότητα του πεδίου.

### 9.2.1 Ερώτημα 1: Κύριοι Αλγόριθμοι Μηχανικής Μάθησης και Χαρακτηριστικά

Οι κυριότεροι αλγόριθμοι μηχανικής μάθησης που αξιοποιούνται στην ανάλυση δεδομένων κοινωνικών δικτύων διακρίνονται σε επιβλεπόμενους, μη επιβλεπόμενους, ενισχυτικής μάθησης και μεθόδους βαθιάς μάθησης. Στους επιβλεπόμενους αλγόριθμους περιλαμβάνονται τεχνικές ταξινόμησης και παλινδρόμησης, όπως η λογιστική παλινδρόμηση, τα δέντρα αποφάσεων, οι τυχαία δάση (Random Forests) και οι μηχανές διανυσματικής υποστήριξης (SVM). Αυτά τα μοντέλα μαθαίνουν από δεδομένα με ετικέτες, παρουσιάζουν υψηλή ακρίβεια σε προβλήματα όπως η ανάλυση συναισθήματος ή η ανίχνευση ανεπιθύμητων μηνυμάτων, όμως ενδέχεται να απαιτούν μεγάλο όγκο ποιοτικών δεδομένων και προσεκτικό ρυθμισμό υπερπαραμέτρων (π.χ. το SVM). Από την άλλη πλευρά, οι μη επιβλεπόμενες μέθοδοι (όπως η ομαδοποίηση k-means ή DBSCAN) χρησιμοποιούνται για την ανίχνευση προτύπων χωρίς προκαθορισμένες κατηγορίες, π.χ. για τη θεματική ομαδοποίηση αναρτήσεων ή τον εντοπισμό κοινοτήτων, με το πλεονέκτημα ότι αποκαλύπτουν κρυφές δομές στα δεδομένα αλλά και με τον περιορισμό ότι η αξιολόγηση των αποτελεσμάτων τους είναι δύσκολη ελλείψει ground truth. Επιπλέον, η ενισχυτική μάθηση εφαρμόζεται σε μικρότερο βαθμό στην ανάλυση κοινωνικών μέσων – ένας αυτόνομος πράκτορας βελτιώνει τις αποφάσεις του μέσω αλληλεπίδρασης με το περιβάλλον και ανταμοιβές, κάτι που ενδείκνυται κυρίως σε δυναμικές εφαρμογές (π.χ. βελτιστοποίηση ροής περιεχομένου) αλλά είναι περίπλοκο να υλοποιηθεί σε πραγματικές πλατφόρμες κοινωνικής δικτύωσης. Συνολικά, η κατηγοριοποίηση και συγκριτική μελέτη των παραπάνω αλγορίθμων είναι απαραίτητη για την κατανόηση της αποδοτικότητας και αξιοπιστίας τους σε διαφορετικά σενάρια κοινωνικών δεδομένων (Mitchell, 1997).

Κάθε κατηγορία αλγορίθμου φέρει συγκεκριμένα χαρακτηριστικά, πλεονεκτήματα και περιορισμούς. Απλά γραμμικά μοντέλα όπως η λογιστική παλινδρόμηση διακρίνονται για την ευκολία ερμηνείας τους και την ταχύτητα εκπαίδευσης, ωστόσο συχνά αδυνατούν να

μοντελοποιήσουν πολύπλοκες μη γραμμικές σχέσεις. Αντίθετα, πιο σύνθετοι αλγόριθμοι ταξινόμησης όπως οι SVM και τα δέντρα αποφάσεων μπορούν να συλλάβουν μη γραμμικά πρότυπα: οι SVM είναι αποδοτικοί σε υψηλής διάστασης δεδομένα (π.χ. χαρακτηριστικά κειμένου) και παρουσιάζουν ισχυρές θεωρητικές εγγυήσεις γενίκευσης, όμως η απόδοσή τους εξαρτάται από την επιλογή του κατάλληλου kernel και δεν προσφέρουν εύκολη ερμηνεία των αποφάσεών τους (black-box). Τα δέντρα αποφάσεων, αντιθέτως, είναι διαφανή και ερμηνεύσιμα, επιτρέποντας την εξαγωγή σαφών κανόνων απόφασης, αλλά τείνουν σε υπερπροσαρμογή (overfitting) σε θορυβώδη δεδομένα. Η συγκέντρωση πολλαπλών δέντρων σε συστήματα ensemble όπως τα *Random Forests* υπερνικά αυτό το μειονέκτημα: συνδυάζοντας πολλές εξασθενημένες προβλέψεις, παράγουν ένα πιο ακριβές και σταθερό μοντέλο, αξιοποιώντας τυχαιοποίηση στην επιλογή δεδομένων και χαρακτηριστικών ώστε να βελτιώσουν τη γενίκευση (Breiman, 2001). Τα ensemble μοντέλα συχνά επιτυγχάνουν κορυφαία ακρίβεια μεταξύ των κλασικών μεθόδων και προσφέρουν ενσωματωμένες εκτιμήσεις σφάλματος και σημασίας μεταβλητών, όμως υστερούν ως προς την αναγνωσιμότητα του συνολικού μοντέλου και απαιτούν μεγαλύτερη υπολογιστική ισχύ για την εκπαίδευση και αποθήκευση πολλαπλών δέντρων. Απλοί ταξινομητές όπως ο *k-Nearest Neighbors* (*k*-NN) χρησιμοποιούνται επίσης για βασικές ταξινομήσεις ή συστάσεις σε κοινωνικά δεδομένα λόγω απλότητας, δεδομένου ότι δεν απαιτούν εκτενή φάση εκπαίδευσης, όμως το κόστος πρόβλεψης τους είναι υψηλό (επειδή υπολογίζουν αποστάσεις από όλα τα δείγματα) και η απόδοσή τους επιδεινώνεται σε μεγάλους όγκους δεδομένων ή όταν τα χαρακτηριστικά δεν είναι ομοιόμορφα κλιμακοποιημένα.

Στο πεδίο των κοινωνικών δικτύων, οι αλγόριθμοι βαθιάς μάθησης έχουν αποκτήσει εξέχοντα ρόλο, ειδικά με την άνοδο των Τεχνητών Νευρωνικών Δικτύων (TNΔ) και των συναφών αρχιτεκτονικών. Τα πολυεπίπεδα νευρωνικά δίκτυα (MLP) και οι σύγχρονες παραλλαγές τους (π.χ. συνελικτικά δίκτυα CNN για εικόνες, επαναλαμβανόμενα δίκτυα RNN/LSTM για σειρές κειμένου) έχουν την ικανότητα να μαθαίνουν σύνθετες συναρτήσεις απόδοσης απευθείας από τα δεδομένα, αυτόματα εξάγοντας χαρακτηριστικά υψηλού επιπέδου από αδόμητο περιεχόμενο (κειμένα, εικόνες, βίντεο). Αυτό τους επιτρέπει να επιτυγχάνουν αποτελέσματα αιχμής σε πλήθος εφαρμογών κοινωνικών μέσων, όπως η ανάλυση συναισθήματος με συνελικτικά/μετασχηματιστικά μοντέλα ή η αναγνώριση προσώπων σε πολυμέσα χρηστών, υπερβαίνοντας κατά πολύ την ακρίβεια των παραδοσιακών μεθόδων όταν υπάρχουν διαθέσιμα μεγάλα σύνολα δεδομένων (Goodfellow et al., 2016). Βασικά χαρακτηριστικά των βαθιών νευρωνικών μοντέλων είναι η παραλληλοποίηση και η ικανότητα προσαρμογής σε μη γραμμικά πρότυπα, ωστόσο συνοδεύονται και από σημαντικούς περιορισμούς. Απαιτούν τεράστιους όγκους εκπαιδευτικών δεδομένων και υπολογιστικών πόρων (GPU/TPU) για να γενικεύσουν αποτελεσματικά, γεγονός που ενίοτε καθιστά δύσκολη την εφαρμογή τους σε πραγματικό χρόνο ή σε εξειδικευμένες κοινότητες όπου τα δεδομένα είναι περιορισμένα. Επιπλέον, συχνά λειτουργούν ως «μαύρα κουτιά», με περιορισμένη ερμηνευσιμότητα: είναι δύσκολο να εξηγηθεί γιατί ένα βαθύ δίκτυο έλαβε μια συγκεκριμένη απόφαση, κάτι που προβληματίζει όταν απαιτείται διαφάνεια ή εντοπισμός

μεροληψίας στα κοινωνικά δεδομένα. Συμπερασματικά, κάθε αλγόριθμος μηχανικής μάθησης φέρει έναν συνδυασμό ιδιοτήτων που τον καθιστά κατάλληλο για συγκεκριμένες αναλυτικές εργασίες στα μέσα κοινωνικής δικτύωσης, ενώ η επιλογή του οφείλει να σταθμίζει τα πλεονεκτήματα (π.χ. ακρίβεια, ταχύτητα, δυνατότητα κλιμάκωσης) έναντι των περιορισμών (π.χ. απαίτηση δεδομένων, πολυπλοκότητα, έλλειψη διαφάνειας).

### 9.2.2 Ερώτημα 2: Κυριότερες Εφαρμογές των Αλγορίθμων στα Κοινωνικά Δίκτυα και Επιλυόμενα Προβλήματα

Η μηχανική μάθηση έχει βρει πληθώρα εφαρμογών στα κοινωνικά δίκτυα, αντιμετωπίζοντας βασικές προκλήσεις και ανάγκες που προκύπτουν από τον τεράστιο όγκο και τη δυναμική των δεδομένων των χρηστών. Ανάμεσα στις σημαντικότερες εφαρμογές διακρίνονται: **(α)** η ανάλυση συναισθήματος, **(β)** η πρόβλεψη συμπεριφοράς χρηστών, **(γ)** η ανίχνευση παραπληροφόρησης και επιβλαβούς περιεχομένου και **(δ)** η θεματική κατηγοριοποίηση περιεχομένου και κοινοτήτων. Κάθε μία από αυτές στοχεύει στην επίλυση διαφορετικών προβλημάτων και αξιοποιεί ειδικές προσεγγίσεις αλγορίθμων.

**Ανάλυση Συναισθήματος (Sentiment Analysis):** Αποτελεί μία από τις δημοφιλέστερες εφαρμογές, όπου μοντέλα μηχανικής μάθησης χρησιμοποιούνται για την αυτόματη κατάταξη του συναισθήματος που εκφράζεται σε αναρτήσεις, σχόλια, κριτικές ή tweets. Με τεχνικές επεξεργασίας φυσικής γλώσσας, όπως ταξινομητές SVM, Naïve Bayes ή βαθιά νευρωνικά δίκτυα, οι αναρτήσεις χαρακτηρίζονται ως θετικές, αρνητικές ή ουδέτερες. Το κύριο πρόβλημα που επιλύεται εδώ είναι η κατανόηση της στάσης και των απόψεων του κοινού σε μεγάλη κλίμακα: οι επιχειρήσεις και οργανισμοί αντλούν πολύτιμη πληροφορία για τη φήμη τους, την ικανοποίηση πελατών ή την αντίδραση σε γεγονότα, χωρίς να απαιτείται χειρωνακτική ανάγνωση χιλιάδων μηνυμάτων. Μελέτες έχουν δείξει ότι η ακρίβεια της ανάλυσης συναισθήματος μπορεί να ενισχυθεί σημαντικά με εξειδικευμένα λεξικά συναισθήματος, χαρακτηριστικά n-grams και σύγχρονες αρχιτεκτονικές deep learning, καθιστώντας την ένα εργαλείο-κλειδί για το στρατηγικό μάρκετινγκ και τη διαχείριση δημοσίων σχέσεων στα κοινωνικά μέσα (Pang & Lee, 2008). Σήμερα η ανάλυση συναισθήματος εφαρμόζεται εκτενώς για την παρακολούθηση αντιδράσεων σε λανσαρίσματα προϊόντων, πολιτικές καμπάνιες και κοινωνικά γεγονότα, επιτρέποντας σε εταιρείες και φορείς να αντιληφθούν έγκαιρα το κοινό συναίσθημα και να προσαρμόσουν ανάλογα τις αποφάσεις τους.

**Πρόβλεψη Συμπεριφοράς Χρηστών:** Ένα ακόμη κρίσιμο πεδίο εφαρμογής είναι η προσπάθεια πρόβλεψης των μελλοντικών ενεργειών και προτιμήσεων των χρηστών στα δίκτυα. Μέσω προγνωστικών μοντέλων (π.χ. ταξινομητές ή συστήματα συστάσεων που εκπαιδεύονται σε ιστορικά δεδομένα αλληλεπίδρασης), επιχειρείται να εκτιμηθεί η πιθανότητα ο χρήστης να προβεί σε συγκεκριμένες ενέργειες – όπως το να κάνει κλικ σε ένα

σύνδεσμο, να αλληλεπιδράσει με ένα περιεχόμενο (like/share) ή να εγκαταλείψει την πλατφόρμα (churn). Αυτή η εφαρμογή στοχεύει στην αντιμετώπιση του ζητήματος διατήρησης και εμπλοκής του κοινού: οι πλατφόρμες και οι διαχειριστές κοινοτήτων μπορούν να εντοπίσουν έγκαιρα χρήστες που παρουσιάζουν σημάδια μείωσης δραστηριότητας και να λάβουν μέτρα (π.χ. προσωποποιημένες ειδοποιήσεις ή προσφορές) ώστε να προλάβουν την αποχώρησή τους. Επιπλέον, η κατανόηση πρότυπων συμπεριφοράς επιτρέπει τον στοχευμένο σχεδιασμό περιεχομένου: οι αλγόριθμοι συστάσεων (όπως το collaborative filtering και τα μοντέλα βαθιάς μάθησης για προτάσεις) προβλέπουν ποιο περιεχόμενο θα ενδιαφέρει κάθε χρήστη, βελτιστοποιώντας την ροή ειδήσεων (news feed) ή τις προτάσεις φίλων/ομάδων. Με τον τρόπο αυτό, η μηχανική μάθηση συνεισφέρει στην εξατομίκευση της εμπειρίας του χρήστη, αντιμετωπίζοντας το πρόβλημα του information overload μέσω προσωποποιημένης παροχής πληροφορίας (Lin, 2016).

