



**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ:
«ΔΙΑΧΕΙΡΙΣΗ ΠΛΗΡΟΦΟΡΙΩΝ ΣΕ ΒΙΒΛΙΟΘΗΚΕΣ, ΑΡΧΕΙΑ, ΜΟΥΣΕΙΑ»**

**ΤΜΗΜΑ ΑΡΧΕΙΟΝΟΜΙΑΣ, ΒΙΒΛΙΟΘΗΚΟΝΟΜΙΑΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΗΣΗΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΤΙΚΩΝ, ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

**DEPARTMENT OF ARCHIVAL, LIBRARY AND INFORMATION STUDIES
SCHOOL OF MANAGEMENT, ECONOMICS AND SOCIAL SCIENCES**

Διπλωματική Εργασία

Τίτλος Εργασίας

Μελέτη και σύγκριση αλγορίθμων, μοντέλων και συστημάτων
εντοπισμού ψευδών ειδήσεων

Study and comparison of fake news detection algorithms, models and
systems

Αικατερίνη-Σαπφώ Κωλέτση (ΑΜ: 206682008)

Επιβλέπων: Δημήτριος Κουής

Αθήνα, Δεκέμβριος 2022

Επιτροπή Εξέτασης

1. Δημήτριος Κουής

2. Αλέξανδρος Κουλούρης

3. Ιωάννης Τριανταφύλλου

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Η κάτωθι υπογεγραμμένη Αικατερίνη-Σαπφώ Κωλέτση του Χρήστου, με αριθμό μητρώου 206682008 φοιτήτρια του Προγράμματος Μεταπτυχιακών Σπουδών «Διαχείριση Πληροφοριών σε Βιβλιοθήκες, Αρχεία, Μουσεία» του Τμήματος Αρχειονομίας, Βιβλιοθηκονομίας και Συστημάτων Πληροφόρησης της Σχολής Διοικητικών, Οικονομικών και Κοινωνικών Επιστημών του Πανεπιστημίου Δυτικής Αττικής, δηλώνω ότι:

«Είμαι συγγραφέας αυτής της μεταπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Επιθυμώ την απαγόρευση πρόσβασης στο πλήρες κείμενο της εργασίας μου μέχρι την παράδοσή της και έπειτα από αίτηση μου στη Βιβλιοθήκη και έγκριση του επιβλέποντα καθηγητή.

Η Δηλούσα



Ευχαριστίες – Αφιερώσεις

Ευχαριστώ τον κ. Κουή και την κ. Ευθυμίου για την πολύτιμη βοήθεια τους στην ολοκλήρωση της παρούσας διπλωματικής εργασίας.

7 Δεκεμβρίου 2022

Αικατερίνη-Σ. Κωλέτση

Περίληψη στα ελληνικά

Η ραγδαία ανάπτυξη στις τεχνολογίες πληροφοριών και επικοινωνιών (ΤΠΕ) συνετέλεσε μεταξύ άλλων, στην αύξηση των μέσων διακίνησης της πληροφορίας και παράλληλα στην εμφάνιση ενός νέου φαινομένου, της διασποράς ψευδών ειδήσεων. Η ανησυχητική τους εξάπλωση προκαλεί σοβαρές συνέπειες τόσο στην κοινωνική ζωή μίας χώρας όσο και στους υπόλοιπους τομείς όπως η υγεία, η οικονομία και η πολιτική. Εξαιτίας αυτού του φαινομένου, τα τελευταία χρόνια, οι ειδικοί στα μέσα και την πληροφορία αναζητούν μεθόδους εντοπισμού των ψευδών ειδήσεων. Οι συγκεκριμένες μέθοδοι εστιάζονται στα αλγοριθμικά μοντέλα και κυρίως σε λύσεις που στηρίζονται στη μηχανική μάθηση (machine learning), καθώς και σε άλλα συστήματα. Οι επιστήμονες της πληροφόρησης ως γνώστες της αποτελεσματικής αναζήτησης και της σωστής αξιολόγησης της πληροφορίας μπορούν να βοηθήσουν στον εντοπισμό και την αντιμετώπιση των ψευδών ειδήσεων, συνεπικουρούμενοι από άλλους κλάδους της επιστήμης όπως η επιστήμη των υπολογιστών. Η παρούσα διπλωματική εργασία στοχεύει στην μελέτη αλγοριθμικών μοντέλων και συστημάτων τα οποία αφορούν στον εντοπισμό των ψευδών ειδήσεων, μέσω βιβλιογραφικής ανασκόπησης.

Λέξεις Κλειδιά: (ψευδείς ειδήσεις, παραπληροφόρηση, αλγόριθμοι, μέθοδοι εντοπισμού ψευδών ειδήσεων, συστήματα εντοπισμού ψευδών ειδήσεων)

Περίληψη στα αγγλικά

The rapid development in information and communication technologies (ICT) has contributed, among other things, to the increase in the means of disseminating information. In parallel there is an emergence of a new phenomenon, the spread of fake news. Its alarming spread is results in serious consequences both for the social life of a country and for other areas such as health, the economy and politics. Because of this phenomenon, in recent years, media and information experts have been searching for methods to detect fake news. These methods focus on machine learning, algorithmic models and other systems. Information scientists as experts, in effective search and proper evaluation of information, can help in identifying and countering fake news, assisted by other disciplines such as computer science. This thesis aims to study algorithmic models and systems that address the detection of fake news, through literature review.

Keywords: (fake news, disinformation, algorithms, fake news detection methods, fake news detection systems)

Πίνακας περιεχομένων

ΕΠΙΤΡΟΠΗ ΕΞΕΤΑΣΗΣ	2
ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ	3
ΕΥΧΑΡΙΣΤΙΕΣ – ΑΦΙΕΡΩΣΕΙΣ	4
ΠΕΡΙΛΗΨΗ ΣΤΑ ΕΛΛΗΝΙΚΑ.....	5
ΠΕΡΙΛΗΨΗ ΣΤΑ ΑΓΓΛΙΚΑ	6
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ	7
ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ	9
ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ	14
1.1 ΠΛΑΙΣΙΟ, ΣΚΟΠΟΣ ΚΑΙ ΣΤΟΧΟΙ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ.....	14
1.2 ΕΡΕΥΝΗΤΙΚΕΣ ΥΠΟΘΕΣΕΙΣ	14
1.3 ΜΕΘΟΔΟΛΟΓΙΑ	15
1.4 ΠΕΡΙΟΡΙΣΜΟΙ	15
1.5 ΟΡΓΑΝΩΣΗ ΚΕΦΑΛΑΙΩΝ	15
ΚΕΦΑΛΑΙΟ 2. ΨΕΥΔΕΙΣ ΕΙΔΗΣΕΙΣ - FAKE NEWS: ΘΕΩΡΗΤΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ.....	17
2.1 ΕΡΜΗΝΕΙΑ ΚΑΙ ΕΞΕΛΙΞΗ ΤΟΥ ΦΑΙΝΟΜΕΝΟΥ.....	17
2.2 ΨΕΥΔΕΙΣ ΕΙΔΗΣΕΙΣ, ΠΛΗΡΟΦΟΡΙΑΚΗ ΠΑΙΔΕΙΑ ΚΑΙ ΕΠΙΣΤΗΜΗ ΤΗΣ ΠΛΗΡΟΦΟΡΗΣΗΣ	22
ΚΕΦΑΛΑΙΟ 3. ΜΕΘΟΔΟΛΟΓΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ	25
3.1 ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ	25
3.2 ΠΕΡΙΓΡΑΦΗ ΥΛΟΠΟΙΗΣΗΣ	33
3.2.1 Δειγματοληψία Χιονοστιβάδας (Snowball Sampling)	34
3.2.2 Ανάλυση Περιεχομένου: Ιστορικές και Εννοιολογικές Προσεγγίσεις.....	36
3.2.3 Χαρακτηριστικά Ανάλυσης Περιεχομένου: Θεματική Ανάλυση.....	39
ΚΕΦΑΛΑΙΟ 4. ΑΠΟΤΕΛΕΣΜΑΤΑ: ΚΥΡΙΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΕΝΤΟΠΙΣΜΟΥ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ ...	41
4.1 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΓΝΩΣΗΣ.....	42
4.1.1 Έλεγχος από Ειδικούς (Expert Based)	43
4.1.2 Έλεγχος Βάση Πληθροπορισμού (Crowdsourcing)	60
4.1.3 Αυτοματοποιημένος Έλεγχος (Automatic Fact-Checking)	62
4.2 ΓΛΩΣΣΟΛΟΓΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ	69
4.3 ΣΤΥΛΟΜΕΤΡΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ.....	82

4.3.1	Χαρακτηριστικά Λεξικών (Μεροληψία και Υποκειμενικότητα).....	90
4.3.2	Σημασιολογικά Χαρακτηριστικά.....	92
4.3.3	Χαρακτηριστικά Ιστοτόπου Ειδήσεων (Domain Specific).....	96
4.3.4	Αναπαράσταση Κειμένου (Μη Παρατηρήσιμα Χαρακτηριστικά).....	97
4.4	ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΑΛΓΟΡΙΘΜΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ.....	109
4.4.1	Αλγόριθμοι Επιβλεπόμενης Μάθησης - Παραδοσιακοί (Supervised Learning).....	110
4.4.2	Αλγόριθμοι Πολλαπλής Εκμάθησης (Ensemble Learning).....	121
4.4.3	Αλγόριθμοι Μη Επιβλεπόμενης Μηχανικής Μάθησης (Unsupervised Learning).....	127
4.4.4	Μεταεμπειρικοί Αλγόριθμοι-Εξελικτικοί (Metaheuristics).....	131
4.4.5	Αλγόριθμοι Βαθιάς Μάθησης (Deep Learning).....	136
4.5	ΘΕΜΑΤΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ.....	162
4.6	ΣΥΝΑΙΣΘΗΜΑΤΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ.....	163
4.7	ΥΒΡΙΔΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ.....	169
4.8	ΟΠΤΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ.....	180
4.9	ΚΟΙΝΩΝΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ.....	185
ΚΕΦΑΛΑΙΟ 5. ΣΥΖΗΤΗΣΗ-ΣΥΜΠΕΡΑΣΜΑΤΑ.....		204
ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ.....		208
ΠΑΡΑΡΤΗΜΑ Α. ΠΙΝΑΚΑΣ ΑΚΡΩΝΥΜΙΩΝ.....		265
ΠΑΡΑΡΤΗΜΑ Β. ΠΙΝΑΚΑΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΕΝΤΟΠΙΣΜΟΥ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ.....		269
ΠΑΡΑΡΤΗΜΑ Γ. ΑΠΟΤΕΛΕΣΜΑΤΑ WEB OF SCIENCE ΚΑΙ SCOPUS.....		270
ΠΑΡΑΡΤΗΜΑ Δ. ΣΥΜΠΛΗΡΩΜΑΤΙΚΕΣ ΠΗΓΕΣ ΜΕ ΤΗ ΜΕΘΟΔΟ SNOWBALL.....		271

Πίνακας Σχημάτων

Εικόνα 1. Στιγμιότυπο εικόνας από ιστοσελίδα Huffington Post	18
Εικόνα 2. Ψευδής πληροφόρηση για πυροβολισμούς σε σχολείο του Μεξικού	18
Εικόνα 3. IFLA infographic. How to spot fake news.	23
Εικόνα 4. Τεχνικές ανάλυσης δεδομένων παραπληροφόρησης και ψευδών ειδήσεων.	24
Εικόνα 5. Εξαγωγή αποτελεσμάτων Scopus.	26
Εικόνα 6. Εξαγωγή αποτελεσμάτων WoS.	26
Εικόνα 7. Αντιστοίχιση WoS και Scopus.....	27
Εικόνα 8. Αντιστοίχιση WoS και Scopus.....	27
Εικόνα 9. Ενοποίηση δεδομένων WoS και Scopus query 1.	28
Εικόνα 10. Ενοποίηση δεδομένων WoS και Scopus query 2.	28
Εικόνα 11. Κοινοί τίτλοι query 3.....	29
Εικόνα 12. Κοινοί τίτλοι ερωτημάτων WoS και Scopus.	29
Εικόνα 13. Διπλοεγγραφές από Scopus.	29
Εικόνα 14. Συγχώνευση εγγραφών query 1.....	30
Εικόνα 15. Συγχώνευση εγγραφών query 2.....	30
Εικόνα 16. Συγχώνευση εγγραφών query 3.....	31
Εικόνα 17. Συγχώνευση εγγραφών query 4.....	31
Εικόνα 18. Αποτελέσματα query 1.....	32
Εικόνα 19. Αποτελέσματα query 2.....	32
Εικόνα 20. Αποτελέσματα query 3.....	33
Εικόνα 21. Αποτελέσματα query 4.....	33
Εικόνα 22. Στιγμιότυπο από MediaBias FactCheck.....	44
Εικόνα 23. Μετρικές και κίτρινος δείκτης επισήμανσης πολιτικής τοποθέτησης.....	44
Εικόνα 24. Στιγμιότυπο από Politifact.....	46
Εικόνα 25. Στιγμιότυπο από Snopes.	49
Εικόνα 26. Στιγμιότυπο από Newsguard.....	50
Εικόνα 27. Στιγμιότυπο από Maldito Bulo.	54
Εικόνα 28. Στιγμιότυπο από SNUFactCheck.....	56
Εικόνα 29. Στιγμιότυπο από Ellinika Hoaxes.	56
Εικόνα 30. Χαρακτηρισμοί σφραγίδες στα Ellinika Hoaxes.	60
Εικόνα 31. Στιγμιότυπο από μοντέλο πληθοπορισμού.	62
Εικόνα 32. Στιγμιότυπο από Beautiful Soup library.	65

Εικόνα 33. Στιγμιότυπο από Rake Library (εξαγωγή λέξεων-κλειδιών).....	66
Εικόνα 34. DBpedia Spotlight.	66
Εικόνα 35. Μοντέλο W3C Prov.....	68
Εικόνα 36. Μοντέλο Komadu.	68
Εικόνα 37. NELA Toolkit.....	72
Εικόνα 38. Κεφαλαία γράμματα σε τίτλο.....	82
Εικόνα 39. Στιγμιότυπο ελληνικού παραπλανητικού άρθρου.....	83
Εικόνα 40. Μετρική Cosine Similarity.....	84
Εικόνα 41. Παράδειγμα αναγνώρισης πρότασης από αλγόριθμο.	87
Εικόνα 42. Διαδικασία εντοπισμού hoax.....	88
Εικόνα 43. Pos Tagging.....	89
Εικόνα 44. Google Perspective API.....	93
Εικόνα 45. Στιγμιότυπο από αναγνώριση ονομαστικών οντοτήτων.	94
Εικόνα 46. Στιγμιότυπο εξαγωγής ονομαστικών οντοτήτων (NER).	94
Εικόνα 47. Αναπαράσταση Bag of words (TriFN framework).	98
Εικόνα 48. Στιγμιότυπο μετρήσεων TF-IDF.....	99
Εικόνα 49. Αρχιτεκτονική μοντέλου CBoW.....	101
Εικόνα 50. Μοντέλο Skipgram.....	102
Εικόνα 51. Αρχιτεκτονική Fast Text.....	102
Εικόνα 52. Υπομοντέλα Fast Text.....	103
Εικόνα 53. Μοντέλο Doc2Vec.....	103
Εικόνα 54. Μετατροπή συνδέσμων στο μοντέλο Link2vec.....	104
Εικόνα 55. Θεωρία ρητορικής δομής.....	105
Εικόνα 56. Θεωρία ρητορικής δομής-RST Webtool.....	106
Εικόνα 57. Παράδειγμα γραμμικής αναπαράστασης Glove.....	106
Εικόνα 58. Αρχιτεκτονική Μοντέλου USE.....	108
Εικόνα 59. Μοντέλο BERT (προεκπαίδευση και προσαρμογή).	109
Εικόνα 60. Γραμμική παρουσίαση χαρακτηριστικών (δημοσιεύσεων, σχολίων, αλληλεπιδράσεων).....	112
Εικόνα 61. Στιγμιότυπο από SVM.....	114
Εικόνα 62. Στιγμιότυπο Linear kernel SVM.....	114
Εικόνα 63. Στιγμιότυπο από SVM-RBF.....	115
Εικόνα 64. Στιγμιότυπο από δέντρο απόφασης.....	116
Εικόνα 65. Δέντρο απόφασης J-48.....	117

Εικόνα 66. Αλγόριθμος CART.	117
Εικόνα 67. Στιγμιότυπο από Gaussian Naive Bayes.	118
Εικόνα 68. Μαθηματικός τύπος Bernoulli Naive Bayes.	119
Εικόνα 69. Δίκτυο Bayes.	120
Εικόνα 70. Αλγόριθμος Stochastic Gradient Descent (SGD).	120
Εικόνα 71. Δάσος Τυχαίας Απόφασης.	122
Εικόνα 72. Αλγόριθμος XGBoost.	123
Εικόνα 73. Παράδειγμα λειτουργίας Gradient Boosting.	124
Εικόνα 74. Αλγόριθμος Gradient Boosting.	124
Εικόνα 75. Στιγμιότυπο από Adaboost.	125
Εικόνα 76. Ensemble Voting.	126
Εικόνα 77. Γραμμική ανάλυση διάκρισης (LDA) και γραμμική ανάλυση τετραγωνικής διακρίσεως (QDA).	128
Εικόνα 78. Μη αρνητική παραγοντοποίηση μήτρας (NMF).	128
Εικόνα 79. Αλγόριθμος K means (οπτικοποίηση).	129
Εικόνα 80. Στιγμιότυπο από αλγόριθμο ομοιότητας (similarity algorithm).	129
Εικόνα 81. Παράδειγμα διαφορών λέξεων.	130
Εικόνα 82. Αλγόριθμος Smith-Waterman.	131
Εικόνα 83. Παράδειγμα μοτίβων συχνότητας Smith Waterman.	131
Εικόνα 84. Αναβάθμιση θέσης γκριζου λύκου βάση θέσης θηράματος (Grey wolf optim. algorithm).	132
Εικόνα 85. Αναβάθμιση θέσης πράκτορα βάση υποψήφια λύσεων α , β , δ	132
Εικόνα 86. Επίθεση και εύρεση θηράματος από αλγόριθμο.	133
Εικόνα 87. Σάλπα.	133
Εικόνα 88. Αλγόριθμος Salp Swarm Optimization.	134
Εικόνα 89. Μετατροπή του αλγόριθμου σμήνους σε δυαδικό.	135
Εικόνα 90. Αλγόριθμος Whale Optimization.	136
Εικόνα 91. FNDEEP.	137
Εικόνα 92. Νευρωνικό δίκτυο.	138
Εικόνα 93. HPA-BLSTM.	139
Εικόνα 94. Hierarchical Attention Network (HAN).	140
Εικόνα 95. 3HAN.	141
Εικόνα 96. Dense Neural Network.	142
Εικόνα 97. Νευρωνικό δίκτυο βαθιάς μάθησης (Deep Neural Network).	143

Εικόνα 98. Αλγόριθμος δικτύου RvNN.....	143
Εικόνα 99. Top-Down, Bottom-Up RvNN.....	144
Εικόνα 100. Αλγόριθμος δικτύου TRNN.....	145
Εικόνα 101. Αλγόριθμος δικτύου RNN.....	146
Εικόνα 102. Αλγόριθμος δικτύου BiRNN.....	146
Εικόνα 103. Αλγόριθμος δικτύου BiRNN.....	147
Εικόνα 104. Αλγόριθμος δικτύου Vanilla RNN.....	147
Εικόνα 105. Αλγόριθμος δικτύου GRU.....	149
Εικόνα 106. Αλγόριθμος δικτύου διπλής κατεύθυνσης (BiGRU).....	149
Εικόνα 107. Μηχανισμός προσοχής GRU.....	149
Εικόνα 108. Αλγόριθμος δικτύου LSTM.....	150
Εικόνα 109. LSTM με μηχανισμό προσοχής.....	151
Εικόνα 110. Αλγόριθμος δικτύου BiLSTM.....	152
Εικόνα 111. Αλγόριθμος δικτύου MLP.....	153
Εικόνα 112. Αλγόριθμος δικτύου FFNN.....	154
Εικόνα 113. Αλγόριθμος δικτύου CNN.....	155
Εικόνα 114. Συνελκτικός αλγόριθμος δικτύου Text-CNN.....	156
Εικόνα 115. Αλγόριθμος δικτύου TCNN-URG.....	156
Εικόνα 116. Συνελκτικός αλγόριθμος γράφου κειμένου TEXT-GCN.....	157
Εικόνα 117. Αλγόριθμος δικτύου προσοχής Graph Attention Network (GAT).....	158
Εικόνα 118. Αλγόριθμος δικτύου κάψουλας (CapsNet).....	159
Εικόνα 119. Αλγόριθμος Multitask Learning.....	160
Εικόνα 120. Μοντέλο FNDNET.....	161
Εικόνα 121. Μοντέλο DEFEND.....	162
Εικόνα 122. Θεματικό Μοντέλο (Biterm Topic Model).....	163
Εικόνα 123. Αντιθέσεις συναισθήματος (Polarity)- TextBlob.....	165
Εικόνα 124. Βαθμός αντίθεσης συναισθήματος (Polarity).....	165
Εικόνα 125. Μετρήσεις λέξεων επιρροής-λεξικό NRC Affect.....	167
Εικόνα 126. Μετρήσεις λεξικού NRC-Vader.....	168
Εικόνα 127. Πολυμοντελικός κωδικοποιητής.....	170
Εικόνα 128. Μοντέλο SAFE.....	171
Εικόνα 129. Μοντέλο ασαφούς λογικής (fuzzy logic).....	171
Εικόνα 130. Κλίμακα μεσαίου επιπέδου.....	172
Εικόνα 131. Μετρικές βαρών για κάθε εναλλακτική.....	173

Εικόνα 132. Μοντέλο CSI.....	174
Εικόνα 133. Πολυμοντελικό συνελκτικό δίκτυο γράφου γνώσης (KMAGCN).	176
Εικόνα 134. Τοπικό πολυμοντελικό θεματικό δίκτυο μνήμης (MTMN).	177
Εικόνα 135. Πολυμοντελικό δίκτυο μηχανισμού προσοχής AttRNN.....	177
Εικόνα 136. Μοντέλο EANN.	178
Εικόνα 137. Μοντέλο Spotfake.	179
Εικόνα 138. Μοντέλο Spotfake+.	179
Εικόνα 139. Μοντέλο Hybrid CNN.....	180
Εικόνα 140. Νευρωνικό δίκτυο VGG-19.....	181
Εικόνα 141. ResNet 50.	182
Εικόνα 142. Πλατφόρμα InVid.	184
Εικόνα 143. Αλγόριθμος εντοπισμού πηγής (παρατηρητές).	193
Εικόνα 144. Αλγόριθμος εντοπισμού πηγής (εντοπιστές).	194
Εικόνα 145. Υψηλή κεντρικότητα δεσμών σε δίκτυο.....	195
Εικόνα 146. Παράδειγμα κεντρικότητας.....	195
Εικόνα 147. Παράδειγμα οντολογίας PROV-O.....	196
Εικόνα 148. Αλγόριθμος Hits.....	197
Εικόνα 149. Εμφάνιση μηνυμάτων μέσω μοντέλου Sardana.....	198
Εικόνα 150. Μοντέλο Ghenai για τον εντοπισμό των φημών.	200

Κεφάλαιο 1. Εισαγωγή

Η ραγδαία εξάπλωση των ψευδών ειδήσεων και η ανησυχητική διασπορά τους στα μέσα κοινωνικής δικτύωσης έχει δυσμενείς επιπτώσεις σε όλους τους τομείς της ζωής, όπως υγεία, πολιτική, οικονομία. Για την αντιμετώπιση του φαινομένου, την τελευταία πενταετία έχουν προταθεί εναλλακτικές προσεγγίσεις για τον εντοπισμό των ψευδών ειδήσεων, που αφορούν είτε σε τεχνικές μηχανικής μάθησης (machine learning), είτε σε ταξινομικούς αλγορίθμους (classification algorithms), είτε σε διάφορα μοντέλα και συστήματα.

Οι επιστήμονες της πληροφόρησης, καθώς είναι ήδη εκπαιδευμένοι στην αναζήτηση και στη σωστή χρήση της πληροφορίας, αλλά και στην αξιολόγηση της εγκυρότητάς της, θα πρέπει να συνεργαστούν με επιστήμονες από άλλους τομείς, όπως της πληροφορικής, ώστε να βοηθήσουν στη δημιουργία εργαλείων εντοπισμού της παραπληροφόρησης. Η επισκόπηση της τρέχουσας βιβλιογραφίας πάνω στο θέμα των αλγορίθμων, των μοντέλων και συστημάτων εντοπισμού ψευδών ειδήσεων, θα μπορούσε να αποτελέσει μία βάση για τη θεωρητική γνωριμία των επιστημόνων της πληροφόρησης σχετικά με το συγκεκριμένο φαινόμενο.

1.1 Πλαίσιο, σκοπός και στόχοι της διπλωματικής εργασίας

Στόχος της παρούσας διπλωματικής εργασίας είναι να διερευνήσει:

- α) τα κύρια χαρακτηριστικά των αλγορίθμων, των μοντέλων και των συστημάτων που χρησιμοποιούνται για τον εντοπισμό των ψευδών ειδήσεων,
- β) τις ομοιότητες και τις διαφοροποιήσεις τους, που προκύπτουν από την παρούσα μελέτη όπως είναι τα κοινά χαρακτηριστικά τους, οι τρόποι λειτουργίας τους, τα είδη συνόλου δεδομένων που χρησιμοποιούν, κλπ.

Απώτερος σκοπός της εργασίας είναι να αποτελέσει έναν ευσύνοπτο οδηγό εξοικείωσης των επιστημόνων της πληροφόρησης και λοιπών ενδιαφερόμενων με το πεδίο ανάπτυξης αλγορίθμων και συστημάτων εντοπισμού ψευδών ειδήσεων.

1.2 Ερευνητικές υποθέσεις

Μέσα από την επισκόπηση της τρέχουσας διεθνούς βιβλιογραφίας για τα έτη 2019-2021 τα ερωτήματα που στοχεύει να απαντήσει η συγκεκριμένη διπλωματική εργασία είναι:

- α) Ποια είναι τα κύρια χαρακτηριστικά των αλγορίθμων, μοντέλων και συστημάτων εντοπισμού των ψευδών ειδήσεων;

- β) Ποιες είναι οι ομοιότητες και ποιες οι διαφορές που παρουσιάζουν;

1.3 Μεθοδολογία

Ο εντοπισμός των πηγών έγινε από τις μηχανές αναζήτησης [Scopus](#) και [Web of Science](#) ενώ για τον ίδιο σκοπό ακολουθήθηκε και η μέθοδος snowball (Heckathorn, 2011). Για την τελική επιλογή ή απόρριψη των πηγών εφαρμόστηκαν τα κριτήρια της συνάφειάς τους με το θέμα της εργασίας και της πληρότητας της πληροφορίας που αυτές εμπεριέχουν. Μετά τη συγκέντρωση των πηγών, η μέθοδος που ακολουθήθηκε είχε ως βάση την ανάλυση περιεχομένου (content analysis), η οποία, σύμφωνα με τον Weber (Weber, 1990, p. 117) αφορά τη χρήση ενός συνόλου κανόνων (διαδικασιών) προκειμένου να εξαχθούν έγκυρα συμπεράσματα από ένα κείμενο.

Πιο συγκεκριμένα, κατά την ανάλυση του περιεχομένου των επιλεγθέντων πηγών έγινε προσπάθεια να εντοπιστούν οι επιμέρους κοινές θεματικές κατηγορίες στις οποίες αναφέρονται οι σχετικές πηγές, όπως για παράδειγμα τα χαρακτηριστικά, ο τρόπος λειτουργίας, τα σύνολα δεδομένων που χρησιμοποιούνται κ.ά. Στη συνέχεια, έγινε συγκριτική παρουσίαση κάθε αλγορίθμου και συστήματος που εξετάστηκε, με βάση τις ανωτέρω κατηγορίες. Η κατηγοριοποίηση της βιβλιογραφίας παρουσιάζεται σε ειδικά διαμορφωμένο πίνακα του προγράμματος MS Excel.

1.4 Περιορισμοί

Στην συγκεκριμένη διπλωματική εργασία η βιβλιογραφική επισκόπηση και η επιλογή των πηγών εστιάστηκε στα έτη 2019 έως 2021 αλλά και στη συνάφεια με το θέμα της διπλωματικής. Πηγές που ανακτήθηκαν μετά τις σχετικές αναζητήσεις αλλά κρίθηκε ότι είχαν ελλιπή πληροφόρηση ή μη σχετική συνάφεια με το θέμα αποκλείστηκαν αν και είχαν υψηλές βιβλιογραφικές παραπομπές.

1.5 Οργάνωση Κεφαλαίων

Στο δεύτερο κεφάλαιο παρουσιάζεται η θεωρητική προσέγγιση των ψευδών ειδήσεων, με σύντομη βιβλιογραφική επισκόπηση των τελευταίων εξελίξεων πάνω στη διασπορά του φαινομένου. Στο τρίτο κεφάλαιο αναλύεται εκτενώς η μεθοδολογία που ακολουθήθηκε στην παρούσα διπλωματική εργασία και ειδικότερα ο τρόπος συλλογής των δεδομένων με τα ερωτήματα που διατυπώθηκαν κατά τη διάρκεια της έρευνας. Στο τέταρτο κεφάλαιο παρουσιάζονται τα αποτελέσματα της έρευνας θεματικά. Το πέμπτο και τελευταίο

κεφάλαιο κλείνει με τη συζήτηση και τα συμπεράσματα της έρευνας καθώς και με πιθανές μελλοντικές της προεκτάσεις.

Κεφάλαιο 2. Ψευδείς ειδήσεις - Fake news:

Θεωρητικές προσεγγίσεις

Στο συγκεκριμένο κεφάλαιο γίνεται μία πρώτη προσπάθεια ερμηνείας του φαινομένου. Στην πρώτη υποενότητα παρουσιάζονται οι εννοιολογικές προσεγγίσεις καθώς και η εξέλιξη των ψευδών ειδήσεων και στη δεύτερη η συμβολή της επιστήμης της πληροφόρησης πάνω στο συγκεκριμένο φαινόμενο.

2.1 Ερμηνεία και εξέλιξη του φαινομένου

Στη σημερινή εποχή κατά την οποία οι Τεχνολογίες Πληροφοριών και Επικοινωνιών (Τ.Π.Ε.) έχουν εξελιχθεί αρκετά η σωστή πληροφόρηση είναι επιβεβλημένη για κάθε ενημερωμένο πολίτη.

Με την εμφάνιση του διαδικτύου και περαιτέρω την εξέλιξη των κοινωνικών μέσων δικτύωσης, αλλά και τη μαζική, γρήγορη πρόσβαση στην πληροφορία, οι πολίτες αντιμετωπίζουν δυσκολίες τόσο στην προσπάθειά τους να εντοπίσουν τη σωστή πληροφορία ανάλογα με την πληροφοριακή τους ανάγκη, όσο και στο να ελέγξουν την αξιοπιστία της.

