



ΣΤΡΑΤΙΩΤΙΚΗ ΣΧΟΛΗ ΕΥΕΛΠΙΔΩΝ
Τμήμα Στρατιωτικών Επιστημών

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΔΙΔΡΥΜΑΤΙΚΟ ΔΙΑΤΜΗΜΑΤΙΚΟ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΑΚΑΔΗΜΑΪΚΟΥ ΕΤΟΥΣ 2016-17
ΣΧΕΔΙΑΣΗ & ΕΠΕΞΕΡΓΑΣΙΑ
ΣΥΣΤΗΜΑΤΩΝ (SYSTEMSENGINEERING)

(ΠΔ 96 / 2015 / ΦΕΚ 163Α' / 20.08.2014)



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
Σχολή Μηχανικών Παραγωγής & Διοίκησης

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΧΕΙΡΑΓΩΓΗΣΗ ΔΕΔΟΜΕΝΩΝ ΜΕΓΑΛΗΣ
ΚΛΙΜΑΚΑΣ “BIG DATA” ΜΕ ΠΡΟΕΛΕΥΣΗ
ΤΑ ΜΕΣΑ ΚΟΙΝΩΝΙΚΗΣ ΔΙΚΤΥΩΣΗΣ
“SOCIAL MEDIA” ΜΕΣΩ ΤΗΣ ΜΕΘΟΔΟΥ
ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ ΚΕΙΜΕΝΟΥ

ΝΙΚΟΛΑΟΣ Γ. ΚΑΪΟΠΟΥΛΟΣ

A.M.: 2017018003

ΜΑΪΟΣ 2019

Η σελίδα αυτή είναι σκόπιμα λευκή

Η Μεταπτυχιακή Διατριβή του Καϊόπουλου Νικολάου εγκρίνεται:

ΤΡΙΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

Ειρήνη Σ. Καρανάσιου

(Επιβλέπουσα)

Αναπληρώτρια Καθηγήτρια Σ.Σ.Ε

Τομέα Μαθηματικών & Επιστημών
Μηχανικού.



Νικόλαος Ι. Δάρας

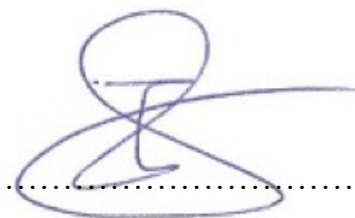
Κοσμήτορας ΣΣΕ

Καθηγητής Τομέα Μαθηματικών &
Επιστημών Μηχανικού.



Στέλιος Τσαφαράκης

Επίκουρος Καθηγητής Πολυτεχνείου
Κρήτης, Σχολής Μηχανικών Παραγωγής &
Διοίκησης.



Η σελίδα αυτή είναι σκόπιμα λευκή

Copyright © Απρ 2019 Νικόλαος Καϊόπουλος

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις της Στρατιωτικής Σχολής Ευελπίδων και του Πολυτεχνείου Κρήτης.

Η παρούσα εργασία αφιερώνεται:

στους γονείς μου Γιώργο και Αμαλία και στη μικρή μου αδελφή Ευτυχία.

ΕΥΧΑΡΙΣΤΙΕΣ

Καταρχάς θα ήθελα να ευχαριστήσω την επιβλέπουσα μου, Αναπληρώτρια Καθηγήτρια κ. Καρνανάσιου Ειρήνη, χωρίς τη βοήθεια της οποίας δεν θα είχε υλοποιηθεί η παρούσα εργασία, καθώς επίσης και για την εμπιστοσύνη που μου επέδειξε.

Θα ήθελα να εκφράσω τον σεβασμό και την εκτίμησή μου στον Κοσμήτορα της ΣΣΕ Καθηγητή κ. Δάρα Νικόλαο καθώς και στον Αντιστράτηγο κ. Λαλούση Χαράλαμπο και Υποστράτηγο κ. Χούπη Δημήτριο για τη βοήθεια που μου παρείχαν αυτά τα δύο χρόνια της προσπάθειάς μου τόσο σε υπηρεσιακό όσο και σε προσωπικό επίπεδο.

Πολλές ευχαριστίες οφείλονται στους γονείς μου, που με στήριζαν τόσο οικονομικά όσο και ψυχολογικά καθ’ όλη την ακαδημαϊκή μου πορεία και παρά τις αντιξοότητες που υπήρχαν.

Επίσης, θέλω να ευχαριστήσω τη θεία μου, καθηγήτρια Φιλολογίας κ. Παπαδοπούλου Ελένη, που παρά το επιβαρημένο πρόγραμμά της, με βοήθησε ιδιαίτερα στη διόρθωση όλου του κειμένου της παρούσας εργασίας, από φιλολογικής πλευράς.

Και, τέλος, οφείλω ένα μεγάλο ευχαριστώ στις κ. Αλβανού Ελένη, Παληού Αικατερίνη και την Υπολοχαγό (ΤΧ) Αλεξοπούλου Ευαγγελία που μου παρείχαν ιδιαίτερη βοήθεια στη διαδικασία συγγραφής και μορφοποίησης της Διπλωματικής μου Εργασίας.

Πίνακας περιεχομένων

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ	10
ΠΕΡΙΛΗΨΗ.....	11
ABSTRACT	12
ΕΙΣΑΓΩΓΗ.....	13
ΚΕΦΑΛΑΙΟ 1.....	15
ΧΕΙΡΑΓΩΓΗΣΗ ΚΑΙ SOCIAL MEDIA	15
1.1 ΕΙΣΑΓΩΓΗ.....	15
1.2 ΟΙ 10 ΤΡΟΠΟΙ ΧΕΙΡΑΓΩΓΗΣΗΣ	15
1.3 ΤΙ ΕΙΝΑΙ ΤΟ TWITTER;.....	19
1.4 Η ΠΟΛΙΤΙΚΗ ΚΑΙ TWITTER	21
1.5 ΕΠΙΛΟΓΟΣ 1 ^{ΟΥ} ΚΕΦΑΛΑΙΟΥ	24
ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ 1^{ΟΥ}.....	25
ΚΕΦΑΛΑΙΟ 2.....	26
ΟΙ ΨΕΥΔΕΙΣ ΕΙΔΗΣΕΙΣ.....	26
2.1 ΕΙΣΑΓΩΓΗ.....	26
2.2 ΟΙ ΠΗΓΕΣ ΤΗΣ ΑΝΑΛΥΣΗΣ	27
2.3 Η ΠΟΡΕΙΑ ΤΟΥ ΦΑΙΝΟΜΕΝΟΥ ΜΕΣΑ ΣΤΟ ΧΡΟΝΟ	27
2.4 Η ΠΕΡΙΠΤΩΣΗ ΤΩΝ FAKE NEWS ΣΤΙΣ ΑΜΕΡΙΚΑΝΙΚΕΣ ΠΡΟΕΔΡΙΚΕΣ ΕΚΛΟΓΕΣ	28
2.5 ΠΡΩΤΟΒΟΥΛΙΕΣ ΑΝΤΙΜΕΤΩΠΙΣΗΣ ΑΠΟ ΤΑ SOCIAL MEDIA	37
2.6 ΑΝΤΙΜΕΤΡΑ ΣΤΟ ΦΑΙΝΟΜΕΝΟ	38
2.7 ΕΠΙΛΟΓΟΣ 2 ^{ΟΥ} ΚΕΦΑΛΑΙΟΥ	41
ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ 2^{ΟΥ}.....	42
ΚΕΦΑΛΑΙΟ 3.....	44
BIG DATA ΚΑΙ NLP.....	44
3.1 ΕΙΣΑΓΩΓΗ.....	44
3.2 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ	45
3.3 BIG DATA ΚΑΙ SOCIAL MEDIA	46
3.4 TWITTER ΚΑΙ BIG DATA.....	49
3.5 NLP ΚΑΙ ΔΕΔΟΜΕΝΑ ΜΕΓΑΛΗΣ ΚΑΙΜΑΚΑΣ.....	50
3.6 ΕΠΙΛΟΓΟΣ 3 ^{ΟΥ} ΚΕΦΑΛΑΙΟΥ	54

ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ 3^{ΟΥ}.....	55
ΚΕΦΑΛΑΙΟ 4.....	57
ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΕΝΟΣ ΚΕΙΜΕΝΟΥ ΜΕ ΒΑΣΗ ΤΗ ΣΥΝΑΙΣΘΗΜΑΤΙΚΗ ΤΟΥ ΑΝΑΛΥΣΗ	57
4.1 ΕΙΣΑΓΩΓΗ.....	57
4.2 ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΚΕΙΜΕΝΟΥ	58
4.3 ΣΥΝΑΙΣΘΗΜΑΤΙΚΗ ΑΝΑΛΥΣΗ.....	61
4.4 ΑΛΓΟΡΙΘΜΟΣ NAIVE BAYES	62
4.5 TWITTER ΚΑΙ NAIVE BAYES	65
4.6 ΕΠΙΛΟΓΟΣ 4 ^{ΟΥ} ΚΕΦΑΛΑΙΟΥ	68
ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ 4^{ΟΥ}.....	70
ΣΥΜΠΕΡΑΣΜΑΤΑ.....	72

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Κεφάλαιο 1°

Εικόνα 1.1 Αναπαράσταση της χειραγώγησης [4].....15

Εικόνα 1.2 Τα δομικά στοιχεία του tweet [7].....17

Κεφάλαιο 2°

Εικόνα 2.1 Διαστρέβλωση της πραγματικότητας από τον επιλεκτικό τηλεοπτικό φακό [12]...28

Εικόνα 2.2 Ταξινόμηση των tweets που σχετίζονται με τη διακυβέρνηση του Προέδρου Donald Trump σχετικά με το περιεχόμενο τους (θετικό, αρνητικό, ουδέτερο) [15].....31

Εικόνα 2.3 Διάταξη των ιστοριών που άπτονται στη διακυβέρνηση του Trump σε σχέση με το εάν συμπεριλαμβάνουν κάποιο tweet του ίδιου του προέδρου ή μέλος της κυβέρνησής του [15].....32

Κεφάλαιο 3°

Εικόνα 3.1 Ο εννοιολογικός χάρτης των Κοινωνικών Μεγάλων Δεδομένων [3].....44

Κεφάλαιο 4°

Εικόνα 4.1 Κατηγοριοποίηση tweet με προσθήκη της παραμέτρου του συναισθήματος του γράφοντος [11].....58

Εικόνα 4.2 Διαδικασία ταξινόμησης Big Data μέσω αλγορίθμου εποπτευόμενης μηχανικής μάθησης [15].....59

Εικόνα 4.3 Αρχικό tweet.....66

Εικόνα 4.4 Στάδιο επεξεργασίας 1°.....67

Εικόνα 4.5 Στάδιο επεξεργασίας 2°.....67

Εικόνα 4.6 Στάδιο επεξεργασίας 3°.....67

Εικόνα 4.7 Στάδιο επεξεργασίας 4°.....68

Εικόνα 4.8 Στάδιο επεξεργασίας 5°.....68

ΠΕΡΙΛΗΨΗ

Στην παρούσα διπλωματική εργασία εξετάζεται η χειραγώγηση της κοινής γνώμης ως ένα καίριο πρόβλημα που αντιμετωπίζει η σημερινή κοινωνία. Ως χειραγώγηση θεωρείται οποιαδήποτε άσκηση επίμονης και σοβαρής πίεσης, κυρίως ψυχολογικής, επάνω σε έναν άνθρωπο με σκοπό να τον οδηγήσει, παρότι τη θέλησή του, σε μία πράξη ή μια παράλειψη που θα του είναι βλαπτική. Ένας από τους σημαντικότερους πρωταγωνιστές αυτής της διαίωξης του συγκεκριμένου προβλήματος είναι τα μέσα ενημέρωσης και, ιδιαιτέρως, τα μέσα κοινωνικής δικτύωσης. Πιο συγκεκριμένα, γίνεται παρουσίαση του Twitter και των τρόπων με τους οποίους η συγκεκριμένη πλατφόρμα δημιουργεί εύφορο έδαφος για να αναπτυχθεί το προαναφερθέν πρόβλημα, με τη γέννηση και διασπορά ψευδών ειδήσεων (fake news).

Επίσης, στην παρούσα διπλωματική εργασία παρουσιάζεται ο ρόλος των δεδομένων μεγάλης κλίμακας (Big Data) και οι τεχνικές με τις οποίες η άντλησή τους συναρτάται με τη χειραγώγηση της κοινής γνώμης επάνω σε ένα ζήτημα. Επιπλέον, επισημαίνεται η μεγάλη σημασία της κατηγοριοποίησης και ταξινόμησης δεδομένων μεγάλης κλίμακας με τη βοήθεια αλγορίθμων εποπτευόμενης μηχανικής μάθησης. Και, τέλος, δίνεται έμφαση στη διαδικασία της ανάλυσης, επεξεργασίας και τελικά ταξινόμησης ενός κειμένου φυσικής γλώσσας, όπως για παράδειγμα ένα tweet με βάση το συναισθηματικό φορτίο του γράφοντος.

Λέξεις Κλειδιά: Χειραγώγηση, Δεδομένα μεγάλης κλίμακας, ψευδείς ειδήσεις, Twitter, Εποπτευόμενη μηχανική μάθηση, Ανάλυση συναισθήματος.

ABSTRACT

The present dissertation focuses on the manipulation of public opinion as a key problem faced by today's society. The term manipulation encapsulates, any deceptive or indirect tactic including persistent and serious pressure, especially psychological, intended to influence and change the behavior or perception of others. One of the most important factors of the perpetuation of public opinion manipulating is the role of the media and, in particular, social media. More specifically, Twitter and the ways in which this platform creates a fertile ground to develop the aforementioned problem, with the birth and spreading of fake news are herein presented.

In this framework, the role of Big Data and the techniques that mediate the manipulation of public opinion on an issue are presented. In addition, the importance of classifying large-scale data (Big Data) with the help of supervised machine learning algorithms, is highlighted. Finally, the center of interest focuses on the process of analyzing, processing and classifying a natural language text, such as a tweet based on the writer's emotional load.

Key words: Manipulation, Big Data, Fake news, Twitter, Supervised machine learning, Sentiment analysis.

ΕΙΣΑΓΩΓΗ

Η ανάπτυξη της τεχνολογίας και ιδιαίτερα στον κλάδο των επικοινωνιών και του διαδικτύου έχει επιφέρει αλματώδεις αλλαγές στη ζωή μας. Οι συνέπειες αυτής της ανάπτυξης δεν είναι πάντα προς το συμφέρον της κοινής γνώμης και πολλές από αυτές στοχεύουν στην χειραγώγησή της. Στην παρούσα διπλωματική εργασία γίνεται αναφορά σε τομείς, όπως το Twitter και τα Social Media γενικά και πώς αυτά συμβάλλουν στη διαδικασία χειρισμού των ανθρώπων σε συγκεκριμένα σημαντικά ζητήματα, όπως η πολιτική σκηνή.

Στο πρώτο κεφάλαιο γίνεται αναφορά στον όρο χειραγώγηση, τι είναι και ποιες είναι οι 10 επιβλαβέστερες τεχνικές που χρησιμοποιούνται παγκοσμίως για να επιτευχθεί η διαδικασία της χειραγώγησης στους ανθρώπους. Επίσης, εμφανίζεται ένα από τα μέσα κοινωνικής δικτύωσης που είναι αρκετά πρόσφορο να υλοποιηθεί η παραπάνω διαδικασία, το Twitter.

Στο δεύτερο κεφάλαιο παρουσιάζονται οι ψευδείς ειδήσεις (fake news) ως μέρος του προβλήματος. Γίνεται ιστορική αναδρομή του όρου fake news και εξετάζονται οι τρόποι με τους οποίους κατά καιρούς έχει επηρεάσει την κοινή γνώμη. Παρατίθεται και σχολιάζεται ένα σημαντικό παράδειγμα πολιτικής χειραγώγησης της κοινής γνώμης που αποκαλύφθηκε μετά τις προεδρικές εκλογές στις ΗΠΑ το 2016. Και, τέλος, παρουσιάζονται αντίμετρα και πρωτοβουλίες διάφορων θεσμικών οργάνων για την καταπολέμηση της ψευδούς είδησης, τόσο σε παγκόσμιο επίπεδο όσο και στη χώρα μας.

Στη συνέχεια, στο κεφάλαιο τρία, αναλύεται ο όρος Big Data δηλαδή δεδομένα μεγάλης κλίμακας. Διερευνάται το περιεχόμενο και τα χαρακτηριστικά αυτών των δεδομένων που οριοθετούν τον χαρακτηρισμό Big και διαφοροποιούν τα απλά δεδομένα από εκείνα που είναι μεγάλης κλίμακας. Επιπλέον, εξετάζεται με ποιο τρόπο το Twitter προσφέρει σε οποιοδήποτε χρήστη ή μη της πλατφόρμας, πρόσβαση σε δεδομένα μεγάλης κλίμακας που σχετίζονται με τους χρήστες και τις αλληλεπιδράσεις τους, ενώ γίνεται λόγος και για την NLP (Natural Language Processing) τη διαδικασία επεξεργασίας κειμένων φυσικής γλώσσας, όπως για παράδειγμα τα tweets.

Στο τέταρτο και τελευταίο κεφάλαιο δηλώνεται ο τρόπος με τον οποίο κατηγοριοποιείται ένα κείμενο φυσικής γλώσσας λαμβάνοντας υπόψη τη συναισθηματική φόρτιση του γράφοντος κατά τη

συγγραφή του. Κατά επέκταση πώς τα tweets μπορούν να κατηγοριοποιηθούν με τη βοήθεια του αλγορίθμου Naive Bayes με βάση το συναίσθημα που εκφράζουν και περιγράφεται με μαθηματικούς όρους η διαδικασία λειτουργίας του αλγορίθμου. Τέλος, παρατίθεται παράδειγμα που αποσαφηνίζει την προεπεξεργασία που υφίσταται ένα tweet, προτού έρθει σε κατάλληλη μορφή για επεξεργασία και ταξινόμησή του από το Naive Bayes και με στόχο να ελαχιστοποιηθεί όσο το δυνατόν ο θόρυβος των δεδομένων που το απαρτίζουν.

ΚΕΦΑΛΑΙΟ 1

Χειραγώγηση και social media

1.1 Εισαγωγή

Στο κεφάλαιο αυτό θα γίνει παρουσίαση της χειραγώγησης που ο άνθρωπος δέχεται καθημερινά από διάφορα μέσα ενημέρωσης και μέσα κοινωνικής δικτύωσης και θα γίνει αναφορά στην επιρροή που ασκείτε και στις υφιστάμενες τεχνικές που καθιστούν τη χειραγώγηση αποτελεσματική [1]. Πριν εξεταστούν οι τεχνικές αυτές, σκόπιμο είναι να οροθετηθεί η έννοια της χειραγώγησης [2]. Ως Χειραγώγηση νοείται η ενέργεια του χειραγωγώ, η συνειδητή δηλαδή ειδική προσπάθεια κάποιου να επηρεάσει ή να διαχειριστεί έξυπνα ή ύπουλα το συνάνθρωπό του με επιδέξιο τρόπο.

1.2 Οι 10 τρόποι χειραγώγησης

Υπάρχουν ποικίλοι τρόποι χειραγώγησης από διάφορα μέσα ενημέρωσης και μέσα κοινωνικής δικτύωσης, ωστόσο δέκα τεχνικές αποτελούν τους πιο αποτελεσματικούς τρόπους της [3]:

1. Η τεχνική της διασκέδασης

Η τεχνική της διασκέδασης βασίζεται στον αποπροσανατολισμό του κοινού από τα ουσιαστικά προβλήματα και από τις αλλαγές που αποφασίστηκαν από συγκεκριμένες πολιτικές και οικονομικές ελίτ, με ένα συνεχές σφυροκόπημα ευφυολογημάτων και ανούσιων λεπτομερειών, για ασήμαντα θέματα. Επίσης, η τεχνική της διασκέδασης είναι ιδιαίτερα χρήσιμη για να απομακρυνθεί το κοινό από νευραλγικές πληροφορίες συγκεκριμένων επιστημονικών τομέων, όπως είναι η Οικονομία, η Ψυχολογία, η Νευροβιολογία και η Κυβερνητική. Σύμφωνα με τον Sylvain Timsit ο χειραγωγημένος

άνθρωπος αποστασιοποιείται από τα πραγματικά προβλήματα και ασχολείται με επουσιώδη ζητήματα, ώστε να μην στοχάζεται και να μην ασκεί κριτική και στην ουσία να αγελοποιείται [3].

2 . Η διαδικασία της πρόκλησης προβλημάτων και της εύρεσης λύσεων

Η συγκεκριμένη μεθοδολογία είναι γνωστή και ως «πρόβλημα αντίδραση- λύση». Με αυτή την τεχνική διαμορφώνεται ένα πλασματικό πρόβλημα για το οποίο έχει προβλεφθεί η ανάλογη αντίδραση των ατόμων ώστε, στη συνέχεια τα άτομα να αναζητήσουν τη λύση που έχει ήδη προδιαγραφεί. Μπορούν να παρουσιαστούν ως παραδείγματα αιματηρά επεισόδια για τα οποία τα άτομα οικειοθελώς απαιτούν τη λήψη μέτρων ασφαλείας με τον ταυτόχρονο περιορισμό των ελευθεριών τους ή μια χαλκευμένη οικονομική κρίση εξαιτίας της οποίας τα άτομα αποδέχονται αδιαμαρτύρητα την απώλεια των κοινωνικών τους δικαιωμάτων και τη διάλυση των δημοσίων υπηρεσιών.

3. Η τεχνική της υπονόμησης

Με γνώμονα το σκεπτικό της επιβολής ενός ανεπίτρεπτου μέτρου, ακολουθείται η διαδικασία της σταδιακής εφαρμογής του κατά φθίνουσα κλίμακα για μια διάρκεια ολίγων ετών. Έτσι στις δεκαετίες του 1980 και 1990 ριζοσπαστικοποιήθηκαν καινοφανείς κοινωνικές και οικονομικές παράμετροι του νεοφιλελευθερισμού. Για παράδειγμα υπερβολικές αλλαγές όπως η υψηλή ανεργία το αβέβαιο επαγγελματικό περιβάλλον, οι ελαστικές εργασιακές σχέσεις, οι δυσμενείς μεταθέσεις και οι πενιχροί μισθοί είχαν ως αποτέλεσμα επαναστατικές κινήσεις αν το χαρακτηριστικό τους ήταν η απρόσμενη και βίαιη επιβολή τους.