**Ανίχνευση Παραπληροφόρησης και Επιβλαβούς Περιεχομένου:** Η ραγδαία διάδοση των κοινωνικών μέσων έχει καταστήσει επιτακτική την ανάπτυξη αλγορίθμων που εντοπίζουν ψευδείς ειδήσεις, παραπλανητικές πληροφορίες, προπαγάνδα ή ρητορική μίσους μέσα στα διαδικτυακά δίκτυα. Τα προβλήματα αυτά έχουν ιδιαίτερα σοβαρές κοινωνικές και πολιτικές προεκτάσεις – η παραπληροφόρηση μπορεί να επηρεάσει εκλογικές διαδικασίες, τη δημόσια υγεία (π.χ. μύθοι για εμβόλια) ή να προκαλέσει πανικό και βία. Οι αλγόριθμοι μηχανικής μάθησης προσφέρουν λύσεις φιλτράροντας και ανιχνεύοντας δυνητικά ψευδές περιεχόμενο: μοντέλα ταξινόμησης (συχνά *συνδυάζοντας χαρακτηριστικά γλώσσας και δικτύου*) μαθαίνουν να διακρίνουν αξιόπιστες από αναξιόπιστες ειδήσεις, ενώ ταυτόχρονα ανιχνεύονται ύποπτοι λογαριασμοί (π.χ. *bots*) που διασπείρουν μαζικά παραπληροφόρηση. Βασική πρόκληση στην εφαρμογή αυτή είναι η χαρακτηριστική πολυπλοκότητα του ψευδούς περιεχομένου: οι δημιουργοί παραπληροφόρησης εξελίσσουν συνεχώς τις πρακτικές τους, επομένως και τα μοντέλα πρέπει να προσαρμόζονται δυναμικά για να παραμένουν αποτελεσματικά. Σύγχρονες προσεγγίσεις συνδυάζουν βαθιά μάθηση με ανάλυση κοινωνικών γραφημάτων – π.χ. γραφικές νευρωνικές δομές που ανιχνεύουν ασυνήθιστα πρότυπα διαμοιρασμού μιας φήμης. Έρευνες δείχνουν ότι ο συνδυασμός σημασιολογικής ανάλυσης κειμένου και μεταδεδομένων δικτύωσης βελτιώνει σημαντικά την ακρίβεια στον εντοπισμό ψευδών ειδήσεων (Shu et al., 2017). Συνολικά, η εφαρμογή αυτή επιχειρεί να μετριάσει το πρόβλημα της αξιοπιστίας της πληροφορίας στα κοινωνικά δίκτυα, παρέχοντας εργαλεία για ασφαλέστερο περιβάλλον ενημέρωσης των χρηστών.

**Θεματική Κατηγοριοποίηση Περιεχομένου και Ανάλυση Κοινοτήτων:** Πέραν των ανωτέρω, σημαντική είναι και η χρήση αλγορίθμων για την ομαδοποίηση και κατανόηση των θεμάτων συζήτησης καθώς και των κοινωνικών δομών στα δίκτυα. Οι τεχνικές Topic Modeling (π.χ. LDA – Latent Dirichlet Allocation) και γενικότερα η θεματική κατηγοριοποίηση εφαρμόζονται σε μεγάλα σύνολα αναρτήσεων ή hashtags, με στόχο να αποκαλύψουν τις κύριες τάσεις και θεματολογίες που απασχολούν το κοινό σε δεδομένη περίοδο. Με αυτόν τον τρόπο αντιμετωπίζεται το πρόβλημα της αδόμητης πληροφορίας:

από ένα χάος καθημερινών αναρτήσεων, εξάγονται συνοπτικά τα επικρατέστερα θέματα (π.χ. μια φυσική καταστροφή, μια δημοφιλής εκδήλωση ή ένα viral meme) που αναδύονται μέσα από τα δεδομένα. Παράλληλα, οι αλγόριθμοι ανάλυσης κοινοτήτων στοχεύουν στον εντοπισμό ομάδων χρηστών με έντονη μεταξύ τους αλληλεπίδραση ή κοινά ενδιαφέροντα. Τεχνικές ομαδοποίησης γραφών (graph clustering) και ανίχνευσης κοινοτήτων στο κοινωνικό γράφημα ενός δικτύου (π.χ. ο αλγόριθμος Louvain ή μέθοδοι εύρεσης υπογραφημάτων) βοηθούν στην αποκάλυψη υπο-δικτύων μέσα στο μεγάλο δίκτυο χρηστών. Αυτές οι κοινότητες μπορεί να αντιστοιχούν σε δημογραφικές ομάδες, φαν κλαμπ, επαγγελματικά δίκτυα ή οποιεσδήποτε συγκροτημένες ομάδες με κοινή θεματολογία. Η χαρτογράφηση τους λύνει το πρόβλημα της κατανόησης της κοινωνικής δομής ενός δικτύου: π.χ. μια εταιρεία μπορεί να διαπιστώσει ότι οι υποστηρικτές ενός προϊόντος συγκροτούν μια συγκεκριμένη κοινότητα στην οποία μπορεί να στοχεύσει ξεχωριστά, ή οι υπεύθυνοι δημόσιας υγείας να εντοπίσουν «θύλακες» χρηστών όπου διαδίδονται ταχύτατα αντιεπιστημονικές αντιλήψεις. Επιπλέον, η κοινωνική δικτύωση και ανάλυση σχέσεων (Social Network Analysis) μετρά δείκτες όπως ο βαθμός κεντρικότητας, η επιρροή συγκεκριμένων κόμβων (influencers) και η ροή πληροφορίας, επιλύοντας πρακτικά προβλήματα όπως η ταυτοποίηση ηγετών γνώμης (opinion leaders) ή η βελτιστοποίηση της διάχυσης περιεχομένου μέσα από το δίκτυο (Weller et al., 2013). Συνολικά, οι εφαρμογές αυτές αναδεικνύουν τον ρόλο της μηχανικής μάθησης ως εργαλείου εξόρυξης γνώσης από τα κοινωνικά δεδομένα: παρέχουν στις επιχειρήσεις, τους ερευνητές και τους φορείς τη δυνατότητα να λαμβάνουν τεκμηριωμένες αποφάσεις, να προσαρμόζουν στρατηγικές μάρκετινγκ ή πολιτικής επικοινωνίας και να κατανοούν βαθύτερα τη δυναμική των online κοινοτήτων.

### 9.2.3 Ερώτημα 3: Τεχνικά, Ηθικά και Κοινωνικά Ζητήματα από την Εφαρμογή Μηχανικής Μάθησης

Η αξιοποίηση αλγορίθμων μηχανικής μάθησης σε δεδομένα μέσων κοινωνικής δικτύωσης εγείρει μια σειρά από τεχνικές προκλήσεις, ηθικά διλήμματα και κοινωνικές επιπτώσεις που πρέπει να αντιμετωπιστούν. Πρώτον, σε τεχνικό επίπεδο, εμφανίζονται ζητήματα σχετιζόμενα με την πολυπλοκότητα των μοντέλων και των δεδομένων. Οι κοινωνικές πλατφόρμες παράγουν *Big Data* με υψηλό ρυθμό (streams πραγματικού χρόνου), μεγάλο όγκο και ποικιλία (κειμενικά μηνύματα, εικόνες, γραφήματα σχέσεων), γεγονός που απαιτεί επεκτάσιμες υποδομές και αλγορίθμους ικανούς να κλιμακώνονται. Η αποδοτικότητα και επεκτασιμότητα των μοντέλων αποτελεί συνεπώς βασική πρόκληση: αλγόριθμοι που αποδίδουν καλά σε μικρά δείγματα μπορεί να καθίστανται αργοί ή ανεπαρκείς σε δεδομένα εκατομμυρίων χρηστών. Επιπλέον, η ποιότητα και προικατάληψη των δεδομένων είναι ένα κρίσιμο τεχνικό ζήτημα. Τα κοινωνικά δεδομένα συχνά περιέχουν θόρυβο, ανακρίβειες ή μη αντιπροσωπευτικά δείγματα (π.χ. υπερεκπροσώπηση ορισμένων ομάδων χρηστών), οδηγώντας σε μεροληπτικά μοντέλα. Έτσι, οι πρακτικές προεπεξεργασίας, η επιλογή

χαρακτηριστικών και η συνεχής επανεικπαίδευση των αλγορίθμων καθίστανται απαραίτητες για τη βελτίωση της γενίκευσης. Ένα επιπλέον τεχνικό ζήτημα είναι η ερμηνευσιμότητα (interpretability) και διαφάνεια των πολύπλοκων μοντέλων: ιδιαίτερα στα βαθιά νευρωνικά δίκτυα και στα μεγάλα συστήματα συστάσεων, οι εσωτερικοί μηχανισμοί απόφασης είναι δυσερμήνευτοι, γεγονός που δυσχεραίνει τον εντοπισμό σφαλμάτων ή μεροληψίας. Η πρόκληση αυτή έχει οδηγήσει στην ανάπτυξη πεδίων όπως το Explainable AI (εξηγήσιμη ΤΝ), προκειμένου να δημιουργηθούν εργαλεία που φωτίζουν τα “μαύρα κουτιά” των αλγορίθμων και καθιστούν τις προβλέψεις τους περισσότερο κατανοητές (Doshi-Velez & Kim, 2017).

Σε ηθικό επίπεδο, το κυριότερο ζήτημα είναι η προστασία της ιδιωτικότητας και των προσωπικών δεδομένων των χρηστών (privacy). Τα κοινωνικά δίκτυα εμπεριέχουν πληθώρα ευαίσθητων πληροφοριών – από προσωπικές απόψεις και συναισθήματα μέχρι τοποθεσίες και κοινωνικές σχέσεις. Η αθροιστική συλλογή και ανάλυση τέτοιων δεδομένων από αλγορίθμους μπορεί να παραβιάσει την ιδιωτικότητα, εάν δεν τηρούνται αυστηρές προδιαγραφές συγκατάθεσης και ανωνυμοποίησης. Ένα σαφές παράδειγμα είναι η εξακρίβωση προτιμήσεων ή χαρακτηριστικών χρηστών (όπως πολιτικών πεποιθήσεων ή σεξουαλικού προσανατολισμού) μέσω ανάλυσης των «likes» και των κοινωνικών τους συνδέσεων – ακόμη κι όταν οι ίδιοι δεν έχουν δημοσιοποιήσει ρητά αυτές τις πληροφορίες. Τέτοιες πρακτικές εγείρουν ηθικά ζητήματα συναίνεσης και χειραγώγησης: ο χρήστης ενδέχεται να μην αντιλαμβάνεται ότι τα δεδομένα του χρησιμοποιούνται για εμπορική στόχευση ή άλλες αναλύσεις, στερούμενος τον έλεγχο επί της ψηφιακής του ταυτότητας (Binns, 2018). Παράλληλα, η συγκέντρωση τεράστιων δεδοτοσυνόλων από εταιρείες κοινωνικών μέσων δημιουργεί κινδύνους ασφαλείας – παραβιάσεις δεδομένων (data breaches) θα εξέθεταν προσωπικά στοιχεία εκατομμυρίων πολιτών. Για τους λόγους αυτούς, τίθεται επιτακτικά η ανάγκη ενσωμάτωσης αρχών δεοντολογίας και συμμόρφωσης με κανονιστικά πλαίσια (όπως ο GDPR στην Ευρώπη) στην ανάπτυξη αλγορίθμων, ώστε να διασφαλιστεί ότι η χρήση των δεδομένων γίνεται με τρόπο δίκαιο και διαφανή.

Σε κοινωνικό επίπεδο, ένα μείζον ζήτημα είναι η αλγοριθμική μεροληψία (algorithmic bias) και οι διακρίσεις που μπορεί να προκύψουν αθέλητα από τα μοντέλα. Δεδομένου ότι οι αλγόριθμοι εκπαιδεύονται πάνω σε ιστορικά δεδομένα, υφιστάμενες κοινωνικές προκαταλήψεις (biases) ενδέχεται να ενσωματωθούν και να αναπαραχθούν από το μοντέλο. Για παράδειγμα, ένας ταξινομητής που μαθαίνει να ανιχνεύει ρητορική μίσους μπορεί να παρουσιάζει χαμηλότερη ακρίβεια σε ιδιώματα ή γλώσσες μειονοτήτων, αποκλείοντας δυσανάλογα τις αναρτήσεις τους ως ύποπτες. Ένα σύστημα σύστασης περιεχομένου μπορεί να προβάλλει λιγότερο περιεχόμενο από υποεκπροσωπούμενες ομάδες, δικαιωνίζοντας έναν κοινωνικό αποκλεισμό στον ψηφιακό λόγο. Τέτοια φαινόμενα αλγοριθμικής μεροληψίας μπορεί να ενισχυθούν όταν λείπει η διαφάνεια: αν τα αποτελέσματα παρουσιάζονται ως ουδέτερα χωρίς οι χρήστες να γνωρίζουν τα κριτήρια, οι συνέπειες στη διαμόρφωση της κοινής γνώμης και στην κοινωνική δικαιοσύνη είναι ανησυχητικές. Εξάλλου, οι αλγόριθμοι

κατάταξης στα news feeds μπορούν να δημιουργήσουν φαινόμενα “φούσκας φίλτρων” (filter bubbles) – εξατομικεύοντας το περιεχόμενο τόσο, ώστε ο χρήστης να εκτίθεται μόνο σε απόψεις με τις οποίες ήδη συμφωνεί, όπερ εντείνει τον κοινωνικό κατακερματισμό και την πόλωση. Η αντιμετώπιση αυτών των ζητημάτων απαιτεί συνδυασμένη προσπάθεια: τεχνικές λύσεις (όπως η εισαγωγή δικλίδων fairness στα μοντέλα, η χρήση πιο αντιπροσωπευτικών δεδομένων εκπαίδευσης και ο τακτικός έλεγχος για bias) αλλά και θεσμικές παρεμβάσεις (κανονισμοί για διαφάνεια αλγορίθμων, εργαλεία λογοδοσίας και αλγοριθμικού ελέγχου) (Floridi et al., 2018). Τέλος, σε κοινωνικό και πολιτικό επίπεδο, η αυξανόμενη εξάρτηση από αυτοματοποιημένες αποφάσεις στα κοινωνικά δίκτυα γεννά ερωτήματα λογοδοσίας: ποιος φέρει την ευθύνη όταν ένας αλγόριθμος λανθασμένα αποκλείει περιεχόμενο ή παραβλέπει ένα κρίσιμο μοτίβο (false negative) που οδηγεί σε βλάβη; Τα ζητήματα αυτά υπογραμμίζουν την ανάγκη η μηχανική μάθηση να προσεγγίζεται με κριτική ματιά και υπευθυνότητα, διασφαλίζοντας ότι η τεχνολογία υπηρετεί το δημόσιο συμφέρον και τα δικαιώματα των χρηστών σε ένα δημοκρατικό ψηφιακό οικοσύστημα.