Στην αναφορά του World Economic Forum του 2013 διατυπώθηκε και επισήμως το μέγεθος του συγκεκριμένου προβλήματος, κυρίως ως προς το ζήτημα της παραπληροφόρησης. Η συγκεκριμένη αναφορά παρουσίαζε ως παραδείγματα τους ανώνυμους λογαριασμούς στο Twitter που ανέφεραν ψευδώς την πληροφορία για πλημμύρα στο κτίριο του Χρηματιστηρίου της Νέας Υόρκης, λόγω του τυφώνα Sandy (2012). Επιπρόσθετα παρουσίαζε και τις ψεύτικες διαδόσεις για πυροβολισμούς σε σχολείο του Μεξικού, γεγονός που είχε ως συνέπεια την απομάκρυνση των παιδιών από το σχολείο (*Digital Wildfires in a Hyperconnected World*, n.d.) (Εικόνα 1 και Εικόνα 2).

Twitter rumor of flooded stock-exchange among a surge of fake storm reports

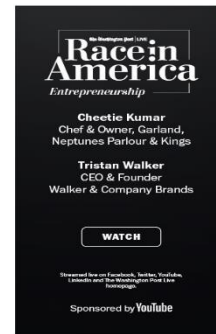


This ominous shot of high seas surrounding a wave-battered Statue of Liberty turned out to be production art from the disaster movie "The Day After Tomorrow." (Twitter image)

By Paul Farhi
October 30, 2012

In the Twitter age, bad news travels fast. Even when the bad news isn't really news at all.

Such was the case at the peak of the storm's wrath Monday night as rumor and fallacy swirled like autumn leaves. Around 8 p.m., as Sandy was belting New Jersey and New York City, a tweet appeared: "BREAKING: Confirmed flooding on NYSE. The trading floor is flooded under more than 3 feet of water."



Εικόνα 1. Στιγμιότυπο εικόνας από ιστοσελίδα Huffington Post



Εικόνα 2. Ψευδής πληροφόρηση για πυροβολισμούς σε σχολείο του Μεξικού

Η έννοια της παραπληροφόρησης και των ψευδών ειδήσεων έως το 2016 δεν είχαν αποτυπωθεί σε μεγάλο βαθμό σε έρευνες, ούτε είχε δοθεί ιδιαίτερη βαρύτητα ως προς την επικινδυνότητα τους από την επιστημονική κοινότητα. Οι Khan et al (Khan et al., 2022), αναφέρουν ενδεικτικά ότι οι βιβλιογραφικές έρευνες για τις ψευδείς ειδήσεις πριν από το 2016 εστίαζονταν κυρίως σε σάτιρα, προπαγάνδα, παρωδίες, διαφήμιση και δημόσιες σχέσεις, φωτογραφική χειραγώγηση, και κατασκευασμένες δημοσιεύσεις.

Χαρακτηριστικά παραδείγματα ανησυχητικής επιρροής των ψευδών ειδήσεων ήταν οι αμερικάνικες εκλογές του 2016 και η εμφάνιση του νέου κορονοϊού SARS-COVID19 στις αρχές του 2020. Και στις δύο περιπτώσεις η παραπληροφόρηση και η ταχύτατη διάδοση τους διαδικτυακά οδήγησε στην πρώτη περίπτωση, σε χειραγώγηση του εκλογικού σώματος, ασκώντας επιρροή ως προς το αποτέλεσμα, και στη δεύτερη περίπτωση, σε κίνδυνο για την υγεία των πολιτών.

Σύμφωνα με τους Gunther et al (Gunther et al., 2018), οι ψευδείς ειδήσεις επηρέασαν το αποτέλεσμα των εκλογών κυρίως ως προς τη συμπεριφορά των ψηφοφόρων. Την ίδια οπτική συμμερίζονται και οι Guo και Vargo (L. Guo & Vargo, 2020) με ποιοτική έρευνα και αξιολόγηση άρθρων ιστοσελίδων για τις αμερικάνικες εκλογές που υποστήριζαν τον Ντόναλντ Τραμπ και την Χίλαρι Κλίντον.

Άλλες βιβλιογραφικές έρευνες αναφέρουν την επέλαση του φαινομένου των ψευδών ειδήσεων στα κοινωνικά δίκτυα, τα οποία ενίσχυσαν τη διάδοσή τους, δεδομένου της πληθώρας των πληροφοριών που κοινοποιούν και μοιράζονται οι χρήστες στους προσωπικούς τους λογαριασμούς. Οι Georgakopoulos & Mores (Georgakopoulos & Mores, 2020) ανέφεραν στην έρευνα τους ότι από τις ψευδείς ειδήσεις που εμφανίστηκαν τρεις μήνες πριν τις εκλογές, οι 30 εκατομμύρια διαμοιρασμένες δημοσιεύσεις ήταν υπέρ του Ντόναλντ Τραμπ (Facebook) και οι 7 εκατομμύρια υπέρ της Χίλαρι Κλίντον. Αντίστοιχες έρευνες πραγματοποιήθηκαν και από τους Bovet και Makse με έμφαση στις ψευδείς ειδήσεις που εμφανίστηκαν κατά τη διάρκεια των εκλογών στο twitter (Bovet & Makse, 2019) αλλά και τη δυναμική διάδοσή τους στο συγκεκριμένο κοινωνικό μέσο από τους Ackland και Gwynn (Ackland & Gwynn, 2020).

Το φαινόμενο των ψευδών ειδήσεων στα κοινωνικά δίκτυα εντάθηκε ακόμα περισσότερο με την εμφάνιση της πανδημίας του κορονοϊού και τη διάδοση θεωριών συνωμοσίας ή εσκεμμένης παραπληροφόρησης. Οι Tagliabue et al (Tagliabue et al., 2020), επισημαίνουν και την έλλειψη επαρκούς πληροφόρησης των ίδιων των επιστημόνων ως προς το νέο ιό, τις διαφορετικές προσεγγίσεις ως προς την επικινδυνότητα του ιού αλλά και τα σχόλια μη ειδικών επιστημόνων (π.χ. γαστρεντερολόγοι). Η συγκεχυμένη πληροφόρηση αλλά και η έκταση της διάδοσης των ψευδών ειδήσεων στα μέσα κοινωνικής δικτύωσης ώθησαν ορισμένους ασθενείς να μη θέλουν να τους χορηγηθεί φαρμακευτική αγωγή με το φόβο ότι θα έχουν αυξημένες πιθανότητες να κολλήσουν τον ιό (Tagliabue et al., 2020).

Αντίστοιχη μελέτη για την παραπληροφόρηση σχετικά με τον Covid19 έγινε και από τους Naeem et al (Naeem et al., 2021) όπου αναλύθηκαν ψευδείς ειδήσεις από “fact-checkers” και

“myth-busters”, δηλαδή από επαγγελματίες ελεγκτές επαλήθευσης ή διάψευσης ορθότητας ειδήσεων.

Η Sciortino, επιπροσθέτως, αναφέρει ότι η αύξηση της παραπληροφόρησης στον τομέα της υγείας εμφάνισε ένα νεολογισμό το «infodemia»¹. Από τη μία πλευρά η ποικιλία πληροφοριών στο διαδίκτυο συντελεί σε έναν πλούσιο διάλογο αλλά από την άλλη η μεγάλη ποσότητα δημιουργεί, σε πολλές περιπτώσεις, ένα είδος «ψηφιακού κελιού πληροφόρησης» για κάθε χρήστη, όπου η διοχέτευση των ειδήσεων και των δεδομένων είναι πολλές φορές ελεγχόμενη. Ένα αντιπροσωπευτικό παράδειγμα θεωρείται η μηχανή αναζήτησης της Google, όπου οι χρήστες κατευθύνονται στο πληροφοριακό περιεχόμενο με βάση τις προτάσεις και τα αποτελέσματα του σχετικού αλγορίθμου (Sciortino, 2020).

Εκτός όμως από την παραπληροφόρηση στα μέσα κοινωνικών δικτύων, η εμφάνιση θεωριών συνωμοσίας που σχετίζονται με τον ιό υπό μορφή βίντεο συνετέλεσαν σε περαιτέρω διαστρέβλωση της ήδη δύσκολης κατάστασης. Αξίζει να αναφερθεί, ενδεικτικά, το αφηγηματικό βίντεο 26 λεπτών, το οποίο έγινε viral το Μάιο του 2020 στην Αμερική, με τον τίτλο «Plandemic» (Home, n.d.) και υιοθετήθηκε από το αντιεμβολιαστικό κίνημα σύμφωνα με τους New York Times (Frenkel et al., 2020). Το συγκεκριμένο βίντεο αναρτήθηκε στα κοινωνικά μέσα Facebook, YouTube και Vimeo και ανέφερε ψευδώς ότι ο συγκεκριμένος ιός και τα εμβόλια χρησιμοποιήθηκαν από μια μικρή ελίτ για να αποκομίσει κέρδη. Μέσα σε μία εβδομάδα είχε πάνω από οκτώ εκατομμύρια θεάσεις στα μέσα κοινωνικής δικτύωσης².

Οι παράγοντες που οδήγησαν στην εξάπλωση των ψευδών ειδήσεων δεν οφείλονται όμως τόσο στην ταχύτητα διάδοσης που επιτρέπει η τεχνολογία, όσο στην πληθώρα διαφορετικών ερμηνευτικών προσεγγίσεων που έχουν διατυπωθεί στη βιβλιογραφική έρευνα.

Σύμφωνα με τους Leonie και Althuis (Leonie & Althuis, 2018) τα **fake news** δεν αναφέρονται αποκλειστικά σε παραπλανητικά πρωτοσέλιδα ειδήσεων, εσφαλμένες πληροφορίες ή «κίτρινο τύπο», αλλά και σε **σάτιρα, ψέματα, προπαγάνδα, παραπληροφόρηση (disinformation), κακή πληροφόρηση (misinformation)**. Συγκεκριμένα:

- Η πολιτική **σάτιρα** αναφέρεται σε προγράμματα διασκέδασης που εμπλέκουν και την πολιτική κατάσταση.
- Τα **ψέματα (bullshit)** αναφέρονται ως «μισές αλήθειες» και παρεξηγήσεις.

¹ infodemia = ο όρος αναφέρεται ως υπερβολική πληροφόρηση πάνω σε ένα πρόβλημα του οποίου η λύση γίνεται πολύ πιο δύσκολη. Ο όρος εμφανίστηκε στην περίοδο της πανδημίας CoVid-19 για να δείξει την αδυναμία των πολιτών ως προς τον διαχωρισμό πηγών πραγματικής πληροφόρησης και παραπληροφόρησης.

² Παραγωγός ήταν μία μη διακεκριμένη επιστήμονας η Judy Mikovits, που δήλωσε ότι η έρευνά της για τα εμβόλια δεν είχε τύχει αναγνωρισιμότητας.

- Η **προπαγάνδα**, σε αντίθεση με τις προηγούμενες έννοιες, δεν αγνοεί την αλήθεια, διαμορφώνει όμως συνειδήσεις χρησιμοποιώντας ένα μέρος της αλήθειας προκειμένου να επιτύχει συγκεκριμένες αντιδράσεις από το κοινό και να ωφεληθεί περαιτέρω ο προπαγανδιστής.
- Η **παραπληροφόρηση** αφορά τον σκοπούμενο έλεγχο και διάδοση ψευδούς πληροφόρησης με απώτερο στόχο την παραπλάνηση και την εξαπάτηση του κοινού ενώ η **κακή πληροφόρηση** είναι αποτέλεσμα είτε άγνοιας είτε λάθους (Leonie & Althuis, 2018).

Η πολιτική σάτιρα έχει αποδειχθεί ότι επηρεάζει τις απόψεις του κοινού, τα πιστεύω και τις αντιλήψεις τους γιατί διαμορφώνει το πως οι άνθρωποι επεξεργάζονται την πολιτική πληροφόρηση (Leonie & Althuis, 2018).

Για τους Mourão και Robertson (Mourão & Robertson, 2019) οι ψευδείς ειδήσεις είναι μια μίξη πολλών διαφορετικών και πολιτικά σχετικών πληροφοριών όπως ο **συναισθηματισμός (sensationalism)**, οι ειδήσεις, η κακή πληροφόρηση, τα **προκλητικά πρωτοσέλιδα (clickbait)** και η **μεροληπτική προπαγάνδα (partisan propaganda)**. Ο **συναισθηματισμός** για τους συγγραφείς είναι η απλοποιημένη πληροφόρηση, κυρίως σε πρωτοσέλιδα σχετικά με ένα πολύπλοκο θέμα η οποία προκαλεί εσκεμμένα συναίσθημα και σοκ. Σύμφωνα με τη Watson (Watson, 2018) το πρώτο χαρακτηριστικό παράδειγμα ψευδούς συναισθηματικής πληροφόρησης καταγράφηκε το 1475 όταν δολοφονήθηκαν 15 άτομα εβραϊκής καταγωγής από φανατικά μέλη της εβραϊκής κοινότητας για το βασανισμό ενός μωρού χριστιανικής καταγωγής στο Τρεντ της Ιταλίας.

Το **clickbait**, αντιθέτως, είναι ένα είδος πρωτοσέλιδης πληροφόρησης η οποία δημιουργείται προκειμένου να παροτρύνει τον αναγνώστη να επιλέξει έναν σύνδεσμο (link) για να συνεχίζει να διαβάζει το δημοσίευμα γνωστό και ως forward referencing. Η συγκεκριμένη τακτική παρουσιάζει αποσπασματικές ή ελλιπείς πληροφορίες στον αναγνώστη, ερωτήματα και άλλες γλωσσολογικές στρατηγικές (Mourão & Robertson, 2019).

Στην ίδια οπτική ο Lim (Lim, 2020) αναφέρει ως ψευδή είδηση την «εκ προθέσεως παραπλανητική και προκατειλημμένη πληροφόρηση με σκοπό το όφελος του αποστολέα, το οποίο περιλαμβάνει τη λάθος πληροφόρηση, είτε με τη μίξη, είτε όχι περισσότερων στοιχείων παραλειπόμενης αλλά σημαντικής πληροφόρησης, ενός αφαιρετικού πλαισίου, παραπλανητικών τίτλων ή πληροφοριών που προκαλούν αίσθηση δηλαδή προκλητικών clickbait».

Ενώ οι Ackland και Gwynn στο βιβλίο *The Psychology of Fake News accepting, sharing, and correcting misinformation* (Ackland & Gwynn, 2020, p. 28) προσθέτουν ότι οι ψευδείς

ειδήσεις είναι ένα «είδος κακής πληροφόρησης όπου η πληροφορία συνδέεται με ένα γεγονός και κακή προαίρεση περισσότερο από το άτομο που τη δημιουργεί παρά από το άτομο που τη διαδίδει». Ως παραδείγματα κακής πληροφόρησης θεωρούνται τα κακόβουλα μηνύματα που αξιοποιούν τη σάτιρα, οι **φήμες (rumours)** που ακόμα δεν έχουν επιβεβαιωθεί αν είναι αληθείς ή εσφαλμένες, οι **θεωρίες συνωμοσίας (conspiracy theories)** που είναι μη αξιόπιστες αλλά διαδίδονται από ανθρώπους που πιστεύουν ότι είναι αληθείς και τα **hoaxes** (απάτες) που είναι κακοπροαίρετες ή χιουμοριστικές πληροφορίες και παραπέμπουν σε έγκυρες πηγές πληροφοριών (Greifeneder et al., 2020).

Η παρούσα διπλωματική εργασία εστιάζεται στην έρευνα των ψευδών ειδήσεων υιοθετώντας τις έννοιες της παραπληροφόρησης, της κακής πληροφόρησης, της προπαγάνδας και της απάτης (hoaxes).

2.2 Ψευδείς ειδήσεις, πληροφοριακή παιδεία και επιστήμη της πληροφόρησης

Η εξάπλωση των ψευδών ειδήσεων και της παραπληροφόρησης δε θα μπορούσε να αφήσει αμέτοχους τους επιστήμονες της πληροφόρησης. Η [IFLA](#) έχει δημοσιεύσει αρκετές ανακοινώσεις ήδη από το 2014 οι οποίες επικεντρώνονται στο επιστημονικό αντικείμενο της πληροφοριακής παιδείας. Το 2017 εξαιτίας της εξάπλωσης των ψευδών ειδήσεων η IFLA σε αντίστοιχη ανακοίνωση επισήμανε την ανάγκη αναβάθμισης των δεξιοτήτων των επιστημόνων πληροφόρησης αλλά και τη δράση που πρέπει να αναλάβουν τα κράτη για την ανάπτυξη των δεξιοτήτων της «πληροφοριακής παιδείας» των πολιτών με έμφαση σε αυτό το φαινόμενο (*Libraries and Fake News – Paper for Zagreb Roundtable on Freedom of Access to Information – IFLA*, n.d.). Ο όρος «πληροφοριακή παιδεία» σύμφωνα με την IFLA είναι «η ικανότητα του πολίτη να ελέγχει τα ψηφιακά εργαλεία χρησιμοποιώντας την τεχνολογία ηθικά και αποτελεσματικά» (*Libraries and Fake News – Paper for Zagreb Roundtable on Freedom of Access to Information – IFLA*, n.d.). Για την αντιμετώπιση του φαινομένου, η IFLA δημιούργησε το παρακάτω πληροφοριακό γράφημα (Εικόνα 3) (Institutions (IFLA), 2017).



Εικόνα 3. IFLA infographic. How to spot fake news.

Στην αναγκαιότητα προώθησης και ανάπτυξης των δεξιοτήτων που υποστηρίζει η πληροφοριακή παιδεία δίνουν έμφαση και σχετικές βιβλιογραφικές έρευνες που έχουν γίνει τα τελευταία χρόνια. Σε μελέτη των De Paor και Heravi (De Paor & Heravi, 2020) παρουσιάστηκαν διαφορετικές πρωτοβουλίες πληροφοριακής παιδείας αλλά και σχέσης των ψευδών ειδήσεων με την επιστήμη της πληροφόρησης. Στην ίδια οπτική το άρθρο των Faix και Fynn (Faix & Fynn, 2020) αναφέρει το πλαίσιο για την πληροφοριακή παιδεία που εφάρμοσε η Ένωση Κολεγίων και Ερευνητικών Βιβλιοθηκών (ACRL) για να αντιμετωπίσει το φαινόμενο της παραπληροφόρησης. Το συγκεκριμένο πλαίσιο καλεί τους επιστήμονες της πληροφόρησης να διδάσκουν τόσο την αξιολόγηση της πηγής ως ερευνητική τακτική όσο και την κριτική σκέψη.

Επιπρόσθετα, οι Young et al (Young et al., 2021), διατυπώνουν μέσω διερευνητικών συνεντεύξεων με τους ίδιους τους βιβλιοθηκονόμους της Ουάσιγκτον την ανάγκη για περαιτέρω βιβλιογραφική έρευνα και εργαλεία που πρέπει να αναπτύξουν οι επιστήμονες της πληροφόρησης για να αποτρέψουν τη διασπορά των ψευδών ειδήσεων.

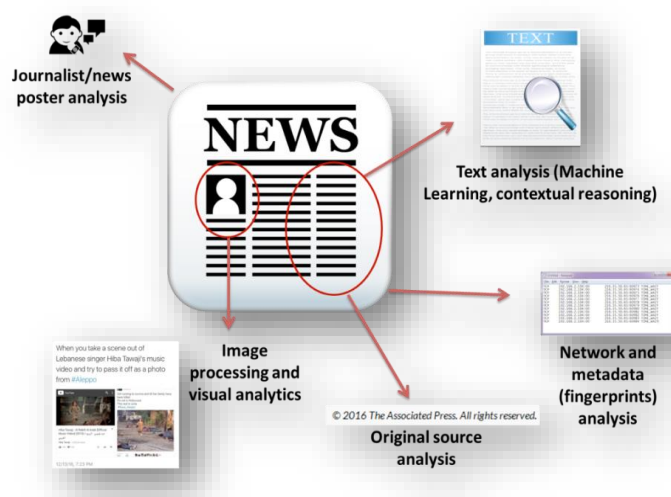
Παράλληλα η Bangani (Bangani, 2021) αναφέρει σε ποσοτική και ποιοτική έρευνά της, τις προσπάθειες των ακαδημαϊκών βιβλιοθηκών της Β. Αφρικής να συνεισφέρουν σε ασφαλή, έγκυρη και ποιοτική πληροφόρηση για τον Covid 19, μέσω της πληροφοριακής παιδείας (Information Literacy), της επαγγελματικής επιμόρφωσης (σεμινάρια, συνέδρια) και των οδηγιών βιβλιοθήκης (Library guides). Οι συγκεκριμένοι οδηγοί παρείχαν πρόσβαση σε πλούσιο υλικό για τον ιό, όπως σε ιστοσελίδες και περιοδικά και είχαν αρκετά μεγάλη απήχηση.

Οι προσπάθειες όμως που αναφέρθηκαν δεν αρκούν για τον εντοπισμό των ψευδών ειδήσεων, δεδομένης της ταχύτατης εξάπλωσής τους από τα μέσα κοινωνικών δικτύων. Οι επιστήμονες της πληροφόρησης οφείλουν να υιοθετήσουν πρακτικές και από άλλους επιστημονικούς κλάδους όπως είναι η πληροφορική και η επιστήμη των δεδομένων (Data Science).

Οι βιβλιογραφικές έρευνες των τελευταίων χρόνων που έχουν μελετήσει το φαινόμενο των ψευδών ειδήσεων έχουν επικεντρωθεί σε μεθόδους μηχανικής μάθησης (τεχνητής νοημοσύνης), αλγορίθμων, μοντέλων και συστημάτων τα οποία βοηθούν σε περαιτέρω αντιμετώπιση του εν λόγω φαινομένου.

Σε άρθρο τους οι Michael Choras και άλλοι συγγραφείς αναφέρουν τις ακόλουθες τεχνικές εντοπισμού των ψευδών ειδήσεων (Choras et al., 2019):

- Συλλογή και ευρετηρίαση πληροφοριών που δημοσιεύονται διαδικτυακά και διασταύρωση με προηγούμενες ειδήσεις
- Προσδιορισμός αξιοπιστίας της πληροφοριακής πηγής και του δημιουργού της όπως ιστοσελίδα παρόχου, άτομο, φήμη περιεχομένου
- Σύγκριση πανομοιότυπων ειδήσεων από διαφορετικές πληροφοριακές πηγές
- Τεχνικές Μηχανικής Μάθησης (Machine Learning) για ανάλυση χαρακτηριστικών περιεχομένου
- Σημασιολογική ανάλυση μέσω εφαρμοσμένων οντολογιών (Εικόνα 4).



Εικόνα 4. Τεχνικές ανάλυσης δεδομένων παραπληροφόρησης και ψευδών ειδήσεων.

Κεφάλαιο 3. Μεθοδολογικές Προσεγγίσεις

Στο συγκεκριμένο κεφάλαιο παρουσιάζεται η μεθοδολογία που ακολουθήθηκε για την εκπόνηση της παρούσας διπλωματικής εργασίας. Στην πρώτη υποενότητα περιγράφεται ο τρόπος συλλογής και διαχείρισης των δεδομένων - καθώς και η παρουσίασή τους μέσω γραφημάτων. Στη δεύτερη υποενότητα διευκρινίζεται ο τρόπος μελέτης και ανάλυσης του περιεχομένου των πηγών που τελικά επιλέχθηκαν, βάση της οποίας προέκυψαν τα αποτελέσματα που καταγράφονται στο κεφάλαιο 4.

3.1 Συλλογή Δεδομένων

Κατά το πρώτο στάδιο της μεθοδολογίας έγινε ο εντοπισμός των πηγών από τις βάσεις δεδομένων της [Scopus](#) και [Web of Science](#) με διατύπωση 4 ερωτημάτων (queries) σε κάθε βάση και με περιορισμούς στα έτη 2019-2021. Τα queries τα οποία διατυπώθηκαν είναι τα εξής **query 1:** (TITLE-ABS-KEY(fake news) AND TITLE-ABS-KEY(algorithms)) AND (LIMIT-TO (PUBYEAR,2021) OR LIMIT-TO (PUBYEAR,2020) OR LIMIT-TO (PUBYEAR,2019)), **query 2:** (TITLE-ABS-KEY (disinformation) AND TITLE-ABS-KEY (algorithms)) AND (LIMIT-TO (PUBYEAR,2021) OR LIMIT-TO (PUBYEAR,2020) OR LIMIT-TO (PUBYEAR,2019)), **query 3:** (TITLE-ABS-KEY (fake AND news) AND TITLE-ABS-KEY (machine AND learning)), **query 4:** (TITLE-ABS-KEY(misinformation) AND TITLE-ABS-KEY(algorithms)) AND (LIMIT-TO (PUBYEAR,2021) OR LIMIT-TO (PUBYEAR,2020) OR LIMIT-TO (PUBYEAR,2019)).

Στη συνέχεια πραγματοποιήθηκε εξαγωγή των δεδομένων και των δύο βάσεων σε csv μορφή με όλα τα περιγραφικά στοιχεία όπως συγγραφέας, τίτλος, έτος δημοσίευσης, βιβλιογραφικές παραπομπές και η καταχώρησή τους σε φύλλο εργασίας (Εικόνα 5,Εικόνα 6)

Authors	Author(s) ID	Title
Sheikhi S.	57211397395;	An effective fake news detection method using WOA-xg
Rassameeroj I., Wu S.F.	26656159500;7407184507;	How do fake news propagators exploit social algorithms
Nagashri K., Sangeetha J.	57226306416;57226303013;	Fake News Detection Using Passive-Aggressive Classifier
Borges P.M., Gamarato R.R.	36544928100;56039359500;	The role of beliefs and behavior on Facebook: A semioti
Shah P., Kobzi Z.	57219283411;6508054194;	Multimodal fake news detection using a Cultural Algori
Ozbay F.A., Alatas B.	57211061747;10341297500;	Adaptive Salp swarm optimization algorithms with inert
Taskin S.G., Kucukille E.U., Topal K.	57282836600;1803773100;57192558295;	Detection of Turkish Fake News in Twitter with Machine
Indrapu S.R.K., Komalla J., Inugala D.R., Reddy Kota G., Sanam A.	57192954930;57226834430;57226834834;57226830528;57226838873;	Comparative analysis of machine learning algorithms to
Huang Y.-F., Chen P.-H.	7501570675;57215427418;	Fake news detection using an ensemble learning model
Nikam S.S., Dalvi R.	57220901648;7004708520;	Machine learning algorithm based model for classificati
Albahr A., Albahr M.	57211668089;57192877385;	An empirical comparison of fake news detection using d
Utmarsh, Sujit, Azeez S.N., Darshan B.C., Chaya Kumari H.A.	57218830862;57211316267;57211324540;57218831552;57212155510;	A study on discernment of fake news using machine lea
Kesarwani A., Chauhan S.S., Nair A.R., Verma G.	57204905282;56288143900;57215633700;57191328307;	Supervised Machine Learning Algorithms for Fake News
Kalvoda B., Stoick B., Snell N., Straub J.	5721121192;57210173433;57208207889;55011324400;	Evaluation of algorithms for fake news identification
Ozbay F.A., Alatas B.	57211061747;10341297500;	Fake news detection within online social media usin s
Abbad M., Kumar G., Samiullah M., Kumar N.S.	6602141840;57202043743;55615773200;55726007800;	A Predominant Advent to Fake News Detection using M
Zimmer F., Scheibe K., Stock M., Stock W.G.	57195064273;57197705506;1189330400;8658221400;	Fake news in social media: Bad algorithms or biased use
Pernagallo G., Torrisi B., Bennato D.	57208691407;6507593700;56366274100;	A Classification Algorithm to Recognize Fake News Webl
Ahmad F., Lokeshkumar R.	57213481713;56407190300;	A comparison of machine learning algorithms in fake ne
Abdulrahman A., Baykara M.	57217834365;36164970800;	Fake News Detection Using Machine Learning and Deep
Selva Birunda S., Kanniga Devi R.	57222050019;49361367000;	A Novel Score-Based Multi-Source Fake News Detection
Bali A.P.S., Fernandes M., Choubey S., Goel M.	57210285262;57210290419;57210283428;57213013121;	Comparative Performance of Machine Learning Algorith
Bharath G., Manikanta K.J., Prakash G.B., Sumathi R., Chinrasamy P.	57223126996;57223131766;57209690968;5722498664;57201503238;	Detecting Fake News Using Machine Learning Algorithm
Mugdha S.B.S., Ferdous S.M., Fahmin A.	57223025052;57223031605;57201688863;	Evaluating Machine Learning Algorithms for Bengali Faki
Kumar A., Singh S., Kaur G.	57211291284;7407874009;57220565935;	Fake news detection of Indian and United States electio
Srinivasa Reddy S., Akhila M., Hinduja S., Hari Durga Reddy R., Vijaya Lakshmi B.	57221332676;57216368815;57216368839;5721637191;57216368074;	Detecting fake news using nlp and machine learning clac
Abdullah-All-Tamvir, Mahir E.M., Akhter S., Huiq M.R.	57215366790;57215388074;5722955850;24328410900;	Detecting Fake News using Machine Learning and Deep
Ozbay F.A., Alatas B.	57211061747;10341297500;	A novel approach for detection of fake news on social m
Ramachandran G., Nemeth D., Neville D., Zhelezov D., Yalcin A., Fohrmann O., Krishnamachari B.	5560680300;57221465728;57221465334;57221458827;57221465074;57210956830;7004879601	WhistleBlower: Towards A Decentralized and Open Plat
Girgis S., Amer E., Gadallah M.	57207884062;54891346400;14022668500;	Deep Learning Algorithms for Detecting Fake News in O
Tiwari V., Lennon R.G., Dowling T.	57219442784;24070832200;57219435908;	Not Everything You Read is True! Fake News Detectio
Dou Y., Shu X., Xia C., Yu P.S., Sun L.	57217744880;57192389892;5720120537;7402366049;57193994946;	User Preference-aware Fake News Detection
Qazi M., Khan M.U.S., Ali M.	57216784075;5631242100;55760540800;	Detection of Fake News Using Transformer Model
Goldani M.H., Momtazi S., Safabakhsh R.	57209101826;24479191100;3677156800;	Detecting fake news with capsule neural networks

Εικόνα 5. Εξαγωγή αποτελεσμάτων Scopus.