4. Η Πρακτική της αναβολής

Μια άλλη πρακτική, για να γίνει ευμενώς αποδεκτό ένα μέτρο με αντιλαϊκό περιεχόμενο, είναι να προβληθεί ως επώδυνο αλλά αδήριτο και με αυτό τον τρόπο να έχει τη σύμφωνη γνώμη των ατόμων στην τρέχουσα πραγματικότητα ώστε να εφαρμοστεί στο άμεσο μέλλον. Σύμφωνα και με την ψυχολογία ο άνθρωπος δέχεται πιο εύκολα μια θυσία σε βάθος χρόνου παρά μια βραχυπρόθεσμη. Και αυτό συμβαίνει επειδή το άτομο έχει μια αισιόδοξη θέαση για το μέλλον, καθώς ελπίζει ότι όλα θα βελτιωθούν και ότι τελικά μπορεί να αποφύγει την επαχθή θυσία. Επιπλέον το άτομο έχει στη

διάθεσή του μια χρονική προθεσμία μέσα στην οποία καλείται να συμφιλιωθεί με την ιδέα της αλλαγής και να την αποδεχθεί παθητικά όταν έρθει η εκπνοή του χρόνου για την εφαρμογή της.

5 . Η αντιμετώπιση των ενηλίκων σαν να είναι νήπια

Όπως είναι γνωστό, το σύνολο των διαφημίσεων που απευθύνονται στο αγοραστικό κοινό μετέρχονται ένα αρκετά ισχυρό λεκτικό οπλοστάσιο που περιλαμβάνει αφηγηματικές τεχνικές, εντυπωσιακά επιχειρήματα, εξαιρετικά απλοϊκό ύφος και επιτηδευμένο περιεχόμενο. Χαρακτηριστικά γνωρίσματα ενός λόγου που στοχεύει σε δέκτες με μικρή ή ελάχιστη κριτική ικανότητα που συγχρόνως είναι και συναισθηματικά ευάλωτοι. Κύριος στόχος τους είναι η παραπλάνηση του δέκτη και η χειραγωγή του ώστε να κατευθυνθεί άκριτα και χωρίς ιδιαίτερα στοχαστική διάθεση στον επιδιωκόμενο σκοπό, την αγορά του προβαλλόμενου προϊόντος.

6. Η επίκληση στο συναίσθημα ως μορφή πειθούς

Η επίκληση στο συναίσθημα είναι ένας διαδεδομένος τρόπος πειθούς για να καμφθεί ο ορθολογισμός και η κριτική σκέψη των ατόμων. Μάλιστα η διέγερση των συναισθημάτων του δέκτη προλαβαίνει το έδαφος για να εμφυτευθούν στη σφαίρα του ασυνείδητου φόβοι, επιθυμίες, ιδέες, παρορμήσεις ή στάσεις και γενικά να ολοκληρωθεί αποτελεσματικά η διαδικασία της προπαγάνδας.

7. Το μέσο για τη δημιουργία ατόμων με εξαρτημένα αντανακλαστικά

Στη συγκεκριμένη διαδικασία το άτομο αδυνατεί να κατανοήσει τις τεχνικές και τις μεθόδους της διενεργούμενης πνευματικής του χειραγωγής. Αυτό επιτυγχάνεται συστηματικά με την παροχή χαμηλής ποιότητας εκπαίδευσης στα κατώτερα κοινωνικά στρώματα με αποτέλεσμα να διευρύνεται το χάσμα ανάμεσα στην κοινωνική ελίτ και στις κατώτερες τάξεις και να παγιώνεται η άγνοια και η πνευματική νωθρότητα των ευάλωτων κοινωνικών τάξεων.

8. Η πρακτική της συμφιλίωσης του κοινού με το μέτριο

Κατά τη διαδικασία αυτή εθίζονται τα άτομα να προκρίνουν και να επιβραβεύουν το κατώτερο, το πιο εύπεπτο και οτιδήποτε δεν απαιτεί πνευματική εγρήγορση και προβληματισμό. Για

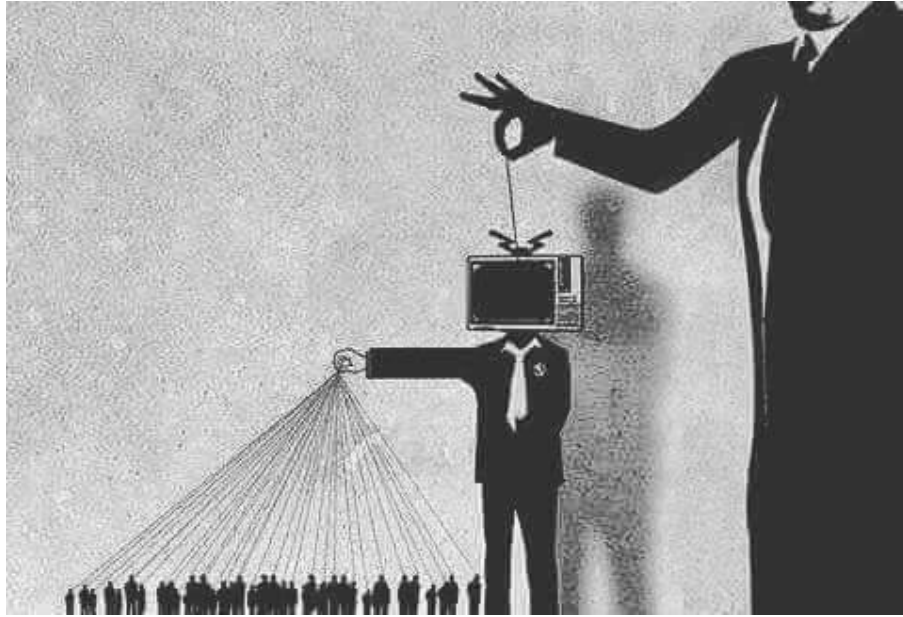
παράδειγμα ο καταγιγισμός των τηλεθεατών κυρίως από τα ιδιωτικά τηλεοπτικά κανάλια, με πρωινές εκπομπές τύπου Life Style.

9. Η πρακτική της δημιουργίας ενοχικών και παθητικών ατόμων

Σε αυτή τη διαδικασία το άτομο έχει την πεποίθηση ότι ευθύνεται το ίδιο για όλα τα προβλήματα που αντιμετωπίζει και εστιάζει στην πηγή των αιτίων, που είναι η ανικανότητα του και η αδυναμία του να βρίσκει αποτελεσματικές λύσεις για τα προβλήματα που τον ταλανίζουν. Με αυτό τον τρόπο αρνείται να αντιδράσει, λειτουργεί σαν αποδιοπομπαίος τράγος που συγκεντρώνει επάνω του όλες τις αιτίες της κακοδαιμονίας του και ως εκ τούτου απομακρύνεται από την ανάληψη πρωτοβουλιών και την οργάνωση δράσης.

10. Το σύστημα γνωρίζει καλύτερα τον άνθρωπο απ’ ότι ο ίδιος τον εαυτό του

Η αλματώδης εξέλιξη της επιστήμης διαμόρφωσε ένα σύστημα που εδράζεται στις σύγχρονες και καινοτόμες γνώσεις της Βιολογίας της Νευροβιολογίας και της εφαρμοσμένης Ψυχολογίας. Οι γνώσεις αυτές διεύρυναν την απόσταση ανάμεσα στο γνωστικό υπόβαθρο των απλών ανθρώπων και στις γνώσεις των ειδημόνων που τις χρησιμοποιούν και τις εφαρμόζουν. Έτσι το σύστημα οδηγήθηκε σε μια εις βάθος γνώση του ανθρώπινου παράγοντα και σε μία μοναδική άσκηση και επιβολή ελέγχου πάνω στα άτομα, χωρίς τα τελευταία να έχουν πλέον τη δυνατότητα του αυτοπροσδιορισμού τους.



Εικόνα 1.1 Αναπαράσταση της Χειραγώγησης [4].

Και οι δέκα τεχνικές που παρουσιάστηκαν συνοπτικά παραπάνω οδηγούν ντετερμινιστικά στο σχεδιαζόμενο αποτέλεσμα και τα πιο δημοφιλή μέσα που χρησιμοποιούνται για αυτό είναι το Facebook και το Twitter, τα μέσα κοινωνικής δικτύωσης που θα αποτελέσουν ιδιαίτερο αντικείμενο μελέτης της εν λόγω διατριβής.

1.3 Τι είναι το Twitter;

Τα τελευταία χρόνια, κοινωνικά δίκτυα όπως το Twitter [5], γίνονται όλο και πιο δημοφιλή. Καθώς το Twitter είναι η πιο πολυσύχναστη ιστοσελίδα για μικροϊστολόγηση (microblogging) με πάνω από 500 εκατομμύρια χρήστες (εκ των οποίων οι 332 εκατομμύρια είναι ενεργοί σύμφωνα με έγκυρα στατιστικά από τον Μάιο του 2018) και 340 εκατομμύρια tweets την μέρα, αποτελεί μία βασική πηγή πληροφορίας, γεγονός που επιβεβαιώνεται και από τα 1,6 δις ερωτήματα αναζήτησης ημερησίως. Τα μηνύματα, ή στη γλώσσα του Twitter τα “tweets”, είναι ένας τρόπος για να μοιράζονται οι άνθρωποι τα ενδιαφέροντα τους δημόσια ή στο πλαίσιο μίας συγκεκριμένης ομάδας.

Μάλιστα το Twitter διαφοροποιείται από τα υπόλοιπα κοινωνικά δίκτυα καθώς σε αυτό περιορίζεται αρκετά το μέγεθος του μηνύματος. Συγκεκριμένα το μέγιστο μέγεθος των 140 χαρακτήρων υποχρεώνει τους χρήστες να διατυπώνουν τις απόψεις τους σε μία ή δύο το πολύ περιόδους.

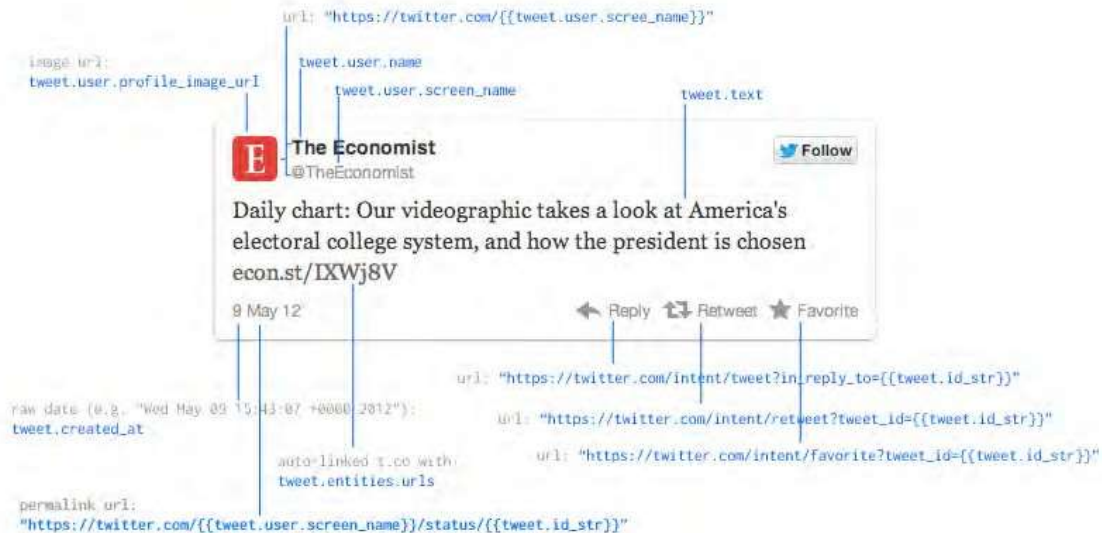
Καθώς το Twitter είναι το πιο προσφιλές σήμερα μέσο κοινωνικής δικτύωσης, συγχρόνως αποτελεί και έναν καθρέφτη των παγκόσμιων γεγονότων. Οι περισσότερες εταιρίες εκδηλώνουν ενδιαφέρον για τις πιο πρόσφατες παγκόσμιες τάσεις που διαμορφώνονται από τα γεγονότα που λαμβάνουν χώρα μέσα σε αυτό. Οι τάσεις αυτές υποβάλλονται σε δυναμική ανάλυση ώστε να ακολουθήσει η ανάλογη εμπορική ανταπόκριση. Όσον αφορά στο Marketing, εταιρίες και οργανισμοί αναλύουν τα tweets, σφυγμομετρώντας τις συνακόλουθες τάσεις και προβαίνουν σε στοχευμένες ενέργειες όπως είναι το λανσάρισμα προϊόντων και υπηρεσιών, με αποτέλεσμα την αύξηση των εσόδων τους και τη διαμόρφωση ενός υγιούς ανταγωνιστικού κλίματος.

Τα δομικά στοιχεία του tweet

Το tweet [6] είναι ο πυλώνας αυτού του μέσου κοινωνικής δικτύωσης καθώς είναι ο αποκλειστικός τρόπος μεταβίβασης και δημοσιοποίησης μιας πληροφορίας ανάμεσα στους χρήστες, με εξαίρεση τα προσωπικά μηνύματα. Το Tweet έχει ιδιαίτερα δομικά στοιχεία εκτός από το περιορισμένο μέγεθός του. Κατ' αρχάς εμπεριέχει σε μεγάλο ποσοστό τα σύμβολα “#” και “@”. Το σύμβολο # καλείται « hashtag » και με την χρήση του γίνεται παρουσιάζεται κάποιο θέμα (topic). Με αυτόν τον τρόπο παρέχεται η δυνατότητα κατηγοριοποίησης πολλών ομόθεμων tweet και η αναζήτηση πληροφοριών με την χρήση π.χ. #iphone10 που θα παρέχει όλα τα ειδικά tweets για καινούριο iPhone. Η αναφορά σε κάποιον άλλο χρήστη γίνεται [6] με τη εισαγωγή του @ (π.χ. @Nikos_kai).

Επίσης ένα άλλο συχνό σύμβολο που παρουσιάζεται στην προμετωπίδα των tweets, είναι το “RT”, το οποίο σημαίνει « retweet », δηλαδή υλοποιεί την ανακοίνωση ενός tweet κάποιου χρήστη από κάποιον άλλο. Τέλος, συχνή είναι και η χρήση ιστοτόπου στα tweets, εφόσον πέραν της προσθήκης αναφοράς σε κάτι προσφέρει και μία λύση στην περιορισμένη δυνατότητα ανάλυσης ενός

θέματος στον περιορισμένο χώρο. Παρακάτω είναι εμφανής η δομή ενός δείγματος tweet του γνωστού περιοδικού "The Economist":



Εικόνα 1.2 Τα δομικά στοιχεία του tweet [7].

1.4 Η Πολιτική και Twitter

Όπως γίνεται κατανοητό το Twitter μπορεί να αξιοποιηθεί σε πολιτικές και διαφημιστικές εκστρατείες και να είναι εξίσου αποτελεσματικό με το Facebook χάρη στην ταχύτητα της διάδοσης τους. Ωστόσο ενώ το Facebook παρέχει στις διάφορες καμπάνιες τα κατάλληλα μέσα για διεξαγωγή μακροσκελούς διαλόγου, γεγονός που διευκολύνει την καλύτερη κατανόηση των θεμάτων, το Twitter με τον περιορισμό του μηνύματος στους 140 χαρακτήρες προσφέρει μία μονόπλευρη και περιορισμένης θεματικής εμβέλειας διάλογο. Ακόμη το Twitter είναι ένα δίκτυο που απαιτεί την εγγραφή των επίδοξων ακόλουθων ώστε να συμμετέχουν και να αλληλοεπιδρούν χωρίς όμως και να

τους παρέχεται η δυνατότητα να εκφράζουν τη συμφωνία ή άλλως τα Likes όπως συμβαίνει στο Facebook [8].

Αυτό σημαίνει πως στο Twitter η απόκτηση ακολούθων είναι ιδιαίτερα σημαντική, εφόσον οι ακολούθοι μπορούν να δούνε τις αναρτήσεις μόνο όταν εγγραφούν στην εκάστοτε σελίδα. Όπως και στις σελίδες του Facebook έτσι και στο Twitter υπάρχει ένα όριο στον αριθμό των ακολούθων. Το όριο αυτό ανέρχεται στα 2000 άτομα ωστόσο εάν κάποιος ακολουθεί μεγάλο αριθμό ατόμων στην πράξη δεν μπορεί να παρακολουθήσει τον καταιγισμό των ενημερώσεων από τις σελίδες στις οποίες είναι εγγεγραμμένος. Ειδικότερα οι υποψήφιοι σε εκλογικές αναμετρήσεις έχουν τη δυνατότητα οι ίδιοι να επιλέξουν έναν αριθμό ακολούθων ώστε να είναι συνεχώς ενημερωμένοι για τις προθέσεις τους και τις επιθυμίες τους τις σχετικές με την ομάδα και με την εκστρατεία τους. Επομένως το Twitter είναι ιδιαίτερα ωφέλιμο για την επιτυχή έκβαση της εκστρατείας ενός υποψηφίου εφόσον μόνο με 140 χαρακτήρες βρίσκεται σε αδιάλειπτη επικοινωνία με τους ψηφοφόρους του [9].

Το Twitter [7] έχει μια υπερβολικά μεγάλη ακολουθία συνδέσεων καθώς αριθμούνται πάνω από 5 εκατομμύρια «tweets» την ημέρα. Έτσι μια εκστρατεία [10] μπορεί να εξατομικεύσει τη δική της σελίδα, να επικοινωνεί συστηματικά και μονόπλευρα με τους ακολούθους της, να προωθεί διαφημιστικά μηνύματα και γενικά να κρατάει μία συνεχή και δυναμική παρουσία στον παγκόσμιο ιστό. Μία άλλη δυνατότητα της σελίδας είναι η βελτίωση της διαφάνειας της πολιτικής διαφήμισης. Αυτό επιτυγχάνεται χάρη στη γνωστοποίηση των κεφαλαίων τα οποία επενδύει ο υποψήφιος προκειμένου να προσεγγίσει τους ψηφοφόρους του. Ουσιαστικά η παρουσίαση της προέλευσης των κεφαλαίων για τη διενέργεια των διαφημίσεων από χορηγούς ή τον ίδιο τον υποψήφιο αφενός έχει κέρδος από τα προωθημένα tweets και αφετέρου καλλιεργείται η ηλεκτρονική Δημοκρατία αφού ο υποψήφιος λογοδοτεί για τη χρηματοδότηση των πρόσθετων διαφημίσεων που προβάλλονται μέσα από τη σελίδα του.

Επιπρόσθετα στο Twitter καταγράφονται τα σχόλια των ψηφοφόρων για τους υποψηφίους και ενθαρρύνεται η μεταξύ τους επικοινωνία έτσι συνεχώς διαπιστώνεται η αποτελεσματικότητα του υποψηφίου και οι προτιμήσεις των ακολούθων σχετικά με την πρόγνωση της ψήφου των. Αυτή η αλληλεπίδραση υποψηφίων και ψηφοφόρων αποτελεί στην πράξη ένα παράδειγμα της δυνατότητας

που παρέχουν τα κοινωνικά δίκτυα για την καλλιέργεια της αμεσότητας της ανθρώπινης σχέσης. Ωστόσο δεν θα πρέπει να λησμονείται και η αρνητική πτυχή του διαδικτύου γενικότερα, η πολιτική χειραγωγήση. Διάφορες πολιτικές οργανώσεις αλλά και παρακρατικοί παράγοντες μονοπωλούν τη χρήση του Ίντερνετ και συγκεκριμένα των μέσων κοινωνικής δικτύωσης για να διαδώσουν ψεύτικες, χαλκευμένες ειδήσεις. Αυτή η σκόπιμη διάδοση κατασκευασμένων ειδήσεων προκαλεί το πρόβλημα της χειραγωγής των ψηφοφόρων και οδηγεί στην αλλοίωση του αποτελέσματος των εκλογών.

Από την άλλη πλευρά, καθίσταται ένα ιδιαίτερο μέσο χειραγωγής. Πολύ συχνά εμφανίζονται ψεύτικα προφίλ που δεν αντιστοιχούν σε υπαρκτά πρόσωπα ή φυσικές υποστάσεις. Με άλλα λόγια πίσω από κάποιους λογαριασμούς υποκρύπτεται ένα bot ή κάποιος άνθρωπος που επιδιώκει την παραπληροφόρησή μας. Σύμφωνα με έρευνα [10] υπάρχουν αρκετά προφίλ τα οποία δημιουργήθηκαν για να εξαπατούν. Η απάτη στα μέσα κοινωνικής δικτύωσης ονομάζεται SMF (Social Media Fraud) και εδράζεται στην δημιουργία τεχνητών Likes και Follows και αποσκοπεί στην αύξηση της ψηφιακής επιρροής ενός λογαριασμού κοινωνικής δικτύωσης. Έτσι παρουσιάζεται να συγκεντρώνει με τεχνητό τρόπο υπερβολικά μεγάλη αποδοχή μελών και οτιδήποτε εκφράζεται σε αυτό φαίνεται σωστό και ακριβώς αποδεκτό απλώς και μόνο επειδή τυγχάνει μεγάλης αποδοχής και εξαιτίας της μαζοποίησης.

Σε έρευνα [10] που διεξήχθη σε 48 χώρες αποδείχθηκε ότι στις 31 από αυτές υπήρχαν τεκμήρια πως οι πολιτικές παρατάξεις χρησιμοποίησαν την διαδικτυακή προπαγάνδα κατά τη διάρκεια εκλογών καθώς και δημοψηφισμάτων. Σε αναπτυσσόμενες δυτικές δημοκρατίες εφαρμόζονται εξελιγμένες αναλύσεις big data καθώς και πολιτικά bots για να διαστρεβλώσουν και να παραχαράξουν την αλήθεια να προωθήσουν το σκεπτικισμό και τη δυσπιστία, να φανατίσουν κομματικά τις εκλογικές περιφέρειες και να φαλκιδεύουν τις δημοκρατικές διαδικασίες.

