

ΜΕΤΑΠΤΥΧΙΑΚΟ ΣΤΗ ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ

Αλγόριθμοι Πρόβλεψης Διαδικτυακής Συμπεριφοράς

Τριμελής Εξεταστική Επιτροπή:

Τσαφάρáκης Στυλιανός, Αναπληρωτής Καθηγητής

Δούμπος Μιχαήλ, Καθηγητής

Κρασάδάκη Ευαγγελία, ΕΔΙΠ

Δρίτσα Αναστασία

A.M.: 2022019015

Χανιά, 2024

MASTER IN BUSINESS ADMINISTRATION - MBA

Algorithms For Predicting Online Behavior

Chania, 2024

ΠΕΡΙΛΗΨΗ

Η επίγνωση της διαδικασίας της συμπεριφοράς των χρηστών στο χώρο του διαδικτύου αποτελεί βασικό στοιχείο του ψηφιακού μάρκετινγκ. Η εμβάθυνση της διαδικασίας έχει τη δυνατότητα ταξινόμησης των χρηστών με βάση την συμπεριφορά τους και μπορεί να αναλυθεί. Για το σκοπό αυτό, η εργασία αναλύει τον τρόπο με τον οποίο οι αλγόριθμοι: Λογιστική Παλινδρόμηση, Δέντρα Αποφάσεων, Τυχαία Δάση, Μηχανές Διανυσμάτων Υποστήριξης και Νευρωνικά Δίκτυα, χρησιμοποιούνται για την ανάλυση και την πρόβλεψη των ενεργειών των χρηστών στο διαδικτυακό περιβάλλον.

Αξιοποιώντας αυτούς τους αλγορίθμους, οι επιχειρήσεις έχουν τη δυνατότητα απόκτησης πολύτιμων πληροφοριών σχετικά με τις αλληλεπιδράσεις, τις προτιμήσεις και τις προθέσεις αγοράς των χρηστών, ενισχύοντας έτσι την αποτελεσματικότητα των στοχευμένων προσπαθειών μάρκετινγκ. Η μελέτη παρέχει συγκριτική ανάλυση των δυνατών σημείων και των περιορισμών κάθε αλγορίθμου, προσφέροντας πρακτικές οδηγίες για την επιλογή της καταλληλότερης μεθόδου για συγκεκριμένες εργασίες πρόβλεψης στην ψηφιακή αγορά. Τα ευρήματα τονίζουν τη σημασία των προηγμένων αναλυτικών τεχνικών για την κατανόηση και την επίδραση στη συμπεριφορά των καταναλωτών, οδηγώντας τελικά σε πιο εξατομικευμένες και αποτελεσματικές στρατηγικές μάρκετινγκ.

Λέξεις Κλειδιά: διαδικτυακή συμπεριφορά, αγοραστική συμπεριφορά, ψηφιακό μάρκετινγκ, αλγόριθμοι πρόβλεψης.

ABSTRACT

The awareness of the process of user behavior in the online space is a key element of digital marketing. Delving into this process allows for the categorization of users based on their behavior, which can then be analyzed. For this purpose, the study examines how algorithms such as Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, and Neural Networks are used to analyze and predict user actions in the online environment.

By utilizing these algorithms, businesses can gain valuable insights into user interactions, preferences, and purchase intentions, thereby enhancing the effectiveness of targeted marketing efforts. The study provides a comparative analysis of the strengths and limitations of each algorithm, offering practical guidance for selecting the most appropriate method for specific prediction tasks in the digital marketplace. The findings highlight the importance of advanced analytical techniques in understanding and influencing consumer behavior, ultimately leading to more personalized and effective marketing strategies.

Keywords: online behavior, purchasing behavior, digital marketing, predictive algorithms.

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω όσους στάθηκαν αρωγοί κατά τη διάρκεια του μεταπτυχιακού προγράμματος αλλά και της μεταπτυχιακής μου εργασίας.

Πρώτα και κύρια, ευχαριστώ τον επιβλέποντα καθηγητή μου κ. Στέλιο Τσαφάρáκη, καθώς καθ' όλη τη διάρκεια της διπλωματικής εργασίας ήταν καθοδηγητής μου, υπήρχε άμεση επικοινωνία αλλά και μου δόθηκε η ευκαιρία να μελετήσω ένα ενδιαφέρον θέμα με τη βοήθεια και τις γνώσεις του.

Τέλος, είμαι ευγνώμων στους γονείς μου γιατί χωρίς τη βοήθεια τους, την παρότρυνσή τους και την εμπιστοσύνη τους δεν θα τα είχα καταφέρει αλλά και στους φίλους μου που ήταν δίπλα μου κάθε στιγμή.

*Αναστασία Δρίτσα
Χανιά, 2024*

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

- **TCP / IP:** Transmission Control Protocol / Internet Protocol – Πρωτόκολλο Ελέγχου Μετάδοσης / Πρωτόκολλο Διαδικτύου
- **WWW:** World Wide Web – Παγκόσμιος Ιστός
- **URL:** Uniform Resource Locators – Ενιαίος Εντοπιστής Πόρων
- **HTTP:** Hypertext Transfer Protocol – Πρωτόκολλο Μεταφοράς Υπερκειμένου
- **HTML:** HyperText Markup Language - Γλώσσα Σήμανσης Υπερκειμένου
- **ISP:** Internet Service Provider – Πάροχος Υπηρεσιών Διαδικτύου
- **FTP:** File Transfer Protocol – Πρωτόκολλο Μεταφοράς Αρχείου
- **AI:** Artificial Intelligence – Τεχνητή Νοημοσύνη (TN)
- **ML:** Machine Learning – Μηχανική Μάθηση
- **SVM:** Support Vector Machines – Μηχανή Διανυσμάτων Υποστήριξης
- **RBF:** Radial Basis Function – Συνάρτηση Ακτινικής Βάσης
- **ReLU:** Rectified Linear Unit – Διορθωμένη Γραμμική Μονάδα
- **DNN:** Deep Neural Networks – Βαθιά Νευρωνικά Δίκτυα
- **K-NN:** K – Nearest Neighbors Algorithm – Αλγόριθμος Πλησιέστερου Γείτονα
- **ROC:** Receiver Operating Characteristic – Χαρακτηριστική Καμπύλη Λειτουργίας Δέκτη
- **DSP:** DemandSide Platform – Πλατφόρμα Ζήτησης
- **CTR:** Click – Through Rate – Αναλογία Κλικ Προς Αριθμό Εμφανίσεων
- **HWRf:** Hybrid Weighted Random Forest – Υβριδικό Σταθμισμένο Τυχαίο Δάσος
- **MLP:** Multi-Layer Perception – Πολυεπίπεδος Αντιληπτικός Νευρώνας
- **MNN:** Multi – Layer Neural Network – Πολυεπίπεδο Νευρωνικό Δίκτυο
- **ANN:** Artificial Neural Networks – Τεχνητά Νευρωνικά Δίκτυα
- **RNN:** Recurrent Neural Networks – Επαναλαμβανόμενα (Αναδρομικά) Νευρωνικά Δίκτυα
- **LSTM:** Long Short-Term Memory – Δίκτυα Μακράς Βραχύχρονης Μνήμης
- **DBN:** Deep Belief Network – Δίκτυο Βαθιάς Πεποίθησης
- **SDA:** Stacked Denoising Auto-Encoders – Αυτόματοι Κωδικοποιητές Stacked Denoising
- **GRU:** Gated Recurrent Unit – Περιφραγμένη Επαναλαμβανόμενη Μονάδα
- **BiRNN:** Bidirectional Recurrent Neural Network – Δικτυακά Νευρωνικά Δίκτυα Διπλής Κατεύθυνσης

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΛΗΨΗ.....	iii
ABSTRACT	iv
ΕΥΧΑΡΙΣΤΙΕΣ.....	v
ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ	vi
ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ	1
1.1. Το Διαδίκτυο	1
1.2. Διαδικτυακή-Αγοραστική Συμπεριφορά του Καταναλωτή	3
1.3. Μάρκετινγκ	7
1.3.1. Στοχευμένο Μάρκετινγκ	9
1.4 Αλγόριθμοι Πρόβλεψης Διαδικτυακής Συμπεριφοράς	11
1.5 Στόχοι της Έρευνας	14
1.6 Δομή της Έρευνας	15
ΚΕΦΑΛΑΙΟ 2: ΑΝΑΛΥΣΗ ΑΛΓΟΡΙΘΜΩΝ ΠΡΟΒΛΕΨΗΣ	16
2.1 Λογιστική Παλινδρόμηση - Logistic Regression	16
2.1.1 Πλεονεκτήματα	18
2.1.2 Περιορισμοί	18
2.1.3 Εφαρμογές στην Πρόβλεψη της Διαδικτυακής Συμπεριφοράς.....	19
2.2 Δέντρα Αποφάσεων - Decision Tree	19
2.2.1 Πλεονεκτήματα	21
2.2.2 Περιορισμοί	22
2.2.3 Εφαρμογές στην Πρόβλεψη της Διαδικτυακής Συμπεριφοράς.....	23
2.3 Τυχαία Δάση - Random Forests.....	23
2.3.1 Πλεονεκτήματα	25
2.3.2 Περιορισμοί	26
2.3.3 Εφαρμογές στην Πρόβλεψη της Διαδικτυακής Συμπεριφοράς.....	26
2.4 Μηχανές Διαनुσμάτων Υποστήριξης (Support Vector Machines - SVM).....	27
2.4.1 Πλεονεκτήματα	29
2.4.2 Περιορισμοί	29
2.4.3 Εφαρμογές στην Πρόβλεψη της Διαδικτυακής Συμπεριφοράς.....	30

2.5 Νευρωνικά Δίκτυα - Neural Network	30
2.5.1 Πλεονεκτήματα	32
2.5.2 Περιορισμοί	33
2.5.3 Εφαρμογές στην Πρόβλεψη της Διαδικτυακής Συμπεριφοράς.....	33
2.6 Σύγκριση Αλγορίθμων Πρόβλεψης	33
 ΚΕΦΑΛΑΙΟ 3: ΕΦΑΡΜΟΓΗ ΤΩΝ ΑΛΓΟΡΙΘΜΩΝ ΣΤΗΝ ΠΡΟΒΛΕΨΗ ΔΙΑΔΙΚΤΥΑΚΗΣ	
ΣΥΜΠΕΡΙΦΟΡΑΣ	35
3.1 Λογιστική Παλινδρόμηση	35
3.1.1 Πρόβλεψη της Αγοραστικής Πρόθεσης των Online Αγοραστών	35
3.1.2 Ταξινόμηση των “Κλικ” σε Διαφημίσεις Πελατών.....	37
3.1.3 Στοχευμένη Διαδικτυακή Διαφήμιση.....	38
3.1.4 Πρόβλεψη της Διαδικτυακής Αγοραστικής Συμπεριφοράς	39
3.1.5 Αποτελέσματα της Λογιστικής Παλινδρόμησης στην Πρόβλεψη των Προθέσεων Αγοράς των Καταναλωτών	39
3.1.6 Υβριδικό Μοντέλο Πρόβλεψης.....	40
3.1.7 Εφαρμογές & Αποτελέσματα	41
3.2 Δέντρο Αποφάσεων	42
3.2.1 Πρόβλεψη Ποσοστού “Κλικ” Μέσω Δέντρου Αποφάσεων	42
3.2.2 Εκτίμηση της Πρόθεσης για Ηλεκτρονικές Αγορές.....	43
3.2.3 Πρόβλεψη Αποχώρησης Πελατών	44
3.2.4 Χρονική Ανάλυση της Διαδικτυακής Συμπεριφοράς των Καταναλωτών	45
3.2.5 Πρόβλεψη της Συμπεριφοράς των Καταναλωτών με τη Χρήση του Αλγορίθμου Δέντρων Αποφάσεων.....	45
3.2.6 Εφαρμογές & Αποτελέσματα	46
3.3 Εφαρμογές του Random Forest για την Πρόβλεψη Διαδικτυακής Συμπεριφοράς...47	
3.3.1 Πρόβλεψη Διαδικτυακής Αγοραστικής Συμπεριφοράς	47
3.3.2 Εκτίμηση της Αποχώρησης Πελατών.....	47
3.3.3 Πρόβλεψη του Ποσοστού “Κλικ” Διαδρομής	48
3.3.4 Ενίσχυση της Ακρίβειας Πρόβλεψης με Υβριδικά Μοντέλα	49
3.3.5 Πρόβλεψη των Προθέσεων Αγοράς των Καταναλωτών.....	49
3.3.6 Συγκριτική Ανάλυση της Λογιστικής Παλινδρόμησης και του Τυχαίου Δάσους.....	51
3.3.7 Εφαρμογές & Αποτελέσματα	52
3.4 Μηχανές Διανυσμάτων Υποστήριξης (SVM) για την Πρόβλεψη της Διαδικτυακής Συμπεριφοράς	53

3.4.1 Πρόβλεψη σε Πραγματικό Χρόνο της Αγοραστικής Πρόθεσης των Online Αγοραστών.....	53
3.4.2 Πρόβλεψη της Συμπεριφοράς των Πελατών στις Ηλεκτρονικές Αγορές	54
3.4.3 Υβριδικό Μοντέλο Πρόβλεψης για την Αγοραστική Συμπεριφορά	54
3.4.4 Συγκριτική Μελέτη των Τεχνικών Μηχανικής Μάθησης	55
3.4.5 Εφαρμογές & Αποτελέσματα	55
3.5 Νευρωνικά Δίκτυα για την Πρόβλεψη της Διαδικτυακής Συμπεριφοράς	56
3.5.1 Πρόβλεψη της Συμπεριφοράς των Πελατών με τη Χρήση Τεχνητών Νευρωνικών Δικτύων	56
3.5.2 Πρόβλεψη σε Πραγματικό Χρόνο της Αγοραστικής Πρόθεσης των Online Αγοραστών με τη Χρήση MLP και LSTM	57
3.5.3 Μείγμα Αναδρομικών Νευρωνικών Δικτύων για την Πρόβλεψη της Αγοραστικής Συμπεριφοράς	58
3.5.4 Προσεγγίσεις Βαθιάς Μάθησης για την Πρόβλεψη Αγοραστικών Προθέσεων	58
3.5.5 RNNs για Δεδομένα Ροής “Κλικ” στο Ηλεκτρονικό Εμπόριο	59
3.5.6 Μοντελοποίηση της Αγοραστικής Συμπεριφοράς με Βάσει τα Μοτίβα “Κλικ”	60
3.5.7 Πρόβλεψη της Συμπεριφοράς Διαδικτυακών Χρηστών με LSTM RNNs	60
3.5.8 Πρόβλεψη της Συμπεριφοράς των Καταναλωτών με ANNs	61
3.5.9 Εφαρμογές & Αποτελέσματα	62
3.6 Πραγματικές Εφαρμογές των Αλγορίθμων Πρόβλεψης	63
3.6.1 Το σύστημα σύστασης προϊόντων της Amazon	63
3.6.2 Αλγόριθμος σύστασης περιεχομένου του Netflix	64
ΚΕΦΑΛΑΙΟ 4: ΣΥΜΠΕΡΑΣΜΑΤΑ	66
4.1 Σύνοψη των Ευρημάτων	66
4.2 Σύγκριση Αλγορίθμων	69
4.3 Πρότυπα Συμπεριφοράς και Τάσεις στη Διαδικτυακή Συμπεριφορά	71
4.4 Εφαρμογές στο Ψηφιακό Μάρκετινγκ	72
4.5 Συνέπειες για τις Στρατηγικές Μάρκετινγκ	74
4.6 Μελλοντικές Ερευνητικές Κατευθύνσεις	76
ΒΙΒΛΙΟΓΡΑΦΙΑ	78

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

1.1. Το Διαδίκτυο

Το Διαδίκτυο είναι ένα παγκόσμιο σύστημα που διασυνδέει εκατομμύρια ιδιωτικά, δημόσια, ακαδημαϊκά, επιχειρηματικά και κυβερνητικά δίκτυα μέσω ενός τυποποιημένου συνόλου πρωτοκόλλων, με το TCP/IP (Transmission Control Protocol/Internet Protocol) να είναι το πιο αξιοσημείωτο από αυτά. Επιτρέπει διάφορες υπηρεσίες και λειτουργίες που είναι αναπόσπαστο κομμάτι της σύγχρονης ζωής, όπως ο Παγκόσμιος Ιστός, το ηλεκτρονικό ταχυδρομείο, η μεταφορά αρχείων και η άμεση ανταλλαγή μηνυμάτων [1].

Ουσιαστικά, το Διαδίκτυο είναι ένα σύνολο δικτύων που επιτρέπει σε διασυνδεδεμένους υπολογιστές να επικοινωνούν και να ανταλλάσσουν δεδομένα σε όλο τον κόσμο. Αυτή η συνεχής επικοινωνία διευκολύνεται από τα πρωτόκολλα, τα οποία είναι τυποποιημένοι κανόνες που εφαρμόζονται τόσο σε επίπεδο υλικού όσο και σε επίπεδο λογισμικού για να εξασφαλίζουν αποτελεσματική και αξιόπιστη ανταλλαγή δεδομένων. Ο βασικός πυλώνας αυτών των πρωτοκόλλων είναι το TCP/IP, το οποίο ρυθμίζει τον τρόπο μετάδοσης και λήψης των πακέτων δεδομένων, διασφαλίζοντας ότι οι πληροφορίες ρέουν ομαλά και φθάνουν με ακρίβεια στον προορισμό τους [2].

Μια από τις πιο σημαντικές υπηρεσίες που παρέχει το Διαδίκτυο είναι ο Παγκόσμιος Ιστός (WWW). Το WWW ξεκίνησε το 1994 και έφερε επανάσταση στην αλληλεπίδραση με τις πληροφορίες στο Διαδίκτυο. Αρχικά, το Διαδίκτυο αποτελούνταν κυρίως από κείμενο σε μαύρες οθόνες, αλλά το WWW εισήγαγε στοιχεία πολυμέσων και υπερσυνδέσμους, μετατρέποντας τις στατικές σελίδες σε δυναμικές, διασυνδεδεμένες πηγές. Οι χρήστες αποκτούν πρόσβαση σε αυτές τις ιστοσελίδες χρησιμοποιώντας URLs (Uniform Resource Locators), οι οποίες καθορίζουν τη διεύθυνση των πόρων στο Διαδίκτυο, διευκολύνοντας την ανάκτηση και την αλληλεπίδραση με πληροφορίες που είναι αποθηκευμένες σε διακομιστές ιστού [2].

Το HTTP (Hypertext Transfer Protocol) είναι το κύριο πρωτόκολλο που χρησιμοποιείται για τη μετάδοση ιστοσελίδων στο WWW. Επιτρέπει την ανάκτηση πόρων, όπως

έγγραφα HTML, εικόνες και άλλα αρχεία πολυμέσων, μέσω της δημιουργίας σύνδεσης μεταξύ του πελάτη (του προγράμματος περιήγησης του χρήστη) και του διακομιστή που φιλοξενεί το περιεχόμενο. Οι υπερσύνδεσμοι που ενσωματώνονται στις ιστοσελίδες επιτρέπουν στους χρήστες να πλοηγούνται ασταμάτητα μεταξύ σχετικών πληροφοριών σε διαφορετικές ιστοσελίδες και δικτυακούς τόπους, δημιουργώντας έναν ιστό διασυνδεδεμένων δεδομένων [3].

Οι πάροχοι υπηρεσιών Διαδικτύου (ISP) έχουν καθοριστική σημασία για την πρόσβαση των χρηστών στο Διαδίκτυο. Οι εταιρείες αυτές προσφέρουν την απαραίτητη υποδομή και τις υπηρεσίες που επιτρέπουν στους χρήστες να συνδεθούν στο Διαδίκτυο, παρέχοντας διάφορες ταχύτητες και πακέτα δεδομένων για την κάλυψη διαφορετικών αναγκών [1].

Ένα άλλο σημαντικό πρωτόκολλο είναι το FTP (File Transfer Protocol), το οποίο χρησιμοποιείται για τη μεταφορά αρχείων μεταξύ υπολογιστών στο Διαδίκτυο. Το FTP είναι ιδιαίτερα χρήσιμο για τη μεταφόρτωση και τη λήψη αρχείων από διακομιστές, γεγονός που το καθιστά απαραίτητο εργαλείο για τους προγραμματιστές ιστοσελίδων και τους επαγγελματίες πληροφορικής. Επιτρέπει την αποτελεσματική και ασφαλή μεταφορά μεγάλων αρχείων, συμβάλλοντας σημαντικά στις δραστηριότητες ανάπτυξης και συντήρησης του διαδικτύου [4].

Ο αντίκτυπος του Διαδικτύου στην επικοινωνία, τις επιχειρήσεις, την έρευνα και την κοινωνική αλληλεπίδραση είναι αρκετά μεγάλος. Έχει μεταμορφώσει τον τρόπο με τον οποίο οι άνθρωποι αλληλεπιδρούν, δραστηριοποιούνται και μοιράζονται τη γνώση, προωθώντας την κατανόηση της καινοτομίας σε διάφορους τομείς. Η ικανότητα του Διαδικτύου να συνδέει διαφορετικά δίκτυα και να παρέχει ένα πλήθος υπηρεσιών το έχει καταστήσει απαραίτητο μέρος της σύγχρονης ζωής, επηρεάζοντας διάφορες πτυχές της κοινωνίας και προωθώντας την καινοτομία σε πολλούς τομείς [5].

Το Διαδίκτυο υποστηρίζει επίσης διάφορα εργαλεία επικοινωνίας, όπως το ηλεκτρονικό ταχυδρομείο και τα άμεσα μηνύματα, τα οποία έχουν φέρει επανάσταση στην προσωπική και επαγγελματική επικοινωνία. Το ηλεκτρονικό ταχυδρομείο επιτρέπει στους χρήστες να στέλνουν και να λαμβάνουν μηνύματα άμεσα,

ανεξάρτητα από τη γεωγραφική τους θέση, ενώ τα άμεσα μηνύματα παρέχουν επικοινωνία σε πραγματικό χρόνο, ενισχύοντας τη συνεργασία και τη συνδεσιμότητα. Εκτός από την επικοινωνία, το Διαδίκτυο έχει επηρεάσει σημαντικά τον επιχειρηματικό κόσμο. Οι πλατφόρμες ηλεκτρονικού εμπορίου επιτρέπουν στις επιχειρήσεις να προσεγγίσουν ένα παγκόσμιο κοινό, προσφέροντας προϊόντα και υπηρεσίες “online” και διευκολύνοντας τις ασφαλείς συναλλαγές. Το Διαδίκτυο υποστηρίζει επίσης την απομακρυσμένη εργασία και την τηλεργασία, παρέχοντας εργαλεία και πλατφόρμες που επιτρέπουν στους εργαζόμενους να εργάζονται από οπουδήποτε, προωθώντας την ευελιξία και την παραγωγικότητα [6].

Στον τομέα της έρευνας και της εκπαίδευσης, το Διαδίκτυο παρέχει πρόσβαση σε τεράστιο όγκο πληροφοριών και πόρων. Οι διαδικτυακές βάσεις δεδομένων, τα περιοδικά και οι εκπαιδευτικές πλατφόρμες προσφέρουν στους εκπαιδευόμενους και τους ερευνητές απaráμιλλη πρόσβαση στη γνώση, προωθώντας τη συνεχή μάθηση και την καινοτομία [7].

Συνεπώς, το Διαδίκτυο αποτελεί ένα ισχυρό και εκτεταμένο παγκόσμιο δίκτυο που έχει φέρει επανάσταση στην επικοινωνία, τις επιχειρήσεις, την έρευνα και την κοινωνική αλληλεπίδραση. Η ικανότητά του να συνδέει διάφορα δίκτυα και να παρέχει πλήθος υπηρεσιών το έχει καταστήσει ουσιαστικό μέρος της σύγχρονης ζωής, επηρεάζοντας διάφορες πτυχές της κοινωνίας και προωθώντας την καινοτομία σε πολλούς τομείς. Η συνεχής ανάπτυξη και εξέλιξη του Διαδικτύου συνεχίζει να διαμορφώνει το μέλλον, προσφέροντας νέες ευκαιρίες και προκλήσεις καθώς μεταβαίνουμε σε μια όλο και περισσότερο ψηφιακή εποχή.

1.2. Διαδικτυακή-Αγοραστική Συμπεριφορά του Καταναλωτή

Η συμπεριφορά των καταναλωτών στο διαδίκτυο είναι ένα πολύπλευρο φαινόμενο που επηρεάζεται από διάφορους κοινωνικούς, πολιτιστικούς, προσωπικούς, ψυχολογικούς και περιβαλλοντικούς παράγοντες. Οι παράγοντες του μάρκετινγκ έχουν ουσιαστική σημασία για την κατανόηση αυτών των επιρροών, προκειμένου να αναπτύξουν αποτελεσματικές στρατηγικές για την επαφή με τους καταναλωτές στο ψηφιακό πεδίο.

Οικονομικοί Παράγοντες

Οι οικονομικοί παράγοντες διαδραματίζουν σημαντικό ρόλο στη διαμόρφωση της διαδικτυακής συμπεριφοράς των καταναλωτών. Τα εισοδήματα, είτε είναι απόλυτα, είτε σχετικά, είτε μόνιμα, επηρεάζουν άμεσα την ικανότητα ενός καταναλωτή να αγοράζει προϊόντα μέσω του διαδικτύου. Η διαθεσιμότητα πιστώσεων και περιουσιακών στοιχείων επηρεάζει περαιτέρω τις αποφάσεις αγοράς, καθώς οι καταναλωτές με περισσότερους οικονομικούς πόρους είναι πιθανό να ξοδέψουν περισσότερα για ηλεκτρονικές αγορές. Η τιμή παραμένει σημαντικός παράγοντας, οι καταναλωτές συχνά συγκρίνουν τις τιμές σε διάφορους ιστότοπους για να βρουν τις καλύτερες προσφορές. Η ύπαρξη υποκατάστατων προϊόντων επηρεάζει επίσης τη συμπεριφορά των διαδικτυακών αγορών, καθώς οι καταναλωτές μπορεί να επιλέξουν παρόμοια προϊόντα σε χαμηλότερες τιμές [8].

Δημογραφικοί Παράγοντες

Τα δημογραφικά χαρακτηριστικά, όπως η ηλικία, το φύλο, το μορφωτικό επίπεδο και η οικογενειακή δομή, επηρεάζουν επίσης τις διαδικτυακές καταναλωτικές συνήθειες. Για παράδειγμα, οι νεότεροι καταναλωτές, οι οποίοι είναι συνήθως πιο εξοικειωμένοι με την τεχνολογία, είναι πιο πιθανό να συμμετέχουν σε διαδικτυακές αγορές σε σύγκριση με τις παλαιότερες γενιές. Το μορφωτικό επίπεδο επηρεάζει τη διαδικτυακή συμπεριφορά επηρεάζοντας την πρόσβαση και την άνεση με την τεχνολογία. Επιπλέον, το στάδιο του κύκλου ζωής μιας οικογένειας, όπως η παρουσία και η ηλικία των παιδιών, μπορεί να υπαγορεύσει τα είδη των προϊόντων που αγοράζονται μέσω διαδικτύου. Οι οικογένειες με μικρά παιδιά μπορεί να δίνουν προτεραιότητα στην ευκολία και την ασφάλεια, με αποτέλεσμα να επιλέγουν τις ηλεκτρονικές αγορές [9].

Ψυχολογικοί Παράγοντες

Οι ψυχολογικοί παράγοντες, που έχουν τις ρίζες τους στις ατομικές ανάγκες και τα κίνητρα, είναι κεντρικής σημασίας για την κατανόηση της συμπεριφοράς των καταναλωτών στο διαδίκτυο. Σύμφωνα με τη φροϋδική ψυχολογία, οι ανάγκες που είναι θαμμένες στο υποσυνείδητο καθοδηγούν τη συμπεριφορά. Η ιεράρχηση των αναγκών του Maslow παρέχει επίσης εικόνα, κατηγοριοποιώντας τις ανθρώπινες ανάγκες από τις βασικές φυσιολογικές ανάγκες έως την αυτοπραγμάτωση. Η

συμπεριφορά των διαδικτυακών αγορών μπορεί να υποκινείται από την ανάγκη για ευκολία, ασφάλεια, κοινωνική ένταξη ή ακόμη και από την επιδίωξη του κύρους και της αυτοβελτίωσης. Οι στρατηγικές μάρκετινγκ που αξιοποιούν αυτές τις ψυχολογικές ανάγκες, όπως οι εξατομικευμένες συστάσεις και οι αποκλειστικές προσφορές, μπορούν να επηρεάσουν σημαντικά τις αποφάσεις των καταναλωτών [10].

Κοινωνικοί και Πολιτιστικοί Παράγοντες

Οι κοινωνικοί και πολιτιστικοί παράγοντες επηρεάζουν σε μεγάλο βαθμό τη διαδικτυακή συμπεριφορά των καταναλωτών. Οι καταναλωτές επηρεάζονται από την οικογένεια, τους φίλους, τους συναδέλφους και τα ευρύτερα κοινωνικά δίκτυα, συμπεριλαμβανομένων των διαδικτυακών κοινοτήτων και των ομάδων συζήτησης. Αυτές οι κοινωνικές αλληλεπιδράσεις συχνά διαμορφώνουν τις προτιμήσεις και τις αγοραστικές αποφάσεις των καταναλωτών. Το πολιτισμικό υπόβαθρο, που περιλαμβάνει τις παραδόσεις, τα έθιμα και τους κοινωνικούς κανόνες, επηρεάζει επίσης τη διαδικτυακή συμπεριφορά. Για παράδειγμα, οι πολιτισμικές στάσεις απέναντι στην τεχνολογία και το εμπόριο μπορούν να επηρεάσουν το βαθμό εμπιστοσύνης και την προθυμία συμμετοχής σε διαδικτυακές συναλλαγές. Η κατανόηση του πολιτισμικού πλαισίου είναι απαραίτητη για τους εμπόρους προκειμένου να προσαρμόσουν τις στρατηγικές τους σε διαφορετικά τμήματα καταναλωτών [11].

Προσωπικά Χαρακτηριστικά

Τα προσωπικά χαρακτηριστικά, όπως ο τρόπος ζωής, οι αξίες και τα χαρακτηριστικά της προσωπικότητας, διαμορφώνουν επίσης τη συμπεριφορά των διαδικτυακών καταναλωτών. Παράγοντες όπως η ηλικία, το φύλο, η οικογενειακή κατάσταση και το μορφωτικό επίπεδο συμβάλλουν σε ξεχωριστά προφίλ καταναλωτών. Για παράδειγμα, τα άτομα με πολυάσχολο τρόπο ζωής μπορεί να προτιμούν την ευκολία των ηλεκτρονικών αγορών, ενώ εκείνοι που εκτιμούν τη βιωσιμότητα μπορεί να αναζητούν φιλικά προς το περιβάλλον προϊόντα. Οι προσωπικές αξίες και τα χαρακτηριστικά της προσωπικότητας, όπως το άνοιγμα σε νέες εμπειρίες ή η αποστροφή του κινδύνου, επηρεάζουν περαιτέρω τις διαδικτυακές αγοραστικές

αποφάσεις. Οι έμποροι πρέπει να λάβουν υπόψη τους αυτά τα προσωπικά χαρακτηριστικά για να δημιουργήσουν στοχευμένες καμπάνιες που θα έχουν απήχηση στις διάφορες ομάδες καταναλωτών [12].

Περιβαλλοντικοί Παράγοντες

Οι περιβαλλοντικοί παράγοντες, συμπεριλαμβανομένων των διαθέσιμων πληροφοριών, των κυβερνητικών κανονισμών και των καταστασιακών παραγόντων, διαδραματίζουν καθοριστικό ρόλο στη διαμόρφωση της συμπεριφοράς των διαδικτυακών καταναλωτών. Η πρόσβαση σε αξιόπιστες πληροφορίες σχετικά με προϊόντα και υπηρεσίες, οι νομικοί περιορισμοί και οι καταστατικοί παράγοντες, όπως η πίεση του χρόνου ή η ευκολία, μπορούν να επηρεάσουν σημαντικά τις αποφάσεις αγοράς. Το ίδιο το διαδίκτυο, με τον τεράστιο όγκο πληροφοριών και το περιεχόμενο που δημιουργείται από τους χρήστες, χρησιμεύει ως κρίσιμη πηγή για τους καταναλωτές που λαμβάνουν ενημερωμένες αποφάσεις. Οι έμποροι πρέπει να διασφαλίζουν ότι οι ακριβείς και ολοκληρωμένες πληροφορίες είναι διαθέσιμες στους καταναλωτές, διευκολύνοντας την εμπιστοσύνη και τις τεκμηριωμένες επιλογές [8].

Ηλεκτρονικό Εμπόριο και Διαδικτυακές Κοινότητες

Στο πλαίσιο του ηλεκτρονικού εμπορίου, οι διαδικτυακές κοινότητες και τα προφίλ χρηστών έχουν ιδιαίτερη επιρροή. Οι ομάδες συζητήσεων, τα "chat rooms" και οι ηλεκτρονικοί πίνακες ανακοινώσεων παρέχουν στους καταναλωτές τη δυνατότητα να ανταλλάσσουν εμπειρίες και συστάσεις, επηρεάζοντας τις αποφάσεις αγοράς. Η κατανόηση της δυναμικής αυτών των διαδικτυακών κοινοτήτων μπορεί να βοηθήσει τους εμπόρους να αναπτύξουν στρατηγικές που εμπλέκουν αποτελεσματικά τους καταναλωτές. Επιπλέον, οι προσωπικές συνήθειες και τα προφίλ χρήσης του διαδικτύου, συμπεριλαμβανομένης της συχνότητας των ηλεκτρονικών αγορών και των προτιμήσεων σε πλατφόρμες, είναι απαραίτητα για το σχεδιασμό στοχευμένων εκστρατειών μάρκετινγκ που απευθύνονται σε συγκεκριμένες ανάγκες και προτιμήσεις των καταναλωτών [9].