No	Publication	Authors	Author Full Names	Article Title	Source Title	Language	Document Co
1	J	Song, CG; Shu, K;	Song, Chenguang; Sh	Temporally evolving graph neural network for fake news detection	INFORMAT	English	Article
2	J	Taskin, SG; Kucuk Taskin, Suleyman	Taskin, SG; Kucuk Taskin, Suleyman	Detection of Turkish Fake News in Twitter with Machine Learning	ARABIAN J	English	Article; Earl
3	J	Ramkissoon, AN; M Ramkissoon, Amit Ne	Ramkissoon, AN; M Ramkissoon, Amit Ne	Determining an Optimal Data Classification Model for Credibility-Ba	REVIEW O	English	Article; Earl
4	J	Huang, GKW; Lee, Huang, Gerald Ki Wei	Huang, GKW; Lee, Huang, Gerald Ki Wei	Hyperpartisan News Classification with ELMO and Bias Feature	JOURNAL	English	Article
5	J	Sheikhi, S	Sheikhi, S	An effective fake news detection method using WOA-xgbTree algo	APPLIED S	English	Article
6	J	Aisaawy, Y; Alkhood Aisaawy, Yazed; Alkh	Aisaawy, Y; Alkhood Aisaawy, Yazed; Alkh	Lightweight Chain for Detection of Rumors and Fake News in Soci	INTERNAT	English	Article
7	J	Ilias, L; Roussaki, I Ilias, Loukas; Roussa	Ilias, L; Roussaki, I Ilias, Loukas; Roussa	Detecting malicious activity in Twitter using deep learning techniqu	APPLIED S	English	Article
8	J	Meel, P; Vishwakar Meel, Priyanka; Vishw	Meel, P; Vishwakar Meel, Priyanka; Vishw	Image captioning, and forensics ensemble multimodal fake n	INFORMAT	English	Article
9	J	Qian, SS; Hu, J; Fa Qian, Shengsheng; H	Qian, SS; Hu, J; Fa Qian, Shengsheng; H	Knowledge-aware Multi-modal Adaptive Graph Convolutional Netw	ACM TRAN	English	Article
10	J	Hansrajh, A; Adelyl Hansrajh, Arvin; Adeli	Hansrajh, A; Adelyl Hansrajh, Arvin; Adeli	Detection of Online Fake News Using Blending Ensemble Learning	SCIENTIFI	English	Article
11	J	Crane, H; Xu, M	Crane, Harry; Xu, Min	Inference on the history of a randomly growing tree	JOURNAL	English	Article
12	J	Vinolin, V; Sucharit Vinolin, V.; Sucharita	Vinolin, V; Sucharit Vinolin, V.; Sucharita	Taylor-RNet: An approach for image forgery detection using Taylor-	INTERNAT	English	Article
13	J	Zervopoulos, A; Alv Zervopoulos, Alexand	Zervopoulos, A; Alv Zervopoulos, Alexand	Deep learning for fake news detection on Twitter regarding the 201	NEURAL C	English	Article; Earl
14	J	Xing, YF; Wang, Xl Xing, Yunfei; Wang, X	Xing, YF; Wang, Xl Xing, Yunfei; Wang, X	Ranking Influential Nodes of Fake News Spreading on Mobile Soci	JOURNAL	English	Article
15	J	Mazzeo, V; Rapisai Mazzeo, Valeria; Rapi	Mazzeo, V; Rapisai Mazzeo, Valeria; Rapi	Detection of Fake News on COVID-19 on Web Search Engines	FRONTIER	English	Article
16	J	George, Y; Karunas George, Yasmeen; Ke	George, Y; Karunas George, Yasmeen; Ke	Real-time spatio-temporal event detection on geotagged social me	JOURNAL	English	Article
17	J	Li, X; Lu, PX; Hu, L Li, Xin; Lu, Peixin; Hu	Li, X; Lu, PX; Hu, L Li, Xin; Lu, Peixin; Hu	A novel self-learning semi-supervised deep learning network to det	MULTIMED	English	Article; Earl
18	J	Al-Ahmad, B; Al-Zo Al-Ahmad, Bilal; Al-Zc	Al-Ahmad, B; Al-Zo Al-Ahmad, Bilal; Al-Zc	An Evolutionary Fake News Detection Method for COVID-19 Pand	SYMMETR	English	Article
19	J	Lin, YP	Lin, Yu-Peng	Media Literacy in an age of Artificial Intelligence: An exploratory res	TAIWAN JC	Chinese	Article
20	J	Madani, Y; Erritali, Madani, Youness; Err	Madani, Y; Erritali, Madani, Youness; Err	Using artificial intelligence techniques for detecting Covid-19 epid	RESULTS I	English	Article
21	J	Nassar, SS; Farage Nassar, Sabry S.; Far	Nassar, SS; Farage Nassar, Sabry S.; Far	Reliable Mark-Embedded Algorithm for Verifying Archived/Encrypte	WIRELESS	English	Article
22	J	Ozbay, FA; Alatas, Ozbay, Feyza Altunbe	Ozbay, FA; Alatas, Ozbay, Feyza Altunbe	Adaptive Salp swarm optimization algorithms with inertia weights fr	MULTIMED	English	Article; Earl
23	J	Pennycook, G; Rar Pennycook, Gordon; I	Pennycook, G; Rar Pennycook, Gordon; I	The Psychology of Fake News	TRENDS IN	English	Review
24	J	Phan, RCW; Low, Y Phan, Raphael C-W; I	Phan, RCW; Low, Y Phan, Raphael C-W; I	Strengthening speech content authentication against tampering	SPEECH C	English	Article
25	J	Baptista, JP; Gradil Baptista, Joao Pedro; Brave	Baptista, JP; Gradil Baptista, Joao Pedro; Brave	New World of Fake News: How it Works	JAVNZOST-	English	Review; Ea
26	J	Pham, CV; Pham, I Pham, Canh V.; Pham	Pham, CV; Pham, I Pham, Canh V.; Pham	Minimum budget for misinformation detection in online social netw	OPTIMIZAT	English	Article; Earl
27	J	Borrelli, C; Bestagil Borrelli, Clara; Bestag	Borrelli, C; Bestagil Borrelli, Clara; Bestag	Synthetic speech detection through short-term and long-term predi	EURASIP J	English	Article
28	J	Silva, CVM; Fontes Meneses Silva, Caio; Int	Silva, CVM; Fontes Meneses Silva, Caio; Int	Intelligent Fake News Detection: A Systematic Mapping	JOURNAL	English	Article
29	J	Dinesh, T; Endran, Dinesh, T.; Endran, T.	Dinesh, T; Endran, Dinesh, T.; Endran, T.	Higher Classification of Fake Political News Using Decision Tree A	REVISTA G	English	Article
30	J	Thaher, T; Saheb, M Thaher, Thae; Saheb	Thaher, T; Saheb, M Thaher, Thae; Saheb	Intelligent Detection of False Information in Arabic Tweets Utilizing	SYMMETR	English	Article
31	J	Turowski, K; Sreed Turowski, Krzysztof; E	Turowski, K; Sreed Turowski, Krzysztof; E	Temporal Ordered Clustering in Dynamic Networks: Unsupervised	IEEE TRAN	English	Article
32	J	Rahman, MS; Hald Rahman, Md Shafiqur;	Rahman, MS; Hald Rahman, Md Shafiqur;	An efficient hybrid system for anomaly detection in social networks	CYBERSEC	English	Article
33	J	Clarke, J; Chen, Hl Clarke, Jonathan; Chl	Clarke, J; Chen, Hl Clarke, Jonathan; Chl	Fake News, Investor Attention, and Market Reaction	INFORMAT	English	Article
34	J	Goldani, MH; Momi Goldani, Mohammad	Goldani, MH; Momi Goldani, Mohammad	Detecting fake news with capsule neural networks	APPLIED S	English	Article

Εικόνα 6. Εξαγωγή αποτελεσμάτων WoS.

Κατά το δεύτερο στάδιο έγινε απαλοιφή των περιττών στηλών και ενοποιήθηκαν για κάθε query τα αποτελέσματα. Στο επόμενο βήμα τα δεδομένα των δύο βάσεων (WoS, Scopus) τέθηκαν σε αντιστοιχία (Εικόνα 7, Εικόνα 8).

Databases	N	Publicat	Authors	Book Editors	Author Full Name	Article Title	S
WoS	1	J	De Aguiar, LA; Rodrigues, CM		De Aguiar, Leonel	Expertise in Journalism: Considerations about professional authority in the con-text of disinformation driven C	
WoS	2	J	Reisach, U		Reisach, Ulrike	The responsibility of social media in times of societal and political manipulation E	
WoS	3	J	Grolauschen, A		Grolauschen, Arke	Virtuality and Literacy in the 21st Century Z	
WoS	4	J	Johnson, KL; Walsh, JL; Amirkhania, YA; Borkowski, J.		Johnson, Kara Lay	Using a novel genetic algorithm to assess peer influence on willingness to use pre-exposure prophylaxis in A	
WoS	5	J	De Blasio, E; Selva, D		De Blasio, Emiliano	Who is Responsible for Disinformation? European Approaches to Social Platforms' Accountability in the Pos A	
WoS	6	J	Robertson, R; Meitjes, C		Robertson, Randy	Towards an Online Risk Mitigation Framework for Political Brands Subject to Computational Propaganda C	
WoS	7	J	Sued, GE		Elisa Sued, Gabriele	YouTube recommendation algorithm and vaccines disinformation during the COVID-19 pandemic C	
WoS	8	J	Khanday, AMUD; Khan, QR; Rabani, ST		Khanday, Akil Mot	Detecting Textual Propaganda Using Machine Learning Techniques E	
WoS	9	J	Mathur, A; Gupta, CP		Mathur, Akanksha	Supervised Machine Learning Methods for Early Detection of Untrue Information IF	
WoS	10	J	Piedrahita-Valdes, H; Piedrahita-Castillo, D; Bermejo-H		Piedrahita-Valdes, H	Vaccine Hesitancy on Social Media: Sentiment Analysis from June 2011 to April 2019 Y	
WoS	11	J	Castellano, PS		Simon Castellano	Trials by social media in the Digital Age. An old acquaintance in a new sphere F	
WoS	12	J	Gehrke, M; Benetti, M		Gehrke, Marilja; Benetti, M	Twitter AS A NEWS SOURCE IN DATA JOURNALISM E	
WoS	13	J	Gritsenko, D; Wood, M		Gritsenko, Daria	Algorithmic governance: A modes of governance approach F	
WoS	14	J	Ahmad, I; Yousef, M; Yousef, S; Ahmad, MO		Ahmad, Iftikhar; Yousef, M	Fake News Detection Using Machine Learning Ensemble Methods C	
WoS	15	J	Savolainen, L; Trilling, D; Liotsiou, D		Savolainen, Laura	Delighting and Detesting Engagement: Emotional Politics of Junk News S	
WoS	16	J	Etter, M; Albu, CB		Etter, Michael; Albu, Cornelia	Activists in the dark: Social media algorithms and collective action in two social movement organizations C	
WoS	17	J	Hoetzel, D; Alchin, D		Hoetzel, Dietmar	Reconceptualizing nature-of-science education in the age of social media S	
WoS	18	J	Ziccardi, G		Ziccardi, Giovanni	The use of social networks in politics between alteration of democratic balance, disinformation, propaganda F	
WoS	19	J	Saurwein, F; Spencer-Smith, C		Saurwein, Florian	Combating Disinformation on Social Media: Multilevel Governance and Distributed Accountability in Europe C	
WoS	20	J	Wilson, N		Wilson, Noah	Algorithmic Dwelling: Ethos as Deformation in Online Spaces F	
WoS	21	J	Innery, D; Colomina, C		Innery, Daniel	Truth in algorithmic democracies C	
WoS	22	J	van Dijk, J		van Dijk, Jose	Governing digital societies: Private platforms, public values F	
WoS	23	J	Wang, Xijun; Dai, J; Li, JX; Yu, JX; Jensen, CS; Yang, X		Wang, Xijun	Efficient targeted influence minimization in big social networks Y	
WoS	24	J	Tomaz, RM; Tomaz, JMT		Tomaz, Raissa M	The Brazilian Presidential Election of 2018 and the relationship between technology and democracy in Latin Y	
WoS	25	J	Alhazbi, S		Alhazbi, Saleh	Behavior-Based Machine Learning Approaches to Identify State-Sponsored Trolls on Twitter IF	
WoS	26	J	Vivar, JMF		Flores Vivar, Jesus	Big data, algorithmization and new media against misinformation and fake news. Bots to minimize the impac C	
WoS	27	J	Fernandez, FR		Ramon Fernandez	Communication and fake news in relation to COVID-19: some reflections on information, disinformation and F Y	
WoS	28	J	Yu, YT; Yu, G; Li, T; Man, QL; Chen, QP		Yu, Yongtian; Yu, G	Quantitative Characterization and Identification of the Company-Related Disinformation Channel Among Meit H	
WoS	29	J	Donovan, J; Boyd, D		Donovan, Joan; Boyd, D	Stop the Presses? Moving From Strategic Silence to Strategic Amplification in a Networked Media Ecosystem A	
WoS	30	J	Papadopoulos, O; Zampoglou, M; Papadopoulos, S; Ko		Garcia Sanz, Rosa	The conspiratorial relationship between the institution of Public Opinion and the public communication proces F	
WoS	31	J	Vivar, JMF		Flores Vivar, Jesús	Artificial intelligence and journalism: diluting the impact of misinformation and fakes news through bots F	
WoS	32	J	Iluassu, A; Capone, L; Firmino, LM; Mannheimer, V; Mu		Iluassu, Arthur	POLITICAL COMMUNICATION, ELECTIONS, AND DEMOCRACY: THE CAMPAIGNS OF DONALD TRUMP F	
WoS	33	J	Qayyum, A; Qadir, J; Janjua, MU; Sher, F		Qayyum, Adnan	Using Blockchain to Rein in the New Post-Truth World and Check the Spread of Fake News IF	
WoS	34	J	Rodriguez-Ferrandiz, R		Rodriguez-Ferrandiz, R	Post-truth and fake news in political communication: A brief genealogy F	
WoS	35	J	Iyengar, S; Massey, DS		Iyengar, Shant; Massey, DS	Scientific communication in a post-truth society F	
WoS	36	J	Nikolov, D; Lalmas, M; Fiammini, A; Mencer, F		Nikolov, Dimitar	Quantifying Biases in Online Information Exposure J	
WoS	37	J	Papadopoulos, O; Zampoglou, M; Papadopoulos, S; Ko		Papadopoulos, O	Corpus of debunked and verified user-generated videos Y	
WoS	38	J	Bradshaw, S		Bradshaw, Samant	Disinformation optimised: gaming search engine algorithms to amplify junk news IF	
WoS	39	J	Walker, S; Mercea, D; Bastos, M		Walker, Shawn; Mercea, D	The disinformation landscape and the lockdown of social platforms IF	
WoS	40	J	Kerysova, K		Kerysova, Katarina	Artificial Intelligence and Disinformation How AI Changes the Way Disinformation is Produced Disseminated S	
WoS	41	J	Bastos, M; Mercea, D		Bastos, Marco; Mercea, D	The public accountability of social platforms: lessons from a study on bots and trolls in the Brexit campaign F	
WoS	42	J	Chulvi, CP		Pauner Chulvi, C	Fake news and freedom of expression and information. The control of information contents on the network T	
WoS	43	J	Soe, SC		Soe, Sile Obelitz	Algorithmic detection of misinformation and disinformation: Gleaner perspectives J	
WoS	44	J	Brousseau, KL; Kant, JD; Sabournt, N; Prenot-Guinard		Brousseau, Kei-Li	From Beliefs to Attitudes: Poles, a Model of Attitude Dynamics Based on Cognitive Modeling and Field Data IF	
WoS	45	S	Last, M; Markov, A; Kandel, A		Chen, H; Wang, FY; Yan Last, M; Markov, A	Multi-lingual detection of terrorist content on the web IF	
Scopus	46	Scopus					
Scopus	47	Proceeding	Yu S., Lo D.			Disinformation detection using passive aggressive algorithms A	
Scopus	48	J	Yu Y., Yu G., Li T., Man Q., Chen Q.			Quantitative Characterization and Identification of the Company-Related Disinformation Channel among Meit H	
Scopus	49	Proceeding	Marinho M., Bastos-Filho C. J. A., J. Lins A.			An Exploratory Analysis on a Disinformation Dataset F	

Εικόνα 7. Αντιστοίχιση WoS και Scopus.

Book Serie	Language	Document	Conference	Conference	Conference	Conference	Conference	Author Key	Keywords	Abstract	Addresses	Reprint Ad	Email A
ZETTE	English	Review							Algorithms; big PRIVACY	We synthesize a burgeoning(Pennycook, C Pennycook, G Gordon, I)			
EMENT	English	Article							The regulation of internet in (Ioannidis, Petros Ioannidis, P (co P Ioannidis)				
M International Conf	English	Conference P	25th ACM SIG4	August 2019	through 8	August 2019			Fake news de RUMOR DETI The proliferation of fake nev(Song, Chengli Wu, B (corres) wubin@				
	English	Article							Fake news de INFORMATION Social media is becoming popular for news consumption due to it: fake news; bo				
	English	Article							CONTEMPOF This commentary on the 20 (Pek, Andrew Peck, A (corres)				
	English	Article							News is a medium that notif (Pandey, Avin Pandey, AC (cavish.ns				
T COMPUTING AND	English	Article							This article goes beyond the (Kalpokas, Igr Kalpokas, I (oikalpokas				
	English	Article							The accelerated spread of f (Farson, Mont Farson, M (co				
MANAGEMENT	English	Article							Using the Data, Information (Sardar, Ziaud Sardar, Z (cor zia@cp				
IDIES	Chinese	Article							Purpose The creation and d (Paschen, Jea Paschen, J (o pasche				
EMENT	English	Article							The advent of artificial intelli (Lin, Yu-Peng) Lin, Yu (corres				
COMPUTING COMM	English	Article							The concept of Fake Media (Chen, Qian; Srinavastava, G cheng44@				
EARCH	English	Article							The spread of fake news po (Baptista, Joao Gradim, A (cor shengaf				
ND ITS APPLICATIO	English	Article							Context: The speed with wh (Meneses Silv Silva, CVM (o kaiovini@				
CIENCE AND ENGIN	English	Article							Along with the development (Ozby, Feyza Alatas, B (cor fallunbe				
M International Conf	English	Conference P	12th ACM Inte	11 February 2019	through 15	February 2019			In temporal ordered cluster (Turowski, Krz Turowski, K (o krzyszc				
	English	Article							A Twitter social network has (Adewole, Kay Sangaihi, AK adewole@				
NS SCIENCES	Chinese	Article							The objective of this study i (Yoon, Byung; Yoon, B (corres postma				
EMENT	English	Article							The explosive growth of fake news and its erosion to democracy, media literacy				
ICATION	English	Article							A white paper on media liter (Lin, Yu-Peng; Wang, WC (o silentist				
	English	Article							Online Event (QUADTREE): A key challenge in mining si (George, Yasin Karunasekera karus@				
NS SECURITY	English	Article	Early A						Propaganda is a mechanism (Barron-Cadden Barron-Cadden a barron				
	English	Article							Online Social Media (OSM) (Abronizio, Hugo Barbon, S (co hugo.ab				
	English	Article							This article addresses the p (Elhadad, Moh Elhadad, MK (melhad				
	English	Article	Early A						This article discusses, from (Borges, Prisc Borges, PM (c primbor				
	English	Article							In this work we develop a s (Casolino, Mai Casolino, M (c casolino				
	English	Article							Due to the widely spreading (Nassar, Sabri El-Bendary, M sabry_r				
	English	Article							Several methods for synthe (Borrelli, Clara Borrelli, C (co clara.bo				
	English	Article							With the development of the (Ying, Long; Y Wang, JG (co wangjin@				
	English	Article	Early A						Evidence of a crisis of insti (Fiew, Terry) Fiew, T (corres t.fiew@				
	English	Article	Early A						The advent of COVID-19 ha (Mutanga, Mai Abayomi, A (o laofee@				
	English	Article							Undoubtedly, social media, (Ilias, Loukas; Roussaki, A (o loukasil				
	English	Article							Social media has become a (Vishwakarma Vishwakarma, dinesh@				
	English	Article							Deep learning in general, the features of f (Huang, Yin-Fi Huang, YF (co huangyf				
	English	Article							Algorithmic detection of fake news is a social phenomenon that is pervasive (Goldani, Moh Momiadi, S (o goldani@				
	Spanish	Article							Communication in Italy in th (Pira, Frances Pira, F (corres f.pira@				
	English	Article							Nowadays, news publicatio (Meel, Priyank Vishwakarma, priyank@				
	English	Article							Does fake news in financial (Clarke, Jonat Chen, HL (cor jonathar				
	English	Article							Fake news has increased d (Goldani, Moh Momiadi, S (o goldani@				
	English	Article	Early A						Artificial intelli (spreading of fake news is a social phenomenon that is pervasive) F				
	English	Article	Early A						The dissemination of fake n (Zervopoulos, Zervopoulos, c19zerv				
	English	Article							The exponential growth of c (Piqueira, Josi Piqueira, JRC piqueira				
	English	Article							What is fake news? Undou (Fox, Jo) Inst I Fox, J (corres				
	Spanish	Article							We are witnessing the emer (Gonzalez de de la Garza, L lmedlag				
	English	Article							The spread of infectious dis (Crane, Harry; Xu, M (corres mx76@				
	English	Article	Early A						With the power of social me (Almass, Mus Agarwal, N (o cmalmas				
	English	Article							These days online networki (Mahabub, Atil Mahabub, A (o atikmah				

Εικόνα 8. Αντιστοίχιση WoS και Scopus.

Κατά το τρίτο στάδιο συγχωνεύτηκαν για κάθε ερώτημα χωριστά οι δύο βάσεις δεδομένων (Εικόνα 9, Εικόνα 10) και εντοπίστηκαν οι κοινούς τίτλοι (Εικόνα 11).

Databases	No	Publication Type	Authors	Author Full Names	Article Title
WoS	124	J	Cooke, P	Cooke, Philip	Digital tech' and the public sector: what new rol
Scopus	356	J	Fox, J.		'Fake news' - The perfect storm: Historical pers
WoS	80	J	Fox, J	Fox, Jo	'Fake news' the perfect storm: historical persp
Scopus	551	J	[No author name available]		10th International Conference on Theory and Ap
Scopus	547	Proceedings	[No author name available]		11th International Conference on Intelligent Net
Scopus	552	J	[No author name available]		12th International Conference on Knowledge Si
Scopus	494	Lecture Notes	[No author name available]		13th International Conference on Computations
Scopus	535	J	[No author name available]		15th International Conference on Hybrid Artificial
Scopus	529	Proceedings	[No author name available]		16th International Conference on Distributed Co
WoS	116	J	Pira, F	Pira, Francesco	2019 EUROPEAN ELECTIONS CONTAMINAT
Scopus	481	J	[No author name available]		2020 17th International Computer Conference
Scopus	538	J	[No author name available]		23rd European Conference on Applications of E
Scopus	566	Proceedings	[No author name available]		24th International Conference on Application of
Scopus	544	J	[No author name available]		25th International Symposium on Methodologie
Scopus	567	J	[No author name available]		25th Pacific-Asia Conference on Knowledge Dis
Scopus	568	J	[No author name available]		25th Pacific-Asia Conference on Knowledge Dis
Scopus	569	J	[No author name available]		25th Pacific-Asia Conference on Knowledge Dis
Scopus	564	J	[No author name available]		2nd EAI International Conference on Data and
Scopus	532	Book	[No author name available]		2nd International Conference on Computing, Co
Scopus	533	Proceedings	[No author name available]		2nd International Conference on Data Science
Scopus	548	J	[No author name available]		2nd International Conference on Machine Learn
Scopus	540	Proceedings	[No author name available]		3rd International Conference on Cognitive Com
Scopus	512	Proceedings	[No author name available]		4th International Conference on Data Managerr
Scopus	565	Proceedings	[No author name available]		4th International Conference on Inventive Comr
Scopus	542	Proceedings	[No author name available]		4th International Conference on Smart Computi
Scopus	543	Proceedings	[No author name available]		4th International Conference on Smart Computi
Scopus	560	Proceedings	[No author name available]		5th International Conference on ICT for Sustain
Scopus	546	Proceedings	[No author name available]		5th International Conference on Recent Develo
Scopus	549	Proceedings	[No author name available]		7th International Conference on Modelling and
Scopus	476	Proceedings	[No author name available]		7th International Conference on Networked Sys
Scopus	482	Proceedings	[No author name available]		9th International Conference on Big Data, BigD
Scopus	559	Lecture Notes	[No author name available]		9th International Conference on Complex Netw
Scopus	463	J	[No author name available]		9th International Conference on Computational
Scopus	531	Proceedings	[No author name available]		9th Mining Humanistic Data Workshop, MHDW
WoS	104	J	Figueira, A; Guimaraes, N; Torgo, L	Figueira, Alvaro; Guimaraes, Nuno; Torgo, Luis	A Brief Overview on the Strategies to Fight Bac
Scopus	472	Proceedings	Figueira, A; Guimaraes N, Torgo L.		A Brief overview on the strategies to fight back t
Scopus	143	J	Pernagallo G, Torrisi B, Bennato D.		A Classification Algorithm to Recognize Fake N
Scopus	454	J	Diaz-Garcia J.A., Ruiz M.D., Martin-Bautista M.J.		A Comparative Study of Word Embeddings for I
Scopus	144	Lecture Notes	Ahmad F, Lokeshkumar R.		A comparison of machine learning algorithms in

Εικόνα 9. Ενοποίηση δεδομένων WoS και Scopus query 1.

WoS	31	J	Vivar, JMF		Flores Vivar, Jests Artificial intelligence and journalism: diluting the impact of misinformation and fakes news through bo
WoS	32	J	Iuassu, A; Capone, L; Firmino, LM; Mannheim, V; Mu		Iuassu, Arthur; Ca POLITICAL COMMUNICATION, ELECTIONS, AND DEMOCRACY: THE CAMPAIGNS OF DONALD
WoS	33	J	Qayyum, A; Qadir, J; Janjua, MU; Sher, F		Qayyum, Adnan; C Using Blockchain to Rein in the New Post-Truth World and Check the Spread of Fake News
WoS	34	J	Rodriguez-Ferrandiz, R		Rodriguez-Ferrandiz, R Post-truth and fake news in political communication: A brief genealogy
WoS	35	J	Iyengar, S; Massey, DS		Iyengar, Shanto; M Scientific communication in a post-truth society
WoS	36	J	Nikolov, D; Lalmas, M; Flammini, A; Menczer, F		Nikolov, Dimitar; L Quantifying Biases in Online Information Exposure
WoS	37	J	Papadopoulou, O; Zampoglou, M; Papadopoulos, S; Ko		Papadopoulou, O; A corpus of debunked and verified user-generated videos
WoS	38	J	Bradshaw, S		Bradshaw, Samant Disinformation optimised: gaming search engine algorithms to amplify junk news
WoS	39	J	Walker, S; Mercea, D; Bastos, M		Walker, Shaw; M The disinformation landscape and the lockdown of social platforms
WoS	40	J	Kertysova, K		Kertysova, Katarin Artificial Intelligence and Disinformation How AI Changes the Way Disinformation is Produced Disser
WoS	41	J	Bastos, M; Mercea, D		Bastos, Marco; M The public accountability of social platforms: lessons from a study on bots and trolls in the Brexit can
WoS	42	J	Chulvi, CP		Pauner Chulvi, Cri Fake news and freedom of expression and information. The control of information contents on the ne
WoS	43	J	Soe, SO		Soe, Silie Obelitz Algorithmic detection of misinformation and disinformation: Gricean perspectives
WoS	44	J	Brousniche, KL; Kant, JD; Sabouret, N; Prenot-Guinard		Brousniche, Kei-L From Beliefs to Attitudes: Polias, a Model of Attitude Dynamics Based on Cognitive Modeling and Fie
WoS	45	S	Last, M; Markov, A; Kandel, A	Chen, H; Wang, FY; Yar Last, M; Markov, A	A Multi-lingual detection of terrorist content on the web
	46	Scopus			
Scopus	47	Proceeding	Yu S, Lo D.		Disinformation detection using passive aggressive algorithms
Scopus	48	J	Yu Y, Yu G, Li T, Man Q, Chen Q.		Quantitative Characterization and Identification of the Company-Related Disinformation Channel am
Scopus	49	Proceeding	Marinho M., Bastos-Filho C.J.A., Lins A.		An Exploratory Analysis on a Disinformation Dataset
Scopus	50	J	Bnkuty-Balogh L.S.		Novel technologies and Geopolitical Strategies: Disinformation Narratives in the Countries of the Vis
Scopus	51	J	Saurwein F., Spencer-Smith C.		Combating Disinformation on Social Media: Multilevel Governance and Distributed Accountability in I
Scopus	52	J	Skarzuskiene A., Maciulene M., Ramasauskaitė O.		The digital media in Lithuania: Combating disinformation and fake news
Scopus	53	Book	Barela S.J., Duberry J.		Understanding disinformation operations in the twenty-first century
Scopus	54	J	Walker S., Mercea D., Bastos M.		The disinformation landscape and the lockdown of social platforms
Scopus	55	Lecture Not	Tardelli S., Avenuti M., Tesconi M., Cresci S.		Characterizing social bots spreading financial disinformation
Scopus	56	J	De Mesa J.C., Dominguez P.M., Muoz D.C., Jacinto L.G.		Homophily, affective polarization and disinformation on Twitter. Case study on the migration crisis #C
Scopus	57	J	Starbird K.		Disinformations spread: bots, trolls and all of us
Scopus	58	J	De Blasio E., Selva D.		Who Is Responsible for Disinformation? European Approaches to Social Platforms Accountability in I
Scopus	59	Proceeding	Nagy K.		Critical media literacy in the Education of Engineers How can we tackle the social risks posed by alg
Scopus	60	Book	ener O.		New literacies for disinformation and manipulation through digital sound and video
Scopus	61	Proceeding	Williams E.M., Novak V., Blackwell D., Platzman P., McCulloh I., Edward Phillips N.		Homophily and Transitivity in Bot Disinformation Networks
Scopus	62	Proceeding	Smith T.D.		Countering Inside Threat Actors in Algorithm-Based Media
Scopus	63	J	Etter M., Albu O.B.		Activists in the dark: Social media algorithms and collective action in two social movement organizati
Scopus	64	J	Orbegozo-Terradillos J., Morales-I-Gras J., Larrondo-Ureta A.		Disinformation in social networks: Closed or dialectical spaces? The case of Luther King, Quim Torra
Scopus	65	J	Said-Hundo E., Mejino-Arribas A., Martinez J.		Evolution of the academic debate on the web of science and scopus about unfakeing news (2014-201

Εικόνα 10. Ενοποίηση δεδομένων WoS και Scopus query 2.

	5th International Conference on ICT for Sustainable Development, ICT4SD 2020	L
	5th International Conference on Recent Developments in Science, Engineering and Technology, REDSET 2019	C
	7th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering, UPCON 2020	7
	7th International Conference on Modelling and Development of Intelligent Systems, MDIS 2020	C
	7th International Conference on Networked Systems, NETYS 2019	L
	9th International Conference on Big Data, BigData 2020, held as part of the Services Conference Federation, SCF 2020	L
	9th International Conference on Computational Data and Social Networks, CSoNet 2020	L
	9th Mining Humanistic Data Workshop, MHDW 2020, and the 5th Workshop on 5G-Putting Intelligence to the Network Edge, 5G-PINE 2020, held	L
reira, Alvaro; Guimaraes	A Brief Overview on the Strategies to Fight Back the Spread of False Information	↓
	A brief overview on the strategies to fight back the spread of false information	↓
.Y., Whangbo T.K., Choi	A Classification method of Fake News based on Ensemble Learning	#
l.	A closer look at fake news detection: A deep learning perspective	F
	A comparison of machine learning algorithms in fake news detection	I
	A comprehensive survey of spam profile detection methods in online social networks	↓
adopoulou, Olga; Zamp	A corpus of debunked and verified user-generated videos	C
Papadopoulos S., Kom	A corpus of debunked and verified user-generated videos	C
haian_Majed; Al-Qirish	A Credibility Analysis System for Assessing Information on Twitter	I

Εικόνα 11. Κοινοί τίτλοι query 3.