#### **9.2.4 Ερώτημα 4: Μελλοντικές Τάσεις και Ανοιχτά Ερευνητικά Ζητήματα στην Ανάλυση Κοινωνικών Μέσων**

Καθώς τα μέσα κοινωνικής δικτύωσης και η τεχνητή νοημοσύνη εξελίσσονται ραγδαία, προκύπτουν νέες τάσεις και ανοιχτά ερευνητικά πεδία που θα διαμορφώσουν την επόμενη δεκαετία στον χώρο της ανάλυσης κοινωνικών μέσων. Μία από τις πλέον σημαντικές τάσεις είναι η ενσωμάτωση των Μεγάλων Γλωσσικών Μοντέλων (Large Language Models, LLMs) και γενικότερα των προηγμένων μοντέλων βαθιάς μάθησης στην κατανόηση του περιεχομένου των κοινωνικών δικτύων. Η εμφάνιση μετασχηματιστικών μοντέλων όπως η οικογένεια GPT και BERT έχει ήδη αναδιαμορφώσει το πεδίο της επεξεργασίας φυσικής γλώσσας, επιτρέποντας την ανάλυση κειμένου με πρωτοφανή βάθος και ακρίβεια. Στα επόμενα χρόνια, αναμένεται ότι τέτοια προεκπαιδευμένα μοντέλα θα αξιοποιηθούν ευρέως για την αυτόματη κατανόηση των social media: από bots κατόπιν AI που θα μπορούν να αλληλεπιδρούν φυσικά με τους χρήστες και να μετριάσουν συζητήσεις, μέχρι εξελιγμένα συστήματα ανάλυσης συναισθήματος που θα ανιχνεύουν λεπτές αποχρώσεις (π.χ. σαρκασμό ή ψευδαισθηση). Παράλληλα, η πολυτροπική μάθηση (multimodal learning) αποτελεί ανερχόμενη ερευνητική κατεύθυνση: τα κοινωνικά δίκτυα συνδυάζουν κείμενο, εικόνες, βίντεο, ήχο και διασυνδέσεις χρηστών – η επόμενη γενιά αλγορίθμων στοχεύει σε ενοποιημένα μοντέλα που θα επεξεργάζονται ταυτόχρονα όλες αυτές τις πηγές δεδομένων. Τέτοια μοντέλα θα μπορούσαν, για παράδειγμα, να αναλύουν ένα *memetic* φαινόμενο συνδυάζοντας την εικόνα του meme, τη λεζάντα κειμένου του και το μοτίβο διάδοσής του στο δίκτυο. Επιπροσθέτως, η ενοποίηση των κοινωνικών μέσων με το περιβάλλον του Metaverse και οι τεχνολογίες εικονικής/επαυξημένης πραγματικότητας αναμένεται να ανοίξουν νέους δρόμους: οι κοινωνικές αλληλεπιδράσεις εντός εικονικών κόσμων θα παράγουν πολύπλοκα δεδομένα χωροχρονικά και ολογραφικά, δημιουργώντας ανάγκη για

νέους αλγορίθμους ανάλυσης που θα λαμβάνουν υπόψη το πλαίσιο της παρουσίας και της τρισδιάστατης αλληλεπίδρασης (Zhang & Lu, 2021). Συνολικά, η συνεχιζόμενη εξέλιξη της TN – από τα μεγάλα γλωσσικά και πολυτροπικά μοντέλα μέχρι τα εξειδικευμένα *graph neural networks* για κοινωνικά γράφηματα – προδιαγράφει ένα μέλλον όπου η ανάλυση κοινωνικών μέσων θα γίνεται πιο ολιστική, *real-time* και ενσωματωμένη σε διάφορες πτυχές της ψηφιακής ζωής.

Πλάι σε αυτές τις τεχνολογικές τάσεις, αναδύεται μια σειρά από ανοιχτά ερευνητικά ζητήματα που θα καθορίσουν την ατζέντα της επόμενης δεκαετίας. Ένα εξ αυτών είναι η βελτίωση της διαφάνειας και της αξιοπιστίας των μοντέλων μηχανικής μάθησης. Καθώς τα κοινωνικά δίκτυα διαδραματίζουν κρίσιμο ρόλο στην ενημέρωση και την κοινωνική αλληλεπίδραση, υπάρχει πιεστική ανάγκη για αλγορίθμους που θα είναι όχι μόνο ακριβείς αλλά και επεξηγήσιμοι, δίκαιοι και υπεύθυνοι. Η ανάπτυξη τεχνικών *Explainable AI* και μετρικών δικαιοσύνης (fairness metrics) θα συνεχίσει να αποτελεί προτεραιότητα, ώστε τα νέα μοντέλα (π.χ. τα LLMs που εφαρμόζονται στην επιμέλεια περιεχομένου) να συνοδεύονται από ερμηνεύσιμες εξόδους και εγγυήσεις ότι δεν ενισχύουν προκαταλήψεις ή παραπληροφόρηση. Επίσης, αναμένεται να δοθεί έμφαση στην προστασία ιδιωτικότητας μέσω τεχνικών μηχανικής μάθησης. Μέθοδοι όπως η ομοσπονδιακή μάθηση (federated learning) και η differential privacy ήδη ερευνώνται ως τρόποι να εκπαιδεύονται μοντέλα πάνω σε αποκεντρωμένα δεδομένα χρήστη (π.χ. στα smartphones) χωρίς να παραβιάζεται η ιδιωτικότητα. Τέτοιες προσεγγίσεις μπορεί να καταστούν στάνταρ στις κοινωνικές εφαρμογές, ελαχιστοποιώντας την ανάγκη συλλογής ακατέργαστων δεδομένων σε κεντρικούς διακομιστές. Επιπλέον, η διεπιστημονική προσέγγιση θα χαρακτηρίσει τις μελλοντικές έρευνες: η συνεργασία ειδικών πληροφορικής, κοινωνιολόγων, νομικών και ψυχολόγων θα είναι απαραίτητη για να αντιμετωπιστούν πολύπλευρα θέματα, όπως η ρύθμιση των αλγορίθμων επιμέλειας περιεχομένου, ο αντίκτυπός τους στην ψυχική υγεία των χρηστών και ο ρόλος τους στη διαμόρφωση της κοινής γνώμης (Dwivedi et al., 2021). Τέλος, μια σημαντική ανοιχτή πρόκληση είναι η επέκταση της έρευνας πέρα από τα αγγλόφωνα ή δυτικά δεδομένα: τα επόμενα χρόνια θα χρειαστούν μοντέλα που να υποστηρίζουν αποτελεσματικά πολλές γλώσσες και πολιτισμικά πλαίσια, ώστε η ανάλυση κοινωνικών μέσων να έχει οικουμενική εφαρμοσιμότητα και να μην αποκλείονται κοινότητες λόγω γλωσσικών φραγμών. Συμπερασματικά, ο τομέας της μηχανικής μάθησης για κοινωνικά δίκτυα βρίσκεται σε φάση ώριμης εξέλιξης: νέες τεχνολογίες δημιουργούν συναρπαστικές δυνατότητες, αλλά και η ανάγκη επίλυσης σύνθετων ζητημάτων (τεχνικών και ηθικών) γίνεται πιο έντονη, καθοδηγώντας μια ενεργή ερευνητική ατζέντα για την επόμενη δεκαετία.

### 9.3 Συμπερασματικά Σχόλια

Συνοψίζοντας, η παρούσα μελέτη ανέδειξε ότι οι σύγχρονοι αλγόριθμοι μηχανικής μάθησης αποτελούν βασικό εργαλείο για την ανάλυση των μεγάλων δεδομένων (big data) στα μέσα κοινωνικής δικτύωσης. Μέσα από τα προηγούμενα κεφάλαια παρουσιάστηκε πώς τεχνικές όπως η βαθιά μάθηση, η ανάλυση δικτύων και προηγμένα μοντέλα επεξεργασίας φυσικής γλώσσας μπορούν να εξάγουν γνώση από τον τεράστιο όγκο και την ποικιλία των δεδομένων χρηστών. Παραδείγματα εφαρμογών – από την ανάλυση συναισθήματος και την ανίχνευση κοινοτήτων μέχρι την πρόβλεψη τάσεων και τη διαχείριση κρίσεων – επιβεβαιώνουν την ευελιξία και την αξία της τεχνητής νοημοσύνης σε πραγματικά σενάρια κοινωνικών μέσων (Acikara et al., 2023). Συνολικά, οι απαντήσεις στα ερευνητικά ερωτήματα κατέδειξαν ότι ο κατάλληλος συνδυασμός αλγορίθμων και δεδομένων μπορεί να προσφέρει πολύτιμες πληροφορίες για την ανθρώπινη συμπεριφορά και τη διάδοση της πληροφορίας στα δίκτυα αυτά.

Διαπιστώθηκε, όμως, ότι οι ευκαιρίες αυτές συνοδεύονται από σημαντικές προκλήσεις τεχνικής και ηθικής φύσεως. Τα κοινωνικά δεδομένα είναι δυναμικά, πολύπλοκα και συχνά «θορυβώδη», γεγονός που δυσχεραίνει την επεξεργασία και ανάλυσή τους. Ζητήματα ιδιωτικότητας και προστασίας προσωπικών δεδομένων αναδείχθηκαν ως κρίσιμα, καθώς η εξόρυξη πληροφορίας από προσωπικές αναρτήσεις εγείρει ανησυχίες περί συγκατάθεσης και επιτήρησης. Επιπλέον, εντοπίστηκε ο κίνδυνος ενσωμάτωσης μεροληψίας (bias) στα δεδομένα και στα μοντέλα – για παράδειγμα, αλγόριθμοι που εκπαιδεύονται σε μονομερείς ή ανεπαρκείς πληροφορίες ενδέχεται να παράγουν στρεβλές προβλέψεις. Ένα επίσης κομβικό συμπέρασμα είναι ότι το περιβάλλον των μέσων κοινωνικής δικτύωσης μεταβάλλεται διαρκώς: τα μοτίβα αλληλεπίδρασης, η γλώσσα (π.χ. νέο αργκό ή memes) και οι τάσεις εξελίσσονται, επιβάλλοντας την ανάγκη για αλγορίθμους προσαρμοστικούς σε πραγματικό χρόνο, ώστε τα μοντέλα ανάλυσης να επικαιροποιούνται και να αναπροσαρμόζονται συνεχώς και να διατηρούν την αποτελεσματικότητά τους παρά τις αλλαγές στο περιβάλλον χρήσης.

Η σημασία και η συμβολή της εργασίας αυτής έγκειται στη συνολική θεώρηση και σύνθεση των παραπάνω θεμάτων, συμβάλλοντας στον ευρύτερο επιστημονικό και κοινωνικό διάλογο για τα big data και την ανάλυση κοινωνικών μέσων μέσω τεχνητής νοημοσύνης. Προσφέροντας μια σφαιρική επισκόπηση – από τους αλγορίθμους και τις εφαρμογές μέχρι τις ηθικές προεκτάσεις – η μελέτη υπογραμμίζει πόσο στενά συνυφασμένη είναι η τεχνική πρόοδος με τις κοινωνικές διαστάσεις της. Το πεδίο αυτό είναι εγγενώς διεπιστημονικό, συνδυάζοντας την επιστήμη υπολογιστών με την κοινωνιολογία, την επικοινωνία, τη νομική και άλλους κλάδους. Ως εκ τούτου, καθίσταται σαφές ότι η ανάπτυξη λύσεων ανάλυσης των μέσων κοινωνικής δικτύωσης απαιτεί διεπιστημονική συνεργασία και ενσωμάτωση ποικίλων οπτικών. Παράλληλα, η ανάγκη για ηθικά και υπεύθυνα συστήματα αναδείχθηκε ως θεμελιώδης προτεραιότητα (Dwivedi et al., 2021). Η διαφάνεια, η λογοδοσία και ο σεβασμός της ιδιωτικότητας πρέπει να αποτελούν πυλώνες σχεδιασμού των αλγορίθμων, ώστε η αξιοποίηση της τεχνητής νοημοσύνης να γίνεται με τρόπο αξιόπιστο και κοινωνικά αποδεκτό.

Βάσει των παραπάνω, διατυπώνονται ορισμένες προτάσεις για μελλοντική έρευνα στο πεδίο. Μια σημαντική κατεύθυνση είναι η ανάπτυξη εξηγήσιμης τεχνητής νοημοσύνης (Explainable AI). Καθώς τα μοντέλα γίνονται ολοένα πιο περίπλοκα, η δημιουργία μεθόδων που επιτρέπουν την ερμηνεία των αποτελεσμάτων τους θα ενισχύσει τη διαφάνεια και την εμπιστοσύνη των χρηστών. Στενά συνυφασμένη είναι και η ανάγκη για καλύτερη ανίχνευση και αντιμετώπιση της μεροληψίας στα κοινωνικά δεδομένα και στους αλγόριθμους. Μελλοντικές μελέτες θα πρέπει να εστιάσουν σε τεχνικές που εντοπίζουν εγκιρίως προκαταλήψεις ή άνισες αντιπροσωπεύσεις στα δεδομένα (π.χ. φυλετικές ή έμφυλες μεροληψίες) και προσαρμόζουν τα μοντέλα ανάλογα, διασφαλίζοντας δίκαια και αξιόπιστα αποτελέσματα (Bennett et al., 2024). Η υιοθέτηση τέτοιων προσεγγίσεων θα συμβάλει στη δημιουργία πιο διαφανούς και υπεύθυνης τεχνητής νοημοσύνης για τα κοινωνικά δίκτυα.