Συμπερασματικά, η διαδικτυακή συμπεριφορά των καταναλωτών διαμορφώνεται από μια σύνθεση οικονομικών, δημογραφικών, ψυχολογικών, κοινωνικών,

πολιτιστικών, προσωπικών και περιβαλλοντικών παραγόντων. Οι έμποροι οφείλουν να κατανοήσουν αυτές τις διάφορες επιρροές για να αναπτύξουν αποτελεσματικές στρατηγικές που θα δεσμεύουν και θα ικανοποιούν τους καταναλωτές στην ψηφιακή αγορά.

1.3. Μάρκετινγκ

Το μάρκετινγκ είναι ένας κλάδος που περιλαμβάνει μια σειρά δραστηριοτήτων που αποσκοπούν στην προώθηση, πώληση και διανομή ενός προϊόντος ή μιας υπηρεσίας. Περιλαμβάνει στρατηγικές και τακτικές που χρησιμοποιούν οι επιχειρήσεις για να εντοπίσουν, να δημιουργήσουν και να διατηρήσουν ικανοποιητικές σχέσεις με τους πελάτες, στοχεύοντας τελικά στην ικανοποίηση των αναγκών και των επιθυμιών τους, επιτυγχάνοντας παράλληλα τους στόχους της επιχείρησης. Αυτός ο ευρύς ορισμός αποτυπώνει την ουσία του μάρκετινγκ, η οποία είναι η σύνδεση των προϊόντων και των υπηρεσιών με το προοριζόμενο κοινό τους κατά τρόπο που να ωφελεί τόσο τον πελάτη όσο και τον οργανισμό.

Η κατανόηση των βασικών εννοιών του μάρκετινγκ είναι απαραίτητη για την αποτελεσματική εφαρμογή του. Οι ανάγκες, οι επιθυμίες και οι απαιτήσεις είναι ουσιαστικής σημασίας για την κατανόηση αυτή. Οι ανάγκες είναι καταστάσεις έλλειψης αναγκαίες για την επιβίωση, ενώ οι επιθυμίες είναι η μορφή που παίρνουν οι ανθρώπινες ανάγκες καθώς διαμορφώνονται από τον πολιτισμό και την ατομική προσωπικότητα. Οι απαιτήσεις είναι επιθυμίες που υποστηρίζονται από αγοραστική δύναμη. Η κατανόηση αυτών των στοιχείων επιτρέπει στους εμπόρους να σχεδιάζουν προσφορές που ανταποκρίνονται στις προσδοκίες των καταναλωτών. Οι προσφορές της αγοράς είναι προϊόντα, υπηρεσίες, πληροφορίες ή εμπειρίες που προσφέρονται σε μια αγορά για να ικανοποιήσουν μια ανάγκη ή μια επιθυμία. Αυτές οι προσφορές είναι κεντρικές στη διαδικασία ανταλλαγής. Η έννοια της αξίας και της ικανοποίησης είναι επίσης σημαντική, όπου η αξία είναι τα αντιληπτά οφέλη σε σχέση με το κόστος απόκτησης αυτών των οφελών και η ικανοποίηση των πελατών είναι ο βαθμός στον οποίο η αντιληπτή απόδοση ενός προϊόντος ανταποκρίνεται στις προσδοκίες του αγοραστή. Το μάρκετινγκ πραγματοποιείται μέσω ανταλλαγών όπου δύο ή

περισσότερα μέρη δίνουν κάτι αξίας το ένα στο άλλο και η οικοδόμηση μακροπρόθεσμων σχέσεων είναι απαραίτητη για τη διαρκή επιχειρηματική επιτυχία [13].

Οι θεωρίες διαχείρισης του μάρκετινγκ παρέχουν διάφορες προσεγγίσεις για την ικανοποίηση των οργανωτικών και καταναλωτικών αναγκών. Η έννοια της παραγωγής επικεντρώνεται στη βελτίωση της αποτελεσματικότητας της παραγωγής και της διανομής και είναι πιο αποτελεσματική όταν η ζήτηση υπερβαίνει την προσφορά ή όταν η μείωση του κόστους είναι ζωτικής σημασίας για την επέκταση της αγοράς. Η έννοια του προϊόντος υποθέτει ότι οι καταναλωτές προτιμούν τα προϊόντα που προσφέρουν τη μεγαλύτερη ποιότητα, απόδοση και καινοτόμα χαρακτηριστικά, ωθώντας τις επιχειρήσεις να επικεντρωθούν στη συνεχή βελτίωση των προϊόντων. Η έννοια της πώλησης βασίζεται στην ιδέα ότι οι καταναλωτές δεν θα αγοράσουν αρκετά από τα προϊόντα της επιχείρησης αν αυτή δεν αναλάβει μια μεγάλης κλίμακας προσπάθεια πώλησης και προώθησης, η οποία συχνά εφαρμόζεται σε μη περιζήτητα αγαθά. Η έννοια του μάρκετινγκ υποστηρίζει ότι η επίτευξη των οργανωτικών στόχων εξαρτάται από τη γνώση των αναγκών και των επιθυμιών των αγορών-στόχων και την παροχή των επιθυμητών ικανοποιήσεων καλύτερα από ό,τι οι ανταγωνιστές. Τέλος, η έννοια του κοινωνικού μάρκετινγκ υποστηρίζει ότι μια εταιρεία πρέπει να λαμβάνει σωστές αποφάσεις μάρκετινγκ λαμβάνοντας υπόψη τις επιθυμίες των καταναλωτών, τις απαιτήσεις της εταιρείας και τα μακροπρόθεσμα συμφέροντα της κοινωνίας [14].

Το μείγμα μάρκετινγκ, το οποίο συχνά αναφέρεται ως τα 4P, περιλαμβάνει το προϊόν, την τιμή, τη θέση και την προώθηση. Το προϊόν αναφέρεται σε αυτό που προσφέρει η εταιρεία στον πελάτη, συμπεριλαμβανομένων των φυσικών αγαθών, των υπηρεσιών και των ιδεών. Η τιμή είναι το χρηματικό ποσό που πρέπει να πληρώσουν οι πελάτες για να αποκτήσουν το προϊόν, με τις στρατηγικές τιμολόγησης να ποικίλλουν σημαντικά ανάλογα με τις συνθήκες της αγοράς και τους επιχειρηματικούς στόχους. Ο τρόπος αναφέρεται στα κανάλια διανομής που χρησιμοποιούνται για να φτάσει το προϊόν στον πελάτη, εξασφαλίζοντας ότι τα προϊόντα είναι διαθέσιμα στις σωστές τοποθεσίες τη σωστή στιγμή. Η προώθηση περιλαμβάνει όλες τις δραστηριότητες που επικοινωνούν τα χαρακτηριστικά και τα οφέλη του προϊόντος και πείθουν τους πελάτες να το αγοράσουν. Περιλαμβάνει τη

διαφήμιση, την προώθηση των πωλήσεων, τις δημόσιες σχέσεις και τις προσωπικές πωλήσεις [15].

Η έλευση των ψηφιακών τεχνολογιών έχει μεταμορφώσει τις παραδοσιακές στρατηγικές μάρκετινγκ. Το ψηφιακό μάρκετινγκ, το μάρκετινγκ στα μέσα κοινωνικής δικτύωσης, το μάρκετινγκ περιεχομένου και η βελτιστοποίηση μηχανών αναζήτησης αποτελούν πλέον αναπόσπαστα στοιχεία του τοπίου του μάρκετινγκ. Τα εργαλεία αυτά επιτρέπουν στις επιχειρήσεις να αλληλεπιδρούν με τους πελάτες σε πραγματικό χρόνο, να εξατομικεύουν τα μηνύματα μάρκετινγκ και να μετρούν την αποτελεσματικότητα των εκστρατειών τους με ακρίβεια [16].

1.3.1. Στοχευμένο Μάρκετινγκ

Το στοχευμένο μάρκετινγκ είναι μια στρατηγική προσέγγιση που επικεντρώνεται στην προσέγγιση συγκεκριμένων ομάδων καταναλωτών εντός της ευρύτερης αγοράς για την αποτελεσματική προώθηση προϊόντων ή υπηρεσιών. Η μέθοδος αυτή βασίζεται σε λεπτομερή δεδομένα πελατών για τον εντοπισμό και τη δέσμευση ενός υποσυνόλου της συνολικής απευθυνόμενης αγοράς, με στόχο τη βελτίωση της επιχειρηματικής ανάπτυξης, των εσόδων και της απόδοσης της επένδυσης. Το στοχευμένο μάρκετινγκ ξεκινά με τον καθορισμό του "ποιος" είναι κατάλληλος για ένα προϊόν ή μια υπηρεσία, αντί να ξεκινά με το "τι" είναι το διαφημιστικό μήνυμα. Αυτό έρχεται σε αντίθεση με τις ευρύτερες στρατηγικές μάρκετινγκ, όπως η διαφήμιση σε διαφημιστικές πινακίδες, οι οποίες στοχεύουν στην προσέγγιση ενός ευρέος κοινού χωρίς συγκεκριμένη στόχευση. Κατανοώντας τα συγκεκριμένα δημογραφικά στοιχεία, τις συμπεριφορές και τις προτιμήσεις του κοινού-στόχου τους, οι έμποροι μπορούν να δημιουργήσουν εξατομικευμένα μηνύματα που έχουν βαθύτερη απήχηση, αυξάνοντας τις πιθανότητες μετατροπής [17].

Παραδοσιακά, τα κριτήρια στοχευμένου μάρκετινγκ περιλάμβαναν δημογραφικά δεδομένα όπως η ηλικία, το φύλο, το εισόδημα, η εκπαίδευση και το επάγγελμα, καθώς και άλλα δεδομένα όπως το μέγεθος της εταιρείας, ο κλάδος και τα έσοδα. Ωστόσο, με την άνοδο των ψηφιακών συσκευών και των προηγμένων αναλύσεων, έχουν προκύψει νέα κριτήρια. Η στόχευση βάσει προθέσεων εντοπίζει το κοινό που

δείχνει ενδιαφέρον για ένα θέμα, προϊόν ή υπηρεσία με βάση τις ψηφιακές δραστηριότητές του, όπως οι όροι αναζήτησης ή οι αλληλεπιδράσεις στα μέσα κοινωνικής δικτύωσης. Τα τεχνολογικά δεδομένα αφορούν τη στόχευση με βάση την τεχνολογία ή το λογισμικό που χρησιμοποιεί ένας καταναλωτής, όπως το λογισμικό προστασίας από ιούς ή συγκεκριμένες κατηγορίες υλικού. Τα αναλυτικά προερχόμενα κοινά χρησιμοποιούν μοντέλα μηχανικής μάθησης για τον εντοπισμό τμημάτων με υψηλή τάση αγοράς με βάση διάφορα χαρακτηριστικά και συμπεριφορές των καταναλωτών [17].

Αρκετές τεχνολογικές πλατφόρμες υποστηρίζουν στοχευμένες εκστρατείες μάρκετινγκ. Τα συστήματα διαχείρισης πελατειακών σχέσεων βοηθούν στη διαχείριση των αλληλεπιδράσεων και των δεδομένων των πελατών, παρέχοντας μια ολοκληρωμένη εικόνα των πελατειακών σχέσεων. Οι πλατφόρμες αυτοματοποίησης μάρκετινγκ βοηθούν στη δημιουργία αυτοματοποιημένων ροών ηλεκτρονικού ταχυδρομείου και τη φροντίδα των προοπτικών μέσω στοχευμένων εκστρατειών. Οι πλατφόρμες διαχείρισης δεδομένων αναλύουν ψηφιακά δεδομένα για τη δημιουργία διαφημιστικών τμημάτων και βοηθούν τους εμπόρους να στοχεύουν το κοινό σε όλο τον ιστό. Οι πλατφόρμες δεδομένων πελατών δημιουργούν μια ενιαία βάση δεδομένων πελατών με εύκολη πρόσβαση σε άλλα συστήματα, επιλύοντας προβλήματα σιλό δεδομένων και ενισχύοντας την τμηματοποίηση και την ανάλυση [18].

Το στοχευμένο μάρκετινγκ μπορεί να εκτελεστεί με αρκετούς τρόπους για την προώθηση της επιχειρηματικής ανάπτυξης. Για παράδειγμα, η διοργάνωση διαδικτυακών σεμιναρίων και εκδηλώσεων προσαρμοσμένων σε σημαντικές προοπτικές με σχετικούς ομιλητές και θέματα μπορεί να οικοδομήσει σχέσεις και αξιοπιστία. Η ανάπτυξη περιεχομένου που απευθύνεται στις συγκεκριμένες ανάγκες και προκλήσεις του κοινού-στόχου μπορεί να ενισχύσει τη δέσμευση και την εμπιστοσύνη. Η κατάρτιση εξατομικευμένων εκστρατειών ηλεκτρονικού ταχυδρομείου με μηνύματα που έχουν απήχηση στους παραλήπτες μπορεί να βελτιώσει τα ποσοστά μετατροπής. Η προσαρμογή της αρχικής σελίδας του ιστότοπου ώστε να αντικατοπτρίζει τα ενδιαφέροντα του επισκέπτη μπορεί να επηρεάσει σημαντικά τη δέσμευσή του και την πιθανότητα μετατροπής [18].

Το στοχευμένο μάρκετινγκ είναι πιο αποτελεσματικό επειδή επιτρέπει στις επιχειρήσεις να κατανοήσουν καλύτερα το κοινό τους, οδηγώντας σε πιο εξατομικευμένη και σχετική επικοινωνία. Αυτή η βαθύτερη κατανόηση επιτρέπει στους υπεύθυνους μάρκετινγκ να ξεπεράσουν τα προβλήματα και να τοποθετήσουν την επιχείρησή τους έναν αξιόπιστο πόρο. Προσαρμόζοντας τα μηνύματα σε συγκεκριμένα τμήματα κοινού, οι επιχειρήσεις μπορούν να επιτύχουν υψηλότερα επίπεδα δέσμευσης, καλύτερα ποσοστά μετατροπής και, τελικά, μεγαλύτερη αύξηση των εσόδων. Για παράδειγμα, η γνώση του ιστορικού αναζήτησης ή της διαδικτυακής συμπεριφοράς ενός υποψήφιου πελάτη επιτρέπει τη δημιουργία στοχευμένων διαφημίσεων που απευθύνονται άμεσα στις ανάγκες του, ενισχύοντας έτσι τη συνολική αποδοτικότητα και αποτελεσματικότητα του μάρκετινγκ [17].

Εν κατακλείδι, το στοχευμένο μάρκετινγκ αξιοποιεί τα δεδομένα των καταναλωτών και την προηγμένη τεχνολογία για την παροχή εξατομικευμένων μηνυμάτων μάρκετινγκ σε συγκεκριμένα τμήματα κοινού. Η προσέγγιση αυτή όχι μόνο αυξάνει την πιθανότητα μετατροπής αλλά και βελτιστοποιεί τις δαπάνες μάρκετινγκ, εξασφαλίζοντας υψηλότερη απόδοση επένδυσης και βιώσιμη επιχειρηματική ανάπτυξη.

1.4 Αλγόριθμοι Πρόβλεψης Διαδικτυακής Συμπεριφοράς

Τα τελευταία χρόνια, η ραγδαία ανάπτυξη του Διαδικτύου και του ηλεκτρονικού εμπορίου έχει προκαλέσει επαναστατικές αλλαγές στον τρόπο με τον οποίο οι καταναλωτές αλληλεπιδρούν με τις επιχειρήσεις, ιδίως σε περιβάλλοντα ηλεκτρονικών αγορών. Η ικανότητα πρόβλεψης της συμπεριφοράς των καταναλωτών έχει καταστεί απαραίτητη για τις επιχειρήσεις προκειμένου να παραμείνουν ανταγωνιστικές σε αυτή την ταχέως εξελισσόμενη, καθοδηγούμενη από δεδομένα αγορά [19]. Οι αλγόριθμοι πρόβλεψης, ιδίως εκείνοι που βασίζονται στην Τεχνητή Νοημοσύνη (AI) και τη Μηχανική Μάθηση (ML), αποτελούν πλέον απαραίτητα εργαλεία για την ανάλυση των αλληλεπιδράσεων των καταναλωτών και την πρόβλεψη των μελλοντικών δράσεων. Αυτές οι τεχνολογίες επιτρέπουν στις εταιρείες

να προσαρμόζουν τις στρατηγικές μάρκετινγκ, να αυξάνουν τη δέσμευση των πελατών και να βελτιώνουν τη συνολική απόδοση των επιχειρήσεων [20].

Η εφαρμογή της Τεχνητής Νοημοσύνης και της Μηχανικής Μάθησης στην πρόβλεψη της διαδικτυακής συμπεριφοράς παρέχει στις επιχειρήσεις τη δυνατότητα να αναλύουν δεδομένα καταναλωτών μεγάλης κλίμακας, επιτρέποντας ακριβέστερη στόχευση και εξατομίκευση στις εκστρατείες μάρκετινγκ. Για παράδειγμα, τα μοντέλα πρόβλεψης μπορούν να προβλέψουν την αποχώρηση των πελατών, να εκτιμήσουν τις προθέσεις αγοράς και να βελτιστοποιήσουν τις συστάσεις σε πραγματικό χρόνο [21]. Αυτό είναι ιδιαίτερα σημαντικό καθώς η συμπεριφορά των καταναλωτών στο διαδίκτυο είναι πολύπλοκη και επηρεάζεται από ένα ευρύ φάσμα παραγόντων, συμπεριλαμβανομένων οικονομικών, δημογραφικών και ψυχολογικών στοιχείων [13].

Η Τεχνητή Νοημοσύνη είναι ένας ευρύτερος τομέας της επιστήμης των υπολογιστών που έχει ως στόχο τη δημιουργία συστημάτων ικανών να εκτελούν εργασίες που συνήθως απαιτούν ανθρώπινη νοημοσύνη, όπως η λήψη αποφάσεων, η επίλυση προβλημάτων και η μάθηση [22]. Στο πλαίσιο της Τεχνητής Νοημοσύνης, η Μηχανική Μάθηση επικεντρώνεται στην ανάπτυξη αλγορίθμων που επιτρέπουν στα συστήματα να μαθαίνουν μοτίβα από δεδομένα και να κάνουν προβλέψεις χωρίς να προγραμματίζονται ρητά [23]. Η Μηχανική Μάθηση έχει αποκτήσει ιδιαίτερη σημασία σε τομείς που περιλαμβάνουν μεγάλα σύνολα δεδομένων, όπως το ηλεκτρονικό εμπόριο και η διαδικτυακή συμπεριφορά των καταναλωτών, όπου επιτρέπει στις επιχειρήσεις να αντλούν αξιοποιήσιμες πληροφορίες από τις αλληλεπιδράσεις των καταναλωτών [24]. Διάφοροι τύποι αλγορίθμων Μηχανικής Μάθησης χρησιμοποιούνται ευρέως για την πρόβλεψη της διαδικτυακής συμπεριφοράς, ο καθένας με ξεχωριστά πλεονεκτήματα ανάλογα με την πολυπλοκότητα και τη φύση του έργου.

Η πρόβλεψη της διαδικτυακής συμπεριφοράς είναι μια βασική πτυχή του ψηφιακού μάρκετινγκ, παρέχοντας πληροφορίες σχετικά με τις αλληλεπιδράσεις, τις προτιμήσεις και τις πιθανές αγοραστικές ενέργειες των χρηστών. Με βάσει μια εκτενή επισκόπηση της υπάρχουσας βιβλιογραφίας, πέντε αλγόριθμοι αναδεικνύονται ως οι

πιο συχνά χρησιμοποιούμενοι για την πρόβλεψη της διαδικτυακής συμπεριφοράς: Λογιστική Παλινδρόμηση (Logistic Regression), Δέντρα Αποφάσεων (Decision Trees), Τυχαία Δάση (Random Forests), Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines - SVM) και Νευρωνικά Δίκτυα (Neural Network). Καθένας από αυτούς τους αλγορίθμους προσφέρει μοναδικά πλεονεκτήματα καθιστώντας τους κατάλληλους για διαφορετικές πτυχές της ανάλυσης συμπεριφοράς.

Λογιστική Παλινδρόμηση: Η Λογιστική Παλινδρόμηση χρησιμοποιείται συχνά για προβλήματα Διωνυμικής ταξινόμησης όπου η μεταβλητή αποτελέσματος είναι διχοτομική. Αυτή η μέθοδος μοντελοποιεί την πιθανότητα ένα δεδομένο σημείο εισόδου να ανήκει σε μια συγκεκριμένη κλάση, παρέχοντας μια σαφή πιθανολογική ερμηνεία των αποτελεσμάτων ταξινόμησης [25]. Είναι ιδιαίτερα χρήσιμη για την πρόβλεψη του κατά πόσον ένας χρήστης θα εκτελέσει μια συγκεκριμένη ενέργεια, όπως το κλικ σε μια διαφήμιση ή η ολοκλήρωση μιας αγοράς.

Δέντρα Απόφασης: Τα Δέντρα Αποφάσεων είναι μια δημοφιλής επιλογή λόγω της απλότητας και της ερμηνευσιμότητας τους. Λειτουργούν με την αναδρομική διάσπαση των δεδομένων με βάση τις τιμές των χαρακτηριστικών για τη δημιουργία ενός δενδροειδούς μοντέλου αποφάσεων [26]. Αυτή η μέθοδος είναι αποτελεσματική για την κατηγοριοποίηση των χρηστών σε διαφορετικά τμήματα με βάση τη συμπεριφορά τους, η οποία μπορεί να είναι καθοριστική για στοχευμένες στρατηγικές μάρκετινγκ και εξατομικευμένες συστάσεις.

Τυχαία Δάση: Τα Τυχαία Δάση, μια επέκταση των Δέντρων Αποφάσεων, περιλαμβάνουν την κατασκευή πολλαπλών δέντρων κατά τη φάση της εκπαίδευσης και την εξαγωγή της κλάσης που είναι ο τρόπος των κλάσεων (ταξινόμηση) ή η μέση πρόβλεψη (παλινδρόμηση) των επιμέρους δέντρων [27]. Αυτή η προσέγγιση συνόλου ενισχύει την ακρίβεια πρόβλεψης και μειώνει τον κίνδυνο υπερπροσαρμογής, καθιστώντας την ισχυρή επιλογή για σύνθετες εργασίες πρόβλεψης, όπως η αποχώρηση χρηστών και η ανίχνευση απάτης.

Μηχανές Διανυσμάτων Υποστήριξης: Οι Μηχανές Διανυσμάτων Υποστήριξης είναι ισχυρά εργαλεία τόσο για τις προκλήσεις της ταξινόμησης όσο και της παλινδρόμησης, ιδίως όταν πρόκειται για δεδομένα υψηλών διαστάσεων.

Λειτουργούν με την εύρεση του βέλτιστου υπερεπιπέδου που διαχωρίζει τα δεδομένα σε διαφορετικές κλάσεις [28]. Οι μηχανές SVM εφαρμόζονται ευρέως σε περιπτώσεις σεναρίων που απαιτούν ακριβή οριοθέτηση, όπως η διάκριση μεταξύ διαφορετικών τύπων συμπεριφοράς χρηστών με βάση λεπτομερή αρχεία καταγραφής δραστηριοτήτων.

Νευρωνικά Δίκτυα: Τα Νευρωνικά Δίκτυα, που είναι εμπνευσμένα από τη δομή και τη λειτουργία του ανθρώπινου εγκεφάλου, αποτελούνται από διασυνδεδεμένα στρώματα κόμβων (νευρώνες) που επεξεργάζονται δεδομένα με πολύπλοκους τρόπους. Είναι ιδιαίτερα αποτελεσματικά για τη σύλληψη μη γραμμικών σχέσεων και μοτίβων σε μεγάλα σύνολα δεδομένων. Τα Νευρωνικά Δίκτυα είναι ιδιαίτερα χρήσιμα για εργασίες όπως η αναγνώριση εικόνων, η επεξεργασία φυσικής γλώσσας και η πρόβλεψη σύνθετων συμπεριφορών χρηστών, μαθαίνοντας από τεράστιες ποσότητες δεδομένων μέσω τεχνικών βαθιάς μάθησης [29].

Η παρούσα έρευνα αποσκοπεί στη διερεύνηση της αποτελεσματικότητας διαφόρων αλγορίθμων Μηχανικής Μάθησης στην πρόβλεψη της συμπεριφοράς των καταναλωτών στο διαδίκτυο, η οποία είναι σημαντική για τους ψηφιακούς έμπορους και τους επιχειρηματικούς αναλυτές που επιθυμούν να αξιοποιήσουν τα δεδομένα για τη λήψη αποφάσεων.

1.5 Στόχοι της Έρευνας

Ο πρωταρχικός στόχος της παρούσας έρευνας είναι να αξιολογήσει διάφορους αλγορίθμους Μηχανικής Μάθησης που χρησιμοποιούνται για την πρόβλεψη της συμπεριφοράς των καταναλωτών στο διαδίκτυο. Συγκεκριμένα, η μελέτη στοχεύει:

- Να αναλύσει την αποτελεσματικότητα διαφόρων αλγορίθμων, συμπεριλαμβανομένων των Λογιστική Παλινδρόμηση- Logistic Regression, Δέντρα Αποφάσεων - Decision Trees, Τυχαία Δέντρα - Random Forests, Μηχανές Διανυσμάτων Υποστήριξης - Support Vector Machines και Νευρωνικά Δίκτυα - Neural Networks, στην πρόβλεψη της διαδικτυακής συμπεριφοράς.

- Να εξετάσει τις εφαρμογές αυτών των αλγορίθμων στο ηλεκτρονικό εμπόριο, εστιάζοντας σε εργασίες όπως η πρόβλεψη της πρόθεσης αγοράς, των ποσοστών κλικ και της αποχώρησης των καταναλωτών.
- Να παρέχει μια συγκριτική ανάλυση των επιδόσεων, των πλεονεκτημάτων και των περιορισμών αυτών των αλγορίθμων για να καθοδηγήσει τις επιχειρήσεις στην επιλογή των κατάλληλων μεθόδων για τις συγκεκριμένες ανάγκες τους.
- Να προσφέρει πληροφορίες σχετικά με τον τρόπο με τον οποίο οι αλγόριθμοι πρόβλεψης μπορούν να παρέχουν πληροφορίες για τις στρατηγικές μάρκετινγκ και να ενισχύουν τη στόχευση των πελατών στην ψηφιακή αγορά.

1.6 Δομή της Έρευνας

Η έρευνα οργανώνεται ως εξής:

Το Κεφάλαιο 2 παρέχει λεπτομερή ανάλυση των αλγορίθμων πρόβλεψης που χρησιμοποιήθηκαν στην παρούσα έρευνα, διερευνώντας τα πλεονεκτήματα, τους περιορισμούς και τις συγκεκριμένες εφαρμογές τους στην πρόβλεψη της διαδικτυακής συμπεριφοράς.

Στο Κεφάλαιο 3 παρουσιάζονται μελέτες περιπτώσεων και εφαρμογές αυτών των αλγορίθμων στην πρόβλεψη διαφόρων πτυχών της διαδικτυακής συμπεριφοράς των καταναλωτών, με έμφαση στην πρόθεση αγοράς, τα ποσοστά κλικ και την αποχώρηση των πελατών.

Το Κεφάλαιο 4 συνοψίζει τα βασικά ευρήματα, συγκρίνει τις επιδόσεις των αλγορίθμων, συζητά τα πρότυπα συμπεριφοράς και τις τάσεις που παρατηρήθηκαν στη διαδικτυακή συμπεριφορά, καθώς τις επιπτώσεις στις στρατηγικές μάρκετινγκ και προτείνει κατευθύνσεις για μελλοντική έρευνα.

ΚΕΦΑΛΑΙΟ 2: ΑΝΑΛΥΣΗ ΑΛΓΟΡΙΘΜΩΝ ΠΡΟΒΛΕΨΗΣ

Σε αυτό το κεφάλαιο, θα πραγματοποιηθεί μια ολοκληρωμένη ανάλυση των αλγορίθμων πρόβλεψης και συγκεκριμένα:

- η Λογιστική Παλινδρόμηση - Logistic Regression,
- τα Δέντρα Αποφάσεων - Decision Trees,
- τα Τυχαία Δάση - Random Forests,
- οι Μηχανές Διανυσμάτων Υποστήριξης - Support Vector Machines (SVM),
- τα Νευρωνικά Δίκτυα - Neural Network.

Κάθε αλγόριθμος θα αναλυθεί λεπτομερώς αναφορικά με τις αρχές του, τα πλεονεκτήματα, πιθανούς περιορισμούς και την εφαρμογή τους στην πρόβλεψη της διαδικτυακής συμπεριφοράς.

2.1 Λογιστική Παλινδρόμηση - Logistic Regression

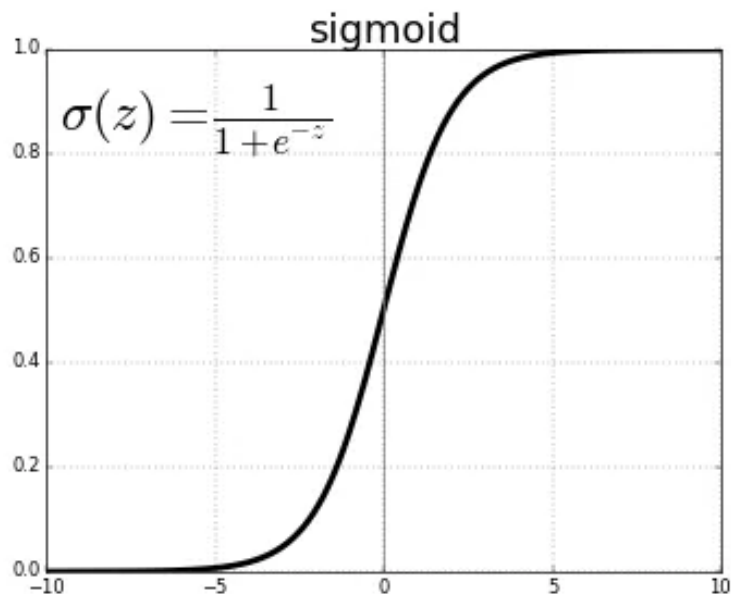
Η Λογιστική Παλινδρόμηση είναι μια μέθοδος που χρησιμοποιείται συχνά για την ταξινόμηση, η οποία είναι μια τεχνική στατιστικής ανάλυσης που εφαρμόζεται για μοντέλα πρόβλεψης. Αυτή η ταξινόμηση είναι ένας από τους πιο δημοφιλείς αλγορίθμους Μηχανικής Μάθησης που υπάγονται στις τεχνικές μάθησης με επίβλεψη. Το μοντέλο ταξινόμησης επιτυγχάνει συνήθως υψηλές επιδόσεις αλγορίθμου, συνεπώς, εφαρμόζεται συχνά στον επιχειρηματικό κόσμο [30].

Υπάρχουν διάφοροι τύποι Λογιστικής Παλινδρόμησης, όπως η Διωνυμική και η Πολυωνυμική Λογιστική Παλινδρόμηση. Το μοντέλο Λογιστικής Παλινδρόμησης μπορεί να επεκταθεί σε πολλαπλές κατηγορίες μέσω τεχνικών, όπως η Πολυωνυμική Λογιστική Παλινδρόμηση για σενάρια όπου η εξαρτημένη μεταβλητή έχει περισσότερες από δύο κατηγορίες. Αυτή η προσαρμοστικότητα καθιστά τη Λογιστική Παλινδρόμηση ένα ευέλικτο εργαλείο σε διάφορες εφαρμογές πέραν της Διωνυμικής ταξινόμησης. Η Διωνυμική Λογιστική Παλινδρόμηση χρησιμοποιείται όταν η μεταβλητή απόκρισης είναι διχοτομική. Δηλαδή, υπάρχουν μόνο δύο κατηγορίες

[31]. Ακόμα, χρησιμοποιείται σε διάφορους τομείς όπως της ιατρικής και των κοινωνικών επιστημών, καθώς και στο μάρκετινγκ, όπως η πρόβλεψη της προδιάθεσης ενός πελάτη να αγοράσει ένα προϊόν ή να διαγραφεί.

Η Λογιστική Παλινδρόμηση μοντελοποιεί την πιθανότητα ένα δεδομένο σημείο εισόδου να ανήκει σε μια συγκεκριμένη κλάση χρησιμοποιώντας τη λογιστική συνάρτηση, η οποία δίνει τιμές μεταξύ 0 και 1. Η λογιστική συνάρτηση ή σιγμοειδής συνάρτηση ορίζεται ως εξής: $\sigma(z) = \frac{1}{1+e^{-z}}$

,όπου z είναι ο γραμμικός συνδυασμός των χαρακτηριστικών εισόδου. Η έξοδος αυτής της συνάρτησης αντιπροσωπεύει την πιθανότητα η είσοδος να ανήκει στη θετική κλάση. Η Λογιστική Παλινδρόμηση εκτιμά τις παραμέτρους ενός λογιστικού μοντέλου, συνδέοντας έτσι την πιθανότητα μιας Διωνυμικής απόκρισης με έναν γραμμικό συνδυασμό μεταβλητών πρόβλεψης [31].



Σχήμα 2.1: Η σιγμοειδής συνάρτηση $\sigma(z)$ στην Λογιστική Παλινδρόμηση [32].

Στο Σχήμα 2.1 απεικονίζεται η σιγμοειδή συνάρτηση, η οποία είναι θεμελιώδης στη Λογιστική Παλινδρόμηση και συμβολίζεται ως $\sigma(z)$. Αναπαριστά οποιονδήποτε αριθμό πραγματικής αξίας στο εύρος μεταξύ 0 και 1. Όπως φαίνεται και στο σχήμα, η σιγμοειδής συνάρτηση δίνει τιμές κοντά στο 1 για μεγάλες θετικές εισόδους και

τιμές κοντά στο 0 για μεγάλες αρνητικές εισόδους, με την έξοδο της συνάρτησης να είναι ακριβώς 0,5 όταν η είσοδος z είναι μηδέν. Αυτή η χαρακτηριστική καμπύλη σχήματος S είναι καθοριστική για τη μετατροπή της εξόδου της γραμμικής παλινδρόμησης σε πιθανότητα, καθιστώντας τη Λογιστική Παλινδρόμηση κατάλληλη για δυαδικές εργασίες ταξινόμησης.