Για κάθε ερώτημα (Εικόνα 12) αφού έγινε εντοπισμός των κοινών τίτλων των πηγών, η κάθε διπλή πηγή συγκρίθηκε με την όμοιά της και διατηρήθηκε η πιο πλήρης σε πληροφορίες πηγή ενώ η άλλη διαγράφηκε (Εικόνα 13). Στη συνέχεια, οι αποδιπλοποιημένες εγγραφές των πηγών από το κάθε ξεχωριστό ερώτημα ενοποιήθηκαν σε νέο φύλλο του excel με τίτλο «totally merged» (Εικόνα 14, Εικόνα 15, Εικόνα 16, Εικόνα 17).

NoS	28	J	Yu, YT; Yu, G; Li, T; Man, QL; Chen, QP	Yu, Yongtian; Yu, C	Quantitative Characterization and Identification of the Company-Related Dis
NoS	17	J	Hoettecke, D; Allichin, D	Hoettecke, Dietma	Reconceptualizing nature-of-science education in the age of social media
Scopus	136	J	Supran G., Oreskes N.		Rhetoric and frame analysis of ExxonMobil's climate change communicator
NoS	35	J	Iyengar, S; Massey, DS	Iyengar, Shanto, M	Scientific communication in a post-truth society
Scopus	80	Proceeding	Iyengar S., Massey D.S.		Scientific communication in a post-truth society
Scopus	138	J	Mills M.C., Sivel J.		Should spreading anti-vaccine misinformation be criminalised?
NoS	29	J	Donovan, J; Boyd, D	Donovan, Joan; Bc	Stop the Presses? Moving From Strategic Silence to Strategic Amplification
Scopus	146	J	Donovan J., boyd D.		Stop the Presses? Moving From Strategic Silence to Strategic Amplification
NoS	9	J	Malthur A; Gupta CP	Malthur Akanksha	Supervised Machine Learning Methods for Early Detection of Untrue Inform

Εικόνα 12. Κοινοί τίτλοι ερωτημάτων WoS και Scopus.

Scopus	333	J	Ganesh P., Priya L., Nandakumar R.		Fake News Detection - A Comparative Study of Advance
Scopus	180	J	Choudhary P., Pandey S., Tripathi S., Chaurasiya S.		Fake News Detection Based on Machine Learning
Scopus	274	Lecture Notes	Desouky Fatoh I.E.L., Mousa F.A.		Fake news detection based on word and document embe
Scopus	235	Proceedings	Lyu S., Lo D.C.-T.		Fake News Detection by Decision Tree
Scopus	320	J	Steinebach M., Gotkowski K., Liu H.		Fake news detection by image montage recognition
Scopus	321	Proceedings	Steinebach M., Gotkowski K., Liu H.		Fake news detection by image montage recognition
Scopus	223	Lecture Notes	Ksieniewicz P., Zybiewski P., Chora M., Kozik R., Gieczyk A., Woniak M.		Fake News Detection from Data Streams
Scopus	283	J	Faustini P.H.A., Coves T.F.		Fake news detection in multiple platforms and languages
Scopus	161	Lecture Notes	Medeiros F.D.C., Braga R.B.		Fake news detection in social media: A systematic review
Scopus	150	J	Kumar A., Singh S., Kaur G.		Fake news detection of Indian and United States election

Εικόνα 13. Διπλοεγγραφές από Scopus.

Scopus	163	Proceedings	Vogel I., Meghana M.			Detecting fake news spreaders on h
Scopus	142	Proceedings	Bharath G., Manikanta K.J., Prakash G.B., Sumathi R., Chinnasamy P.			Detecting Fake News Using Machin
Scopus	146	Book	Abdullah-Ali-Tamir, Mahir E.M., Akhter S., Huq M.R.			Detecting Fake News using Machin
Scopus	145	Proceedings	Srinivasa Reddy S., Akhila M., Hinduja S., Hari Durga Reddy R., Vijaya Lakshmi B.			Detecting fake news using rip and n
WoS	34	J	Goldani, MH; Momtazi, S; Safabakhsh, R	Goldani, Mohammad Hadi; Momtazi, Saeedeh; Safab		Detecting fake news with capsule ne
Scopus	212	J	Aneja N., Aneja S.			Detecting Fake News with Machine
WoS	7	J	Ilias, L, Roussaki, I	Ilias, Loukas; Roussaki, Ioanna		Detecting malicious activity in Twitte
Scopus	449	Proceedings	Shi P., Zhang Z., Choo K.-K.R.			Detecting Malicious Social Bots Bas
WoS	85	J	Elhadad, MK; Li, KF; Gebali, F	Elhadad, Mohamed K.; Li, Kin Fun; Gebali, Fayez		Detecting Misleading Information on
Scopus	347	Proceedings	Deepika N., Guruprasad N.			Detecting text anomalies in social ne
Scopus	400	J	Miao L., Last M., Litvak M.			Detecting troll tweets in a bilingual c
Scopus	307	Proceedings	Homsí A., Nemri J.A., Naimat N., Kareem H.A., Al-Fayoumi M., Snober M.A.			Detecting Twitter fake accounts usin
WoS	93	J	Vishwakarma, DK; Varshney, D; Yadav, A	Vishwakarma, Dinesh Kumar; Varshney, Deepika; Yar		Detection and veracity analysis of fa
Scopus	339	Proceedings	Anjali B., Reshma R., Geetha Lekshmy V.			Detection of Counterfeit News Using
WoS	15	J	Mazzeo, V; Rapisarda, A; Giuffrida, G	Mazzeo, Valeria; Rapisarda, Andrea; Giuffrida, Giovar		Detection of Fake News on COVID-
Scopus	346	Proceedings	Sharma D., Singhal S.			Detection of fake news on social me
Scopus	151	J	Qazi M., Khan M.U.S., Ali M.			Detection of Fake News Using Trans
Scopus	308	J	Ayidin I., Sevi M., Salur M.U.			Detection of Fake Twitter Accounts v
WoS	10	J	Hansrajah, A; Adeliyi, TT; Wing, J	Hansrajah, Arvin; Adeliyi, Timothy T.; Wing, Jeanette		Detection of Online Fake News Usin
Scopus	338	J	Dhanya N.M., Harish U.C.			Detection of Rumors in Tweets Usin
Scopus	374	Lecture Notes	Oroh A.J., Bandung Y., Zagi L.M.			Detection of the Key Actor of Issues
WoS	2	J	Taskin, SG; Kucuksille, EU; Topal, K	Taskin, Suleyman Gokhan; Kucuksille, Epir Ugur; Top		Detection of Turkish Fake News in T
WoS	3	J	Ramkissoon, AN; Mohammed, S; Goodridge, W	Ramkissoon, Amil Neli; Mohammed, Shareeda; Good		Determining an Optimal Data Classi
Scopus	315	Proceedings	Singh R., Chun S.A., Atluri V.			Developing machine learning model
Scopus	264	J	Fleek W., Snell N., Traylor T., Straub J.			Development of a 'fake news' machi
Scopus	292	J	Kaneva O.N., Goncharenko A.I.			Development of software for fake ne
Scopus	239	J	Maci S.			Discourse strategies of fake news in
Scopus	310	J	Anandhan K., Damodharan D., Lakhanpal A., Manoj Sagar K., Murugan K., Shanker Singh A.			Discover Pretend Disease News Mis
Scopus	221	Lecture Notes	Chen Z., Freire J.			Discovering and Measuring Maliciou
Scopus	296	J	Yu S., Lo D.			Disinformation detection using passi
WoS	67	J	Orbegozo-Terradillos, J; Morales-I-Gras, J; Larrondo-Ureta, A	Orbegozo-Terradillos, Julien; Morales-I-Gras, Jordi; L		Disinformation in social networks: ch
Scopus	227	Proceedings	Gereme F.B., Zhu W.			Early detection of fake news before
Scopus	247	J	Silva A., Han Y., Luo L., Karunasekera S., Leckie C.			Embedding partial propagation netw
WoS	55	J	De Marzo, G, Zaccaria, A, Castellano, C	De Marzo, Giordano; Zaccaria, Andrea; Castellano, C		Emergence of polarization in a voter
Scopus	428	Proceedings	Battaily H.			Engaging closed-mindedly with your
Scopus	467	J	Paluch R., Gajewski J., Suchecki K., Szymaski B., Hoyst J.A.			Enhancing Maximum Likelihood Est

Εικόνα 14. Συγχώνευση εγγραφών query 1.

Databas	N	Publicat	Authors	Book Editors	Author Full Nar	Article Title
WoS	17	J	Hoettecke, D; Alchin, D		Hoettecke, Dietma	Reconceptualizing nature-of-science education in the age of social media
Scopus	118	J	Ahmed W., Vidal-Alaball J., Downing J., Segu F.L.			COVID-19 and the 5G conspiracy theory: Social network analysis of twitter data
WoS	2	J	Reisach, U		Reisach, Ulrike	The responsibility of social media in times of societal and political manipulation
WoS	6	J	Robertson, R; Meintjes, C		Robertson, Randy	Towards an Online Risk Mitigation Framework for Political Brands Subject to Computational Propaganda
WoS	19	J	Saurwein, F; Spencer-Smith, C		Saurwein, Florian	Combating Disinformation on Social Media: Multilevel Governance and Distributed Accountability in Europe
WoS	13	J	Gritsenko, D; Wood, M		Gritsenko, Daria; V	Algorithmic governance: A modes of governance approach
WoS	40	J	Kertysova, K		Kertysova, Katarin	Artificial Intelligence and Disinformation How AI Changes the Way Disinformation is Produced Disseminated
WoS	32	J	Ituassu, A; Capone, L; Firmino, LM; Mannheim, V; Mu		Ituassu, Arthur; Ca	POLITICAL COMMUNICATION, ELECTIONS, AND DEMOCRACY: THE CAMPAIGNS OF DONALD TRUMP
WoS	34	J	Rodriguez-Ferrandiz, R		Rodriguez-Ferran	Post-truth and fake news in political communication: A brief genealogy
Scopus	67	Proceeding	Iyengar S., Massey D.S.			Scientific communication in a post-truth society
WoS	11	J	Castellano, PS		Simon Castellano,	Trials by social media in the Digital Age. An old acquaintance in a new sphere
WoS	41	J	Bastos, M; Mercea, D		Bastos, Marco; Me	The public accountability of social platforms: lessons from a study on bots and trolls in the Brexit campaign
WoS	29	J	Donovan, J; Boyd, D		Donovan, Joan; B;	Stop the Presses? Moving From Strategic Silence to Strategic Amplification in a Networked Media Ecosystem
WoS	39	J	Walker, S; Mercea, D; Bastos, M		Walker, Shawn; M;	The disinformation landscape and the lockdown of social platforms
WoS	43	J	Soe, SO		Soe, Silie Obelitz	Algorithmic detection of misinformation and disinformation: Gricean perspectives
WoS	16	J	Etter, M; Albu, OB		Etter, Michael; Albu	Activists in the dark: Social media algorithms and collective action in two social movement organizations
WoS	38	J	Bradshaw, S		Bradshaw, Samant	Disinformation optimised: gaming search engine algorithms to amplify junk news
WoS	5	J	De Blasio, E; Selva, D		De Blasio, Emilian	Who Is Responsible for Disinformation? European Approaches to Social Platforms' Accountability in the Pos
WoS	27	J	Fernandez, FR		Ramon Fernandez	Communication and fake news in relation to COVID-19: some reflections on information, disinformation and
WoS	3	J	Grotluschen, A		Grotluschen, Anki	Vulnerability and Literacy in the 21st Century
WoS	44	J	Brousmiche, KL; Kant, JD; Sabouret, N; Prenot-Guinard		Brousmiche, Kei-L	From Beliefs to Attitudes: Polias, a Model of Attitude Dynamics Based on Cognitive Modeling and Field Data
WoS	25	J	Alhazbi, S		Alhazbi, Saleh	Behavior-Based Machine Learning Approaches to Identify State-Sponsored Trolls on Twitter
WoS	24	J	Tomaz, RM; Tomaz, JMT		Tomaz, Raissa Me	The Brazilian Presidential Election of 2018 and the relationship between technology and democracy in Latin
WoS	36	J	Nikolov, D; Lalmas, M; Flammini, A; Menczer, F		Nikolov, Dimitar; I	Quantifying Biases in Online Information Exposure
WoS	7	J	Sued, GE		Eilea Sued, Gabrie	YouTube recommendation algorithm and vaccines disinformation during the COVID-19 pandemic
WoS	28	J	Yu, YT; Yu, G; Li, T; Man, QL; Chen, QP		Yu, Yongtian; Yu, C	Quantitative Characterization and Identification of the Company-Related Disinformation Channel Among Me
WoS	15	J	Savolainen, L; Trilling, D; Liotsiou, D		Savolainen, Laura;	Delighting and Detesting Engagement: Emotional Politics of Junk News
WoS	10	J	Piedrahita-Valdes, H; Piedrahita-Castillo, D; Bermejo-Hi		Piedrahita-Valdes,	Vaccine Hesitancy on Social Media: Sentiment Analysis from June 2011 to April 2019
WoS	14	J	Ahmad, I; Yousaf, M; Yousaf, S; Ahmad, MO		Ahmad, Ifikhar; Yc	Fake News Detection Using Machine Learning Ensemble Methods
Scopus	53	J	Starbird K.			Disinformations spread: bots, trolls and all of us
WoS	20	J	Wilson, N		Wilson, Noah	Algorithmic Dwelling: Ethos as Deformance in Online Spaces
WoS	45	S	Last, M; Markov, A; Kandel, A	Chen, H; Wang, FY; Yar	Last, M; Markov, A	Multi-lingual detection of terrorist content on the web
WoS	1	J	De Aguiar, LA; Rodrigues, CM		De Aguiar, Leone;	Expertise in Journalism: Considerations about professional authority in the con-text of disinformation driven
WoS	4	J	Johnson, KL; Walsh, JL; Amirkhania, YA; Borkowski, J.		Johnson, Kara Lay	Using a novel genetic algorithm to assess peer influence on willingness to use pre-exposure prophylaxis in r
WoS	42	J	Chulvi, CP		Pauner Chulvi, Ori	Fake news and freedom of expression and information. The control of information contents on the network

Εικόνα 15. Συγχώνευση εγγραφών query 2.

Databas	No	Publicat	Authors	Author Full Names	Article Title
Scopus	358	J	Anjad M., Sidorov G., Zhila A., Gmez-Adorno H., Voronko	Gmez-Adorno H., Voronko	Bend the truth: Benchmark dataset for fake news detection in Urdu language and its evaluation
WoS	112	J	Cooke, P	Cooke, Philip	'Digital tech' and the public sector: what new role after public funding?
Scopus	584	Book	[No author name available]		11th International Conference on Intelligent Networking and Collaborative Systems, INCoS 2019
Scopus	590	Book	[No author name available]		12th International Conference of the Cross-Language Evaluation Forum for European Languages, CLE
Scopus	577	Book	[No author name available]		13th International Conference on Computational Intelligence in Security for Information Systems, CISIS
Scopus	581	Book	[No author name available]		13th International Conference on Computational Processing of the Portuguese Language, PROPOR 20
Scopus	569	Book	[No author name available]		14th International Conference on Computational Processing of the Portuguese Language, PROPOR 21
Scopus	572	Book	[No author name available]		15th International Conference on Hybrid Artificial Intelligent Systems, HAIS 2020
Scopus	552	Book	[No author name available]		16th Ibero-American Conference on Artificial Intelligence, IBERAMIA 2018
Scopus	586	Book	[No author name available]		17th International Conference on Computer Information Systems and Industrial Management, CISIM 21
Scopus	591	J	[No author name available]		1st BenchCouncil International Federated Intelligent Computing and Block Chain Conferences, FICC 21
Scopus	585	Book	[No author name available]		1st Doctoral Symposium on Natural Computing Research, DSNCR 2020
Scopus	546	Book	[No author name available]		1st International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud
Scopus	537	Book	[No author name available]		2018 4th International Conference on Computing Communication and Automation, ICCCA 2018
Scopus	559	Book	[No author name available]		2020 17th International Computer Conference on Wavelet Active Media Technology and Information Pr
Scopus	595	Book	[No author name available]		23rd International Conference on Text, Speech, and Dialogue, TSD 2020
Scopus	589	Book	[No author name available]		24th Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD 2020
Scopus	555	Book	[No author name available]		25th International Symposium on Methodologies for Intelligent Systems, ISMIS 2020
Scopus	538	Book	[No author name available]		2nd International Conference on Computing, Communications, and Cyber-Security, IC4S 2020
Scopus	458	J	[No author name available]		2nd International Conference on Machine Learning and Information Processing, ICMLIP 2020
Scopus	508	Book	[No author name available]		2nd Multidisciplinary International Symposium on Disinformation in Open Online Media, MISDOOM 202
Scopus	526	Book	[No author name available]		3rd International Conference on Smart IoT Systems: Innovations and Computing, SSIC 2021
Scopus	520	Book	[No author name available]		4th International Conference on Data Management, Analytics and Innovation, ICDMAI 2020
Scopus	567	Book	[No author name available]		4th International Conference on Smart Computing and Informatics, SCI 2020
Scopus	568	Book	[No author name available]		4th International Conference on Smart Computing and Informatics, SCI 2020
Scopus	558	Book	[No author name available]		5th International Conference on ICT for Sustainable Development, ICT4SD 2020
Scopus	592	J	[No author name available]		5th International Conference on Recent Developments in Science, Engineering and Technology, REDS
Scopus	530	Book	[No author name available]		7th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engi
Scopus	561	J	[No author name available]		7th International Conference on Modelling and Development of Intelligent Systems, MDIS 2020
Scopus	550	Book	[No author name available]		7th International Conference on Networked Systems, NETYS 2019
Scopus	557	Book	[No author name available]		9th International Conference on Big Data, BigData 2020, held as part of the Services Conference Fede
Scopus	524	Book	[No author name available]		9th International Conference on Computational Data and Social Networks, CSoNet 2020
Scopus	588	Book	[No author name available]		9th Mining Humanistic Data Workshop, MHDW 2020, and the 5th Workshop on 5G-Putting Intelligence
WoS	100	J	Figueira, A; Guimaraes, N; Tor Figueira, Alvaro; Guimaraes	Figueira, A; Guimaraes, N; Tor Figueira, Alvaro; Guimaraes	A Brief Overview on the Strategies to Fight Back the Spread of False Information
Scopus	272	Book	Lee S.B., Lim J.S., Cho J.S., Oh S.Y., Whangbo T.K., Choi	Lee S.B., Lim J.S., Cho J.S., Oh S.Y., Whangbo T.K., Choi	A Classification method of Fake News based on Ensemble Learning
Scopus	235	Proceeding	Abdelhalla A., Al-Sadi A., Abdullah M.	Abdelhalla A., Al-Sadi A., Abdullah M.	A closer look at fake news detection: A deep learning perspective
Scopus	138	J	Ahmad F., Lokeskumar R.	Ahmad F., Lokeskumar R.	A comparison of machine learning algorithms in fake news detection
Scopus	542	J	Krithiga R., Ilavarasan E.	Krithiga R., Ilavarasan E.	A comprehensive survey of spam profile detection methods in online social networks
WoS	104	J	Papadopoulou, O; Zampoglou Papadopoulou, Olga; Zamp	Papadopoulou, O; Zampoglou Papadopoulou, Olga; Zamp	A corpus of debunked and verified user-generated videos
Scopus	532	J	Alrubaihan M., Al-Qurishi M., Hassan M.M., Alami A.	Alrubaihan M., Al-Qurishi M., Hassan M.M., Alami A.	A Credibility Analysis System for Assessing Information on Twitter
WoS	14	J	Ferreira, S; Antunes, M; Corre Ferreira, Sara; Antunes, Ma	Ferreira, S; Antunes, M; Corre Ferreira, Sara; Antunes, Ma	A Dataset of Photos and Videos for Digital Forensics Analysis Using Machine Learning Processing

Εικόνα 16. Συγχώνευση εγγραφών query 3.

Databas	No	Publicat	Authors	Book Edito	Author Full	Article Title
WoS	109	J	Zhao, J; Cao, N; Wer		Zhao, Jian; Cac	#FluxFlow: Visual Analysis of Anomalous Information Spreading on Social Media
Scopus	242	Book	[No author name available]			6th International Symposium on Signal Processing and Intelligent Recognition Systems, SIRS
Scopus	158	J	zgbek ., Kille B., Gulla J.A., Larson M., Lommatzsch			6th international workshop on news recommendation and analytics (INRA 2018)
Scopus	194	Book	Oni O.O., Zyl T.L.V.			A Comparative Study of Ensemble Approaches to Fact-Checking for the FEVER Shared Task
Scopus	202	Book	Diaz-Garcia J.A., Ruiz M.D., Martin-Bautista M.J.			A Comparative Study of Word Embeddings for the Construction of a Social Media Expert Filter
WoS	32	J	Zeng, J; Chan, CH		Zeng, Jing; Che	A cross-national diagnosis of infodemics: comparing the topical and temporal features of misin
Scopus	199	Proceeding	Kaubiyal J., Jain A.K.			A feature based approach to detect fake profiles in twitter
WoS	76	J	Yuliani, SY; Bin Abdol		Yuliani, S. Y.; Bi	A Framework for Hoax News Detection and Analyzer used Rule-based Methods
WoS	119	S	Demirbas, M; Arora, f Higashino, T		Demirbas, M; A	A hierarchy-based fault-local stabilizing algorithm for tracking in sensor networks
WoS	30	J	Guo, JX; Chen, TT; W		Guo, Jianxiang;	A Multi-Feature Diffusion Model: Rumor Blocking in Social Networks
WoS	89	J	Abuzainab, N; Saad,		Abuzainab, Nof	A Multiclass Mean-Field Game for Thwarting Misinformation Spread in the Internet of Battlefield
WoS	5	J	Safarnejad, L; Xu, Q;		Safarnejad, Lid	A Multiple Feature Category Data Mining and Machine Learning Approach to Characterize and
Scopus	263	J	Luo X., Zimet G., Shah S.			A natural language processing framework to analyse the opinions on HPV vaccination reflecte
Scopus	232	Book	Saikh T., Anand A., Ekbal A., Bhattacharyya P.			A Novel Approach Towards Fake News Detection: Deep Learning Augmented with Textual Ent
Scopus	192	Book	Rao A. S., Shetty A., Uphade A., Thawani P., Priya R.L.			A proposal for a novel approach to analyze and detect the fake news using AI techniques
Scopus	179	Book	Kumar S., Arora B.			A Review of Fake News Detection Using Machine Learning Techniques
Scopus	241	Book	Zhang M., Wang X., Qin X.			A rumor blocking method based on tolerance of users
WoS	58	J	Zhu, JM; Ni, PK; Wan		Zhu, Jianming;	Activity Minimization of Misinformation Influence in Online Social Networks
WoS	50	J	Cohen, R; Moffatt, K;		Cohen, Robin;	Addressing Misinformation in Online Social Networks: Diverse Platforms and the Potential of M
Scopus	203	Book	Jain I., Goel N.			Advancements in image splicing and copy-move forgery detection techniques: A survey
WoS	20	J	Gallo, A; Perregaux, J		Gallo, Antonio;	Advances in the management of statin myopathy
Scopus	245	J	Ballesteros M.F., Sumner S.A., Law R., Wolkin A., Jo			Advancing injury and violence prevention through data science
Scopus	174	Book	Umer M., Polikar R.			Adversarial Targeted Forgetting in Regularization and Generative Based Continual Learning M
WoS	82	J	Tan, ZH; Wu, DK; Gai		Tan, Zhenhua;	AIM: Activation increment minimization strategy for preventing bad information diffusion in OSN
WoS	70	J	Gorwa, R; Binns, R; h		Gorwa, Robert;	Algorithmic content moderation: Technical and political challenges in the automation of platform
WoS	96	J	Soe, SO		Soe, Sille Obell	Algorithmic detection of misinformation and disinformation: Gricean perspectives
Scopus	184	J	Hunt R., McKelvey F.			Algorithmic regulation in media and cultural policy a framework to evaluate barriers to account
WoS	63	J	Shin, J; Valente, T		Shin, Jieun; Val	Algorithms and Health Misinformation: A Case Study of Vaccine Books on Amazon
WoS	25	J	Tundis, A; Mukherjee		Tundis, Andrea;	An Algorithm for the Detection of Hidden Propaganda in Mixed-Code Text over the Internet
Scopus	151	Book	Arazkhani N., Meybodi M.R., Rezvanian A.			An Efficient Algorithm for Influence Blocking Maximization based on Community Detection
WoS	24	J	Rahman, MS; Halder,		Rahman, Md St	An efficient hybrid system for anomaly detection in social networks
WoS	66	J	Tong, GM; Wu, WL; C		Tong, Guangmc	An Efficient Randomized Algorithm for Rumor Blocking in Online Social Networks
Scopus	239	J	Anuyah O., Milton A., Green M., Pera M.S.			An empirical analysis of search engines response to web search queries associated with the cl
WoS	55	J	Albahr, A; Albahar, M		Albahr, Abdulaz	An Empirical Comparison of Fake News Detection using different Machine Learning Algorithms
WoS	14	J	Al-Ahmad, B; Al-Zout		Al-Ahmad, Bilal	An Evolutionary Fake News Detection Method for COVID-19 Pandemic Information
Scopus	130	J	Tran T., Valecha R., Rad P., Rao H.R.			An investigation of misinformation harms related to social media during humanitarian crises
WoS	87	J	Liu, Y; Yu, K; Wu, XF;		Liu, Yue; Yu, Ke	Analysis and Detection of Health-Related Misinformation on Chinese Social Media
WoS	111	J	Nguyen, NP; Yan, GH		Nguyen, Nam F	Analysis of misinformation containment in online social networks
Scopus	156	J	Samanta P., Jain S.			Analysis of Perceptual Hashing Algorithms in Image Manipulation Detection

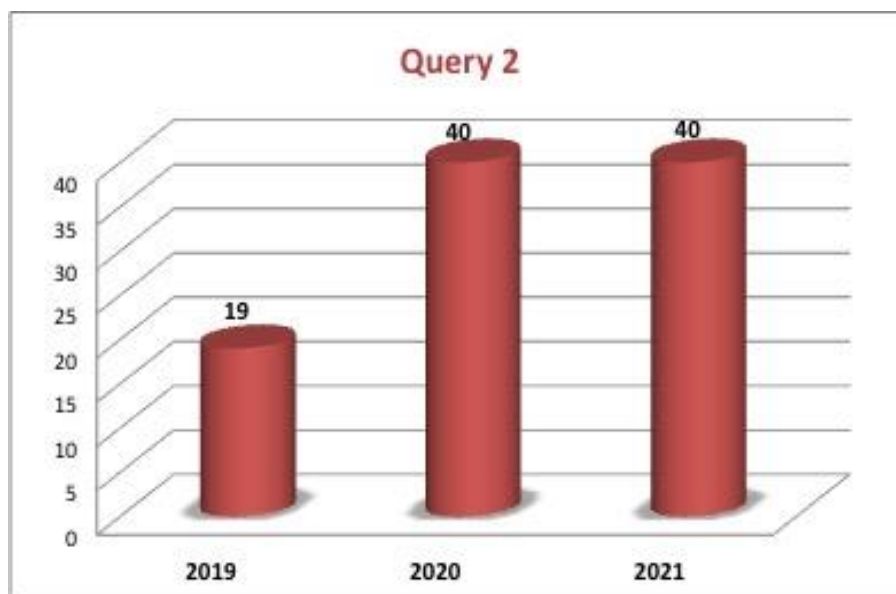
Εικόνα 17. Συγχώνευση εγγραφών query 4.

Στο τελευταίο βήμα πραγματοποιήθηκε εκ νέου αποδιπλοποίηση των δεδομένων των συνολικά τεσσάρων ερωτημάτων από όπου έμειναν οι μοναδικές εγγραφές.

Ακολουθούν τα γραφήματα με τα ποσοτικά δεδομένα πηγών ανά έτος για κάθε ερώτημα (query) (Εικόνα 18, Εικόνα 19, Εικόνα 20, Εικόνα 21).



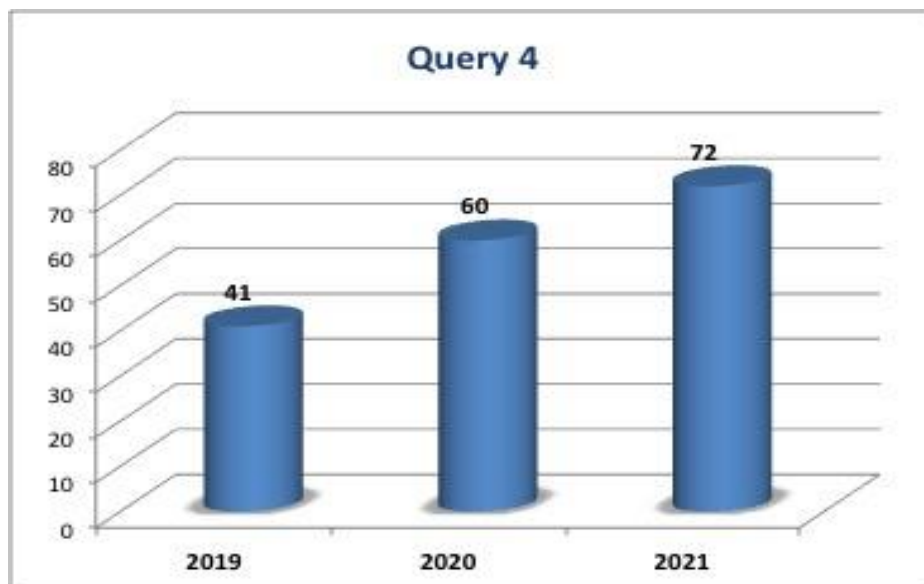
Εικόνα 18. Αποτελέσματα query 1.



Εικόνα 19. Αποτελέσματα query 2.



Εικόνα 20. Αποτελέσματα query 3.



Εικόνα 21. Αποτελέσματα query 4.

3.2 Περιγραφή Υλοποίησης

Η φάση της υλοποίησης, ουσιαστικά αποτελεί το τέταρτο στάδιο της έρευνας. Αρχικά επιλέχθηκαν για μελέτη οι πηγές που έφεραν το μεγαλύτερο αριθμό αναφορών, ουσιαστικά με 100 αναφορές και πάνω. Διαπιστώθηκε ότι σε κάποιες πηγές δεν υπήρχε πρόσβαση ή ήταν εκτός αντικείμενου της παρούσας διπλωματικής εργασίας αλλά και σε γλώσσα μη κατανοητή (π.χ. Κινέζικα). Εξαιτίας των συγκεκριμένων λόγων επιλέχθηκαν και πηγές με λιγότερες αναφορές (π.χ. με 73 αναφορές). Έπειτα από την απόρριψη των παραπάνω πηγών μελετήθηκαν οι περιλήψεις των υπόλοιπων πηγών με τις περισσότερες αναφορές και

επιλέχτηκαν οι 60 πιο συναφείς, με βάση τα θέματα που εξετάζει η παρούσα διπλωματική. Οι επιλεγμένες πηγές παρουσιάζονται σε διαμορφωμένο πίνακα (πίνακας χαρακτηριστικών ψευδών ειδήσεων) με βάση την αρίθμηση του excel αρχείου totally merged (π.χ. 108).