Αξίζει να αναφερθεί ότι στα αυταρχικά καθεστώτα οι κυβερνώντες εφαρμόζουν πρακτικές χειραγωγής με σκοπό την πλήρη ανατροπή της εκλογικής αναμέτρησης. Σε αυτή την περίπτωση διαμορφώνεται η δύναμη μίας κυβερνητικής ελίτ η οποία στηρίζεται στο χειρισμό των κοινωνικών μέσων δικτύωσης, στον έλεγχο των μέσων ενημέρωσης και στον αστυνομοκρατικό εκφοβισμό. Βέβαια, οι κυβερνήσεις σε όλο τον κόσμο σήμερα εκφράζουν έντονες ανησυχίες για τη διογκούμενη

απειλή της παραπλανητικής διάδοσης ειδήσεων είτε αυτές είναι πραγματικές είτε αυτές είναι κατασκευασμένες. Εξάλλου από το 2016 σε 30 και πλέον χώρες εφαρμόζεται νομοθεσία που ορίζει την καταπολέμηση των ψεύτικων ειδήσεων στο Διαδίκτυο [11].

1.5 Επίλογος 1^ο κεφαλαίου

Συμπερασματικά λοιπόν, διαφαίνεται πως, άνθρωποι του πολιτικού κόσμου έχουν τη δυνατότητα να κάνουν χρήση μιας ποικιλίας τεχνικών χειραγώγησης. Κάθε πολιτική καμπάνια χρησιμοποιεί ένα διαφορετικό σύνολο εργαλείων για αυτή τη δουλειά. Οι περισσότεροι χρησιμοποιούν online σχολιαστές και ψεύτικους λογαριασμούς κοινωνικών μέσων ενημέρωσης και δικτύωσης για τη διάδοση φιλοκυβερνητικών ή φιλοπολιτικών μηνυμάτων σε πληθυσμούς τόσο στο εσωτερικό όσο και στο εξωτερικό. Τα πολιτικά αυτά bots πλημμυρίζουν hashtags με αυτοματοποιημένα μηνύματα που προωθούν η επιτίθενται σε συγκεκριμένους πολιτικούς. Σίγουρα μπορούμε να πούμε πως η χειραγώγηση των δεδομένων των μέσων κοινωνικής δικτύωσης δημιουργεί τη χειραγώγηση των μαζών η οποία με τη σειρά της μπορεί να είναι επιζήμια για το διαδικτυακό οικοσύστημα.

ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ 1^{ου}

[1] Alice Marwick, Rebecca Lewis (2016), Media manipulation and disinformation online, Data and Society Research Institute.

[2] Χειραγώγηση (2019), Βικιλεξικό, Η ελεύθερη Εγκυκλοπαίδεια,
<https://el.wiktionary.org/w/index.php?>.

[3] Sylvain Timsit (2002), Strategies de manipulation,
<https://www.syti.net/GB/SilentWeaponsGB.html>.

[4] <https://www.facebook.com/onlythinkingart/photos/a.1381534705232967/1811970285522738/?type=3&theater>

[5] “Twitter.com Traffic, Demographics and competitors” (2019), www.alexa.com.

[6] Issie Lapowsky (2013), Ev Williams on Twitter’s Early Years, Inc,
<http://www.inc.com/issie-lapowsky/ev-williams-twitter-early-years.html>.

[7] Κουναδέας Θεόδωρος (2017), Τα Νέα Μέσα ως εργαλεία επικοινωνίας των πολιτικών κομμάτων στην Ελλάδα, Διπλωματική Εργασία Μεταπτυχιακού Διπλώματος Ειδίκευσης.

[8] “Following Rules and Best Practices.” Twitter. Web.

[9] Dugan, L. (2012), Facebook, Twitter, Google+, Pinterest: The Users of Social Media”,
http://www.mediabistro.com/alltwitter/social-media-users_b22556.

[10] Masarah Paquet-Clouston, Olivier Bilodeau, David Decary-Hetu (2017), Can we Trust Social Media? Social Network Manipulation by an IoT Botnet, SM Society, Toronto, Canada.

[11] Philip N. Howard (2018), Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation, University of Oxford, Oxford

ΚΕΦΑΛΑΙΟ 2

Οι Ψευδείς Ειδήσεις

2.1 Εισαγωγή

Η λέξη fake news δηλαδή οι ψεύτικες και εντυπωσιακές πληροφορίες που διαδίδονται προσχηματικά σαν είδηση είναι η λέξη που, σύμφωνα με το λεξικό Collins [1] επιλέχθηκε ως λέξη της χρονιάς του 2017 αφού η χρήση της αυξήθηκε κατά 365% σε σύγκριση με το 2016.

Στο κεφάλαιο αυτό θα καταδειχθεί η ιδιάζουσα σημασία που έχουν αποκτήσει σήμερα τα fake news με τη χρήση ενός γεγονότος από τα κορυφαία πολιτικά γεγονότα, αυτό των προεδρικών εκλογών στις ΗΠΑ, που διαδραματίστηκε σε μία περίοδο κατά την οποία σημειώθηκε μία διαρκώς αυξανόμενη χρήση των μέσων κοινωνικής δικτύωσης από τους πολίτες, είτε ατομικά είτε συλλογικά από οργανωμένες ομάδες.

Εκτός από την αναφορά στα πολιτικά γεγονότα και στο ρόλο των fake news, θα επιχειρηθεί και μία εστίαση στις ατομικές και ομαδικές προσπάθειες που έχουν γίνει μέχρι σήμερα ως αντίδραση στο φαινόμενο των fake news και για την προστασία από την επιρροή τους. Το συνεχές και έντονα εξελισσόμενο φαινόμενο έχει οδηγήσει σε πρακτικές για την εξεύρεση λύσεων του προβλήματος τόσο από άτομα και ομάδες επιστημόνων όσο και από τα εθνικά κοινοβούλια και τα θεσμικά όργανα, όπως είναι η Ευρωπαϊκή Επιτροπή. Χαρακτηριστική είναι η περίπτωση ενός προπτυχιακού φοιτητή στο τμήμα πληροφορικής του ΠΑ.ΜΑΚ., του Βαλεντίνου Τζέικα [2], ο οποίος πέτυχε να δημιουργήσει έναν αλγόριθμο που συμβάλλει στον εντοπισμό μιας παραπλανητικής είδησης και στο διαχωρισμό της από μία πραγματική σε όλη την έκταση του Διαδικτύου άμεσα και γρήγορα.

2.2 Οι πηγές της ανάλυσης

Η παρούσα ανάλυση εμπειριστατωμένα παρουσιάζει πως τα fake news εντός των social media ασκούν επιρροή στους αναγνώστες τους εξαιτίας της ιλιγγιώδους μετάδοσης των πληροφοριών και των πολιτικοοικονομικών εξελίξεων διεθνώς. Είναι αριστό ένα tweet ή ένα post, μέσω του οποίου μπορεί να διαστρεβλωθεί εντελώς η άποψη μερίδας ανθρώπων σχετικά με κάποιο ζήτημα, μέσω των δημοφιλών μέσων κοινωνικής δικτύωσης, οι περισσότεροι από τους οποίους ασπάζονται χωρίς κριτική σκέψη μια άποψη και χωρίς να επιδιώκουν να εξετάσουν την αξιοπιστία της πληροφορίας. Η ανάλυση στηρίχθηκε σε δεδομένα και πληροφορίες που είχαν ως πηγή άντλησης: δημοσιευμένα άρθρα στο διαδίκτυο, δημοσιευμένα άρθρα σε διάφορες εφημερίδες και επιστημονικά περιοδικά διεθνούς ενδιαφέροντος και δημοσιεύματα διεθνών ειδησεογραφικών κολοσσών στον τομέα της ενημέρωσης.

2.3 Η πορεία του φαινομένου μέσα στο χρόνο

Ο Tom Holland [3] σε άρθρο του στην Wall Street Journal στις 30 Μαρτίου 2017 και με ευκαιρία της επετείου της γεννήσεως, του αδιαμφισβήτητου πατέρα της Ιστορίας Ηρόδοτου πριν 2500 χιλιάδες χρόνια δήλωσε εμφατικά πως η επέτειος αυτή συμπίπτει με την επέτειο των 2500 χρόνων από τη δημιουργία των ψευδών ειδήσεων. Ο αρθρογράφος και μεταφραστής του Ηρόδοτου εκθειάζει με αυτό τον τρόπο τον πατέρα της ιστορίας αλλά και τον λάτρη των παραμυθικών αφηγήσεων. Στη σημερινή εποχή της διάδοσης της πληροφορίας σε κλάσματα του δευτερολέπτου ο κύριος σκοπός του πομπού εξακολουθεί να είναι η επιρροή και η κατεύθυνση της κοινής γνώμης. Στην ουσία αυτό που έχει αλλάξει είναι η ταχύτητα διάδοσης της πληροφορίας η αδιάλειπτα αυξανόμενη συμμετοχή των χρηστών στα μέσα κοινωνικής δικτύωσης και η άκριτη υιοθέτηση των ειδήσεων και πρακτικών που τους δίνονται. Η καθηγήτρια φιλοσοφίας στο ΕΚΠΑ Βάσω Κιντή σε άρθρο [4] της στην εφημερίδα το ΒΗΜΑ προσυπογράφει πως: «Το πρόθεμα μετα- (post)- πριν από τις λέξεις «αλήθεια» και «γεγονότα», σημαίνει δύο πράγματα: πρώτον, κάτι που έρχεται χρονικά μετά την αλήθεια ή το

γεγονός και δεύτερον, κάτι που το ακολουθεί και συγχρόνως τα υπερβαίνει, τα καταργεί, τα προσπερνά ως ασήμαντα. Όταν λέμε, λοιπόν, «post – truth», ή «post - fact» politics, δηλαδή πολιτική της μετα- αλήθειας [5] και του μετα- γεγονότος εννοούμε ότι κάνουμε πολιτική χωρίς να μας απασχολεί τι είναι αληθές και τι ψευδές, τι είναι γεγονός και τι όχι, ότι ζούμε και ενεργούμε χωρίς να μας ενδιαφέρει αυτός ο χαρακτηρισμός για τα πράγματα. Αυτή η περιφρόνηση της αλήθειας είναι πιο σοβαρό πρόβλημα από το να λες απλώς πολλά ψέματα».

Το πολυσυζητημένο φαινόμενο των fake news και η επίδρασή τους διογκώθηκε στις απαρχές του 21^{ου} αιώνα και παρακάτω θα γίνει εστίαση στην πιο πρόσφατη χρονικά ουσιαστική περίπτωση επίδρασης των fake news, τις αμερικανικές προεδρικές εκλογές και επιλεκτική παρουσίαση ορισμένων περιπτώσεων fake news.

2.4 Η περίπτωση των Fake News στις αμερικανικές προεδρικές εκλογές

Μια μεγάλη έρευνα [6] για τα fake news, που διεξήγαγε και δημοσίευσε η Buz Feed News, ενισχύει την άποψη για μεταβολή της πατροπαράδοτης προπαγάνδας στην ψηφιακή εποχή σε διασπορά των ψευδών ειδήσεων. Σημειωτέον η Buz Feed News είναι μια διεθνής πλατφόρμα με ειδίκευση στα νέα μέσα και κοινωνικά δίκτυα και πέρα από ειδήσεις δημιουργεί ταυτόχρονα οπτικοακουστικό υλικό, εφαρμογές κι έρευνες σε σχέση με τα πιο διαδεδομένα ευρέως μέσα κοινωνικής δικτύωσης παγκοσμίως, το Facebook και το Twitter. Σύμφωνα με τα απίστευτα αλλά πραγματικά αποτελέσματα της εν λόγω έρευνας, τους τελευταίους 3 μήνες της προεκλογικής περιόδου των αμερικανικών προεδρικών εκλογών, ο διασκορπισμός και διανομή των fake news μέσα από τις ηλεκτρονικές πλατφόρμες του Facebook και του Twitter υπερκέρασε την αντίστοιχη ροή των συμβατικών τυπικών ειδήσεων. Τα αποτελέσματα της έρευνας αποδεικνύουν την αντιστοιχία 20 αληθινών ειδήσεων οι οποίες μεταδόθηκαν από ειδησεογραφικά μέσα διεθνούς μετάδοσης και κύρους όπως οι NY Times, Washington Post, Huffington Post, NBC News και οι οποίες αφορούσαν τις

προεδρικές εκλογές και τη διαφημιστική καμπάνια των υποψηφίων, με 20 fake news δηλαδή με ψευδείς ειδήσεις που αφορούσαν στην προεκλογική εκστρατεία των υποψηφίων και οι οποίες αναρτήθηκαν σε πλατφόρμες και ιστοσελίδες αμφιβόλου αξιοπιστίας. Αυτές οι ψευδείς ειδήσεις είχαν εξαιρετικά μεγαλύτερη αποδοχή, καθώς αναγνώστηκαν από περισσότερους χρήστες του Facebook και του Twitter σε σχέση με την αναγνωσιμότητα των αληθινών ειδήσεων.

Στην εμπεριστατωμένη έρευνα της συγκεκριμένης ανάλυσης επιβεβαιώνεται πανηγυρικά η επιρροή της διασποράς και διακίνησης των ψευδών ειδήσεων στην διαμόρφωση του τελικού εκλογικού αποτελέσματος. Είναι αισθητό ότι από τις 20 αναληθείς ειδήσεις, οι 17 ήταν υπέρ του Donald Trump είτε ζημιώνοντας άμεσα την Hillary Clinton, είτε συμπληρώνοντας στον ίδιο ευμενή στοιχεία που στην πραγματικότητα δεν είχε. Συνοπτικά, κάποιες από τις παρατραβηγμένα παραποιημένες ειδήσεις που διαδραμάτισαν θεμελιώδη ρόλο στην προεκλογική καμπάνια, ήταν πως ο ISIS προμηθευόταν οπλισμό ύστερα από αγοραπωλησίες με την Hillary Clinton, ότι ο Barack Obama δεν είχε γνήσιο πιστοποιητικό γέννησης δημιουργώντας και εμπαιγμό για την καταγωγή του και ότι ο Πάπας ήταν φίλια προσκείμενος στην εκλογή του Trump.

Άξιες μελέτης είναι οι μέθοδοι που μετέρχονται οι διαμορφωτές των fake news, προκειμένου να πετύχουν το στόχο τους που δεν είναι άλλος από την αποτελεσματική κάλυψή τους από θεσμικούς παράγοντες και αρχές που επιφορτίζονται με το καθήκον για την προστασία από τα fake news. Κατόπιν της διαδικασίας των αμερικανικών προεδρικών εκλογών αποκαλύφθηκε ότι το αρχηγείο των επιχειρήσεων και της δημιουργίας των ψευδών δημοσιευμάτων της γραμμής Trump, στην προεκλογική του εκστρατεία εντοπιζόταν σε έναν server στη Β. Μακεδονία και πιο συγκεκριμένα στα Σκόπια. Στην πόλη αυτή γινόταν ο σχεδιασμός και η παραγωγή των fake news με τα οποία τροφοδοτούνταν η αμερικανική κοινή γνώμη μέσω email, ιστοσελίδων και μέσων κοινωνικής δικτύωσης.

Η ομολογία του Ront Willer [7] συμβούλου του τηλεοπτικού δικτύου Fox news για ζητήματα ασφαλείας, εκτέλεσης του νόμου και δικαιοσύνης είναι βοηθητική της μεθοδολογία και της στρατηγικής που ακολούθησε το επιτελείο του προέδρου Trump. Ο Ront Willer επέρριψε κατηγορίες στο συγκεκριμένο τηλεοπτικό δίκτυο για την χάλκευση μιας αναληθούς ιστορίας

συναρτώμενης με την υπόθεση του Seth Rich [8], ενός υπαλλήλου στην εθνική επιτροπή του δημοκρατικού κόμματος, ο οποίος θανατώθηκε με τη χρήση όπλου τον Ιούλιο του 2016 δίπλα στο σπίτι του στην Washington. Αν και η αστυνομική έρευνα κατέληξε στο απαύγασμα ότι σκοτώθηκε κατά τη διάρκεια ένοπλης ληστείας, πολλές ιστοσελίδες και μέσα ενημέρωσης φιλικά προσκείμενα στο συντηρητικό κόμμα άφησαν περιθώρια να θεωρηθεί πως πρόκειται για δολοφονία πολιτικών σκοπιμοτήτων. Το ίδιο επίσης ειδησεογραφικό πρακτορείο το Fox news ανέβασε στον ιστότοπό του δημοσίευμα με το οποίο επικύρωνε αυτή την υπόθεση στηριζόμενο σε πλασματικές δηλώσεις του ίδιου του Ροντ Γουίλερ. Ο Γουίλερ στην προσπάθειά του να θωρακιστεί αυτός και η οικογένειά του από τις ψευδείς ειδήσεις προσέφυγε στο ομοσπονδιακό δικαστήριο του Manhattan και εκεί δήλωσε απερίφραστα ότι οι δηλώσεις του ήταν αποκυήματα της φαντασίας του τηλεοπτικού σταθμού [7].

Σε συνάρτηση και με τα παραπάνω οι πολιτικοί αναλυτές των αποτελεσμάτων των Αμερικανικών προεδρικών εκλογών, οδηγήθηκαν στο συμπέρασμα ότι η γένεση και η διάδοση των fake news από το επιτελείο του Trump ήταν μία καλά σχεδιασμένη στρατηγική και διαλογή του επιτελείου του. Η ίδια τον βοήθησε να προσεταιριστεί την πλειονότητα των Πολιτειών που είχαν μετέωρο εκλογικό εξαγόμενο αποτέλεσμα και προξένησε τη αποκαλούμενη Μυστική ψήφο[9].

Παρατίθεται παρακάτω μελέτη [10] ερευνητών του πανεπιστημίου του Στάνφορντ των Hunt Allcott και Matthew Gentzkow με τίτλο: «Κοινωνικά μέσα μαζικής ενημέρωσης και ψεύτικες ειδήσεις στις εκλογές του 2016», και με πιο τρόπο οι ψεύτικες ειδήσεις που εμπεριέχονται σε ειδησεογραφικά άρθρα και ενσωματώνουν επαληθεύσιμα ψεύδη μπορούν να παραπλανήσουν τους αναγνώστες. Οι ψεύτικες ειδήσεις συνθέτονται με υλικά, όπως είναι η φήμες, οι θεωρίες συνωμοσίας, σατιρικές εκφράσεις και πλασματικές δηλώσεις των ίδιων των πολιτικών και αναφορές που είναι μεροληπτικές ή παραπλανητικές, σε καμία περίπτωση όμως πλαστές ή αποκυήματα της φαντασίας τους.

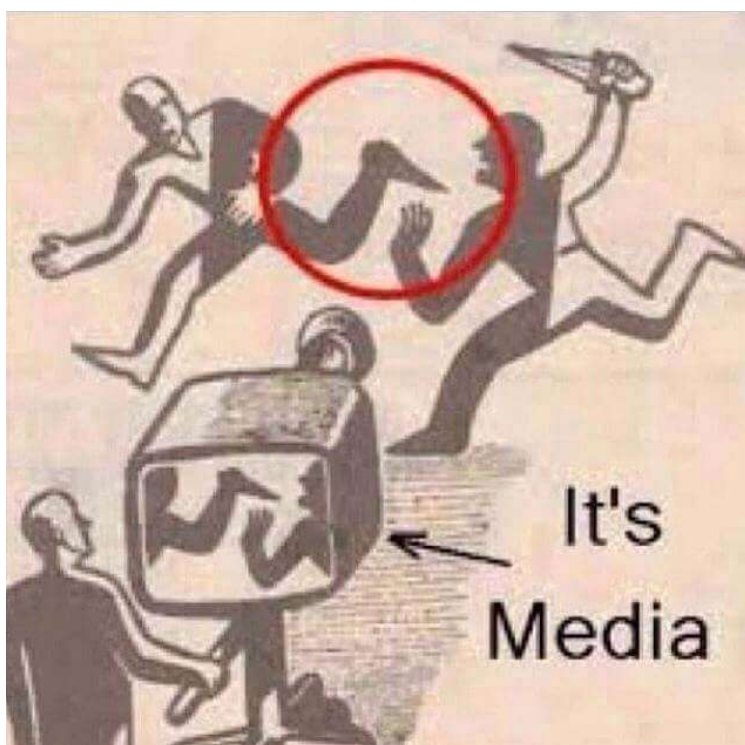
Οι προαναφερθέντες στην έρευνά τους [10] επιχειρήσαν να διακρίνουν πόσο αποτελεσματικά επηρέασαν τους Αμερικανούς ψηφοφόρους συγκεκριμένες από τις ψευδείς ειδήσεις. Στην έρευνα που πραγματοποιήθηκε συμμετείχαν 1208 πολίτες με δικαίωμα ψήφου στις απερχόμενες εκλογές και οι οποίοι, κλήθηκαν να ανταποκριθούν στο ερώτημα αν είχαν πληροφορηθεί διάφορες ειδήσεις που

είχαν θετικά ή αρνητικά συμφραζόμενα σε κάθε έναν από τους 2 υποψηφίους ανάμεσα από 3 κατηγορίες ειδήσεων. Η αρχική κατηγορία ειδήσεων περιείχε αληθινές ειδήσεις ενώ η δευτερεύουσα κατηγορία περιείχε ψευδείς ειδήσεις όπως αυτές αποσπάστηκαν από ελεγχόμενες ιστοσελίδες, όπως το Snopes και τον Politifact. Επίσης, υπήρχε και τρίτη κατηγορία ειδήσεων που ήταν επινόηση των 2 ερευνητών και περιείχε κίβδηλες αναληθείς ειδήσεις. Οι 2 ερευνητές δημιούργησαν κάποια πρωτοσέλιδα που έμοιαζαν με τα είδη των δημοσιευμάτων των παραπλανητικών ιστοσελίδων που παράγουν αναληθείς ειδήσεις, και δεν είχαν δημοσιευτεί κατά την διάρκεια της προεκλογικής εκστρατείας. Το πιο θετικό αποτέλεσμα στις απαντήσεις των συμμετεχόντων στην έρευνα είναι ότι οι περισσότεροι δήλωσαν ότι άκουσαν και πίστεψαν τις αληθινές ειδήσεις και όχι τις ψευδείς. Μόνο το 15,3% των ερωτηθέντων θυμήθηκε την αναφορά των ψευδών ειδήσεων και το 7,9% τις ανακάλεσε στη μνήμη του και τις θεώρησε αξιόπιστες.