2.1.1 Πλεονεκτήματα

Το κύριο πλεονέκτημα της Λογιστικής Παλινδρόμησης είναι η απλότητα και η ευκολία στο να ερμηνευτεί. Οι συντελεστές του μοντέλου παρέχουν πληροφορίες για τις σχέσεις μεταξύ των ανεξάρτητων μεταβλητών και της μεταβλητής αποτελέσματος. Αυτή η ευκολία εξήγησης του αλγορίθμου είναι σημαντική για την κατανόηση του τρόπου με τον οποίο τα διάφορα χαρακτηριστικά συμβάλλουν στην πιθανότητα εμφάνισης του αποτελέσματος. Ένα άλλο πλεονέκτημα του μοντέλου Λογιστικής Παλινδρόμησης είναι η δυνατότητα επεξεργασίας μεγάλου όγκου δεδομένων με υψηλή ταχύτητα, επειδή απαιτεί μικρότερη υπολογιστική ικανότητα, όπως μνήμη και επεξεργαστική ισχύ [31].

Η Λογιστική Παλινδρόμηση παρέχει επίσης αξιόπιστα αποτελέσματα υπό διάφορες συνθήκες και μπορεί να υλοποιηθεί με σχετικά απλούς αλγορίθμους. Δεν απαιτεί τις παραδοχές της κανονικότητας και της ομοσκεδαστικότητας, όπως η γραμμική παλινδρόμηση, γεγονός που προσθέτει στην ευρωστία της στο χειρισμό δεδομένων σε πραγματικό χρόνο [31].

2.1.2 Περιορισμοί

Ένας σημαντικός περιορισμός της Λογιστικής Παλινδρόμησης είναι η υπόθεση της γραμμικής σχέσης μεταξύ των ανεξάρτητων μεταβλητών και των λογαριθμικών αποδόσεων της εξαρτημένης μεταβλητής. Αυτή η υπόθεση μπορεί να μην ισχύει σε πολλά σενάρια του πραγματικού κόσμου, οδηγώντας ενδεχομένως σε μεροληπτικές

ή ανακριβείς προβλέψεις [33]. Η Λογιστική Παλινδρόμηση συνήθως δεν χρησιμοποιείται με πολύπλοκες, μη γραμμικές σχέσεις μεταξύ των χαρακτηριστικών.

Επιπλέον, η Λογιστική Παλινδρόμηση είναι λιγότερο αποτελεσματική όταν έχει να κάνει με μεγάλο αριθμό κατηγορικών χαρακτηριστικών ή όταν τα δεδομένα έχουν σύνθετες αλληλεπιδράσεις και μη γραμμικότητες που δεν μπορεί να συλλάβει χωρίς εκτεταμένη μηχανική ή μετασχηματισμό των χαρακτηριστικών.

2.1.3 Εφαρμογές στην Πρόβλεψη της Διαδικτυακής Συμπεριφοράς

Η Λογιστική Παλινδρόμηση χρησιμοποιείται ευρέως στην πρόβλεψη του κατά πόσον ένας χρήστης θα κάνει κλικ σε μια διαφήμιση ή θα προβεί σε αγορά ή ακόμα και για δημογραφική στόχευση σε διαδικτυακές διαφημίσεις κ.α. [31, 34, 35, 36]. Έχει εφαρμοστεί για τη μοντελοποίηση των ποσοστών κλικ στη διαδικτυακή διαφήμιση και τον προσδιορισμό της πιθανότητας αλληλεπίδρασης των χρηστών με τις διαφημίσεις. Τα πιθανοτικά αποτελέσματα του αλγορίθμου είναι ιδιαίτερα πολύτιμα για τη λήψη αποφάσεων με βάση την πιθανότητα ενεργειών των χρηστών, επιτρέποντας στοχευμένες στρατηγικές μάρκετινγκ και εξατομικευμένες εμπειρίες χρηστών.

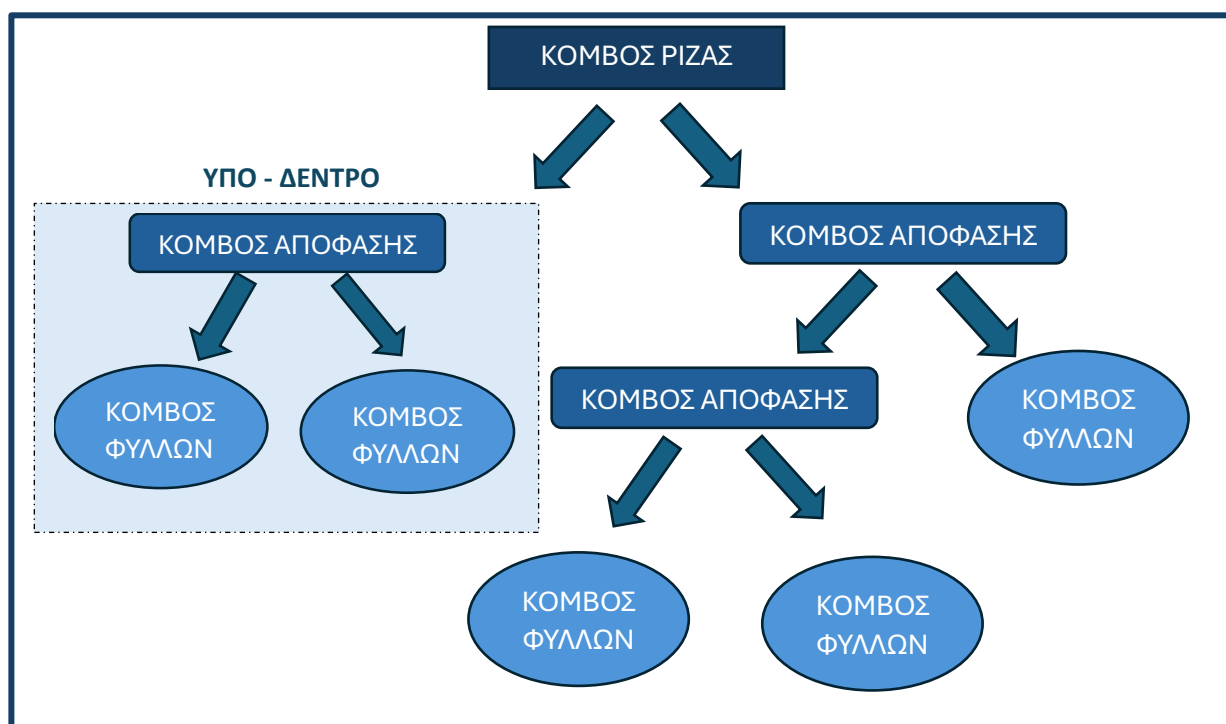
2.2 Δέντρα Αποφάσεων - Decision Tree

Τα Δέντρα Αποφάσεων είναι μια μη παραμετρική μέθοδος μάθησης με επίβλεψη που χρησιμοποιείται για ταξινόμηση και παλινδρόμηση. Το μοντέλο λειτουργεί με αναδρομική διαίρεση των δεδομένων σε υποσύνολα με βάση την τιμή των χαρακτηριστικών εισόδου, δημιουργώντας μια δενδροειδή δομή αποφάσεων. Κάθε εσωτερικός κόμβος του δέντρου αναπαριστά μια απόφαση με βάση ένα χαρακτηριστικό, κάθε κλάδος αναπαριστά το αποτέλεσμα αυτής της απόφασης και κάθε κόμβος φύλλου αναπαριστά μια ετικέτα κλάσης ή μια συνεχή τιμή.

Η διαδικασία ξεκινά με ολόκληρο το σύνολο των δεδομένων και επιλέγει το χαρακτηριστικό που διαχωρίζει καλύτερα τα δεδομένα σύμφωνα με ένα

συγκεκριμένο κριτήριο, όπως το κέρδος πληροφορίας για εργασίες ταξινόμησης ή το μέσο τετραγωνικό σφάλμα για εργασίες παλινδρόμησης. Αυτή η διαδικασία διαχωρισμού συνεχίζεται αναδρομικά μέχρι τα υποσύνολα σε κάθε κόμβο να είναι καθαρά, δηλαδή να περιέχουν μόνο περιπτώσεις μιας κλάσης ή να πληρούν ένα κριτήριο διακοπής π.χ. μέγιστο βάθος δέντρου ή ελάχιστος αριθμός δειγμάτων ανά κόμβο [26].

Τα Δέντρα Αποφάσεων αναλύουν συστηματικά τα δεδομένα για να εξαγάγουν πολύτιμους κανόνες και σχέσεις από μεγάλα σύνολα δεδομένων. Αυτός ο αλγόριθμος χρησιμοποιείται συχνά σε εργασίες ταξινόμησης ή πρόβλεψης, αξιοποιώντας τόσο συνεχείς όσο και κατηγορικές μεταβλητές. Η δομή του δέντρου εκτείνεται από τον κόμβο ρίζας έως τους κόμβους φύλλων, με κάθε κόμβο να αντιπροσωπεύει έναν κανόνα απόφασης που προκύπτει από τα δεδομένα [37].



Σχήμα 2.2: Η δομή του Αλγόριθμου Πρόβλεψης “Δέντρα Αποφάσεων” [38].

Το Σχήμα 2 απεικονίζει τη δομή ενός Δέντρου Αποφάσεων, το οποίο χρησιμοποιείται τόσο για εργασίες ταξινόμησης όσο και για εργασίες παλινδρόμησης. Το δέντρο αρχίζει με έναν **κόμβο ρίζας**, ο οποίος αντιπροσωπεύει το αρχικό σημείο απόφασης με βάση ένα συγκεκριμένο χαρακτηριστικό. Από τον κόμβο ρίζας, το δέντρο

διακλαδίζεται σε διάφορους **κόμβους απόφασης**, καθένας από τους οποίους αντιπροσωπεύει μια επακόλουθη απόφαση με βάση διαφορετικές τιμές χαρακτηριστικών. Οι διακλαδώσεις συνεχίζονται σε **κόμβους φύλλων**, οι οποίοι αντιπροσωπεύουν τα τελικά αποτελέσματα ή τις ετικέτες κλάσης. Αυτή η ιεραρχική δομή παρουσιάζει τον τρόπο με τον οποίο το δέντρο αποφάσεων λαμβάνει συστηματικές αποφάσεις, διαχωρίζοντας τα δεδομένα σε κάθε κόμβο σύμφωνα με το καλύτερο χαρακτηριστικό σε αυτό το επίπεδο. Η διαγραμματισμένη ενότητα **Sub-Tree** εντός του σχήματος δείχνει πώς ένα υποσύνολο των κόμβων απόφασης και των κόμβων φύλλων σχηματίζει ένα μικρότερο δέντρο εντός της ευρύτερης δομής, δίνοντας έμφαση στην αναδρομική φύση των Δέντρων Απόφασης. Η σαφής απεικόνιση των διαδρομών απόφασης στο δέντρο επιτρέπει την εύκολη ερμηνεία και κατανόηση του τρόπου με τον οποίο τα διάφορα χαρακτηριστικά συμβάλλουν στις τελικές προβλέψεις, καθιστώντας τα Δέντρα Αποφάσεων μια δημοφιλή επιλογή για εργασίες που απαιτούν ευκολία στην ερμηνεία τους και διαφάνεια.

2.2.1 Πλεονεκτήματα

Τα δέντρα απόφασης προσφέρουν αρκετά πλεονεκτήματα έναντι άλλων μεθόδων ταξινόμησης, όπως οι Μηχανές Διανυσμάτων Υποστήριξης, τα Νευρωνικά Δίκτυα και η Λογιστική Παλινδρόμηση:

- Απλότητα: Τα Δέντρα Αποφάσεων είναι εξαιρετικά εύκολο να απεικονιστούν και να ερμηνευτούν. Μπορούν να αναπαρασταθούν γραφικά, επιτρέποντας στους χρήστες να δουν τη δομή του ταξινομητή και να κατανοήσουν όλα τα ενδιάμεσα βήματα της διαδικασίας ταξινόμησης, συμπεριλαμβανομένων των μεταβλητών που χρησιμοποιούνται και με ποια σειρά [39].
- Ευελιξία: Μπορούν να χειριστούν τόσο αριθμητικά όσο και κατηγορικά δεδομένα και μπορούν να χρησιμοποιηθούν για εργασίες ταξινόμησης και παλινδρόμησης.
- Μη γραμμικότητα: Τα Δέντρα Αποφάσεων μπορούν να καταγράψουν μη γραμμικές σχέσεις μεταξύ των χαρακτηριστικών και της μεταβλητής-στόχου

χωρίς να απαιτείται μετασχηματισμός των χαρακτηριστικών ή όροι αλληλεπίδρασης.

- Ταχύτητα: Τα Δέντρα Αποφάσεων εκπαιδεύονται σε σχετικά σύντομο χρονικό διάστημα και είναι ιδιαίτερα γρήγορα στην ταξινόμηση νέων δεδομένων [39].

2.2.2 Περιορισμοί

Παρά τα πλεονεκτήματά τους, τα Δέντρα Αποφάσεων έχουν αξιοσημείωτους περιορισμούς:

- Υπερπροσαρμογή: Τα Δέντρα Αποφάσεων είναι επιρρεπή στην υπερπροσαρμογή, ειδικά όταν είναι βαθιά και πολύπλοκα, καταγράφοντας το θόρυβο στα δεδομένα εκπαίδευσης και όχι το υποκείμενο μοτίβο. Μέθοδοι κανονικοποίησης, όπως το pruning, είναι μια τεχνική που αφαιρεί τα τμήματα του Δέντρου Αποφάσεων που το εμποδίζουν να αναπτυχθεί σε όλο του το βάθος) ή η επιβολή ενός ελάχιστου αριθμού δειγμάτων εκπαίδευσης ανά φύλλο, μπορούν να μετριάσουν αυτό το ζήτημα [39].
- Αστάθεια: Μικρές αλλαγές στα δεδομένα μπορούν να οδηγήσουν σε μια εντελώς διαφορετική δομή δέντρου, καθιστώντας τα ευαίσθητα στη μεταβλητότητα των δεδομένων.
- Μεροληψία: Τα Δέντρα Αποφάσεων μπορεί να μεροληπτούν προς τα χαρακτηριστικά με περισσότερα επίπεδα ή κατηγορίες λόγω των κριτηρίων διαχωρισμού τους, υπερτονίζοντας ενδεχομένως τη σημασία αυτών των μεταβλητών .
- Υπολογιστική Πολυπλοκότητα: Η δημιουργία ενός βέλτιστου Δέντρου Απόφασης είναι NP-hard, πράγμα που σημαίνει ότι δεν είναι δυνατόν να δημιουργηθεί αποτελεσματικά ένα συνολικά βέλτιστο δέντρο [39].

2.2.3 Εφαρμογές στην Πρόβλεψη της Διαδικτυακής Συμπεριφοράς

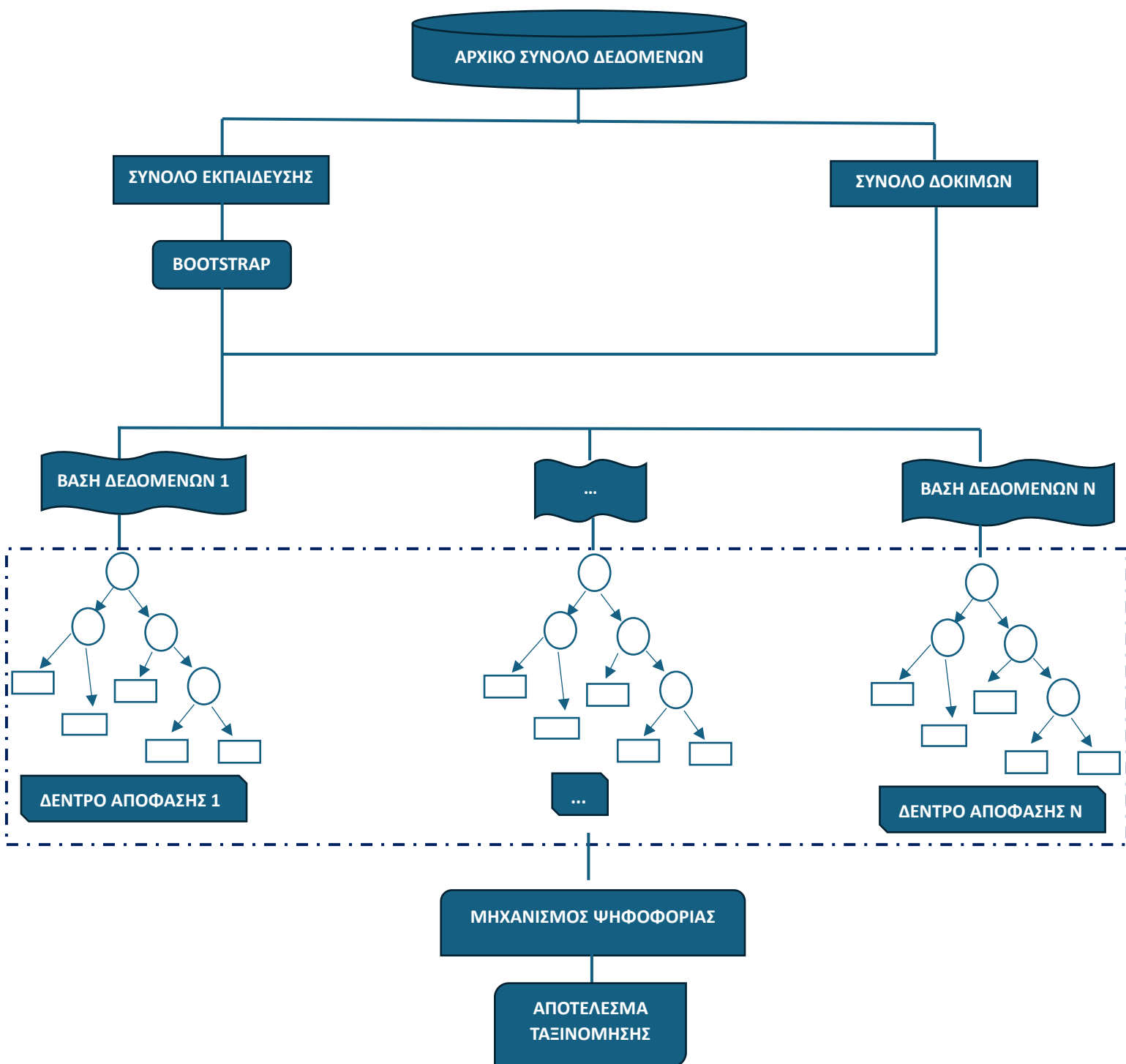
Τα Δέντρα Αποφάσεων χρησιμοποιούνται ευρέως στην πρόβλεψη της διαδικτυακής συμπεριφοράς για εργασίες όπως η τμηματοποίηση των χρηστών, η πρόβλεψη της αποχώρησης, η εκτίμηση του ποσοστού κλικ, η εκτίμηση της πρόθεσης για διαδικτυακή αγορά, η πρόβλεψη της συμπεριφοράς των διαδικτυακών χρηστών κ.α. [37, 39, 40, 41]. Η ερμηνευσιμότητα των αποτελεσμάτων τους επιτρέπει στους υπεύθυνους μάρκετινγκ να κατανοήσουν τα βήματα λήψης αποφάσεων που οδηγούν σε συγκεκριμένες ενέργειες των χρηστών, βοηθώντας στην ανάπτυξη στοχευμένων στρατηγικών μάρκετινγκ.

2.3 Τυχαία Δάση - Random Forests

Τα Τυχαία Δάση είναι μια μέθοδος μάθησης συνόλου που κατασκευάζει πολλαπλά δέντρα απόφασης κατά την εκπαίδευση και συγχωνεύει τις εξόδους τους για να βελτιώσει την ακρίβεια πρόβλεψης και να ελέγξει την υπερπροσαρμογή. Κάθε δέντρο σε ένα τυχαίο δάσος εκπαιδεύεται σε ένα τυχαίο υποσύνολο των δεδομένων εκπαίδευσης και κάθε διάσπαση στο δέντρο καθορίζεται χρησιμοποιώντας ένα τυχαίο υποσύνολο των χαρακτηριστικών. Αυτή η διαδικασία, γνωστή ως *bagging*¹, συμβάλλει στη μείωση της διακύμανσης του μοντέλου και ενισχύει την ευρωστία του [27].

Ο αλγόριθμος του Τυχαίου Δάσους λειτουργεί με τη μέση τιμή των προβλέψεων των μεμονωμένων δέντρων (για εργασίες παλινδρόμησης) ή με τη λήψη της πλειοψηφίας (για εργασίες ταξινόμησης). Αυτή η προσέγγιση του συνόλου συμβάλλει στον μετριασμό του προβλήματος της υπερπροσαρμογής στο οποίο είναι επιρρεπή τα μεμονωμένα δέντρα απόφασης, παρέχοντας καλύτερη γενίκευση σε αφανή δεδομένα [43].

¹ είναι μια τεχνική συνδυασμού μοντέλων που περιλαμβάνει την εκπαίδευση πολλών μοντέλων ανεξάρτητα σε τυχαία επιλεγμένα υποσύνολα δεδομένων και τη συνένωση των προβλέψεών τους μέσω ψηφοφορίας ή υπολογισμού του μέσου όρου [42].



Σχήμα 2.3: Η δομή του Αλγορίθμου Πρόβλεψης “Τυχαία Δάση” [44].

Το Σχήμα 2.3 απεικονίζει τη δομή και τη διαδικασία ενός Τυχαίου Δάσους, το οποίο είναι μια μέθοδος μάθησης συνόλου που χρησιμοποιείται για εργασίες ταξινόμησης και παλινδρόμησης. Το διάγραμμα ξεκινά με το Αρχικό Σύνολο Δεδομένων, το οποίο χωρίζεται σε ένα “Σύνολο Εκπαίδευσης” και ένα “Σύνολο Δοκιμών”. Το σύνολο

εκπαίδευσης υφίσταται μια διαδικασία Bootstrap, όπου παράγονται πολλαπλά δείγματα, σύνολα δεδομένων, με τυχαία δειγματοληψία με αντικατάσταση. Κάθε δείγμα Bootstrap (σύνολο δεδομένων 1, σύνολο δεδομένων 2, ..., σύνολο δεδομένων N) χρησιμοποιείται για την εκπαίδευση ενός ξεχωριστού Δέντρου Απόφασης (δέντρο απόφασης 1, δέντρο απόφασης 2, ..., δέντρο απόφασης N). Αυτά τα δέντρα απόφασης κατασκευάζονται με αναδρομική κατάτμηση των δεδομένων σε υποσύνολα με βάση τις τιμές των χαρακτηριστικών, με κάθε δέντρο να είναι μοναδικό λόγω των διαδικασιών τυχαίας δειγματοληψίας και επιλογής χαρακτηριστικών που εμπλέκονται. Αφού εκπαιδευτούν όλα τα δέντρα απόφασης, κάνουν συλλογικά προβλέψεις στο σύνολο δοκιμών. Το τελικό αποτέλεσμα ταξινόμησης προσδιορίζεται με τη συγκέντρωση των προβλέψεων όλων των Δέντρων Απόφασης χρησιμοποιώντας έναν μηχανισμό ψηφοφορίας (για εργασίες ταξινόμησης) ή μέσο όρο (για εργασίες παλινδρόμησης). Αυτή η προσέγγιση συνόλου βοηθά στη βελτίωση της ακρίβειας πρόβλεψης και της ευρωστίας του μοντέλου μειώνοντας τη διακύμανση που σχετίζεται με τα μεμονωμένα δέντρα απόφασης.

2.3.1 Πλεονεκτήματα

Τα Τυχαία Δάση προσφέρουν αρκετά πλεονεκτήματα σε σχέση με άλλες μεθόδους Μηχανικής Μάθησης, συμπεριλαμβανομένων των Δέντρων Αποφάσεων:

- Βελτιωμένη Ακρίβεια: Με τον μέσο όρο πολλαπλών δέντρων, τα Τυχαία Δάση επιτυγχάνουν υψηλότερη ακρίβεια πρόβλεψης από τα μεμονωμένα δέντρα απόφασης [27].
- Ανθεκτικότητα στην Υπερπροσαρμογή: Η φύση του συνόλου των Τυχαίων Δασών μειώνει την υπερπροσαρμογή, καθιστώντας τα πιο ανθεκτικά στο θόρυβο των δεδομένων εκπαίδευσης [27].
- Σημασία Χαρακτηριστικών: Τα Τυχαία Δάση παρέχουν εκτιμήσεις της σημασίας των χαρακτηριστικών, επιτρέποντας την καλύτερη κατανόηση των παραγόντων που επηρεάζουν τις προβλέψεις [43].

- Ευελιξία: Μπορούν να χειριστούν τόσο εργασίες ταξινόμησης όσο και παλινδρόμησης και λειτουργούν καλά τόσο με αριθμητικά όσο και με κατηγορικά δεδομένα [45].
- Χειρισμός Ελλειπόντων Δεδομένων: Τα Τυχαία Δάση μπορούν να χειριστούν αποτελεσματικά τις ελλείπουσες τιμές και να διατηρήσουν την ακρίβεια χρησιμοποιώντας διάφορες μεθόδους υπολογισμού [43].

2.3.2 Περιορισμοί

Παρά τα πλεονεκτήματά τους, τα Τυχαία Δάση έχουν ορισμένους περιορισμούς:

- Υπολογιστική Δύναμη: Η εκπαίδευση πολλαπλών Δέντρων Απόφασης και η ομαδοποίηση των αποτελεσμάτων τους μπορεί να είναι υπολογιστικά δαπανηρή, ειδικά με μεγάλα σύνολα δεδομένων [45].
- Έλλειψη Ερμηνευσιμότητας: Ενώ τα μεμονωμένα Δέντρα Αποφάσεων είναι εύκολο να ερμηνευτούν, το σύνολο των δέντρων σε ένα τυχαίο δάσος καθιστά πιο δύσκολη την κατανόηση της διαδικασίας λήψης αποφάσεων [43].
- Χρήση Μνήμης: Τα Τυχαία Δάση απαιτούν περισσότερη μνήμη για την αποθήκευση πολλαπλών δέντρων σε σύγκριση με τα μοντέλα μεμονωμένων δέντρων [43].

2.3.3 Εφαρμογές στην Πρόβλεψη της Διαδικτυακής Συμπεριφοράς

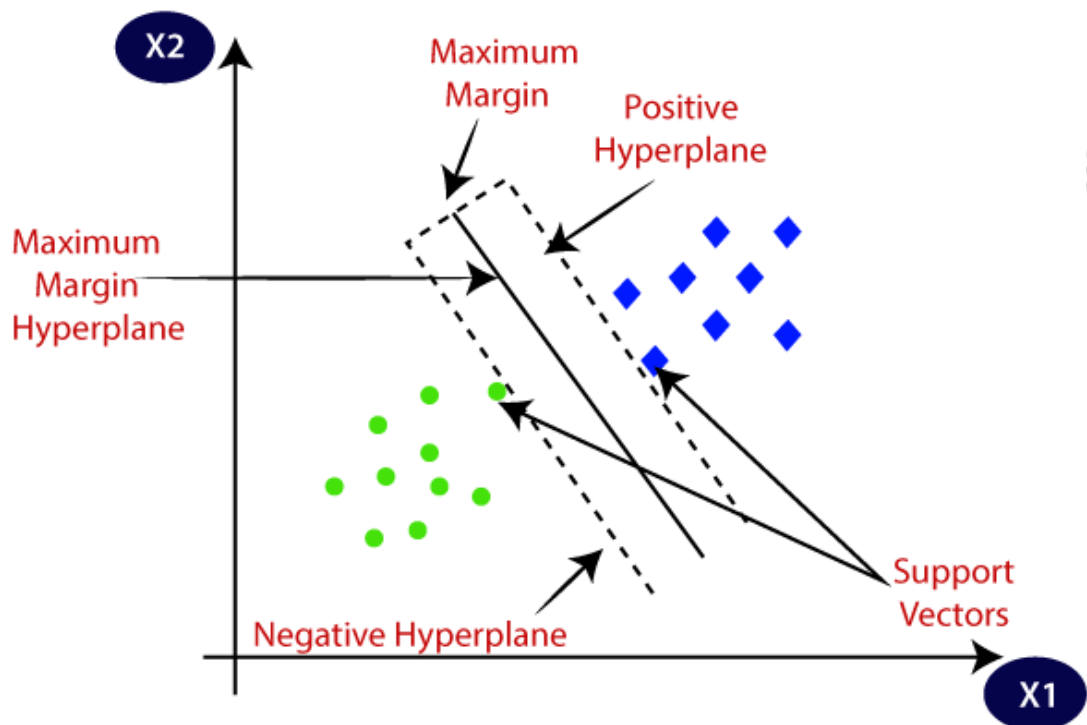
Τα Τυχαία Δάση χρησιμοποιούνται συχνά στην πρόβλεψη της διαδικτυακής συμπεριφοράς για εργασίες όπως η πρόβλεψη της αποχώρησης, η εκτίμηση του ποσοστού κλικ, η διαδικτυακή αγοραστική συμπεριφορά, η εκτίμηση της πρόθεσης για διαδικτυακή αγορά κ.α. [37, 39, 45, 46, 47]. Η ικανότητά τους να χειρίζονται μεγάλο αριθμό χαρακτηριστικών και να παρέχουν ισχυρές προβλέψεις τις καθιστά ιδιαίτερα κατάλληλες για σύνθετα σύνολα δεδομένων που είναι συνηθισμένα στην ανάλυση της διαδικτυακής συμπεριφοράς.

2.4 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines - SVM)

Οι Μηχανές Διανυσμάτων Υποστήριξης είναι μοντέλα μάθησης με επίβλεψη που χρησιμοποιούνται για εργασίες ταξινόμησης και παλινδρόμησης. Ο πρωταρχικός στόχος των SVM είναι η εύρεση του υπερεπιπέδου που διαχωρίζει καλύτερα τα δεδομένα σε διαφορετικές κλάσεις. Για Διωνυμική ταξινόμηση, αυτό το υπερεπίπεδο είναι ένα όριο απόφασης που μεγιστοποιεί το περιθώριο μεταξύ δύο κλάσεων, εξασφαλίζοντας ότι τα πλησιέστερα σημεία δεδομένων (διανύσματα υποστήριξης) βρίσκονται όσο το δυνατόν πιο μακριά από το υπερεπίπεδο [28].

Η συνάρτηση απόφασης ενός SVM αναπαρίσταται ως εξής: $f(\mathbf{x}) = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b)$, όπου w είναι το διάνυσμα βάρους, x είναι το διάνυσμα χαρακτηριστικών εισόδου και b είναι η προκατάληψη. Το βέλτιστο υπερεπίπεδο βρίσκεται με την επίλυση του ακόλουθου προβλήματος βελτιστοποίησης: $\min_{\frac{1}{2}} \|\mathbf{w}\|^2$, με τον περιορισμό ότι όλα τα σημεία δεδομένων ταξινομούνται σωστά: $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1$, όπου το y_i αντιπροσωπεύει τις ετικέτες της κλάσης [48].

Σε περιπτώσεις όπου τα δεδομένα δεν είναι γραμμικά διαχωρίσιμα, το SVM μπορεί να χρησιμοποιήσει συναρτήσεις πυρήνα για να απεικονίσει τα δεδομένα εισόδου σε έναν χώρο υψηλότερων διαστάσεων όπου είναι δυνατός ο γραμμικός διαχωρισμός. Οι συνήθεις συναρτήσεις πυρήνα περιλαμβάνουν τον πολυωνυμικό πυρήνα, τον πυρήνα ακτινικής συνάρτησης βάσης (RBF) και τον σιγμοειδή πυρήνα [49].



Σχήμα 2.4: Η δομή του Αλγορίθμου Πρόβλεψης “Μηχανές Διανυσμάτων Υποστήριξης” [50].

Το Σχήμα 2.4 απεικονίζει τη βασική ιδέα μιας Μηχανής Διανυσμάτων Υποστήριξης για Διωνυμική ταξινόμηση. Το διάγραμμα δείχνει δύο κλάσεις σημείων δεδομένων (πράσινοι κύκλοι και μπλε τετράγωνα) που διαχωρίζονται από ένα υπερεπίπεδο. Τα βασικά στοιχεία είναι τα εξής:

Διανύσματα Υποστήριξης – Support Vectors: Σημεία δεδομένων που βρίσκονται πλησιέστερα στο υπερεπίπεδο, τα οποία είναι καθοριστικά για τον καθορισμό της θέσης και του προσανατολισμού του.

Υπερεπίπεδο Μέγιστου Περιθωρίου – Maximum Margin Hyperplane: Το βέλτιστο υπερεπίπεδο που μεγιστοποιεί την απόσταση μεταξύ των δύο κλάσεων.

Θετικό και Αρνητικό Υπερεπίπεδο – Positive & Negative Hyperplane: Παράλληλα υπερεπίπεδα που βρίσκονται πλησιέστερα στις θετικές (μπλε τετράγωνα) και αρνητικές (πράσινοι κύκλοι) κλάσεις. Η μεταξύ τους απόσταση αντιπροσωπεύει το περιθώριο.

Μέγιστο Περιθώριο – Maximum Margin: Η απόσταση μεταξύ του θετικού και του αρνητικού υπερεπιπέδου, την οποία μεγιστοποιεί το SVM για να βελτιώσει τη γενίκευση.

Αυτό το σχήμα δείχνει πώς οι SVM μεγιστοποιούν το περιθώριο μεταξύ των κλάσεων για να δημιουργήσουν έναν ισχυρό ταξινομητή.