Επιπλέον, ανοίγονται νέοι ορίζοντες με την ενσωμάτωση της μηχανικής μάθησης σε επαυξημένα και εικονικά κοινωνικά περιβάλλοντα. Η εμφάνιση του metaverse και των πλατφορμών εικονικής πραγματικότητας δημιουργεί πρόσφορο έδαφος για συστήματα τεχνητής νοημοσύνης που θα αναλύουν σύνθετες, πολυτροπικές αλληλεπιδράσεις σε πραγματικό χρόνο (Almeman et al., 2025). Η μελλοντική έρευνα μπορεί να διερευνήσει πώς οι αλγόριθμοι θα αντιμετωπίσουν αυτές τις προκλήσεις, από την κατανόηση της συμπεριφοράς των χρηστών σε εικονικές κοινότητες μέχρι τη διαχείριση νέων μορφών ψηφιακής επικοινωνίας. Τέλος, υπογραμμίζεται η σημασία μιας συμμετοχικής και συμπεριληπτικής τεχνητής νοημοσύνης στην ανάλυση κοινωνικών μέσων. Αυτό σημαίνει ότι ο σχεδιασμός των αλγορίθμων θα πρέπει να γίνεται με τη συμμετοχή των ίδιων των χρηστών και ποικίλων ενδιαφερομένων μερών, ώστε τα συστήματα να αντανακλούν τις ανάγκες της κοινωνίας στο σύνολό της. Μια τέτοια προσέγγιση θα μπορούσε να μεγιστοποιήσει την αποδοτικότητα και την αποδοχή των νέων τεχνολογιών, διασφαλίζοντας ότι η πρόοδος της τεχνητής νοημοσύνης υπηρετεί το κοινωνικό σύνολο. Με αυτόν τον τρόπο, το πεδίο θα συνεχίσει να εξελίσσεται με τρόπο καινοτόμο αλλά και υπεύθυνο, γεφυρώνοντας το χάσμα μεταξύ τεχνολογικής εξέλιξης και κοινωνικής προόδου.

Σε τελική ανάλυση, η παρούσα μελέτη ανέδειξε την ισχυρή δυναμική που αναπτύσσεται στη διασταύρωση μηχανικής μάθησης και κοινωνικών μέσων, προσφέροντας ένα πολυδιάστατο πλαίσιο κατανόησης, εφαρμογής και προβληματισμού. Η έρευνα τεκμηρίωσε ότι η αξιοποίηση των κοινωνικών δικτύων ως πεδίο εξόρυξης γνώσης δεν είναι μόνο τεχνικά εφικτή, αλλά και κοινωνικά κρίσιμη, καθώς αγγίζει ζητήματα διαφάνειας, δημόσιου διαλόγου και τεχνολογικής ηθικής. Η ανάγκη για διαρκή επικαιροποίηση, διαφάνεια και συμμετοχικότητα στην ανάπτυξη αλγορίθμων καθίσταται πιο επιτακτική από ποτέ. Το ερευνητικό αυτό εγχείρημα φιλοδοξεί να συμβάλει σε έναν ώριμο και υπεύθυνο διάλογο γύρω από τη χρήση της τεχνητής νοημοσύνης στον ψηφιακό δημόσιο χώρο και να εμπνεύσει περαιτέρω διεπιστημονικές προσεγγίσεις για την κατανόηση των κοινωνικών επιπτώσεων της τεχνολογίας.

# Βιβλιογραφικές αναφορές

## Ελληνόγλωσσα Βιβλία

1. **Βασιλειάδης, Γ. (2025).** *Πρακτικός Οδηγός Social Media: Facebook*. Αθήνα: Εκδόσεις Διαυλος.
2. **Βασιλειάδης, Γ. (2025b).** *Πρακτικός Οδηγός Social Media: YouTube*. Αθήνα: Εκδόσεις Διαυλος.
3. **Βλαχοπούλου, Μ. (2020).** *Ψηφιακό Μάρκετινγκ από τη Θεωρία στην Πράξη*. Αθήνα: Εκδόσεις Rosili.

## Ξενόγλωσσα Βιβλία

1. **Aggarwal, C. C. (2011).** *Social Network Data Analytics*. Springer.
2. **Alpaydin, E. (2020).** *Introduction to Machine Learning (4th ed.)*. MIT Press.
3. **Barabási, A-L. (2016).** *Network Science*. Cambridge University Press.
4. **Barocas, S., Hardt, M. & Narayanan, A. (2019).** *Fairness and Machine Learning*. MIT Press.
5. **Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009).** *Introduction to Meta-Analysis*. Wiley.
6. **Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. (1984).** *Classification and Regression Trees*. Belmont, CA: Wadsworth International.
7. **Bryman, A. (2016).** *Social research methods (5th ed.)*. Oxford University Press.
8. **Cadwalladr, C. & Graham-Harrison, E. (2018).** *The Cambridge Analytica Files*. The Guardian.
9. **Chaffey, D. & Ellis-Chadwick, F. (2019).** *Digital Marketing (7th ed.)*. Pearson Education.
10. **Churchland, P. S. & Sejnowski, T. J. (1992).** *The Computational Brain*. MIT Press.
11. **Deboeck, G. & Kohonen, T. (Eds.). (1998).** *Visual Explorations in Finance: With Self-Organizing Maps*. Springer.
12. **Duda, R. O., Hart, P. E. & Stork, D. G. (2000).** *Pattern classification*. John Wiley & Sons.
13. **Géron, A. (2019).** *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow (2nd ed.)*. O'Reilly Media.
14. **Gillespie, T. (2018).** *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.
15. **Goodfellow, I., Bengio, Y. & Courville, A. (2016).** *Deep Learning*. MIT Press.
16. **Creswell, J. W. & Creswell, J. D. (2018).** *Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.)*. Sage publications.
17. **Gujarati, D. N. & Porter, D. C. (2009).** *Basic Econometrics (5th ed.)*. McGraw-Hill Education.
18. **Hastie, T., Tibshirani, R. & Friedman, J. (2009).** *The Elements of Statistical Learning (2nd ed.)*. Springer.

19. **Heath, C. & Heath, D. (2007).** *Made to Stick: Why Some Ideas Survive and Others Die*. Random House.
20. **James, G., Witten, D., Hastie, T. & Tibshirani, R. (2021).** *An Introduction to Statistical Learning (2nd ed.)*. Springer.
21. **Kassambara, A. (2017).** *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning (1st ed.)*. STHDA.
22. **Katz, E. & Lazarsfeld, P. F. (1955).** *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Free Press.
23. **Kingsnorth, S. (2019).** *Digital Marketing Strategy: An Integrated Approach to Online Marketing (2nd ed.)*. Kogan Page.
24. **Kutner, M. H., Nachtsheim, C. J. & Neter, J. (2004).** *Applied Linear Regression Models (4th ed.)*. McGraw-Hill/Irwin.
25. **Manning, C. D., Raghavan, P. & Schütze, H. (2008).** *Introduction to Information Retrieval*. Cambridge University Press.
26. **Minsky, M. & Papert, S. (1987).** *Perceptrons: An Introduction to Computational Geometry, Expanded Edition*. MIT Press.
27. **Mitchell, T. M. (1997).** *Machine learning*. McGraw Hill.
28. **Molnar, C. (2022).** *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2nd ed.)*. Leanpub.
29. **Montgomery, D. C., Peck, E. A. & Vining, G. G. (2012).** *Introduction to Linear Regression Analysis (5th ed.)*. Wiley.
30. **Newman, M. E. J. (2010).** *Networks: An Introduction*. Oxford University Press.
31. **Nielsen, M. (2015).** *Neural Networks and Deep Learning*. Determination Press.
32. **Pariser, E. (2011).** *The Filter Bubble: What the Internet is Hiding from You*. Penguin Press.
33. **Platt, J. (1998).** *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines*. Tech. Rep., Microsoft Research, Technical Report msr-tr-98-14.
34. **Provost, F. & Fawcett, T. (2013).** *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.
35. **Russell, S. J. & Norvig, P. (2020).** *Artificial Intelligence: A Modern Approach (4th ed.)*. Pearson.
36. **Schölkopf, B. & Smola, A. J. (2002).** *Learning with kernels: Support vector machines, regularization, optimization and beyond*. MIT Press.
37. **Scott, J. (2017).** *Social Network Analysis (4th ed.)*. SAGE.
38. **Spath, H. (1989).** *Cluster Analysis Algorithms*. Ellis Horwood, Chichester.
39. **Steinhauer, J. (2021).** *History, Disrupted: How Social Media and the World Wide Web Have Changed the Past*. Palgrave Macmillan.
40. **Stock, J. H. & Watson, M. W. (2019).** *Introduction to Econometrics (4th ed.)*. Pearson.
41. **Sutton, R. S. & Barto, A. G. (2018).** *Reinforcement Learning: An Introduction (2nd ed.)*. MIT Press.
42. **Tan, P.-N., Steinbach, M. & Kumar, V. (2018).** *Introduction to Data Mining (2nd ed.)*. Pearson.

43. Tuten, T. L. & Solomon, M. R. (2018). *Social Media Marketing (3rd ed.)*. SAGE Publications.
44. Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley-Interscience.
45. Weller, K., Bruns, A., Burgess, J., Mahrt, M. & Puschmann, C. (Eds.). (2013). *Twitter and Society*. Peter Lang.
46. Yegnanarayana, B. (2009). *Artificial Neural Networks*. PHI Learning.
47. Zuboff, S. (2019). *The Age of Surveillance Capitalism*. Profile Books.

#### Άρθρα σε Επιστημονικά Περιοδικά και Συλλογικούς Τόμους

1. Abiodun, O., Jantan, A., Omolara, O., Dada, K., Umar, A., Linus, O., Arshad, H., Aminu Kazaure, A., Gana, U. & Kiru, M.. (2019). *Comprehensive Review of Artificial Neural Network Applications to Pattern Recognition*. IEEE Access. PP. 1-1. Διαθέσιμο στο: <https://doi.org/10.1109/ACCESS.2019.2945545>
2. Acikara, T., Xia, B., Yigitcanlar, T. & Hon, C. (2023). *Contribution of Social Media Analytics to Disaster Response Effectiveness: A Systematic Review of the Literature*. Sustainability, 15(11), 8860. Διαθέσιμο στο: <https://doi.org/10.3390/su15118860>
3. Acquisti, A., Brandimarte, L. & Loewenstein, G. (2015). *Privacy and human behavior in the age of information*. Science, 347(6221), 509-514. Διαθέσιμο στο: <https://doi.org/10.1126/science.aaa1465>
4. Aggarwal, C.C. (2015). *Outlier Analysis*. In: *Data Mining*. Springer, Cham, 237-263. Διαθέσιμο στο: [https://doi.org/10.1007/978-3-319-14142-8\\_8](https://doi.org/10.1007/978-3-319-14142-8_8)
5. Aggarwal, C. C. (2016). *Model-Based Collaborative Filtering*. In: *Recommender Systems*, Springer, Cham, 71-138. Διαθέσιμο στο: [https://doi.org/10.1007/978-3-319-29659-3\\_3](https://doi.org/10.1007/978-3-319-29659-3_3)
6. Aggarwal, C.C. (2018). *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing, Cham. Διαθέσιμο στο: <https://doi.org/10.1007/978-3-319-94463-0>
7. Alassafi, M., Alghamdi, W., Naveena, S., Alkhayyat, A., Tolib, A. & Ugli, I. (2023). *Machine Learning for Predictive Analytics in Social Media Data*. E3S Web of Conferences. 399. Διαθέσιμο στο: <https://doi.org/10.1051/e3sconf/202339904046>
8. Alayrac, J.-B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., ... & Zisserman, A. (2022). *Flamingo: a Visual Language Model for Few-Shot Learning*. arXiv preprint arXiv:2204.14198. Διαθέσιμο στο: <https://arxiv.org/abs/2204.14198>
9. Almeman, K., EL Ayeb, F., Berrima, M., Issaoui, B. & Morsy, H. (2025). *The Integration of AI and Metaverse in Education: A Systematic Literature Review*. Applied Sciences, 15(2), 863. Διαθέσιμο στο: <https://doi.org/10.3390/app15020863>
10. Annetta, L., Johnson, A., Newton, M. H. & Franco, Y. (2024). *Immersive Spatial Computing: How Technology Can Improve Science Content Reading and Vocabulary in Elementary Schools*. Education Sciences, 14(12), 1355. Διαθέσιμο στο: <https://doi.org/10.3390/educsci14121355>

11. **Ariga, K. (2025).** *Layer-by-Layer Nanoarchitectonics: A Method for Everything in Layered Structures.* Materials, 18(3), 654. Διαθέσιμο στο: <https://doi.org/10.3390/ma18030654>
12. **Arora, P., Deepali, k. & Varshney, S. (2016).** *Analysis of K-Means and K-Medoids Algorithm For Big Data.* Procedia Computer Science. 78. 507-512. Διαθέσιμο στο: <https://doi.org/10.1016/j.procs.2016.02.095>
13. **Arthur, D. & Vassilvitskii, S. (2007).** *K-Means++: The Advantages of Careful Seeding.* In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, Philadelphia, 1027-1035. Διαθέσιμο στο: <https://theory.stanford.edu/~sergei/papers/kMeansPP-soda.pdf>
14. **Bai, Y., Kadavath, S., Kundu, S., Asbell, A., Kernion, J., Jones, A., ... & Amodei, D. (2022).** *Constitutional AI: Harmlessness from AI feedback.* arXiv:2212.08073 [cs.CL]. Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.2212.08073>
15. **Belinkov, Y. & Glass, J. (2019).** *Analysis Methods in Neural Language Processing: A Survey.* Transactions of the Association for Computational Linguistics, 7, 49–72. Διαθέσιμο στο: [https://doi.org/10.1162/tacl\\_a\\_00254](https://doi.org/10.1162/tacl_a_00254)
16. **Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell, S. (2021).** *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?.* Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21), 610–623. Διαθέσιμο στο: <https://doi.org/10.1145/3442188.3445922>
17. **Bengio, Y. (2009).** *Learning Deep Architectures for AI (PDF).* Foundations and Trends in Machine Learning, 2, 1-127. Διαθέσιμο στο: <https://doi.org/10.1561/22000000006>
18. **Bennett, C., Cole-Lewis, H., Farquhar, S., Haamel, N., Babenko, B., Lang, O., Fleck, M., Traynis, I., Lau, C., Horn, I. & Lyles, C. (2024).** *Interdisciplinary Expertise to Advance Equitable Explainable AI.* Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.2406.18563>
19. **Berger, J. & Milkman, K. L. (2012).** *What makes online content viral?* Journal of Marketing Research, 49(2), 192–205. Διαθέσιμο στο: <https://doi.org/10.1509/jmr.10.0353>
20. **Binns, R. (2018).** *Fairness in Machine Learning: Lessons from Political Philosophy.* In Proceedings of the 2018 Conference on Fairness, Accountability and Transparency (pp. 149-159). Διαθέσιμο στο: <https://proceedings.mlr.press/v81/binns18a.html>
21. **Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003).** *Latent Dirichlet Allocation.* Journal of Machine Learning Research, 3, 993–1022. Διαθέσιμο στο: <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
22. **Blondel, V.D., Guillaume, J.L., Lambiotte, R. & Lefebvre, E. (2008).** *Fast Unfolding of Communities in Large Networks.* Journal of Statistical Mechanics: Theory and Experiment, 2008, Article ID: P10008. Διαθέσιμο στο: <https://doi.org/10.1088/1742-5468/2008/10/P10008>