Από τα περισσότερα άξια λόγου συμπεράσματα της ανάλυσης είναι ότι τα ποσοστά αυτά σχεδόν συμπίπτουν με τα ποσοστά των ανθρώπων που άκουσαν (14,1%) και εμπιστεύθηκαν (8,3%) τις κίβδηλες ψευδείς ειδήσεις που δημιούργησαν οι δύο ερευνητές. Το αποτέλεσμα της έρευνας στοιχειοθετεί τον τρόπο λειτουργίας των δημιουργών των ψευδών ειδήσεων. Είναι αδιαμφισβήτητο πλέον ότι ψευδείς ειδήσεις κατευθύνουν την άποψη των ανθρώπων για σημαντικές αποφάσεις και ζητήματα. Το πιο αξιοσημείωτο είναι ότι το 8% του ενήλικου πληθυσμού είναι διατεθειμένο και πρόθυμο να πιστέψει οτιδήποτε μοιάζει πιθανό και ταυτίζεται με τις προκαταλήψεις των ανθρώπων για ήρωες και παράνομους στην πολιτική.

Ένα άλλο ενδιαφέρον συμπέρασμα που ανέκυψε από την εν λόγω έρευνα είναι ότι αντίστροφα με την κυρίαρχη άποψη τα κοινωνικά μέσα ενημέρωσης δεν αποτελούν την πιο σημαντική πηγή εκλογικών επικαίρων. Έτσι μόνο το 14% του δείγματος συμφωνούσε πως αυτή η πηγή είναι σημαντική και αξιόπιστη. Η πιο αξιόλογη πηγή ειδήσεων ήταν σύμφωνα με το 23% η καλωδιακή τηλεόραση και ιδιαίτερα το 19% εξυμνούσε ως σημαντική πηγή πληροφόρησης την τηλεόραση. Εκ των υστέρων της ερευνητικής διαδικασίας, ο Gentzkow σημείωσε ότι «η ψεύτικη μνήμη είναι απόλυτα ταυτισμένη με τα ένστικτα και τις προκαταλήψεις των ανθρώπων» [10]. Έτσι αναντίρρηση προκύπτει το συμπέρασμα πως τελικά όντως η ανθρώπινη μνήμη είναι ατελείς.

Εν κατακλείδι, το πιο επίφοβο συμπέρασμα της μελέτης είναι ότι οι πιο απλά κατασκευασμένες αναληθείς ειδήσεις καθοδηγούν την συμπεριφορά και τον τρόπο με τον οποίο ο ανθρώπινος νους κατανοεί τον κόσμο γύρω του. Ένα μέρος του πληθυσμού επιζητεί ειδήσεις που προσαρμόζονται στις προκαταλήψεις τους και αυτή η διαδικασία συνεχίζεται και μετεξελίσσεται στην εποχή των μέσων ενημέρωσης με έντονες και εκρηκτικές τάσεις, ούτως ώστε να ικανοποιηθεί αυτή η ζήτηση. Η καθηγήτρια Κοινωνικών σπουδών Katie Quartuch [11] στην πραγματεία που εκπόνησε με θέμα την παιδεία στα κοινωνικά μέσα σημειώνει: «όταν σκέφτομαι τον όρο fake news σκέφτομαι κυριολεκτικά τους ανθρώπους που πληρώνονται για να παράγουν ψεύτικες ειδήσεις για πολιτικό ή οικονομικό κέρδος».



Εικόνα 2.1 Διαστρέβλωση της πραγματικότητας από τον επιλεκτικό τηλεοπτικό φακό [12].

Το συμπέρασμα στο οποίο κατέληξε η μελέτη των Allcott και Gentzkow του Στάνφορντ [10] αποτυπώνει με γλαφυρότατο τρόπο πως οι ευάρεστες ψευδείς ειδήσεις πρωταγωνίστησαν στην επιλογή του Trump από το εκλογικό δυναμικό που συμμετείχε στις εκλογές του 2016. Τα αναληθή

δημοσιεύματα και ειδήσεις που διαδόθηκαν συστηματικά και ακραιφνώς υπέρ του Donald Trump, αναρτήθηκαν στο Facebook και Twitter 30.000.000 φορές ενώ οι ψεύτικες ειδήσεις πριν από την εποχή του Bill Clinton αναρτήθηκαν 7,6 εκατομμύρια φορές. Αυτό αντιστοιχεί σε 760 εκατομμύρια χρήσεις ενός χρήστη που επισκέπτεται το διαδίκτυο και διαβάζει μία ψεύτικη ιστορία, ή περίπου 3 ψεύτικες ιστορίες ανά Αμερικανό ενήλικα χρήστη.

Πολλοί σχολιαστές υποστήριζαν ότι ο Donald Trump, δεν θα είχε εκλεγεί αν δεν είχε κάνει χρήση των fake news και ότι η πρόθεση ψήφου των ψηφοφόρων δεν θα είχε επηρεαστεί από αυτά. Το πόσο καθοριστικό ρόλο έπαιζαν τα fake news στις αμερικάνικες προεδρικές εκλογές εξαρτάται από την σημασία που είχαν στην επιλογή των ψηφοφόρων σχετικά με την ψήφο τους. Μάλιστα, σε πολλές περιπτώσεις το άκουσμα των fake news, έπαιξε πιο σημαντικό ρόλο από ένα επιτυχημένο και ευφάνταστο τηλεοπτικό εκλογικό spot.

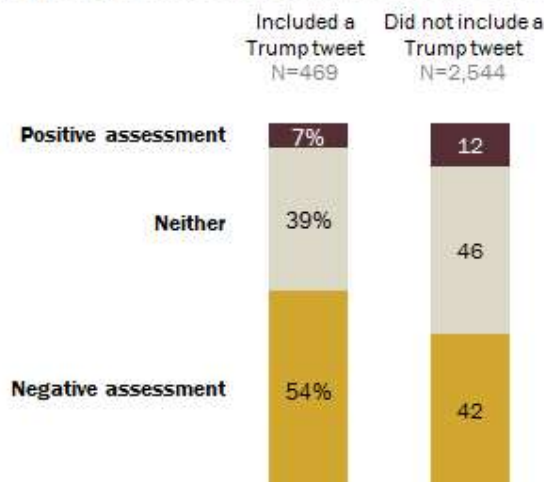
Σε άρθρο της η Angie Drobnic Holan [13] τον Οκτώβριο του 2017 που δημοσιεύτηκε στην ιστοσελίδα του Politifact, τονίζει ότι ο Πρόεδρος Trump κατά τη διάρκεια του έτους 2017 έκανε χρήση των λέξεων fake και news σε συνδυασμό 153 φορές. Παρά την έννοια του κατασκευασμένου περιεχομένου, οι αναλυτές διαφωνούν στο γεγονός ότι ο ορισμός, όπως τον καταλογίζει ο Donald Trump, έχει αξιόλογες διαφοροποιήσεις. Κάτι που επιβεβαιώνεται πανηγυρικά από το παρακάτω παράδειγμα όπου στις αρχές του ίδιου μήνα όταν ο τυφώνας Maria προκάλεσε μεγάλες καταστροφές στο Πουέρτο Ρίκο ο Donald Trump, κάνοντας χρήση του προσωπικού του λογαριασμού στο Twitter σημείωσε εκτός των άλλων πως: «έχουμε κάνει πολύ καλή δουλειά με την σχεδόν αδύνατη κατάσταση στο Πουέρτο Ρίκο», υπαινισσόμενος κατάφωρα πως τα μέσα ενημέρωσης δεν τον αντιμετωπίζουν αμερόληπτα και αποφεύγουν να δημοσιοποιήσουν τις θετικές του παρεμβάσεις σε διάφορα ζητήματα. Επίσης, σε ομιλία του κατά τη διάρκεια της προεκλογικής του εκστρατείας στο Φοίνιξ της Αριζόνα και με παρουσία ψηφοφόρων που τον υποστήριζαν και τον επευφημούσαν, εφόρμησε λεκτικά κατά των παρόντων δημοσιογράφων λέγοντας: «Οι δημοσιογράφοι είναι πραγματικά ανέντιμοι άνθρωποι. Και όχι όλοι. Έχετε μερικούς πολύ καλούς δημοσιογράφους. Αλλά οι περισσότεροι είναι ανέντιμοι άνθρωποι και νομίζω ότι δεν τους αρέσει η χώρα μας. Οι μόνοι άνθρωποι που δίνουν μια πλατφόρμα για να μισούν ομάδες είναι τα ίδια τα μέσα ενημέρωσης και τα fake news».

Το 2017 πραγματοποιήθηκε δημοσκόπηση [14] εξετάζοντας την σχέση Trump και ΜΜΕ. Η ερώτηση που κλήθηκαν να απαντήσουν οι συμμετέχοντες ήταν κατά πόσο πιστεύουν πως τα μέσα μαζικής ενημέρωσης δημιουργούν ειδήσεις με ψευδή ή αναληθή στοιχεία για την προεδρία και διακυβέρνηση Trump. Το αποτέλεσμα καταφανώς υπέδειξε με ποσοστό 46% πως οι συμμετέχοντες συμφωνούσαν με την παραπάνω ερώτηση, ποσοστό 37% συγκέντρωσε η αρνητική απάντηση ενώ οι αναποφάσιστοι για το εξεταζόμενο ερώτημα ανερχόταν σε 17%. Κατόπιν ανάλυσης των αποτελεσμάτων σε βάθος, εξ ορύχθηκε ένα ποσοστό 85% που υποστήριζε πως τα μέσα ενημέρωσης κατασκεύαζαν ψευδείς ειδήσεις για τον Πρόεδρο Trump και φυσικά άνηκε σε ψηφοφόρους θετικά προσκείμενους σε αυτόν. Όσον αφορά την πολιτική ταυτότητα των συμμετεχόντων οι Ρεπουμπλικάνοι σε ποσοστό 76% υποστήριζαν την άποψη πως τα μέσα ενημέρωσης δημιουργούσαν και διέσπειραν fake news για τον Trump, ενώ για τους Ανεξάρτητους και τους Δημοκρατικούς τα ποσοστά διαμορφώνονταν σε 44% και 20% αντίστοιχα.

Μία ακόμη έρευνα [15] με αξιόλογα αποτελέσματα διεξήχθη το 2017 από το Pew Research Center [16]. Το συγκεκριμένο ερευνητικό κέντρο εδρεύει στην Washington και θεωρείται ως το πιο ανεξάρτητο κομματικά. Παρέχει πληροφορίες για κοινωνικά θέματα, την κοινή γνώμη και δημογραφικές τάσεις που διαμορφώνονται στις ΗΠΑ καθώς και το διεθνές στερέωμα. Επίσης, διεξάγει δημοσκοπήσεις, δημογραφική έρευνα, ανάλυση περιεχομένου των μέσων ενημέρωσης και κοινωνικής δικτύωσης χωρίς να λαμβάνει συγκεκριμένη πολιτική τοποθέτηση. Στο επίκεντρο της ανάλυσής του το 2017, το ερευνητικό κέντρο έβαλε τα μέσα κοινωνικής δικτύωσης και πιο συγκεκριμένα το Twitter σε σχέση με τον Πρόεδρο Trump. Εξέτασε πάνω από 3000 κοινοποιήσεις και ειδήσεις που βρέθηκαν σε 24 μέσα ενημέρωσης και έβγαλε συμπεράσματα για τις πρώτες 100 μέρες διακυβέρνησης του Προέδρου Trump. Στη συνέχεια παρατίθεται σχολιασμός για κάθε ένα από τα αποτελέσματα-γραφήματα της έρευνας.

Stories with Trump tweets more likely to have negative assessment of Trump/ administration

% of Trump administration stories that had an overall ...



Note: Stories coded as positive had at least twice as many positive as negative statements; negative stories had at least twice as many negative as positive statements. All other stories coded as neither.
Source: Pew Research Center content analysis of news stories from national newspaper websites, radio, cable and network broadcasts and websites, and digital outlets about President Trump or his administration, Jan. 21-April 30, 2017 (Monday-Friday).

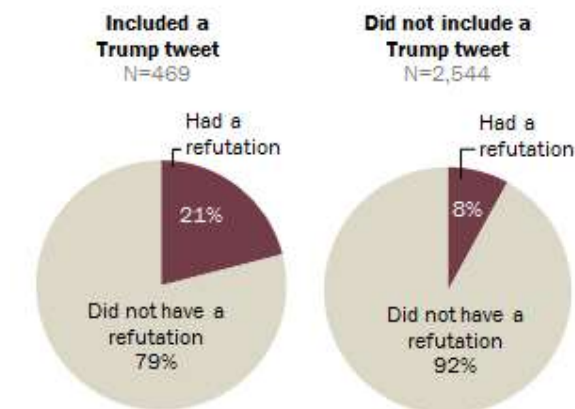
PEW RESEARCH CENTER

Εικόνα 2.2: Ταξινόμηση των tweets που σχετίζονται με τη διακυβέρνηση του Προέδρου Donald Trump σχετικά με το περιεχόμενό τους (θετικό, αρνητικό ή ουδέτερο) [15].

Σύμφωνα με όσες δημοσιεύσεις είχαν συμπεριλάβει κάποιο tweet του Trump το 54% είχε αρνητικό περιεχόμενο, το 39% ούτε θετικό ούτε αρνητικό και το 7% θετικό, ενώ όσες δημοσιεύσεις δεν περιείχαν tweet του Trump το 46% είχε ουδέτερο περιεχόμενο, το 42% αρνητικό και το 12% θετικό.

**Stories that included Trump tweets
more than twice as likely to refute
statement by Trump/administration**

*% of Trump administration stories that refuted a
Trump/administration statement*



Source: Pew Research Center content analysis of news stories from national newspaper websites, radio, cable and network broadcasts and websites, and digital outlets about President Trump or his administration, Jan. 21-April 30, 2017 (Monday-Friday).

PEW RESEARCH CENTER

Εικόνα 2.3: Διάταξη των ιστοριών που άπτονται στη διακυβέρνηση του Trump σε σχέση με το εάν συμπεριλαμβάνουν κάποιο tweet του ιδίου του προέδρου ή μέλος της κυβέρνησής του [15].

Σύμφωνα με τις δημοσιεύσεις που περιείχαν ένα tweet του Trump το 79% δεν περιείχε κάποια αναίρεση από δημοσιογράφο για θέματα που είχαν ειπωθεί από τον ίδιο ή από κάποιο μέλος της κυβέρνησής του ενώ το 21% είχε σαφή αναίρεση. Αντίθετα, όσες δημοσιεύσεις δεν περιείχαν ένα tweet του Trump μόνο ένα 8% είχαν σαφή αναίρεση.

2.5 Πρωτοβουλίες αντιμετώπισης από τα Social Media

Το 2016 το Δεκέμβριο είχε ανακοινωθεί από την κοινότητα του Facebook [17] μια καινοτόμα εφαρμογή που έδινε τη δυνατότητα στους χρήστες να αναφέρουν αναληθείς ειδήσεις στην πλατφόρμα του, με σκοπό να ελεγχθούν μέσω του διεθνούς δικτύου ελέγχου γεγονότων (IFCN). Η πρωτοβουλία του Facebook βρήκε σύμφωνα τα μέσα ενημέρωσης στα κράτη μέλη της Ε.Ε. και ξεκίνησε την λειτουργία της το Μάρτιο του 2017. Επίσης το Facebook παρουσίασε στη Βαρσοβία μια καινοτόμα μέθοδο που θα χρησιμοποιούσε ώστε να θωρακιστούν οι χρήστες σε παγκόσμια εμβέλεια από την παραπληροφόρηση και τη διάδοση αναληθών ειδήσεων. Αυτό συνέβη κάτω από την πίεση των αποκαλύψεων του σκανδάλου Cambridge Analytica [18] που έλαβε χώρα το Μάρτιο του 2018, όταν έγινε γνωστό ότι η εν λόγω εταιρεία είχε χρησιμοποιήσει ευαίσθητα δεδομένα προσωπικού χαρακτήρα 87 εκατομμυρίων ατόμων που έκανα χρήση της πλατφόρμας εκ των οποίων 2,7 εκατομμύρια ήταν Ευρωπαίοι πολίτες.

Όσον αφορά την Google [19] βασικό μέλημά της είναι η εξάλειψη της παραπληροφόρησης ιδιαίτερα όταν συμβαίνουν γεγονότα που αφορούν μεγαλύτερη μερίδα του κοινού. Όπως αναφέρεται [20] οι πηγές παραπληροφόρησης καπηλεύονται αυτά τα γεγονότα και δημοσιεύουν περιεχόμενο με ανακρίβειες με στόχο την έκθεση των χρηστών σε αυτό. Εν κατακλείδι, επισημαίνεται από την εταιρεία ότι θα συνεχίσει να συνεργάζεται με τα παγκόσμια ειδησεογραφικά δίκτυα με στόχο να αναπτυχθούν τεχνολογίες τέτοιες που θα μεγιστοποιήσουν τις επιδόσεις των ομάδων σύνταξης και θα προφυλάξουν τους δημοσιογράφους από τις κυβερνοεπιθέσεις διεθνώς.

Τέλος, σύμφωνα με συνέντευξη [21] που δόθηκε στο δίκτυο CNN από τον CEO του Twitter Jack Dorsey αναφέρθηκε ότι δεν έχει εξευρεθεί τρόπος ανακάλυψης των fake news και για αυτό δεν αφαιρούνται δημοσιεύσεις που ενδέχεται να είναι ψευδείς. Επισημάνε ότι η εταιρεία του είναι μικρότερη σε σύγκριση με άλλες πλατφόρμες όπως το Facebook και επομένως δεν έχει την ίδια οικονομική εμβέλεια ώστε να αφαιρεί παραπλανητικό περιεχόμενο. Επίσης, θα αντιμετωπιζόταν δυσκολία να κερδηθεί η εμπιστοσύνη των χρηστών αν γινόταν αφαίρεση μελών της πλατφόρμας που απλά εκφράζουν τα συναισθήματα της κοινής γνώμης ένα συγκεκριμένο χρονικό διάστημα. Τέλος

διέψευσε την επιλογή δημοσιεύσεων βάσει πολιτικών απόψεων και επισήμανε ότι η στάση του είναι ουδέτερη.

2.6 Αντίμετρα στο φαινόμενο

Στην Ελλάδα έχουν αναληφθεί τα τελευταία χρόνια πρωτοβουλίες για την αντιμετώπιση των fake news. Συγκεκριμένα, η ίδρυση της πλατφόρμας Ellinika Hoaxes [22] επιδιώκει την αποσόβηση ψευδών δημοσιευμάτων στο διαδίκτυο. Στην εν λόγω πλατφόρμα αναρτώνται σε καθημερινή βάση κείμενα της συγγραφικής ομάδας που ανατρέπουν μια μεγάλη σειρά υποτιθέμενων ειδήσεων που δημοσιεύονται στο Ελληνικό διαδίκτυο. Πριν 5 χρόνια περίπου εμφανίστηκε με τη μορφή του blog με αντικειμενικό σκοπό τη διερεύνηση των ειδήσεων που κυκλοφορούν στο διαδίκτυο και την ανάδειξη αυτών που δεν είναι αληθινές. Επειδή και στην Ελλάδα τα fake news κυκλοφορούν με μεγάλη ευκολία και σε υπερβολικό αριθμό, γρήγορα το blog έγινε προσφιλές στο αναγνωστικό κοινό και μετατράπηκε σε ιστοσελίδα που αριθμεί χιλιάδες υποστηρικτές και αναγνώστες. Σύμφωνα με τον ιδρυτή της ιστοσελίδας κ. Θοδωρή Δανιηλίδη [23] υπάρχουν συγκεκριμένοι δείκτες που βρίσκονται στην ανάρτηση που διαβάζουμε και μας βοηθούν να καταλάβουμε ότι πρόκειται για ψευδή είδηση.

Αρχικά, η απουσία συντάκτη [23,24] σε ένα κατ’ επίφαση σημαντικό κείμενο και η συνακόλουθη προτροπή για αναπαραγωγή του αποτελούν στοιχεία που υποβάλλουν τον δέκτη σε υποψίες για την αξιοπιστία του. Ακόμη, η επανάληψη χρονικών προσδιορισμών χωρίς τον ακριβή ημερολογιοχρονολογικό εντοπισμό του γεγονότος σε συνδυασμό με αναφορές σε κοινότοπα ονόματα χωρίς σαφή παρουσίαση της ταυτότητας των προσώπων που πρωταγωνιστούν στην ιστορία ενέχουν τεκμήρια για τη δημιουργία πλαστών ιστοριών. Από την άλλη η εμφάνιση πολλών ορθογραφικών λαθών και οι ασυνταξίες καταδεικνύουν ότι πρόκειται για προϊόν αυτόματης μετάφρασης και αν μάλιστα το κείμενο συνοδεύεται από άσχετες με το εξεταζόμενο θέμα φωτογραφίες χαμηλής ανάλυσης επαληθεύεται ότι το περιεχόμενο της προβαλλόμενης ιστορίας είναι πλαστό.

Εξάλλου, και το κειμενικό περιεχόμενο με θέματα σεξουαλικού περιεχομένου που εξάπτουν την περιέργεια των αναγνωστών ή με φιλανθρωπικό περιεχόμενο που επιδιώκουν να εγείρουν τα

ανθρωπιστικά αισθήματα των αναγνωστών ώστε να προσφέρουν ένα χρηματικό ποσό για την υποτιθέμενη θεραπεία κάπου ασθενή συνιστούν πλαστά κείμενα. Όσο πιο απίθανοι ισχυρισμοί διατυπώνονται σε ένα κείμενο τόσο πιο ακλόνητες αποδείξεις πρέπει να τους υποστηρίζουν. Ωστόσο, όταν ένα κείμενο δεν βασίζεται σε λογικά επιχειρήματα αλλά σε παράλογες αιροβασίες ή σε συνωμοσιολογίες και σε μη λειτουργικούς υπερσυνδέσμους (links) που έχουν διαγραφεί εξαιτίας του αβάσιμου περιεχομένου τους, τότε καταρρίπτεται η αξιοπιστία του κειμένου.