Προκειμένου για μεγαλύτερη εμβάθυνση στην έρευνα υιοθετήθηκε και η μέθοδος snowball που επιτρέπει τον περαιτέρω εντοπισμό συναφών πηγών. Η συγκεκριμένη μέθοδος περιγράφεται αναλυτικά στην ενότητα 3.2.1 και αποτελεί το πέμπτο στάδιο υλοποίησης της διπλωματικής. Οι πηγές, με βάση τη συγκεκριμένη μέθοδο, παρουσιάζονται στον πίνακα με τα χαρακτηριστικά των ψευδών ειδήσεων μέσα σε αγκύλες και με βάση τις παραπομπές των άρθρων (π.χ. με την αρίθμηση [68]). Κατά το έκτο και τελευταίο στάδιο της υλοποίησης έγινε η κατηγοριοποίηση των θεμάτων του περιεχομένου των πηγών για την οποία ακολουθήθηκε η μέθοδος της ανάλυσης περιεχομένου. Σε αυτό το στάδιο δημιουργήθηκαν κωδικοποιημένες κατηγορίες με βάση το περιεχόμενο των πηγών (δείτε αναλυτικά κεφάλαιο 4).

3.2.1 Δειγματοληψία Χιονοστιβάδας (Snowball Sampling)

Η μέθοδος συλλογής δεδομένων που χρησιμοποιείται στις κοινωνικές επιστήμες, είναι η μη πιθανοτική δειγματοληψία χιονοστιβάδας. Η δειγματοληψία χιονοστιβάδας αναφέρεται από τους Handcock και Gile ως «η συλλογή ενός δείγματος από έναν πληθυσμό όπου η συνήθης δειγματοληπτική προσέγγιση είναι είτε αδύνατη είτε απαγορευτικά κοστοβόρα, για τον σκοπό της μελέτης χαρακτηριστικών των ατόμων του πληθυσμού. Αυτού του είδους οι καταστάσεις αφορούν πληθυσμούς που είναι δύσκολα προσεγγίσιμοι και χαρακτηρίζονται από έλλειψη καθορισμένου δειγματοληπτικού πλαισίου» (Handcock & Gile, 2011).

Σε πολλές περιπτώσεις πληθυσμών, δύσκολα ανιχνεύσιμων, η δειγματοληψία που βασίζεται στην ανίχνευση συνδέσμων είναι αποτελεσματικό μέσο συλλογής δεδομένων ανάμεσα στα μέλη ενός πληθυσμού. Για το λόγο αυτό, η μη πιθανοτική χρήση της δειγματοληψίας χιονοστιβάδας αποτελεί συνηθισμένη πρακτική, αν και λιγότερο συνηθισμένη στην στατιστική, που υποστηρίζει πιθανοτικές εφαρμογές (Handcock & Gile, 2011).

Σύμφωνα με τους συγγραφείς, η δειγματοληψία χιονοστιβάδας μπορεί να πραγματοποιηθεί με δύο τρόπους οι οποίοι βασίζονται σε δειγματοληψία δικτύων. Στην πρώτη περίπτωση μέλη, ενός πληθυσμού προς μελέτη, προτείνουν άλλα μέλη του πληθυσμού και εκείνα με την σειρά τους άλλα, έως ότου διαμορφωθεί ένα πλαίσιο ικανού δείγματος συλλογής δεδομένων. Στην δεύτερη περίπτωση ακολουθείται η ίδια διαδικασία

βασιζόμενη όμως στο επίπεδο των σχέσεων και στο δίκτυο των επαφών που διατηρούν μεταξύ τους τα άτομα όπως για παράδειγμα κοινωνικές σχέσεις (οικογενειακές, φιλικές, επαγγελματικές). Με αυτό τον τρόπο είναι δυνατή η πρόσβαση σε πληροφορίες που αφορούν πληθυσμούς δύσκολα ανιχνεύσιμους ενώ αξιοποιείται η δυναμική των ήδη διαμορφωμένων σχέσεων και του υπό μελέτη πληθυσμού.

Η συγκεκριμένη μέθοδος δειγματοληψίας αναφέρεται και ως «αλυσίδα» λόγω του αλυσιδωτού τρόπου προσέγγισης του δείγματος και ένταξης των συλλεχθέντων δεδομένων στην ερευνητική διαδικασία. Αξιοποιείται τόσο στην ποσοτική όσο και στην ποιοτική έρευνα μέσα από την συλλογή αρχειακού υλικού, κειμένων, συνεντεύξεων και βιογραφιών (Noy, 2008).

Η αποτύπωση των δεδομένων που προκύπτει από μία δειγματοληψία χιονοστιβάδας (ή αλυσίδας) πραγματοποιείται μέσα από «δειγματοληπτικά δέντρα ή στελέχη». Η δυναμική διαδικασία αυτής της μεθόδου συλλογής, αλλά και παραγωγής δεδομένων, επιτρέπει την ανάδειξη κοινωνικής γνώσης έτσι όπως προκύπτει από την αξιοποίηση πληροφοριών που προέρχονται από την κοινωνική δικτύωση (Noy, 2008).

Όπως αναφέρουν οι Heckathorn και Cameron η δειγματοληψία χιονοστιβάδας βασίζεται στην δικτυακή δειγματοληψία (network sampling) αξιοποιώντας τα ίχνη των συνδέσμων μεταξύ των κόμβων (άτομα, κείμενα) που παρέχουν αρχικά και εν συνεχεία τις πληροφορίες στους ερευνητές (Heckathorn & Cameron, 2017).

Η δειγματοληψία χιονοστιβάδας αξιοποιείται, επίσης, και για την συλλογή δεδομένων που αφορούν σε πληροφορίες επιστημονικών άρθρων σχετικών με ένα θεματικό πεδίο μελέτης. Η έρευνα των Dobronolskyi και Keberle αξιοποίησε την δειγματοληψία χιονοστιβάδας σε περιλήψεις επιστημονικών άρθρων ενός γνωστικού αντικειμένου με στόχο να αναζητήσουν τα κριτήρια ένταξης, σε μία βιβλιογραφική επισκόπηση, της πλέον σχετικής αρθρογραφίας στο υπό συζήτηση θεματικό πεδίο μελέτης (Dobronolskyi & Keberle, 2019).

Οι ερευνητές συμπεριέλαβαν στην ερευνητική τους στρατηγική, μεταξύ άλλων μεθόδων, την συγκεκριμένη μέθοδο δειγματοληψίας προκειμένου να συλλέξουν αναφορές των πλέον σχετικών επιστημονικών, με το θεματικό πεδίο, άρθρων. Επιπλέον, τονίζουν ότι *«η δειγματοληψία χιονοστιβάδας επιτρέπει την συλλογή δημοσιεύσεων σε ένα συγκεκριμένο θέμα έρευνας και την σύνδεσή τους με το δίκτυο των αναφορών»* (Dobronolskyi & Keberle, 2019).

Για την πραγματοποίηση της δειγματοληψίας χιονοστιβάδας αξιοποίησαν ως εργαλείο το πιθανοτικό τοπικό μοντέλο (Probabilistic Topic Model) *«που χρησιμοποιεί ένα μεγάλο σύνολο εγγράφων και στατιστική προσέγγιση για την δημιουργία μοντέλων λέξεων*

και εγγράφων ως διανύσματα σε έναν υψηλών διαστάσεων σημασιολογικό χώρο R^n όπου το n είναι πολύ μικρότερο από τον αριθμό των λέξεων και των εγγράφων. Η βασική ιδέα του PTM είναι η δημιουργία περιορισμένων θεματικών που αποτελούν ομαδοποιήσεις στενά συνδεδεμένων λέξεων. Εν συνεχεία οι λέξεις του κειμένου αναπαρίστανται ως αποτέλεσμα τυχαίας δειγματοληψίας διπλού επιπέδου» (Dobronolski & Keberle, 2019). Στην έρευνά τους αξιοποίησαν τις πληροφορίες που ήταν προσβάσιμες στις βάσεις δεδομένων όπως περιλήψεις των άρθρων, τους τίτλους τους αλλά και λέξεις-κλειδιά που περιλαμβάνονταν σε αυτές, όπου αυτή η πληροφορία ήταν διαθέσιμη (Dobronolski et al., 2017).

Προηγούμενες σχετικές έρευνες, έχουν αξιοποιήσει επίσης την δειγματοληψία χιονοστιβάδας, αλλά με μια πιο περιορισμένη προσέγγιση έτσι ώστε να οριοθετήσουν τα πλέον σχετικά άρθρα με τις σχετικές αναφορές τους, που συζητούν ένα θεματικό πεδίο. Στόχος της έρευνας, και σε αυτή την περίπτωση ήταν να διευκολυνθούν οι ερευνητές στην αναζήτηση και εύρεση της πλέον σχετικής βιβλιογραφίας έτσι ώστε η βιβλιογραφική επισκόπηση, σε ένα γνωστικό αντικείμενο που τους ενδιαφέρει, να είναι πληρέστερη και πιο εστιασμένη.

Οι Lecy και Beatty εφάρμοσαν στην έρευνά τους περιορισμένη δειγματοληψία χιονοστιβάδας για να αναδείξουν το δίκτυο των αναφορών (citation network) που προκύπτει από ένα αρχικό άρθρο (seed article), αναφορικά με ένα εξειδικευμένο θέμα ή γνωστικό αντικείμενο. Στην συγκεκριμένη έρευνα, παρατηρήθηκε ότι άρθρα τα οποία φαίνεται να απομακρύνονται από το αρχικό, κύριο, άρθρο, που αποτελεί σημείο θεματικής αναφοράς σε μια βιβλιογραφική επισκόπηση, εμφανίζονται όλο και περισσότερο περιφερειακά, τότε έχουν μικρότερη πιθανότητα συσχέτισης με το υπό συζήτηση θέμα (Lecy & Beatty, 2012).

3.2.2 Ανάλυση Περιεχομένου: Ιστορικές και Εννοιολογικές Προσεγγίσεις

Ιστορικά σύμφωνα με τον Krippendorff, η ανάλυση περιεχομένου συναντάται στις διερευνητικές αναζητήσεις της Εκκλησίας τον 17^ο αιώνα, όπου μετά την ανακάλυψη και εξάπλωση του τύπου, θορυβημένη από την πληθώρα μη εκκλησιαστικών δημοσιεύσεων ασχολήθηκε επισταμένα με την ηθική του περιεχομένου των εφημερίδων (Krippendorff, 2004). Η πρώτη αξιόλογη ποσοτική ανάλυση περιεχομένου με βάση την αναφορά της Donring χρησιμοποιήθηκε τον 18^ο αιώνα από τη σουηδική εκκλησία με τη δημοσίευση των «Τραγουδιών της Ζίον», ενός συνόλου από ύμνους αγνώστου συγγραφέα προκειμένου να ελεγχθούν εάν ήταν «βλάσφημοι» ή σύμφωνοι με τα εκκλησιαστικά πρότυπα (Donring, 1954; Macnamara, 2018).

Τον 19ο αιώνα η ανάλυση περιεχομένου πρωτοεμφανίστηκε στις εφημερίδες με τη μέτρηση στηλών σε ένα δεδομένο θέμα αλλά και στη συνέχεια το 1893 σε φοιτητικές μελέτες της Σαιξπηρικής λογοτεχνίας (Sumpter, 2001).

Η καθιέρωση όμως της ανάλυσης περιεχομένου στον τομέα των επιστημών, σύμφωνα με την Κυριαζή οφείλεται κατά ένα μεγάλο μέρος στην ολοκληρωμένη μέθοδο των Berrelson & Lazarsfeld, *The Analysis Communication Content*, η οποία εκδόθηκε το 1948 από τα Πανεπιστήμια του Σικάγου και της Νέας Υόρκης αλλά και στη μετέπειτα ξεχωριστή συμβολή του Berelson με τίτλο «*Content Analysis in Communication Research*» το 1952 (Κυριαζή, 2009). Αποτέλεσε, επιπροσθέτως, και ένα πολύτιμο εργαλείο αναγνώρισης της συγκεκριμένης τεχνικής στην κοινωνική επιστήμη αλλά και στους ερευνητές των Μέσων Επικοινωνίας.

Ο Berrelson στην έκδοση του 1952 έδωσε έναν πρώτο ορισμό της ανάλυσης περιεχομένου ως τη «μέθοδο χρήσης των δεδομένων των μέσων επικοινωνίας για την εξέταση επιστημονικών υποθέσεων και αξιολόγηση των περιοδικών» (Burns & Berelson, 1953).

Για τον Weber, η ανάλυση περιεχομένου ορίζεται ως «μία μέθοδος έρευνας η οποία χρησιμοποιεί συγκεκριμένους κανόνες για να εξάγει τα απαραίτητα συμπεράσματα μέσω της ανάλυσης γραπτών κειμένων» (Weber, 1990, p. 117).

Για τον Neuendorff ορίζεται ως «*η συνοπτική και ποσοτική ανάλυση των μηνυμάτων η οποία βασίζεται σε επιστημονική μέθοδο (συμπεριλαμβανομένης της υποκειμενικότητας, διυποκειμενικότητας, θεωρητικού σχεδιασμού, αξιοπιστίας, εγκυρότητας, καθολικότητας, αναπαραγωγικότητας και υποθετικών τεστ) και δεν περιορίζεται ως προς τα είδη των μεταβλητών οι οποίες μπορούν να μετρηθούν ή στο πλαίσιο όπου δημιουργούνται ή παρουσιάζονται τα μηνύματα*» (Neuendorf, 2002).

Για τον Krippendorff η ανάλυση περιεχομένου είναι «*μία ερευνητική τεχνική προκειμένου να δημιουργηθούν παράγωγα και αξιόπιστα συμπεράσματα από κείμενα (ή άλλα αξιολογικά θέματα) στα πλαίσια της χρήσης τους*» (Krippendorff, 2004).

Ενώ οι Riff, Lacy, Watson και Fico προσδιορίζουν την ανάλυση περιεχομένου ως ένα επιστημονικό εργαλείο το οποίο υιοθετεί την γνώση και την οπτική της κοινωνικής επιστήμης δηλαδή το σύνολο των προτύπων και των κανόνων προκειμένου να διατυπώσει σχεσιακές εκφράσεις οι οποίες περιγράφουν και εξηγούν τις ανθρώπινες συμπεριφορές και διαδικασίες (Riffe et al., 2019).

Οι ερευνητικές προσεγγίσεις που ακολουθούν την συγκεκριμένη μέθοδο βασίζονται στα θέματα που διερευνώνται στο κείμενο, στη σύγκριση μεταξύ τους και σε διαφορετικά στοιχεία του υπό διερεύνηση κειμένου τα οποία θα βοηθήσουν τον ερευνητή να διατυπώσει το μήνυμα του κειμένου με βάση τις κατηγορίες που εκείνος θα επιλέξει (συστηματική διερεύνηση). Άρα, μέσω της ανάλυσης κειμένου κωδικοποιούνται γραπτά και προφορικά κείμενα τα οποία έχουν αρκετά κοινά χαρακτηριστικά με τις ερωτήσεις που διατυπώνονται στα ερωτηματολόγια αλλά και στις άτυπες συνεντεύξεις (Κυριαζή, 2009).

Με βάση τα όσα αναφέρθηκαν παραπάνω η ανάλυση περιεχομένου διακρίνεται σε ποσοτική και ποιοτική. Οι έρευνες που έγιναν από τις αρχές του 20ου αιώνα έως και σήμερα χρησιμοποιούσαν είτε τη μία από τις δύο οπτικές ή και τις δύο. Από το 1920 έως το 1950 οι έρευνες ανάλυσης περιεχομένου ασχολήθηκαν με διαφορετικές θεματικές όπως πολιτική, εγκλήματα, κοινωνικά ζητήματα.

Ενδεικτικά και με βάση τον Krippendorff η πρώτη ποσοτική ανάλυση έγινε το 1893 σε εφημερίδες και ακολούθως το 1910 όπου οι τότε ερευνητές θέλησαν να δείξουν τον χώρο που καταλάμβαναν οι ήσσονος σημασίας ειδήσεις στην Αμερική όπως σπορ και σκάνδαλα, σε σχέση με το πραγματικό ρεπορτάζ (επιστημονικές ειδήσεις). Τη δεκαετία του μεσοπολέμου '30-'40 και με την εμφάνιση των κοινωνικών επιστημών η ποσοτική ανάλυση μετατράπηκε σε ποσοτική ανάλυση περιεχομένου και διευρύνθηκε στον τομέα των ερωτηματολογίων και δημοσκοπήσεων μέσω και των στατιστικών εργαλείων τα οποία υιοθέτησαν από άλλες επιστήμες όπως η ψυχολογία. Με την εμφάνιση των νέων τότε μέσων ενημέρωσης όπως η τηλεόραση και το ραδιόφωνο αλλά και τους μετέπειτα παγκόσμιους πολέμους (Α' και Β'), η ανάλυση περιεχομένου εστιάστηκε στις προπαγανδιστικές ομιλίες των τότε απολυταρχικών καθεστώτων (Χίτλερ, Στάλιν). Στόχος ήταν η έγκαιρη πρόβλεψη των αλλαγών που συντελούνταν στα Ναζιστικά στρατεύματα καθώς και στην ιεραρχία τους με κύριο παράδειγμα την ανάλυση της ομιλίας του Γκέμπελς από τους Βρετανούς επιστήμονες για τα όπλα που επρόκειτο να χρησιμοποιηθούν εναντίον της χώρας τους (Krippendorff, 2004).

Από τη δεκαετία του '80 και μετά και με την έλευση των ΤΠΕ (υπολογιστές, ίντερνετ, κοινωνικά δίκτυα) αλλά και την ανάπτυξη των βάσεων δεδομένων, η ανάλυση περιεχομένου διευρύνθηκε στον συγκεκριμένο τομέα πολύ γρήγορα μέσω ειδικών προγραμμάτων ανάλυσης μηνυμάτων κειμένου σε ηλεκτρονικό υπολογιστή (CATA). Έκτοτε η ανάλυση περιεχομένου υιοθετήθηκε ως μέθοδος και από άλλους κλάδους όπως η ανθρωπολογία και η επιστήμη της πληροφόρησης.

Ο Chu σε μελέτη του 2015 αναφέρει την ανάλυση περιεχομένου ως την πλέον διαδεδομένη μέθοδο στον τομέα της επιστήμης της πληροφόρησης πέραν των θεωρητικών προσεγγίσεων και των πειραμάτων, αντικαθιστώντας τη χρήση ερωτηματολογίων (Chu, 2015). Το ίδιο έτος η έρευνα των Al-Daihani και Alawadhi εστιάστηκε σε ανάλυση και μελέτη του περιεχομένου των αναρτήσεων (tweets) των λογαριασμών των ακαδημαϊκών βιβλιοθηκών στο twitter (Al-Daihani & AlAwadhi, 2015).

Ενώ οι Franz et al, χρησιμοποίησαν την ανάλυση περιεχομένου για την ανάλυση των αναρτήσεων των χρηστών του Facebook (Franz et al., 2019).

Η Lim τη χρησιμοποίησε για να αναλύσει οδηγούς ακαδημαϊκών βιβλιοθηκών προκειμένου να αναδείξει τις στρατηγικές που έχουν υιοθετήσει οι βιβλιοθηκονόμοι για τον έλεγχο των ψευδών ειδήσεων (Lim, 2020).

Οι Jarvelin και Vakkarī χρησιμοποίησαν την ανάλυση περιεχομένου για τη μελέτη της εξέλιξης της Επιστήμης της Βιβλιοθηκονομίας σε άρθρα περιοδικών κατά τα έτη 1965 έως 2015 (Järvelin & Vakkarī, 2021).

Οι Sahoo et al, ακολούθησαν τη μέθοδο ανάλυσης περιεχομένου για ποιοτική και ποσοτική αξιολόγηση της βιβλιογραφίας κατά τα έτη 2017-2020, με ευρεία κάλυψη θεμάτων όπως πληροφοριακή παιδεία, εντοπισμό ψευδών ειδήσεων, θεωρίες, μοντέλα, κοινωνικά μέσα δικτύωσης (Sahoo et al., 2021).

3.2.3 Χαρακτηριστικά Ανάλυσης Περιεχομένου: Θεματική Ανάλυση

Σκοπός της ανάλυσης περιεχομένου είναι να προσδιορίσει τα εξής χαρακτηριστικά: περιεχόμενο (ανάλυση & κατηγοριοποίηση), πομπός επικοινωνίας (παράγωγο επικοινωνίας π.χ. έργο τέχνης), δέκτης (προσωπικά χαρακτηριστικά αποδέκτη π.χ. συναισθήματα, προσωπικότητα, αντίδραση) (Τζάνη & Κεχαγιάς, 2005).

Εκτός από τα χαρακτηριστικά που αναφέρθηκαν η συγκεκριμένη μέθοδος χωρίζεται σε τρεις κατηγορίες: τη **λεξικολογική**, τη **φραστική** και τη **θεματική ανάλυση**.

Στην παρούσα διπλωματική εργασία η κατηγοριοποίηση των άρθρων υλοποιήθηκε βάση της **θεματικής ανάλυσης** περιεχομένου.

Σύμφωνα με τους Kiger και Vargio η θεματική ανάλυση έχει ερμηνευτεί από πολλές επιστημονικές μελέτες ως μία «ευέλικτη ποιοτική μέθοδος η οποία περιγράφει τα δεδομένα, τα κωδικοποιεί και δημιουργεί κοινά θέματα» (Aronson, 1995; Attride-Stirling, 2001; Boyatzis, 1998; Braun & Clarke, 2006; Merton, 1975). Τα θέματα αποτελούν ενεργά μοτίβα τα οποία προέρχονται από τα δεδομένα και απαντούν σε ερευνητικές υποθέσεις και τα οποία

δημιουργούνται με βάση τα δεδομένα (επαγωγικά) ή παραπομπή σε προηγούμενες θεωρίες, ή μελέτες (συμπερασματικά) (Braun & Clarke, 2012; Kiger & Varpio, 2020; Varpio et al., 2020).

Σύμφωνα με την Clarke η θεματική ανάλυση εκτός από θέματα παράγει και κωδικοποιημένες πληροφορίες (κώδικες) οι οποίες αποτελούν μικρότερες μονάδες δεδομένων. Οι «κώδικες» συμβάλλουν στη δημιουργία των θεμάτων (ευρύτερων εννοιών) μίας κεντρικής κοινής έννοιας. Στόχος της θεματικής ανάλυσης είναι όχι μόνο η συμπύκνωση των δεδομένων αλλά η ερμηνεία και ο προσδιορισμός βασικών στοιχείων από αυτά (Clarke & Braun, 2017).

Η θεματική ανάλυση περιεχομένου όπως και η απλή ανάλυση περιεχομένου εφαρμόστηκε κατά καιρούς σε αρκετές έρευνες.

Οι S. Wang et al, χρησιμοποίησαν την θεματική ανάλυση περιεχομένου για ποσοτική μελέτη άρθρων του διεπιστημονικού περιοδικού Πληροφόρηση, Γνώση και Διαχείριση (S. Wang et al., 2018).

Αντίστοιχη μελέτη πραγματοποίησαν και οι Shemaieva και Shevtsova σε θεματική ανάλυση Ουκρανικών περιοδικών της επιστήμης της πληροφόρησης (Shevtsova & Shemaieva, 2020).

Ο Kosciejew ανέλυσε θεματικά τις αναφορές των πολιτικών των βιβλιοθηκών για τον Covid19 και στη συνέχεια σύγκρινε τα θέματα από όλες τις αναφορές για τον προσδιορισμό των ομοιοτήτων και των διαφορών ως προς το περιεχόμενο (Kosciejew, 2021).

Η θεματική ανάλυση περιεχομένου υιοθετήθηκε από τους Zuiderwijk et al, προκειμένου να αναλύσει τα κοινά χαρακτηριστικά των ανοιχτών δεδομένων (Zuiderwijk et al., 2020).

Η Hill τη χρησιμοποίησε για την ανάλυση των υπηρεσιών τριών βιβλιοθηκονομικών οργανώσεων σε σχέση με τα άτομα με αναπηρία καθώς και τα περιφερειακά θέματα που προέκυψαν κατά τη διάρκεια της έρευνας (Hill, 2021).

Κεφάλαιο 4. Αποτελέσματα: Κύρια Χαρακτηριστικά Εντοπισμού Ψευδών Ειδήσεων

Σε αυτό το κεφάλαιο παρουσιάζονται τα κυριότερα χαρακτηριστικά που χρησιμοποιούν διάφοροι αλγόριθμοι, μοντέλα και συστήματα για τον εντοπισμό των ψευδών ειδήσεων. Μετά την ανάλυση περιεχομένου της κάθε σχετικής πηγής που επιλέχθηκε για μελέτη, έγινε μία προσπάθεια ώστε τα χαρακτηριστικά αυτά να παρουσιαστούν ενταγμένα σε συγκεκριμένες κατηγορίες³.

Αρχικά, τα χαρακτηριστικά εντάσσονται σε δύο γενικές κατηγορίες: (Α) **στο περιεχόμενο ειδήσεων** και (Β) **στο κοινωνικό περιεχόμενο**. Η συγκεκριμένη κατηγοριοποίηση βασίζεται στο άρθρο των Shu et al., (Shu et al., 2017) το οποίο αποτέλεσε τη βάση για την περαιτέρω ανάπτυξη της κατηγοριοποίησης που πραγματοποιήθηκε στο πλαίσιο της παρούσας διπλωματικής. Ο λόγος που επιλέχθηκε ως βάση το εν λόγω άρθρο είναι όχι μόνο γιατί αποτυπώνει και εξηγεί με εξαιρετική ευκρίνεια ποικίλα χαρακτηριστικά που φέρουν μοντέλα, αλγόριθμοι και συστήματα εντοπισμού ψευδών ειδήσεων, αλλά γιατί οι ερευνητές του συγκεκριμένου πεδίου επίσης το αναφέρουν πολύ συχνά (Abonizio et al., 2020; Ahmad et al., 2020; Amoruso et al., 2020; Barrón-Cedeño et al., 2019; Braşoveanu & Andonie, 2021; Gravanis et al., 2019; Kaliyar et al., 2020; Kar, 2020; Ksieniewicz et al., 2019; Y. Liu et al., 2019; Mahabub, 2020; Ozbay & Alatas, 2020; Posadas-Durán et al., 2019; S. Qian et al., 2021; Reis et al., 2019; Shu, Wang, et al., 2019; Shu, Cui, et al., 2019; Sotirakou et al., 2021; Souza Freire et al., 2021; Varshney & Vishwakarma, 2021; Weiss et al., 2019; Ying et al., 2021; Y. Yu et al., 2020; Yuliani et al., 2019).

Ειδικότερα, η κατηγορία (Α) **περιεχόμενο ειδήσεων** αναφέρεται στον εντοπισμό της ψευδούς πληροφόρησης που προκύπτει μετά από την ανάλυση του περιεχομένου ενός κειμένου ή/και εικόνας. Μάλιστα, σύμφωνα με τους Hangloo και Arora (Hangloo & Arora, 2021) για τον εντοπισμό των ψευδών ειδήσεων αρκετές έρευνες έχουν επικεντρωθεί είτε σε *μη παρατηρήσιμα χαρακτηριστικά* (latent features)⁴ είτε σε *χαρακτηριστικά που έχουν*

³ Ο αναλυτικός πίνακας με τα χαρακτηριστικά παρατίθεται στο παράρτημα Β της παρούσας διπλωματικής εργασίας.

⁴ Μη παρατηρήσιμα (Latent features): η σημασιολογία των συγκεκριμένων χαρακτηριστικών έχει υιοθετηθεί από τις μη παρατηρήσιμες μετρικές των κοινωνικών επιστημών. Πρόκειται για δεδομένα των οποίων η σημασία δεν είναι άμεσα παρατηρήσιμη από το κείμενο αλλά «υπαινίσσονται» (μέσα από μαθηματικά μοντέλα) από τις υπόλοιπες μεταβλητές. Τα μη παρατηρήσιμα χαρακτηριστικά χρησιμοποιούνται για την απομείωση των δεδομένων. Παραδείγματα μοντέλων που χρησιμοποιούν μη ορατά χαρακτηριστικά θεωρούνται η *Ανάλυση Κύριας Συνιστώσας* (Principal Component Analysis) (χρήση βασικών στοιχείων δεδομένων), *Λανθάνουσα Σημασιολογική Ανάλυση* (Latent Semantic analysis) (ανάλυση λέξεων και μερών κειμένου χρήση σε Φυσική Επεξεργασία Γλώσσας^(NLP) κ.α.

εξαχθεί από άλλα χαρακτηριστικά (hand-crafted features)⁵ μέσα σε κείμενα και εικόνες. Τα χαρακτηριστικά που εντοπίζονται εντός του κειμένου/ή και εικόνας εντάσσονται σε τέσσερις (4) υπο-κατηγορίες : **Χαρακτηριστικά γνώσης (Knowledge-based)**, **Στυλομετρικά χαρακτηριστικά** (βασιζόμενα στη γραφή του συγγραφέα και στον τρόπο που επηρεάζει το κοινό δηλαδή επίπεδο σύνταξης, λεξικού-bias lexicon, υποκειμενικότητας, σημασιολογίας, αναπαραστάσεων κειμένου), **Γλωσσολογικά** (επίπεδο λεκτικό, μετρικές πολυπλοκότητας, αναγνωσιμότητας, τεχνικές NLP), **Οπτικά** (Shu et al., 2017), οι οποίες θα αναλυθούν παρακάτω.

Εκτός από τις κατηγορίες που αναφέρθηκαν παραπάνω θα παρουσιαστούν και τρία επιπλέον χαρακτηριστικά τα οποία προκύπτουν από τη διερεύνηση της τρέχουσας βιβλιογραφίας και είναι **τα συναισθηματικά χαρακτηριστικά** (Barrón-Cedeño et al., 2019; Braşoveanu & Andonie, 2021; Cohen et al., 2020; B. D. Horne et al., 2020; Kausar et al., 2020; Meel & Vishwakarma, 2020; Papadopoulou et al., 2019; Reis et al., 2019; Souza Freire et al., 2021) , **τα υβριδικά χαρακτηριστικά** (Baeth & Aktas, 2019; Gôlo et al., 2021; S. Qian et al., 2021; Shu, Cui, et al., 2019; Ying et al., 2021), και **τα θεματικά χαρακτηριστικά** (Du et al., 2021).

4.1 Χαρακτηριστικά Γνώσης

Με βάση τη βιβλιογραφία, μία από τις βασικές κατηγορίες χαρακτηριστικών που εντοπίζονται σε αλγόριθμους, μοντέλα και συστήματα εντοπισμού ψευδών ειδήσεων είναι τα λεγόμενα χαρακτηριστικά «γνώσης» (knowledge based). Στην ουσία τα χαρακτηριστικά «γνώσης» θα μπορούσαν να περιγραφούν ως μία προσέγγιση που υιοθετεί τη μέθοδο ελέγχου γεγονότων (fact-checking) κατά την οποία ένας ισχυρισμός συγκρίνεται με εξωτερικές πηγές προκειμένου να πιστοποιηθεί η αυθεντικότητά του. Η μέθοδος ελέγχου γεγονότων περιλαμβάνει τρεις βασικές υποκατηγορίες: **(α)** τον έλεγχο από ειδικούς/επαγγελματίες (manual fact checking) **(β)** τον έλεγχο βάση πληθοπορισμού (crowdsourcing)⁶ και **(γ)** τον αυτοματοποιημένο έλεγχο (automatic fact-checking). Ειδικότερα

⁵ handcrafted features: τα συγκεκριμένα χαρακτηριστικά αφορούν χαρακτηριστικά εικόνας τα οποία εξάγονται με αλγόριθμους (π.χ. άκρες ή γωνίες εικόνας). Παράδειγμα χρήσης είναι η αναγνώριση εικόνας (Nanni et al., 2017).