23. Bojarski, M, Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L., Monfort, M., Muller, U., Zhang, J., Zhang, X., Zhao, J. & Zieba, K. (2016). *End to End Learning for Self-Driving Cars*. Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.1604.07316>
24. Bonacich, P. (1987). *Power and Centrality: A Family of Measures*. American Journal of Sociology, 92, 1170-1182. Διαθέσιμο στο: <http://dx.doi.org/10.1086/228631>
25. Boser, B.E., Guyon, I.M. & Vapnik, V.N. (1992). *A Training Algorithm for Optimal Margin Classifiers*. Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, July 1992, 144-152. Διαθέσιμο στο: <https://doi.org/10.1145/130385.130401>
26. Bovet, A. & Makse, H. A. (2019). *Influence of fake news in Twitter during the 2016 US presidential election*. Nature Communications, 10(1), 7. Διαθέσιμο στο: <https://doi.org/10.1038/s41467-018-07761-2>
27. Boyd, D. & Crawford, K. (2012). *Critical questions for big data: Provocations for a cultural, technological and scholarly phenomenon*. Information, Communication & Society, 15(5), 662–679. Διαθέσιμο στο: <https://doi.org/10.1080/1369118X.2012.678878>
28. Breiman, L. (1996). *Bagging Predictors*. Machine Learning, Vol. 24, No. 2, 1996, pp. 123-140. Διαθέσιμο στο: <http://dx.doi.org/10.1007/BF00058655>
29. Breiman, L. (2001). *Random Forests*. Machine Learning, 45, 5-32. Διαθέσιμο στο: <http://dx.doi.org/10.1023/A:1010933404324>
30. Brin, S. & Page, L. (1998). *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Computer Networks and ISDN Systems, 30, 107-117. Διαθέσιμο στο: [http://dx.doi.org/10.1016/S0169-7552\(98\)00110-X](http://dx.doi.org/10.1016/S0169-7552(98)00110-X)
31. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D. et al. (2020). *Language Models Are Few-Shot Learners*. Advances in Neural Information Processing Systems, 33, 1877-1901. Διαθέσιμο στο: <https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf>
32. Bruns, A. (2019). *After the 'APIcalypse': social media platforms and their fight against critical scholarly research*. Information, Communication & Society, 22(11), 1544–1566. Διαθέσιμο στο: <https://doi.org/10.1080/1369118X.2019.1637447>
33. Bruns, A. & Stieglitz, S. (2013). *Towards more systematic Twitter analysis: metrics for tweeting activities*. International Journal of Social Research Methodology, 16(2), 91–108. Διαθέσιμο στο: <https://doi.org/10.1080/13645579.2012.756095>
34. Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). *Sparks of Artificial General Intelligence: Early experiments with GPT-4*. arXiv preprint arXiv:2303.12712. Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.2303.12712>
35. Burges, C.J.C. (1998). *A Tutorial on Support Vector Machines for Pattern Recognition*. Data Mining and Knowledge Discovery, 2, 121-167. Διαθέσιμο στο: <https://doi.org/10.1023/A:1009715923555>
36. Burrell, J. (2016). *How the machine 'thinks': Understanding opacity in machine learning algorithms*. Big Data & Society, 3(1). Διαθέσιμο στο: <https://doi.org/10.1177/2053951715622512>

37. Cambria, E., Schuller, B., Xia, Y. & Havasi, C. (2013). *New avenues in opinion mining and sentiment analysis*. IEEE Intelligent Systems 28, 2 (March 2013), 15–21. Διαθέσιμο στο: <https://doi.org/10.1109/MIS.2013.30>
38. Cambria, E. & White, B. (2014). *Jumping NLP curves: A review of natural language processing research*. IEEE Computational Intelligence Magazine, 9(2), 48–57. Διαθέσιμο στο: <https://doi.org/10.1109/MCI.2014.2307227>
39. Cambria, E., Poria, S., Gelbukh, A. & Thelwall, M. (2017). *Sentiment Analysis Is a Big Suitcase*. IEEE Intelligent Systems. 32. 74-80. Διαθέσιμο στο: <https://doi.org/10.1109/MIS.2017.4531228>
40. Cauwenberghs, G. & Poggio, T. (2001). *Incremental and Decremental Support Vector Machine Learning*. Advances in Neural Information Processing Systems, 13. Διαθέσιμο στο: [https://proceedings.neurips.cc/paper\\_files/paper/2000/file/155fa09596c7e18e50b58eb7e0c6ccb4-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2000/file/155fa09596c7e18e50b58eb7e0c6ccb4-Paper.pdf)
41. Cesare, N., Lee, H., McCormick, T., Spiro, E. & Zagheni, E. (2018). *Promises and pitfalls of using digital trace data for demographic research*. Demography, 55(5), 1979–1999. Διαθέσιμο στο: <https://doi.org/10.1007/s13524-018-0715-2>
42. Chandola, V., Banerjee, A. & Kumar, V. (2009). *Anomaly Detection: A Survey*. ACM Computing Surveys, 41, 1-58. Διαθέσιμο στο: <https://doi.org/10.1145/1541880.1541882>
43. Chang, C.C. & Lin, C.J. (2011). *LIBSVM: A Library for Support Vector Machines*. ACM Transactions on Intelligent Systems and Technology, 2, 27. Διαθέσιμο στο: <http://dx.doi.org/10.1145/1961189.1961199>
44. Chapelle, O., Vapnik, V., Bousquet, O. & Mukherjee, S. (2002). *Choosing multiple parameters for Support Vector Machines*. Machine Learning, 46(1-3), 131–159. Διαθέσιμο στο: <https://doi.org/10.1023/A:1012450327387>
45. Chen, T. & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, San Francisco, 13-17 August 2016, 785-794. Διαθέσιμο στο: <https://doi.org/10.1145/2939672.2939785>
46. Chinta, U., Goel, P. & Renuka, A. (2023). *Leveraging AI and Machine Learning in Salesforce for Predictive Analytics and Customer Insights*. Universal Research Reports. Διαθέσιμο στο: <https://doi.org/10.246-258.10.36676/urr.v10.i1.1328>
47. Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F. & Flammini, A. (2011). *Political Polarization on Twitter*. Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 5, 89-96. Διαθέσιμο στο: <https://ojs.aaai.org/index.php/ICWSM/article/view/14126>
48. Cortes, C. & Vapnik, V. (1995). *Support-Vector Networks*. Machine Learning, 20, 273-297. Διαθέσιμο στο: <http://dx.doi.org/10.1007/BF00994018>

49. **Cover, T. & Hart, P. (1967).** *Nearest Neighbor Pattern Classification*. IEEE Transactions on Information Theory, 13, 21-27. Διαθέσιμο στο: <http://dx.doi.org/10.1109/TIT.1967.1053964>
50. **Covington, P., Adams, J. & Sargin, E. (2016).** *Deep Neural Networks for YouTube Recommendations*. Proceedings of the 10th ACM Conference on Recommender Systems, 191–198. Διαθέσιμο στο: <https://doi.org/10.1145/2959100.2959190>
51. **Cristianini, N. & Shawe-Taylor, J. (2000).** *An Introduction to Support Vector Machines and Other Kernelbased Learning Methods*. Cambridge University Press, Cambridge. Διαθέσιμο στο: <https://doi.org/10.1017/CBO9780511801389>
52. **Dang, N. C., Moreno-García, M. N. & De la Prieta, F. (2020).** *Sentiment Analysis Based on Deep Learning: A Comparative Study*. Electronics, 9(3), 483. Διαθέσιμο στο: <https://doi.org/10.3390/electronics9030483>
53. **Devlin, J., Chang, M. W., Lee, K. & Toutanova, K. (2019).** *BERT: Pre-training of deep bidirectional transformers for language understanding*. In Proceedings of NAACL-HLT 2019 (pp. 4171–4186). Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.1810.04805>
54. **Dietterich, T.G. (2000).** *Ensemble Methods in Machine Learning*. In: Multiple Classifier Systems. MCS 2000. Lecture Notes in Computer Science, vol 1857. Springer, Berlin, Heidelberg. Διαθέσιμο στο: [https://doi.org/10.1007/3-540-45014-9\\_1](https://doi.org/10.1007/3-540-45014-9_1)
55. **Dimitri, G. M. (2022).** *A Short Survey on Deep Learning for Multimodal Integration: Applications*. Future Perspectives and Challenges. Computers, 11(11), 163. Διαθέσιμο στο: <https://doi.org/10.3390/computers11110163>
56. **Ding, Q., Jahanshahi, H., Wang, Y., Bekiros, S. & Alassafi, M. O. (2022).** *Optimal Reinforcement Learning-Based Control Algorithm for a Class of Nonlinear Macroeconomic Systems*. Mathematics, 10(3), 499. Διαθέσιμο στο: <https://doi.org/10.3390/math10030499>
57. **Donahue, J., Hendricks, L.A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K. & Darrell, T. (2014).** *Long-Term Recurrent Convolutional Networks for Visual Recognition and Description*. arXiv: 1411.4389. Διαθέσιμο στο: <https://doi.org/10.48550/ARXIV.1411.4389>
58. **Doshi-Velez, F. & Kim, B. (2017).** *Towards a rigorous science of interpretable machine learning*. arXiv preprint arXiv:1702.08608. Διαθέσιμο στο: <https://arxiv.org/abs/1702.08608>
59. **Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A. & Donaldson, L. (2013).** *Harnessing the cloud of patient experience: using social media to detect poor quality healthcare*. BMJ Quality & Safety, 22(3), 251–255. Διαθέσιμο στο: <https://doi.org/10.1136/bmjqs-2012-001527>
60. **Dritsas, E. & Trigka, M. (2025).** *Exploring the Intersection of Machine Learning and Big Data: A Survey*. Machine Learning and Knowledge Extraction, 7(1), 13. Διαθέσιμο στο: <https://doi.org/10.3390/make7010013>
61. **Du, X., Yu, M., Zhang, Z., Tong, M., Zhu, Y. & Xue, C. (2025).** *A Task- and Role-Oriented Design Method for Multi-User Collaborative Interfaces*. Sensors, 25(6), 1760. Διαθέσιμο στο: <https://doi.org/10.3390/s25061760>

62. Dwivedi, Y. K., Hughes, D. L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T. & Williams, M. D. (2021). *Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy*. International Journal of Information Management, 57, Article 101994. Διαθέσιμο στο: <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>
63. Erhan, D., Bengio, Y., Courville, A., Manzagol, P. A., Vincent, P. & Bengio, S. (2010). *Why does unsupervised pre-training help deep learning?* Journal of Machine Learning Research, 11, 625–660. Διαθέσιμο στο: <https://doi.org/10.1145/1756006.1756025>
64. Ester, M., Kriegel, H.P., Sander, J. & Xu, X. (1996). *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, 96, 226-231. Διαθέσιμο στο: <https://file.biolab.si/papers/1996-DBSCAN-KDD.pdf>
65. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M. & Thrun, S. (2017). *Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks*. Nature, 542, 115-118. Διαθέσιμο στο: <https://doi.org/10.1038/nature21056>
66. Ferrara, E., Varol, O., Davis, C., Menczer, F. & Flammini, A. (2016). *The Rise of Social Bots*. Communications of the ACM, 59, 96-104. Διαθέσιμο στο: <https://doi.org/10.1145/2818717>
67. Ferrara, E. (2017). *Disinformation and social bot operations in the run up to the 2017 French presidential election*. First Monday, 22(8). Διαθέσιμο στο: <https://doi.org/10.5210/fm.v22i8.8005>
68. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V. & Luetge, C. (2018). *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles and Recommendations*. Minds and Machines, 28, 689-707. Διαθέσιμο στο: <https://doi.org/10.1007/s11023-018-9482-5>
69. Fortunato, S. & Hric, D. (2016). *Community detection in networks: A user guide*. Physics Reports, 659, 1–44. Διαθέσιμο στο: <https://doi.org/10.1016/j.physrep.2016.09.002>
70. Fournier-Tombs, E. (2020). *Book Review: Twitter and tear gas: The power and fragility of networked protest*. Convergence, 26(5-6), 1417-1419. Διαθέσιμο στο: <https://doi.org/10.1177/1354856520923965>
71. François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G. & Pineau, J. (2018). *An Introduction to Deep Reinforcement Learning*. Foundations and Trends in Machine Learning, 11, 219-354. Διαθέσιμο στο: <https://doi.org/10.1561/22000000071>
72. Freeman, L. C. (1979). *Centrality in social networks conceptual clarification*. Social Networks, 1(3), 215–239. Διαθέσιμο στο: [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)