Εκτός όμως από τα λογικά σφάλματα και τις ασυνέπειες που εντοπίζουμε σε ένα κείμενο (για παράδειγμα ένα κείμενο είναι αναληθές όταν ενώ αναφέρεται σε Ευρωπαϊκή χώρα, περιέχει τιμές σε δολάρια Καναδά) και οι βαρύγδουποι και ηχηροί τίτλοι με έντονα συναισθηματικό περιεχόμενο επιβεβαιώνουν την αναξιοπιστία της ιστορίας. Επιπλέον, οι επικλήσεις σε προγενέστερες έρευνες και σε ειδικούς επιστήμονες χωρίς να γίνεται ταυτόχρονα η αναγκαία παραπομπή σε αυτούς καθώς και η υπερβολικά μεγάλη παραγωγή και ανάρτηση άρθρων, που συχνά συνοδεύονται από σχόλια με ισχυρισμούς που αποκαθλώνουν την αξιοπιστία τους, αυξάνουν τις πιθανότητες η ανάρτηση να είναι ψευδής.

Οι Έλληνες χρήστες [23] είναι πλέον περισσότερο φιλόποτοι και επιφυλακτικοί σε σύγκριση με το παρελθόν. Οι ειδήσεις που γίνονται πιο εύκολα πιστευτές είναι εκείνες που περιλαμβάνουν βία, σκάνδαλα, ξενοφοβία, πολιτική σπέκουλα και δήθεν θαυματουργές θεραπείες για ανίατες ασθένειες. Από την άλλη πλευρά όμως λόγω των πλαστών ειδήσεων, δεν πρέπει να κατηγορούμε το διαδίκτυο. Το Internet είναι ένα αναντικατάστατο εργαλείο ενημέρωσης και επικοινωνίας. Παράλληλα όμως ελλοχεύουν πολλοί κίνδυνοι και οι χρήστες θα πρέπει να είναι προσεκτικοί, γιατί η ταχύτητα που προσφέρει στην διάδοση μίας είδησης είναι ασύλληπτη και μπορεί να οδηγήσει σε εσφαλμένες διαδρομές.

Όπως προαναφέρθηκε μια άλλη πολύ αξιόλογη προσπάθεια έχει γίνει από τον προπτυχιακό φοιτητή του ΠΑ.ΜΑ.Κ., στο Τμήμα Πληροφορικής, ο οποίος [2] κατάφερε να δημιουργήσει ένα πρόγραμμα με παραμέτρους και ξεχωριστό αλγόριθμο, το FightHoax.Com [25], προκειμένου να εντοπίζει άμεσα και με ταχύτητα μία ψευδή είδηση και να την διαχωρίζει από μία πραγματική σε όλη την έκταση του διαδικτύου. Ο αλγόριθμός [2] του Fight Hoax έχει την δυνατότητα μέσα σε κλάσματα

του δευτερολέπτου να φιλτράρει κάθε είδους πληροφορία που είναι προβεβλημένη στο διαδίκτυο. Είναι ένα ηλεκτρονικό μέσο που αναλύει άρθρα ειδήσεων και παρουσιάζει την αξιοπιστία του κειμένου που διαβάζει ο αναγνώστης. Ενισχύει τους δημοσιογράφους και τους καθημερινούς ανθρώπους να αναλύουν τις ειδήσεις που διαβάζουν. Με την βοήθεια των μηχανών αναζήτησης, μπορεί να εκτελέσει μαζικές σαρώσεις στο Internet και να ανιχνεύσει τις ειδήσεις, κατατάσσοντάς τις στην κατηγορία όπου ανήκουν. Στο στόχαστρο της εφαρμογής τίθενται παραπλανητικοί τίτλοι, παραπληροφόρηση αλλά και αρθρογράφοι που παραβιάζουν συστηματικά τον κώδικα δημοσιογραφικής δεοντολογίας. Μάλιστα η συγκεκριμένη πρακτική έχει ήδη παρουσιαστεί στο Ευρωπαϊκό Κοινοβούλιο [26], και εντάσσεται και στο δίκτυο συνεργασιών του τμήματος Δημοσιογραφίας του Χάρβαρντ.

Σύμφωνα με τον δημιουργό του Βαλεντίνο Τζέκα, ο Fight Hoax στόχο έχει τον έλεγχο της αξιοπιστίας οποιουδήποτε ειδησεογραφικού κειμένου, που είναι γραμμένο στην Αγγλική γλώσσα αλλά και σε ορισμένες περιπτώσεις και στα Ελληνική, σε κλάσματα του δευτερολέπτου και με ποσοστό ακρίβειας που κυμαίνεται στο 89%. Ακόμα παρέχεται η δυνατότητα μέσω της ιστοσελίδας για δωρεάν δοκιμή ελέγχου του βαθμού αξιοπιστίας ενός διαδικτυακού κειμένου έκτασης έως 1000 λέξεων. Πρακτικά αντιπαραβάλλει το περιεχόμενο του επιλεγμένου δημοσιεύματος με άλλα δημοσιεύματα από έγκυρες πηγές, όπως το Associated Press, και εντοπίζει τυχόν σημεία απόκλισης. Για παράδειγμα, εξετάζει αν ο τίτλος είναι παραπλανητικός ή η γλώσσα παραπέμπει σε προπαγανδιστικά στοιχεία. Επιπλέον, αξιολογεί το μέσο στο οποίο αναρτήθηκε. Παράλληλα, μελετά και την αξιοπιστία του αρθρογράφου (αν έχει κατηγορηθεί για παραβίαση της δεοντολογίας στο παρελθόν ή αν έχει εργαστεί σε άλλα αξιόπιστα μέσα κλπ.). Όταν λοιπόν υπάρχουν ύποπτες ενδείξεις, προτείνει στον χρήστη δημοσιεύσεις που παρατίθενται που ενδέχεται να είναι πιο έγκυρες, στις οποίες εκείνος μπορεί να αναζητήσει αντικειμενικότερη ενημέρωση στο ζήτημα που τον ενδιαφέρει. Ο Fight Hoax δεν υποβάλει σε έλεγχο την αξιοπιστία ως προς την ορθότητα άρθρων που εκφράζουν τη γνώμη του γράφοντος ή έχουν σατυρικό περιεχόμενο, γιατί αυτό το γεγονός θα μπορούσε να θεωρηθεί λογοκρισία. Καινοτόμο βέβαια είναι ότι προτείνονται στο χρήστη άρθρα και δημοσιεύματα που

έχουν παρόμοιο περιεχόμενο με αυτά που έχει ήδη διαβάσει και παράλληλα έχει πρόσβαση σε πηγές με αντίθετη άποψη ώστε να δημιουργηθεί μια ολιστική εικόνα σε αυτόν.

2.7 Επίλογος 2^ο κεφαλαίου

Εν κατακλείδι, το φαινόμενο της παραπληροφόρησης μέσω των ψευδών ειδήσεων σε όλα τα μέσα ενημέρωσης είναι κάτι που απασχολεί τη σημερινή κοινωνία. Υποστηρικτές του φαινομένου φαίνεται να είναι κυρίως άνθρωποι που έχουν ένα δημόσιο προφίλ και προσπαθούν να επηρεάσουν την κοινή γνώμη για να αποσπάσουν ευμενή σχόλια ή να κατευθύνουν τις απόψεις των ανθρώπων για ένα ζήτημα, τη μερίδα του λέοντος καταλαμβάνουν οι πολιτικοί και εταιρίες marketing. Βέβαια εξαιτίας της έντασης του φαινομένου έχουν πραγματοποιηθεί αρκετές προσπάθειες για τη λήψη αντιμέτρων στο φαινόμενο τόσο σε επίπεδο κρατών όσο και από την πλευρά των ευρέως διαδεδομένων εταιριών στο χώρο των social media.

ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ 2^ο

- [1] <https://www.independent.co.uk/news/uk/home-news/fake-news-word-of-the-year-2017-collins-dictionary-donald-trump-kellyanne-conway-antifa-corbynmania-a8032751.html>
- [2] <https://www.cnn.gr/tech/story/92197/valentinos-tzekas-o-20xronos-larisaios-poy-polema-tis-pseydeis-eidiseis-sto-diadiktyo-vid>
- [3] Tom Holland (2017), A 2,500th Birthday for the Father of History-and fake News, article, Wall Street Journal, <https://www.wsj.com/articles/a-2-500th-birthday-for-the-father-of-history-and-fake-news-1490884822>
- [4] Κιντή Βάσω (2017), Μετά – αλήθεια και μετά – γεγονότα, άρθρο, Το Βήμα, <https://www.tovima.gr/opinions/article/?aid=856920>
- [5] Daniel J. Levithin (2017), Weaponized Lies: How to Think Critically in the Post – Truth Era, Penguin Random House LLC
- [6] <https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook>
- [7] <https://www.mediaite.com/tv/rod-wheeler-why-is-the-president-even-involved-in-a-murder-investigation/>
- [8] <https://www.theguardian.com/media/2017/aug/07/seth-rich-trump-white-house-fox-news>
- [9] <https://www.theguardian.com/commentisfree/2016/nov/14/fake-news-donald-trump-election-alt-rigth-social-media-tech-companies>
- [10] Hunt Allcot, Matthew Gentzkow (2017), Social Media and Fake News in the 2016 Election, Journal of Economic Perspectives-Volume31, Number 2, pages 211-236
- [11] <https://slspotlight.com/our-world/2017/12/18/how-fake-news-affects-american-politics/>
- [12] <https://www.facebook.com/onlythinkingart/photos/a.1381534705232967/2089347547785009/?type=3&theater>
- [13] <https://www.politifact.com/truth-o-meter/article/2017/oct/18/deciding-whats-fake-medias-definition-fake-news-vs/>

[14] <https://www.politico.com/story/2017/10/18/trump-media-fake-news-poll-243884>

[15] Pew Research Center (2017), <https://www.pewresearch.org/fact-tank/2017/10/18/in-trumps-first-100-days-news-stories-citing-his-tweets-werw-more-likely-to-be-negative>

[16] <https://www.pewresearch.org/about/>

[17] Andrew Guess, Jonathan Nagler, Joshua Tucker (2019), Less than you think: Prevalence and predictors of fake news dissemination on Facebook, Science Advances, Vol.5, no.1, eaau4586

[18] <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>

[19] <https://www.nytimes.com/2018/03/20/business/media/google-false-news.html>

[20] <https://newsinitiative.withgoogle.com/>

[21] <https://money.cnn.com/2018/08/18/media/twitter-jack-dorsey-trump-social-media/index.html>

[22] <https://ellinikahoaxes.gr/about-us/>

[23] Θοδωρής Δανιηλίδης (2014), Πως να διακρίνουμε μια Hoax δημοσίευση, άρθρο, <https://www.ellinikahoaxes.gr/about-us/>

[24] Κωνσταντίνος Αμπατζής (2017), Πώς αναχαιτίζεις τα fake new στο ελληνικό ιντερνετ, άρθρο, Oneman, <https://www.oneman.gr/keimena/diabasma/megala/keimena/pws-anaxaitizeis-ta-fake-news-sto-ellhniko-internet.4840522.html>

[25] <https://fighthoax.com/about/>

[26] https://el.wikipedia.org/wiki/%CE%92%CE%B1%CE%BB%CE%B5%CE%BD%CF%84%CE%AF%CE%BD%CE%BF%CF%82_%CE%A4%CE%B6%CE%AD%CE%BA%CE%B1%CF%82#cite_note-2

ΚΕΦΑΛΑΙΟ 3

Big Data και NLP

3.1 Εισαγωγή

Ο όρος Big Data [1], μεγάλης κλίμακας δεδομένα, χρησιμοποιείται για να περιγράψει δεδομένα τα οποία χαρακτηρίζονται από εξαιρετικά μεγάλο όγκο. Αυτό το χαρακτηριστικό του όγκου δημιουργεί προβλήματα όσον αφορά τις διαδικασίες της επεξεργασίας και ανάλυσής τους. Για να χρησιμοποιηθούν και να αποδώσουν τα μέγιστα σε θέματα που απασχολούν τους εικάστοτε ερευνητές πρέπει να ακολουθηθεί μια διαδικασία για τη συλλογής τους. Τα βήματα που ακολουθούνται κατά τη διάρκεια ολόκληρης της διαδικασίας ανάλυσης δεδομένων μεγάλης κλίμακας είναι, η εξόρυξη, αποθήκευση, διαχείριση και τελικά ανάλυση και αξιολόγησή τους.

Η ανάλυση [2] των Big Data συνδέεται άρρηκτα με το φαινόμενο των ψεύτικων ειδήσεων που αποσκοπούν στη χειραγώγηση των ανθρώπινων κοινωνιών μιας και αν γίνει ορθολογικά εξάγει αποτελέσματα από κρυφές πληροφορίες που κρύβονται εντός των δημοσιεύσεων, άρθρων, και όλων γενικά των πηγών πληροφόρησης. Πραγματοποιείται με την ανάλυση δεδομένων μεγάλης κλίμακας που προέρχονται από μέσα κοινωνικής δικτύωσης όπως το Twitter και χρησιμοποιώντας την NLP (Natural Language Processing) τη φυσική γλώσσα επεξεργασίας κειμένων από αναρτήσεις και δημοσιεύματα. Επίσης, τα μεγάλης κλίμακας δεδομένα σε πολλούς τομείς σήμερα κυμαίνονται από μερικές χιλιάδες gigabytes ως δεκάδες χιλιάδες terabytes ψηφιακών δεδομένων και εδραιώνονται πλέον σε κάθε κλάδο είτε οικονομικό είτε ευρύτερα διαχειριστικό σε θέματα συμπεριφοράς, γνώμης και τάσεων σε κάθε οργανισμό και συνασπισμό ανθρώπων. Επομένως, γίνεται εύκολα αντιληπτό πως αποτελεί κλάδο με μεγάλη άνθιση και σε συνδυασμό με την εξέλιξη της ψηφιακής τεχνολογίας, τον καθιστά ιδιαίτερα ελκυστικό στους ερευνητές παγκοσμίως.

3.2 Χαρακτηριστικά

Η έννοια [3] των μεγάλων δεδομένων έχει καθοριστεί μέσω του 3V model, μοντέλο το οποίο καθορίστηκε το 2001 από τον Laney [4] ως: «υψηλής περιεκτικότητας, υψηλής ταχύτητας και υψηλής ποικιλίας πληροφοριακά στοιχεία που απαιτούν οικονομικά αποδοτικές καινοτόμες μορφές επεξεργασίας πληροφοριών για να βελτιωθεί η διορατικότητα στη λήψη πιο ορθών αποφάσεων». Πιο πρόσφατα, το 2012, η Gartner [5] επικαιροποίησε τον ορισμό ως εξής: «Τα μεγάλα δεδομένα είναι υψηλής έντασης, υψηλής ταχύτητας ή / και υψηλής περιεκτικότητας στοιχεία πληροφοριών που απαιτούν νέες μορφές επεξεργασίας για την ενίσχυση της λήψης αποφάσεων». Και οι δύο ορισμοί αναφέρονται στα τρία βασικά χαρακτηριστικά των μεγάλων δεδομένων: Όγκος, Ποικιλία και Ταχύτητα. Άλλοι οργανισμοί και ελεύθεροι επαγγελματίες (π.χ. ερευνητές, μηχανικοί, κ.α.) έχουν επεκτείνει αυτό το μοντέλο 3V σε μοντέλο 4V, συμπεριλαμβάνοντας μια νέα τιμή "V" [6]. Αυτό το μοντέλο μπορεί να επεκταθεί ακόμη και σε 5Vs εάν ενσωματωθεί η έννοια Veracity (φιλαλήθεια) στον κύριο ορισμό μεγάλων δεδομένων. Συνοψίζοντας, αυτό το σύνολο * V-μοντέλων παρέχει έναν απλό και ευρέως αποδεκτό ορισμό που σχετίζεται με αυτό που είναι (καθώς και τι δεν είναι) ένα πρόβλημα που στηρίζεται σε big data, μία εφαρμογή ένα λογισμικό κλπ.. Αυτές οι έννοιες μπορούν να περιγραφούν σύντομα ως εξής [4,6]:

Όγκος (Volume): αναφέρεται σε μεγάλες ποσότητες οποιουδήποτε είδους δεδομένων από διαφορετικές πηγές, συμπεριλαμβανομένων κινητών ψηφιακών συσκευών δημιουργίας δεδομένων και ψηφιακών συσκευών. Το όφελος από τη συγκέντρωση, την επεξεργασία και την ανάλυση αυτών των μεγάλων ποσοτήτων δεδομένων δημιουργεί μια σειρά προκλήσεων για την απόκτηση πολύτιμων γνώσεων από ελεύθερους επαγγελματίες και επιχειρήσεις.

Ταχύτητα (Velocity): αναφέρεται στην ταχύτητα μεταφοράς δεδομένων. Τα περιεχόμενα των δεδομένων μεταβάλλονται διαρκώς μέσω της απορρόφησης συμπληρωματικών συλλογών δεδομένων, της εισαγωγής προηγούμενων συλλογών δεδομένων ή παλαιότερων συλλογών και των διαφόρων μορφών ροής δεδομένων από πολλαπλές πηγές. Από την άποψη αυτή, απαιτούνται νέοι αλγόριθμοι

και μέθοδοι για την επαρκή επεξεργασία και ανάλυση των δεδομένων σε απευθείας σύνδεση και συνεχούς ροής.

Ποικιλία (Variety): αναφέρεται σε διαφορετικούς τύπους δεδομένων που συλλέγονται μέσω αισθητήρων, smartphone ή κοινωνικών δικτύων, όπως βίντεο, εικόνες, κείμενο, ήχο, data logs και άλλα πολυμέσα. Επιπλέον, αυτά τα δεδομένα μπορούν να δομηθούν (όπως τα δεδομένα που προέρχονται από σχεσιακές βάσεις δεδομένων) ή να μην δομηθούν σε κάποια συγκεκριμένη μορφή.

Αξία (Value): αναφέρεται στη διαδικασία εξαγωγής πολύτιμων πληροφοριών από μεγάλα σύνολα κοινωνικών δεδομένων και συνήθως αναφέρεται ως ανάλυση μεγάλης κλίμακας δεδομένων Big Data. Η αξία είναι το πιο σημαντικό χαρακτηριστικό οποιασδήποτε εφαρμογής ανάλυσης μεγάλης κλίμακας δεδομένων, διότι επιτρέπει τη δημιουργία χρήσιμων πληροφοριών από τα εξαγόμενα αποτελέσματα.

Φιλαλήθεια (Veracity): αναφέρεται στην ορθότητα και ακρίβεια των πληροφοριών. Πίσω από οποιαδήποτε πρακτική διαχείρισης πληροφοριών βρίσκονται τα βασικά δόγματα της ποιότητας των δεδομένων, της διακυβέρνησης των δεδομένων και της διαχείρισης των μεταδεδομένων, καθώς και τα ζητήματα ιδιωτικού απορρήτου και νομικών ανησυχιών

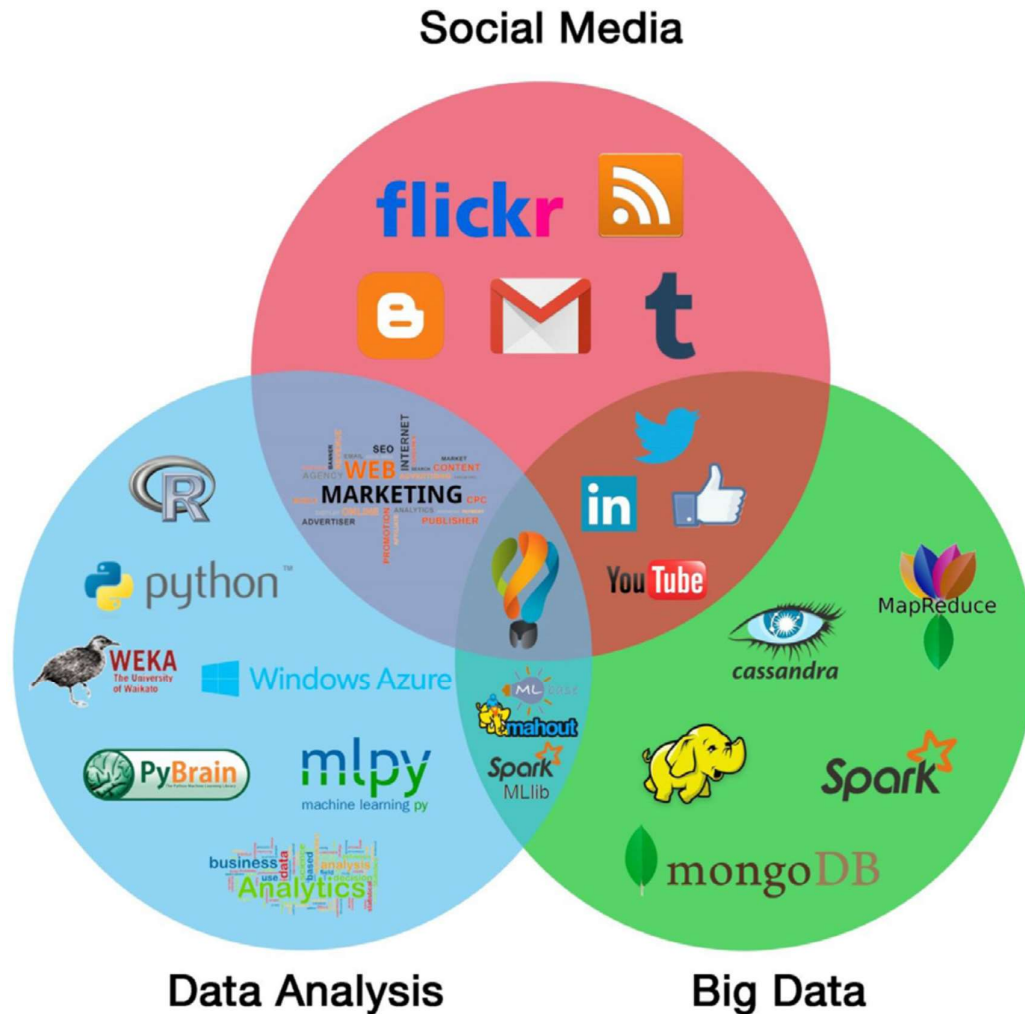
3.3 Big data και Social Media

Η αναφορά των μεγάλων δεδομένων στην παρούσα εργασία προέρχεται από την ένωση δύο τομέων: τα κοινωνικά μέσα δικτύωσης και τα μεγάλα δεδομένα. Ως εκ τούτου, τα κοινωνικά μεγάλα δεδομένα βασίζονται στην ανάλυση τεράστιων ποσοτήτων δεδομένων που θα μπορούσαν να προέρχονται από πολλαπλές κατανεμημένες πηγές αλλά με έντονη εστίαση στα κοινωνικά μέσα. Έτσι, η ανάλυση κοινωνικών μεγάλων δεδομένων [7,8] είναι εγγενώς διεπιστημονική και καλύπτει τομείς όπως εξόρυξη (άντληση) δεδομένων, μηχανική μάθηση, στατιστικά μοντέλα, εξόρυξη γραφημάτων, ανάκτηση πληροφοριών, γλωσσολογία, επεξεργασία φυσικής γλώσσας (NLP), σημασιολογικό ιστό, κ.α. Οι εφαρμογές τους μπορούν να επεκταθούν σε πολλούς τομείς όπως η υγεία και η πολιτική τάση

και πρόβλεψη, τα χόμπι, το ηλεκτρονικό επιχειρείν, το έγκλημα στον κυβερνοχώρο, η καταπολέμηση της τρομοκρατίας, η εξελισσόμενη άντληση γνώμης και η ανάλυση των κοινωνικών δικτύων. Η έννοια των μεγάλων κοινωνικών δεδομένων μπορεί να οριστεί ως εξής:

Είναι οι διαδικασίες και οι μέθοδοι που έχουν σχεδιαστεί, για να παρέχουν ευαίσθητες και σχετικές γνώσεις σε οποιονδήποτε χρήστη ή εταιρεία από πηγές δεδομένων των κοινωνικών μέσων δικτύωσης, όταν οι πηγές δεδομένων μπορούν να χαρακτηρίζονται από διαφορετικές μορφές και περιεχόμενο πολύ μεγάλου μεγέθους, καθώς και η δημιουργία τους σε απευθείας σύνδεση.