2.4.1 Πλεονεκτήματα

Οι SVM προσφέρουν αρκετά πλεονεκτήματα έναντι άλλων μεθόδων ταξινόμησης:

- Υψηλή Ακρίβεια: Οι SVM είναι γνωστές για την υψηλή τους ακρίβεια ταξινόμησης, ιδίως σε χώρους υψηλών διαστάσεων.
- Αποτελεσματική σε Υψηλές Διαστάσεις: Οι SVM λειτουργούν καλά με μεγάλο αριθμό χαρακτηριστικών και μπορούν να χειριστούν αποτελεσματικά δεδομένα υψηλών διαστάσεων [28].
- Ανθεκτικότητα στην Υπερπροσαρμογή: Εστιάζοντας στα διανύσματα υποστήριξης, οι SVM μειώνουν τον κίνδυνο υπερπροσαρμογής, ιδίως σε περιπτώσεις με σαφές περιθώριο διαχωρισμού.
- Ευελιξία: Η χρήση συναρτήσεων πυρήνα επιτρέπει στα SVM να μοντελοποιούν πολύπλοκες μη γραμμικές σχέσεις.

2.4.2 Περιορισμοί

Παρά τα πλεονεκτήματά τους, οι SVM έχουν ορισμένους περιορισμούς:

- Υπολογιστική Δύναμη: Η εκπαίδευση των SVM μπορεί να είναι υπολογιστικά εντατική, ειδικά με μεγάλα σύνολα δεδομένων και πολύπλοκους πυρήνες [49].
- Χρήση Μνήμης: Οι SVM απαιτούν σημαντική μνήμη για την αποθήκευση διανυσμάτων υποστήριξης και τους υπολογισμούς του πυρήνα.
- Ρύθμιση Παραμέτρων: Η απόδοση των SVM εξαρτάται σε μεγάλο βαθμό από την επιλογή του πυρήνα και των παραμέτρων (όπως η παράμετρος κανονικοποίησης C και οι παράμετροι του πυρήνα), οι οποίες απαιτούν προσεκτικό συντονισμό.

- Περιορισμένη ερμηνευσιμότητα: Σε αντίθεση με τα Δέντρα Αποφάσεων, τα SVM δεν παρέχουν εύκολα ερμηνεύσιμα μοντέλα, καθιστώντας δύσκολη την κατανόηση των ορίων απόφασης.

2.4.3 Εφαρμογές στην Πρόβλεψη της Διαδικτυακής Συμπεριφοράς

Οι SVM χρησιμοποιούνται ευρέως στην πρόβλεψη της διαδικτυακής συμπεριφοράς για εργασίες όπως η ταξινόμηση χρηστών, προβλέψεις για ηλεκτρονικές αγορές κ.α. [51, 52]. Η ικανότητά τους να χειρίζονται δεδομένα υψηλών διαστάσεων και να μοντελοποιούν πολύπλοκες σχέσεις τις καθιστά ιδιαίτερα χρήσιμες σε αυτές τις εφαρμογές.

2.5 Νευρωνικά Δίκτυα - Neural Network

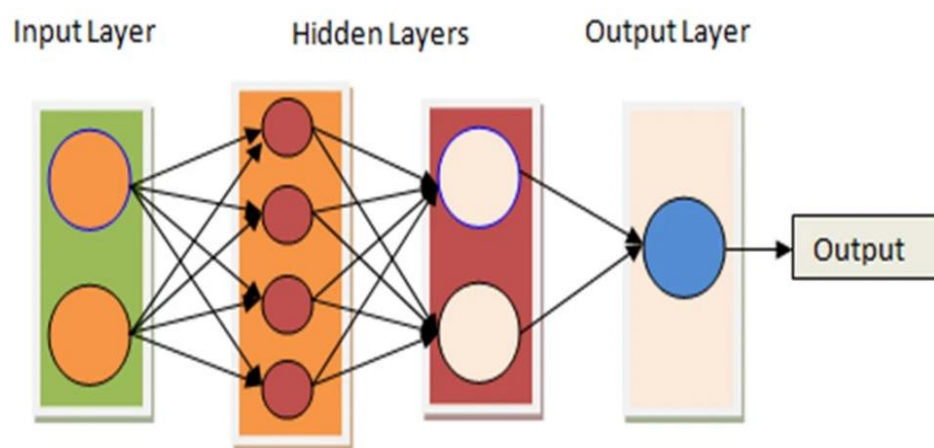
Τα Νευρωνικά Δίκτυα είναι υπολογιστικά μοντέλα που είναι εμπνευσμένα από τη δομή και τη λειτουργία του ανθρώπινου εγκεφάλου. Αποτελούνται από διασυνδεδεμένα στρώματα κόμβων (νευρώνων), που περιλαμβάνουν ένα στρώμα εισόδου, ένα ή περισσότερα κρυφά στρώματα και ένα στρώμα εξόδου. Κάθε κόμβος επεξεργάζεται δεδομένα εισόδου και τα περνάει από μια συνάρτηση ενεργοποίησης για να παράγει μια έξοδο που μεταδίδεται στο επόμενο επίπεδο. Ο πυρήνας της λειτουργίας ενός Νευρωνικού Δικτύου είναι η ικανότητά του να μαθαίνει από τα δεδομένα μέσω μιας διαδικασίας που ονομάζεται εκπαίδευση. Κατά τη διάρκεια της εκπαίδευσης, το δίκτυο προσαρμόζει τα βάρη και τις προκαταλήψεις του για να ελαχιστοποιήσει το σφάλμα μεταξύ των προβλέψεών του και των πραγματικών αποτελεσμάτων. Αυτή η προσαρμογή γίνεται συνήθως με τη χρήση μιας μεθόδου που ονομάζεται backpropagation, σε συνδυασμό με έναν αλγόριθμο βελτιστοποίησης όπως η gradient descent [29].

Η μέθοδος οπισθοδιάδοσης περιλαμβάνει τον υπολογισμό της κλίσης της συνάρτησης απώλειας ως προς κάθε βάρος με τον κανόνα της αλυσίδας, διαδίδοντας ουσιαστικά το σφάλμα προς τα πίσω από το στρώμα εξόδου στο στρώμα εισόδου.

Αυτό επιτρέπει στο δίκτυο να ενημερώνει τα βάρη του με τρόπο που μειώνει το συνολικό σφάλμα. Το Gradient Descent είναι ένας επαναληπτικός αλγόριθμος βελτιστοποίησης που χρησιμοποιείται για την εύρεση του ελαχίστου μιας συνάρτησης- σε αυτό το πλαίσιο, χρησιμοποιείται για την ελαχιστοποίηση του σφάλματος του Νευρωνικού Δικτύου με επαναληπτική προσαρμογή των βαρών [53].

Επιπλέον, τα Νευρωνικά Δίκτυα μπορούν να χρησιμοποιούν διάφορους τύπους συναρτήσεων ενεργοποίησης, όπως η σιγμοειδής συνάρτηση, η υπερβολική εφαπτομένη (tanh) και η διορθωμένη γραμμική μονάδα (ReLU²). Αυτές οι συναρτήσεις εισάγουν μη γραμμικότητα στο μοντέλο, επιτρέποντας στο δίκτυο να μαθαίνει σύνθετα μοτίβα στα δεδομένα [54].

Τα Νευρωνικά Δίκτυα μπορούν επίσης να επεκταθούν σε μοντέλα βαθιάς μάθησης, όπου περιέχουν πολλά κρυφά στρώματα και είναι γνωστά ως Βαθιά Νευρωνικά Δίκτυα (DNN). Αυτά τα μοντέλα έχουν επιδείξει μεγάλη επιτυχία σε διάφορες σύνθετες εργασίες, όπως η αναγνώριση εικόνας και ομιλίας, η επεξεργασία φυσικής γλώσσας και άλλα [55].



Σχήμα 2.5: Η δομή του Αλγορίθμου Πρόβλεψης “Νευρωνικά Δίκτυα” [51].

Το Σχήμα 2.5 απεικονίζει τη βασική αρχιτεκτονική ενός Νευρωνικού Δικτύου, που αποτελείται από τρεις κύριους τύπους στρωμάτων:

² ReLu: συνάρτηση ενεργοποίησης που χρησιμοποιείται συχνότερα στα Βαθιά Νευρωνικά Δίκτυα [54].

Στρώμα Εισόδου (Input Layer): Αυτό είναι το πρώτο στρώμα του Νευρωνικού Δικτύου, όπου τα δεδομένα εισόδου εισάγονται στο μοντέλο. Κάθε κόμβος (κύκλος) σε αυτό το επίπεδο αντιπροσωπεύει ένα χαρακτηριστικό των δεδομένων εισόδου.

Κρυφά Στρώματα (Hidden Layers): Τα στρώματα αυτά τοποθετούνται μεταξύ των στρωμάτων εισόδου και εξόδου. Κάθε κρυφό στρώμα περιέχει πολλαπλούς κόμβους και τα στρώματα συνδέονται μεταξύ τους μέσω σταθμισμένων ακμών. Τα κρυφά στρώματα εκτελούν διάφορους μετασχηματισμούς και υπολογισμούς στα δεδομένα εισόδου, επιτρέποντας στο δίκτυο να μαθαίνει πολύπλοκα μοτίβα.

Στρώμα Εξόδου (Output Layer): Αυτό είναι το τελευταίο στρώμα του Νευρωνικού Δικτύου, όπου οι επεξεργασμένες πληροφορίες από τα κρυφά στρώματα μετατρέπονται στην έξοδο. Το στρώμα εξόδου παράγει το τελικό αποτέλεσμα πρόβλεψης ή ταξινόμησης.

Σε αυτό το σχήμα, τα βέλη υποδεικνύουν τη ροή των δεδομένων μέσω του δικτύου, ξεκινώντας από το στρώμα εισόδου, περνώντας μέσα από τα κρυφά στρώματα και φτάνοντας τελικά στο στρώμα εξόδου. Η διασυνδεδεμένη δομή των στρωμάτων δείχνει πώς τα νευρωνικά δίκτυα επεξεργάζονται και μαθαίνουν από τα δεδομένα μέσω πολλαπλών σταδίων υπολογισμού.

2.5.1 Πλεονεκτήματα

Τα Νευρωνικά Δίκτυα προσφέρουν αρκετά σημαντικά πλεονεκτήματα:

- **Δυνατότητα Εκμάθησης Μη Γραμμικών Σχέσεων:** Τα Νευρωνικά Δίκτυα μπορούν να μοντελοποιήσουν σύνθετες, μη γραμμικές σχέσεις μεταξύ εισόδων και εξόδων, καθιστώντας τα κατάλληλα για ένα μεγάλο αριθμό εργασιών.
- **Επεκτασιμότητα:** Μπορούν να διαχειριστούν μεγάλο όγκο δεδομένων και να αποδώσουν καλά με εισόδους υψηλών διαστάσεων.

- Ευελιξία: Τα Νευρωνικά Δίκτυα μπορούν να προσαρμοστούν για διάφορες εφαρμογές, όπως ταξινόμηση, παλινδρόμηση και εργασίες μάθησης χωρίς επίβλεψη.

2.5.2 Περιορισμοί

Παρά τα πλεονεκτήματά τους, τα Νευρωνικά Δίκτυα έχουν ορισμένους περιορισμούς:

- Υπολογιστικές Απαιτήσεις: Η εκπαίδευση των Νευρωνικών Δικτύων, ιδίως των Deep Neural Networks, απαιτεί σημαντική υπολογιστική ισχύ και χρόνο.
- Απαιτήσεις Δεδομένων: Συνήθως απαιτούν μεγάλο όγκο δεδομένων με ετικέτες για να αποδώσουν καλά.
- Ερμηνευσιμότητα: Τα Νευρωνικά Δίκτυα θεωρούνται συχνά "μαύρα κουτιά" επειδή οι διαδικασίες λήψης αποφάσεων δεν είναι εύκολα ερμηνεύσιμες [56].

2.5.3 Εφαρμογές στην Πρόβλεψη της Διαδικτυακής Συμπεριφοράς

Τα Νευρωνικά Δίκτυα χρησιμοποιούνται ευρέως στην πρόβλεψη της διαδικτυακής συμπεριφοράς των καταναλωτών, όπως οι προτιμήσεις των χρηστών, τα ποσοστά κλικ, οι προθέσεις αγοράς, τα δεδομένα ροής κλικ [57, 58, 59]. Η ικανότητά τους να μαθαίνουν από μεγάλα σύνολα δεδομένων και να αποτυπώνουν σύνθετα μοτίβα τα καθιστά ιδιαίτερα χρήσιμα σε αυτά τα πλαίσια.

2.6 Σύγκριση Αλγορίθμων Πρόβλεψης

Πίνακας 2.1: Σύγκριση αλγορίθμων πρόβλεψης.

ΑΛΓΟΡΙΘΜΟΣ	ΠΛΕΟΝΕΚΤΗΜΑΤΑ	ΜΕΙΟΝΕΚΤΗΜΑΤΑ
Λογιστική Παλινδρόμηση	Απλό και εύκολο στην ερμηνεία	Υποθέτει γραμμική σχέση μεταξύ των ανεξάρτητων μεταβλητών και των λογαριθμικών αποδόσεων, η οποία ενδέχεται να μην ισχύει σε πραγματικές περιπτώσεις
	Υπολογιστικά αποτελεσματική, ιδίως για μεγάλα σύνολα δεδομένων	Δεν είναι κατάλληλη για πολύπλοκα, μη γραμμικά δεδομένα
	Παρέχει πιθανολογικές προβλέψεις χρήσιμες για τη λήψη αποφάσεων	Ευαίσθητη στις ακραίες τιμές, οι οποίες μπορούν να αλλοιώσουν τις προβλέψεις
Δέντρα Αποφάσεων	Εύκολη οπτικοποίηση και ερμηνεία	Επιρρεπείς σε υπερπροσαρμογή, ιδίως με δέντρα μεγάλου βάθους
	Χειρίζεται τόσο αριθμητικά όσο και κατηγορικά δεδομένα	Μικρές αλλαγές στα δεδομένα μπορούν να αλλάξουν δραστικά τη δομή του δέντρου
	Καταγράφει μη γραμμικές σχέσεις χωρίς να χρειάζεται μετασχηματισμός των χαρακτηριστικών	Μπορεί να μεροληπτεί προς χαρακτηριστικά με περισσότερες κατηγορίες ή επίπεδα
Τυχαία Δάση	Μειώνει την υπερπροσαρμογή με τον μέσο όρο των προβλέψεων πολλαπλών Δέντρων Απόφασης	Υπολογιστικά δαπανηρό λόγω της ανάγκης εκπαίδευσης πολλαπλών Δέντρων Απόφασης
	Αντιμετωπίζει αποτελεσματικά τα δεδομένα που λείπουν και τα μεγάλα σύνολα δεδομένων	Λιγότερο ερμηνεύσιμο σε σύγκριση με τα μεμονωμένα Δέντρα Αποφάσεων, καθιστώντας δυσκολότερη την κατανόηση της διαδικασίας λήψης αποφάσεων
	Παρέχει κατατάξεις σπουδαιότητας χαρακτηριστικών, χρήσιμες για την επιλογή χαρακτηριστικών	Απαιτεί σημαντική μνήμη για την αποθήκευση και τη διαχείριση πολλαπλών δέντρων
Μηχανές Διανυσμάτων Υποστήριξης	Αποτελεσματική σε χώρους υψηλών διαστάσεων και χειρίζεται καλά μη γραμμικά δεδομένα χρησιμοποιώντας συναρτήσεις πυρήνα	Υπολογιστική δαπάνη, ιδίως για μεγάλα σύνολα δεδομένων
	Ανθεκτικός έναντι υπερπροσαρμογής, ιδίως σε σενάρια με σαφές διαχωρισμό περιθωρίων	Απαιτεί προσεκτικό συντονισμό παραμέτρων (π.χ. επιλογή πυρήνα, παράμετρος κανονικοποίησης) για την επίτευξη βέλτιστης απόδοσης
	Λειτουργεί καλά τόσο με εργασίες ταξινόμησης όσο και με εργασίες παλινδρόμησης	Περιορισμένη ερμηνευσιμότητα λόγω της πολυπλοκότητας του ορίου απόφασης
Νευρωνικά Δίκτυα	Δυνατότητα μοντελοποίησης εξαιρετικά πολύπλοκων, μη γραμμικών σχέσεων στα δεδομένα	Υπολογιστικά απαιτητικό και χρονοβόρο, ιδίως με τα βαθιά δίκτυα
	Εξαιρετικά επεκτάσιμη για μεγάλα σύνολα δεδομένων και μπορεί να μαθαίνει από τεράστιες ποσότητες δεδομένων	Απαιτεί μεγάλες ποσότητες επισημασμένων δεδομένων για αποτελεσματική εκπαίδευση
	Ευέλικτη και προσαρμόσιμη σε διάφορες εργασίες, όπως ταξινόμηση, παλινδρόμηση και μάθηση χωρίς επίβλεψη	Χαρακτήρας «μαύρου κουτιού»- δύσκολο να ερμηνευτεί και να κατανοηθεί ο τρόπος με τον οποίο γίνονται οι προβλέψεις

ΚΕΦΑΛΑΙΟ 3: ΕΦΑΡΜΟΓΗ ΤΩΝ ΑΛΓΟΡΙΘΜΩΝ ΣΤΗΝ ΠΡΟΒΛΕΨΗ ΔΙΑΔΙΚΤΥΑΚΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ

3.1 Λογιστική Παλινδρόμηση

Η Λογιστική Παλινδρόμηση είναι μια στατιστική μέθοδος που χρησιμοποιείται συχνά για την πρόβλεψη δυαδικών αποτελεσμάτων με βάσει μία ή περισσότερες μεταβλητές πρόβλεψης. Είναι ιδιαίτερα αποτελεσματική για εργασίες ταξινόμησης, όπου ο στόχος είναι η κατηγοριοποίηση των παρατηρήσεων σε μία από δύο κλάσεις. Η εφαρμογή της Λογιστικής Παλινδρόμησης στην πρόβλεψη της πρόθεσης αγοράς των ηλεκτρονικών αγοραστών και της αποτελεσματικότητας της ηλεκτρονικής διαφήμισης έχει αποδειχθεί σε αρκετές μελέτες.

3.1.1 Πρόβλεψη της Αγοραστικής Πρόθεσης των Online Αγοραστών

Η Διωνυμική Λογιστική Παλινδρόμηση εφαρμόστηκε για την πρόβλεψη της πρόθεσης αγοράς των διαδικτυακών αγοραστών. Οι Muda et al. (2020) εφάρμοσαν Διωνυμική Λογιστική Παλινδρόμηση για την πρόβλεψη της πρόθεσης αγοράς των διαδικτυακών αγοραστών. Η μελέτη περιλάμβανε βήματα προεπεξεργασίας, όπως η υπερ-δειγματοληψία για την αντιμετώπιση των ανισόρροπων δεδομένων και η επιλογή χαρακτηριστικών με τη χρήση του τεστ "Chi-Square"³ για την εξάλειψη περιττών μεταβλητών. Το τελικό μοντέλο Λογιστικής Παλινδρόμησης παρουσίασε ότι οι ανεξάρτητες μεταβλητές όπως η "Διοικητική Διάρκεια", η "Ενημερωτική Διάρκεια", η "Διάρκεια που σχετίζεται με το προϊόν", τα "Ποσοστά Αναπήδησης" και οι "Τιμές Σελίδας" επηρέασαν σημαντικά τις προθέσεις αγοράς. Το μοντέλο πέτυχε ακρίβεια 83.87% με 5-πλό "Cross-Validation" ⁴και 86.56% με 10-πλό "Cross-Validation", υποδεικνύοντας την ευρωστία και την αποτελεσματικότητά του στην πρόβλεψη της διαδικτυακής αγοραστικής συμπεριφοράς [36].

³ Chi-square: είναι μια στατιστική μέθοδος που ελέγχει τη σχέση μεταξύ δύο κατηγορηματικών μεταβλητών.

⁴ Cross-Validation (Διασταυρωμένη Επικύρωση): είναι μια τεχνική που μπορεί να χρησιμοποιηθεί στην μηχανική μάθηση και στην στατιστική με σκοπό την αξιολόγηση της απόδοσης ενός μοντέλου πρόβλεψης.

Για να αντιμετωπιστεί το ζήτημα της ανισορροπίας των δεδομένων, χρησιμοποιήθηκε η τεχνική συνθετικής υπερ-δειγματοληψίας μειοψηφίας (SMOTE)⁵ για την εξισορρόπηση των κλάσεων. Η επιλογή των χαρακτηριστικών πραγματοποιήθηκε με τη χρήση του τεστ "Chi-Square", το οποίο εντόπισε σημαντικούς προγνωστικούς παράγοντες, όπως "Διοικητική Διάρκεια", "Ενημερωτική Διάρκεια", "Διάρκεια σχετική με το προϊόν", "Ποσοστά Αναπήδησης" και "Τιμές Σελίδας". Στο μοντέλο Λογιστικής Παλινδρόμησης, το οποίο αξιολογήθηκε με τη χρήση 5πλής και 10πλής διασταυρούμενης επικύρωσης, χρησιμοποίησαν την Διωνυμική Λογιστική Παλινδρόμηση για να προβλέψουν την πρόθεση αγοράς των ηλεκτρονικών αγοραστών. Η μελέτη περιλάμβανε βήματα προεπεξεργασίας, όπως υπερ-δειγματοληψία για την αντιμετώπιση ασύμμετρων δεδομένων και επιλογή χαρακτηριστικών με τη χρήση τεστ chi-square για την εξάλειψη περιττών μεταβλητών. Το τελικό μοντέλο Λογιστικής Παλινδρόμησης έδειξε ότι μεταβλητές όπως η «Διοικητική Διάρκεια», η «Ενημερωτική Διάρκεια», η «Σχετική με το προϊόν διάρκεια», τα «Ποσοστά Αναπήδησης» και τα «Ποσοστά Σελίδας» επηρέασαν σημαντικά την πρόθεση αγοράς. Το μοντέλο πέτυχε ακρίβεια 83.87% με 5-πλό "Cross-Validation" και 86,56% με 10-πλό "Cross-Validation", υποδεικνύοντας την ευρωστία και την αποτελεσματικότητά του στην πρόβλεψη της διαδικτυακής αγοραστικής συμπεριφοράς [36].

Για να αντιμετωπιστεί το ζήτημα της ετερογένειας των δεδομένων, χρησιμοποιήθηκε η SMOTE για την εξισορρόπηση των κλάσεων. Η επιλογή των χαρακτηριστικών πραγματοποιήθηκε με τη χρήση του τεστ chi-square, το οποίο εντόπισε σημαντικούς προγνωστικούς παράγοντες, όπως η «Διοικητική Διάρκεια», η «Ενημερωτική Διάρκεια», η «Διάρκεια που σχετίζεται με το προϊόν», τα «Ποσοστά Αναπήδησης» και τα «Ποσοστά Σελίδων». Το μοντέλο Λογιστικής Παλινδρόμησης, το οποίο αξιολογήθηκε με τη χρήση 5πλής και 10πλής διασταυρούμενης επικύρωσης, έδειξε υψηλή ακρίβεια και γενικευσιμότητα. Η παρούσα μελέτη ανέδειξε τη σημασία των βημάτων προεπεξεργασίας για τη βελτίωση της απόδοσης των μοντέλων Λογιστικής Παλινδρόμησης στην πρόβλεψη των προθέσεων για ηλεκτρονικές αγορές.

⁵ SMOTE: είναι μια μέθοδος υπερ-δειγματοληψίας που χρησιμοποιείται σε περιπτώσεις μη ισορροπημένων δεδομένων σε προβλήματα ταξινόμησης [36].

3.1.2 Ταξινόμηση των “Κλικ” σε Διαφημίσεις Πελατών

Οι Dani & Ginting (2023) χρησιμοποίησαν τη Λογιστική Παλινδρόμηση για να προβλέψουν τα διαφημιστικά κλικ των πελατών σε ιστότοπους και πλατφόρμες κοινωνικής δικτύωσης. Η μελέτη τους συνέκρινε τη Λογιστική Παλινδρόμηση με τον αλγόριθμο k-πλησιέστερων γειτόνων (k-NN). Η Λογιστική Παλινδρόμηση υπερείχε του k-NN όσον αφορά την ακρίβεια, την ανάκληση, την ακρίβεια, το ⁶F1-Score και το AUC-ROC⁷. Οι ερευνητές χρησιμοποίησαν ένα σύνολο δεδομένων από τον ιστότοπο του Kaggle σχετικά με τη διαδικτυακή διαφήμιση μιας εταιρείας μάρκετινγκ. Συγκεκριμένα, το σύνολο δεδομένων αποτελούταν από 1000 παρατηρήσεις και 10 χαρακτηριστικά τα οποία ήταν ο ημερήσιος χρόνος παραμονής στον ιστότοπο, η ηλικία, το εισόδημα περιοχής, η ημερήσια χρήση του διαδικτύου, η θεματική γραμμή διαφήμισης, η πόλη, το φύλο Άνδρας, η χώρα, η χρονοσφραγίδα και η χρήση κλικ στη διαφήμιση. Το χαρακτηριστικό απόκρισης ήταν το Clicked on Ad. Αυτό το χαρακτηριστικό είχε δύο πιθανά αποτελέσματα που είναι 0 και 1, όπου το 0 αναφέρεται στην περίπτωση όπου ο χρήστης δεν έκανε κλικ στη διαφήμιση (κλάση 0), ενώ η κλάση 1 αναφέρεται στο σενάριο όπου ο χρήστης κάνει κλικ στη διαφήμιση (κλάση 1). Το μοντέλο Λογιστικής Παλινδρόμησης παρείχε πιθανοτικά αποτελέσματα, καθιστώντας ευκολότερη την ερμηνεία και την εφαρμογή του σε πραγματικές περιπτώσεις. Η μελέτη ανέδειξε τη σημασία της στόχευσης του σωστού κοινού για τη βελτιστοποίηση του κόστους διαφήμισης και την αύξηση των ποσοστών κλικ [31].

Η πιθανολογική φύση του μοντέλου Λογιστικής Παλινδρόμησης επιτρέπει στους διαφημιστές να αξιολογούν την πιθανότητα ένας χρήστης να κάνει κλικ σε μια διαφήμιση, επιτρέποντας έτσι πιο στοχευμένες και οικονομικά αποδοτικές στρατηγικές διαφήμισης [35, 60]. Η μελέτη υπογράμμισε επίσης τη σημασία της χρήσης κατάλληλων μετρικών αξιολόγησης, για την ολοκληρωμένη αξιολόγηση της απόδοσης των μοντέλων ταξινόμησης.

⁶ F1-Score: χρησιμοποιείται ως μέτρο αξιολόγησης απόδοσης ενός μοντέλου ταξινόμησης που συνδυάζει δύο κύριες μετρικές του Precision (Ακρίβεια) και Recall (Ανάκληση) [31].

⁷ AUC-ROC: είναι μια σημαντική μέτρηση που χρησιμοποιείται για την αξιολόγηση της απόδοσης των μοντέλων ταξινόμησης, σε περιπτώσεις μη ισορροπημένες.

3.1.3 Στοχευμένη Διαδικτυακή Διαφήμιση

Η μελέτη των Šoltés et al. (2020) επικεντρώθηκε στη χρήση της Λογιστικής Παλινδρόμησης για τη δημογραφική στόχευση στη διαδικτυακή διαφήμιση. Η μελέτη χρησιμοποίησε τη διωνυμική Λογιστική Παλινδρόμηση για την πρόβλεψη του φύλου των χρηστών του διαδικτύου και την πολυωνυμική Λογιστική Παλινδρόμηση για την εκτίμηση των ηλικιακών κατηγοριών τους. Οι αναλύσεις χρησιμοποίησαν δεδομένα από την DSP (DemandSide Platform) - μια πραγματική εκστρατεία μάρκετινγκ που πραγματοποιήθηκε μέσα σε ένα μήνα. Οι ερευνητές είχαν 160.544 παρατηρήσεις. Το σύνολο δεδομένων περιλάμβανε τη δραστηριότητα και τα δημογραφικά στοιχεία του χρήστη, συμπεριλαμβανομένων των USER_DAY (καθημερινή ημέρα = 1, Σαββατοκύριακο = 0), USER_HOUR (ώρα χρήσης του διαδικτύου), OPERATING_SYSTEM (αναγνωριστικό της έκδοσης του λειτουργικού συστήματος), BROWSER (αναγνωριστικό του προγράμματος περιήγησης), SITE_DOMAIN (τομέας ιστοτόπου), INTEREST (περιοχή ενδιαφέροντος του χρήστη), GENDER (γυναίκα = 1, άνδρας = 0), AGE (ηλικιακή ομάδα: 13-24, 25-44, 45-65), και TARGET (εάν ο χρήστης ανήκει στην ομάδα-στόχο, γυναίκες ηλικίας 25-44 ετών). Αναλύοντας τις προτιμήσεις των χρηστών για ιστότοπους και τομείς ενδιαφέροντος, τα λογιστικά μοντέλα παρείχαν ακριβείς προβλέψεις που βοήθησαν στην αποτελεσματικότερη στόχευση των διαφημίσεων. Τα μοντέλα επικυρώθηκαν χρησιμοποιώντας ένα μεγάλο σύνολο δεδομένων από μια πραγματική διαφημιστική εκστρατεία, αποδεικνύοντας την πρακτική χρησιμότητά τους για την ενίσχυση της ακρίβειας των στρατηγικών διαδικτυακού μάρκετινγκ [31].

Η συγκεκριμένη μελέτη αξιοποίησε τα μοντέλα Λογιστικής Παλινδρόμησης για την τμηματοποίηση των χρηστών του διαδικτύου βάσει δημογραφικών χαρακτηριστικών, επιτρέποντας στους διαφημιστές να παρέχουν εξατομικευμένο περιεχόμενο. Τα αποτελέσματα έδειξαν ότι οι προβλέψεις φύλου και ηλικίας με βάση τη διαδικτυακή συμπεριφορά των χρηστών βελτίωσαν σημαντικά την αποτελεσματικότητα των στοχευμένων διαφημίσεων. Η χρήση δεδομένων μεγάλης κλίμακας σε πραγματικό χρόνο εξασφάλισε την αξιοπιστία και τη δυνατότητα εφαρμογής των μοντέλων σε διάφορα πλαίσια του διαδικτυακού μάρκετινγκ. Γενικότερα, τα μεγάλα δεδομένα επιτρέπουν και παρέχουν στις επιχειρήσεις τη δυνατότητα να κατανοήσουν τις

απαιτήσεις και προσδοκίες των καταναλωτών και να επιτύχουν μια βιώσιμη επιχειρηματική επιτυχία [61].

3.1.4 Πρόβλεψη της Διαδικτυακής Αγοραστικής Συμπεριφοράς

Ο Billio (2017) εξέτασε τη συμβολή των διαφόρων τύπων προβλεπτικών παραγόντων στη συμπεριφορά αγορών σε ένα ηλεκτρονικό κατάστημα χρησιμοποιώντας μοντελοποίηση "Logit"⁸. Χρησιμοποίησαν μεταβλητές από τέσσερις κατηγορίες: *γενική συμπεριφορά ροής κλικ, λεπτομερείς πληροφορίες ροής κλικ, δημογραφικά στοιχεία πελατών και ιστορική αγοραστική συμπεριφορά*. Τα αποτελέσματα έδειξαν ότι οι λεπτομερείς μεταβλητές ροής κλικ ήταν οι πιο σημαντικές για την ταξινόμηση των πελατών ανάλογα με τη συμπεριφορά τους στις ηλεκτρονικές αγορές. Η μελέτη τους επισήμανε τη σημασία της συμπερίληψης μεγάλου εύρους μεταβλητών για την ενίσχυση της προβλεπτικής ισχύος του μοντέλου Λογιστικής Παλινδρόμησης [62].

Η έρευνα χρησιμοποίησε τεχνικές επιλογής μεταβλητών προς τα εμπρός και προς τα πίσω και τον αλγόριθμο αναζήτησης συνολικής βαθμολογίας των Furnival και Wilson, μέτρα αξιολόγησης ποιότητας μοντέλων παλινδρόμησης, για τον εντοπισμό του καλύτερου υποσυνόλου προβλεπτικών παραγόντων. Το τελικό μοντέλο περιλάμβανε μεταβλητές όπως «Recency of Visit», «Click Speed», «Page Views», «Gender» και «Trust». Η ανάλυση έδειξε ότι τα λεπτομερή δεδομένα ροής κλικ, όπως ο αριθμός των σελίδων που έχουν προβληθεί και η ταχύτητα κλικ, ήταν καθοριστικής σημασίας για την πρόβλεψη των ηλεκτρονικών αγορών. Αυτή η ολοκληρωμένη προσέγγιση για την επιλογή των μεταβλητών εξασφάλισε ότι συμπεριλήφθηκαν οι πιο σχετικοί προγνωστικοί παράγοντες, με αποτέλεσμα ένα ισχυρό και ακριβές μοντέλο.

3.1.5 Αποτελέσματα της Λογιστικής Παλινδρόμησης στην Πρόβλεψη των Προθέσεων Αγοράς των Καταναλωτών

Η εφαρμογή της Λογιστικής Παλινδρόμησης στην πρόβλεψη των προθέσεων αγοράς των καταναλωτών έδειξε ποικίλη αποτελεσματικότητα σε σύγκριση με πιο προηγμένα

⁸ Logit: είναι μια μοντελοποίηση που χρησιμοποιείται για την ανάλυση και πρόβλεψη δυαδικών αποτελεσμάτων (0 ή 1, επιτυχία ή αποτυχία, ναι ή όχι κ.α.) [63].