73. **Gadekallu, T. R., Khare, N., Bhattacharya, S., Singh, S., Reddy, M. P. K., Alazab, M. & Maddikunta, P. K. R. (2023).** *Blockchain for the Metaverse: A Review*. Διαθέσιμο στο: <https://doi.org/10.1016/j.future.2023.02.008>
74. **Gal, Y. & Ghahramani, Z. (2016).** *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*. Proceedings of the 33rd International Conference on Machine Learning, 1050–1059. Διαθέσιμο στο: <https://arxiv.org/abs/1506.02142>
75. **Gama, J. Žliobaitė, I. Bifet, A. Pechenizkiy, M. & Bouchachia, H. (2014).** *A Survey on Concept Drift Adaptation*. ACM Computing Surveys (CSUR). 46. Διαθέσιμο στο: <https://doi.org/10.1145/2523813>
76. **Gandomi, A. & Haider, M. (2015).** *Beyond the hype: Big data concepts, methods and analytics*. International Journal of Information Management, 35(2), 137–144. Διαθέσιμο στο: <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
77. **Goel, S., Anderson, A., Hofman, J. & Watts, D. J. (2015).** The structural virality of online diffusion. Management Science, 62(1), 180–196. Διαθέσιμο στο: <https://doi.org/10.1287/mnsc.2015.2158>
78. **González-Bailón, S. (2013).** *Social science in the era of big data*. Policy & Internet, 5(2), 147–160. Διαθέσιμο στο: <https://doi.org/10.1002/1944-2866.POI328>
79. **Goodman, B. & Flaxman, S. (2017).** *European Union regulations on algorithmic decision-making and a “right to explanation”*. AI Magazine, 38(3), 50-57. Διαθέσιμο στο: <https://doi.org/10.1609/aimag.v38i3.2741>
80. **Gulan, M., Feng, Y., Mvondo, G. F. N. & Niu, B. (2025).** *How Do Virtual Influencers Affect Consumer Brand Evangelism in the Metaverse? The Effects of Virtual Influencers’ Marketing Efforts, Perceived Coolness, and Anthropomorphism*. Journal of Theoretical and Applied Electronic Commerce Research, 20(1), 36. Διαθέσιμο στο: <https://doi.org/10.3390/jtaer20010036>
81. **Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. (2002).** *Gene selection for cancer classification using Support Vector Machines*. Machine Learning, 46(1-3), 389–422. Διαθέσιμο στο: <https://doi.org/10.1023/A:1012487302797>
82. **Han, B., Cook, P. & Baldwin, T. (2013).** *Lexical normalization for social media text*. ACM Transactions on Intelligent Systems and Technology (TIST), 4(1), 5. Διαθέσιμο στο: <https://doi.org/10.1145/2414425.2414430>
83. **Hassabi, D., Kumaran, D., Summerfield, C. & Botvinick, M. (2017).** *Neuroscience-Inspired Artificial Intelligence*. Neuron, 95, 245-258. Διαθέσιμο στο: <https://doi.org/10.1016/j.neuron.2017.06.011>
84. **Hassan, A., Abbasi, A. & Zeng, D. (2013).** *Twitter Sentiment Analysis: A Bootstrap Ensemble Framework*. Proceedings – SocialCom /PASSAT /BigData /EconCom / BioMedCom 2013. 357-364. Διαθέσιμο στο: <https://doi.org/10.1109/SocialCom.2013.56>
85. **Haykin, S. (2009).** *Neural Networks and Learning Machines*. Third Edition, Pearson Education, Inc., McMaster University, Hamilton. Διαθέσιμο στο: <http://dai.fmph.uniba.sk/courses/NN/haykin.neural-networks.3ed.2009.pdf>

86. **He, H. & Garcia, E.A. (2009).** *Learning from Imbalanced Data*. IEEE Transactions on Knowledge and Data Engineering, 21, 1263-1284. Διαθέσιμο στο: <http://dx.doi.org/10.1109/TKDE.2008.239>
87. **Heaton, J. B., Polson, N. G. & Witte, J. H. (2017).** *Deep learning in finance*. Annual Review of Financial Economics, 9, 145–181. Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.1602.06561>
88. **Hendrycks, D., Basart, S., Mu, N., Kadavath, S., Wang, F., Dorundo, E., ... & Song, D. (2021).** *The Many Faces of Robustness: A Critical Analysis of Out-of-Distribution Generalization*. Proceedings of the International Conference on Computer Vision (ICCV). Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.2006.16241>
89. **Hinton, G., Osindero, S. & Teh, Y. (2006).** *A Fast Learning Algorithm for Deep Belief Nets*. Neural Computation, 18, 1527-1554. Διαθέσιμο στο: <http://dx.doi.org/10.1162/neco.2006.18.7.1527>
90. **Ho, T.K. (1998).** *The Random Subspace Method for Constructing Decision Forests*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20, 832-844. Διαθέσιμο στο: <http://dx.doi.org/10.1109/34.709601>
91. **Hoi, S. C. H., Sahoo, D., Lu, J. & Zhao, P. (2018).** Online learning: A comprehensive survey. Neurocomputing, 459, 249–289. Διαθέσιμο στο: <https://doi.org/10.1016/j.neucom.2021.04.112>
92. **Hosmer Jr., D.W., Lemeshow, S. & Sturdivant, R.X. (2013).** *Applied Logistic Regression*. 3rd Edition, John Wiley & Sons, Hoboken, NJ. Διαθέσιμο στο: <https://doi.org/10.1002/9781118548387>
93. **Hsu, C.W. & Lin, C.J. (2002).** *A Comparison of Methods for Multiclass Support Vector Machines*. IEEE Transactions on Neural Networks, 13, 415-425. Διαθέσιμο στο: <http://dx.doi.org/10.1109/72.991427>
94. **Huang, A., Xu, R., Chen, Y. & Guo, M. (2023).** *Research on multi-label user classification of social media based on ML-KNN algorithm*. Technological Forecasting and Social Change. 188. 122271. Διαθέσιμο στο: <https://doi.org/10.1016/j.techfore.2022.122271>
95. **Imran, M., Castillo, C., Diaz, F. & Vieweg, S. (2015).** *Processing Social Media Messages in Mass Emergency: A Survey*. ACM Computing Surveys, 47(4), 1–38. Διαθέσιμο στο: <https://doi.org/10.1145/2771588>
96. **Imran, M., Zaman, U., Imran, Imtiaz, J., Fayaz, M. & Gwak, J. (2021).** *Comprehensive Survey of IoT, Machine Learning, and Blockchain for Health Care Applications: A Topical Assessment for Pandemic Preparedness, Challenges and Solutions*. Electronics, 10(20), 2501. Διαθέσιμο στο: <https://doi.org/10.3390/electronics10202501>
97. **Isaak, J. & Hanna, M. J. (2018).** *User data privacy: Facebook, Cambridge Analytica and privacy protection*. Computer, 51(8), 56-59. Διαθέσιμο στο: <https://doi.org/10.1109/MC.2018.3191268>

98. **Isnain, A. R., Supriyanto, J. & Kharisma, M. P. (2021).** Implementation of K-Nearest Neighbor (K-NN) Algorithm for Public Sentiment Analysis of Online Learning. Indonesian Journal of Computing and Cybernetics Systems, 15(2), 121–132. Διαθέσιμο στο: <https://doi.org/10.22146/ijccs.65176>
99. **Jain, A.K. (2010).** *Data Clustering: 50 Years beyond K-Means*. Pattern Recognition Letters, 31, 651-666. Διαθέσιμο στο: <http://dx.doi.org/10.1016/j.patrec.2009.09.011>
100. **Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023).** *Survey of Hallucination in Natural Language Generation*. ACM Computing Surveys, 55(12), 1-38. Διαθέσιμο στο: <https://doi.org/10.1145/3571730>
101. **Jiao, X., Yin, Y., Shang, L., Jiang, X., Chen, X., Li, L., ... & Liu, Q. (2020).** *TinyBERT: Distilling BERT for natural language understanding*. In Findings of EMNLP 2020 (pp. 4163–4174). Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.1909.10351>
102. **Jones, M. (2015).** *The Complete History of Social Media: The Founding of Online Networking*. Διαθέσιμο στο: <https://historycooperative.org/the-history-of-social-media>
103. **Jordan, M.I. & Mitchell, T.M. (2015).** Machine Learning: Trends, Perspectives, and Prospects. Science (New York, N.Y.), 349, 255-260. Διαθέσιμο στο: <https://doi.org/10.1126/science.aaa8415>
104. **Kaelbling, L. P., Littman, M. L. & Moore, A. W. (1996).** *Reinforcement Learning: A Survey*. Journal of Artificial Intelligence Research, 4, 237–285. Διαθέσιμο στο: <https://doi.org/10.1613/jair.301>
105. **Kaliyar, R. K., Goswami, A., Narang, P. & Sinha, K. (2021).** *FakeBERT: Fake news detection in social media with a BERT-based deep learning approach*. Multimedia Tools and Applications, 80, 11765–11788. Διαθέσιμο στο: <https://doi.org/10.1007/s11042-020-10183-2>
106. **Kaplan, A.M. & Haenlein, M. (2010).** *Users of the World, Unite! The Challenges and Opportunities of Social Media*. Business Horizons, 53, 59-68. Διαθέσιμο στο: <http://dx.doi.org/10.1016/j.bushor.2009.09.003>
107. **Kapur, K. & Harikrishnan, R. (2022).** *Comparative Study of Sentiment Analysis for Multi-Sourced Social Media Platforms*. arXiv preprint. Διαθέσιμο στο: <https://arxiv.org/abs/2212.04688>
108. **Kaufman, L. & Rousseeuw, P. J. (1990).** *Finding Groups in Data—An Introduction to Cluster Analysis*. New York: John Wiley & Sons Inc. Διαθέσιμο στο: <https://doi.org/10.1002/9780470316801>
109. **Kaye, D. B. V., Chen, X., & Zeng, J. (2020).** *The co-evolution of two Chinese mobile short video apps: Parallel platformization of Douyin and TikTok*. Mobile Media & Communication. 9. 205015792095212. Διαθέσιμο στο: <https://doi.org/10.1177/2050157920952120>
110. **Keerthi, S.S., Shevade, S.K., Bhattacharyya, C. & Murthy, K.R.K. (2001).** *Improvements to Platt's SMO Algorithm for SVM Classifier Design*. Neural Computation, 13, 637-649. Διαθέσιμο στο: <https://doi.org/10.1162/089976601300014493>

111. **Kempe, D., Kleinberg, J. & Tardos, É. (2003).** *Maximizing the spread of influence through a social network*. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 137–146. Διαθέσιμο στο: <https://doi.org/10.1145/956750.956769>
112. **Khosla, A., Das Sarma, A. & Hamid, R. (2014).** What makes an image popular? WWW 2014 Proceedings, 867–876. Διαθέσιμο στο: <https://doi.org/10.1145/2566486.2567996>
113. **Kingma, D.P. & Ba, J. (2014).** *Adam: A Method for Stochastic Optimization*. Proceedings of the 3rd International Conference on Learning Representations (ICLR), Banff, 14-16 April 2014. arXiv preprint arXiv:1412.6980. Διαθέσιμο στο: <https://arxiv.org/abs/1412.6980>
114. **Kipf, T. N. & Welling, M. (2016).** *Semi-Supervised Classification with Graph Convolutional Networks*. International Conference on Learning Representations (ICLR). Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.1609.02907>
115. **Kitchenham, B. & Charters, S. (2007).** *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report. Διαθέσιμο στο: [https://legacyfileshare.elsevier.com/promis\\_misc/525444systematicreviewsguide.pdf](https://legacyfileshare.elsevier.com/promis_misc/525444systematicreviewsguide.pdf)
116. **Kleinberg, J. (2003).** *Bursty and hierarchical structure in streams*. Data Mining and Knowledge Discovery, 7(4), 373–397. Διαθέσιμο στο: <https://doi.org/10.1023/A:1024940629314>
117. **Ko, T., Rhiu, I., Yun, M. H. & Cho, S. (2020).** *A Novel Framework for Identifying Customers' Unmet Needs on Online Social Media Using Context Tree*. Applied Sciences, 10(23), 8473. Διαθέσιμο στο: <https://doi.org/10.3390/app10238473>
118. **Kohonen, T. (1982).** *Self-Organized Formation of Topologically Correct Feature Maps*. Biological Cybernetics, 43, 59-69. Διαθέσιμο στο: <http://dx.doi.org/10.1007/BF00337288>
119. **Kohonen, T. (1990).** *The Self-Organizing Map*. Proceedings of the IEEE, 78, 1464-1480. Διαθέσιμο στο: <http://dx.doi.org/10.1109/5.58325>
120. **Kohonen, T. (2001).** *Self-Organizing Maps. 3rd Edition*. Springer, New York. Διαθέσιμο στο: <http://dx.doi.org/10.1007/978-3-642-56927-2>
121. **Kong, L., Chen, G., Huang, B., Xing, E., Chi, Y. & Zhang, K. (2024).** *Learning Discrete Concepts in Latent Hierarchical Models*. Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.2406.00519>
122. **Kora, R. & Mohammed, A. (2023).** *An enhanced approach for sentiment analysis based on meta-ensemble deep learning*. Social Network Analysis and Mining. 13. 38. Διαθέσιμο στο: <https://doi.org/10.1007/s13278-023-01043-6>
123. **Koren, Y., Bell, R. & Volinsk, C. (2009).** *Matrix Factorization Techniques for Recommender Systems*. Computer, 42, 30-37. Διαθέσιμο στο: <https://doi.org/10.1109/MC.2009.263>

124. **Kosinski, M., Stillwell, D. & Graepel, T. (2013).** *Private Traits and Attributes Are Predictable from Digital Records of Human Behavior*. Proceedings of the National Academy of Sciences, 110, 5802-5805. Διαθέσιμο στο: <http://dx.doi.org/10.1073/pnas.1218772110>
125. **Krizhevsky, A., Sutskever, I. & Hinton, G.E. (2012).** *ImageNet Classification with Deep Convolutional Neural Networks*. In: Pereira, F., Burges, C.J.C., Bottou, L. and Weinberger, K.Q., Eds., Advances in Neural Information Processing Systems, Vol. 25, Curran Associates, Inc., Red Hook, NY, 1097-1105. Διαθέσιμο στο: [https://proceedings.neurips.cc/paper\\_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf)
126. **Kyrlitsias, C. & Michael-Grigoriou, D. (2022).** *Social Interaction With Agents and Avatars in Immersive Virtual Environments: A Survey*. Frontiers in Virtual Reality. 2. 10.3389/frvir.2021.786665. Διαθέσιμο στο: <https://doi.org/10.3389/frvir.2021.786665>
127. **Larochelle, H., Bengio, Y., Louradour, J. & Lamblin, P. (2012).** *Learning algorithms for the classification restricted Boltzmann machine*. Journal of Machine Learning Research, 13, 643–669. Διαθέσιμο στο: <http://www.jmlr.org/papers/volume13/larochelle12a/larochelle12a.pdf>
128. **Lazer, D. M. J., Baum, M. A., Grinberg, N., Friedland, L., Joseph, K. & Hobbs, W. (2018).** *The science of fake news*. Science, 359(6380), 1094–1096. Διαθέσιμο στο: <https://doi.org/10.1126/science.aao2998>
129. **LeCun, Y., Bengio, Y. & Hinton, G. (2015).** *Deep Learning*. Nature, 521, 436-444. Διαθέσιμο στο: <http://dx.doi.org/10.1038/nature14539>
130. **Lee, D. D. & Seung, H. S. (2000).** *Algorithms for Non-negative Matrix Factorization*. Advances in Neural Information Processing Systems, 13. Διαθέσιμο στο: [https://proceedings.neurips.cc/paper\\_files/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf)
131. **Lee, H., Grosse, R., Ranganath, R. & Ng, A.Y. (2009).** *Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations*. Proceedings of the 26th Annual International Conference on Machine Learning, 609-616. Διαθέσιμο στο: <http://dx.doi.org/10.1145/1553374.1553453>
132. **Li, J., Guo, W., Liu, H., Chen, X., Yu, A. & Li, J. (2021).** *Predicting User Activity Intensity Using Geographic Interactions Based on Social Media Check-In Data*. ISPRS International Journal of Geo-Information, 10(8), 555. Διαθέσιμο στο: <https://doi.org/10.3390/ijgi10080555>
133. **Lin, C. (2016).** *Predicting user response to sponsored advertising on social media via the technology acceptance model*. Computers in Human Behavior. 64. 710–718. Διαθέσιμο στο: <https://doi.org/10.1016/j.chb.2016.07.027>
134. **Lipton, Z. C. (2018).** *The mythos of model interpretability*. Communications of the ACM, 61(10), 36–43. Διαθέσιμο στο: <https://doi.org/10.1145/3233231>