⁶ πληθοπορισμός: η εξωτερική ανάθεση καθηκόντων σε ένα άτομο ή μία ομάδα ατόμων από έναν οργανισμό, άτομο ή ίδρυμα προκειμένου να φέρουν εις πέρας μία εργασία. Η εργασία εξαρτάται από την πολυπλοκότητα και τα άτομα εργάζονται βάση γνώσεων, εμπειρίας, χρημάτων για προσωπική ανέλιξη, ικανοποίηση αναγκών, ατομικών δεξιοτήτων, οικονομικών αιτιών (Estellés-Arolas & González-Ladrón-de-Guevara, 2012, p. 7). Παραδείγματα χρήσης πληθοπορισμού συναντώνται στην ανάπτυξη νέων τεχνολογιών και στην υλοποίηση ενός αλγόριθμου. Πιο συγκεκριμένα το Amazon Mechanical Turk και τα Wikis θεωρούνται παραδείγματα πληθοπορισμού.

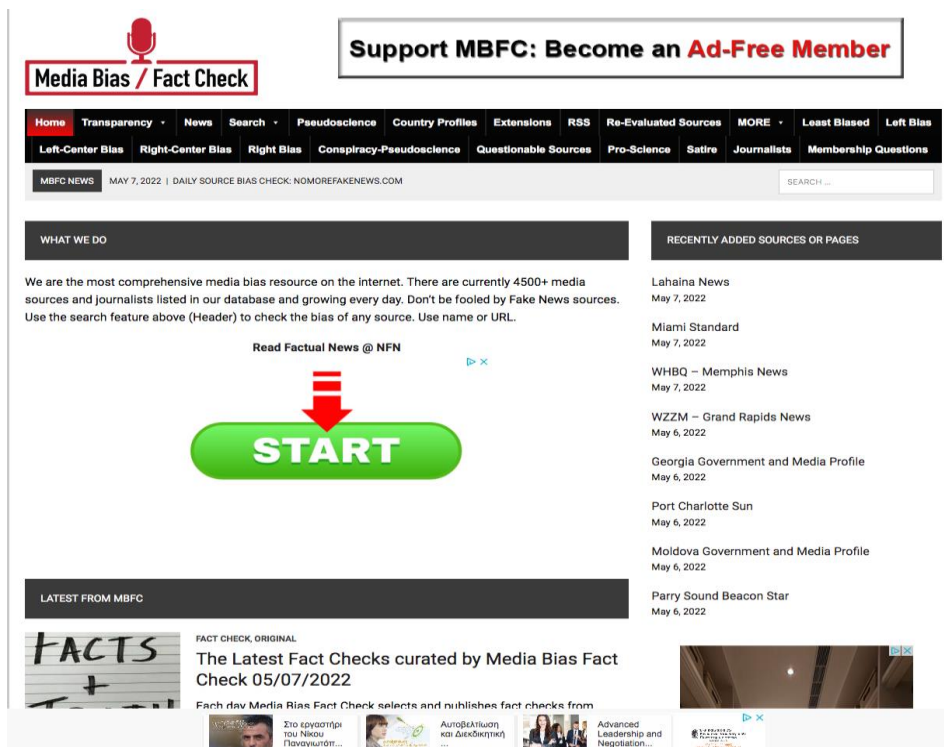
η τρίτη υποκατηγορία αφορά τα ακόλουθα τρία στοιχεία: 1. αυτοματοποιημένες ιστοσελίδες ελέγχου γεγονότων (π.χ. Claimbuster, Hoaxy), 2. σύγκριση γεγονότων με γνωσιακή βάση (π.χ. FEVER) και δημιουργία αυτόματων μοντέλων προσδιορισμού παραπληροφόρησης με χρήση αλγορίθμων μηχανικής μάθησης και 3. την εξαγωγή γνώσης από τον ιστό (knowledge ή fact extraction) (Shu et al., 2017; X. Zhou & Zafarani, 2020), τα οποία θα αναλυθούν εκτενώς στις επόμενες υποενότητες.

4.1.1 Έλεγχος από Ειδικούς (Expert Based)

Στην **προσέγγιση ειδικών (experts-based)** υιοθετείται η οπτική επαγγελματιών οι οποίοι εργάζονται σε συγκεκριμένους τομείς για τη λήψη των αποφάσεων (de Beer & Matthee, 2021). Ιστοσελίδες όπως το Politifact (*PolitiFact*, n.d.), Snopes (*Snopes.Com*, n.d.), FullFact (*Full Fact - Full Fact Is the UK's Independent Fact Checking Organisation*, n.d.), MediaBias Fact Check (*Media Bias/Fact Check News*, n.d.), Newsguard (*NewsGuard - Combating Misinformation with Trust Ratings for News*, n.d.), Maldito Buló (*Portada Maldita.Es - Periodismo Para Que No Te La Cuelen*, n.d.), SNU Fact Check (*Factcheck*, n.d.) αλλά και τα Ellinika Hoaxes (*Ellinika Hoaxes*, n.d.) υιοθετούν τη συγκεκριμένη οπτική.

Η προσέγγιση αυτή αναφέρεται στα περισσότερα άρθρα της βιβλιογραφίας παράλληλα με αλγόριθμους μηχανικής μάθησης (Barrón-Cedeño et al., 2019; Braşoveanu & Andonie, 2021; B. D. Horne et al., 2020; Isaakidou et al., 2021; Kausar et al., 2020; Khanday et al., 2021; Posadas-Durán et al., 2019; Reis et al., 2019; Shim et al., 2021; Shu, Wang, et al., 2019; Silva et al., 2020; Sotirakou et al., 2021; Souza Freire et al., 2021; Varshney & Vishwakarma, 2021; Vishwakarma et al., 2019).

Στο **Media Bias Factcheck** (Εικόνα 22) χρησιμοποιούνται μετρικές κλίμακας για την μεροληψία. Η συγκεκριμένη οπτική υιοθετεί μία κλίμακα με την εισαγωγή μιας κίτρινης τελείας ως δείκτη για συγκεκριμένες ιστοσελίδες που υιοθετούν μεροληπτικές απόψεις. Οι μετρικές μεροληψίας εστιάζονται στην πολιτική τοποθέτηση, δηλαδή αριστερός, δεξιός, κεντρώος, στο κατά πόσο αληθής είναι η πληροφόρηση, ή και τα δύο (Εικόνα 23).



Εικόνα 22. Στιγμιότυπο από MediaBias FactCheck.



Εικόνα 23. Μετρικές και κίτρινος δείκτης επισήμανσης πολιτικής τοποθέτησης.

Οι συγκεκριμένες μετρήσεις έχουν υιοθετηθεί από την πολιτική κλίμακα των Η.Π.Α. και είναι επικεντρωμένες σε 4 βασικές κατηγορίες:

- **Μεροληπτικές λέξεις/τίτλοι (Biased wording/headlines):** Η πηγή χρησιμοποιεί μεροληπτικές λέξεις για να προκαλέσει το συναίσθημα του αναγνώστη (Οι τίτλοι ταιριάζουν με την ιστορία;)
- **Αληθοφάνεια/Πηγή (Factual/Sourcing):** Έμφαση στις αναφορές της πηγής (τεκμηρίωση) και υποστήριξη των ισχυρισμών με αποδεικτικά στοιχεία (έγκυρες πηγές). Στη συγκεκριμένη κατηγορία δίδεται μία αναφορά «εγκυρότητας» (βαθμολογία) ως προς τις πηγές (πολύ υψηλή, υψηλή, κυρίως πραγματική, μικτή, χαμηλή, πολύ χαμηλή)
- **Επιλογές ιστορίας (Story choices):** Η πηγή αναφέρει την είδηση αντικειμενικά (και από τις δύο πλευρές) ή μονομερώς;

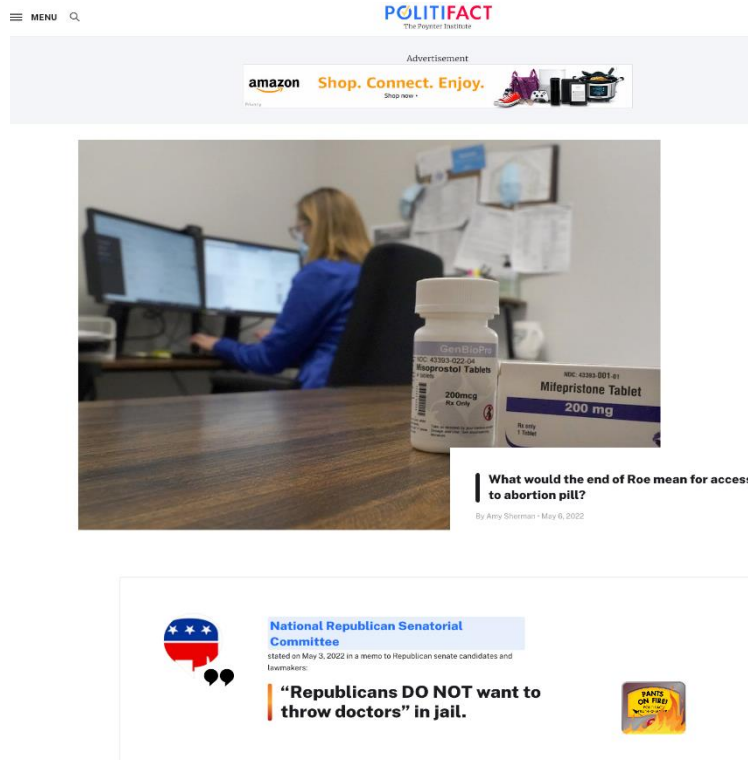
- **Πολιτική σχέση (Political affiliation):** Πόσο ευνοεί μία συγκεκριμένη πολιτική ιδεολογία; Πόσο ακραίες είναι οι απόψεις;

Η πηγή αξιολογείται από 0-10 όπου 0 η μη μεροληπτική και 10 το μέγιστο της μεροληψίας (*Methodology*, n.d.). Με βάση την πολιτική τοποθέτηση η βαθμολογημένη κλίμακα είναι η εξής: 0-2: λιγότερο μεροληπτική, 2-5: μεροληψία κεντροαριστερού/κεντροδεξιού, 5-8: μεροληψία αριστερού/δεξιού, 8-10: μέγιστη μεροληψία.

Με βάση την «αναφορά τεκμηρίωσης» από την κατηγορία *factual/sourcing* οι αξιολογήσεις είναι οι ακόλουθες:

- **Πολύ υψηλή:** αν η αξιολόγηση είναι 0 τότε η πηγή είναι πάντα αληθής, οι πηγές πληροφόρησης που παραθέτει είναι αξιόπιστες, διορθώνει εσφαλμένες πληροφορίες, δεν έχει αποτύχει σε κάθε έλεγχο αναφοράς ειδήσεων
- **Υψηλή (αξιολόγηση 1-2):** η πηγή είναι πάντα αξιόπιστη, οι πηγές πληροφόρησης που παραθέτει έχουν χαμηλή μεροληψία ή υψηλή αξιόπιστη πληροφόρηση, διορθώνει εσφαλμένες πληροφορίες, έχει αποτύχει μία φορά σε έλεγχο αναφοράς ή δύο
- **Κυρίως πραγματική (αξιολόγηση 3-4):** η πηγή είναι συνήθως αξιόπιστη αλλά έχει αποτύχει σε έλεγχο αξιοπιστίας (1 ή 2 φορές). Χρησιμοποιεί κατά περιόδους μεροληπτικές πηγές που όμως οδηγούν σε αξιόπιστο περιεχόμενο τις περισσότερες φορές
- **Μεικτή (αξιολόγηση 5-6):** η πηγή δεν χρησιμοποιεί πάντα αξιόπιστες πηγές ή οδηγεί σε μεροληπτικές/μεικτές πηγές. Οι μεικτές πηγές έχουν αποτύχει σε ένα ή περισσότερους ελέγχους αξιοπιστίας και δεν διορθώνουν λανθασμένες ή παραποιημένες πληροφορίες
- **Χαμηλή (αξιολόγηση 7-9):** η πηγή σπάνια χρησιμοποιεί αξιόπιστες πηγές και δεν είναι έμπιστη για αξιόπιστη πληροφόρηση. Αυτές οι πηγές πρέπει να ελεγχθούν για σκόπιμη διασπορά ψευδών ειδήσεων, προπαγάνδα, θεωρίες συνωμοσίας
- **Πολύ Χαμηλή (αξιολόγηση 10):** η πηγή σπάνια χρησιμοποιεί αξιόπιστες πηγές και δεν είναι έμπιστη για αξιόπιστη πληροφόρηση. Οι πηγές πρέπει να ελεγχθούν για σκόπιμη διασπορά ψευδών ειδήσεων, προπαγάνδα, θεωρίες συνωμοσίας

Στο *PolitiFact* (Drobnic-Holan, 2018) (Εικόνα 24) υιοθετείται η «μέτρηση αληθείας» (*Truth-O-Meter*) σχετικά με τους ισχυρισμούς (*claims*) που διαδίδονται στο διαδίκτυο και επιλέγονται οι πιο αξιόλογοι και σημαντικοί.



Εικόνα 24. Στιγμιότυπο από Politifact.

Για τον έλεγχο και την επιλογή των ισχυρισμών εξετάζονται τα παρακάτω ερωτήματα:

- Είναι ο ισχυρισμός βασισμένος σε επιβεβαιωμένο γεγονός; (Δεν ελέγχονται γνώμες δεδομένου ότι στην ομιλία και στην πολιτική ρητορική διακρίνεται η έννοια της υπερβολής)
- Η συγκεκριμένη δήλωση φαίνεται λανθασμένη ή παραπλανητική;
- Είναι η δήλωση αξιόλογη; (μικρά λάθη στην ομιλία αποφεύγονται να ελεγχθούν)
- Η δήλωση μπορεί να μεταφερθεί και να αναπαραχθεί από άλλους;
- Μπορεί ένας συνηθισμένος άνθρωπος να ακούσει ή να διαβάσει τη δήλωση και να αναρωτηθεί: «Αληθεύει;»

Η επιλογή των δηλώσεων γίνεται βάση των θεμάτων των ειδήσεων ενώ η επιλογή των γεγονότων πραγματοποιείται και από τις δύο πολιτικές παρατάξεις των Η.Π.Α. (Δημοκρατικοί και Συντηρητικοί). Επιπλέον, τις περισσότερες φορές ελέγχεται το κόμμα εξουσίας ή τα άτομα που επανειλημμένα κάνουν δηλώσεις για να παραπλανήσουν ή να τραβήξουν την προσοχή.

Η διαδικασία ελέγχου είναι η ακόλουθη: ανασκόπηση ως προς το τι έχουν βρει οι προηγούμενοι αξιολογητές, ενδελεχή έρευνα στη μηχανή αναζήτησης Google, έρευνα σε online βάσεις δεδομένων, συμβουλευτική με διαφορετικούς ειδικούς, ανασκόπηση των δημοσιεύσεων και μια τελική ολική ανασκόπηση των διαθέσιμων στοιχείων.

Η κλίμακα προσδιορισμού του «μέτρου αληθείας» αντανακλά την ακρίβεια της δήλωσης και περιλαμβάνει 6 βαθμολογήσεις:

- **Αληθές (True):** ακρίβεια δήλωσης και δεν υπάρχει κάτι ελλιπές
- **Σχεδόν αληθές (Mostly True):** η δήλωση είναι αληθής αλλά χρειάζεται διευκρίνιση ή πρόσθετη πληροφόρηση
- **Κατά το ήμισυ αληθές (Half-True):** η δήλωση είναι σχεδόν αληθής αλλά παραλείπει σημαντικές λεπτομέρειες ή λαμβάνει υπόψη γεγονότα εκτός πλαισίου
- **Σχεδόν λανθασμένη (Mostly False):** η δήλωση περιλαμβάνει ένα μέρος της αλήθειας αλλά αγνοεί σημαντικά γεγονότα που θα έδιναν διαφορετική εντύπωση
- **Λανθασμένη (False):** η δήλωση δεν είναι αληθής
- **Ψέματα (Pants on Fire):** η δήλωση δεν είναι αληθής και δημιουργεί έναν εξωπραγματικό ισχυρισμό

Η βαθμολογία της δήλωσης βασίζεται στην πληροφορία που είναι γνωστή τη στιγμή που γίνεται η δήλωση και βαρύνει τον ομιλητή. Ο δημοσιογράφος που ερευνά και γράφει την αξιολόγηση προτείνει μία κλίμακα όταν η αναφορά επιστρέφεται στον επιβλέποντα επιμελητή. Ο επιμελητής και ο δημοσιογράφος αξιολογούν μαζί την αναφορά, αποσαφηνίζουν και προσθέτουν λεπτομέρειες και συμφωνούν για την βαθμολόγηση. Μετέπειτα, ο επιβλέπων επιμελητής δίνει τον έλεγχο των γεγονότων σε δύο επιπλέον επιμελητές.

Οι τρεις επιμελητές και ο δημοσιογράφος αξιολογούν τον έλεγχο μέσω συζήτησης των ερωτήσεων:

- Είναι η δήλωση πραγματικά αληθής;
- Η δήλωση μπορεί να αποτυπωθεί διαφορετικά; Μπορεί να ερμηνευτεί;
- Μπορεί ο ομιλητής να παρέχει αποδείξεις; Ο ομιλητής απέδειξε ότι η δήλωση είναι αληθής;
- Πώς έγινε η διαχείριση παρόμοιων δηλώσεων στο παρελθόν; Ποια είναι η νομική διαδικασία του Politifact;

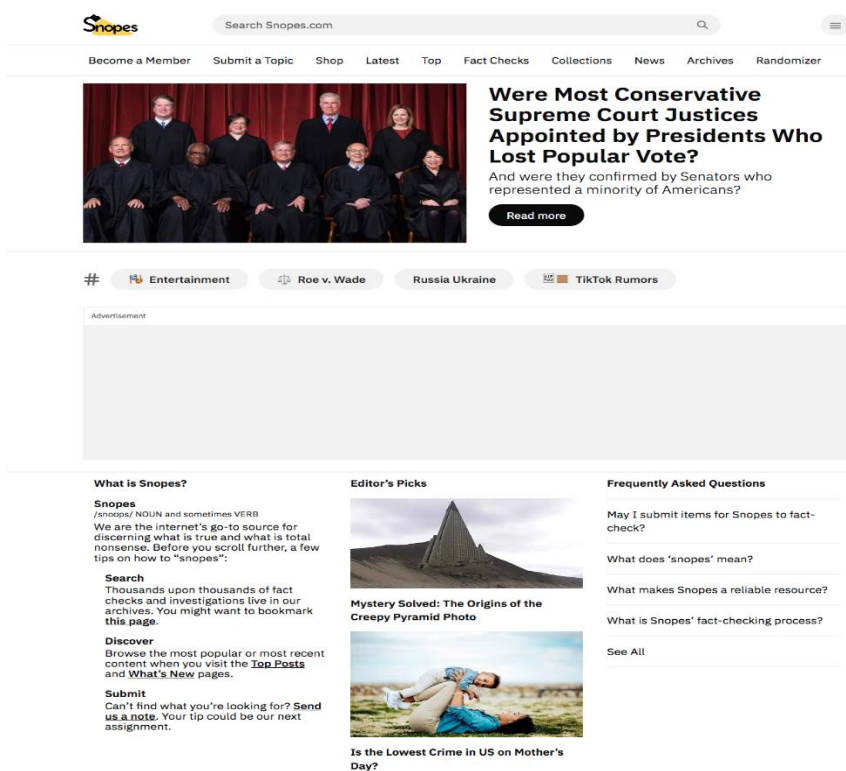
Στη συνέχεια, οι τρεις επιμελητές ψηφίζουν για την βαθμολόγηση (δύο ψήφοι καθορίζουν την απόφαση), κάποιες φορές αφήνοντάς τη όπως πρότεινε ο δημοσιογράφος και άλλες φορές αλλάζοντας τη σε άλλη βαθμολόγηση. Πραγματοποιούνται περισσότερες τροποποιήσεις και κατόπιν η αναφορά δημοσιεύεται.

Στην ίδια περίπτωση οπτική κινείται και το **Snores** (*Our Fact Check Ratings, Explained*, n.d.) (Εικόνα 25) το οποίο ελέγχει τους ισχυρισμούς με βάση την ακόλουθη κλίμακα:

- **Αληθές (True):** τα αρχικά στοιχεία των ισχυρισμών είναι αληθή
- **Σχεδόν αληθές (Mostly True):** τα αρχικά στοιχεία των ισχυρισμών είναι αληθή αλλά πρόσθετες λεπτομέρειες γύρω από τον ισχυρισμό είναι αναληθείς
- **Μεικτό (Mixture):** ο ισχυρισμός έχει κάποια στοιχεία που είναι αληθή και κάποια ψευδή έτσι ώστε να μην μπορεί να αποτυπωθεί από καμία άλλη βαθμολογία
- **Κυρίως Ψευδές (Mostly false):** τα κύρια στοιχεία ενός ισχυρισμού είναι ψευδή αλλά κάποιες αλλά πρόσθετες λεπτομέρειες γύρω από τον ισχυρισμό είναι αληθείς
- **Ψευδές (False):** τα κύρια στοιχεία ενός ισχυρισμού είναι ψευδή
- **Μη αποδεδειγμένο (Unproven):** δεν υπάρχουν επαρκή στοιχεία ότι ο ισχυρισμός είναι αληθής αλλά ο ισχυρισμός δεν μπορεί να αποδειχθεί ότι είναι ψεύτικος
- **Εκτός χρονικού πλαισίου (Outdated):** η βαθμολόγηση αφορά σε στοιχεία όπου τα μεταγενέστερα γεγονότα έχουν καταστήσει την πρωτότυπη βαθμολόγηση μη σχετική (π.χ. μία παράγραφος ενός νόμου έχει καταργηθεί)
- **Λανθασμένο (Miscaption):** χρησιμοποιείται σε φωτογραφίες και βίντεο που είναι πραγματικά (όχι προϊόν μερικής ή ολικής ψηφιακής χειραγώγησης) αλλά που κατά κύριο λόγο είναι παραπλανητικά γιατί το συμπληρωματικό υλικό περιγράφει ψευδώς την καταγωγή, το πλαίσιο και τη σημασία
- **Σωστή αναφορά (Correct Attribution):** το αναφερόμενο υλικό (υλικό ή κείμενο) έχει αποδοθεί σωστά στο άτομο που την ανέφερε ή την έγραψε
- **Μύθος (Legend):** γεγονότα που περιγράφονται πολύ γενικά ή είναι ελλιπή ως προς κάποιες λεπτομέρειες, όπως ως προς το αν έχουν συμβεί σε κάποιον, κάποια στιγμή και γι' αυτό είναι αναξιόπιστα
- **Εσφαλμένη απόδοση (Misattributed):** το αναφερόμενο υλικό (ομιλία ή κείμενο) έχει αποδοθεί εσφαλμένα σε ένα άτομο που δεν το είπε ή δεν το έγραψε
- **Απάτη (Scam):** δεν είναι μία αξιολόγηση αλήθειας, αλλά υποδεικνύει ιστοσελίδες που περιγράφουν λεπτομέρειες από επαληθευμένες απάτες
- **Νόμιμο (Legit):** περιγράφει αυθεντικές, νόμιμες προσφορές ή πρωτοβουλίες που είναι διαθέσιμες σε καταναλωτές ή μέλη του κοινού
- **Κατηγοριοποιημένα ως σάτιρα (Labeled satire):** ένας ισχυρισμός προέρχεται από περιεχόμενο που περιγράφεται από τον δημιουργό του ή/και το ευρύτερο κοινό ως σάτιρα. Δεν είναι απαραίτητο να είναι σάτιρα όλο το περιεχόμενο που περιγράφεται από τον δημιουργό ή το κοινό και δεν κάνει διάκριση ανάμεσα στην «πραγματική»

σάτιρα και το περιεχόμενο που μπορεί να μην αναγνωρίζεται αποτελεσματικά ή να μην γίνεται κατανοητό ως σάτιρα παρόλο που χαρακτηρίζεται ως τέτοιο

- **Προέλευση σάτιρας (Originated satire):** αναφέρεται σε περιεχόμενο που προήλθε αρχικά από έναν ιστότοπο που περιγράφεται ως σάτιρα, αλλά αργότερα αφαιρέθηκαν ορισμένες από τις σατιρικές του επισημάνσεις, επαναδιατυπώθηκαν και δημοσιεύτηκαν αλλού. Ισχύει επίσης για περιεχόμενο που δεν χαρακτηρίζεται απαραίτητα ως σάτιρα, αλλά για το κοινό θεωρείται σατιρικό, όπως το περιεχόμενο της «[The Onion](#)»



Εικόνα 25.Στιγμιότυπο από Snopes.

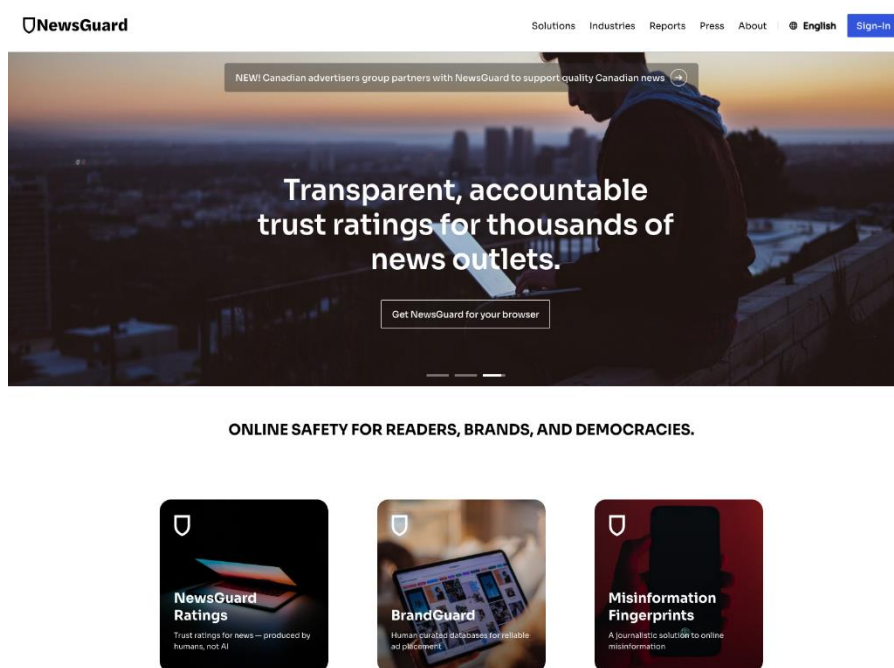
Το News Guard (*Rating Process and Criteria*, 2022) (Εικόνα 26) χρησιμοποιεί 9 κριτήρια για την βαθμολόγηση των πληροφοριών σε μια ιστοσελίδα και παράλληλα μία βαθμολογία εμπιστοσύνης 0-100, αποδεικνύοντας κατά αυτόν τρόπο την αξιοπιστία ή την αναξιοπιστία της. Με βάση την απόδοση κάθε ιστοσελίδα χαρακτηρίζεται με:

Πράσινο: εάν ακολουθεί τα βασικά πρότυπα της αξιοπιστίας και της διαφάνειας (εάν ακολουθεί και τα εννέα, σημειώνεται στη βαθμολόγηση, εάν έχει εξαιρέσεις, σημειώνεται και αυτό).

Κόκκινο: εάν σε γενικές γραμμές αποτύχει να ακολουθήσει τα βασικά πρότυπα αξιοπιστίας και διαφάνειας (εάν η ιστοσελίδα αποτύχει σε αρκετά από τα εννέα ή εάν παραβιάζει δημοσιογραφικά πρότυπα με το να αποτύχει σε σημαντικό αριθμό κριτηρίων)

Σάτιρα: η συγκεκριμένη ιστοσελίδα αν και χρησιμοποιεί ειδήσεις ως βάση της σάτιρας δεν αποτελεί παραδοσιακό μέσο ειδήσεων. Οι ιστοσελίδες σάτιρας δεν βαθμολογούνται αλλά παρέχεται για αυτές μία περιγραφή σχετικά με το ποιος είναι ο δημιουργός τους, εάν αυτό βέβαια είναι δυνατό

Πλατφόρμα: εάν φιλοξενεί πρωτίστως περιεχόμενο απευθείας δημοσιευμένο από χρήστες με περιορισμένο έλεγχο. Οι πληροφορίες τέτοιων ιστοσελίδων μπορεί να είναι ή να μην είναι αξιόπιστες. Οι συγκεκριμένες ιστοσελίδες δεν βαθμολογούνται με βάση τα εννέα κριτήρια αλλά για κάθε μία παρέχεται περιγραφή για τις πρακτικές της



Εικόνα 26. Στιγμιότυπο από NewsGuard.

Κάθε κριτήριο εμπιστοσύνης έχει βαρύνουσα σημασία με βάση τους αριθμητικούς δείκτες (0-100). Όλα τα κριτήρια είναι επιτυχία (pass) - αποτυχία (fail). Δηλαδή, αν λάβει όλους τους βαθμούς που αντιστοιχούν στα κριτήρια ή εάν δεν λάβει κανένα βαθμό για συγκεκριμένο κριτήριο. Μία ιστοσελίδα με βαθμολογία 60 ή υψηλότερη λαμβάνει **πράσινη** αξιολόγηση. Ιστοσελίδα με βαθμολογία κάτω από 60 λαμβάνει **κόκκινη** αξιολόγηση. Σε πλατφόρμες και σατιρικές ιστοσελίδες δίδονται ξεχωριστές περιγραφές και δεν

αξιολογούνται με βάση τα εννέα κριτήρια. Κάθε βαθμολόγηση καταγράφεται σε μία «Ετικέτα πληροφόρησης» (Nutrition Label).