μοντέλα. Στη μελέτη που διεξήχθη από τους Zheng και Liu (2018), αναπτύχθηκε ένα σύστημα υποστήριξης αποφάσεων για την πρόβλεψη των προθέσεων αγοράς των καταναλωτών κατά τη διάρκεια περιήγησης. Στην έρευνα χρησιμοποιήθηκε ένα σύνολο δεδομένων που συλλέχθηκε από μια μεγάλη ευρωπαϊκή επιχείρηση λιανικού εμπορίου ηλεκτρονικού εμπορίου που πουλάει διάφορα καταναλωτικά προϊόντα. Το σύνολο δεδομένων περιείχε 9,249,729 ανώνυμες συνεδρίες περιήγησης και 509,696 (5.51%) συνεδρίες περιήγησης οδηγούν σε τελική αγορά από τους πελάτες. Καταγράφηκαν 52,739 συναλλαγές προβολής και αγοράς προϊόντων. Η μελέτη περιλάμβανε ως μεταβλητές τον αριθμό των κλικ, τον αριθμό των κατηγοριών, των εμπορικών σημάτων, των προϊόντων και των προωθητικών ενεργειών που είδαν οι χρήστες σε μια περίοδο περιήγησης. Πρόσθετα χαρακτηριστικά κατέγραψαν την εντροπία κατηγορίας περιήγησης, την εντροπία μάρκας και την εντροπία προϊόντος, που αντικατοπτρίζουν την ποικιλομορφία των προβαλλόμενων αντικειμένων. Περιλαμβάνονται επίσης η διάρκεια της συνεδρίας σε δευτερόλεπτα και οι χρονικές πτυχές, όπως η ώρα, η ημέρα και ο μήνας έναρξης της συνεδρίας. Επιπλέον, προσδιορίστηκε το ποσοστό κλικ της πιο συχνά προβαλλόμενης κατηγορίας, μάρκας και προϊόντος, καθώς και το ποσοστό αγοράς του πιο συχνά προβαλλόμενου προϊόντος. Το σύστημα αυτό κατηγοριοποίησε τις δραστηριότητες περιήγησης στο διαδίκτυο σε διαδικτυακές περιηγήσεις που προσανατολίζονται στην αγορά και σε γενικές περιόδους χρησιμοποιώντας μηχανές ακραίας διαβάθμισης (extreme gradient boosting). Ενώ το μοντέλο Λογιστικής Παλινδρόμησης χρησίμευσε ως μέτρο σύγκρισης, οι μηχανές ακραίας διαβάθμισης το ξεπέρασαν σημαντικά, επιτυγχάνοντας ανάκληση 41.81% και F-score 34.35% [64]. Αυτό παρείχε ένα σημαντικό εργαλείο υποστήριξης αποφάσεων για τη χάραξη στρατηγικών σχεδίων ψηφιακού μάρκετινγκ από τους διαδικτυακούς έμπορους και διαφημιστές.

3.1.6 Υβριδικό Μοντέλο Πρόβλεψης

Οι Hu et al. (2020) πρότειναν ένα υβριδικό μοντέλο πρόβλεψης που συνδυάζει αλγόριθμους Λογιστικής Παλινδρόμησης και Μηχανής Διανυσμάτων Υποστήριξης για την πρόβλεψη της αγοραστικής συμπεριφοράς των χρηστών σε μια πλατφόρμα ηλεκτρονικού εμπορίου. Τα δεδομένα του δείγματος σε αυτό το άρθρο προέκυψαν

από την πραγματική αγοραστική συμπεριφορά των χρηστών στην πλατφόρμα κινητού ηλεκτρονικού εμπορίου της Alibaba, συμπεριλαμβανομένων πέντε πεδίων: αναγνωριστικό χρήστη, αναγνωριστικό προϊόντος, αναγνωριστικό κατηγορίας προϊόντος, τύπος συμπεριφοράς του χρήστη στο προϊόν (κλικ, αγαπημένο, προσθήκη στο καλάθι, αγορά) και χρόνος συμπεριφοράς. Το δείγμα περιείχε 16,880 χρήστες, 393,798 προϊόντα και 929 κατηγορίες προϊόντων μεταξύ της 1ης Ιουλίου 2017 και της 20ής Ιουλίου 2017. Το μοντέλο Λογιστικής Παλινδρόμησης χρησιμοποιήθηκε για την πιθανολογική του ερμηνεία, ενώ το SVM χρησιμοποιήθηκε για την ευρωστία του στο χειρισμό δεδομένων υψηλής διάστασης. Το υβριδικό μοντέλο ξεπέρασε τα μεμονωμένα μοντέλα όσον αφορά την ακρίβεια πρόβλεψης, αποδεικνύοντας την αποτελεσματικότητα του συνδυασμού πολλαπλών αλγορίθμων για τη βελτίωση της απόδοσης πρόβλεψης [65].

Το υβριδικό μοντέλο αξιοποίησε τα πλεονεκτήματα τόσο της Λογιστικής Παλινδρόμησης όσο και του SVM. Η Λογιστική Παλινδρόμηση παρείχε ένα πιθανοτικό πλαίσιο για την ταξινόμηση, ενώ το SVM παρείχε ευρωστία έναντι δεδομένων υψηλών διαστάσεων. Ο συνδυασμός αυτών των αλγορίθμων μέσω του "Soft Voting" ⁹ ενίσχυσε τη συνολική απόδοση πρόβλεψης. Τα εμπειρικά αποτελέσματα έδειξαν ότι το υβριδικό μοντέλο πέτυχε υψηλότερη ακρίβεια και καλύτερη γενίκευση σε σύγκριση με τα μεμονωμένα μοντέλα, καθιστώντας το ένα ισχυρό μέσο για την πρόβλεψη της αγοραστικής συμπεριφοράς των χρηστών.

3.1.7 Εφαρμογές & Αποτελέσματα

Οι εφαρμογές της Λογιστικής Παλινδρόμησης στην πρόβλεψη της διαδικτυακής συμπεριφοράς έχουν δείξει πολύ ενθαρρυντικά αποτελέσματα. Για παράδειγμα, το μοντέλο πρόθεσης αγοράς των Muda (2020) εντόπισε βασικούς παράγοντες που επηρεάζουν τη συμπεριφορά των διαδικτυακών αγορών, όπως ο χρόνος παραμονής στις σελίδες που σχετίζονται με το προϊόν και το ποσοστό αναπήδησης [36]. Ομοίως, οι Dani & Ginting (2023) ανέδειξαν την αποτελεσματικότητα της Λογιστικής

⁹ Soft Voting: είναι τεχνική συναρμολόγησης που χρησιμοποιείται σε αλγορίθμους μηχανικής μάθησης συνδυάζοντας τις προβλέψεις πολλών μοντέλων καταλήγοντας σε μια απόφαση [66].

Παλινδρόμησης στην πρόβλεψη των διαφημιστικών κλικ, η οποία είναι σημαντική για τη βελτιστοποίηση των διαδικτυακών διαφημιστικών προσπαθειών [31]. Οι Soltes (2020) παρείχαν πληροφορίες σχετικά με τον τρόπο με τον οποίο η Λογιστική Παλινδρόμηση μπορεί να χρησιμοποιηθεί για τη δημογραφική στόχευση, ενισχύοντας την αποτελεσματικότητα των διαδικτυακών διαφημίσεων με την εστίαση στα πιο συναφή τμήματα χρηστών [35]. Ο Billio (2017) παρουσίασαν την σημασία λεπτομερών δεδομένων ροής κλικ για την πρόβλεψη διαδικτυακών αγορών, ενώ προβλήθηκαν τα οφέλη των υβριδικών μοντέλων που συνδυάζουν τη Λογιστική Παλινδρόμηση με άλλους αλγορίθμους [62, 64].

3.2 Δέντρο Αποφάσεων

Οι αλγόριθμοι των Δέντρων Αποφάσεων χρησιμοποιούνται ευρέως στη Μηχανική Μάθηση για εργασίες ταξινόμησης λόγω της απλότητας, της ερμηνευσιμότητας και της αποτελεσματικότητάς τους. Αυτοί οι αλγόριθμοι διαιρούν τα δεδομένα σε υποσύνολα με βάση τις τιμές των χαρακτηριστικών εισόδου, παράγοντας ένα δενδροειδές μοντέλο αποφάσεων και τις πιθανές συνέπειές τους. Στο πλαίσιο της πρόβλεψης της διαδικτυακής αγοραστικής συμπεριφοράς και των διαδικτυακών ποσοστών κλικ, τα μοντέλα Δέντρων Αποφάσεων έχουν δείξει μεγάλες δυνατότητες.

3.2.1 Πρόβλεψη Ποσοστού “Κλικ” Μέσω Δέντρου Αποφάσεων

Οι Kolan et al. (2020) διερεύνησαν την εφαρμογή των Δέντρων Απόφασης για την πρόβλεψη του ποσοστού “κλικ” (CTR¹⁰) στη διαδικτυακή διαφήμιση. Η μελέτη τους επικεντρώθηκε στην κατανόηση του τρόπου με τον οποίο διάφορα χαρακτηριστικά, όπως ο εκδότης, τα δημογραφικά στοιχεία των χρηστών και τα στατιστικά στοιχεία του περιεχομένου της διαφήμισης, επηρεάζουν την πιθανότητα να γίνει κλικ σε μια διαφήμιση. Τα σύνολα δεδομένων για την εκπαίδευση και τη δοκιμή του αλγορίθμου αντλήθηκαν από τον ιστότοπο kaggle. Ο ταξινομητής Δέντρων Απόφασης

¹⁰ CTR: είναι μια βασική μετρική στο πεδίο της ψηφιακής διαφήμισης και του μάρκετινγκ, με την οποία μετρά την αποτελεσματικότητα των διαφημιστικών καμπανιών και την επιτυχία των διαφημίσεων στη προσέλκυση “κλικ” από τους χρήστες [41].

χρησιμοποιήθηκε λόγω της ικανότητάς του να χειρίζεται αποτελεσματικά τόσο κατηγορικά όσο και αριθμητικά δεδομένα [41].

Η μεθοδολογία περιλάμβανε τον διαχωρισμό του συνόλου δεδομένων σε σύνολα εκπαίδευσης και δοκιμής και την κατασκευή του Δέντρου Απόφασης με χρήση της Python¹¹. Ο αλγόριθμος του Δέντρου Αποφάσεων χώρισε με επαναλαμβανόμενο τρόπο τους κόμβους με βάση τις τιμές των χαρακτηριστικών, δημιουργώντας ένα μοντέλο που προέβλεπε με ακρίβεια το CTR. Η μελέτη έδειξε ότι ο ταξινομητής Δέντρου Αποφάσεων βελτίωσε την ακρίβεια πρόβλεψης, αναδεικνύοντας τη σημασία της χρήσης σχετικών δεδομένων χρηστών, διαφημιστών και πλατφόρμας για την πρόβλεψη της απόδοσης των διαφημίσεων. Συγκεκριμένα, η πρόβλεψη του ποσοστού κλικ σε διαφημίσεις είναι χρήσιμη για τη μείωση του συνολικού κόστους που επενδύεται στις διαφημίσεις. Η οπτικοποίηση του Δέντρου Αποφάσεων βοήθησε στην κατανόηση της ροής αποφάσεων και των αποτελεσμάτων, καθιστώντας το πολύτιμο μέσο για τους διαφημιστές.

3.2.2 Εκτίμηση της Πρόθεσης για Ηλεκτρονικές Αγορές

Ο Toral (2019) εξέτασε τη λειτουργία των Δέντρων Αποφάσεων για την εκτίμηση της πρόθεσης αγοράς στο διαδίκτυο. Ο στόχος της μελέτης ήταν η ανάπτυξη ενός μοντέλου πρόβλεψης με λιγότερα δεδομένα, χρησιμοποιώντας την επιλογή χαρακτηριστικών Fisher Score και τη διασταυρούμενη επικύρωση μοντέλων K-fold Cross-Validation. Το σύνολο δεδομένων, που προήλθε από το UCI Machine Learning Repository¹², περιλάμβανε μεταβλητές όπως οι προβολές σελίδων, τα ποσοστά αναπήδησης, τα ποσοστά εξόδου και τα δημογραφικά στοιχεία των επισκεπτών. Το σύνολο δεδομένων περιλάμβανε 12,330 συνεδρίες, με 1,908 να καταλήγουν σε αγορές και 10,422 χωρίς, αποκαλύπτοντας μια ανισοβαρή αναλογία αγορών προς μη αγορές, περίπου 15% προς 85%. Η βασική μεταβλητή αποτελέσματος ήταν το κατά

¹¹ Python: γλώσσα προγραμματισμού σε πεδία πληροφορικής και μηχανικής μάθησης.

¹² UCI Machine Learning Repository: είναι μια πηγή δεδομένων που χρησιμοποιείται για έρευνα και ανάπτυξη αλγορίθμων μηχανικής μάθησης.

πόσον μια συνεδρία καταλήγει σε αγορά, η οποία ταξινομήθηκε ως Διωνυμική (αληθές/λάθος).

Το μοντέλο Δέντρων Απόφασης εκπαιδεύτηκε με τη χρήση του MATLAB¹³ και αξιολογήθηκε με τη χρήση 10πλής διασταυρούμενης επικύρωσης για να διασφαλιστεί η ευρωστία. Τα αποτελέσματα έδειξαν ότι ένα μόνο χαρακτηριστικό, η «Αξία σελίδας», ήταν ιδιαίτερα προγνωστικό της αγοραστικής συμπεριφοράς. Το δέντρο αποφάσεων παρείχε ένα σαφές και ερμηνεύσιμο μοντέλο που οι επιχειρήσεις μπορούσαν να χρησιμοποιήσουν για να κατανοήσουν και να προβλέψουν τη συμπεριφορά των πελατών, ενισχύοντας έτσι τις στρατηγικές μάρκετινγκ και πωλήσεων.

3.2.3 Πρόβλεψη Αποχώρησης Πελατών

Ο Zhao (2023) εφάρμοσε τα μοντέλα Δέντρων Αποφάσεων και Τυχαίων Δασών για την πρόβλεψη της αποχώρησης πελατών σε τραπεζικό πλαίσιο. Το σύνολο δεδομένων αποτελούνταν από 23 στήλες που αντιπροσώπευαν 23 μεταβλητές, συμπεριλαμβανομένης της ηλικίας των πελατών, του μισθού, της οικογενειακής κατάστασης, του ορίου της πιστωτικής κάρτας, της κατηγορίας της πιστωτικής κάρτας κ.λπ., και 10,000 γραμμές που αντιπροσώπευαν 10,000 πελάτες, συμπεριλαμβανομένων των υφιστάμενων και των αποχωρήσεων. Επιπλέον περιλάμβανε διάφορα χαρακτηριστικά συμπεριφοράς των πελατών και ο στόχος ήταν να εντοπιστούν οι δυνητικοί αποχωρήσαντες, ώστε η τράπεζα να λάβει προληπτικά μέτρα. Η μελέτη περιλάμβανε προεπεξεργασία δεδομένων, επιλογή χαρακτηριστικών και διαχωρισμό του συνόλου δεδομένων σε σύνολα εκπαίδευσης και δοκιμής [40].

Εκπαιδεύτηκαν και αξιολογήθηκαν τόσο το δέντρο απόφασης όσο και το τυχαίο δάσος, με το δέντρο απόφασης να επιτυγχάνει βαθμολογία ακρίβειας περίπου 90.8%. Η ανάλυση της σημασίας των χαρακτηριστικών αποκάλυψε ότι τα χαρακτηριστικά που σχετίζονται με τις συναλλαγές ήταν τα πιο προγνωστικά για την απόσυρση. Το μοντέλο Δέντρου Αποφάσεων παρείχε πολύτιμες πληροφορίες σχετικά με τη

¹³ MATLAB: πλατφόρμα υπολογισμών και προγραμματισμού.

συμπεριφορά των πελατών, βοηθώντας την επιχείρηση να διατηρήσει τους πελάτες αποτελεσματικότερα.

3.2.4 Χρονική Ανάλυση της Διαδικτυακής Συμπεριφοράς των Καταναλωτών

Οι Pitka et al. (2024) χρησιμοποίησαν μοντέλα Δέντρων Αποφάσεων για να αναλύσουν τη συμπεριφορά των διαδικτυακών καταναλωτών σε βάθος χρόνου. Η μελέτη χρησιμοποίησε ένα σύνολο δεδομένων με 9,123 παραγγελίες αθλητικής διατροφής, υγιεινών τροφίμων, ρούχων και αξεσουάρ γυμναστικής, που κάλυψαν οκτώ έτη (2014-2021). Τα αρχικά δεδομένα περιλάμβαναν βάσεις δεδομένων που περιείχαν λεπτομέρειες σχετικά με τις παραγγελίες της σχετικής περιόδου αναφοράς, προφίλ πελατών, λεπτομέρειες προϊόντων και πληροφορίες για την προμήθεια αγαθών. Τα μοντέλα Δέντρων Αποφάσεων χρησιμοποιήθηκαν για την ταξινόμηση των παραγγελιών ανά τεταρτημόριο του έτους και μέρη της ημέρας [67].

Η ανάλυσή τους αποκάλυψε ότι τα πιο σημαντικά χαρακτηριστικά για την ταξινόμηση των παραγγελιών στο τέταρτο τρίμηνο του έτους ήταν η χώρα, η κατηγορία ακαθάριστου κέρδους και η μέθοδος παράδοσης. Για τις πρωινές παραγγελίες, το φύλο και η χώρα ήταν τα πιο σημαντικά χαρακτηριστικά. Η μελέτη κατέληξε στο συμπέρασμα ότι τα Δέντρα Αποφάσεων θα μπορούσαν να βοηθήσουν τις επιχειρήσεις να προσαρμόσουν τις στρατηγικές μάρκετινγκ σε συγκεκριμένα τμήματα πελατών με βάση χρονικούς και δημογραφικούς παράγοντες, ενισχύοντας το στοχευμένο μάρκετινγκ και τις εξατομικευμένες συστάσεις.

3.2.5 Πρόβλεψη της Συμπεριφοράς των Καταναλωτών με τη Χρήση του Αλγορίθμου Δέντρων Αποφάσεων

Οι Hu et al. (2024) διερεύνησαν την αποτελεσματικότητα των αλγορίθμων Δέντρων Αποφάσεων στην πρόβλεψη της συμπεριφοράς των καταναλωτών. Η μελέτη σύγκρινε τα Δέντρα Αποφάσεων με άλλες τεχνικές Μηχανικής Μάθησης, όπως τα Αναδρομικά Νευρωνικά Δίκτυα (RNN) και τα μοντέλα που βασίζονται στη λειτουργική απεικόνιση μαγνητικού συντονισμού. Η έρευνα βασίστηκε σε εμπειρικά δεδομένα από μια μεγάλη διαδικτυακή πλατφόρμα μόδας στην Ευρώπη. Τα Δέντρα

Αποφάσεων βρέθηκαν ιδιαίτερα αποτελεσματικά στον εντοπισμό προτύπων στη συμπεριφορά των καταναλωτών, παρέχοντας αξιοποιήσιμες πληροφορίες για τις επιχειρήσεις [65].

Η μελέτη έδειξε ότι τα Δέντρα Αποφάσεων μπορούν να ταξινομήσουν με ακρίβεια τις ομάδες καταναλωτών με βάση την πιθανότητα πραγματοποίησης αγορών. Το μοντέλο Δέντρων Απόφασης υπερείχε έναντι άλλων αλγορίθμων όσον αφορά την πρακτική εφαρμογή και την ευκολία ερμηνείας, καθιστώντας το πολύτιμο εργαλείο για τις επιχειρήσεις που στοχεύουν στην κατανόηση και την πρόβλεψη της συμπεριφοράς των πελατών.

3.2.6 Εφαρμογές & Αποτελέσματα

Η εφαρμογή των αλγορίθμων Δέντρων Αποφάσεων για την πρόβλεψη της διαδικτυακής συμπεριφοράς έχει οδηγήσει σε πολλά υποσχόμενα αποτελέσματα σε διάφορες μελέτες. Αποδείχθηκε ότι τα Δέντρα Αποφάσεων μπορούν να προβλέψουν με ακρίβεια το CTR λαμβάνοντας υπόψη ένα εύρος χαρακτηριστικών χρηστών και διαφημίσεων [41]. Παρουσιάστηκε ότι τα Δέντρα Αποφάσεων θα μπορούσαν να εκτιμήσουν αποτελεσματικά τις προθέσεις αγοράς στο διαδίκτυο με υψηλή ακρίβεια χρησιμοποιώντας λιγότερα σημεία δεδομένων, καθιστώντας το μια πρακτική προσέγγιση για επιχειρήσεις με περιορισμένους πόρους δεδομένων [37]. Επιβεβαιώθηκε η χρησιμότητα των Δέντρων Αποφάσεων στην πρόβλεψη της αποχώρησης των πελατών, παρέχοντας πολύτιμες πληροφορίες για στρατηγικές διατήρησης πελατών στον τραπεζικό κλάδο [40]. Ακόμα, τα Δέντρα Αποφάσεων έχουν τη δυνατότητα να χρησιμοποιηθούν για την ανάλυση χρονικών προτύπων στη συμπεριφορά των καταναλωτών στο διαδίκτυο, βοηθώντας στην ανάπτυξη στοχευμένων στρατηγικών μάρκετινγκ [67]. Τέλος, επισημάνθηκε η ευελιξία των Δέντρων Αποφάσεων στην πρόβλεψη της συμπεριφοράς των καταναλωτών, τονίζοντας την πρακτική εφαρμογή και την ερμηνευσιμότητά τους [65].

3.3 Εφαρμογές των Τυχαίων Δασών για την Πρόβλεψη Διαδικτυακής Συμπεριφοράς

Οι αλγόριθμοι Τυχαία Δάση έχουν αναγνωριστεί για την αξιοπιστία, την ακρίβεια και την ικανότητά τους να χειρίζονται μεγάλα σύνολα δεδομένων με σύνθετες αλληλεπιδράσεις. Αυτοί οι αλγόριθμοι κατασκευάζουν πολλαπλά δέντρα απόφασης κατά τη διάρκεια της εκπαίδευσης και εξάγουν τον τρόπο ταξινόμησης (ταξινόμηση) ή τη μέση πρόβλεψη (παλινδρόμηση) των μεμονωμένων δέντρων, βελτιώνοντας έτσι την απόδοση πρόβλεψης και μειώνοντας την υπερπροσαρμογή. Στον τομέα της πρόβλεψης της διαδικτυακής συμπεριφοράς, τα μοντέλα Τυχαίων Δασών έχουν επιδείξει σημαντικές δυνατότητες, ιδίως στους τομείς της διαδικτυακής αγοραστικής συμπεριφοράς, της απομάκρυνσης πελατών και της πρόβλεψης του ποσοστού κλικ.

3.3.1 Πρόβλεψη Διαδικτυακής Αγοραστικής Συμπεριφοράς

Οι Joshi et al (2018) διερεύνησαν τη χρήση μοντέλων Random Forest για την πρόβλεψη της συμπεριφοράς των Ινδών πελατών στις διαδικτυακές αγορές σε διάφορες κατηγορίες προϊόντων. Χρησιμοποιώντας μια έρευνα βάσει ερωτηματολογίου με δεδομένα που συλλέχθηκαν από 124 ερωτηθέντες σε 18 πολιτείες, η μελέτη είχε ως στόχο να κατανοήσει τους παράγοντες που επηρεάζουν τις προτιμήσεις για ηλεκτρονικές αγορές. Το μοντέλο Τυχαία Δάση χρησιμοποιήθηκε για την ανάλυση αυτών των δεδομένων, αποκαλύπτοντας υψηλή ευαισθησία για κατηγορίες όπως τα βιβλία και τα ηλεκτρονικά είδη, υποδεικνύοντας ισχυρή τάση για ηλεκτρονικές αγορές. Αντίθετα, κατηγορίες όπως οι ταινίες και ο αθλητικός εξοπλισμός παρουσίασαν υψηλότερη εξειδίκευση, υποδεικνύοντας προτίμηση για αγορές εκτός σύνδεσης. Η μελέτη αυτή αναδεικνύει την αποτελεσματικότητα του μοντέλου Τυχαία Δάση στην τμηματοποίηση της αγοράς και την πρόβλεψη των προτιμήσεων των καταναλωτών, βοηθώντας έτσι τους λιανοπωλητές στην ανάπτυξη στοχευμένων στρατηγικών μάρκετινγκ [40].

3.3.2 Εκτίμηση της Αποχώρησης Πελατών

Ο Zhao (2023) εφάρμοσε τα μοντέλα Random Forest και Decision tree για την πρόβλεψη της αποχώρησης πελατών στο τραπεζικό τομέα. Το σύνολο δεδομένων

συλλέχθηκε από τον ιστότοπο Kaggle. Το σύνολο δεδομένων αποτελούνταν από 23 στήλες που αντιπροσώπευαν 23 μεταβλητές, συμπεριλαμβανομένης της ηλικίας των πελατών, του μισθού, της οικογενειακής κατάστασης, του ορίου της πιστωτικής κάρτας, της κατηγορίας της πιστωτικής κάρτας, κ.λπ., και 10,000 γραμμές που αντιπροσώπευαν 10,000 πελάτες, συμπεριλαμβανομένων των υφιστάμενων και των αποχωρήσεων. Η μελέτη περιλάμβανε την προεπεξεργασία ενός συνόλου δεδομένων που περιείχε διάφορα χαρακτηριστικά συμπεριφοράς των πελατών, το διαχωρισμό του σε σύνολα εκπαίδευσης και δοκιμής και στη συνέχεια την εκπαίδευση των μοντέλων. Το μοντέλο Τυχαία Δάση πέτυχε βαθμολογία ακρίβειας περίπου 91%, ξεπερνώντας το μοντέλο Δέντρου Αποφάσεων. Η ανάλυση της σημαντικότητας των χαρακτηριστικών αποκάλυψε ότι τα χαρακτηριστικά που σχετίζονται με τις συναλλαγές ήταν τα πιο προγνωστικά για τη αποχώρηση. Αυτή η εφαρμογή δείχνει πώς τα μοντέλα Τυχαία Δάση μπορούν να παρέχουν πολύτιμες πληροφορίες για τη συμπεριφορά των πελατών, επιτρέποντας στις επιχειρήσεις να εφαρμόζουν αποτελεσματικές στρατηγικές διατήρησης [40].

3.3.3 Πρόβλεψη του Ποσοστού “Κλικ” Διαδρομής

Στον τομέα της διαδικτυακής διαφήμισης, οι Muda et al., (2020) χρησιμοποίησαν τα μοντέλα Random Forest για να προβλέψουν τα ποσοστά κλικ αναλύοντας δεδομένα επισκέψεων χρηστών σε μια πλατφόρμα ηλεκτρονικού εμπορίου. Η μελέτη χρησιμοποίησε το σύνολο δεδομένων για την πρόθεση αγοράς online αγοραστών από το UCI Machine Learning Repository, το οποίο περιλαμβάνει 12,330 παρατηρήσεις και 18 μεταβλητές. Τα δεδομένα, τα οποία αποτυπώνουν διάφορες πτυχές της συμπεριφοράς των διαδικτυακών αγορών, χωρίστηκαν σε 70% για εκπαίδευση και 30% για δοκιμή για την ανάπτυξη προγνωστικών μοντέλων. Οι βασικές μεταβλητές περιλάμβαναν παράγοντες όπως τα έσοδα (η μεταβλητή-στόχος), τα ποσοστά αναπήδησης, τα ποσοστά εξόδου και διάφορες κατηγορικές μεταβλητές όπως ο μήνας, τα λειτουργικά συστήματα και ο τύπος επισκέπτη. Ακόμα η έρευνα χρησιμοποίησε βήματα προεπεξεργασίας υπερ-δειγματοληψίας και επιλογής χαρακτηριστικών για να βελτιώσει την απόδοση του μοντέλου. Εφαρμόζοντας 5 φορές διασταυρωμένη επικύρωση, το συγκεκριμένο μοντέλο πέτυχε ακρίβεια

88.21%, αναδεικνύοντας την ικανότητά του να προβλέπει με ακρίβεια τη συμπεριφορά των “online” αγοραστών και παρέχοντας ένα πολύτιμο μέσο για τη βελτιστοποίηση των διαφημιστικών τοποθετήσεων και στρατηγικών [36].

3.3.4 Ενίσχυση της Ακρίβειας Πρόβλεψης με Υβριδικά Μοντέλα

Οι Lilhore et al. (2021) πρότειναν ένα υβριδικό σταθμισμένο τυχαίο δάσος (HWRF - Hybrid Weighted Random Forest) για τη βελτίωση της πρόβλεψης και της ταξινόμησης της συμπεριφοράς των ηλεκτρονικών αγορών. Αυτό το μοντέλο ενσωματώνει βάρη σε επίπεδο δέντρου για τη βελτίωση της ακρίβειας και χρησιμοποιεί τον αλγόριθμο C4.5 για την πρόβλεψη της συμπεριφοράς των καταναλωτών. Το σύνολο δεδομένων συλλέχθηκε από το διαδικτυακό σύνολο δεδομένων του Kaggle (πελάτες που αγοράζουν online). Αυτό το σύνολο δεδομένων περιελάμβανε κυρίως την ηλικιακή ομάδα του πελάτη, το εισόδημα, τον χρόνο που δαπανάται για ηλεκτρονικές αγορές, το φύλο, την κατάσταση των δύο τελευταίων αγορών, τον τύπο του πελάτη κ.λπ. Αυτό το σύνολο δεδομένων περιείχε 80,000 καταχωρίσεις διαφόρων πελατών. Τα πειραματικά αποτελέσματα έδειξαν ότι το μοντέλο HWRF ξεπέρασε τα παραδοσιακά μοντέλα Τυχαία Δάση και Naïve Bayes¹⁴, επιτυγχάνοντας μεγαλύτερη ακρίβεια στην πρόβλεψη της αγοραστικής συμπεριφοράς των πελατών. Αυτή η υβριδική προσέγγιση αναδεικνύει τις δυνατότητες των προηγμένων εκδόσεων του μοντέλου Τυχαία Δάση για την παροχή ακριβέστερων και αξιόπιστων προβλέψεων στο πλαίσιο της διαδικτυακής συμπεριφοράς των καταναλωτών [68].

3.3.5 Πρόβλεψη των Προθέσεων Αγοράς των Καταναλωτών

Η εφαρμογή των αλγορίθμων Τυχαίου Δάσους στην πρόβλεψη των προθέσεων αγοράς των καταναλωτών έχει επιδείξει σημαντική αποτελεσματικότητα, ιδίως στον χειρισμό της πολυπλοκότητας των περιβαλλόντων ηλεκτρονικού εμπορίου. Στην έρευνα που διεξήχθη από τους Sang & Wu (2022), αναπτύχθηκε ένα σύστημα πρόβλεψης της συμπεριφοράς των διαδικτυακών αγοραστών σε πραγματικό χρόνο

¹⁴ Naïve Bayes: τεχνική μηχανική μάθησης για ταξινόμηση δεδομένων, βασισμένη στο θεώρημα Bayes.

για την πρόβλεψη των προθέσεων αγοράς των επισκεπτών του ιστότοπου. Το σύστημα αυτό χρησιμοποίησε δεδομένα περιήγησης και επισκεπτών για τη δημιουργία ενός ισχυρού μοντέλου πρόβλεψης χρησιμοποιώντας τον αλγόριθμο Τυχαίου Δάσους. Συγκεκριμένα, το σύνολο δεδομένων που χρησιμοποιήθηκε περιλάμβανε πάνω από 10,000 δείγματα, με σημαντική ανισορροπία: σχεδόν το 90% των συνεδριών δεν οδήγησε σε συναλλαγές και λιγότερο από το 10% οδήγησε σε αγορές. Οι συνεχείς και διακριτές μεταβλητές υποβλήθηκαν σε επεξεργασία με τη χρήση αλγορίθμου διαχωρισμού για να εξασφαλιστεί η βέλτιστη απόδοση του μοντέλου. Η μελέτη είχε ως στόχο να αντιμετωπίσει τον έντονο ανταγωνισμό μεταξύ των πλατφόρμων ηλεκτρονικού εμπορίου παρέχοντας στους εμπόρους πληροφορίες σχετικά με τη συμπεριφορά των πελατών, επιτρέποντάς τους έτσι να κατανοήσουν καλύτερα και να ικανοποιήσουν τις ανάγκες των καταναλωτών. Η μελέτη ενσωμάτωσε τεχνικές υπερ-δειγματοληψίας για την ενίσχυση της απόδοσης και της επεκτασιμότητας του ταξινομητή, διασφαλίζοντας ότι το μοντέλο θα μπορούσε να χειριστεί αποτελεσματικά ανισόρροπα δεδομένα. Τα αποτελέσματα έδειξαν ότι ο αλγόριθμος Τυχαίου Δάσους πέτυχε υψηλή ακρίβεια 86.78% και βαθμολογία F1 0,60, ξεπερνώντας σημαντικά άλλα παραδοσιακά μοντέλα [69]. Το βασικό συμπέρασμα ήταν ότι το Random Forest αποδείχθηκε ότι είναι ένας αποτελεσματικός αλγόριθμος για την ταξινόμηση της συμπεριφοράς των πελατών σε πραγματικό χρόνο, με σημαντικές συνέπειες για την καθοδήγηση των επιχειρηματικών αποφάσεων στο ηλεκτρονικό εμπόριο.

Σε άλλη μελέτη, των Kabir et al, (2019), αναλύθηκαν διάφορα μοντέλα πρόβλεψης χρησιμοποιώντας εμπειρικά δεδομένα από "online" αγοραστές για να προσδιοριστεί ο πιο αποτελεσματικός αλγόριθμος για την πρόβλεψη των προθέσεων αγοράς. Το σύνολο δεδομένων αποτελούνταν από 12,330 σειρές, με αριθμητικές και κατηγορικές μεταβλητές και δύο κατηγορίες χρηστών: αυτούς που πραγματοποίησαν αγορά και αυτούς που δεν πραγματοποίησαν. Η μελέτη συνέκρινε την απόδοση των Δέντρων Απόφασης, των Τυχαιών Δασών, του Naive Bayes και του SVM, με το τυχαίο δάσος να αναδεικνύεται ως ο πιο ακριβής προγνωστικός δείκτης. Το τυχαίο δάσος πέτυχε ακρίβεια 89,55% στο σύνολο δεδομένων, ξεπερνώντας τους άλλους αλγορίθμους όσον αφορά την ακρίβεια και το "F1-Score". Ο συνδυαστικός χαρακτήρας του Τυχαιού

Δάσους, ο οποίος περιλαμβάνει τη δημιουργία πολλαπλών Δέντρων Απόφασης και τη μέση τιμή των προβλέψεών τους, συνέβαλε στην ευρωστία και την ικανότητά του να χειρίζεται πολύπλοκες σχέσεις δεδομένων. Ενώ τα Δέντρα Αποφάσεων παρουσίασαν καλύτερη ανάκληση λόγω των χαρακτηριστικών του συνόλου δεδομένων, το τυχαίο δάσος υπερείχε σε συνολική ισορροπία, ιδίως όσον αφορά την ακρίβεια και την εξειδίκευση [70].