Η συγκέντρωση, αποθήκευση, επεξεργασία και ανάλυση των δεδομένων από τα κοινωνικά μέσα δικτύωσης που προέρχεται είτε από αδόμητες είτε από ημιδομημένες πηγές, οπότε η απόκτηση γνώσεων σχετικά με την αξία τους είναι ένα εξαιρετικά δύσκολο έργο που δεν έχει λυθεί πλήρως. Οι κλασσικές μέθοδοι, αλγόριθμοι και εργαλεία διαχείρισης δεδομένων έχουν καταστεί ανεπαρκή για την επεξεργασία του τεράστιου όγκου δεδομένων. Το ζήτημα αυτό δημιούργησε ένα μεγάλο αριθμό ανοιχτών προβλημάτων και προκλήσεων σχετικά με τον μεγάλο τομέα κοινωνικών μεγάλων δεδομένων που σχετίζεται με διαφορετικές πτυχές όπως η διαχείριση δεδομένων, η επεξεργασία δεδομένων, η ανάλυση δεδομένων και η απεικόνιση τους [9].



Εικόνα 3.1 Ο εννοιολογικός χάρτης των Κοινωνικών Μεγάλων Δεδομένων [3].

Μερικές από αυτές τις προκλήσεις περιλαμβάνουν την πρόσβαση σε πολύ μεγάλες ποσότητες μη δομημένων δεδομένων (θέματα διαχείρισης), τον προσδιορισμό του αριθμού των δεδομένων που επαρκούν για τη λήψη μιας μεγάλης ποσότητας δεδομένων υψηλής ποιότητας (ποιότητα έναντι ποσότητας), την επεξεργασία της ροής δεδομένων που αλλάζει δυναμικά ή τη διασφάλιση επαρκούς προστασίας (ιδιοκτησία και ασφάλεια), μεταξύ άλλων. Ωστόσο, δεδομένης της πολύ μεγάλης ετερογενούς δέσμης δεδομένων από τα κοινωνικά μέσα δικτύωσης, προκύπτει μία από τις σημαντικότερες προκλήσεις, ο προσδιορισμός των πολύτιμων δεδομένων και ο τρόπος ανάλυσης

αυτών για να ανακαλυφθούν χρήσιμες γνώσεις που βελτιώνουν τη λήψη αποφάσεων από μεμονωμένους χρήστες όπως ελεύθεροι επαγγελματίες καθώς και επιχειρήσεις όλων των κλάδων [10].

Είναι άξια καταγραφής ορισμένα σημαντικά παραδείγματα [1] που διευκρινίζουν καλύτερα το μέγεθος του όρου Big Data. Κάποιος για παράδειγμα, μπορεί να αγοράσει μία μονάδα δίσκου με ικανότητα να αποθηκεύσει την παγκόσμια μουσική. Επίσης, τα μέσα άντλησης γνώσης από τα δεδομένα σημειώνουν σημαντική βελτίωση καθώς τα διαθέσιμα λογισμικά για την εφαρμογή τεχνικών αυξανόμενης πολυπλοκότητας συνδυάζονται με την αυξανόμενη υπολογιστική ισχύ. Επιπλέον, η δυνατότητα παραγωγής επικοινωνίας και πρόσβασης σε δεδομένα έχει εκτοξευθεί από την αύξηση του αριθμού των ατόμων, συσκευών και αισθητήρων που συνδέονται σήμερα σε ψηφιακά δίκτυα. Το 2018 περισσότερα από τέσσερα δισεκατομμύρια άνθρωποι ή το 70% του παγκόσμιου πληθυσμού χρησιμοποιούσαν κινητά τηλέφωνα και περίπου 65% από αυτούς τους ανθρώπους είχαν smartphone των οποίων η χρήση αυξάνεται κατά περισσότερο από 20% το χρόνο.

3.4 Twitter και Big Data

Το μέσο κοινωνικής δικτύωσης που έχει την ευκολότερη και δωρεάν πρόσβαση σε δεδομένα μεγάλης κλίμακας είναι το Twitter. Μέσω μιας διαδικασίας ο καθένας μπορεί να έχει πρόσβαση σε δεδομένα χρηστών του Twitter απλά κάνοντας κλικ στην ιστοσελίδα του και συμπληρώνοντας μια αίτηση. Η διαδικασία αυτή επιτυγχάνεται με χρήση εφαρμογών τύπου API (Application Programming Interfaces) [11] την εφαρμογή αυτή χρησιμοποιούν επιχειρήσεις, προγραμματιστές και οποιοσδήποτε χρήστης για να συλλέξει δωρεάν δεδομένα μεγάλης κλίμακας από την πλατφόρμα για να τα χρησιμοποιήσει όπως επιθυμεί.

Τα δεδομένα προέρχονται από σχόλια και αναρτήσεις ενώ με την κατάλληλη επιλογή μπορούν να συλλεχθούν και προσωπικά κρυφά μηνύματα. Με τη συλλογή, επεξεργασία και ανάλυσή τους μπορούν να προσφέρουν αξιόλογες και αρκετά αξιόπιστες πληροφορίες για την κοινή γνώμη και τις τάσεις που επικρατούν καθώς και για απόψεις των χρηστών που προκύπτουν από θέματα της

επικαιρότητας. Όλα αυτά τα δεδομένα μπορούν να συλλεχθούν με την εισαγωγή λέξεων κλειδιών στην αναζήτηση ενός θέματος με δυνατότητα επιλογής φίλτρου γλώσσας (όπως για παράδειγμα: fashion trends in shirts, election 2016 USA κ.α.) και να ληφθούν σε μορφή αρχείου JSON (JavaScript Object Notation) [12]. Τα αρχεία με αυτή την επέκταση είναι ανοιχτά μορφότυπα τα οποία χρησιμοποιούν κείμενο σε φυσική γλώσσα που μπορεί να αναγνωσθεί από τον άνθρωπο και βασίζονται σε ένα υποσύνολο της γλώσσας JavaScript.

3.5 NLP και δεδομένα μεγάλης κλίμακας

Τα δεδομένα που συλλέγονται όπως προαναφέρθηκε είναι μεγάλης κλίμακας, βρίσκονται σε φυσική γλώσσα και πρέπει να κατηγοριοποιηθούν και να επεξεργαστούν κατάλληλα για να αποδώσουν την πληροφορία που κρύβουν μέσα τους. Η επεξεργασία φυσικής γλώσσας [13] είναι ένας κλάδος της επιστήμης της πληροφορικής, της τεχνητής νοημοσύνης και της υπολογιστικής γλωσσολογίας που αφορά στην αλληλεπίδραση μεταξύ υπολογιστών και της ανθρώπινης (φυσικής) γλώσσας. Συνεπώς, η NLP (Natural Language Processing) αποσαφηνίζει την αλληλεπίδραση ανθρώπου και μηχανής στη μετάδοση πληροφοριών και δεδομένων κειμένου. Κύριο αντικείμενο έρευνας αυτής της επιστήμης είναι η κατανόηση της φυσικής γλώσσας με το να παρέχει έτσι τη δυνατότητα στους υπολογιστές να εξάγουν συμπεράσματα από τέτοιου είδους εισόδους δεδομένων. Κάτι τέτοιο θα μπορούσε να χρησιμοποιηθεί για να ερμηνευθεί και να καταπολεμηθεί μια ψευδή είδηση που στόχο έχει την παραπληροφόρηση και την χειραγώγηση του ανθρώπου.

Οι σύγχρονοι NLP αλγόριθμοι εδράζονται στην μηχανική μάθηση, και πιο συγκεκριμένα στην στατιστική μηχανική μάθηση. Παλαιότερα όλη αυτή η διαδικασία επεξεργασίας φυσικής γλώσσας γινόταν χειροκίνητα με τη βοήθεια πληθώρας κανόνων και υποπεριπτώσεων. Η μηχανική μάθηση όμως έχει συντελέσει βοηθητικά στη διαδικασία αυτή χρησιμοποιώντας αλγορίθμους μάθησης όλων αυτών των κανόνων που χρειάζονται για να κατηγοριοποιήσουν στοιχεία φυσικής γλώσσας κατόπιν εισόδου σε αυτούς δεδομένων κειμένου. Στην ουσία είναι μια διαδικασία στην οποία δημιουργείται ένας αλγόριθμος μηχανικής μάθησης, κατόπιν τροφοδοτείται με δεδομένα μεγάλης κλίμακας

(κείμενο στην προκειμένη περίπτωση) από διάφορες πηγές όσο το δυνατόν πιο αμερόληπτα και ορθά και αφού εκπαιδευθεί κατάλληλα, σε δεδομένη είσοδο (ερώτηση) δεδομένου είναι σε θέση να το κατηγοριοποιήσει και να εξάγει ένα αποτέλεσμα για την ταυτότητα του (ετικέτα).

Η παρακάτω λίστα [14] περιέχει μερικούς από τους τομείς έρευνας στην επιστήμη της NLP. Να σημειωθεί ότι μερικά από αυτά τα πεδία έχουν άμεση εφαρμογή σε πραγματικά προβλήματα, ενώ άλλα αξιοποιούνται κυρίως ως επιμέρους υποενέργειες και χρησιμεύουν για την επίλυση μεγαλύτερων προβλημάτων επίσης, οι διεργασίες επεξεργασίας φυσικής γλώσσας είναι άρρηκτα συνδεδεμένες μεταξύ τους, συχνά όμως υποδιαιρούνται σε κατηγορίες για λόγους ευκολότερης κατανόησης:

Σύνταξη [15]

Κατάταξη Γραμματικής: Δημιουργεί μια τυπική γραμματική που περιγράφει τη σύνταξη μιας γλώσσας.

Μορφολογικός κατακερματισμός: Ξεχωρίζει λέξεις από ένα κείμενο σε μεμονωμένες οντότητες και προσδιορίζει την τάξη των μορφωμάτων. Η δυσκολία αυτής της διαδικασίας εξαρτάται σε μεγάλο βαθμό από την πολυπλοκότητα που έχει η μορφολογία (δηλαδή η δομή των λέξεων) της γλώσσας που εξετάζεται. Η αγγλική γλώσσα για παράδειγμα έχει μια σχετικά απλή μορφολογία και έτσι μπορεί αυτή η διαδικασία να παραληφθεί και να αντικατασταθεί με μια απλή μοντελοποίηση όλων των πιθανών μορφών μιας λέξης (π.χ. open, opening, opens, opened). Από την άλλη μεριά η ελληνική γλώσσα έχει δυσκολότερη μορφολογία και μια μοντελοποίηση δεν είναι εφικτή διότι το κάθε λεξικό έχει δεκάδες πιθανές διαφοροποιήσεις της ίδιας λέξης.

Ετικέτα σε μέρη του λόγου: Δίνοντας μια πρόταση κειμένου (είσοδος) καθορίζει το μέρος του λόγου για κάθε λέξη της ξεχωριστά. Πολλές λέξεις ιδιαίτερα κοινές μπορούν να χρησιμοποιηθούν ως πολλαπλά μέρη του λόγου για παράδειγμα η λέξη book στην αγγλική που μπορεί να χρησιμοποιηθεί ως ουσιαστικό για να δηλώσει το βιβλίο αλλά και ως ρήμα για να εκφράσει το ότι κλείνω, κανονίζω εισιτήρια φέρ' ειπείν. Στην ελληνική γλώσσα δεν υπάρχουν τέτοια φαινόμενα καθώς οι λέξεις είναι

μοναδικά μέρη του λόγου και για να μεταπηδήσουμε από ένα μέρος σε ένα άλλο θα πρέπει να διαφοροποιήσουμε σύνταξη της λέξης, υπάρχουν μόνο ομόηχες που έχουν διαφορετικό νόημα.

Στοχαστική γραμματική: Προσδιορίζει τη γραμματική ανάλυση μιας δοθείσας πρότασης κειμένου. Η γραμματική για τις φυσικές γλώσσες που χρησιμοποιούν οι άνθρωποι είναι διαφορετική και οι τυπικές προτάσεις έχουν πολλές πιθανές αναλύσεις. Έτσι ίσως απροσδόκητα για μια τυπική πρόταση ενδέχεται να υπάρχουν χιλιάδες αναλύσεις από έναν υπολογιστή πολλές από τις οποίες θα φαινόταν αφελής σε κάποιον άνθρωπο.

Εντοπισμός προτάσεων: Λαμβάνοντας ένα κομμάτι κειμένου βρίσκει τα όρια των προτάσεων που το αποτελούν. Τα όρια των προτάσεων συχνά χαρακτηρίζονται από περιόδους ή σημεία στίξης, αλλά για έναν υπολογιστή οι ίδιοι χαρακτήρες μπορεί να σημαίνουν μια συντομογραφία για παράδειγμα.

Εύρεση ρίζας: Δίνοντας προτάσεις βρίσκει για την κάθε λέξη την ρίζα από την οποία προήλθε.

Τμηματοποίηση λέξεων: Ξεχωρίζει ένα κομμάτι συνεχόμενου κειμένου σε ξεχωριστές λέξεις, στην αγγλική η την ελληνική γλώσσα αυτό το πρόβλημα δεν συναντάται αφού οι λέξεις χωρίζονται με κενό. Σε άλλες γλώσσες όμως, όπως τα κινέζικα δεν επισημαίνονται τα όρια των λέξεων με τον ίδιο τρόπο επομένως χρησιμοποιείται αυτός ο διαχωρισμός των λέξεων από τα κείμενα.

Εξαγωγή ορολογίας: Έχει ως στόχο τον αυτόματο εντοπισμό των λέξεων που αποτελούν κομμάτια ορολογίας σε ένα κείμενο φυσικής γλώσσας και υπόκεινται σε συγκεκριμένους κανόνες κατανόησης.

Σημασιολογία

Λεξική σημασιολογία: εξετάζει τη λεξιλογική σημασία των λέξεων που αποτελούν το δοσμένο κείμενο.

Σημασιολογική κατανομή: εξετάζει πώς μπορούμε να ειλάβουμε τις σημασιολογικές αναπαραστάσεις των λέξεων από τα δεδομένα και τις κατανέμει κατάλληλα.

Μηχανική μετάφραση: Μεταφράζει αυτόματα ένα κείμενο από μία ανθρώπινη φυσική γλώσσα σε κάποια άλλη. Αυτό αποτελεί ένα από τα πιο δύσκολα προβλήματα και καλείται ΑΙ (Artificial Intelligence) Complete διότι απαιτεί την γνώση από τον υπολογιστή όλων των ανθρώπινων δεξιοτήτων και γνώσεων σε θέματα γραμματικής, σημασιολογίας και γεγονότων σχετικά με τον πραγματικό κόσμο ώστε να επιλυθεί σωστά.

Παραγωγή φυσικής γλώσσας: Μετατρέπει πληροφορίες από βάσεις δεδομένων υπολογιστών σε αναγνώσιμη ανθρώπινη γλώσσα.

Κατανόηση φυσικής γλώσσας: Μετατρέπει κομμάτια κειμένου σε στοιχεία όπως λογικές δομές πρώτης τάξης που είναι ευκολότερα κατανοητές από τους υπολογιστές. Η κατανόηση της φυσικής γλώσσας περιλαμβάνει τον προσδιορισμό της σημασιολογικής σημασίας του κειμένου, από πολλαπλές πιθανές που μπορεί να προκύψουν και τις οργανώνει σε γλωσσικές φυσικές έννοιες.

Οπτική αναγνώριση χαρακτήρων: Καθορίζει το κείμενο που υπάρχει μέσα σε μια εικόνα [16].

Απάντηση ερωτημάτων: Λαμβάνει μια ερώτηση από έναν άνθρωπο και αποδίδει μια απάντηση σε αυτό. Οι τυπικές ερωτήσεις έχουν μια συγκεκριμένη σωστή απάντηση όπως για παράδειγμα (Ποια η πρωτεύουσα της Ελλάδας;), υπάρχουν όμως και ερωτήσεις πάνω σε ανοικτά ζητήματα όπως ποιο είναι το νόημα της θρησκείας ή της ζωής που δεν επιδέχονται μία απάντηση [17].

Εξαγωγή σχέσεων: Λαμβάνοντας ένα κομμάτι κειμένου προσδιορίζει τις σχέσεις ανάμεσα στις δοθείσες οντότητες που το απαρτίζουν (ποιο παιδί ανήκει σε ποια οικογένεια).

Ανάλυση συναισθήματος: Αποσπά υποκειμενικές πληροφορίες (κείμενα) συνήθως από ένα σύνολο δεδομένων μεγάλης κλίμακας και τις κατηγοριοποιεί, χρησιμοποιώντας το συναίσθημα που κρύβονται πίσω από τις λέξεις τους. Με αυτό τον τρόπο μπορεί να παρουσιάσει και να καθορίσει με σχετικά μεγάλη επιτυχία την πολικότητα γύρω από ένα θέμα. Αυτό αποτελεί ένα ιδιαίτερα χρήσιμο εργαλείο στον τομέα της έρευνας διότι εντοπίζει τις τάσεις της κοινής γνώμης στα μέσα κοινωνικής δικτύωσης.

3.6 Επίλογος 3^{ου} κεφαλαίου

Κατόπιν μακροσκελούς έρευνας διαπιστώθηκε πως τα δεδομένα μεγάλης κλίμακας πρωταγωνιστούν στη ζωή μας και παρουσιάστηκαν τα χαρακτηριστικά τους και τρόποι που χρησιμοποιούνται. Στο συγκεκριμένο κεφάλαιο αφού παρατάθηκαν τα χαρακτηριστικά στοιχεία των μεγάλων δεδομένων έγινε αναφορά σε έναν πιο ειδικό κλάδο των social media που συναντώνται, το Twitter. Παρουσιάστηκε ένας νέος καινοτόμος τομέας έρευνας που προέρχεται από τέτοιου είδους δεδομένα, η επεξεργασία φυσικής γλώσσας και πώς αυτή αναλύεται στην πράξη. Τέλος, έγινε αναφορά στην ανάλυση συναισθήματος που είναι και το αντικείμενο μελέτης του επόμενου κεφαλαίου.

ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ 3^{ου}

- [1] McKinsey (2011), Big data: The next frontier for innovation, competition and productivity. [ed.] Global Institute.
- [2] Juniper networks (2012), Introduction to Big Data: Infrastructure and Networking Considerations, p.11.
- [3] Gema Bello-Orgaz, Jason J. Jung, David Camacho (2015), Social big data: Recent achievements and new challenges. Information Fusion.
- [4] D. Laney (2001), 3D Data Management: Controlling Data Volume, Velocity, and Variety, Technical Report, <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
- [5] M.A. Beyer, D. Laney (2012), The Importance of ‘Big Data’: A Definition, Gartner, Stamford, CT.
- [6] I.A.T. Hashema, I. Yaqooba, N.B. Anuara, S. Mokhtara, A. Gania, S.U. Khanb (2015), The rise of big data on cloud computing: review and open research issues, Inf. Syst. 47 pp98–115.
- [7] E. Cambria, D. Rajagopal, D. Olsher, D.Das (2013), Big social data analysis, Big Data Comput.13 pp 401–414.
- [8] L. Manovich (2011), Trending: the promises and the challenges of big social data, Debates Digit. Hum. pp 460–475.
- [9] S. Kaisler, F. Armour, J.A. Espinosa, W. Money (2013), Big data: Issues and challenges moving forward, in: Proceedings of 46th Hawaii International Conference on System Sciences (HICSS), IEEE, , pp.995–1004.
- [10] H. Chen, R.H. Chiang, V.C. Storey (2012), Business intelligence and analytics: from big data to big impact, MISQ. 36 (4) pp1165–1188.
- [11] <https://help.twitter.com/en/rules-and-policies/twitter-api>.
- [12] Doug Crockford (2009), Google Tech Talks: JavaScript: The Good Parts, <http://talks.ui-patterns.com/video/javascript-the-good-parts>.

- [13] Ellen Riloff, Janyce Wiebe (2003), Learning extraction patterns for subjective expressions,
Proceedings of the Conference on Empirical Methods in Natural Language Processing, (2003).
- [14] Liddy, E.D. (2001), Natural Language Processing, In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc.
- [15] Klein Dan, Christopher D. Manning (2002), Natural language grammar induction using a constituent-context model, Advances in neural information processing systems.
- [16] Yucong Duan, Christophe Cruz (2011), Formalizing Semantic of Natural Language through Conceptualization from Existence. Archived 2011-10-09 at the Wayback Machine International Journal of Innovation, Management and Technology, pp. 37-42.
- [17] Mittal (2011), Versatile question answering systems: seeing in synthesis, IJIIDS, pp119-142.