Τα κριτήρια τα οποία ακολουθούνται από το **NewsGuard** για τους ιστοτόπους καθώς και τα κριτήρια διαφάνειας είναι τα ακόλουθα:

- **Δεν δημοσιεύει επανειλημμένα ψευδές περιεχόμενο:** δεν παράγει επανειλημμένα περιεχόμενο που έχει διαπιστωθεί ότι είναι ξεκάθαρα και σημαντικά ψευδές και το οποίο δεν έχει διορθωθεί γρήγορα και σωστά. Υπάρχει ένα «υψηλό ποσοστό» αποτυχίας αυτού του κριτηρίου. Οποιαδήποτε, δηλαδή, δεδομένη ημέρα ο ιστότοπος είναι πιθανό να εμφανίζει ψευδές περιεχόμενο
- **Συλλέγει και παρουσιάζει πληροφορίες υπεύθυνα:** Οι πάροχοι περιεχομένου είναι γενικά δίκαιοι και ακριβείς στην αναφορά και την παρουσίαση πληροφοριών. Αναφέρονται σε πολλές πηγές, κατά προτίμηση σε εκείνες που παρουσιάζουν άμεσες πληροφορίες, από πρώτο χέρι για ένα θέμα ή ένα γεγονός ή από αξιόπιστες δευτερεύουσες πηγές ειδήσεων και δεν παραποιούν κατάφωρα ή δεν παρουσιάζουν λανθασμένα πληροφορίες για να δημιουργήσουν ένα επιχείρημα ή να αναφέρουν ένα θέμα
- **Διορθώνει ή διευκρινίζει τακτικά λάθη:** Ο ιστότοπος διαθέτει αποτελεσματικές πρακτικές για τον εντοπισμό λαθών και για τη δημοσίευση διευκρινίσεων και διορθώσεων, αναγνωρίζει με διαφάνεια τα λάθη και δεν αφήνει αδιόρθωτο ψευδές περιεχόμενο
- **Χειρίζεται τη διαφορά μεταξύ ειδήσεων και γνώμης υπεύθυνα:** Οι πάροχοι περιεχομένου που δίνουν την εντύπωση ότι αναφέρουν ειδήσεις ή ένα μείγμα ειδήσεων και γνώμης ξεχωρίζουν τη γνώμη από την αναφορά ειδήσεων και όταν αναφέρουν ειδήσεις, δεν επιλέγουν σκανδαλώδη γεγονότα ή ιστορίες για να προωθήσουν απόψεις. Πάροχοι περιεχομένου που προβάλλουν μια συγκεκριμένη άποψη αποκαλύπτουν αυτήν την άποψη
- **Αποφεύγει τους παραπλανητικούς τίτλους:** Ο ιστότοπος δεν δημοσιεύει τίτλους που περιλαμβάνουν ψευδείς πληροφορίες, που προκαλούν σημαντική αίσθηση ή δεν αντικατοπτρίζουν αυτό που πραγματικά υπάρχει στο περιεχόμενο κάτω από την επικεφαλίδα
- **Ο ιστότοπος αποκαλύπτει την ιδιοκτησία και τη χρηματοδότηση:** Ο ιστότοπος αποκαλύπτει με τρόπο φιλικό προς τον χρήστη την ιδιοκτησία ή/και τη χρηματοδότησή, καθώς και τυχόν αξιοσημείωτες πολιτικές σχέσεις ή οικονομικά συμφέροντα. Ο αναγνώστης πρέπει να γνωρίζει ποιος χρηματοδοτεί το περιεχόμενο

και ποια σχετικά ενδιαφέροντα μπορεί να έχει ο ιδιοκτήτης ή ο χρηματοδότης στο περιεχόμενο

- **Επισημαίνει με σαφήνεια τη διαφήμιση:** Ο ιστότοπος καθιστά σαφές ποιο περιεχόμενο πληρώνεται και ποιο όχι
- **Αποκαλύπτει ποιος είναι ο υπεύθυνος, συμπεριλαμβανομένων πιθανών συγκρούσεων συμφερόντων:** Οι υπεύθυνοι για το περιεχόμενο προσδιορίζονται στον ιστότοπο και υπάρχει τρόπος για τους αναγνώστες να επικοινωνήσουν με τον ιστότοπο σχετικά με θέματα σύνταξης. Αναφέρεται, δηλαδή, στον αναγνώστη ποιος είναι αρμόδιος να αποφασίσει ποιο περιεχόμενο να δημοσιεύεται
- **Ο ιστότοπος παρέχει τα ονόματα των δημιουργών περιεχομένου, είτε με τα στοιχεία επικοινωνίας ή τα βιογραφικά στοιχεία:** πληροφορίες σχετικά με αυτούς που παράγουν το περιεχόμενο είναι προσβάσιμες στον ιστότοπο και είναι σαφές ποιος παράγει, ποιο περιεχόμενο

Εκτός από τα κριτήρια που αναφέρθηκαν παραπάνω το **NewsGuard** υιοθετεί και διαδικασία βαθμολόγησης (αξιολόγησης) των κριτηρίων.

Η διαδικασία αξιολόγησης έχει σχεδιαστεί κατά τέτοιο τρόπο ώστε να διασφαλίζονται ότι τα εννέα δημοσιογραφικά κριτήρια εφαρμόζονται δίκαια και με σαφήνεια σε όλους τους ιστοτόπους χωρίς να λαμβάνει υπόψη το θέμα, το ύφος, ή την πολιτική τοποθέτηση (εάν υπάρχει).

- **Ένας αναλυτής του NewsGuard αξιολογεί το περιεχόμενο του ιστότοπου με βάση τα εννέα κριτήρια.** Οι αναλυτές είναι εκπαιδευμένοι δημοσιογράφοι που πραγματοποιούν ρεπορτάζ για να καθορίσουν λεπτομέρειες σχετικά με την ιδιοκτησία, τη χρηματοδότηση, την αξιοπιστία και τις πρακτικές διαφάνειας του ιστοτόπου
- **Ο αναλυτής συντάσσει μία γραπτή «Ετικέτα Πληροφόρησης» (Nutrition Label) για τον ιστότοπο με βάση τις αναφορές του.** Οι «επιστημονικές» ετικέτες αποτελούνται από ένα πλέγμα που δείχνει την απόδοση του ιστοτόπου σε καθένα από τα εννέα κριτήρια και μία γραπτή εξήγηση του περιεχομένου του ιστοτόπου (ποιος βρίσκεται πίσω από αυτό και γιατί έλαβε τη βαθμολογία).
- **Κάλεσμα στον ιστότοπο για σχόλια.** Εάν ένας αναλυτής του NewsGuard πιστεύει ότι ένας ιστότοπος μπορεί να αποτύχει σε ένα ή περισσότερα από τα εννέα κριτήρια, η πρακτική του NewsGuard είναι να επικοινωνήσει με τον ιδιοκτήτη του ιστοτόπου για να προσπαθήσει να ζητήσει σχόλιο πριν δημοσιεύσει την αξιολόγηση. Εάν ο

- ιστότοπος παρέχει ένα σχόλιο, αυτό το σχόλιο περιλαμβάνεται στη γραπτή αξιολόγηση του ιστοτόπου για να παρέχει στους χρήστες την άποψη του ιστοτόπου
- **Η βαθμολογία ελέγχεται και αξιολογείται από έμπειρους συντάκτες.** Τουλάχιστον ένας ανώτερος συντάκτης και οι συν-διευθύνοντες σύμβουλοι του NewsGuard εξετάζουν κάθε «επιστημονική ετικέτα» πριν από τη δημοσίευση για να διασφαλίσουν ότι η βαθμολογία είναι όσο το δυνατόν πιο δίκαιη και ακριβής
 - **Ο ιστότοπος λαμβάνει βαθμολογία Κόκκινο ή Πράσινο με βάση τα εννέα δημοσιογραφικά κριτήρια.** Οι αξιολογήσεις του NewsGuard περιλαμβάνουν το ανθρώπινο κριτήριο για την αξιολόγηση της απόδοσης ενός ιστοτόπου σε καθένα από τα εννέα κριτήρια. Αλλά η βαθμολογία του ιστοτόπου για το κόκκινο ή το πράσινο καθορίζεται κυρίως με βάση τα εννέα κριτήρια
 - **Περιοδική ενημέρωση των «Επιστημονικών» Ετικετών.** Το NewsGuard ενημερώνει περιοδικά τη βαθμολογία του για κάθε ιστοτόπο. Εάν ένας ιστοτόπος αλλάξει τις πρακτικές του, η απόδοσή του σε ένα ή περισσότερα από τα εννέα κριτήρια του NewsGuard ενδέχεται να αλλάξει. Τέτοιες ενημερώσεις σημειώνονται σε μια σημείωση του συντάκτη στην «επιστημονική» ετικέτα
 - **Ασκείται υπευθυνότητα και παρουσιάζεται το έργο.** Κάθε «επιστημονική» ετικέτα περιέχει τα ονόματα του συγγραφέα και των συντακτών που εργάστηκαν για την αξιολόγηση. Τα ιστορικά των αναλυτών και των συντακτών που κατονομάζονται, καθώς και αυτά των εποπτικών συντακτών μπορούν να αναζητηθούν επιλέγοντας τα ονόματά τους ή μεταβαίνοντας στη σελίδα της ομάδας του ιστοτόπου NewsGuard. Εάν ένας ιστοτόπος διαφωνεί με τη βαθμολογία του NewsGuard μπορεί να σχολιάσει ως μέρος της διαδικασίας αξιολόγησης και τα αναφερόμενα σχόλια θα συμπεριληφθούν στην «επιστημονική» ετικέτα του ιστοτόπου. Ο εκδότης μπορεί επίσης να γράψει μια πιο λεπτομερή απάντηση ή παράπονο χρησιμοποιώντας τη σελίδα επικοινωνίας του NewsGuard

Το **Maldito Buló** (*Portada · Maldita.Es - Periodismo Para Que No Te La Cuelen*, n.d.) (Εικόνα 27) αποτελεί τον ισπανόφωνο ιστοτόπο ελέγχου γεγονότων (hoax) το οποίο επιλέγει τη μέθοδο αξιολόγησης των ψευδών ειδήσεων, βάση δύο μεταβλητών: της διάχυσης (viralidad) και της επικινδυνότητας (peligrosidad) (*Metodología de Maldito Buló · Maldita.es - Periodismo para que no te la cuelen*, 2020).

Ως προς τη διάχυση εντοπίζονται μόνο όσες δημοσιεύσεις έχουν μεγάλο αντίκτυπο και όσες ενέχουν κίνδυνο (όχι απλές φάρσες από Facebook, Twitter). Ως προς την επικινδυνότητα λαμβάνονται υπόψη εκείνες που εμφανίζονται σε κρίσιμες στιγμές (π.χ. καταστροφές,

δημόσια υγεία, κοινωνική συνύπαρξη). Γίνεται έλεγχος μόνο σε δεδομένα και γεγονότα που μπορούν να επαληθευτούν και όχι σε απόψεις. Η διαδικασία της επαλήθευσης έχει ως εξής:

- Η ομάδα ερευνά την ύπαρξη πιθανής παραπληροφόρησης με επικοινωνία πρωτογενών πηγών, επαληθεύει την προέλευσή τους, τις συγκρίνει με άλλες βάσεις δεδομένων από επίσημες πηγές αλλά και μέσω τεχνικών επαλήθευσης εικόνας, ήχου ή βίντεο αν κριθεί απαραίτητο
- Η ομάδα με τη βοήθεια και της ομάδας Telegram ελέγχει με ποιες πηγές έχει αξιολογηθεί και συγκριθεί, πως έχει φτάσει σε αυτό το συμπέρασμα, ποιες τεχνικές έχει ακολουθήσει για να αντικρούσει την συγκεκριμένη ημερομηνία και ζητά το λάθος
- Μετά τον έλεγχο του αξιολογητή η απόφαση δίδεται σε ψηφοφορία 9 επιμελητών. Εάν δεν υπάρχει αντίθετη γνώμη, η απόφαση δημοσιεύεται με 4 ψήφους υπέρ. Εάν υπάρχει μία αντίθετη ψήφος, η διαδικασία αξιολόγησης λαμβάνει υπόψη τα αιτήματα που έχουν παρουσιαστεί ή τον λόγο της αντίθετης ψήφου και ψηφίζει ξανά
- Όλα τα μέλη με δικαίωμα ψήφου έχουν δικαίωμα βέτο. Εάν κάποιος ασκήσει βέτο σε μία αντίθετη ψήφο με αιτιολόγηση, και μετά από τη συζήτηση των μελών που έχουν δικαίωμα ψήφου, το βέτο του παραμένει αμετάκλητο, δεν δημοσιεύεται. Δεν υπάρχει βέτο ως προς την ιδεολογία και θα πρέπει να υποστηρίζεται από επιχειρήματα ή άλλους αποδεκτούς λόγους



Εικόνα 27. Στιγμιότυπο από Maldito Buló.

Οι κατηγορίες παραπληροφόρησης του Maldito Bulo είναι οι εξής:

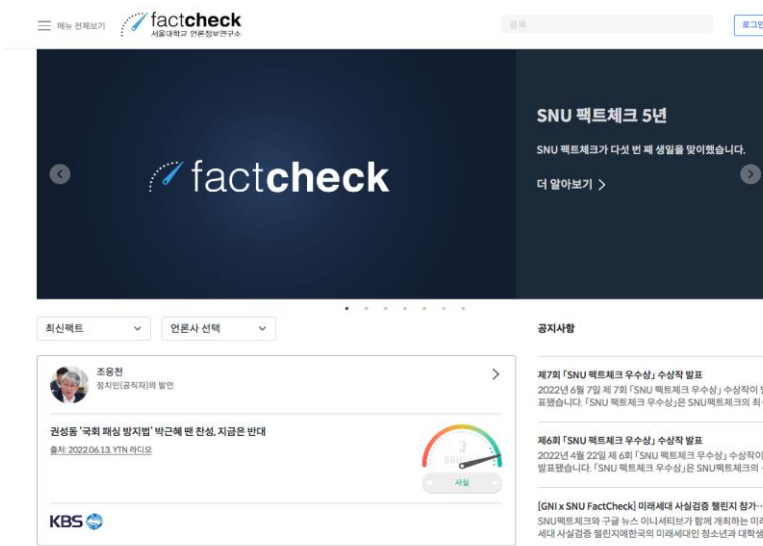
- **Hoax (bulo):** πληροφορίες που έχουν επαληθευτεί ότι είναι ψευδείς και αντιμετωπίζονται ως πραγματικές, δηλαδή ημερομηνίες, γεγονότα, βίντεο, εικόνες, ηχητικές πληροφορίες, παραποιημένες εικόνες και βίντεο, πραγματικές εικόνες, βίντεο και ηχητικές πληροφορίες όπου υπάρχει ένα πλαίσιο ψεύδους είτε χρονικό, είτε τοποθεσίας, είτε απόδοσης, ή άλλου εσφαλμένου πλαισίου και σατιρικό περιεχόμενο το οποίο διαδίδεται ως πραγματικό και το κοινό το πιστεύει.
- **Δεν υπάρχουν αποδείξεις:** οι πληροφορίες που δεν έχουν αποδειχτεί ως αληθείς αλλά και εκείνες για τις οποίες δεν υπάρχουν αξιόπιστες αποδείξεις ότι είναι hoax και όταν κανείς δεν παρουσιάζει αποδείξεις αξιόπιστες της ύπαρξής τους, όπως δηλώσεις ή αποσπάσματα από άτομα δημόσιου κύρους, πρόσφατα ή ιστορικά, των οποίων η προέλευση δεν μπορεί να εντοπιστεί εκτός από το περιεχόμενο που υπόκειται σε επαλήθευση ή άλλα συναφή, χωρίς να βρεθούν προηγούμενες ή παράλληλες αναφορές, οι οποίες να υποστηρίζουν το συγκεκριμένο απόσπασμα, γεγονότα για τα οποία δεν υπάρχουν αποδείξεις, μάρτυρες, ή online υλικό, αλλά δεν μπορεί να πιστοποιηθεί ότι δεν είναι αληθή για να χαρακτηριστούν ως hoax.

Επιπλέον, έχει υιοθετηθεί και άλλο ένα κριτήριο το «**Τι Γνωρίζουμε**». Το συγκεκριμένο κριτήριο εφαρμόζεται όταν δεν υπάρχουν αρκετές ανεξάρτητες πληροφορίες ως προς το περιεχόμενο ή συμβάν για να χαρακτηριστεί ως hoax ή δεν υπάρχουν αποδείξεις, αλλά έχουν ανακτηθεί πληροφορίες από διαφορετικούς παράγοντες και που εκτίθενται για παράθεση του πλαισίου, περιεχόμενα που έχουν διαφορετικές εκδοχές και είναι αντιφατικές, αξιολογήσεις που εμπίπτουν σε μη τελεσίδικες δικαστικές προτάσεις και που για αυτό το λόγο υπόκεινται σε μελλοντική τροποποίηση.

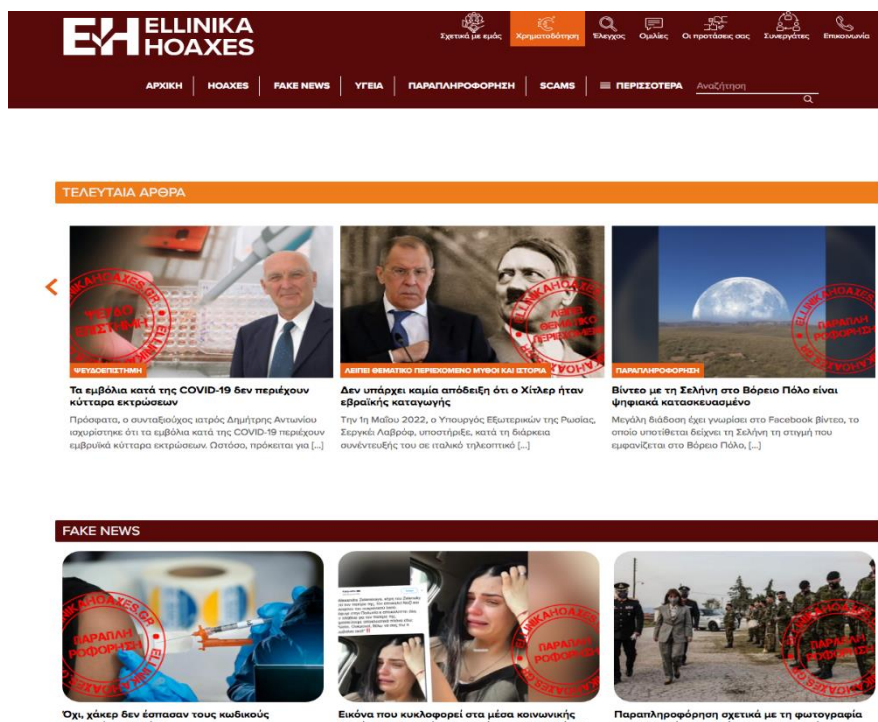
Ως προς το κοινωνικό περιεχόμενο ο ιστότοπος αναγνωρίζει ως hoax τα tweet που έχουν ένδειξη τρολ και εκείνα που μιμούνται τον λογαριασμό ενός ατόμου ή οργανισμού αλλά και όταν δεν υπάρχουν σύνδεσμοι που αποδεικνύουν την δημοσίευση του tweet, όταν δεν υπάρχει αρχειοθετημένη έκδοση ή αρχεία μνήμης (cache) της δημοσίευσης, όταν δεν υπάρχουν αναδημοσιεύσεις ή σχόλια, αυτό που συνοδεύει τη δημοσίευση είναι το ίδιο στιγμιότυπο οθόνης, με τον ίδιο αριθμό retweets, likes, ή δεν εμφανίζεται ο αριθμός των likes, retweets, όταν ο δημιουργός το αρνείται κ.ά.

Το **SNU FactCheck** (Εικόνα 28) είναι η πρώτη κορεάτικη πλατφόρμα για τον έλεγχο των ισχυρισμών στη Σεούλ, η οποία έχει αρκετά κοινά στοιχεία με το Politifact με υιοθέτηση και εδώ του «μέτρου αληθείας». Περιλαμβάνει πάνω 1000 πιστοποιημένα γεγονότα τα οποία έχουν ελεγχθεί από εταιρείες μέσω ενημέρωσης. Το SNU Fact Check αξιολογεί τις δηλώσεις

πολιτικών, υποψηφίων εκλογών, ντιμπέϊτ, ομιλιών και τις δηλώσεις που δημοσιεύονται στο κοινό από τον τύπο και τα μέσα (SNUFactCheck, n.d.). Η κλίμακα πιστοποίησης βασίζεται σε 5 κριτήρια, από εκείνα που δεν είναι «καθόλου έγκυρα» έως εκείνα που είναι «πλήρως αληθή». Σε περίπτωση που μία δημοσίευση έχει μεγάλη απόκλιση σε βαθμολογία τότε κατηγοριοποιείται ως «αμφισβητήσιμη» (SNU FactCheck Wins Human Technology Award at Asia Future Forum - News - Newsroom - SNU NOW, n.d.). Συγκεκριμένα τα κριτήρια που χρησιμοποιεί είναι: **Μη αληθές, Σχεδόν μη αληθές, Ημιαληθές, Αρκετά Αληθές, Αληθές, Αμφισβητήσιμο.**



Εικόνα 28. Στιγμιότυπο από SNUFactCheck.



Εικόνα 29. Στιγμιότυπο από Ellinika Hoaxes.

Στην Ελλάδα ο ιστότοπος ελέγχου γεγονότων είναι τα **Ellinika Hoaxes** (Εικόνα 29). Τα βήματα που ακολουθούνται για τη μεθολογία αξιολόγησης των ειδήσεων είναι:

- **Εντοπισμός δυνητικά ύποπτου υλικού:** κάθε μορφή πληροφορίας εξετάζεται, είτε πρόκειται για άρθρο, φωτογραφία ή βίντεο. Αναζητούνται συγκεκριμένες λεπτομέρειες οι οποίες οδηγούν στο συμπέρασμα ως προς τη μη εγκυρότητα του υλικού. Χαρακτηριστικά παραδείγματα αποτελούν οι τίτλοι με κεφαλαία γράμματα και έντονα συναισθηματική γλώσσα, η απουσία αρθρογράφου, η πρόταση για άμεση κοινοποίηση. Τα συγκεκριμένα χαρακτηριστικά δεν καθιστούν αυτόματα το υπό εξέταση υλικό ψευδές, δεδομένου όμως του τεράστιου όγκου πληροφοριών, μπορούν να λειτουργήσουν ως ενδείξεις για εντοπισμό δυνητικά ύποπτου υλικού, το οποίο χρειάζεται περαιτέρω ανάλυση. Με αυτό το τρόπο, μειώνεται ο χρόνος αναζήτησης του υπό εξέταση υλικού
- **Ανάλυση περιεχομένου:** εφόσον επιλεγεί το υλικό που χρειάζεται αξιολόγηση, διερευνάται προσεκτικά το περιεχόμενο κρατώντας μία μικρή λίστα με όλους τους ισχυρισμούς που εμφανίζονται μέσα σε αυτή ώστε να ελεγχθούν ξεχωριστά. Σε περίπτωση που είναι εφικτό ή αναγκαίο γίνεται επικοινωνία με την αρχική πηγή του υπό εξέταση ισχυρισμού ώστε να συμπεριληφθεί σε δηλώσεις ή σχόλια που θα είναι χρήσιμα στο άρθρο (Sitistas & Pournis, 2019)
- **Έρευνα στο οπτικοακουστικό υλικό:** στην περίπτωση που το υλικό είναι ένα άρθρο το οποίο περιλαμβάνει εικόνες ή βίντεο σχετικά με τους ισχυρισμούς που γίνονται εντός του κειμένου πρέπει να εντοπιστεί αν το οπτικοακουστικό υλικό σχετίζεται με τους εν λόγω ισχυρισμούς. Για τις εικόνες χρησιμοποιείται η υπηρεσία Google Images η οποία επιτρέπει να ενσωματωθεί μία εικόνα στη μηχανή αναζήτησης και να διερευνηθεί που και πότε έχει χρησιμοποιηθεί ξανά. Αυτό το εργαλείο έχει φανεί χρήσιμο για περιπτώσεις στις οποίες το άρθρο που εξετάζεται περιλαμβάνει φωτογραφίες οι οποίες ήταν από πιο παλιά ημερομηνία από αυτή που ισχυριζόταν ο αρθρογράφος ή από φωτογραφίες οι οποίες δεν είχαν ουδεμία σχέση με το περιστατικό που περιέγραφε ο αρθρογράφος. Επίσης, μπορούν να δημιουργηθούν στιγμιότυπα (screenshots) από διαφορετικά σημεία μέσα στο βίντεο και να ανέβουν στο Google Images ώστε να εντοπιστεί η πηγή μέσω των thumbnails του βίντεο.
- Σε περίπτωση που το υπό εξέταση άρθρο περιλαμβάνει έναν ισχυρισμό για συγκεκριμένη τοποθεσία, π.χ. το άρθρο το οποίο ισχυριζόταν ότι ξεκίνησε η κατασκευή εκκλησίας σατανιστών στην Αθήνα, είναι δυνατό να χρησιμοποιηθούν εφαρμογές όπως το Google Maps και το Google Street View για να παρουσιαστεί με

πραγματικά στοιχεία ότι ισχυρισμός αληθεύει. Όταν οι ισχυρισμοί αφορούν φωτογραφίες για τις οποίες υπάρχουν ενδείξεις ότι είναι αλλοιωμένες μέσω της χρήση ειδικού προγράμματος επεξεργασίας, εκτός από τη χρήση του Google Images για να εντοπιστεί η αρχική πηγή, μπορεί να πραγματοποιηθεί έρευνα της πρωτότυπης φωτογραφίας για να πραγματοποιηθεί δοκιμή ανάλυσης της αλλοιωμένης εικόνας

- **Εξέταση επιστημονικών μελετών σε περιπτώσεις ψευδοεπιστημονικών ισχυρισμών:** δημιουργία λίστας με κάθε ισχυρισμό και έρευνα καθενός με χρήση δημοσιευμένης επιστημονικής βιβλιογραφίας (*Ellinika Hoaxes-Μεθοδολογία Ελέγχου Και Επεξήγηση Χαρακτηρισμών*, n.d.). Μέσω της ακόλουθης μεθόδου διερευνώνται οι επιστημονικές μελέτες προκειμένου να εξακριβωθεί αν αληθεύει ο ισχυρισμός.
 - Εύρεση μελέτης (λέξεις-κλειδιά)
 - Εξέταση αξιοπιστίας επιστημονικού περιοδικού (διπλή-τυφλή αξιολόγηση από ομότιμους). Αξίζει να επισημανθεί και περαιτέρω προσοχή, βάση των συντακτών του ιστοτόπου, ως προς τη μεθοδολογία της έρευνας στο περιοδικό. Δεδομένου ότι πολλά περιοδικά δεν υιοθετούν την διπλή τυφλή αξιολόγηση αλλά δίνουν επί πληρωμή θέσεις για συγγραφή ερευνών (αρπακτικά περιοδικά) (*Discussion Document*, 2019) με αποτέλεσμα να υπάρχουν μεθοδολογικά λάθη στη έρευνα
 - Εστίαση σε συγκεκριμένα τμήματα της μελέτης (περίληψη, συζήτηση, συμπεράσματα)
 - Επαλήθευση αποτελεσμάτων. Σύγκριση με άλλες επιστημονικές μελέτες των συμπερασμάτων των μελετών
- **Επικοινωνία με άλλες ομάδες ελέγχου ειδήσεων:** όταν για παράδειγμα μια ψευδής είδηση από την Κένυα εμφανιστεί στην ελληνική πλευρά του διαδικτύου. Σε αυτές τις περιπτώσεις, αφού εξαντληθούν όλα τα άλλα εργαλεία για να εξακριβωθεί αν ο ισχυρισμός είναι αληθής ή ψευδής, πραγματοποιείται επικοινωνία με άλλες ομάδες ελέγχου ειδήσεων οι οποίες έχουν πιστοποιηθεί από το Παγκόσμιο Δίκτυο Ελέγχου Ειδήσεων (IFCN) και γίνεται επερώτηση αν υπάρχει επαφή με τον συγκεκριμένο ισχυρισμό ώστε να εντοπιστούν στοιχεία που θα βοηθήσουν στην αξιολόγηση του υπό εξέταση άρθρου (Daniilides, 2019)

Στα **Ellinika Hoaxes** υιοθετούνται και χαρακτηρισμοί-σφραγίδες για την αξιολόγηση των ψευδών ειδήσεων . Στην αρχική εικόνα κάθε άρθρου εμφανίζεται μία σχετική “σφραγίδα”,

προκειμένου να γίνει από την αρχή κατανοητό τι είδους παραπληροφόρηση θα διαβάσει το κοινό. Στη συνέχεια παρουσιάζονται οι αξιολογήσεις των ισχυρισμών που διαμορφώνονται με βάση τη σφραγίδα αξιολόγησης από τους συντάκτες των ελληνικών hoaxes αλλά και κάποιοι χαρακτηρισμοί-σφραγίδες (Εικόνα 30):

ΑΛΗΘΕΙΑ: μία είδηση ή ένα δημοσίευμα είναι στην ολότητά του έγκυρο

ΨΕΥΔΕΣ: ένα άρθρο ή μία είδηση είναι συνολικά ψευδής

ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜΟΣ: ένα δημοσίευμα περιλαμβάνει ψευδή ισχυρισμό και πραγματικά γεγονότα

ΜΙΞΗ ΓΕΓΟΝΟΤΩΝ ΚΑΙ ΠΑΡΑΠΟΙΗΣΕΩΝ: ένα δημοσίευμα περιέχει ψευδή και πραγματικά στοιχεία

ΚΙΝΔΥΝΟΛΟΓΙΑ: ένα άρθρο ή ένας ισχυρισμός περιλαμβάνει στοιχεία τρομολαγνείας ή πρόκλησης πανικού

ΨΕΥΔΟΕΠΙΣΤΗΜΗ: ένα άρθρο ή ένας ισχυρισμός είναι στην ολότητά του ή εν μέρη βασισμένος σε μη επιστημονικά αποδεκτά στοιχεία

LIKE FARMING: ένα άρθρο, ένας ισχυρισμός, ή μία φωτογραφία, χρησιμοποιείται στα κοινωνικά δίκτυα με σκοπό να προκαλέσει (αδικαιολόγητη) αίσθηση προκειμένου να παροτρύνει τους χρήστες να δηλώσουν ότι «τους αρέσει» (like) ή να το κοινοποιήσουν

CLICKBAIT: ένα άρθρο ή ένας ισχυρισμός είναι διατυπωμένα με υπερβολικούς χαρακτηρισμούς που προκαλούν αίσθηση για να παροτρύνουν τους χρήστες να το επιλέξουν





Εικόνα 30. Χαρακτηρισμοί σφραγίδες στα Ellinika Hoaxes.

4.1.2 Έλεγχος Βάση Πληθοπορισμού (Crowdsourcing)

Για τη συγκεκριμένη οπτική υιοθετείται η ευφυΐα του πλήθους (wisdom of crowd) προκειμένου να ελεγχθεί η εγκυρότητα των ψευδών ειδήσεων (Ibrishimova & Li, 2020; Shu, Cui, et al., 2019; Souza Freire et al., 2021). Η συγκεκριμένη οπτική παρότι λιγότερο αξιόπιστη και ακριβής (προκατελιημμένες πολιτικές απόψεις), έχει καλύτερη απόδοση καθώς ο χρήστης μπορεί να φιλτράρει τους μη αξιόπιστους χρήστες με επίλυση διασταυρούμενων αποτελεσμάτων (X. Zhou & Zafarani, 2020). Παραδείγματα τέτοιας οπτικής είναι: το Fiskkit (*Fiskkit- A Better Way to Discuss News*, n.d.) όπου διαχωρίζονται τα άρθρα ανάλογα με το περιεχόμενο (ειδήσεις ή μη) και αξιολογούνται από τους αναγνώστες. Επίσης υπάρχει και η βάση δεδομένων CredBank (Mitra & Gilbert, 2015) όπου οι πληροφορίες έχουν κατηγοριοποιηθεί ως ψευδής ή αληθής από fact-checkers μέσω της πλατφόρμας Amazon Mechanical Turk.