Η μελέτη βελτίωσε περαιτέρω την απόδοση του μοντέλου Τυχαίου Δάσους μέσω μεθόδων συνόλου. Το Gradient Boosting¹⁵, που εφαρμόστηκε στον αλγόριθμο Τυχαίου Δάσους, αύξησε την ακρίβεια στο 90.34%, αποδεικνύοντας την αποτελεσματικότητα του συνδυασμού αυτών των τεχνικών. Άλλες μέθοδοι, όπως η συσσωμάτωση και η στοίβαξη, βελτίωσαν επίσης την απόδοση του μοντέλου, με την συσσωμάτωση να αυξάνει την ακρίβεια στο 90.25%. Το κύριο συμπέρασμά για τη χρήση Gradient Boosting και Random Forest για την πρόβλεψη της πρόθεσης αγοράς των online αγοραστών μπορεί να κάνει το ηλεκτρονικό εμπόριο πιο εύκολο, προσιτό και άνετο για όλους τους ενδιαφερόμενους.

3.3.6 Συγκριτική Ανάλυση της Λογιστικής Παλινδρόμησης και του Τυχαίου Δάσους

Οι Sunarya et al. (2024) διεξήγαγαν μια συγκριτική μελέτη των μοντέλων Λογιστικής Παλινδρόμησης και Τυχαίου Δάσους για την πρόβλεψη της αποχώρησης πελατών ηλεκτρονικού εμπορίου. Η έρευνα χρησιμοποίησε ένα σύνολο δεδομένων πελατών ηλεκτρονικού εμπορίου που προήλθε από το Kaggle και περιλάμβανε τόσο αριθμητικές όσο και κατηγορικές μεταβλητές, όπως η διάρκεια ζωής των πελατών, ο χρόνος παραμονής στην πλατφόρμα, το ιστορικό παραγγελιών και δημογραφικές πληροφορίες. Χρησιμοποιώντας το σύνολο δεδομένων πελατών ηλεκτρονικού εμπορίου, η μελέτη ανέδειξε τα πλεονεκτήματα και τις αδυναμίες και των δύο μοντέλων μέσω αυστηρής προεπεξεργασίας και αξιολόγησης των δεδομένων. Το Τυχαίο Δάσος επέδειξε ανώτερη απόδοση με 95% ακρίβεια, 95% ακρίβεια για την κατηγορία 0 και 93% για την κατηγορία 1.99% ανάκληση για την κατηγορία 0 και 74%

¹⁵ Gradient Boosting: τεχνική μάθησης για την κατασκευή ισχυρών μοντέλων πρόβλεψης, συνδυάζοντας ποικιλία αδύναμων μοντέλων για τη δημιουργία ενός τελικού ισχυρότερου.

για την κατηγορία 1, F1-Score 97% για την κατηγορία 0 και 82% για την κατηγορία 1 και AUC-ROC 0,97. Αντίθετα, η Λογιστική Παλινδρόμηση πέτυχε ακρίβεια 90%, ακρίβεια 91% για την κατηγορία 0 και 82% για την κατηγορία 1, ανάκληση 98% για την κατηγορία 0 και 50% για την κατηγορία 1, F1-Score 94% για την κατηγορία 0 και 62% για την κατηγορία 1 και AUC-ROC 0,88. Η υψηλότερη συνολική απόδοση του Τυχαίου Δάσους υποδηλώνει την καταλληλότητά του για την καταγραφή πολύπλοκων προτύπων στη συμπεριφορά των πελατών, παρέχοντας πολύτιμες πληροφορίες για την ενίσχυση των στρατηγικών διατήρησης πελατών στο ηλεκτρονικό εμπόριο [71]. Στους σημαντικούς προγνωστικούς παράγοντες περιλαμβάνονταν η δέσμευση των πελατών, η ικανοποίηση και το ιστορικό συναλλαγών.

3.3.7 Εφαρμογές & Αποτελέσματα

Η εφαρμογή των αλγορίθμων Τυχαίου Δάσους για την πρόβλεψη της διαδικτυακής συμπεριφοράς έχει δώσει πολλά υποσχόμενα αποτελέσματα σε διάφορες μελέτες. Καταδείχθηκε η ικανότητα του μοντέλου να προβλέπει με ακρίβεια τη συμπεριφορά των διαδικτυακών αγορών σε διάφορες κατηγορίες προϊόντων, παρέχοντας πολύτιμες πληροφορίες για τους εμπόρους λιανικής πώλησης [45]. Τονίστηκε η χρησιμότητα του Τυχαίου Δάσους στην πρόβλεψη της αποχώρησης πελατών στον τραπεζικό τομέα, παρέχοντας ένα ισχυρό εργαλείο για στρατηγικές διατήρησης πελατών [40]. Παρουσιάστηκε η αποτελεσματικότητα του μοντέλου στην πρόβλεψη των ποσοστών κλικ, ενισχύοντας την αποτελεσματικότητα της στόχευσης διαφημίσεων [36]. Δημιουργήθηκε μια υβριδική προσέγγιση, βελτιώνοντας περαιτέρω την ακρίβεια και την αξιοπιστία της πρόβλεψης [68]. Επιβεβαιώθηκε η ανώτερη απόδοση ταξινόμησης του μοντέλου σε σχέση με τα παραδοσιακά δέντρα απόφασης, αναδεικνύοντας την ανθεκτικότητα και την προσαρμοστικότητά του [43]. Παρουσιάστηκε συγκριτική ανάλυση, προβάλλοντας την ανώτερη απόδοση του Τυχαίου Δάσους έναντι της Λογιστικής Παλινδρόμησης στην πρόβλεψη της αποχώρησης πελατών ηλεκτρονικού εμπορίου. Αυτές οι μελέτες αναδεικνύουν συλλογικά την προσαρμοστικότητα και την αποτελεσματικότητα του Τυχαίου Δάσους σε διάφορες εφαρμογές πρόβλεψης διαδικτυακής συμπεριφοράς [71].

3.4 Μηχανές Διανυσμάτων Υποστήριξης (SVM) για την Πρόβλεψη της Διαδικτυακής Συμπεριφοράς

Οι Μηχανές Διανυσμάτων Υποστήριξης είναι μια ισχυρή και ευέλικτη τεχνική ταξινόμησης που χρησιμοποιείται στη Μηχανική Μάθηση. Είναι ιδιαίτερα αποτελεσματικές για δυαδικές εργασίες ταξινόμησης, παρέχοντας ισχυρή απόδοση με τη δημιουργία βέλτιστων υπερεπιπέδων που διαχωρίζουν τα σημεία δεδομένων που ανήκουν σε διαφορετικές κλάσεις. Στο πλαίσιο της πρόβλεψης της διαδικτυακής συμπεριφοράς, οι SVM έχουν εφαρμοστεί σε διάφορα προβλήματα, όπως η πρόθεση αγοράς, η απομάκρυνση πελατών και η πρόβλεψη αποχώρησης σε πραγματικό χρόνο.

3.4.1 Πρόβλεψη σε Πραγματικό Χρόνο της Αγοραστικής Πρόθεσης των Online Αγοραστών

Οι Sakar et al. (2019) δημιούργησαν ένα σύστημα ανάλυσης της συμπεριφοράς των “online” αγοραστών σε πραγματικό χρόνο, το οποίο προβλέπει την πρόθεση αγοράς και την πιθανότητα εγκατάλειψης του ιστότοπου χρησιμοποιώντας πολλαπλούς αλγορίθμους Μηχανικής Μάθησης, συμπεριλαμβανομένων των SVM. Η μελέτη χρησιμοποίησε συγκεντρωτικά δεδομένα προβολής σελίδας που αποτελούνταν από 12.330 συνεδρίες, με αριθμητικές και κατηγορικές μεταβλητές, πληροφορίες συνεδρίας και πληροφορίες χρήστη για να ενισχύσει την ακρίβεια της πρόβλεψης. Περίπου το 84,5% των συνεδριών δεν κατέληξαν σε αγορά. Ο ταξινομητής SVM συγκρίθηκε με τους ταξινομητές Τυχαίου Δάσους και Multilayer Perceptron (MLP)¹⁶. Ενώ ο MLP πέτυχε την υψηλότερη ακρίβεια και βαθμολογία F1, ο SVM εξακολουθούσε να παρουσιάζει ανταγωνιστικές επιδόσεις. Τα ευρήματα τόνισαν ότι τα χαρακτηριστικά από τα δεδομένα ροής κλικ, όπως οι προβολές σελίδων και τα ποσοστά εγκατάλειψης, συνέβαλαν σημαντικά στην ακρίβεια πρόβλεψης. Ακόμα η μελέτη ανέδειξε η αξία της ενσωμάτωσης χαρακτηριστικών που βασίζονται στη

¹⁶ MultiLayer Perception: είναι ένας τύπος τεχνητού Νευρωνικού Δικτύου που αποτελείται από πολλαπλά επίπεδα νευρώνων, συμπεριλαμβανομένου ενός εισόδου, ενός ή περισσότερων κρυφών επιπέδων και ενός επιπέδου εξόδου [72].

λειτουργία συνεδρίας με δεδομένα ροής κλικ για τη βελτίωση των μοντέλων πρόβλεψης σε περιβάλλοντα ηλεκτρονικού εμπορίου. Η μελλοντική έρευνα μπορεί να ενσωματώσει συστήματα συστάσεων για να προσφέρει εξατομικευμένο περιεχόμενο σε επισκέπτες που είναι πιθανό να μετατραπούν αλλά κινδυνεύουν να εγκαταλείψουν τον ιστότοπο. [72].

3.4.2 Πρόβλεψη της Συμπεριφοράς των Πελατών στις Ηλεκτρονικές Αγορές

Οι Maheswari & Priya (2017) χρησιμοποίησαν ταξινομητές SVM για να αναλύσουν και να κατηγοριοποιήσουν την αγοραστική συμπεριφορά των πελατών με βάση διάφορα χαρακτηριστικά όπως η ποιότητα, τα κίνητρα, το επάγγελμα και το επίπεδο εισοδήματος. Με την εφαρμογή τεχνικών εξόρυξης δεδομένων και την ταξινόμηση SVM, η μελέτη προέβλεψε αποτελεσματικά τη συμπεριφορά των πελατών, αναδεικνύοντας την ανθεκτικότητα του μοντέλου στο χειρισμό διαφορετικών και πολύπλοκων συνόλων δεδομένων. Τα ευρήματα έδειξαν ότι οι ταξινομητές SVM μπορούν να τμηματοποιήσουν αποτελεσματικά τους πελάτες με βάση τα αγοραστικά τους πρότυπα, βοηθώντας σε στοχευμένες στρατηγικές μάρκετινγκ. Συγκεκριμένα, από τα πειραματικά αποτελέσματα, ο πελάτης που έχει ηλικία μικρότερη των 7 ετών προσελκύεται περισσότερο να αγοράσει προϊόντα μέσω του διαδικτύου τα τελευταία χρόνια. Επιπλέον, κατά τη διάρκεια των προσφορών οι πελάτες φαίνεται να αγοράζουν περισσότερο. [73].

3.4.3 Υβριδικό Μοντέλο Πρόβλεψης για την Αγοραστική Συμπεριφορά

Οι Hu et al. (2020) εξέτασαν ένα υβριδικό μοντέλο πρόβλεψης που συνδυάζει Logistic Regression και SVM για την πρόβλεψη της αγοραστικής συμπεριφοράς των χρηστών στην πλατφόρμα ηλεκτρονικού εμπορίου της Alibaba. Το σύνολο δεδομένων, που περιλάμβανε 16,880 χρήστες, 393,798 προϊόντα και 929 κατηγορίες προϊόντων, διήρκεσε από την 1η Ιουλίου έως τις 20 Ιουλίου 2017 και περιλάμβανε χαρακτηριστικά όπως η συμπεριφορά των χρηστών και η αλληλεπίδραση με τα προϊόντα. Η εμπειρική ανάλυση αποκάλυψε ότι το υβριδικό μοντέλο υπερείχε των μεμονωμένων αλγορίθμων όσον αφορά την ακρίβεια πρόβλεψης. Αξιοποιώντας τα

πλεονεκτήματα τόσο της Λογιστικής Παλινδρόμησης όσο και του SVM, η μελέτη έδειξε βελτιωμένη γενικευσιμότητα και αποτελεσματικότητα στην πρόβλεψη της αγοραστικής συμπεριφοράς, προσφέροντας μια πολύτιμη προσέγγιση για το χειρισμό ασύμμετρων συνόλων δεδομένων που είναι κοινά σε σενάρια ηλεκτρονικού εμπορίου [64].

3.4.4 Συγκριτική Μελέτη των Τεχνικών Μηχανικής Μάθησης

Οι Charbuty & Abdulazeez (2021) πραγματοποίησαν μια συγκριτική μελέτη διαφόρων αλγορίθμων Μηχανικής Μάθησης, συμπεριλαμβανομένου του SVM, για την πρόβλεψη της αγοραστικής συμπεριφοράς των πελατών. Η μελέτη αξιολόγησε την απόδοση των διαφόρων ταξινομητών χρησιμοποιώντας την τεχνική της διασταυρούμενης επικύρωσης και τις καμπύλες ROC¹⁷. Οι SVM έδειξαν αξιοσημείωτη ακρίβεια, αναδεικνύοντας την αποτελεσματικότητά τους στην πρόβλεψη της συμπεριφοράς των πελατών σε σύγκριση με άλλους αλγορίθμους, όπως τα Δέντρα Αποφάσεων, ο KNN και ο Naïve Bayes. Η μελέτη αυτή ανέδειξε την ευελιξία των SVMs στην παροχή ακριβών και αξιόπιστων προβλέψεων σε διαφορετικά σύνολα δεδομένων [38].

3.4.5 Εφαρμογές & Αποτελέσματα

Η εφαρμογή των αλγορίθμων SVM στην πρόβλεψη της διαδικτυακής συμπεριφοράς έχει δώσει σημαντικά αποτελέσματα σε αρκετές μελέτες. Οι Sakar (2019) απέδειξαν ότι οι SVM, αν και υπερέχουν των MLP σε ορισμένες μετρήσεις, εξακολουθούν να παρέχουν πολύτιμες πληροφορίες για την πρόβλεψη των προθέσεων αγοράς χρησιμοποιώντας δεδομένα συνεδρίας και ροής κλικ [72]. Οι Maheswari & Priva (2017) απέδειξαν ότι η ευρωστία των SVMs στην κατηγοριοποίηση της συμπεριφοράς των πελατών με βάση διάφορα χαρακτηριστικά [73]. Οι Hu et al. (2020) παρουσίασαν τα πλεονεκτήματα ενός υβριδικού μοντέλου που συνδυάζει SVM με Λογιστική Παλινδρόμηση, επιτυγχάνοντας καλύτερη ακρίβεια πρόβλεψης για

¹⁷ ROC: είναι διαγνωστικό εργαλείο που εφαρμόζεται στην ανάλυση της απόδοσης δυαδικών ταξινομητών [38]

τη συμπεριφορά αγοράς σε μια μεγάλη πλατφόρμα ηλεκτρονικού εμπορίου [64]. Επιπλέον, οι Chrabuty & Abdulazeez (2021) επιβεβαίωσαν την αποτελεσματικότητα των SVM σε μια συγκριτική μελέτη, αποδεικνύοντας την ανώτερη απόδοσή τους στην πρόβλεψη της συμπεριφοράς των πελατών σε πολλαπλές τεχνικές Μηχανικής Μάθησης [38]. Αυτές οι μελέτες αναδεικνύουν συλλογικά την ευελιξία και την αποτελεσματικότητα των SVMs σε διάφορες εφαρμογές πρόβλεψης διαδικτυακής συμπεριφοράς.

3.5 Νευρωνικά Δίκτυα για την Πρόβλεψη της Διαδικτυακής Συμπεριφοράς

Τα Νευρωνικά Δίκτυα είναι μια ευέλικτη και αποτελεσματική τεχνική Μηχανικής Μάθησης, ικανή να μοντελοποιεί πολύπλοκες σχέσεις στα δεδομένα. Είναι ιδιαίτερα κατάλληλα για εργασίες που περιλαμβάνουν μεγάλα σύνολα δεδομένων και μη γραμμικά σχήματα, γεγονός που τα καθιστά ιδανικά για την πρόβλεψη της διαδικτυακής συμπεριφοράς. Τα Νευρωνικά Δίκτυα έχουν εφαρμοστεί εκτενώς σε διάφορους τομείς, όπως η πρόβλεψη της συμπεριφοράς των πελατών, η πρόθεση αγοράς σε πραγματικό χρόνο και η εγκατάλειψη του ιστότοπου.

3.5.1 Πρόβλεψη της Συμπεριφοράς των Πελατών με τη Χρήση Τεχνητών Νευρωνικών Δικτύων

Οι Zheng et al. (2013) διερεύνησαν την εφαρμογή Τεχνητών Νευρωνικών Δικτύων (ANN) για την πρόβλεψη της συμπεριφοράς των πελατών, χρησιμοποιώντας δεδομένα από το Foursquare.com, τα οποία περιλάμβαναν check-ins χρηστών με χρονοσφραγίδες, γεωγραφική θέση και σχόλια (βλ. Εικόνα 1). Το σύνολο δεδομένων περιλάμβανε 120,825 check-ins από διάφορους χρήστες, συμπεριλαμβανομένων τόσο δομημένων όσο και μη δομημένων δεδομένων, όπως γεωγραφικό μήκος, γεωγραφικό πλάτος και πληροφορίες κειμένου από τις κοινωνικές αλληλεπιδράσεις των χρηστών. Το μοντέλο ANN, ειδικότερα, επέδειξε ανώτερη απόδοση στην καταγραφή μη γραμμικών σχέσεων, ειδικά για χρήστες με επαρκή ιστορικά δεδομένα, δηλαδή εκείνους με πάνω από 100 check-ins που σχετίζονται με εστιατόρια. Το ANN πέτυχε ακρίβεια πρόβλεψης 93.13% χρησιμοποιώντας 5 φορές διασταυρούμενη

επικύρωση, ξεπερνώντας σημαντικά το μοντέλο SVM, το οποίο πέτυχε μόνο 54%. Η μελέτη υπογραμμίζει ότι τα ANN μπορούν να επεξεργαστούν αποτελεσματικότερα πολύπλοκα πρότυπα συμπεριφοράς πελατών ενσωματώνοντας τις χρονικές και κοινωνικές πτυχές των check-ins, προσφέροντας έτσι δυνατότητες για την ενίσχυση της πρόβλεψης της συμπεριφοράς των καταναλωτών στους τομείς του λιανικού εμπορίου και του ηλεκτρονικού εμπορίου. Οι μελλοντικές ερευνητικές κατευθύνσεις που προτείνονται από τους συγγραφείς περιλαμβάνουν την ενσωμάτωση της ανάλυσης συναισθήματος με βάση δεδομένα κειμένου και τη βελτιστοποίηση των χρόνων εκπαίδευσης των ANN [64].

3.5.2 Πρόβλεψη σε Πραγματικό Χρόνο της Αγοραστικής Πρόθεσης των Online Αγοραστών με τη Χρήση MLP και LSTM

Οι Sakar et al. (2018) ανέπτυξαν ένα σύστημα πραγματικού χρόνου για την ανάλυση της συμπεριφοράς των διαδικτυακών αγοραστών χρησιμοποιώντας Νευρωνικά Δίκτυα Πολλαπλών Επιπέδων “perceptron” (MLP) και Δίκτυα Μακράς Βραχύχρονης Μνήμης (LSTM). Το σύστημα αποτελείται από δύο ενότητες: η πρώτη προβλέπει την πρόθεση αγοράς χρησιμοποιώντας πληροφορίες συνεδρίας και χρήστη, ενώ η δεύτερη εκτιμά την πιθανότητα αποχώρησης από τον ιστότοπο χρησιμοποιώντας διαδοχικά δεδομένα ροής κλικ. Το MLP, με τον εύρωστο αλγόριθμο “backpropagation” που χρησιμοποιεί, ξεπέρασε τους άλλους ταξινομητές σε ακρίβεια και F1 Score. Το LSTM-RNN εκτίμησε αποτελεσματικά την πιθανότητα εγκατάλειψης του ιστότοπου, εξισορροπώντας την ακρίβεια και την εφαρμογή σε πραγματικό χρόνο. Η μελέτη κατέδειξε ότι ένα ελάχιστο υποσύνολο χαρακτηριστικών, το οποίο προέρχεται τόσο από δεδομένα περιόδου λειτουργίας όσο και από δεδομένα ροής κλικ, βελτιστοποιεί την απόδοση του συστήματος χωρίς να θυσιάζει την ακρίβεια. Η μελλοντική έρευνα πρότεινε την ενσωμάτωση ενός συστήματος συστάσεων για την προσφορά εξατομικευμένου περιεχομένου, με στόχο τη βελτίωση των ποσοστών μετατροπής μέσω της προσέλκυσης χρηστών που είναι πιθανό να εγκαταλείψουν τον ιστότοπο, ενώ έχουν πρόθεση αγοράς. [72].

3.5.3 Μείγμα Αναδρομικών Νευρωνικών Δικτύων για την Πρόβλεψη της Αγοραστικής Συμπεριφοράς

Οι Toth et al. (2017) μελέτησαν η χρήση ενός μείγματος Αναδρομικών Νευρωνικών Δικτύων για την πρόβλεψη της συμπεριφοράς των αγοραστών. Η μελέτη χρησιμοποίησε δεδομένα ροής κλικ που συλλέχθηκαν σε διάστημα δύο εβδομάδων και περιλάμβαναν 1,560,830 συνεδρίες. Αφού αποκλείστηκαν οι συνεδρίες με λιγότερες από τέσσερις προβολές σελίδων, παρέμειναν 198.936 συνεδρίες για ανάλυση. Κάθε συνεδρία περιλάμβανε μεταβλητές όπως ο τύπος της σελίδας (π.χ. προβολές προϊόντων, σελίδες αναζήτησης) και ο χρόνος παραμονής, χρόνος παραμονής σε κάθε σελίδα, οι οποίες αποτέλεσαν βασικά χαρακτηριστικά των μοντέλων. Η μελέτη σύγκρινε τις επιδόσεις των Αλυσίδων Markov υψηλής τάξης και των RNNs στην πρόβλεψη αποτελεσμάτων όπως οι αγορές, τα εγκαταλελειμμένα καλάθια αγορών και οι περιηγήσεις μόνο για πλοήγηση. Τα RNNs, ιδίως εκείνα που χρησιμοποιούν αρχιτεκτονικές μακράς βραχυπρόθεσμης μνήμης, έδειξαν ανώτερες επιδόσεις στη σύλληψη μακροπρόθεσμων εξαρτήσεων σε δεδομένα ροής κλικ. Η μελέτη έδειξε ότι τα RNNs θα μπορούσαν να βελτιώσουν σημαντικά την έγκαιρη πρόβλεψη της αγοραστικής συμπεριφοράς και ηλεκτρονικού εμπορίου σε πραγματικό χρόνο, βοηθώντας τις έγκαιρες και στοχευμένες παρεμβάσεις μάρκετινγκ [73].

3.5.4 Προσεγγίσεις Βαθιάς Μάθησης για την Πρόβλεψη Αγοραστικών Προθέσεων

Ο Vieira (2016) πραγματοποίησε μια μελέτη που συγκρίνει τις παραδοσιακές τεχνικές Μηχανικής Μάθησης με προηγμένες προσεγγίσεις βαθιάς μάθησης, όπως τα Δίκτυα Βαθιάς Πεποίθησης (DBN) και οι αυτόματοι κωδικοποιητές Stacked Denoising Auto-Encoders (SDA), για την πρόβλεψη των προθέσεων αγοράς. Το σύνολο δεδομένων αποτελούνταν από έξι μήνες αλληλεπιδράσεων των χρηστών με έναν ιστότοπο ηλεκτρονικού εμπορίου, συμπεριλαμβανομένων συμβάντων όπως προβολές σελίδων προϊόντων, προβολές καταλόγου και αγορές, αλλά αποκλείστηκαν τα διαφημιστικά κλικ και οι προβολές διαφημίσεων ως άσχετα με τους στόχους της μελέτης. Το σύνολο δεδομένων περιείχε περίπου 25,000 διαφορετικούς τύπους προϊόντων και για γεγονότα που αφορούσαν αγορές ή προβολές καταλόγου, ήταν

διαθέσιμες πρόσθετες λεπτομέρειες, όπως η τιμή. Η μελέτη έδειξε ότι τα μοντέλα βαθιάς μάθησης, μέσω της αποτελεσματικής εξαγωγής χαρακτηριστικών κατά την προ-εκπαίδευση, παρείχαν σημαντικές βελτιώσεις στο χειρισμό δεδομένων υψηλών διαστάσεων και στην ανισορροπία κλάσεων. Τεχνικές όπως τα DBNs και τα SDAs βελτίωσαν σημαντικά την ακρίβεια πρόβλεψης σε σύγκριση με γραμμικά μοντέλα. Η έρευνα κατέληξε στο συμπέρασμα ότι οι αρχιτεκτονικές βαθιάς μάθησης προσφέρουν σημαντικές βελτιώσεις σε εργασίες πρόβλεψης ηλεκτρονικού εμπορίου. Τα μοντέλα αυτά πέτυχαν υψηλότερη ακρίβεια και αξιοπιστία σε εργασίες πρόβλεψης, αναδεικνύοντας την αποτελεσματικότητά τους σε εφαρμογές ηλεκτρονικού εμπορίου [39].

3.5.5 RNNs για Δεδομένα Ροής “Κλικ” στο Ηλεκτρονικό Εμπόριο

Οι Koehn et al. (2020) διερεύνησαν η χρήση Αναδρομικών Νευρωνικών Δικτύων για την πρόβλεψη της συμπεριφοράς σε διαδικτυακές αγορές από δεδομένα ροής κλικ. Η μελέτη χρησιμοποίησε δεδομένα από έναν διαδικτυακό λιανοπωλητή μόδας, τα οποία διήρκεσαν από τις 20 Μαΐου έως τις 20 Ιουλίου 2018, και χώρισε τις συνεδρίες σε σύνολα εκπαίδευσης και δοκιμών με βάση αυτή τη δίμηνη περίοδο. Τα δεδομένα περιλάμβαναν πάνω από 550,000 συνεδρίες και 8.6 εκατομμύρια προβολές σελίδων. Η μελέτη κατέδειξε τη σημασία της μοντελοποίησης των αλληλεξαρτήσεων που ενυπάρχουν στα δεδομένα ροής κλικ. Συγκρίθηκαν διάφορες αρχιτεκτονικές RNN, συμπεριλαμβανομένων των μοντέλων LSTM και Gated Recurrent Unit (GRU), μηχανισμός πύλης που αφορά Αναδρομικά Νευρωνικά Δίκτυα, με τα RNN να υπερτερούν έναντι των παραδοσιακών προσεγγίσεων Μηχανικής Μάθησης με επίβλεψη. Τα αποτελέσματα έδειξαν ότι, ενώ τα RNN υπερείχαν σε ακρίβεια πρόβλεψης (AUC) σε σύγκριση με τα παραδοσιακά μοντέλα, δεν οδήγησαν σε υψηλότερα έσοδα εκστρατείας στη συγκεκριμένη εφαρμογή της ταξινόμησης μετατροπών. Ωστόσο, τα RNNs είχαν ιδιαίτερα καλές επιδόσεις στην πρόβλεψη των τιμών παραγγελίας, ξεπερνώντας τους παραδοσιακούς παλινδρομείς με μεγάλη διαφορά. Η μελλοντική έρευνα θα μπορούσε να διερευνήσει τη χρήση συνελκτικών στρωμάτων ή το διαχωρισμό δυναμικών και στατικών χαρακτηριστικών για τη

βελτίωση της αποτελεσματικότητας των αρχιτεκτονικών RNN σε εργασίες πρόβλεψης ηλεκτρονικού εμπορίου. [74].

3.5.6 Μοντελοποίηση της Αγοραστικής Συμπεριφοράς με Βάσει τα Μοτίβα “Κλικ”

Οι Wu et al. (2015) μελέτησαν την αποτελεσματικότητα διαφορετικών μοντέλων Νευρωνικών Δικτύων στην πρόβλεψη της αγοραστικής συμπεριφοράς με βάση τα μοτίβα κλικ στον ιστότοπο “YOOCHOOSE¹⁸”. Το σύνολο δεδομένων περιέχει ακολουθίες γεγονότων κλικ κατά τη διάρκεια συνόδων χρήστη, με κάθε γεγονός να περιλαμβάνει αναγνωριστικό συνόδου, αναγνωριστικό στοιχείου, χρονοσφραγίδα και ετικέτα τιμής, εάν το στοιχείο αγοράστηκε. Οι περισσότερες συνεδρίες περιέχουν λιγότερα από 30 συμβάντα κλικ, αν και ορισμένες έχουν έως και 200. Η έρευνα χρησιμοποίησε αμφίδρομα RNNs (BiRNNs), τύπος Αναδρομικού Νευρωνικού Δικτύου, για τη μοντελοποίηση της ακολουθίας των συμβάντων κλικ, αποτυπώνοντας τη χρονική δυναμική πιο αποτελεσματικά από τα παραδοσιακά μοντέλα. Τα αποτελέσματα έδειξαν ότι τα BiRNNs μπορούσαν να επιτύχουν συγκρίσιμες ή ανώτερες επιδόσεις με λιγότερα τεχνητά χαρακτηριστικά, ενισχύοντας την αποτελεσματικότητα της διαδικασίας πρόβλεψης, διατηρώντας παράλληλα υψηλή ακρίβεια. Αυτό αναδεικνύει την αποτελεσματικότητα και την ακρίβεια των BiRNNs στην πρόβλεψη του κλικ στην αγορά στο ηλεκτρονικό εμπόριο, ιδίως όσον αφορά τη μείωση της ανάγκης για εκτεταμένα μηχανικά τεχνητών χαρακτηριστικών. [75].

3.5.7 Πρόβλεψη της Συμπεριφοράς Διαδικτυακών Χρηστών με LSTM RNNs

Ο Batool (2023) ανέπτυξε ένα Beton μοντέλο βαθιάς μάθησης χρησιμοποιώντας Δίκτυα Μακράς Βραχύχρονης Μνήμης για την πρόβλεψη της συμπεριφοράς των διαδικτυακών χρηστών από δεδομένα ροής κλικ που συλλέχθηκαν επί έξι μήνες από έναν ιστότοπο ηλεκτρονικού εμπορίου. Το σύνολο δεδομένων περιλάμβανε λεπτομερή δεδομένα συνόδου χρήστη, όπως σελίδες που επισκέφθηκαν, προϊόντα

¹⁸ YOOCHOOSE: είναι μια διαδικτυακή πλατφόρμα που εμπεριέχει εξατομικευμένες λύσεις για συστάσεις προϊόντων και περιεχομένου. (χαρακτηριστικά του ιστοτόπου: προσωποποιημένες συστάσεις, ανάλυση συμπεριφοράς χρηστών, πολλαπλά κανάλια, ευκολία ενσωμάτωσης) [75].

που είδαν, αντικείμενα που προστέθηκαν στο καλάθι και αγορές. Πραγματοποιήθηκε εκτεταμένη προεπεξεργασία για την εξαγωγή βασικών χαρακτηριστικών και βελτιστοποίηση των αρχιτεκτονικών LSTM μέσω μεθόδων “Bayes”. , όπως η διάρκεια της συνεδρίας, ο χρόνος μεταξύ των κλικ, οι κατηγορίες προϊόντων που είδαμε και ο αριθμός των αναζητήσεων και των προσθηκών στο καλάθι. Το βελτιστοποιημένο μοντέλο LSTM πέτυχε ακρίβεια άνω του 85% στην πρόβλεψη ενεργειών των χρηστών, όπως αναζητήσεις, προσθήκες στο καλάθι και αγορές. Η έρευνα κατέδειξε την αποτελεσματικότητα των δικτύων LSTM στη μοντελοποίηση χρονικών προτύπων και παρείχε ένα ισχυρό πλαίσιο για την κατανόηση και την πρόβλεψη της συμπεριφοράς των διαδικτυακών χρηστών σε πραγματικό χρόνο. Ωστόσο, οι περιορισμοί περιλάμβαναν χαμηλή απόδοση σε σύντομες συνεδρίες χαμηλής δραστηριότητας και την απουσία δεδομένων πλαισίου, όπως τα δημογραφικά στοιχεία των χρηστών και οι εκστρατείες μάρκετινγκ, τα οποία θα μπορούσαν να επηρεάσουν τη συμπεριφορά. Οι μελλοντικές μελέτες θα μπορούσαν να ενσωματώσουν πλουσιότερες πηγές δεδομένων, όπως αρχιτεκτονικές βασισμένες στην προσοχή, για να βελτιώσουν τις προβλέψεις για μεγαλύτερες, πιο σύνθετες διαδρομές των χρηστών. Επιπλέον, η πρόβλεψη των χρονικών διαστημάτων μεταξύ των ενεργειών και της μακροπρόθεσμης συμμετοχής των χρηστών θα μπορούσε να βελτιώσει περαιτέρω τις δυνατότητες του μοντέλου [57].