ΚΕΦΑΛΑΙΟ 4

Κατηγοριοποίηση ενός κειμένου με βάση τη συναισθηματική του ανάλυση

4.1 Εισαγωγή

Όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο το φαινόμενο των δεδομένων μεγάλης κλίμακας διογκώθηκε με την αναβάθμιση της πληροφορίας και την εξέλιξη της τεχνολογίας. Δεδομένα μπορούν να δημιουργηθούν και να συλλεχθούν πολύ γρήγορα και είναι τόσο μεγάλης ποσότητας, που έχουν ξεπεράσει κατά πολύ τη δυνατότητα επεξεργασίας τους από τον άνθρωπο [1]. Εξαιτίας αυτού του προβλήματος καθώς και των χαρακτηριστικών τους, που έχουν προαναφερθεί, τα δεδομένα μεγάλης κλίμακας χρήζουν διαχείρισης από υπολογιστικές μηχανές. Ο κλάδος δεδομένων ο οποίος αναφέρεται στην παρούσα εργασία ασχολείται με τα μέσα κοινωνικής δικτύωσης τα οποία είναι μία πολύ καλή πηγή εξόρυξης δεδομένων μεγάλης κλίμακας στις μέρες μας εξαιτίας της εξέλιξης του διαδικτύου.

Οι άνθρωποι μπορούν να επικοινωνούν όποτε θέλουν σε οποιοδήποτε μέρος στον κόσμο, να δημοσιεύουν τις σκέψεις τους, να μεταφορτώνουν φωτογραφίες και να διαμοιράζονται ενδιαφέροντα βίντεο. Όλα αυτά τα δεδομένα που περιέχονται στα κοινωνικά δίκτυα είναι δημιουργήματα των χρηστών και αποτελούν δεδομένα υψηλής αξίας διότι είναι επιστέγασμα των απόψεών τους. Υπάρχει μία νέα φράση που τείνει να γίνει viral και ονομάζεται το ‘Internet minute’ που αναφέρεται, σε ό,τι συμβαίνει, σε ένα λεπτό στο Internet. Σύμφωνα με αναφορά της Intel [2] 347,222 tweets δημοσιεύτηκαν στο Twitter, 138,889 ώρες βίντεο παρακολούθηθηκε στο YouTube και 38,194 φωτογραφίες μεταφορτώθηκαν στην πλατφόρμα του Facebook σε ένα μόνο λεπτό. Τα μέσα κοινωνικής δικτύωσης έχουν γίνει απαραίτητο μέρος της καθημερινής μας ζωής και το μέγεθος των

δεδομένων που παράγονται μέσα σε αυτά ολόένα και αυξάνεται. Στην παρούσα εργασία γίνεται αναφορά στο Twitter που είναι πολύ σημαντική και αξιόλογη πηγή δεδομένων προσωπικού χαρακτήρα. Σε αυτά τα δεδομένα κειμένου περιέχεται το συναίσθημα των συντακτών τους και γενικότερα η ψυχική τους κατάσταση τη στιγμή που πληκτρολογούν.

4.2 Κατηγοριοποίηση κειμένου

Όταν οι άνθρωποι δημοσιεύουν ένα tweet διακατέχονται [3] από τα δικά τους συναισθήματα και απόψεις για το θέμα στο οποίο αναφέρονται, όπως είναι η ικανοποίηση, η απογοήτευση, θετικά ή αρνητικά συναισθήματα. Αυτά τα δεδομένα από τα συναισθήματα των χρηστών του Twitter για παράδειγμα είναι μία εξαιρετική πηγή δεδομένων για εταιρείες ή οργανισμούς που εδράζονται στο μάρκετινγκ ώστε να κάνουν επισκόπηση των πελατών και των προτιμήσεών τους. Παράλληλα όμως μπορεί να γίνουν και πολύ επιζήμια αν χρησιμοποιηθούν για πολιτική σκοπιμότητα ή χειραγώγηση της κοινής γνώμης.

Στα δεδομένα κειμένου [4], ενδημούν δύο τύποι πληροφορίας: η εμφανής, στην οποία μπορούν να υπαχθούν όλα τα χαρακτηριστικά γνωρίσματα του κειμένου και της πλατφόρμας που αυτή τα υποδέχεται όπως το χρονικό σημάδι δημοσίευσής τους (είναι μια ακολουθία χαρακτήρων που προσδιορίζουν πότε συνέβη ένα συγκεκριμένο συμβάν, δίνοντας ημερομηνία και ώρα, ορισμένες φορές ακριβείς σε ένα μικρό κλάσμα του δευτερολέπτου), ο χρήστης που τα επινόησε, στοιχεία τοποθεσίας κ.α. και η κρυφή πληροφορία. Το κειμενοκίνητο κομμάτι της πληροφορίας κειμένου έχει ένα αξιόλογο σημαντικό εύρος και αρκετές φορές μπορεί να χρησιμοποιηθεί σε μια αφθονία εφαρμογών όπως: λιτές θεματολογικές αναπτύξεις, ερμηνεύσεις της βούλησης και πρόθεσης του συγγραφέα και το μοτίβο προσέγγισης και χρήσης ηλεκτρονικών υπηρεσιών ενός χρήστη, είναι μερικές από τους αρκετούς προσανατολισμούς έρευνας και υλοποίησης πληθώρας εφαρμογών και λογισμικού.

Η επεξήγηση και η σπουδή αυτής της πληροφορίας, εισέρχεται σε μια γενικότερη κατηγορία ανάλυσης που καλείται Βαθιά Μάθηση [5]. Η Βαθιά Μάθηση, αποπειράται να διερευνήσει την

πληροφορία που ενδημεί στα βαθύτερα κομμάτια των δεδομένων και τις αφανέστες αλληλουχίες αναφοράς μεταξύ τους. Σημειώνεται πως η κατηγορία αυτή ουσιαστικά εδραιώθηκε στις απαρχές του 1980 με τα νευρωνικά δίκτυα [3], μόλις όμως την τελευταία 15ετία και εξαιτίας της μεγάλης ποσότητας της πληροφορίας που αναφέρθηκε ανωτέρω, έχει γνωρίσει πληθώρα εφαρμογών και νεοσύστατων προσεγγίσεων.

Σ’ αυτές, πέραν των αφαιρετικών μεθόδων που είναι άξιες μελέτης, εισέρχονται και εφαρμοσμένοι κλάδοι διαχωρισμού, ανάλυσης, ερμηνείας και κατηγοριοποίησης κειμένου, οι οποίοι είναι ειδικά διαμορφωμένοι να συμβαδίζουν επάνω στους ίδιους τους χρήστες και τις επιθυμίες τους, το καταναλωτικό προφίλ τους, τρόπους σερφαρίσματος στο Ίντερνετ, την άποψή και γνώμη τους επάνω σε ένα θέμα της επικαιρότητας και της πολιτικής σκηνής κλπ. Πώς όμως γίνεται η ανάλυση και η κατηγοριοποίηση του κειμένου φυσικής γλώσσας;

Η κατηγοριοποίηση του κειμένου [6] είναι ένα μέρος της επεξεργασίας φυσικής γλώσσας που κατηγοριοποιεί το κείμενο βασιζόμενο στο περιεχόμενό του. Μπορούν εύκολα να κατηγοριοποιηθούν δεδομένα από εφημερίδες άρθρα ή βιβλία σε διαφορετικές κατηγορίες σύμφωνα με τα χαρακτηριστικά που τους προσδίδει ο αναλυτής. Όταν γίνεται λόγος για κατηγοριοποίηση κειμένου συνήθως αναφέρεται σε μεθόδους που χρησιμοποιούν μηχανική εποπτευόμενη μάθηση κατηγοριοποίησης δηλαδή υλοποιείται με τη χρήση αλγορίθμων. Η διαδικασία της κατηγοριοποίησης περιλαμβάνει δύο στάδια: το πρώτο στάδιο είναι αυτό της εκμάθησης από τον αλγόριθμο, δηλαδή της τροφοδότησης δεδομένων με ετικέτες και χαρακτηριστικά που επιθυμεί ο αναλυτής να εξετάσει μέσω προσθήκης, κατά τη διάρκεια της μάθησης Big Data. Το δεύτερο στάδιο είναι αυτό της δοκιμής, όπου περιλαμβάνεται η προεπεξεργασία του αρχικού περιεχομένου δεδομένων που έχουν εισαχθεί κατά την εκμάθηση, η ανάλυσή και η κατηγοριοποίησή τους με βάση κάποια χαρακτηριστικά. Παρακάτω φαίνεται η διαδικασία αυτή παραθέτοντας τα 5 βήματα που ακολουθούνται.

1. Δημιουργία [7]

Συλλογή κειμένου βάσει κατηγοριών. Κάθε κείμενο ανήκει σε μία κατηγορία και του αποδίδεται μία ετικέτα με έναν χαρακτηρισμό. Μερικές φορές αυτό το στάδιο επαναχρησιμοποιείται και στα δύο στάδια (εκπαίδευσης, δοκιμών) ιδίως αν δεν υπάρχουν μεγάλα ποσοστά επιτυχίας στις δοκιμές.

2. Προεπεξεργασία

Αφαιρούνται όλα τα περιττά στοιχεία του κειμένου, όπως οι τελείες και γενικά τα σημεία στίξης ή μη αναγνώσιμο κείμενο(εικόνα). Αυτό το βήμα είναι πολύ σημαντικό γιατί επηρεάζει την εκπαίδευση του ταξινομητή.

3. Μετατροπή κειμένου σε διάνυσμα χαρακτηρισμών

Κατόπιν το κείμενο μετατρέπεται σε γλώσσα μηχανής ούτως ώστε να μπορεί να αναγνωριστεί από την υπολογιστική μηχανή. Όλα τα κείμενα αναπαρίστανται σαν διανύσματα όρων βάσει των χαρακτηριστικών που επιλέχθηκαν.

4. Κατάρτιση του ταξινομητή

Επιλέγεται ένας από τους αλγορίθμους ταξινόμησης (υπάρχουν έτοιμες βιβλιοθήκες ακόμη και στο Matlab που χρησιμοποιούν τέτοιους αλγορίθμους) τροφοδοτείται με δεδομένα και ξεκινάει η εκμάθηση του.

5. Ταξινόμηση

Αφού το μοντέλο ταξινόμησης εκπαιδευτεί τότε τροφοδοτείται εκ νέου με δεδομένα και αποδίδει τις προβλέψεις του για την ταξινόμησή τους.

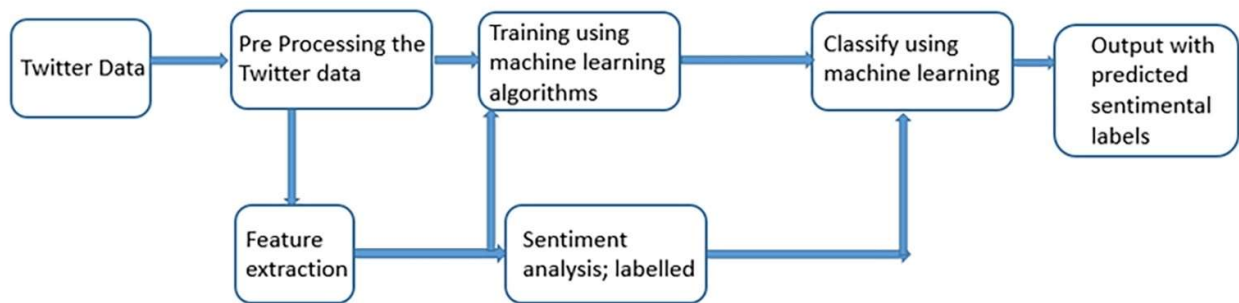
4.3 Συναισθηματική ανάλυση

Η ανάλυση συναισθήματος [8] είναι μία μέθοδος που καθορίζει την άποψη ενός ομιλητή η συντάκτη για κάποιο θέμα ή για τη συνολική πολιτικότητα των συμφραζομένων του σε ένα έγγραφο κειμένου. Η ανάλυση συναισθημάτων ή αλλιώς στην περίπτωση μας η εξόρυξη γνώμης, αναφέρεται στην επεξεργασία φυσικής γλώσσας (NLP) και έχει ως στόχο να αξιοποιήσει μερικές βασικές μεθόδους κατάταξης κειμένων σε κατηγορίες και υπολογιστικής γλωσσολογίας για την εκτίμηση και εξαγωγή αντικειμενικής πληροφορίας από διάφορες πηγές. Η ανάλυση συναισθημάτων συναντάται σε μεγάλη κλίμακα και στα μέσα κοινωνικής δικτύωσης και έχει μια πληθώρα εφαρμογών που σχετίζονται κυρίως με τη διαφήμιση, εξυπηρέτηση πελατών μέχρι και χειραγώγηση της κοινής γνώμης.

Η έκταση του αποσπάσματος λόγου που χαρακτηρίζεται κείμενο δεν σχετίζεται με την απόδοση του χαρακτηρισμού αυτού, άρα μπορεί να υπάρχουν σύντομα κείμενα ή ακόμα και μονολεκτικά. Επίσης, για να χαρακτηριστεί ένα κομμάτι λόγου κείμενο δεν είναι απαραίτητο να περιέχει γραμματικά και συντακτικά πλήρεις προτάσεις που να υπακούουν στους κανόνες ορθογραφίας. Με τον τελευταίο χαρακτηρισμό σκιαγραφείται το είδος κειμένων που συναντώνται σε όλες σχεδόν τις αναρτήσεις των μέσων κοινωνικής δικτύωσης όπως το Twitter.

Αρκετές αναλύσεις συναισθήματος σε κείμενο έχουν αναπτυχθεί τα τελευταία χρόνια από τις οποίες οι πιο απλές κατηγοριοποιούν το κείμενο σε μία συναισθηματική κλίμακα αρνητικού και θετικού βασιζόμενες στο υπό εξέταση κείμενο. Η βασική προσέγγιση βασίζεται σε λεξικό [9] με το οποίο αναλύεται ένα κείμενο όπως ένα tweet για παράδειγμα, βασιζόμενη στις λέξεις που το απαρτίζουν. Το tweet λοιπόν σαρώνεται και γίνεται έλεγχος για το αν περιέχει λέξεις συναισθηματικά φορτισμένες. Τέλος, σε ένα ταξινομητή κειμένου έχει εισαχθεί λεξιλόγιο που περιλαμβάνει αρκετές λέξεις με αρνητική και θετική συναισθηματική φόρτιση όπως τις (μισώ, αγαπώ κλπ.) και οι οποίες καταλαμβάνουν μία συναισθηματική βαθμονόμηση, με βάση αυτό, ολόκληρο το κείμενο ταξινομείται σε μία κλίμακα αρνητικού θετικού.

Βέβαια, υπάρχει μία αντικειμενική δυσκολία στο να υπολογιστεί με μεγάλη ακρίβεια το συναισθηματικό αυτό αποτέλεσμα. Έτσι έχουν αναπτυχθεί ιδιαίτερα τα τελευταία χρόνια αλγόριθμοι με επίβλεψη ή μη μάθησης οι οποίοι χρησιμοποιούνται για την κατηγοριοποίηση κειμένων με βάση το συναισθημά [10] κάποιοι από αυτούς είναι ο αφελής Bayes (Naïvy Bayes) το δέντρο αποφάσεων (Decision tree) και Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines). Για να πραγματοποιηθεί η διαδικασία κατηγοριοποίησης με αυτούς τους αλγόριθμους μηχανικής μάθησης πρέπει πρώτα να τροφοδοτηθούν με λέξεις κειμένων που έχουν συγκεκριμένη συναισθηματική ταυτότητα ώστε να εκπαιδευτούν. Έτσι, ο εκάστοτε ταξινομητής δημιουργεί ένα μοντέλο πιθανοτήτων βασιζόμενος στα δεδομένα που δέχτηκε κατά την εκπαίδευσή του και αποδίδει πρόβλεψη για κάθε νέα είσοδο δεδομένων. Παρακάτω παρατίθεται σχηματικά εικόνα που αναπαριστά την κατηγοριοποίηση κειμένου του Twitter με προσθήκη της παραμέτρου του συναισθήματος στην διαδικασία της εκπαίδευσης του αλγορίθμου μηχανικής μάθησης.



Εικόνα 4.1 Κατηγοριοποίηση tweet με προσθήκη της παραμέτρου του συναισθήματος του γράφοντος [11].

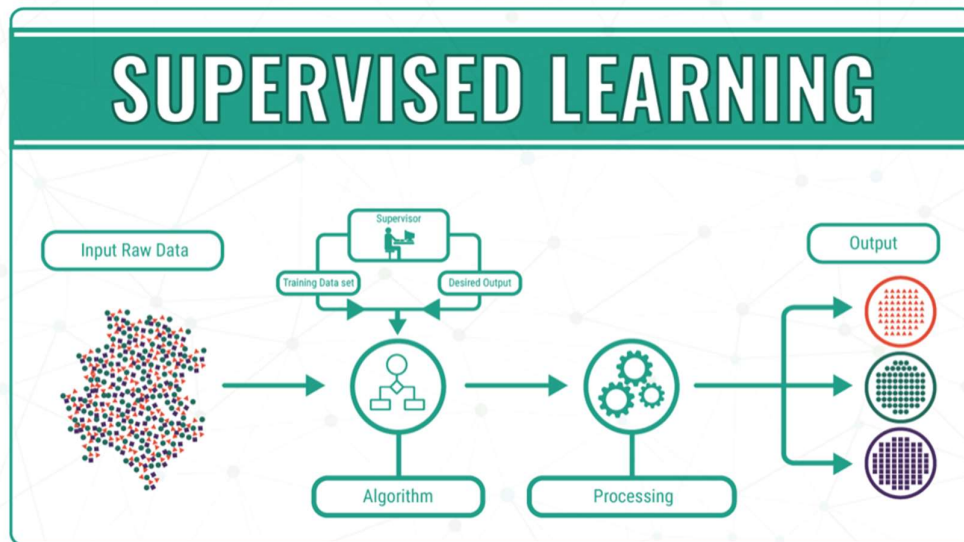
4.4 Αλγόριθμος Naive Bayes

Ο αλγόριθμος Naive Bayes [12, 13, 14] έχει μελετηθεί εκτενώς στο παρελθόν σχεδόν από το 1950. Είχε εισαχθεί με διαφορετική ονομασία αρχικά στην κοινότητα της ανάλυσης κειμένου στις αρχές της δεκαετίας του 1960, αλλά παραμένει μια δημοφιλής πρωταρχική τεχνική για την κατηγοριοποίηση κειμένων, (όπως διαφημίσεις, αθλητικά νέα, πολιτικά, κτλ.), με κύρια γνωρίσματα

τις συχνότητες παρουσίασης των λέξεων. Συνιστά μία εύχρηστη τεχνική για την δημιουργία ταξινομητών, πρότυπων δηλαδή που οριοθετούν ετικέτες κατηγορίας σε οντότητες κειμένων που παρουσιάζονται σαν διανύσματα με τιμές χαρακτηριστικών, όπου οι ετικέτες αυτές καθορίζονται από ένα συγκεκριμένο σύνολο. Δεν είναι ένας μεμονωμένος αλγόριθμος για την εκπαίδευση τέτοιων ταξινομητών, αλλά ανήκει σε μια οικογένεια αλγορίθμων που συναρτώνται σε ένα κοινό πρότυπο.

Όλοι οι Naïve Bayes ταξινομητές εικάζουν ότι η αξία ενός καθορισμένου χαρακτηριστικού είναι ανεξάρτητη και αυτοδιοίκητη από την τιμή οποιουδήποτε άλλου χαρακτηριστικού δεδομένης της μεταβλητής κλάσης. Επι παραδείγματι, ένα φρούτο μπορεί να θεωρηθεί ότι είναι ένα αχλάδι αν είναι πράσινο, στρογγυλό και έχει περίπου 8 cm διάμετρο. Ένας Naïve Bayes ταξινομητής εικάζει πως καθένα από αυτά τα χαρακτηριστικά συμβάλλει ανεξάρτητα στην πιθανότητα ότι το συγκεκριμένο φρούτο είναι ένα αχλάδι, ανεξαρτήτως κάθε πιθανοφανούς αναφοράς μεταξύ χρώματος, σχήματος και χαρακτηριστικών μεγέθους.

Για να εκτελεστεί μία διαδικασία κατηγοριοποίησης κειμένου αρχικά πρέπει να επιλεγεί το διάνυσμα χαρακτηριστικών που θα εξεταστεί, κάτι που αποτελεί το πιο σημαντικό κομμάτι κατά τη διάρκεια και της εκπαίδευσης και της διαδικασίας ταξινόμησης του αλγόριθμου. Αν υποτεθεί πως εξετάζονται tweets για παράδειγμα θα πρέπει όλα τα δεδομένα κειμένου να μετασχηματιστούν σε διανύσματα όρων (λέξεις) $X=(x_1...x_n)$. Συνήθως τα διανύσματα που περιλαμβάνουν τις λέξεις αυτές δημιουργούνται με βάση ένα μοναδικό λεξικό το οποίο παράγεται κατά τη διαδικασία της εκπαίδευσης του αλγορίθμου μηχανικής μάθησης, έτσι είναι δύσκολο να λάβει υπόψη του μία λέξη δύο φορές. Τέλος, το μέγεθος του διανύσματος όρων πρέπει να συμφωνεί με το μέγεθος του λεξιλογίου.



Εικόνα 4.2 Διαδικασία ταξινόμησης Big Data μέσω αλγορίθμου εποπτευόμενης μηχανικής μάθησης[15].