135. **Liu, F.T., Ting, K.M. & Zhou, Z.H. (2008).** *Isolation Forest*. 2008 8th IEEE International Conference on Data Mining, Pisa, 15-19 December 2008, 413-422. Διαθέσιμο στο: <https://doi.org/10.1109/ICDM.2008.17>
136. **Liu, N. (2009).** *Topic Detection and Tracking*. In: LIU, L., ÖZSU, M.T. (eds) *Encyclopedia of Database Systems*. Springer, Boston, MA. Διαθέσιμο στο: [https://doi.org/10.1007/978-0-387-39940-9\\_430](https://doi.org/10.1007/978-0-387-39940-9_430)
137. **Liu, B. (2012).** *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies, 5, 1-167. Διαθέσιμο στο: <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
138. **Liu, C. & Liu, Q. (2018).** *Community Detection Based on Differential Evolution Using Modularity Density*. *Information*, 9(9), 218. Διαθέσιμο στο: <https://doi.org/10.3390/info9090218>
139. **Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2020).** *RoBERTa: A robustly optimized BERT pretraining approach*. Διαθέσιμο στο: <https://arxiv.org/abs/1907.11692>
140. **Lloyd, S.P. (1982).** *Least Squares Quantization in PCM*. *IEEE Transactions on Information Theory*, 28, 129-137. Διαθέσιμο στο: <https://doi.org/10.1109/TIT.1982.1056489>
141. **Lu, J., Liu, A., Dong, F., Gu, F., Gama, J. & Zhang, G. (2019).** *Learning under Concept Drift: A Review*. *IEEE Transactions on Knowledge and Data Engineering*, 31(12), 2346–2363. Διαθέσιμο στο: <https://doi.org/10.1109/TKDE.2018.2876857>
142. **Lundberg, S.M. & Lee, S.I. (2017).** *A Unified Approach to Interpreting Model Predictions*. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 4768-4777. Διαθέσιμο στο: <https://dl.acm.org/doi/pdf/10.5555/3295222.3295230>
143. **McCulloch, W.S. & Pitts, W.H. (1943).** *A Logical Calculus of the Ideas Immanent in Nervous Activity*. *Bulletin of Mathematical Biophysics*, 5, 115-133. Διαθέσιμο στο: <http://dx.doi.org/10.1007/BF02478259>
144. **MacQueen, J.B. (1967).** *Some Methods for Classification and Analysis of Multivariate Observations*. In: *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, University of California Press, Berkeley, 281-297. Διαθέσιμο στο: <http://projecteuclid.org/euclid.bsmsp/1200512992>
145. **Medhat, W., Hassan, A. & Korashy, H. (2014).** *Sentiment analysis algorithms and applications: A survey*. *Ain Shams Engineering Journal*, 5(4), 1093–1113. Διαθέσιμο στο: <https://doi.org/10.1016/j.asej.2014.04.011>
146. **Mercer, J. (1909).** *Xvi. Functions of Positive and Negative Type and Their Connection the Theory of Integral Equations*. *Philosophical Transactions of the Royal Society of London. Series A*, 209, 415-446. Διαθέσιμο στο: <https://doi.org/10.1098/rsta.1909.0016>

147. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019). *Model Cards for Model Reporting*. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 220–229). Διαθέσιμο στο: <https://doi.org/10.1145/3287560.3287596>
148. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. & Hassabis, D. (2015). *Human-Level Control through Deep Reinforcement Learning*. Nature, 518, 529-533. Διαθέσιμο στο: <https://doi.org/10.1038/nature14236>
149. Müller, M. & Salathé, M. (2019). *Crowdbreaks: Tracking health trends using public social media data and crowdsourcing*. Frontiers in Public Health, 8, 375. Διαθέσιμο στο: <https://doi.org/10.3389/fpubh.2019.00081>
150. Murthy, D. (2008). *Digital ethnography: An examination of the use of new technologies for social research*. Sociology, 42(5), 837–855. Διαθέσιμο στο: <https://doi.org/10.1177/0038038508094565>
151. Mystakidis, S. (2022). *Metaverse*. Encyclopedia, 2, 486-497. Διαθέσιμο στο: <https://doi.org/10.3390/encyclopedia2010031>
152. Nasteski, V. (2017). *An overview of the supervised machine learning methods*. HORIZONS.B. 4. 51-62. Διαθέσιμο στο: <https://doi.org/10.20544/HORIZONS.B.04.1.17.P05>
153. Newman, M. E. J. (2006). *Modularity and community structure in networks*. Proc Natl Acad Sci USA 103: 8577-8582. Proceedings of the National Academy of Sciences of the United States of America. 103. 8577-82. Διαθέσιμο στο: <https://doi.org/10.1073/pnas.0601602103>
154. Newman, D., Lau, J. H., Grieser, K. & Baldwin, T. (2010). *Automatic evaluation of topic coherence*. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL, 100–108. Διαθέσιμο στο: <https://aclanthology.org/N10-1012.pdf>
155. Ng, R & Han, J. (2002). *CLARANS: A method for clustering objects for spatial data mining*. Knowledge and Data Engineering, IEEE Transactions on. 14. 1003-1016. Διαθέσιμο στο: <https://doi.org/10.1109/TKDE.2002.1033770>
156. Nguyen, A., Yosinski, J. & Clune, J. (2014). *Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 427–436. Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.1412.1897>
157. Nguyen, T. T., Nguyen, C. M., Nguyen, D. T., Nguyen, D. T. & Nahavandi, S. (2019). *Deep learning for deepfakes creation and detection: A survey*. Computers & Security, 102, 101746. Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.1909.11573>
158. Okoli, C. & Schabram, K. (2010). *A Guide to Conducting a Systematic Literature Review of Information Systems Research*. Sprouts: Working Papers on Information Systems, 10, 1-49. Διαθέσιμο στο: <https://doi.org/10.2139/ssrn.1954824>

159. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). *Training language models to follow instructions with human feedback*. In Advances in Neural Information Processing Systems, 35, 27730–27744. Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.2203.02155>
160. Pacheco, D., Hui, P.-M., Torres-Lugo, C., Truong, B., Flammini, A. & Menczer, F. (2021). *Uncovering Coordinated Networks on Social Media: Methods and Case Studies*. Proceedings of the International AAAI Conference on Web and Social Media. 15. 455-466. Διαθέσιμο στο: <https://doi.org/10.1609/icwsm.v15i1.18075>
161. Page, L., Brin, S., Motwani, R. & Winograd, T. (1998). *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report SIDL-WP-1999-0120, Stanford Digital Library Technologies Project. Διαθέσιμο στο: <https://www.cis.upenn.edu/~mkearns/teaching/NetworkedLife/pagerank.pdf>
162. Pak, A. & Paroubek, P. (2010). *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*. In Proceedings of LREC 2010 (pp. 1320–1326). Διαθέσιμο στο: [http://www.lrec-conf.org/proceedings/lrec2010/pdf/385\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf)
163. Pan, Y., Chen, Z., Suzuki, Y., Fukumoto, F. & Nishizaki, Hi. (2020). *Sentiment analysis using semi-supervised learning with few labeled data*. 231-234. Διαθέσιμο στο: <https://doi.org/10.1109/CW49994.2020.00044>
164. Pang, B. & Lee, L. (2008). *Opinion Mining and Sentiment Analysis*. Foundations and Trends® in Information Retrieval, 2, 1-135. Διαθέσιμο στο: <https://doi.org/10.1561/15000000011>
165. Park, H.S. & Jun, C.H. (2009). *A Simple and Fast Algorithm for k-Medoids Clustering*. Expert Systems with Applications, 36, 3336-3341. Διαθέσιμο στο: <https://doi.org/10.1016/j.eswa.2008.01.039>
166. Parkhi, O.M., Vedaldi, A. & Zisserman, A. (2015). *Deep Face Recognition*. Proceedings of the British Machine Vision Conference (BMVC). Διαθέσιμο στο: <https://doi.org/10.5244/c.29.41>
167. Patil, S. & Mailcontractor, R. (2024). *Impact of AI and Machine Learning on Financial Services*. ITM Web of Conferences. 68. Διαθέσιμο στο: <https://doi.org/10.1051/itmconf/20246801021>
168. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. (2011). *Scikit-Learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825-2830. Διαθέσιμο στο: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
169. Poria, S., Cambria, E., Hazarika, D. & Vij, P. (2016). *A deeper look into sarcastic tweets using deep convolutional neural networks*. COLING, 1601–1612. Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.1610.08815> ,

170. **Powers, D.M.W. (2011).** *Evaluation: From Precision, Recall and F-Measure to Roc, Informedness, Markedness & Correlation*. Journal of Machine Learning Technologies, 2, 37-63. Διαθέσιμο στο: <https://arxiv.org/pdf/2010.16061>
171. **Quinlan, J.R. (1986).** Induction of Decision Trees. Machine Learning, 1, 81-106. Διαθέσιμο στο: <http://dx.doi.org/10.1007/BF00116251>
172. **Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021).** *Learning Transferable Visual Models From Natural Language Supervision*. arXiv preprint arXiv:2103.00020. Διαθέσιμο στο: <https://arxiv.org/abs/2103.00020>
173. **Rajaraman, A. & Ullman, J.D. (2011).** *Mining of Massive Datasets*. Cambridge University Press, Cambridge. Διαθέσιμο στο: <http://dx.doi.org/10.1017/cbo9781139058452>
174. **Recht, B., Roelofs, R., Schmidt, L. & Shankar, V. (2019).** *Do ImageNet Classifiers Generalize to ImageNet?* International Conference on Machine Learning (ICML), 5389–5400. Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.1902.10811>
175. **Ribeiro, M.T., Singh, S. & Guestrin, C. (2016).** *"Why Should I Trust You?" Explaining the Predictions of Any Classifier*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, 13-17 August 2016, 1135-1144. Διαθέσιμο στο: <https://doi.org/10.1145/2939672.2939778>
176. **Ribeiro, F. Benevenuto, F. & Zagheni, E. (2020).** *How Biased is the Population of Facebook Users? Comparing the Demographics of Facebook Users with Census Data to Generate Correction Factors*. 325-334. Διαθέσιμο στο: <https://doi.org/10.1145/3394231.3397923>
177. **Riccio, C., Menanno, M., Zennaro, I. & Savino, M. M. (2024).** *A New Methodological Framework for Optimizing Predictive Maintenance Using Machine Learning Combined with Product Quality Parameters*. Machines, 12(7), 443. Διαθέσιμο στο: <https://doi.org/10.3390/machines12070443>
178. **Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N. & Huang, R. (2013).** *Sarcasm as Contrast between a Positive Sentiment and Negative Situation*. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP), 704–714. Διαθέσιμο στο: <https://aclanthology.org/D13-1066.pdf>
179. **Rodríguez-Ibáñez, M., Casanez-Ventura, A., Castejón-Mateos, F. & Cuenca-Jiménez, P-M. (2023).** *A Review on Sentiment Analysis from Social Media Platforms*. Expert Systems with Applications. 223. 119862. Διαθέσιμο στο: <https://doi.org/10.1016/j.eswa.2023.119862>
180. **Rosenblatt, F. (1958).** *The perceptron: A probabilistic model for information storage and organization in the brain*. Psychological Review, 65(6), 386–408. Διαθέσιμο στο: <https://doi.org/10.1037/h0042519>
181. **Ross, T.J. (2010).** *Fuzzy Logic with Engineering Applications*. 3rd Edition. Wiley, Hoboken. Διαθέσιμο στο: <http://dx.doi.org/10.1002/9781119994374>
182. **Rubin, V. L., Chen, Y. & Conroy, N. J. (2016).** *Deception detection for news: Three types of fakes*. Proceedings of the Association for Information Science and

- Technology, 52(1), 1–4. Διαθέσιμο στο: <https://doi.org/10.1002/praz.2015.145052010083>
183. **Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986).** *Learning representations by back-propagating errors.* Nature, 323(6088), 533–536. Διαθέσιμο στο: <https://doi.org/10.1038/323533a0>
  184. **Saito, T. & Rehmsmeier, M. (2015).** *The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets.* PLoS ONE. Διαθέσιμο στο: <https://doi.org/10.1371/journal.pone.0118432>
  185. **Sanh, V., Debut, L., Chaumond, J. & Wolf, T. (2019).** *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.* arXiv preprint arXiv:1910.01108. Διαθέσιμο στο: <https://arxiv.org/abs/1910.01108>
  186. **Salazar, C., Montoya-Múnera, E. & Aguilar, J. (2022).** *Sentiment analysis in learning resources.* Journal of Computers in Education. Διαθέσιμο στο: <https://doi.org/10.101007/s40692-022-00237-9>
  187. **Sarwar, B., Karypis, G., Konstan, J. & Riedl, J. (2001).** *Item-Based Collaborative Filtering Recommendation Algorithms.* Proceedings of the 10th International Conference on World Wide Web, Hong Kong, 1-5 May 2001, 285-295. Διαθέσιμο στο: <https://doi.org/10.1145/371920.372071>
  188. **Schmidhuber, J. (2015).** *Deep Learning in Neural Networks: An Overview.* Neural Networks, 61, 85-117. Διαθέσιμο στο: <https://doi.org/10.1016/j.neunet.2014.09.003>
  189. **Schmidt, A. & Wiegand, M. (2017).** *A Survey on Hate Speech Detection using Natural Language Processing.* Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media, 1–10. Διαθέσιμο στο: <https://doi.org/10.18653/v1/W17-1101>
  190. **Schubert, E., Zimek, A. & Kriegel, H. P. (2012).** *Local outlier detection reconsidered: a generalized view on locality with applications to spatial clustering.* Data Mining and Knowledge Discovery, 28(1), 190–237. Διαθέσιμο στο: <https://doi.org/10.1007/s10618-012-0300-z>
  191. **Schubert, E., Sander, J., Ester, M., Kriegel, H. P. & Xu, X. (2017).** *DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN.* ACM Transactions on Database Systems (TODS), 42(3), 1–21. Διαθέσιμο στο: <https://doi.org/10.1145/3068335>
  192. **Shah, E. (2024).** *Ethical Considerations in Use of Artificial Intelligence in Digital Marketing.* Journal of Peace Development & Communication. Volume 08. 340-351. Διαθέσιμο στο: <https://doi.org/10.36968/JPCDC-V08-I02-25>
  193. **Shen, D., Wu, G. & Suk, H. I. (2017).** *Deep learning in medical image analysis.* Annual Review of Biomedical Engineering, 19, 221–248. Διαθέσιμο στο: <https://doi.org/10.1146/annurev-bioeng-071516-044442>
  194. **Shu, K., Sliva, A., Wang, S., Tang, J. & Liu, H. (2017).** *Fake News Detection on Social Media: A Data Mining Perspective.* ACM SIGKDD Explorations