3.5.8 Πρόβλεψη της Συμπεριφοράς των Καταναλωτών με ANNs

Ο Badea (2014) διερεύνησε την εφαρμογή των Τεχνητών Νευρωνικών Δικτύων (ANN) στην πρόβλεψη της συμπεριφοράς των καταναλωτών με βάση δεδομένα από την έρευνα «Romania - Financial Literacy and Financial Services Survey 2010». Η έρευνα, που αποτελείται από 2,389 παρατηρήσεις σε επίπεδο νοικοκυριού, φιλτράρεται σε 1,671 εγγραφές για ανάλυση, εστιάζοντας σε άτομα με και χωρίς τραπεζικές καταθέσεις. Οι βασικές μεταβλητές περιλάμβαναν την ηλικία, το επάγγελμα, το εισόδημα και την εκπαίδευση, μαζί με αρκετές άλλες που σχετίζονται με την πρόβλεψη της κατοχής τραπεζικών καταθέσεων. Η έρευνα εντόπισε τις σημαντικότερες μεταβλητές που επηρεάζουν την πιθανότητα τα άτομα να έχουν τραπεζικές καταθέσεις. Η μελέτη συνέκρινε τις επιδόσεις των ANN με την κλασική

διακριτική ανάλυση, αποδεικνύοντας ότι τα ANN παρείχαν υψηλότερα ποσοστά ανίχνευσης και καλύτερη ακρίβεια στον εντοπισμό ατόμων που είναι πιθανό να προβούν σε τραπεζικές καταθέσεις. Η μελέτη ανέδειξε την ευελιξία και την προβλεπτική ικανότητα των ANN, υποδηλώνοντας τη δυνητική χρησιμότητά τους σε διάφορες χρηματοοικονομικές και οικονομικές εφαρμογές. Τα ευρήματα υποστήριξαν την ευρύτερη υιοθέτηση των ANN για την ενίσχυση των διαδικασιών λήψης αποφάσεων και των στρατηγικών μάρκετινγκ μέσω βελτιωμένων προβλέψεων της συμπεριφοράς των καταναλωτών. Συγκεκριμένα, η ευελιξία και η προβλεπτική ικανότητα των ANNs τα κατέστησαν πολύτιμο εργαλείο για τη βελτίωση των στρατηγικών μάρκετινγκ και των διαδικασιών λήψης αποφάσεων στις χρηματοπιστωτικές υπηρεσίες. Αν και η διαμόρφωση των ANN είναι πιο χρονοβόρα από τα παραδοσιακά μοντέλα, απαιτούν λιγότερους μετασχηματισμούς δεδομένων και έλεγχο υποθέσεων. [58].

3.5.9 Εφαρμογές & Αποτελέσματα

Η εφαρμογή αλγορίθμων Νευρωνικών Δικτύων για την πρόβλεψη της διαδικτυακής συμπεριφοράς έχει δώσει σημαντικά αποτελέσματα σε διάφορες μελέτες. Οι Zheng et al. (2013) υπογράμμισαν τις δυνατότητες των ANNs για τη σύλληψη σύνθετων μοτίβων σε δεδομένα συμπεριφοράς πελατών [59]. Οι Sakar et al. (2018) κατέδειξαν την αποτελεσματικότητα του συνδυασμού δικτύων MLP και LSTM για την πρόβλεψη σε πραγματικό χρόνο των προθέσεων αγοράς και της εγκατάλειψης του ιστότοπου [72]. Οι Toth et al. (2017) παρουσίασαν ότι τα RNN μπορούν να ενισχύσουν την έγκαιρη πρόβλεψη της αγοραστικής συμπεριφοράς [73]. Ακόμα, οι Vieira (2016) & Batool (2023) παρουσίασαν τα πλεονεκτήματα των προσεγγίσεων βαθιάς μάθησης στο χειρισμό δεδομένων υψηλής διάστασης και στην ακριβή πρόβλεψη των ενεργειών των χρηστών [39, 57]. Οι Wu et al. (2015) τόνισαν την αποτελεσματικότητα των BiRNNs στην καταγραφή της χρονικής δυναμικής με λιγότερα χαρακτηριστικά [75]. Τέλος, ο Badea (2014) επιβεβαίωσε η ανώτερη απόδοση των ANNs έναντι των παραδοσιακών μεθόδων στην πρόβλεψη της συμπεριφοράς των καταναλωτών. Αυτές οι μελέτες αναδεικνύουν συλλογικά την ευελιξία και την αποτελεσματικότητα των

Νευρωνικών Δικτύων σε διάφορες εφαρμογές πρόβλεψης διαδικτυακής συμπεριφοράς [58].

3.6 Πραγματικές Εφαρμογές των Αλγορίθμων Πρόβλεψης

3.6.1 Το σύστημα σύστασης προϊόντων της Amazon

Το σύστημα συστάσεων προϊόντων της Amazon είναι ένα από τα πιο αναγνωρισμένα παραδείγματα εφαρμογών Μηχανικής Μάθησης στο ηλεκτρονικό εμπόριο. Μέσω της εξατομικευμένης μηχανής συστάσεων, η Amazon έχει καταφέρει να αυξήσει σημαντικά τα έσοδά της προσφέροντας προτάσεις προσαρμοσμένες στις προηγούμενες συμπεριφορές και προτιμήσεις των χρηστών. Οι αλγόριθμοι πίσω από αυτές τις συστάσεις αξιοποιούν διάφορες τεχνικές Μηχανικής Μάθησης, με έμφαση στο συνεργατικό φιλτράρισμα, το φιλτράρισμα βάσει περιεχομένου και τα μοντέλα βαθιάς μάθησης [76].

Συνεργατικό Φιλτράρισμα

Η Amazon χρησιμοποιεί αλγορίθμους συνεργατικού φιλτραρίσματος για να προτείνει προϊόντα με βάση τις αλληλεπιδράσεις των χρηστών, όπως το ιστορικό περιήγησης, τις προηγούμενες αγορές και τα αντικείμενα που έχουν προβληθεί ή αξιολογηθεί. Το συνεργατικό φιλτράρισμα λειτουργεί συνήθως με δύο τρόπους: προσεγγίσεις με βάση τον χρήστη και προσεγγίσεις με βάση τα αντικείμενα. Στην προσέγγιση με βάση τον χρήστη, οι συστάσεις γίνονται με τον εντοπισμό χρηστών με παρόμοιες αγοραστικές συνήθειες και την υπόδειξη ειδών που έχουν αγοράσει. Αντίθετα, η προσέγγιση που βασίζεται σε στοιχεία εντοπίζει παρόμοια στοιχεία με βάση τις αξιολογήσεις ή τις αλληλεπιδράσεις των χρηστών και τα προτείνει στον χρήστη [77].

Φιλτράρισμα Βάσει Περιεχομένου

Εκτός από το συνεργατικό φιλτράρισμα, η Amazon χρησιμοποιεί επίσης το φιλτράρισμα βάσει περιεχομένου, όπου τα χαρακτηριστικά των προϊόντων (όπως οι περιγραφές, οι κατηγορίες και τα χαρακτηριστικά) αντιστοιχίζονται με το προφίλ ή τις προτιμήσεις του χρήστη [78]. Αυτή η προσέγγιση είναι χρήσιμη όταν εισάγονται νέα

αντικείμενα στην πλατφόρμα, τα οποία δεν έχουν ακόμη συγκεντρώσει επαρκή δεδομένα αλληλεπίδρασης των χρηστών για συνεργατικό φιλτράρισμα.

Συνδυάζοντας το συνεργατικό φιλτράρισμα και το φιλτράρισμα βάσει περιεχομένου, το υβριδικό σύστημα συστάσεων της Amazon μπορεί όχι μόνο να προτείνει δημοφιλή ή υψηλά βαθμολογημένα προϊόντα, αλλά και να προσαρμόζει τις συστάσεις για εξειδικευμένα είδη που ευθυγραμμίζονται με συγκεκριμένες προτιμήσεις των χρηστών [79]. Αυτό είναι ιδιαίτερα σημαντικό στις αγορές μακράς διάρκειας, όπου οι χρήστες μπορεί να ενδιαφέρονται για λιγότερο συνηθισμένα προϊόντα.

Βαθιά Μάθηση & Εξατομίκευση

Τα τελευταία χρόνια, μοντέλα βαθιάς μάθησης, ιδίως Νευρωνικά Δίκτυα, έχουν ενσωματωθεί στη μηχανή συστάσεων της Amazon για την περαιτέρω εξατομίκευση των προτάσεων προϊόντων. Τα μοντέλα βαθιάς μάθησης μπορούν να αναλύσουν τη συμπεριφορά των χρηστών σε πολύ βαθύτερο επίπεδο, εντοπίζοντας κρυμμένα μοτίβα στις αλληλεπιδράσεις των χρηστών και δίνοντας νόημα στις πολύπλοκες, μη γραμμικές σχέσεις των δεδομένων [76]. Τα Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks, RNN) και τα Συνελκτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks, CNN) είναι ιδιαίτερα χρήσιμα στην πρόβλεψη του τι θα ενδιαφέρει τους χρήστες με βάση τόσο την τρέχουσα όσο και την ιστορική συμπεριφορά [80].

Η ικανότητα της Amazon να κάνει συστάσεις σε πραγματικό χρόνο έχει επίσης βελτιωθεί σημαντικά μέσω της χρήσης βαθιάς μάθησης. Με τεράστιες ποσότητες δεδομένων συναλλαγών, τα μοντέλα βαθιάς μάθησης μπορούν να προβλέψουν την πιθανότητα ένας χρήστης να αγοράσει ένα προϊόν μέσα σε μια συγκεκριμένη περίοδο περιήγησης.

3.6.2 Αλγόριθμος σύστασης περιεχομένου του Netflix

Το σύστημα συστάσεων του Netflix υπήρξε θεμελιώδες για την επιτυχία του, επιτρέποντας στην πλατφόρμα να διατηρεί τους χρήστες και να αυξάνει τη δέσμευση παρέχοντας εξατομικευμένες προτάσεις περιεχομένου. Ο αλγόριθμος βασίζεται σε διάφορες προηγμένες τεχνικές Μηχανικής Μάθησης, όπως συνεργατικό

φιλτράρισμα, φιλτράρισμα βάσει περιεχομένου και μοντέλα βαθιάς μάθησης για την πρόβλεψη των ταινιών ή τηλεοπτικών εκπομπών που οι χρήστες είναι πιο πιθανό να παρακολουθήσουν.

Συνεργατικό Φιλτράρισμα

Όπως και η Amazon, έτσι και το Netflix βασίζεται σε μεγάλο βαθμό στο συνεργατικό φιλτράρισμα για να προτείνει περιεχόμενο στους χρήστες του. Το συνεργατικό φιλτράρισμα χρησιμοποιεί ιστορικά δεδομένα προβολής, όπως βαθμολογίες και ιστορικό παρακολούθησης, για να εντοπίσει χρήστες με παρόμοιες προτιμήσεις. Ο αλγόριθμος του Netflix, γνωστός ως αλγόριθμος Cinematch, κατασκευάστηκε αρχικά χρησιμοποιώντας παραδοσιακές τεχνικές παραγοντοποίησης πινάκων, οι οποίες αναλύουν τις σχέσεις μεταξύ χρηστών και ταινιών για τη δημιουργία συστάσεων [81].

Με την πάροδο του χρόνου, το Netflix προχώρησε πέρα από το βασικό συνεργατικό φιλτράρισμα σε πιο προηγμένες τεχνικές, όπως οι μηχανές παραγοντοποίησης, οι οποίες επέτρεπαν ακριβέστερη πρόβλεψη λαμβάνοντας υπόψη τις αλληλεπιδράσεις μεταξύ διαφορετικών παραγόντων, όπως οι προτιμήσεις είδους, τα μοτίβα παρακολούθησης των χρηστών και τα μεμονωμένα χαρακτηριστικά των σειρών [82]. Αυτή η καινοτομία επέτρεψε στο Netflix να δημιουργήσει μια πιο εξατομικευμένη και δυναμική εμπειρία για κάθε χρήστη.

Φιλτράρισμα Βάσει Περιεχομένου & Υβριδική Προσέγγιση

Το φιλτράρισμα βάσει περιεχομένου εξακολουθεί να παίζει ρόλο στη μηχανή συστάσεων του Netflix. Σε αυτή την προσέγγιση, το Netflix αναλύει τα χαρακτηριστικά των εκπομπών και των ταινιών (π.χ. είδος, ηθοποιοί, σκηνοθέτες) που έχει παρακολουθήσει προηγουμένως ένας χρήστης και στη συνέχεια συνιστά παρόμοιο περιεχόμενο. Αυτό είναι ιδιαίτερα χρήσιμο για νέους χρήστες ή όταν προστίθεται νέο περιεχόμενο στην πλατφόρμα [83]. Όπως συζητήθηκε από τους Gomez-Uribe & Hunt (2015), το Netflix χρησιμοποιεί ένα υβριδικό σύστημα συστάσεων που συνδυάζει συνεργατικό και βασισμένο στο περιεχόμενο φιλτράρισμα για τη βελτιστοποίηση των συστάσεων για όλους τους χρήστες, ανεξάρτητα από το ιστορικό παρακολούθησης [84].

Το υβριδικό σύστημα του Netflix μπορεί να προσφέρει συστάσεις ακόμη και σε περιπτώσεις αραιών δεδομένων (δηλαδή για νέους χρήστες με μικρό ιστορικό παρακολούθησης), χρησιμοποιώντας χαρακτηριστικά περιεχομένου και με τα δεδομένα σχετικά με τις ταινίες και τις εκπομπές.

Βαθιά Μάθηση & Εξατομίκευση

Το Netflix έχει ενσωματώσει μοντέλα βαθιάς μάθησης στη μηχανή συστάσεων του για να ενισχύσει περαιτέρω την εξατομίκευση. Αυτά τα μοντέλα μπορούν να αναλύουν μεγάλες ποσότητες συμπεριφοράς προβολής και χαρακτηριστικών περιεχομένου, εντοπίζοντας σύνθετα μοτίβα και τάσεις σε ό,τι προτιμούν οι χρήστες. Τα Νευρωνικά Δίκτυα, όπως τα CNN και τα RNN, είναι ιδιαίτερα αποτελεσματικά στην καταγραφή των διαχρονικών προτύπων στη συμπεριφορά θέασης των χρηστών. Αυτό επιτρέπει στο Netflix να κατανοήσει όχι μόνο ποιο περιεχόμενο μπορεί να αρέσει σε έναν χρήστη, αλλά και πότε είναι πιο πιθανό να παρακολουθήσει ορισμένους τύπους περιεχομένου [85].

Για παράδειγμα, τα RNN χρησιμοποιούνται για την πρόβλεψη της επόμενης επιλογής προβολής ενός χρήστη με βάση τη συμπεριφορά της τρέχουσας συνεδρίας και το μακροπρόθεσμο ιστορικό προβολής του. Το Netflix έχει επίσης υιοθετήσει αυτοκωδικοποιητές, έναν τύπο βαθύ Νευρωνικού Δικτύου, για να αναλύει τη συμπεριφορά των χρηστών σε λεπτομερές επίπεδο και να προτείνει περιεχόμενο που ευθυγραμμίζεται με τις λεπτές προτιμήσεις [86].

ΚΕΦΑΛΑΙΟ 4: ΣΥΜΠΕΡΑΣΜΑΤΑ

4.1 Σύνοψη των Ευρημάτων

Στην παρούσα μελέτη αναλύθηκαν σφαιρικά διάφοροι αλγόριθμοι πρόβλεψης για την κατανόηση και την πρόβλεψη της διαδικτυακής συμπεριφοράς στο πλαίσιο του ψηφιακού μάρκετινγκ. Το κύριο βάρος δόθηκε σε πέντε βασικούς αλγορίθμους: *Λογιστική Παλινδρόμηση*, *Δέντρα Αποφάσεων*, *Τυχαία Δάση*, *Μηχανές Διανυσμάτων Υποστήριξης* και *Νευρωνικά Δίκτυα*. Κάθε αλγόριθμος παρουσίασε διακριτά πλεονεκτήματα και αδυναμίες στην πρόβλεψη διαφόρων πτυχών της διαδικτυακής συμπεριφοράς, συμπεριλαμβανομένης της συμπεριφοράς των διαδικτυακών

αγορών, της απομάκρυνσης πελατών, των ποσοστών κλικ και άλλων αλληλεπιδράσεων των χρηστών.

Η *Λογιστική Παλινδρόμηση* έχει αποδειχθεί ιδιαίτερα αποτελεσματική σε εργασίες Διωνυμικής ταξινόμησης, οι οποίες είναι συνηθισμένες σε περιπτώσεις ψηφιακού μάρκετινγκ, όπως η πρόβλεψη αν ένας χρήστης θα κάνει κλικ σε μια διαφήμιση ή θα προβεί σε μια ηλεκτρονική αγορά. Η πιθανολογική ερμηνεία της Λογιστικής Παλινδρόμησης επιτρέπει στους υπεύθυνους μάρκετινγκ να αξιολογούν την πιθανότητα συγκεκριμένων ενεργειών των χρηστών, επιτρέποντας έτσι πιο στοχευμένες και αποτελεσματικές στρατηγικές μάρκετινγκ.

Τα *Δέντρα Αποφάσεων* είναι πολύτιμα για την απλότητα και την ερμηνευσιμότητά τους, γεγονός που τα καθιστά χρήσιμα για την τμηματοποίηση των χρηστών με βάση τη διαδικτυακή τους συμπεριφορά. Αυτή η τμηματοποίηση μπορεί να βοηθήσει τους υπεύθυνους μάρκετινγκ να προσαρμόσουν τις στρατηγικές τους σε διαφορετικές ομάδες χρηστών, ενισχύοντας την εξατομίκευση και τη συνάφεια των εκστρατειών μάρκετινγκ.

Τα *Τυχαία Δάση*, συγκεντρώνοντας τις προβλέψεις πολλαπλών Δέντρων Αποφάσεων, έχουν βελτιώσει την ακρίβεια και την ευρωστία των προβλέψεων. Αυτό ήταν ιδιαίτερα χρήσιμο σε σύνθετες περιπτώσεις ψηφιακού μάρκετινγκ που περιλαμβάνουν μεγάλα σύνολα δεδομένων και πολλαπλά χαρακτηριστικά χρηστών. Τα Τυχαία Δάση έχουν αποδείξει την αποτελεσματικότητά τους στην πρόβλεψη της αποχώρησης πελατών, της πρόθεσης αγοράς και των ποσοστών κλικ, που αποτελούν κρίσιμες μετρήσεις για τη βελτιστοποίηση των στρατηγικών μάρκετινγκ και την κατανομή των πόρων.

Οι *Μηχανές Διανυσμάτων Υποστήριξης* έχουν διαπρέψει σε χώρους υψηλών διαστάσεων, παρέχοντας ακριβείς ταξινομήσεις που είναι απαραίτητες για τη διάκριση μεταξύ διαφορετικών τύπων συμπεριφοράς των χρηστών. Στο ψηφιακό μάρκετινγκ, οι SVMs έχουν αποδειχθεί αποτελεσματικές στην τμηματοποίηση των πελατών και στην πρόβλεψη των ενεργειών τους με βάση λεπτομερή αρχεία

καταγραφής δραστηριοτήτων, επιτρέποντας έτσι την ακριβέστερη στόχευση και τη βελτίωση των αποτελεσμάτων του μάρκετινγκ.

Τα *Νευρωνικά Δίκτυα*, ιδίως τα μοντέλα βαθιάς μάθησης, όπως τα Αναδρομικά Νευρωνικά Δίκτυα και τα Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης, καταγράφουν πολύπλοκες, μη γραμμικές σχέσεις στα δεδομένα. Αυτά τα μοντέλα έχουν αποδειχθεί ιδιαίτερα αποτελεσματικά σε εργασίες πρόβλεψης σε πραγματικό χρόνο, όπως η πρόβλεψη προθέσεων αγοράς και η μοντελοποίηση της συμπεριφοράς των χρηστών με βάση μοτίβα κλικ. Η ικανότητα των Νευρωνικών Δικτύων να επεξεργάζονται τεράστιες ποσότητες δεδομένων και να εντοπίζουν περίπλοκα μοτίβα έχει σημαντικές συνέπειες για το εξατομικευμένο μάρκετινγκ και τη δέσμευση πελατών σε πραγματικό χρόνο.

Η ενσωμάτωση των βημάτων προεπεξεργασίας, όπως η επιλογή χαρακτηριστικών και η εξισορρόπηση δεδομένων, έχει βελτιώσει σημαντικά την απόδοση όλων των μοντέλων. Τα υβριδικά μοντέλα, τα οποία συνδυάζουν τα πλεονεκτήματα πολλαπλών αλγορίθμων, έχουν επιδείξει ανώτερες επιδόσεις σε σύγκριση με τα μεμονωμένα μοντέλα, παρέχοντας ακόμη πιο ισχυρές προβλέψεις. Αυτές οι εφαρμογές των αλγορίθμων πρόβλεψης παρείχαν πολύτιμες πληροφορίες για τη συμπεριφορά των διαδικτυακών καταναλωτών, επιτρέποντας στους υπευθύνους μάρκετινγκ να αναπτύξουν αποτελεσματικότερες στρατηγικές στοχευμένου μάρκετινγκ και να βελτιστοποιήσουν τη δέσμευση των πελατών.

Ωστόσο, ορισμένες συμπεριφορές αποδείχθηκαν πιο δύσκολο να προβλεφθούν. Τα Νευρωνικά Δίκτυα αντιμετώπισαν δυσκολίες στην πρόβλεψη της συμπεριφοράς των καταναλωτών σε συνεδρίες με περιορισμένες ή σπάνιες αλληλεπιδράσεις, όπου η σπανιότητα των δεδομένων οδήγησε σε λιγότερο ακριβή αποτελέσματα. Ομοίως, η Λογιστική Παλινδρόμηση αντιμετώπισε δυσκολίες με πιο σύνθετες, μη γραμμικές συμπεριφορές λόγω της εγγενούς παραδοχής της γραμμικότητας στις σχέσεις δεδομένων.

4.2 Σύγκριση Αλγορίθμων

Κατά τη σύγκριση των πέντε αλγορίθμων είναι προφανές ότι ο καθένας έχει μοναδικά πλεονεκτήματα ανάλογα με το συγκεκριμένο πλαίσιο εφαρμογής της πρόβλεψης της διαδικτυακής συμπεριφοράς στο ψηφιακό μάρκετινγκ.

Η Λογιστική Παλινδρόμηση χαρακτηρίζεται από την απλότητα, την ερμηνευσιμότητα και την πιθανολογική της έξοδο. Αυτό την καθιστά ιδιαίτερα κατάλληλη για εργασίες Διωνυμικής ταξινόμησης, όπως η πρόβλεψη προθέσεων αγοράς και κλικ σε διαφημίσεις, όπου το αποτέλεσμα είναι ένα σαφές ναι ή όχι. Η απλή εφαρμογή της και η ευκολία ερμηνείας της επιτρέπουν στους υπεύθυνους μάρκετινγκ να κατανοήσουν γρήγορα την επιρροή των διαφόρων προβλεπτικών παραγόντων. Ωστόσο, η Λογιστική Παλινδρόμηση υποθέτει μια γραμμική σχέση μεταξύ των ανεξάρτητων μεταβλητών και των λογαριθμικών αποδόσεων της εξαρτημένης μεταβλητής, η οποία μπορεί να μην ισχύει σε πολλά σενάρια του πραγματικού κόσμου που περιλαμβάνουν σύνθετες, μη γραμμικές αλληλεπιδράσεις. Καταδείχθηκε, σύμφωνα με έρευνα, η αποτελεσματικότητα σε σενάρια με σαφώς καθορισμένες, γραμμικές σχέσεις, παρέχοντας σαφείς πληροφορίες για στοχευμένες στρατηγικές μάρκετινγκ [36, 31]. Αντίθετα, τα ευρήματα άλλης έρευνας κατέδειξε τους περιορισμούς της Λογιστικής Παλινδρόμησης στο χειρισμό σύνθετων και υψηλών διαστάσεων δεδομένων που συναντώνται συνήθως σε περιβάλλοντα ηλεκτρονικού εμπορίου [64].

Τα *Δέντρα Αποφάσεων* προσφέρουν υψηλή ερμηνευσιμότητα και είναι απλά στην εφαρμογή, χειριζόμενα αποτελεσματικά τόσο αριθμητικά όσο και κατηγορικά δεδομένα. Είναι ιδιαίτερα χρήσιμα για την τμηματοποίηση της συμπεριφοράς των χρηστών και την κατανόηση της σημασίας των χαρακτηριστικών, καθιστώντας τα ιδανικά για σενάρια που απαιτούν διαφάνεια του μοντέλου. Ωστόσο, τα Δέντρα Αποφάσεων είναι επιρρεπή στην υπερπροσαρμογή, ιδίως με θορυβώδη δεδομένα, και ενδέχεται να μην είναι τόσο ακριβή όσο οι πιο σύνθετες μέθοδοι συνόλου. Τονίστηκε η χρησιμότητα τους στην πρόβλεψη των προθέσεων ηλεκτρονικών αγορών και των ποσοστών κλικ, αντίστοιχα [37, 41]. Παρά τους περιορισμούς τους, τα Δέντρα Αποφάσεων παρέχουν μια σαφή οπτική αναπαράσταση των διαδρόμων λήψης

αποφάσεων, η οποία είναι πολύτιμη για τους marketers που επιδιώκουν να κατανοήσουν και να εξηγήσουν τη συμπεριφορά των χρηστών.

Τα *Τυχαία Δάση* βελτιώνουν τα Δέντρα Αποφάσεων με τον μέσο όρο των προβλέψεων πολλαπλών δέντρων, ενισχύοντας την ακρίβεια και την ευρωστία και μειώνοντας παράλληλα την υπερπροσαρμογή. Αυτό τα καθιστά ιδιαίτερα αποτελεσματικά για σύνθετες εργασίες πρόβλεψης, όπως η αποχώρηση πελατών και η συμπεριφορά σε διαδικτυακές αγορές σε διάφορες κατηγορίες προϊόντων. Η ικανότητα να χειρίζονται καλά μεγάλα σύνολα δεδομένων και χώρους υψηλών διαστάσεων είναι ένα άλλο σημαντικό πλεονέκτημα. Παρουσιάστηκε η ανώτερη απόδοση των Τυχαίων Δασών σε διάφορες διεργασίες πρόβλεψης, παρέχοντας στους υπεύθυνους μάρκετινγκ αξιόπιστα εργαλεία για την ανάπτυξη έγκυρων και βασισμένων στα δεδομένα στρατηγικών. Ωστόσο, η φύση του συνόλου των Τυχαίων Δασών τα καθιστά λιγότερο ερμηνεύσιμα από τα μεμονωμένα Δέντρα Αποφάσεων και απαιτητικά σε υπολογισμούς, γεγονός που μπορεί να αποτελέσει μειονέκτημα για εφαρμογές πραγματικού χρόνου [40, 45, 70].

Οι *Μηχανές Διανυσμάτων Υποστήριξης* είναι ιδιαίτερα αποτελεσματικές σε χώρους υψηλών διαστάσεων, παρέχοντας ισχυρή απόδοση με την εύρεση βέλτιστων υπερεπιπέδων που διαχωρίζουν τα σημεία δεδομένων που ανήκουν σε διαφορετικές κλάσεις. Αυτό καθιστά τις SVM άριστες για εργασίες Διωνυμικής ταξινόμησης, όπως η διάκριση προτύπων συμπεριφοράς χρηστών και η πρόβλεψη τμημάτων πελατών με βάση λεπτομερή χαρακτηριστικά. Ωστόσο, οι SVM είναι υπολογιστικά ακριβές για μεγάλα σύνολα δεδομένων και απαιτούν προσεκτικό συντονισμό των παραμέτρων, ο οποίος μπορεί να αποτελέσει πρόκληση. Η αποτελεσματικότητα των SVMs αναδείχθηκε στην πρόβλεψη της συμπεριφοράς των πελατών και των προθέσεων αγοράς, αντίστοιχα. Παρόλο που είναι λιγότερο ερμηνεύσιμα από ορισμένα άλλα μοντέλα, τα SVM προσφέρουν ακριβείς και αξιόπιστες προβλέψεις, καθιστώντας τα πολύτιμα για λεπτομερείς εργασίες τμηματοποίησης [64, 73].

Τα *Νευρωνικά Δίκτυα* είναι ικανά να αποτυπώνουν πολύπλοκες, μη γραμμικές σχέσεις και είναι ιδιαίτερα αποτελεσματικά για μεγάλα σύνολα δεδομένων, ιδίως σε παραλλαγές βαθιάς μάθησης, όπως τα Αναδρομικά Νευρωνικά Δίκτυα και τα Δίκτυα

Μακράς Βραχυπρόθεσμης Μνήμης. Αυτά τα μοντέλα είναι ιδανικά για την πρόβλεψη συμπεριφοράς σε πραγματικό χρόνο, τη μοντελοποίηση διαδοχικών δεδομένων και εργασίες που περιλαμβάνουν σύνθετα πρότυπα, όπως η ανάλυση ροής κλικ και η πρόβλεψη συμπεριφοράς αγοράς. Παράλληλα, αναδείχθηκε η ανώτερη απόδοση των Νευρωνικών Δικτύων στην πρόβλεψη σε πραγματικό χρόνο και στη μοντελοποίηση της συμπεριφοράς των χρηστών από δεδομένα ροής κλικ. Ωστόσο, τα Νευρωνικά Δίκτυα απαιτούν μεγάλες ποσότητες δεδομένων και υπολογιστικών πόρων και οι διαδικασίες λήψης αποφάσεων τους είναι λιγότερο ερμηνεύσιμες, γεγονός που μπορεί να αποτελέσει περιορισμό. Παρά τις προκλήσεις αυτές, η ικανότητα των Νευρωνικών Δικτύων να επεξεργάζονται τεράστιες ποσότητες δεδομένων και να εντοπίζουν περίπλοκα μοτίβα τα καθιστά απαραίτητα για το εξατομικευμένο μάρκετινγκ και τη δέσμευση πελατών σε πραγματικό χρόνο [57, 72].

Κάθε αλγόριθμος έχει αποδείξει τη χρησιμότητά του σε διάφορες πτυχές του ψηφιακού μάρκετινγκ, παρέχοντας στους υπεύθυνους μάρκετινγκ μια σειρά εργαλείων για την ακριβέστερη κατανόηση και πρόβλεψη της συμπεριφοράς των διαδικτυακών καταναλωτών. Επιλέγοντας τον κατάλληλο αλγόριθμο με βάση το συγκεκριμένο πλαίσιο εφαρμογής, οι marketers μπορούν να βελτιστοποιήσουν τις στρατηγικές τους, να ενισχύσουν τη δέσμευση των πελατών και να επιτύχουν καλύτερα επιχειρηματικά αποτελέσματα.

4.3 Πρότυπα Συμπεριφοράς και Τάσεις στη Διαδικτυακή Συμπεριφορά

Η εφαρμογή αλγορίθμων Μηχανικής Μάθησης ανέδειξε διάφορα βασικά πρότυπα συμπεριφοράς στη διαδικτυακή δραστηριότητα των καταναλωτών:

- Προθέσεις Αγοράς: Η Λογιστική Παλινδρόμηση και τα Δέντρα Αποφάσεων προέβλεψαν αποτελεσματικά τις προθέσεις αγοράς των χρηστών, ιδίως όταν αναλύθηκαν χαρακτηριστικά που βασίζονται σε συνεδρίες, όπως οι προβολές σελίδων, οι κατηγορίες προϊόντων που προβλήθηκαν και η διάρκεια συνεδρίας. Ωστόσο, τα Νευρωνικά Δίκτυα ήταν καταλληλότερα για την

πρόβλεψη της αγοραστικής συμπεριφοράς σε συνεδρίες με υψηλή συμμετοχή των χρηστών ή όταν εμπλέκονταν δεδομένα ροής κλικ.

- Αποχώρηση Πελατών: Τα Τυχαία Δάση και τα Δέντρα Αποφάσεων ήταν ιδιαίτερα επιτυχημένα στην πρόβλεψη της αποχώρησης πελατών αναλύοντας μοτίβα στο ιστορικό περιήγησης, σε προηγούμενες αγορές και σε δημογραφικά δεδομένα. Βασικοί παράγοντες, όπως η συχνότητα συνεδρίας και το βάθος ροής κλικ, αναδείχθηκαν ως ισχυροί δείκτες αποχώρησης.
- Ποσοστά Κλικ: Τα μοντέλα Δέντρων Απόφασης και Τυχαίου Δάσους ανέδειξαν ότι μεταβλητές όπως ο χρόνος παραμονής σε συγκεκριμένες σελίδες, η τοποθέτηση διαφημίσεων και τα δημογραφικά στοιχεία των χρηστών επηρέασαν σημαντικά τα ποσοστά κλικ. Αυτά τα μοντέλα αποδείχθηκαν ιδιαίτερα ισχυρά στον εντοπισμό προτύπων στην αλληλεπίδραση των χρηστών με τις διαδικτυακές διαφημίσεις.
- Εγκατάλειψη Διαφημίσεων: Τα SVM και τα Νευρωνικά Δίκτυα ήταν ιδιαίτερα αποτελεσματικά στην πρόβλεψη του κατά πόσον οι χρήστες θα εγκατέλειπαν μια διαφημιστική καμπάνια. Τα μοντέλα πρόβλεψης κατέγραψαν τις λεπτές αλληλεπιδράσεις μεταξύ των χαρακτηριστικών του χρήστη και του περιεχομένου της διαφήμισης, βοηθώντας στον εντοπισμό του πότε ένας χρήστης είναι πιθανό να εγκαταλείψει μια διαφήμιση.

Παρά τις επιδόσεις αυτές, η πρόβλεψη **συμπεριφορών σύντομης διάρκειας** ή ενεργειών που συνεπάγονται ελάχιστη αλληλεπίδραση με ιστότοπους (π.χ. σύνοδοι με λίγα κλικ ή επισκέψεις) αποδείχθηκε πρόκληση για όλους τους αλγορίθμους, ιδίως για τα Νευρωνικά Δίκτυα και τα SVM. Η έλλειψη δεδομένων σε αυτές τις περιπτώσεις περιόρισε την ικανότητα των αλγορίθμων να διακρίνουν σημαντικές τάσεις.