Υπάρχουν δύο βασικοί τύποι του Naive Bayes που χρησιμοποιούνται ο Naive Bayes Bernulli και ο Naive Bayes Multinomial, η κύρια διαφορά τους είναι ο τρόπος που εξάγουν τα χαρακτηριστικά του κειμένου. Ο Naive Bayes Bernulli για παράδειγμα, σε μία πρόταση από ένα tweet θα βαθμονομήσει όλα τα στοιχεία του διανύσματος όρων ίσα με μηδέν και μετά θα ελέγξει κάθε λέξη από το λεξικό του για να δει αν υπάρχει μέσα στο tweet που εξετάζεται, αν βρει κάποια λέξη τότε θα την επισημάνει και θα αλλάξει τη βαθμονόμησή της σε 1. Αν το λεξιλόγιο του εκπαιδευμένου αλγορίθμου είναι αρκετά μεγάλο τότε κάθε tweet που θα εξεταστεί θα παρουσιαστεί με διάνυσμα όρων που θα περιλαμβάνει τους αριθμούς 0 και 1. Στα στοιχεία του διανύσματος όρων του κειμένου δεν εξετάζεται μόνο η παρουσία ή απουσία λέξεων αλλά και η συχνότητα εμφάνισης αυτών. Για παράδειγμα έστω ότι έχουμε ένα κείμενο D και έχουμε C κλάσεις οι οποίες εμπεριέχουν και αυτές κάποιες κλάσεις, τότε για να εξεταστεί σε ποια κλάση ανήκει το κείμενο με βάση το συναίσθημα του υπολογίζεται η δεσμευμένη πιθανότητα $P(C|D)$ που είναι ο βαθμός αλήθειας έχοντας υπολογίσει το D και τότε λαμβάνεται υπόψη η πιθανότητα με το μεγαλύτερο ποσοστό.

$$P(C|D) = \frac{P(D|C)P(C)}{P(D)} \propto P(D|C)P(C)$$

Στην παραπάνω εξίσωση η κύρια πιθανότητα και η πιθανοφάνεια μπορούν να υπολογιστούν από τα δεδομένα χαρακτηριστικά που έχουν λάβει ετικέτα με συγκεκριμένο συναισθημα. Αναλυτικότερα, κάθε κλάση είναι ισόποσα πιθανή, έτσι εύκολα βρίσκουμε την πιθανότητα $P(C_i)$: Υποθέτοντας πως V είναι το λεξιλόγιο και W_j είναι j -οστή λέξη στο λεξιλόγιο από τα δεδομένα εκπαίδευσης του αλγορίθμου υπολογίζεται η πιθανότητα μία λέξη W_j να ανήκει σε C_i κλάσεις συναισθημάτων $P(W_j | C_i)$.

Ένα άλλο παράδειγμα με δεδομένα κειμένου από το Twitter που περιλαμβάνουν πάνω από μία λέξη: Γίνεται η υπόθεση πως S_i είναι i αριθμός tweets αποθηκευμένων με συγκεκριμένη ετικέτα και T_j είναι το διάνυσμα λέξεων που περιλαμβάνει το tweet το οποίο όπως προαναφέρθηκε αποτελείται από στοιχεία που παίρνουν τιμές 0 και 1. Κατά τη διαδικασία εκτέλεσης του αλγορίθμου αν η αντίστοιχη λέξη στο λεξικό υπάρχει και στο tweet S_i τότε η πιθανότητα το tweet να ανήκει στην κλάση C_i είναι σύμφωνα με το θεώρημα του Bayes:

$$(P(S_i) | P(C_i) \propto P(T_j | C_i)) = \prod_{j=1}^{|v|} [T_j P(w_j | C_i) + (1 - T_j)(1 - P(w_j | C_i))]$$

Δηλαδή δεδομένης μια κλάσης S_i , η πιθανότητα παρατήρησης των λέξεων του tweet είναι ίση με το γινόμενο των πιθανοτήτων να παρατηρηθεί κάθε μία λέξη του ξεχωριστά.

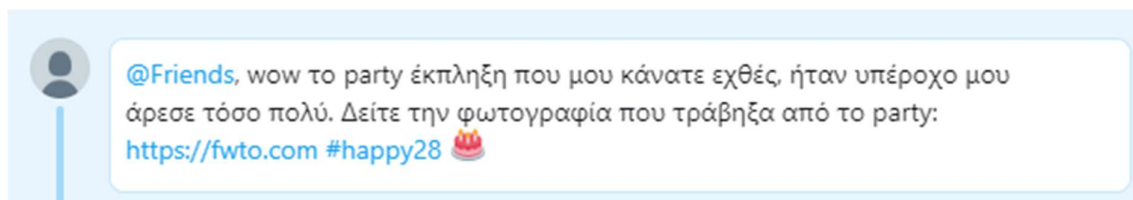
4.5 Twitter και Naive Bayes

Το Twitter επιλέχθηκε ως πηγή των δεδομένων προς αξιολόγηση σε αυτή την εργασία διότι σύμφωνα με το API (Application Programming Interfaces) [16, 17] που διαθέτει και αναφέρθηκε στο 3^ο κεφάλαιο της παρούσας έχει πάνω από 200 εκατομμύρια χρήστες ενεργούς κάθε μήνα οι οποίοι διανέμουν 500 εκατομμύρια «τιτιβίσματα» καθημερινά. Ένας άλλος σημαντικός λόγος είναι ότι η επικοινωνία στο Twitter πραγματοποιείται ως επί το πλείστον με κείμενο φυσικής γλώσσας σε

αντίθεση με άλλες πλατφόρμες όπως το Facebook ή άλλα μέσα κοινωνικής δικτύωσης όπου γίνεται εκτενής χρήση οπτικοακουστικού περιεχομένου όπως video, εικόνα ή κοινή χρήση συνδέσμων ιστοσελίδων. Στο Twitter επιτρέπεται η δημοσίευση κειμένου με περιορισμό στους 140 χαρακτήρες, κάτι που αναγκάζει τους χρήστες να εκφράζονται πιο συγκεκριμένα για κάποιο θέμα και αποδίδοντας περισσότερο συναίσθημα [18] στα λεγόμενά τους. Ωστόσο, το Twitter έχει και κάποια μειονεκτήματα, διότι το κείμενο που παράγεται και δημοσιεύεται εξαιτίας του περιορισμού της έκτασης χαρακτήρων περιέχει πληθώρα σολοικισμών και βαρβαρισμών επίσης, περιέχει πολλές συντομογραφίες και emoticons σε μια προσπάθεια οι χρήστες να είναι όσο πιο σύντομοι και περιεκτικοί μπορούν. Τέλος, σχεδόν σε όλα τα tweets εμπεριέχονται οι χαρακτήρες @ και #, ο πρώτος δηλώνει τη δυνατότητα να γίνει παραπομπή σε κάποιον άλλο χρήστη ενώ ο δεύτερο χρησιμοποιείται για να αποδώσει μια ετικέτα σε κάποιο θέμα. Όλα αυτά δυσχεραίνουν τη διαδικασία ανάλυσης και κατηγοριοποίησης συναισθήματος από τα δεδομένα των κειμένων. Παρακάτω έγινε μια προσπάθεια να αποδοθεί σχηματικά μια διαδικασία αφαίρεσης του περιττού περιεχομένου από τα δεδομένα του παραδείγματος, ώστε να μεταβούν στην μορφή που πρέπει για να είναι επεξεργάσιμα, σύμφωνα με τα προαναφερθέντα.

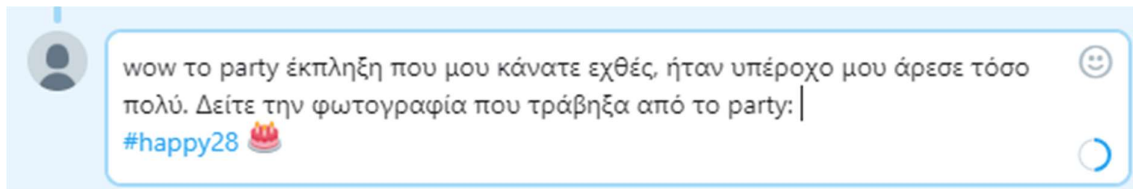
Προπαρασκευή και προετοιμασία tweet για να γίνει η επεξεργασία από αλγόριθμο επιβλεπόμενης μάθησης:

Αρχικό tweet πριν την επεξεργασία



Εικόνα 4.3 Αρχικό tweet.

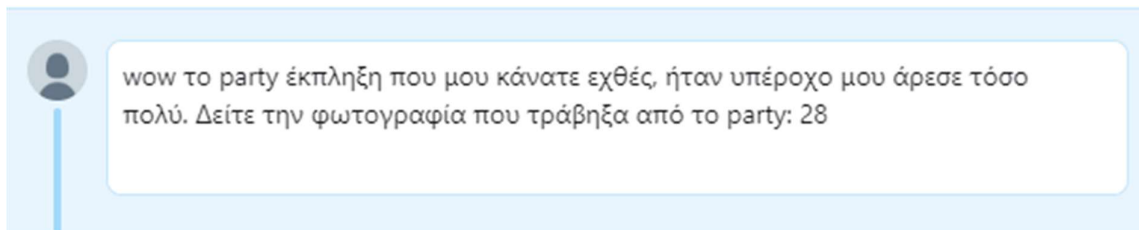
1) Στάδιο αφαίρεση συμβόλου @ και URLs



Εικόνα 4.4 Στάδιο επεξεργασίας 1°.

Στο πρώτο στάδιο αφαιρείται οποιοδήποτε URL και emoticon υπάρχει, το σύμβολο @ καθώς και ό,τι είναι ακριβώς μετά από αυτό διότι είναι άχρηστη πληροφορία.

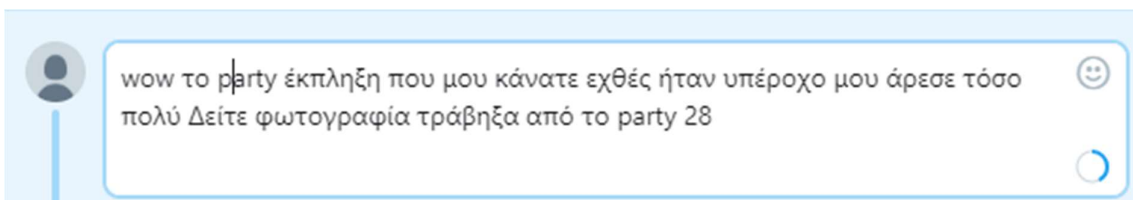
2) Στάδιο Αφαίρεση Hashtag



Εικόνα 4.5 Στάδιο επεξεργασίας 2°.

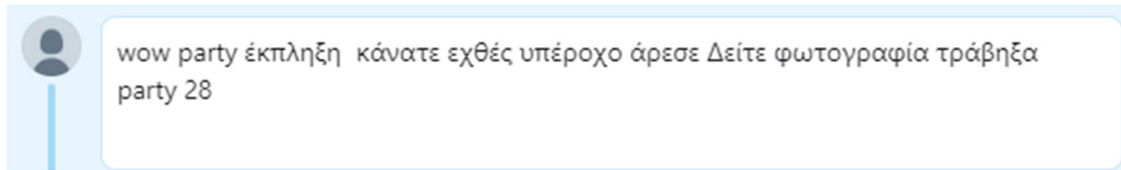
Το σύμβολο # αφαιρείται γιατί προσδίδει μια ετικέτα ή ένα θέμα στο tweet, συνήθως αφορά θέματα που σχετίζονται με κάποια κατάσταση και δεν εκφράζει κάποια άποψη του γράφοντος έτσι δεν είναι κάποια σημαντική πληροφορία επίσης, η λέξη που ακολουθείται δεν αφαιρείται, αλλά αντιμετωπίζεται σαν μια κανονική λέξη μέσα στο κείμενο.

3) Αφαίρεση των σημείων στίξης



Εικόνα 4.6 Στάδιο επεξεργασίας 3°.

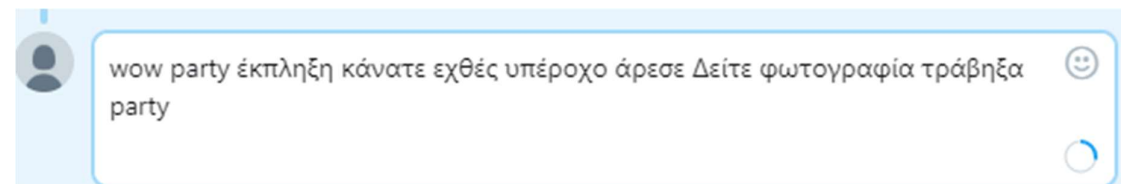
4) Αφαίρεση συνδεδειγμένων λέξεων



Εικόνα 4.7 Στάδιο επεξεργασίας 4°.

Σε αυτό το στάδιο αφαιρούνται συνδεδειγμένες λέξεις, όπως σύνδεσμοι, άρθρα κλπ. Διότι σε μια ανάλυση συναισθήματος δεν είναι χρήσιμα.

5) Αφαίρεση λέξεων που ξέκινούν με κάποιο ψηφίο



Εικόνα 4.8 Στάδιο επεξεργασίας 5°.

Διάφορα ψηφία, όπως η ώρα ή μια χρονολογία, συναντώνται πριν από μια λέξη, αυτά τα ψηφία δεν έχουν κάποια συσχέτιση με συναισθήματα ή συμπεριφορές και αφαιρούνται.

Τέλος, μέσα από αυτή την προεπεξεργασία για το συγκεκριμένο παράδειγμα στην ουσία, μειώνεται ο θόρυβος των δεδομένων και έτσι ο αλγόριθμος επιβλεπόμενης μάθησης είναι δυνατόν να εκπαιδευτεί με μεγαλύτερη ακρίβεια.

4.6 Επίλογος 4^{ου} κεφαλαίου

Αυτό το κεφάλαιο αναφέρεται στον τρόπο που κατηγοριοποιείται ένα κείμενο φυσικής γλώσσας καθώς και η διαδικασία επεξεργασίας του. Η κατηγοριοποίηση αυτή γίνεται με βάση τη συναισθηματική ανάλυση του κειμένου που προέρχεται από δημοσιεύσεις που γίνονται στο Twitter.

Η διαδικασία της κατηγοριοποίησης χρησιμοποιεί έναν αλγόριθμο ο οποίος βασίζεται στη θεωρία του Bayes. Τέλος, δίδεται ένα παράδειγμα όπου αφαιρούνται περιττά στοιχεία που υπάρχουν στα tweets έτσι ώστε να έρθουν στη σωστή κατάσταση ώστε να γίνουν αντιληπτά από την υπολογιστική μηχανή προκειμένου να ταξινομηθούν.

ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ 4^{ου}

- [1] Maniyka, James, et al. (2011), Big Data: The next frontier for innovation, competition and productivity.
- [2] Temple Krystal (2012), What happens in an Internet Minute?, Inside Scoop.
- [3] Gropp William et al (1996), A high-performance, portable implementation of the MPI message passing interface standard, Parallel computing, pp. 789-828.
- [4] Chatzakou, D. , Vakali, A. (2015), Harvesting Opinions and Emotions from Social Media Textual Resources, IEEE Computer Society, pp. 46-50.
- [5] Bengio, Y. (2009), Learning Deep Architectures for AI Foundations and Trends in Machine Learning, 2(1), pp. 1-127.
- [6] George Forman (2004), An Extensive Empirical study of feature selection Metrics for Text Classification, Journal of Machine Learning Research, vol.3.
- [7] Zaharia Matei (2012), Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing, Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation, USENIX Association.
- [8] Paltoglou, G. , Thelwall M. (2012), Twitter, MySpace, Digg: Unsupervised Sentiment Analysis in Social Media, ACM.
- [9] Taboada, Maite, et al. (2011), Lexicon-based methods for sentiment analysis, Computational linguistics 37.2 pp. 267-307.
- [10] Pang, Bo, Lillian Lee, Shivakumar Vaithy Anathan (2002), Thumbs up?: sentiment classification using machine learning techniques, Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, Association for Computational Linguistics.
- [11] <http://scalar.usc.edu/works/c2c-digital-magazine-spring--summer-2017/sentiment-analysis-of-real-time-twitter-data>
- [12] Text Classification and Naïve Bayes, <https://web.stanford.edu/class/cs124/lec/naivebayes.pdf>.

- [13] McCallum, Andrew, Kamal Nigam (1998), A comparison of event models for naïve bayes text classification, AAAI-98 workshop on learning for text categorization, Vol 752.
- [14] Yousef, Ahmed Hassan, Walaa Medhat, Hoda Korashy Mohamed (2014), Sentiment Analysis Algorithms and Applications: A survey.
- [15] <https://bigdata-madesimple.com/machine-learning-explained-understanding-supervised-unsupervised-and-reinforcement-learning/>
- [16] Kim Susanna (2013), Twitter’s IPO Filing Shows 215 Million Monthly Active users, ABC news 3, <http://abcnewsradioonline.com/business-news/twitters-ipo-filing-shows-215-million-monthly-active-users.html>.
- [17] John G. H. , Pat L. (1995), Estimating continuous distributions in Bayesian classifiers, Proceedings of the 11th conference on Uncertainty in artificial intelligence.
- [18] E. Gilbert, K Karahalios (2010), Widespread worry and the stock market, in: Fourth International AAAI Conference on Weblogs and Social Media, Washington DC, pp. 58-65.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Τα συμπεράσματα τα οποία εξάχθηκαν στην παρούσα εργασία είναι ότι τα δεδομένα μεγάλης κλίμακας εδώ και μερικά χρόνια έκαναν την εμφάνισή τους και αποτελούν ένα φαινόμενο που διογκώνεται ολοένα και περισσότερο όσο εξελίσσεται η τεχνολογία και η ηλεκτρονική μας προσυπογραφή. Θεμιτό θα ήταν να χρησιμοποιούνται μόνο σε κοινωφελείς σκοπούς, που στόχο θα έχουν την εξόρυξη γνώμης και συμπερασμάτων για βαρυσήμαντα θέματα που απασχολούν την κοινή γνώμη και όχι για τη χειραγώγηση αυτής. Ένας από τους τομείς στον οποίο μπορούν να παρέχουν σημαντική βοήθεια είναι αυτός της υγείας και της περίθαλψης. Επίσης, συνάγεται το συμπέρασμα πως συνετό θα ήταν εμείς οι ίδιοι να γίνουμε περισσότερο υποψιασμένοι για το τι διαβάζουμε και να οξύνουμε την κριτική μας ικανότητα, ώστε να αντιπαρερχόμαστε τυχόν ψεύδη που συναντάμε καθημερινά στα μέσα κοινωνικής δικτύωσης. Εξάλλου, στην παρούσα εργασία αναφέρθηκαν τρόποι που βοηθούν στον εντοπισμό ψευδών ειδήσεων που στοχεύουν στον αποπροσανατολισμό και στη χειραγώγησης μας.

Η συγκεκριμένη έρευνα έδειξε ότι υπάρχουν πολλοί τρόποι με τους οποίους ένας χρήστης μέσω κοινωνικής δικτύωσης χειραγωγείται διαδικτυακά. Δεδομένα μεγάλης κλίμακας συγκεντρώνονται και γίνονται αντικείμενο επεξεργασίας με την εφαρμογή τεχνολογίας υψηλού επιπέδου, ώστε να αποσπούν τις προσωπικές πληροφορίες των χρηστών και να τις χρησιμοποιούν για να χειραγωγήσουν το περιεχόμενο που μοιράζονται μαζί τους. Έτσι, επηρεάζουν τις απόψεις των χρηστών και εγείρουν συναισθήματα καθώς και γνωστικές επιτυχίες χωρίς την επίγνωση και την άδειά τους. Όσοι εδράζονται στη διαδικασία της χειραγώγησης των μέσων κοινωνικής δικτύωσης δεν αποσκοπούν μόνο σε τομείς, όπως αυτός των οικονομικών, αλλά αποβλέπουν τόσο στο συναισθηματικό όσο και στο πρακτικό πεδίο, κυρίως εξαιτίας της άγνοιας που σημειώνεται από την πλευρά των χρηστών.

Ένα μεγάλο μέρος της άγνοιας των χρηστών μπορεί να οφείλεται στην ελλιπή εκπαίδευσης που τους παρέχεται σε αυτόν τον συγκεκριμένο τομέα, καθώς και στο γεγονός ότι αποφεύγουν να επαληθεύσουν το περιεχόμενο που λαμβάνουν διαδικτυακά και μοιράζονται με άλλους χρήστες.

Παρακάτω παρατίθενται μερικοί τρόποι με τους οποίους οι χρήστες των μέσων κοινωνικής δικτύωσης μπορούν να αποφεύγουν τη διαδικτυακή χειραγώγηση:

1. Εκπαιδεύοντας τους ανθρώπους για τις τακτικές χειραγώγησης και τους τρόπους προστασίας των δεδομένων τους, οι χρήστες πρέπει να καταλάβουν τον τρόπο με τον οποίο μπορούν να περιορίσουν τις πληροφορίες που μοιράζονται στα κοινωνικά μέσα δικτύωσης και αφορούν σε προσωπικά θέματα, έτσι οι διάφορες εταιρείες δεν θα τις αποσπούν με σκοπό να εξατομικεύσουν το περιεχόμενό τους και να επηρεάσουν τις απόψεις των χρηστών.

2. Όλοι οι χρήστες πρέπει να κατανοήσουν πόσο σημαντικό είναι να ελέγχουν πολλές φορές το περιεχόμενο που ακούν, βλέπουν και μοιράζονται σε απευθείας σύνδεση και σε σχέση με την ορθότητά του, εξαιτίας των δεδομένων παραπληροφόρησης που υπάρχουν, διακινούνται και μοιράζονται συνεχώς σε απευθείας σύνδεση από τα μέσα κοινωνικής δικτύωσης, ειδικά όσον αφορά σε πολιτικά ζητήματα.

3. Χρειάζεται περαιτέρω ανάπτυξη και ενίσχυση της προστασίας της ιδιωτικής ζωής στα μέσα κοινωνικής δικτύωσης από νόμους που να προβλέπουν την ακριβή διαδικασία και τα δικαιώματα που θα έχουν διάφορες εταιρείες που χρησιμοποιούν δεδομένα μεγάλης κλίμακας, προερχόμενα από χρήστες οποιασδήποτε ηλεκτρονικής πλατφόρμας.

4. Επιβάλλεται να δημιουργηθούν συστήματα που να εντοπίζουν ψεύτικες δημοσιεύσεις που κοινοποιούνται με τη χρήση Bots καθώς και την ύπαρξη πλαστών Likes σε αυτές.

5. Οι χρήστες θα πρέπει να αναλάβουν την ευθύνη της προστασίας των προσωπικών δεδομένων τους με τη βοήθεια των αντιικών (anti-virus) και άλλων λογισμικών.

6. Οι διαχειριστές των κοινωνικών μέσων δικτύωσης πρέπει να ξεκινήσουν με τη διενέργεια του καθαρισμού των πλαστών προφίλ που διαχειρίζονται διάφορες εταιρείες και βάζουν κατά των πολιτών. Επίσης, πρέπει να γίνεται τακτικός έλεγχος της γνησιότητας των υπόπτων σχολιαστών με προβοκατόριο ύφος trolls ούτως ώστε να προστατεύονται οι πληροφορίες των χρηστών.