- Newsletter, 19(1), 22–36. Διαθέσιμο στο: <https://doi.org/10.1145/3137597.3137600>
195. **Specht, D.F. (1991).** *A General Regression Neural Network*. IEEE Transactions Neural Network, 2, 568-576. Διαθέσιμο στο: <http://dx.doi.org/10.1109/72.97934>
  196. **Subrahmanian, V. S., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., ... & Menczer, F. (2016).** *The DARPA Twitter bot challenge*. Computer, 49(6), 38–46. Διαθέσιμο στο: <https://doi.org/10.1109/MC.2016.183>
  197. **Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019).** *How to Fine-Tune BERT for Text Classification?*. Proceedings of the China National Conference on Chinese Computational Linguistics, 194–206. Διαθέσιμο στο: <https://arxiv.org/abs/1905.05583>
  198. **Syed, A. S., Sierra-Sosa, D., Kumar, A. & Elmaghraby, A. (2021).** *IoT in Smart Cities: A Survey of Technologies, Practices and Challenges*. Smart Cities, 4(2), 429-475. Διαθέσιμο στο: <https://doi.org/10.3390/smartcities4020024>
  199. **Taboada, M., Brooke, J., Tofiloski, M., Voll, K. & Stede, M. (2011).** *Lexicon-based methods for sentiment analysis*. Computational Linguistics, 37(2), 267–307. Διαθέσιμο στο: [https://doi.org/10.1162/COLI\\_a\\_00049](https://doi.org/10.1162/COLI_a_00049)
  200. **Taherdoost, H. (2023).** *Enhancing Social Media Platforms with Machine Learning Algorithms and Neural Networks*. Algorithms, 16(6), 271. Διαθέσιμο στο: <https://doi.org/10.3390/a16060271>
  201. **Tay, Y., Dehghani, M., Bahri, D. & Metzler, D. (2020).** *Efficient Transformers: A Survey*. Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.2009.06732>
  202. **Tian, Y., Hettiarachchi, D. & Kamijo, S. (2022).** *Transportation Mode Detection Combining CNN and Vision Transformer with Sensors Recalibration Using Smartphone Built-In Sensors*. Sensors, 22(17), 6453. Διαθέσιμο στο: <https://doi.org/10.3390/s22176453>
  203. **Townsend, L. & Wallace, C. (2016).** *Social media research: A guide to ethics*. University of Aberdeen. Διαθέσιμο στο: [https://www.gla.ac.uk/media/Media\\_487729\\_smxx.pdf](https://www.gla.ac.uk/media/Media_487729_smxx.pdf)
  204. **Tsai, Y. H. H., Bai, S., Yamada, M., Morency, L. P. & Salakhutdinov, R. (2019).** *Multimodal transformer for unaligned multimodal language sequences*. In Proceedings of ACL 2019 (pp. 6558–6569). Διαθέσιμο στο: <https://doi.org/10.48550/arXiv.1906.00295>
  205. **Tsur, O. & Rappoport, A. (2012).** *What's in a Hashtag? Content based Prediction of the Spread of Ideas in Microblogging Communities*. WSDM 2012 - Proceedings of the 5th ACM International Conference on Web Search and Data Mining. 643-652. Διαθέσιμο στο: <https://doi.org/10.1145/2124295.2124320>
  206. **Tufekci, Z. (2014).** *Big questions for social media big data: Representativeness, validity and other methodological pitfalls*. Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, 505–514. Διαθέσιμο στο: <https://ojs.aaai.org/index.php/ICWSM/article/view/14517>

207. **Tumasjan, A., Sprenger, T. O., Sandner, P. G. & Welpe, I. M. (2010).** *Predicting elections with Twitter: What 140 characters reveal about political sentiment.* Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Washington DC. Διαθέσιμο στο: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1441>
208. **Vapnik, V.N. (1995).** *The Nature of Statistical Learning Theory.* Springer, New York. Διαθέσιμο στο: <http://dx.doi.org/10.1007/978-1-4757-2440-0>
209. **Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017).** *Attention is All You Need.* In Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS 2017) (pp. 6000–6010). Διαθέσιμο στο: <https://arxiv.org/abs/1706.03762>
210. **Vesanto, J. & Alhoniemi, E. (2000).** *Clustering of the Self-Organizing Map.* IEEE Transactions on Neural Networks, 11, 586–600. Διαθέσιμο στο: <https://doi.org/10.1109/72.846731>
211. **Viktoratos, Iosif & Tsadiras, Athanasios. (2021).** *Personalized Advertising Computational Techniques: A Systematic Literature Review.* Findings, and a Design Framework. Information. 12. 480. Διαθέσιμο στο: <https://doi.org/10.3390/info12110480>
212. **Vinyals, O., Toshev, A., Bengio, S. & Erhan, D. (2015).** *Show and tell: A neural image caption generator.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3156–3164. Διαθέσιμο στο: <https://doi.org/10.1109/CVPR.2015.7298935>
213. **Vosoughi, S. Roy, D., & Aral, S. (2018).** *The spread of true and false news online.* Science, 359(6380), 1146–1151. Διαθέσιμο στο: <https://doi.org/10.1126/science.aap9559>
214. **Wang, H., Can, D., Kazemzadeh, A., Bar, F. & Narayanan, S. (2012).** *A system for real-time Twitter sentiment analysis of 2012 US presidential election cycle.* Proceedings of the ACL 2012 System Demonstrations, 115–120. Διαθέσιμο στο: <https://aclanthology.org/P12-3020.pdf>
215. **Wasserman, S. & Faust, K. (1994).** *Social network analysis: Methods and applications.* Cambridge University Press. Διαθέσιμο στο: <https://doi.org/10.1017/CBO9780511815478>
216. **Webster, J. & Watson, R. T. (2002).** *Analyzing the Past to Prepare for the Future: Writing a Literature Review.* MIS Quarterly, 26(2), xiii–xxiii. Διαθέσιμο στο: <http://www.jstor.org/stable/4132319>
217. **Weinberger, K. Q., Blitzer, J. & Saul, L. K. (2006).** *Distance Metric Learning for Large Margin Nearest Neighbor Classification.* Advances in Neural Information Processing Systems (NIPS). Διαθέσιμο στο: <https://jmlr.csail.mit.edu/papers/volume10/weinberger09a/weinberger09a.pdf>

218. **Xu, D. & Tian, Y.A. (2015).** *Comprehensive Survey of Clustering Algorithms.* Annals of Data Science, 2, 165-193. Διαθέσιμο στο: <http://dx.doi.org/10.1007/s40745-015-0040-1>
219. **Yang, Y., Zhou, F., Samsurijan, M. S. & Ma, R. (2025).** *Recognizing fake news spreading in social networks using ensemble learning.* Journal of Information Technology & Politics, 1–19. Διαθέσιμο στο: <https://doi.org/10.1080/19331681.2025.2458686>
220. **Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. (2014).** *How Transferable Are Features in Deep Neural Networks?* In: Advances in Neural Information Processing Systems, Springer, New York, 3320-3328. arXiv preprint. Διαθέσιμο στο: <https://arxiv.org/abs/1411.1792>
221. **Yu, M., Liu, T., Yin, J. & Chai, P. (2022).** *Deep Interest Context Network for Click-Through Rate.* Applied Sciences, 12(19), 9531. Διαθέσιμο στο: <https://doi.org/10.3390/app12199531>
222. **Zadeh, L.A. (1965).** *Fuzzy Sets.* Information Control, 8, 338-353. Διαθέσιμο στο: [http://dx.doi.org/10.1016/S0019-9958\(65\)90241-X](http://dx.doi.org/10.1016/S0019-9958(65)90241-X)
223. **Zhang, G., Patuwo, B.E. & Hu, M.Y. (1998).** *Forecasting with Artificial Neural Networks: The State of the Art.* International Journal of Forecasting, 14, 35-62. Διαθέσιμο στο: [http://dx.doi.org/10.1016/S0169-2070\(97\)00044-7](http://dx.doi.org/10.1016/S0169-2070(97)00044-7)
224. **Zhang, M. L. & Zhou, Z. H. (2007).** ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognition, 40(7), 2038–2048. Διαθέσιμο στο: <https://doi.org/10.1016/j.patcog.2006.12.019>
225. **Zhang, Z., Robinson, D. & Tepper, J. (2018).** *Detecting Hate Speech on Twitter Using a Convolution-GRU Based Deep Neural Network.* ESWC 2018 Workshops, 745–760. Διαθέσιμο στο: [https://doi.org/10.1007/978-3-319-93417-4\\_48](https://doi.org/10.1007/978-3-319-93417-4_48)
226. **Zhang, C. & Lu, Y. (2021).** *Study on Artificial Intelligence: The State of the Art and Future Prospects.* Journal of Industrial Information Integration. 23. 100224. Διαθέσιμο στο: <https://doi.org/10.1016/j.jii.2021.100224>
227. **Zhao, W., Jiang, J., Weng, J., He, J., Lim, E-P., Yan, H. & Li, X. (2011).** *Comparing Twitter and Traditional Media Using Topic Models.* Advances in Information Retrieval. 6611/2011. 338-349. Διαθέσιμο στο: [https://doi.org/10.1007/978-3-642-20161-5\\_34](https://doi.org/10.1007/978-3-642-20161-5_34)
228. **Zhao, Q., Erdogdu, M. A., He, H. Y., Rajaraman, A. & Leskovec, J. (2015).** *SEISMIC: A Self-Exciting Point Process Model for Predicting Tweet Popularity.* Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1513–1522. Διαθέσιμο στο: <https://doi.org/10.1145/2783258.2783401>
229. **Zhou, X. & Zafarani, R. (2020).** *A survey of fake news: Fundamental theories, detection methods and opportunities.* ACM Computing Surveys, 53(5), 1–40. Διαθέσιμο στο: <https://doi.org/10.1145/3395046>
230. **Zhou, X., Jain, A., Phoha, V. V. & Zafarani, R. (2020).** *Fake news early detection: A theory-driven model.* Digital Threats: Research and Practice, 1(1), 1–25. Διαθέσιμο στο: <https://doi.org/10.1145/3377478>

231. **Zliobaite, I. (2017).** *Measuring discrimination in algorithmic decision making*. Data Mining and Knowledge Discovery, 31. 1-30. Διαθέσιμο στο: <https://doi.org/10.1007/s10618-017-0506-1>
232. **Zou, J. Y. & Schiebinger, L. (2018).** *AI can be sexist and racist — it's time to make it fair*. Nature, 559(7714), 324–326. Διαθέσιμο στο: <https://doi.org/10.1038/d41586-018-05707-8>

### Ιστογραφία (Web References)

1. **Bounteous. (2010).** *The Evolution of Social Media Analytics*. Διαθέσιμο στο: <https://www.bounteous.com/insights/2010/11/23/evolution-social-media-analytics/>
2. **Britannica. (2025).** *Social networking*. Encyclopædia Britannica. Διαθέσιμο στο: <https://www.britannica.com/topic/data-protection>
3. **Coursera. (2025).** *Machine Learning vs. AI: Differences, Uses and Benefits*. Διαθέσιμο στο: <https://www.coursera.org/articles/machine-learning-vs-ai>
4. **DataRobot. (2021).** *Statistics and machine learning: what's the difference?* Διαθέσιμο στο: <https://www.datarobot.com/blog/statistics-and-machine-learning-whats-the-difference/>
5. **DeepMind. (2022).** *Flamingo: a new model that can learn visual tasks with very little training data*. DeepMind Blog. Διαθέσιμο στο: <https://www.deepmind.com/blog/tackling-multiple-tasks-with-a-single-visual-language-model>
6. **Google. (2023).** *PaLM-E: An Embodied Multimodal Language Model*. Διαθέσιμο στο: <https://palm-e.github.io/>
7. **Google Cloud. (2025).** *What is Artificial Intelligence (AI)?* Διαθέσιμο στο: <https://cloud.google.com/learn/what-is-artificial-intelligence>
8. **Grazitti Interactive. (2023).** *How Data Analytics Empowers Your Digital Marketing*. Διαθέσιμο στο: <https://www.grazitti.com/blog/impact-of-data-analytics-on-digital-marketing/>
9. **Hex.Tech. (2023).** *When to Choose Density-Based Methods*. Διαθέσιμο στο: <https://hex.tech/blog/comparing-density-based-methods/>
10. **IBM. (2025).** *AI healthcare benefits*. Διαθέσιμο στο: <https://www.ibm.com/think/insights/ai-healthcare-benefits>
11. **NetSuite. (2024).** *16 Applications of Machine Learning in Manufacturing in 2024*. Διαθέσιμο στο: <https://www.netsuite.com/portal/resource/articles/erp/machine-learning-in-manufacturing.shtml>
12. **Our World in Data. (2019).** *The rise of social media*. Διαθέσιμο στο: <https://ourworldindata.org/rise-of-social-media>
13. **Scikit-learn Developers. (2025).** *sklearn.cluster.DBSCAN*. Διαθέσιμο στο: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html>

14. **TechTarget.** (2025). *What is social media analytics?* Διαθέσιμο στο: <https://www.techtarget.com/searchbusinessanalytics/definition/social-media-analytics>