4.4 Εφαρμογές στο Ψηφιακό Μάρκετινγκ

Τα αποτελέσματα της παρούσας μελέτης παρέχουν πρακτικές εφαρμογές στις στρατηγικές ψηφιακού μάρκετινγκ, προσφέροντας διάφορες οδούς για τη βελτιστοποίηση των εκστρατειών και των προσπάθειών δέσμευσης:

- Στοχευμένη Διαφήμιση: Αλγόριθμοι όπως η Λογιστική Παλινδρόμηση και τα Τυχαία Δάση μπορούν να χρησιμοποιηθούν για τη βελτιστοποίηση της στόχευσης των διαφημίσεων, προβλέποντας ποιοι χρήστες είναι πιο πιθανό να ασχοληθούν με τις διαφημίσεις με βάση τη διαδικτυακή τους συμπεριφορά. Αυτά τα μοντέλα μπορούν να βοηθήσουν στον εντοπισμό βασικών μεταβλητών όπως η διάρκεια συνεδρίας, οι προβολές προϊόντων και οι προηγούμενες αλληλεπιδράσεις, επιτρέποντας στους εμπόρους να εστιάσουν τους πόρους τους σε κοινά με υψηλότερες πιθανότητες μετατροπής.
- Εξατομίκευση: Τα Δέντρα Αποφάσεων και τα Νευρωνικά Δίκτυα μπορούν να εφαρμοστούν για την ενίσχυση της εξατομίκευσης στο μάρκετινγκ σε πραγματικό χρόνο. Με τη χρήση αυτών των αλγορίθμων, οι έμποροι μπορούν να τμηματοποιούν τους χρήστες πιο αποτελεσματικά, παρέχοντας προσαρμοσμένο περιεχόμενο που ανταποκρίνεται στις ατομικές προτιμήσεις. Για παράδειγμα, τα Νευρωνικά Δίκτυα μπορούν να προβλέψουν ποια προϊόντα είναι πιθανό να αγοράσουν οι χρήστες με βάση το ιστορικό περιήγησής τους, επιτρέποντας εξατομικευμένες συστάσεις και βελτιώνοντας τα ποσοστά μετατροπής.
- Βελτιστοποίηση της Τοποθέτησης Διαφημίσεων: Τα SVM και τα Δέντρα Αποφάσεων αποδείχθηκαν αποτελεσματικά στην πρόβλεψη των ποσοστών κλικ και της εγκατάλειψης διαφημίσεων. Οι έμποροι μπορούν να χρησιμοποιήσουν αυτές τις πληροφορίες για να βελτιστοποιήσουν τις τοποθετήσεις διαφημίσεων στοχεύοντας το σωστό κοινό τη σωστή στιγμή. Για παράδειγμα, τα Δέντρα Αποφάσεων μπορούν να βοηθήσουν στον προσδιορισμό της καλύτερης ώρας της ημέρας ή της καλύτερης δημογραφικής ομάδας χρηστών για την τοποθέτηση διαφημίσεων, αυξάνοντας τα ποσοστά κλικ και μειώνοντας τη σπατάλη διαφημιστικών δαπανών.
- Διατήρηση Πελατών: Τα Τυχαία Δάση ήταν ιδιαίτερα αποτελεσματικά στην πρόβλεψη της απομάκρυνσης πελατών, επιτρέποντας στις επιχειρήσεις να εντοπίζουν τους πελάτες που διατρέχουν κίνδυνο και να παρεμβαίνουν πριν

φύγουν. Εστιάζοντας σε συγκεκριμένους παράγοντες που συμβάλλουν στην απομάκρυνση, όπως η συχνότητα των συνεδριών ή τα πρότυπα προηγούμενων αγορών, οι έμποροι μπορούν να εφαρμόσουν στοχευμένες στρατηγικές διατήρησης για την επαναπροσέγγιση των χρηστών και τη βελτίωση της αφοσίωσης των πελατών.

- Μάρκετινγκ σε Πραγματικό Χρόνο: Τα Νευρωνικά Δίκτυα, ιδίως τα μοντέλα Μακράς Βραχυπρόθεσμης Μνήμης – LSTM, μπορούν να εφαρμοστούν σε σενάρια πραγματικού χρόνου όπου απαιτούνται άμεσες προβλέψεις, όπως ο εντοπισμός χρηστών που ενδέχεται να εγκαταλείψουν τα καλάθια αγορών τους ή η πρόβλεψη επικείμενων αγορών. Με την ενσωμάτωση αυτών των μοντέλων σε συστήματα μάρκετινγκ, οι επιχειρήσεις μπορούν να κάνουν έγκαιρες παρεμβάσεις, όπως η προσφορά εκπτώσεων ή εξατομικευμένων συστάσεων, για να διατηρήσουν τους πελάτες και να προωθήσουν τις μετατροπές.

4.5 Συνέπειες για τις Στρατηγικές Μάρκετινγκ

Τα συμπεράσματα της παρούσας μελέτης έχουν σημαντικές συνέπειες για την ανάπτυξη αποτελεσματικών στρατηγικών στοχευμένου μάρκετινγκ. Η κατανόηση των πλεονεκτημάτων και των περιορισμών κάθε αλγορίθμου πρόβλεψης επιτρέπει στους υπεύθυνους μάρκετινγκ να επιλέξουν την καταλληλότερη μέθοδο για τις συγκεκριμένες ανάγκες τους. Η Λογιστική Παλινδρόμηση, για παράδειγμα, μπορεί να χρησιμοποιηθεί για απλές εργασίες Διωνυμικής ταξινόμησης, όπως ο προσδιορισμός του αν ένας χρήστης θα κάνει κλικ σε μια διαφήμιση ή θα ολοκληρώσει μια αγορά. Αυτή η προσέγγιση παρέχει σαφείς και αξιοποιήσιμες πιθανότητες, επιτρέποντας στους marketers να λαμβάνουν αποφάσεις βάσει δεδομένων που ενισχύουν την ακρίβεια της στόχευσης [31, 36].

Τα Δέντρα Αποφάσεων και τα Τυχαία Δάση προσφέρουν πολύτιμες πληροφορίες σχετικά με τη σημασία των χαρακτηριστικών, βοηθώντας τους marketers να κατανοήσουν ποια χαρακτηριστικά επηρεάζουν περισσότερο τη συμπεριφορά των

πελατών. Εντοπίζοντας τους βασικούς οδηγούς των ενεργειών των χρηστών, οι αλγόριθμοι αυτοί επιτρέπουν πιο στοχευμένες παρεμβάσεις, βελτιστοποιώντας τις στρατηγικές μάρκετινγκ για την αντιμετώπιση συγκεκριμένων αναγκών και προτιμήσεων των καταναλωτών. Για παράδειγμα, παρουσιάστηκε πώς τα Τυχαία Δάση μπορούν να χρησιμοποιηθούν αποτελεσματικά για την πρόβλεψη της αποχώρησης των πελατών και της συμπεριφοράς τους στις διαδικτυακές αγορές, αντίστοιχα, παρέχοντας αξιοποιήσιμες πληροφορίες για στρατηγικές διατήρησης και δέσμευσης [40, 45].

Οι Μηχανές Διανυσμάτων Υποστήριξης παρέχουν ακριβή τμηματοποίηση της συμπεριφοράς των χρηστών, επιτρέποντας στους υπεύθυνους μάρκετινγκ να αναπτύσσουν εξαιρετικά εστιασμένες εκστρατείες για ξεχωριστές ομάδες πελατών. Η ικανότητα των SVMs να χειρίζονται δεδομένα υψηλής διάστασης τις καθιστά ιδιαίτερα χρήσιμες για τη διάκριση σύνθετων προτύπων συμπεριφοράς και την προσαρμογή των προσπαθειών μάρκετινγκ σε συγκεκριμένα τμήματα [60, 68]). Αυτή η ακρίβεια επιτρέπει τη δημιουργία εξατομικευμένων μηνυμάτων μάρκετινγκ που έχουν βαθιά απήχηση στο κοινό-στόχο, ενισχύοντας την αποτελεσματικότητα των εκστρατειών.

Τα *Νευρωνικά Δίκτυα* είναι ιδιαίτερα χρήσιμα για δυναμικές εφαρμογές πραγματικού χρόνου, όπως οι εξατομικευμένες συστάσεις και η πρόβλεψη άμεσων προθέσεων αγοράς. Η ικανότητά τους να επεξεργάζονται μεγάλες ποσότητες δεδομένων και να συλλαμβάνουν περίπλοκα μοτίβα τα καθιστά ιδανικά για τη βελτίωση της εμπειρίας των πελατών και την αύξηση των ποσοστών μετατροπής [57, 72]. Αξιοποιώντας τα Νευρωνικά Δίκτυα, οι επιχειρήσεις μπορούν να παρέχουν έγκαιρο και σχετικό περιεχόμενο στους χρήστες, προωθώντας τη δέσμευση και οδηγώντας σε υψηλότερα έσοδα.

Αξιοποιώντας τους κατάλληλους αλγόριθμους πρόβλεψης, οι επιχειρήσεις μπορούν να βελτιστοποιήσουν τις προσπάθειες μάρκετινγκ, να βελτιώσουν τη δέσμευση των πελατών και τελικά να οδηγήσουν σε υψηλότερα έσοδα. Η ενσωμάτωση αυτών των προηγμένων αναλυτικών τεχνικών στις στρατηγικές μάρκετινγκ επιτρέπει μια πιο

διαφοροποιημένη κατανόηση της συμπεριφοράς των καταναλωτών, οδηγώντας σε πιο αποτελεσματικές και αποδοτικές διαφημιστικές ενέργειες marketing [61].

4.6 Μελλοντικές Ερευνητικές Κατευθύνσεις

Η μελλοντική έρευνα σχετικά με τους αλγορίθμους πρόβλεψης και τη συμπεριφορά των διαδικτυακών καταναλωτών θα πρέπει να επικεντρωθεί σε διάφορους βασικούς τομείς. Ένας πολλά υποσχόμενος δρόμος είναι η περαιτέρω διερεύνηση υβριδικών μοντέλων που συνδυάζουν τα πλεονεκτήματα πολλαπλών αλγορίθμων για τη βελτίωση της ακρίβειας και της ευρωστίας της πρόβλεψης.

Τα υβριδικά μοντέλα, τα οποία ενσωματώνουν διαφορετικές τεχνικές Μηχανικής Μάθησης, έχουν δείξει δυνατότητες βελτίωσης της απόδοσης σε διάφορες εργασίες πρόβλεψης [65].

Η ανάπτυξη και βελτίωση αλγορίθμων ικανών να χειρίζονται ροές δεδομένων σε πραγματικό χρόνο είναι ένας άλλος σημαντικός τομέας για μελλοντική έρευνα. Η επεξεργασία δεδομένων σε πραγματικό χρόνο ενισχύει την αμεσότητα και τη συνάφεια των προβλέψεων, επιτρέποντας στις επιχειρήσεις να ανταποκρίνονται γρήγορα στις αλλαγές στη συμπεριφορά των καταναλωτών και στις συνθήκες της αγοράς [72]. Αυτή η ικανότητα είναι ιδιαίτερα σημαντική σε δυναμικά περιβάλλοντα, όπου η έγκαιρη λήψη αποφάσεων μπορεί να προσφέρει ανταγωνιστικό πλεονέκτημα.

Η ενίσχυση της ερμηνευσιμότητας πολύπλοκων μοντέλων, όπως τα Νευρωνικά Δίκτυα, είναι επίσης απαραίτητη. Η βελτιωμένη ερμηνευσιμότητα μπορεί να παρέχει σαφέστερες γνώσεις σχετικά με τη συμπεριφορά των καταναλωτών, διευκολύνοντας τους marketers να κατανοήσουν και να εμπιστευτούν τις προβλέψεις που παράγονται από αυτά τα μοντέλα [56]. Αυτή η διαφάνεια είναι αναγκαία για την αποδοχή των ενδιαφερομένων μερών και τη διασφάλιση της ηθικής χρήσης των τεχνολογιών Μηχανικής Μάθησης [75].

Η διασφάλιση της επεκτασιμότητας και της αποτελεσματικότητας στις αλγοριθμικές εφαρμογές είναι μια άλλη σημαντική ερευνητική κατεύθυνση. Καθώς τα σύνολα δεδομένων συνεχίζουν να αυξάνονται σε μέγεθος και πολυπλοκότητα, απαιτούνται κλιμακούμενες λύσεις για την επεξεργασία μεγάλου όγκου δεδομένων σε λογικά χρονικά πλαίσια χωρίς να διακυβεύεται η ακρίβεια [53].

Αυτή η επεκτασιμότητα θα επιτρέψει στις επιχειρήσεις να αξιοποιήσουν αποτελεσματικά τα μεγάλα δεδομένα, εξάγοντας πολύτιμες πληροφορίες από εκτεταμένα σύνολα δεδομένων. Η ενσωμάτωση πληροφοριών που σχετίζονται με το πλαίσιο, όπως χρονικά και χωρικά δεδομένα, μπορεί να βελτιώσει σημαντικά την ακρίβεια και τη συνάφεια των προβλέψεων συμπεριφοράς. Η ανάλυση πλαισίου επιτρέπει τη βαθύτερη κατανόηση του τρόπου με τον οποίο διάφοροι παράγοντες επηρεάζουν τις ενέργειες των καταναλωτών, οδηγώντας σε πιο ακριβείς και αποτελεσματικές στρατηγικές μάρκετινγκ [54].

Η διερεύνηση του τρόπου με τον οποίο η συμπεριφορά των καταναλωτών ποικίλλει σε διαφορετικές διαδικτυακές πλατφόρμες και συσκευές είναι ένας άλλος κρίσιμος τομέας για μελλοντική έρευνα. Η ανάλυση της συμπεριφοράς μεταξύ διαφορετικών πλατφόρμων μπορεί να προσφέρει μια ολοκληρωμένη εικόνα των αλληλεπιδράσεων των χρηστών, επιτρέποντας την ανάπτυξη πιο ολιστικών και αποτελεσματικών μοντέλων μάρκετινγκ [55, 87].

Τέλος, η αντιμετώπιση ηθικών προβληματισμών που σχετίζονται με την προστασία της ιδιωτικής ζωής των δεδομένων και την αλγοριθμική μεροληψία είναι απαραίτητη για τη διασφάλιση της υπεύθυνης χρήσης των τεχνολογιών πρόβλεψης. Η διασφάλιση ότι οι αλγόριθμοι είναι διαφανείς, δίκαιοι και σέβονται την ιδιωτικότητα των χρηστών θα συμβάλει στην οικοδόμηση εμπιστοσύνης και στη συμμόρφωση με τις κανονιστικές απαιτήσεις [88].

Αντιμετωπίζοντας αυτούς τους τομείς, η μελλοντική έρευνα μπορεί να συνεχίσει να εξελίσσει τον τομέα της πρόβλεψης της διαδικτυακής συμπεριφοράς, παρέχοντας ακόμη πιο ισχυρά εργαλεία για τους εμπόρους και τις επιχειρήσεις για την κατανόηση και τη συνεργασία με τους πελάτες τους. Αυτή η συνεχής εξέλιξη θα επιτρέψει τη

δημιουργία όλο και πιο εξελιγμένων και αποτελεσματικών στρατηγικών μάρκετινγκ, οδηγώντας στην ανάπτυξη και την καινοτομία στην ψηφιακή αγορά.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] ITU, "ITU Internet Reports. The Internet of Things," *Int. Telecommun. Union*, p. 212, 2005.
- [2] B. M. Leiner *et al.*, "A brief history of the internet," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 5, pp. 22–31, 2009, doi: 10.1145/1629607.1629613.
- [3] D. Bressoud, Th. & White, "The HyperText Transfer Protocol," 2020, doi: 10.1007/978-3-030-54371-6_20.
- [4] R. Rahim *et al.*, "Prototype file transfer protocol application for LAN and Wi-Fi

- communication,” *Int. J. Eng. Technol.*, vol. 7, no. 2.13 Special Issue 13, pp. 345–347, 2018, doi: 10.14419/ijet.v7i2.13.16917.
- [5] “Τι είναι το Internet - TO BHMA.” <https://www.tovima.gr/2008/11/24/culture/ti-einai-to-internet/> (accessed Jun. 27, 2024).
- [6] N. Hua, C. Morosan, and A. DeFranco, “The other side of technology adoption: Examining the relationships between e-commerce expenses and hotel performance,” *Int. J. Hosp. Manag.*, vol. 45, pp. 109–120, 2015, doi: 10.1016/j.ijhm.2014.12.001.
- [7] “Το διαδίκτυο ως εκπαιδευτικό εργαλείο | HuffPost Greece ΚΟΙΝΩΝΙΑ.” https://www.huffingtonpost.gr/entry/to-diadikteo-os-ekpaideetiko-eryaleio_gr_5cb5992ae4b05ed55a3558b5 (accessed Jun. 27, 2024).
- [8] M. De Mooij, “Consumer Behavior and Culture: Consequences for Global Marketing and Advertising: Consequences for Global Marketing and Advertising. Sage.” 2010.
- [9] G. Kotler, P., Armstrong, *Αρχές Μάρκετινγκ*. Αθήνα: Εκδόσεις Γκιούρδας, 2001.
- [10] D. D. Rucker, A. D. Galinsky, and D. Dubois, “Power and consumer behavior: How power shapes who and what consumers value,” *J. Consum. Psychol.*, vol. 22, no. 3, pp. 352–368, 2012, doi: 10.1016/j.jcps.2011.06.001.
- [11] Γ. Σιώμκος, *Συμπεριφορά Καταναλωτή και Στρατηγική Μάρκετινγκ*. Αθήνα: Εκδόσεις Σταμούλης, 2002.
- [12] M. Šostar and V. Ristanović, “Assessment of Influencing Factors on Consumer Behavior Using the AHP Model,” *Sustain.*, vol. 15, no. 13, 2023, doi: 10.3390/su151310341.
- [13] K. L. Kotler, P., & Keller, *Marketing Management (14th edition)*. Shanghai: Shanghai People’s Publishing House, 2016.
- [14] Μάντζαρη Γ., *Δυναμικό Μάρκετινγκ Αγαθών και Υπηρεσιών*. Αθήνα: Β. Γκιούρδας, 2003.
- [15] D. Fahy, J., Jobber, *Αρχές Μάρκετινγκ*. Αθήνα: Εκδόσεις Κριτική, 2014.
- [16] Μ. Γ. Πέτρου, *Εισαγωγή στο Μάρκετινγκ*. Αθήνα: εκδόσεις “Α. Σταμούλης”, 2001.
- [17] R. Angel Rosario, Roshini R., and Vidhya Pillai, “A Study on Digital Marketing and Its Types: A Deep Review of Pros and Cons,” *Integr. J. Res. Arts Humanit.*, vol. 2, no. 4, pp. 140–145, 2022, doi: 10.55544/ijrah.2.4.76.
- [18] J. Shin, J., Yu, *Targeted Advertising and Consumer Inference. Marketing Science*. 2021.
- [19] A. T. Rosário and J. C. Dias, “How has data-driven marketing evolved: Challenges and opportunities with emerging technologies,” *Int. J. Inf. Manag. Data Insights*, vol. 3, no. 2, 2023, doi: 10.1016/j.jjime.2023.100203.
- [20] A. Haleem, M. Javaid, M. Asim Qadri, R. Pratap Singh, and R. Suman, “Artificial intelligence (AI) applications for marketing: A literature-based study,” *Int. J. Intell. Networks*, vol. 3, no. July, pp. 119–132, 2022, doi: 10.1016/j.ijin.2022.08.005.

- [21] V. Kumar, A. R. Ashraf, and W. Nadeem, "AI-powered marketing: What, where, and how?," *Int. J. Inf. Manage.*, vol. 77, no. March, p. 102783, 2024, doi: 10.1016/j.ijinfomgt.2024.102783.
- [22] P. Russell, S., & Norvig, *Artificial Intelligence: A Modern Approach (3rd ed.)*. Pearson. 2016.
- [23] T. M. Mitchell, *Machine Learning*. McGraw Hill. 1997.
- [24] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 3, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.
- [25] R. X. Hosmer, D. W., Lemeshow, S., & Sturdivant, *Applied Logistic Regression*. Wiley, 2013.
- [26] J. R. Quinlan, *Induction of decision trees*. Machine Learning. 1986.
- [27] L. Breiman, *Random forests*. Machine Learning. 2001.
- [28] V. Cortes, C., & Vapnik, *Support-vector networks*. Machine Learning. 1995.
- [29] G. LeCun, Y., Bengio, Y., & Hinton, *Deep learning*. Nature. 2015.
- [30] Y. Dani and M. A. Ginting, "Classification of Predicting Customer Ad Clicks Using Logistic Regression and k-Nearest Neighbors," *Int. J. Informatics Vis.*, vol. 7, no. 1, pp. 98–104, 2023, doi: 10.30630/joiv.7.1.1017.
- [31] "Demystifying Logistic Regression: A Simple Guide | by WeiQin Chuah | Becoming Human: Artificial Intelligence Magazine."
<https://becominghuman.ai/demystifying-logistic-regression-a-simple-guide-8f762767a775> (accessed Jul. 09, 2024).
- [32] M. S., *Applied Logistic Regression Analysis*. Sage Publications, 2002.
- [33] K. Rahman, M., & Vasimalla, "Prediction of online shopper's purchasing intention using logistic regression.," Central University of Kerala, 2021.
- [34] E. Šoltés, J. Tábořecká-Petrovičová, and R. Šipoldová, "Targeting of online advertising using logistic regression," *E a M Ekon. a Manag.*, vol. 23, no. 4, pp. 197–214, 2020, doi: 10.15240/tul/001/2020-4-013.
- [35] M. A. Muda and R. A. Iswari, "Prediction of Online Shopper's Purchasing Intention Using Binary Logistic Regression, Decision Tree, and Random Forest," no. May, pp. 0–14, 2020, doi: 10.13140/RG.2.2.16567.55209.
- [36] İ. Topal, "Estimation of Online Purchasing Intention Using Decision Tree," *Yönetim ve Ekon. Araştırmaları Derg.*, vol. 17, no. 4, pp. 269–280, 2019, doi: 10.11611/yea.542249.
- [37] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.
- [38] A. Vieira, "Predicting online user behaviour using deep learning algorithms," pp. 1–21, 2015, [Online]. Available: <http://arxiv.org/abs/1511.06247>.
- [39] S. Zhao, "Customer Churn Prediction Based on the Decision Tree and Random

- Forest Model,” *BCP Bus. Manag.*, vol. 44, pp. 339–344, 2023, doi: 10.54691/bcpbm.v44i.4840.
- [40] S. Pallavi, K. R. Laxmi, N. Ramya, and R. Raja, *Study and analysis of modified mean shift method and kalman filter for moving object detection and tracking -> Chapter: Click-Through Rate Prediction Using Decision Tree*, vol. 1090. 2020.
 - [41] “A Guide to Bagging in Machine Learning: Ensemble Method to Reduce Variance and Improve Accuracy | DataCamp.” <https://www.datacamp.com/tutorial/what-bagging-in-machine-learning-a-guide-with-examples> (accessed Jul. 10, 2024).
 - [42] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, “Random forests and decision trees,” *IJCSI Int. J. Comput. Sci. Issues*, vol. 9, no. 5, pp. 272–278, 2012.
 - [43] “Architecture of the Random Forest algorithm. | Download Scientific Diagram.” https://www.researchgate.net/figure/Architecture-of-the-Random-Forest-algorithm_fig1_337407116 (accessed Jul. 09, 2024).
 - [44] R. Joshi, R. Gupte, and P. Saravanan, “A Random Forest Approach for Predicting Online Buying Behavior of Indian Customers,” *Theor. Econ. Lett.*, vol. 08, no. 03, pp. 448–475, 2018, doi: 10.4236/tel.2018.83032.
 - [45] W. Wei, “User Behavior Analysis of E-commerce Platforms Under Random Forest Recommendation Algorithm,” *Int. Conf. Distrib. Comput. Optim. Tech. ICDCOT 2024*, pp. 1–5, 2024, doi: 10.1109/ICDCOT61034.2024.10516196.
 - [46] M. Kamal and T. A. Bablu, “International Journal of Applied Machine Learning and Computational Intelligence Machine Learning Models for Predicting Click-through Rates on social media: Factors and Performance Analysis,” *Int. J. Appl. Mach. Learn. Comput. Intell.*, vol. 4, no. 11, pp. 1–14, 2022.
 - [47] W. Wang, W., Men, C., & Lu, *Online prediction model based on support vector machine*. Neurocomputing, 2008.
 - [48] A. Patle and D. S. Chouhan, “SVM kernel functions for classification,” *2013 Int. Conf. Adv. Technol. Eng. ICATE 2013*, 2013, doi: 10.1109/ICAdTE.2013.6524743.
 - [49] “Support Vector Machine (SVM) Algorithm.” <https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/> (accessed Jul. 09, 2024).
 - [50] G. Chaubey, P. R. Gavhane, D. Bisen, and S. K. Arjaria, “Customer purchasing behavior prediction using machine learning classification techniques,” *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 12, pp. 16133–16157, 2023, doi: 10.1007/s12652-022-03837-6.
 - [51] A. Srivastava et al., “Customer Behaviour Prediction for Online Shopping: a Review,” *Int. J. Eng. Appl. Sci. Technol.*, vol. 8, no. 2, pp. 261–269, 2023, doi: 10.33564/ijeast.2023.v08i02.039.
 - [52] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
 - [53] A. Goodfellow, I., Bengio, Y., & Courville, *Deep Learning*. MIT Press, 2016.
 - [54] J. Schmidhuber, “Deep Learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015, doi: 10.1016/j.neunet.2014.09.003.
 - [55] V. Hassija et al., “Interpreting Black-Box Models: A Review on Explainable

- Artificial Intelligence,” *Cognit. Comput.*, vol. 16, no. 1, pp. 45–74, 2024, doi: 10.1007/s12559-023-10179-8.
- [56] K. Sahaja and L. Pudukarapu, “A Deep Learning Approaches for online shopping Behavior prediction Using Clickstream Data,” vol. 6, no. 6, pp. 1–4, 2019.
 - [57] L. M. Badea (Stroie), “Predicting Consumer Behavior with Artificial Neural Networks,” *Procedia Econ. Financ.*, vol. 15, no. 14, pp. 238–246, 2014, doi: 10.1016/s2212-5671(14)00492-4.
 - [58] B. Zheng, K. Thompson, S. S. Lam, S. W. Yoon, and N. Gnanasambandam, “Customers’ behavior prediction using artificial neural network,” *IIE Annu. Conf. Expo 2013*, pp. 700–709, 2013.
 - [59] X. Hu, Y. Yang, S. Zhu, and L. Chen, “Research on a Hybrid Prediction Model for Purchase Behavior Based on Logistic Regression and Support Vector Machine,” *2020 3rd Int. Conf. Artif. Intell. Big Data, ICAIBD 2020*, pp. 200–204, 2020, doi: 10.1109/ICAIBD49809.2020.9137484.
 - [60] Gold Nmesoma Okorie *et al.*, “Leveraging Big Data for Personalized Marketing Campaigns: a Review,” *Int. J. Manag. Entrep. Res.*, vol. 6, no. 1, pp. 216–242, 2024, doi: 10.51594/ijmer.v6i1.778.
 - [61] P. A. Sunarya, U. Rahardja, S. C. Chen, Y. M. Li, and M. Hardini, “Deciphering Digital Social Dynamics: A Comparative Study of Logistic Regression and Random Forest in Predicting E-Commerce Customer Behavior,” *J. Appl. Data Sci.*, vol. 5, no. 1, pp. 100–113, 2024, doi: 10.47738/jads.v5i1.155.
 - [62] R. S. Monica Billio, Lorian Pelizzon, *Systemic Risk Tomograhpy*. 2017.
 - [63] B. Zheng and B. Liu, “A scalable purchase intention prediction system using extreme gradient boosting machines with browsing content entropy,” *2018 IEEE Int. Conf. Consum. Electron. ICCE 2018*, vol. 2018-Janua, pp. 1–4, 2018, doi: 10.1109/ICCE.2018.8326351.
 - [64] Z. Hu, *Consumer Behavior Prediction Based on Machine Learning Scenarios*. Atlantis Press International BV, 2023.
 - [65] R. Verma and S. Chandra, “RepuTE: A soft voting ensemble learning framework for reputation-based attack detection in fog-IoT milieu,” *Eng. Appl. Artif. Intell.*, vol. 118, no. June 2022, p. 105670, 2023, doi: 10.1016/j.engappai.2022.105670.
 - [66] T. Pitka *et al.*, “Time analysis of online consumer behavior by decision trees, GUHA association rules, and formal concept analysis,” *J. Mark. Anal.*, no. 0123456789, 2024, doi: 10.1057/s41270-023-00274-y.
 - [67] U. K. Lilhore, S. Simaiya, D. Prasad, and D. K. Verma, “Hybrid weighted random forests method for prediction & classification of online buying customers,” *J. Inf. Technol. Manag.*, vol. 13, no. 2, pp. 245–259, 2021, doi: 10.22059/jitm.2021.310062.2607.
 - [68] G. Sang and S. Wu, “Predicting the Intention of Online Shoppers’ Purchasing,” *Proc. - 2022 5th Int. Conf. Adv. Electron. Mater. Comput. Softw. Eng. AEMCSE 2022*, pp. 333–337, 2022, doi: 10.1109/AEMCSE55572.2022.00074.
 - [69] M. R. Kabir, F. Bin Ashraf, and R. Ajwad, “Analysis of different predicting model for online shoppers’ purchase intention from empirical data,” *2019 22nd Int.*

- Conf. Comput. Inf. Technol. ICCIT 2019*, pp. 1–6, 2019, doi: 10.1109/ICCIT48885.2019.9038521.
- [70] C. O. Sakar, S. O. Polat, M. Katircioglu, and Y. Kastro, “Real-time prediction of online shoppers’ purchasing intention using multilayer perceptron and LSTM recurrent neural networks,” *Neural Comput. Appl.*, vol. 31, no. 10, pp. 6893–6908, 2019, doi: 10.1007/s00521-018-3523-0.
 - [71] K. Maheswari and P. P. A. Priya, “Predicting customer behavior in online shopping using SVM classifier,” *Proc. 2017 IEEE Int. Conf. Intell. Tech. Control. Optim. Signal Process. INCOS 2017*, vol. 2018-Febru, pp. 1–5, 2017, doi: 10.1109/ITCOSP.2017.8303085.
 - [72] A. Toth, L. Tan, G. Di Fabbri, and A. Datta, “Predicting shopping behavior with mixture of RNNs,” *CEUR Workshop Proc.*, vol. 2311, 2017.
 - [73] D. Koehn, S. Lessmann, and M. Schaal, “Predicting online shopping behaviour from clickstream data using deep learning,” *Expert Syst. Appl.*, vol. 150, p. 113342, 2020, doi: 10.1016/j.eswa.2020.113342.
 - [74] Z. Wu and B. H. Tan, “Neural Modeling of Buying Behaviour for E-Commerce from Clicking Patterns,” doi: 10.1145/2813448.2813521.
 - [75] L. Mirghaderi, M. Sziron, and E. Hildt, “Ethics and Transparency Issues in Digital Platforms: An Overview,” *AI*, vol. 4, no. 4, pp. 831–843, 2023, doi: 10.3390/ai4040042.
 - [76] I. Altintas *et al.*, “Amazon Recommender System.”
 - [77] G. Linden, B. Smith, and J. York, “Amazon.com recommendations: Item-to-item collaborative filtering,” *IEEE Internet Comput.*, vol. 7, no. 1, pp. 76–80, 2003, doi: 10.1109/MIC.2003.1167344.
 - [78] L. Chen, G. Chen, and F. Wang, “Recommender systems based on user reviews: the state of the art,” *User Model. User-adapt. Interact.*, vol. 25, no. 2, pp. 99–154, 2015, doi: 10.1007/s11257-015-9155-5.
 - [79] G. Geetha, M. Safa, C. Fancy, and D. Saranya, “A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System,” *J. Phys. Conf. Ser.*, vol. 1000, no. 1, 2018, doi: 10.1088/1742-6596/1000/1/012101.
 - [80] A. Da’u and N. Salim, *Recommendation system based on deep learning methods: a systematic review and new directions*, vol. 53, no. 4. Springer Netherlands, 2020.
 - [81] D. Jannach and M. Zanker, *Value and Impact of Recommender Systems*. 2022.
 - [82] D. D. Sharma, D. D. Aggarwal, and D. A. B. Saxena, “Content Based Recommendation System on Netflix Data,” *Int. J. Res. Sci. Eng.*, no. 42, pp. 19–26, 2024, doi: 10.55529/ijrise.42.19.26.
 - [83] C. A. Gomez-Urbe and N. Hunt, “The netflix recommender system: Algorithms, business value, and innovation,” *ACM Trans. Manag. Inf. Syst.*, vol. 6, no. 4, 2015, doi: 10.1145/2843948.
 - [84] H. Steck, L. Baltrunas, E. Elahi, D. Liang, Y. Raimond, and J. Basilico, “Deep learning for recommender systems: A Netflix case study,” *AI Mag.*, vol. 42, no. 3,

- pp. 7–18, 2021, doi: 10.1609/aimag.v42i3.18140.
- [85] A. Torkashvand, S. M. Jameii, and A. Reza, *Deep learning-based collaborative filtering recommender systems: a comprehensive and systematic review*, vol. 35, no. 35. Springer London, 2023.
- [86] J. Bennett and S. Lanning, “The Netflix Prize,” *KDD Cup Work.*, pp. 3–6, 2007.
- [87] H. Huang *et al.*, “A cross-platform consumer behavior analysis of large-scale mobile shopping data,” *Web Conf. 2018 - Proc. World Wide Web Conf. WWW 2018*, no. April 2018, pp. 1785–1794, 2018, doi: 10.1145/3178876.3186169.
- [88] N. Balasubramaniam, M. Kauppinen, A. Rannisto, K. Hiekkänen, and S. Kujala, “Transparency and explainability of AI systems: From ethical guidelines to requirements,” *Inf. Softw. Technol.*, vol. 159, no. February, p. 107197, 2023, doi: 10.1016/j.infsof.2023.107197.