Το Fiskkit είναι ένας ιστότοπος συζήτησης σημαντικών θεμάτων ειδήσεων και εύρεσης εάν το περιεχόμενο είναι ακριβές σε κάθε άρθρο στο διαδίκτυο. Ο διαχωρισμός των άρθρων βασίζεται στο περιεχόμενο και η ανατροφοδότηση της εγκυρότητας του άρθρου επισημαίνεται από τις γνώμες των χρηστών (αληθές, ψευδές, γενικόλογο, υπεραπλουστευμένο, μεροληπτικό κ.λ.π.) (*Fiskkit- A Better Way to Discuss News*, n.d.).

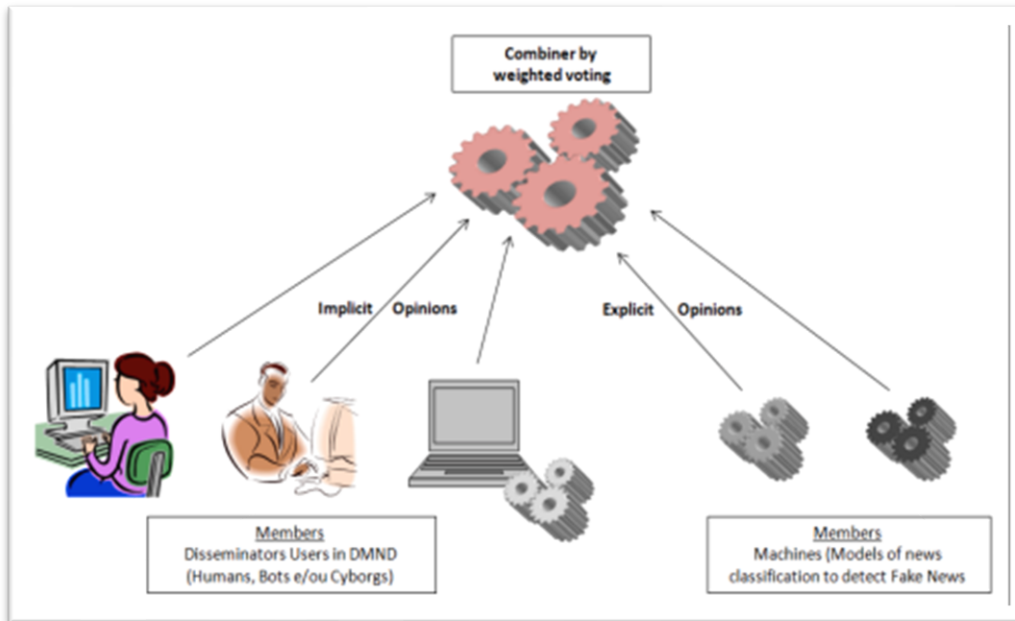
Το Amazon Mechanical Turk (*Amazon Mechanical Turk*, n.d.) είναι μια ανοιχτή πλατφόρμα πληθοπορισμού για επιχειρήσεις όπου οι εργασίες ανατίθενται σε διαδικτυακούς «επαγγελματίες-ερευνητές» σε όλο τον κόσμο (π.χ. αξιολόγηση δεδομένων, συμμετοχή σε έρευνα, τροποποίηση περιεχομένου), μειώνοντας κατά αυτόν τρόπο μεγάλα

και χρονοβόρα έργα σε μικρότερες υποεργασίες. Η Shu χρησιμοποίησε το Amazon Mechanical Turk για την αξιολόγηση της κατηγοριοποιημένης λίστας σχολίων των ψευδών ειδήσεων (Shu, Cui, et al., 2019).

Ο Souza Freire χρησιμοποίησε την ευφυΐα πλήθους σε ένα υβριδικό μοντέλο ανθρώπων και μηχανικής μάθησης για τον εντοπισμό των ψευδών ειδήσεων (Souza Freire et al., 2021) (Εικόνα 31). Υιοθετώντας τη μέθοδο Optimal detective από τον Tschiatzschek, οι απαντήσεις των χρηστών αξιολογούνται από ειδικούς ανάλογα της φήμης τους (π.χ. πιθανότητα ένας χρήστης να δώσει γνώμη καλή, αδιάφορη, κακή για τις ειδήσεις) (Tschiatzschek et al., 2018). Στη συγκεκριμένη μέθοδο υπολογίζεται ότι η «έμμεση γνώμη των χρηστών» (πλήθος) υπονοείται από τη συμπεριφορά τους ως προς τη διάδοση των ειδήσεων. Η οπτική της έμμεσης γνώμης των χρηστών βασίζεται στην αρχή ότι όταν ο χρήστης διαδίδει ειδήσεις σε ψηφιακά μέσα, είτε κακόβουλα είτε όχι, θέλει να δείξει ότι η είδηση είναι αληθής, εμπνεόμενη από τη φιλοσοφία του Habermas ότι κάθε επικοινωνιακή πράξη μεταφέρει αδιαμφισβήτητα έναν αληθοφανή ισχυρισμό (Habermas, 2020). Στη συγκεκριμένη μέθοδο χρησιμοποιούνται και αλγόριθμοι (μηχανές) συμπληρωματικά με τις γνώμες των χρηστών (άμεση γνώμη μηχανής). Οι αλγόριθμοι επιλέγονται ανάλογα με τα είδη των χαρακτηριστικών (εικόνα, κείμενο, δομημένα δεδομένα).

Σύμφωνα με τους Lofi και Maary το υβριδικό πλήθος μπορεί να έχει τα εξής μοτίβα:

- **Μαγικό φίλτρο:** Αλγόριθμος που φιλτράρει δεδομένα τα οποία θα σταλούν στο ανθρώπινο πλήθος. Σκοπός του η αποστολή σχετικών δεδομένων στο πλήθος, ο εντοπισμός δημοσίευσης ειδήσεων από χρήστες ή προπαγανδιστές αλλά και ο εντοπισμός του τρόπου δημιουργίας της φήμης, π.χ. μέσα από το προφίλ χρήστη
- **Εκπαιδευτής Πλήθους:** Το ανθρώπινο πλήθος παρέχει δεδομένα στον αλγόριθμο. Σκοπός του η χρήση του ανθρώπινου πλήθους για τη δημιουργία βάσεων δεδομένων οι οποίες θα δοθούν αργότερα ως δεδομένα εκπαίδευσης των αλγορίθμων
- **Βελτίωση μηχανής:** οι αλγόριθμοι αρχικά προσπαθούν να βρουν λύσεις και μετέπειτα να προωθήσουν τα αποτελέσματα τα οποία δεν είναι ικανοποιητικά στο ανθρώπινο πλήθος. Σκοπός η αποστολή μόνο χαμηλής ποιότητας αποτελεσμάτων στο ανθρώπινο πλήθος από τους αλγόριθμους
- **Δυνητικός εργάτης:** Αλγόριθμοι και ανθρώπινο πλήθος λαμβάνουν δεδομένα και ενεργούν για την εύρεση εξατομικευμένων λύσεων. Στόχος η παράλληλη και ενσωματωμένη λειτουργία αλγορίθμων και ανθρώπινου πλήθους (Lofi & Maary, 2014)



Εικόνα 31. Στιγμιότυπο από μοντέλο πληθοπορισμού.

4.1.3 Αυτοματοποιημένος Έλεγχος (Automatic Fact-Checking)

Η οπτική που βασίζεται στην ανθρώπινη προσέγγιση δεν ενδείκνυται για μεγάλα δεδομένα όπου ο όγκος των πληροφοριών είναι αρκετά μεγάλος, κυρίως από βάσεις που έχουν πάρει δεδομένα από κοινωνικά μέσα δικτύωσης. Γι' αυτό η συγκεκριμένη προσέγγιση υιοθετεί πρακτικές βασιζόμενες στην επιστήμη της μηχανικής μάθησης (Machine Learning), στην επεξεργασία φυσικής γλώσσας (Natural Language Processing), στην εξόρυξη δεδομένων (Data Mining) αλλά και στη θεωρία δικτύου/γράφων (Barrón-Cedeño et al., 2019; Ibrishimona & Li, 2020; Kausar et al., 2020; Meel & Vishwakarma, 2020; Shu, Cui, et al., 2019; X. Zhou & Zafarani, 2020).

Αποτελείται από δύο επιμέρους επίπεδα: α) εξαγωγή γεγονότων (συλλογή από γεγονότα και δημιουργία της Γνωσιακής Βάσης) και β) έλεγχος γεγονότων (fact-checking) όπου ελέγχεται η αυθεντικότητα των ειδήσεων μέσω σύγκρισης τους με γεγονότα της Γνωσιακής Βάσης⁷ (Hangloo & Arora, 2021).

Στον έλεγχο των γεγονότων γίνεται σύγκριση της γνώσης (περιεχόμενο τριπλετών) με τα γεγονότα (αληθή γεγονότα). Οι γνωσιακές βάσεις είναι απαραίτητες για τον έλεγχο των γεγονότων γιατί παρέχουν κατηγοριοποιημένες ψευδής και αληθής πληροφορίες. Ανάλογο παράδειγμα γνωσιακής βάσης είναι η Fever (*Fact Extraction and VERification*, n.d.-a; Ibrishimona & Li, 2020). Οι ισχυρισμοί έχουν δημιουργηθεί από τροποποίηση προτάσεων που

⁷ Χρήση ανοιχτού λογισμικού και γνωσιακού γράφου για έλεγχο των ισχυρισμών.

έχουν εξαχθεί από τη Wikipedia και έχουν κατηγοριοποιηθεί ως «Υποστηριζόμενοι», «Αναληθής» και «Χωρίς αρκετή πληροφορία». Στη συνέχεια έγινε επανέλεγχος για την ορθή κατηγοριοποίηση των ισχυρισμών χωρίς προηγούμενη γνώση της πρότασης από την οποία προήλθαν (*Fact Extraction and VERification*, n.d.-b).

Οι Ibrishimova και Li υιοθέτησαν ένα μοντέλο γνωσιακής αξιολόγησης το οποίο συνδύαζε τρία ερωτήματα σχετικά με τη διερεύνηση (εξέταση), διάδοση, και την αξιοπιστία πηγής (π.χ. τίτλος όμοιος με πηγή; ίδιος τίτλος σε περισσότερες από x πηγές; περιλαμβάνει ίδιους ισχυρισμούς;) με τρεις απαντήσεις χρηστών (μεταβλητές) (π.χ. ναι, όχι, άγνωστο) οι οποίες έχουν πιστοποιηθεί από το σύστημα (Ibrishimova & Li, 2020).

Η Yu για τον εντοπισμό της διασποράς των ψευδών ειδήσεων συνέλεξε δεδομένα από μέσα που αναπαράγουν ή δημιουργούν παραπληροφόρηση (π.χ. αναδημοσιεύσεις tweet) (disinformation channel). Για τον εντοπισμό του καναλιού παραπληροφόρησης υιοθετήθηκαν μεταβλητές σε σχέση με το μέγεθος της παραπληροφόρησης, τη βαθμολόγηση της παραπληροφόρησης (3 για τις ψευδείς ειδήσεις, 2 για εκείνες που είναι κατά το ήμισυ αληθείς, 1 για τις αληθείς που ενέχουν σκοπιμότητα και αποπροσανατολίζουν το κοινό) καθώς και το μέγεθος παραπληροφόρησης γεγονότων.

Για τις αρνητικές ειδήσεις έγινε συλλογή λέξεων κλειδιών από εταιρείες παραπληροφόρησης και μετέπειτα από ειδήσεις που είχαν τις συγκεκριμένες λέξεις στους τίτλους τους. Οι αρνητικές μεταβλητές που χρησιμοποιήθηκαν είναι οι ακόλουθες: μέγεθος αρνητικών ειδήσεων σχετικά με την εταιρεία, αρνητική επιρροή σχετικά με την εταιρεία (αναλογία μεγέθους αρνητικών ειδήσεων σε συνολικό αριθμό ειδήσεων), εχθρικότητα (αρνητικό μέγεθος ειδήσεων για εταιρεία σε σχέση με το συνολικό αρνητικό μέγεθος), αρνητικός βαθμός ειδήσεων από μέσα και αναδημοσιεύσεις, και αναλογία μεγέθους αρνητικών ειδήσεων σε σχέση με τον συνολικό αριθμό ειδήσεων και αναδημοσιεύσεων.

Στη συνέχεια, διαπιστώθηκε ότι όσο περισσότερο ένα κανάλι αναπαράγει ή δημιουργεί παραπληροφόρηση τόσο τείνει να είναι κανάλι παραπληροφόρησης, αλλά και ότι το μέγεθος της παραπληροφόρησης που δημιουργούν ή αναδημοσιεύουν τα μέσα καθώς και οι αρνητικές ειδήσεις αποτελούν και αυτά κριτήρια παραπληροφόρησης (Y. Yu et al., 2020).

Εκτός από τον έλεγχο των γεγονότων χρησιμοποιούνται και αυτοματοποιημένες διαδικτυακές πλατφόρμες για τον εντοπισμό των ψευδών ειδήσεων όπως είναι οι: Emergent, Twitter Trails, Claimbuster, Google FactCheck Tools, Hoaxy, InVid, Mever, Bullshit Detector (** *Twitter Trails: Tool for Monitoring the Propagation of Rumors* **, n.d.; *ClaimBuster: Automated Live Fact-Checking*, n.d.; *Emergent*, n.d.; *Fact Check Tools*, n.d.; *Hoaxy*, n.d.; *InVID*

Project - Video Verification, n.d.; *MeVer - Image Verification Assistant*, n.d.; *Stiahnite si Bullshit detector*, n.d.), (Kar, 2020; Ksieniewicz et al., 2019; Meel & Vishwakarma, 2020; Papadopoulou et al., 2019; Yuliani et al., 2019). Οι συγκεκριμένες πλατφόρμες εξειδικεύονται είτε σε επιστημονικές ισχυρισμών στα κοινωνικά μέσα (Emergent) και φημών (Twitter Trails), είτε σε πλατφόρμες ή εφαρμογές εντοπισμού και ανάλυσης παραπληροφόρησης (Hoaxy, Bullshit Detector), είτε σε online fact-checking για εύρεση πραγματικών ισχυρισμών (Claimbuster), είτε για την αξιοπιστία και αξιολόγηση των βίντεο καθώς και για τον εντοπισμό απάτης (InVid).

Στην *εξαγωγή των γεγονότων* (fact extraction) για να γίνει η συλλογή τους είναι απαραίτητη η εξόρυξη γνώσης από τον ιστό. Με τον όρο γνώση εννοούμε ένα σύνολο από τριπλέτες (υποκείμενο, κατηγορία, αντικείμενο ή πράγμα-σχέση-πράγμα). Παραδείγματα τέτοιων σχέσεων θεωρούνται τα Uris (υποκείμενο-αντικείμενο-σχέση) (π.χ. viaf), το RDF (περιγραφή οντολογιών, τριπλετών), η DBpedia κ.α. (Braşoveanu & Andonie, 2021; Lukas Koster, 2011). Σύμφωνα με τους Zhou και Zafarani η εξαγωγή γνώσης μπορεί να βασίζεται είτε σε μία πηγή η οποία θα είναι έμπιστη (π.χ. Wikipedia) είτε σε γνώση από διαφορετικές πηγές (ανοιχτή πηγή), η οποία θα είναι πλήρης (π.χ. Knowledge Vault, ScraperWiki, Scrapy) (*Scrapy | A Fast and Powerful Scraping and Web Crawling Framework*, n.d.) (X. Zhou & Zafarani, 2020).

Άλλα εργαλεία εξαγωγής πληροφοριών τα οποία χρησιμοποιούνται συμπληρωματικά με μοντέλα αλγοριθμικής μάθησης και παρουσιάζονται στη βιβλιογραφία είναι: το Wart Tool (wget2) (*Wget - GNU Project - Free Software Foundation*, n.d.) το οποίο ανακτά αρχεία http, https, ftp (ακόμα και αν ο χρήστης δεν είναι ενεργός) μέσω γραμμής εντολών, το dragnet tool το οποίο εξάγει περιεχόμενο από άρθρα ιστοσελίδων και σχολιασμό χρηστών από την ιστοσελίδα (προαιρετικά), το Beautiful soup library (*Beautiful Soup Documentation — Beautiful Soup 4.4.0 Documentation*, n.d.) (Εικόνα 32) όπου μέσω της ανάλυσης εγγράφων από html/xml δημιουργείται ένα δέντρο σύνταξης (Parse tree) προκειμένου να εξάγει πληροφορίες από τις ιστοσελίδες (π.χ. εξαγωγή παραγράφων). Λειτουργεί μαζί με τον συντάκτη ανάλυσης (parser). Ο συντάκτης ανάλυσης είναι ένα πρόγραμμα εντολών που διαχωρίζει τα δεδομένα εισόδου σε μέρη (ουσιαστικά-αντικείμενα, ρήματα-μέθοδοι). Το Rake library (csurfer, n.d.) (Εικόνα 33) όπου ο αλγόριθμος εξάγει λέξεις κλειδιά και προσδιορίζει σημαντικές φράσεις σε ένα σώμα κειμένου (π.χ. false, falsely, death hoax) μέσα από την ανάλυση της συχνότητας των λέξεων. Η ανάλυση της συχνότητας των λέξεων συσχετίζεται με άλλες λέξεις στο κείμενο. Το Selenium (*The Selenium Browser*

Automation Project, n.d.) συλλέγει ιστοσελίδες προκειμένου να εξάγει ισχυρισμούς (Cremisini et al., 2019; Vishwakarma et al., 2019).

```
from bs4 import BeautifulSoup
soup = BeautifulSoup(html_doc, 'html.parser')

print(soup.prettify())
# <html>
# <head>
# <title>
#   The Dormouse's story
# </title>
# </head>
# <body>
# <p class="title">
# <b>
#   The Dormouse's story
# </b>
# </p>
# <p class="story">
#   Once upon a time there were three little sisters; and their names were
#   <a class="sister" href="http://example.com/elsie" id="link1">
#     Elsie
#   </a>
#   ,
#   <a class="sister" href="http://example.com/lacie" id="link2">
#     Lacie
#   </a>
#   and
#   <a class="sister" href="http://example.com/tillie" id="link2">
#     Tillie
#   </a>
#   ; and they lived at the bottom of a well.
# </p>
# <p class="story">
#   ...
# </p>
# </body>
# </html>
```

Εικόνα 32. Στιγμιότυπο από BeautifulSoup library.

```
In [1]: from rake_nltk import Rake

In [2]: r = Rake()

In [3]: mytext = '''Compatibility of systems of linear constraints over the set of
...: natural numbers. Criteria of compatibility of a system of linear
...: Diophantine equations, strict inequations, and nonstrict inequations are
...: considered. Upper bounds for components of a minimal set of solutions
...: and algorithms of construction of minimal generating sets of solutions
...: for all types of systems are given. These criteria and the corresponding
...: algorithms for constructing a minimal supporting set of solutions can be
...: used in solving all the considered types of systems and systems of mixed
...: types.'''

In [4]: r.extract_keywords_from_text(mytext)
```

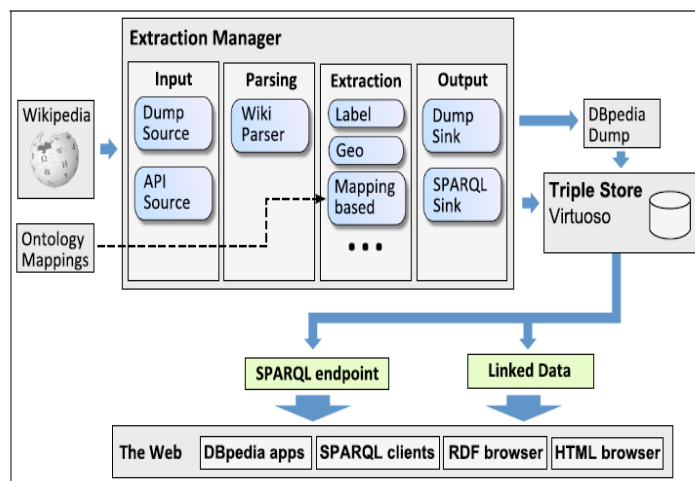
```

Out[5]:
['minimal generating sets',
'linear diophantine equations',
'minimal supporting set',
'minimal set',
'linear constraints',
'upper bounds',
'strict inequations',
'nonstrict inequations',
'natural numbers',
'mixed types',
'corresponding algorithms',
'considered types',
'set',
'types',
'considered',
'algorithms',
'used',
'systems',
'system',
'solving',
'solutions',
'given',
'criteria',
'construction',
'constructing',
'components',
'compatibility']
In [6]: r

```

Εικόνα 33. Στιγμιότυπο από Rake Library (εξαγωγή λέξεων-κλειδιών).

Η *DBpedia Spotlight* (*DBpedia Spotlight - Shedding Light on the Web of Documents*, n.d.) είναι ένα εργαλείο εξαγωγής, εντοπισμού και ανάλυσης οντοτήτων και χρησιμοποιείται και στην αναγνώριση ονομαστικών οντοτήτων. Μέσω του συγκεκριμένου εργαλείου ο χρήστης μπορεί να διασυνδέσει, να επαναχρησιμοποιήσει, να δημιουργήσει σημασιολογικά ερωτήματα από υψηλής ποιότητας βάσης δεδομένων και να εξάγει πληροφορίες από αδόμητα δεδομένα (Brașoveanu & Andonie, 2021) (Εικόνα 34).



Εικόνα 34. DBpedia Spotlight.

Ο αλγόριθμος *Boilerpipe* εξάγει περιεχόμενο από ιστοσελίδες με τη χρήση βιβλιοθήκης ανοιχτού κώδικα. Ο αλγόριθμος εντοπίζει και διαγράφει τα πρόσθετα μέρη ενός περιεχομένου μίας ιστοσελίδας (template) και περιλαμβάνει 4 στρατηγικές: εξαγωγή

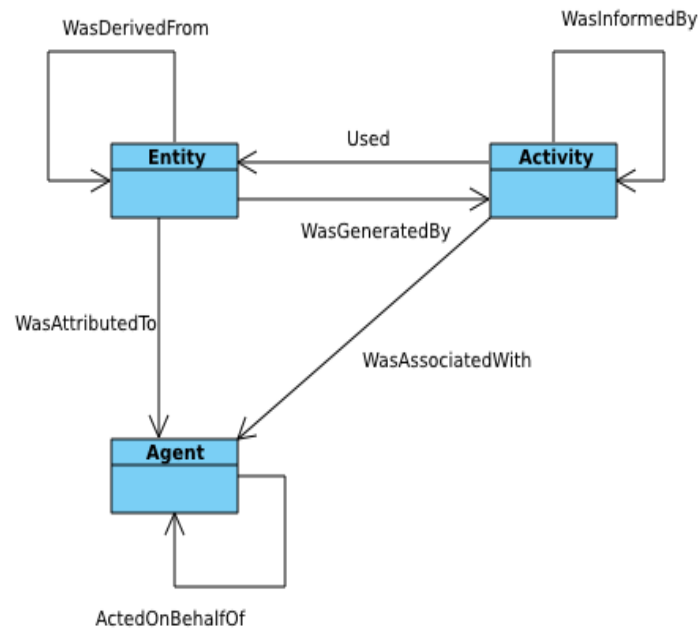
άρθρου, κύρια εξαγωγή, εξαγωγή μεγάλου περιεχομένου (κατάλληλο για κείμενα που δεν είναι άρθρα), ολόκληρη εξαγωγή (θεώρηση όλων ως περιεχόμενο). Ο συγγραφέας χρησιμοποίησε τον εξαγωγέα άρθρου προκειμένου να εξάγει το κύριο περιεχόμενο από τις ιστοσελίδες (επιλογή ιστοσελίδων με ελάχιστα 256 bytes περιεχομένου Urdu) (Kausar et al., 2020).

Σύμφωνα με τους Baeth και Aktas όμως τα απλά εργαλεία εξόρυξης ιστού αλλά και τα κοινωνικά δίκτυα δεν προβλέπουν εντοπισμό και προσδιορισμό αντικειμένων και δεδομένων για αυτό και είναι απαραίτητη η χρήση ειδικών συστημάτων. Οι συγγραφείς χρησιμοποίησαν το σύστημα ανοιχτού εξομοιωτή και δημιουργού γράφων Workflow Sim και Komadu για να δημιουργήσουν ένα υβριδικό μοντέλο ποιοτικής αξιολόγησης της πληροφορίας το οποίο συνέλεξε χαρακτηριστικά χρηστών και περιεχομένου. Το συγκεκριμένο μοντέλο είχε τη δυνατότητα να υπολογίζει την ποιότητα του αντικειμένου μέσω μετρικών των χρηστών (Baeth & Aktas, 2019).

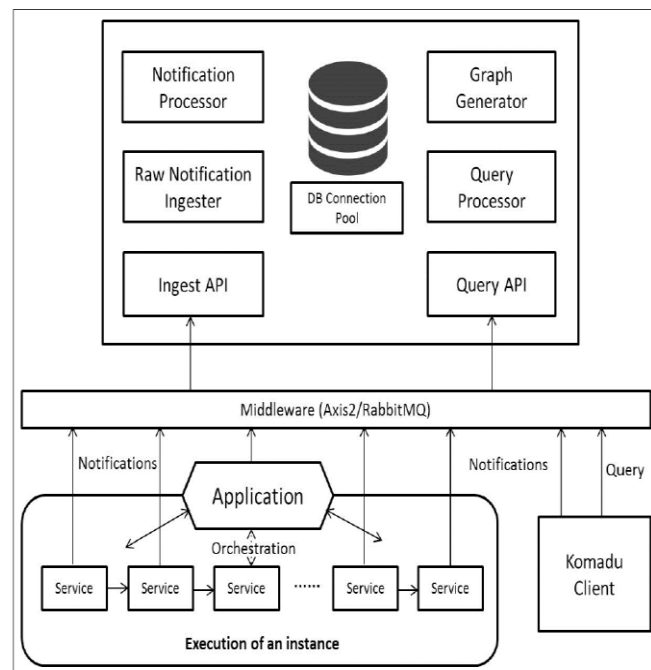
Το *Workflow Sim* και *Generator* είναι ένα σύστημα ανοιχτών ροών εργασίας και εξομοίωσης το οποίο χρησιμοποιεί ακυκλικούς γράφους (direct acyclic graph) (DAX), υποστηρίζει εφαρμογές από γνωστές δυναμικές ροές εργασίας αλλά και αλγόριθμους ομαδοποίησης εργασιών. Τα αρχεία DAX αναπαριστούν αφηρημένες περιγραφές μοναδικών ροών εργασίας σε μορφότυπο XML. Τα συγκεκριμένα αρχεία αποτελούν αρχεία καταγραφής των χρηστών (provenance data)⁸ στα κοινωνικά δίκτυα τα οποία αποτυπώθηκαν μέσω ειδικού εργαλείου (PROVToolbox) (Moreau, n.d.) και αποτελούν χαρακτηριστικά εισόδου στο σύστημα αναπαράστασης, οπτικοποίησης και διαχείρισης Komadu .

Το *Komadu* χρησιμοποιεί το πρότυπο W3C PROV. Το πρότυπο W3C είναι οργανωμένο με βάση θεματικές ομαδοποιήσεις εννοιών (*PROV-DM: The PROV Data Model*, n.d.) (Εικόνα 35/Εικόνα 36). Τα αρχεία καταγραφής είναι σημαντικά σύμφωνα με τον Moreau γιατί βοηθούν ως προς τον προσδιορισμό του εάν η πληροφορία είναι έγκυρη, εάν και πώς μπορεί να ενσωματωθεί με διαφορετικές πηγές πληροφόρησης, και πώς θα την πιστώσει στους δημιουργούς της όταν επαναχρησιμοποιηθεί (Moreau, n.d.).

⁸ provenance data: εγγραφές που περιγράφουν πρόσωπα, οντότητες, οργανισμούς και δραστηριότητες που παράγουν, επηρεάζουν ή μεταφέρουν μέρος δεδομένων ή αντικειμένων δεσ (Moreau, n.d.)



Εικόνα 35. Μοντέλο W3C Prov.



Εικόνα 36. Μοντέλο Komadu.

Οι Βραζονεαυ και Andonie υιοθέτησαν μία ολιστική προσέγγιση για τον εντοπισμό των ψευδών ειδήσεων χρησιμοποιώντας εκτός από την εξαγωγή και εντοπισμό των ονομαστικών οντοτήτων (DBpedia Spotlight) και τον γνωσιακό τους γράφο. Ο γράφος περιλάμβανε τριπλέτες οντοτήτων από τη βάση δεδομένων DBpedia. Οι σχέσεις που εξήγαγαν από τον γνωσιακό γράφο ήταν σε μορφή οντολογίας και είχαν εξαχθεί από

δηλώσεις ειδήσεων με προσδιορισμό των σχέσεων μέσω επισήμανσης POS (Parts of Speech) (Braşoveanu & Andonie, 2021).

Ο Horne αντίθετα επισήμανε ότι η παρουσία ιστοσελίδας στη Wikipedia υποδηλώνει την εγκυρότητα της πηγής υιοθετώντας την άποψη του Baly (Baly et al., 2018) ότι η απουσία ανάλογης ιστοσελίδας καταδεικνύει την μη εγκυρότητά της (B. D. Horne et al., 2020).

4.2 Γλωσσολογικά Χαρακτηριστικά

Οι ψευδείς ειδήσεις συνήθως υιοθετούν σκανδαλώδη (προβοκατόρικη) γλώσσα γνωστή και ως «clickbait» προκειμένου να ωθήσουν τον αναγνώστη να επιλέξει τον σύνδεσμο που οδηγεί στο άρθρο (Shu et al., 2017). Για αυτό το λόγο είναι απαραίτητη η χρήση των γλωσσολογικών χαρακτηριστικών τα οποία θα βοηθήσουν να αναγνωριστούν λεκτικά χαρακτηριστικά, χαρακτηριστικά πολυπλοκότητας και μετρικές αναγνωσιμότητας από το περιεχόμενο των ψευδών ειδήσεων και των σκανδαλοθηρικών πρωτοσέλιδων (clickbait).

Επιπλέον, σύμφωνα με τους Sotirakou και Sheikhi, *το μέγεθος κειμένου σε χαρακτήρες* (π.χ. τα ψευδή άρθρα έχουν λιγότερες τεχνικές λέξεις, μικρότερες λέξεις, λιγότερη στίξη, αποσιωπητικά και περισσότερους λεκτικούς πλεονασμούς), *το μέγεθος χαρακτήρων του τίτλου* (π.χ. στις ψευδείς ειδήσεις στο σύνολο των λέξεων στους τίτλους είναι πιο μεγάλοι οι τυπογραφικοί χαρακτήρες από ότι στις αληθείς) (B. D. Horne & Adali, 2017; Sotirakou et al., 2021), *αλλά και το σύνολο από διαστήματα, ειδικούς χαρακτήρες, σύμβολα (-, ..)* (π.χ. Οι τίτλοι των αληθινών ειδήσεων κατά μέσο όρο χρησιμοποιούν 34 χαρακτήρες) (Sheikhi, 2021) είναι κάποια χαρακτηριστικά των ψευδών ειδήσεων. Ειδικότερα στα clickbait οι επικεφαλίδες είναι διαφορετικές από την κύρια ιστορία (Sotirakou et al., 2021).

Αρκετές μελέτες της τρέχουσας βιβλιογραφικής επισκόπησης έχουν υιοθετήσει συγκεκριμένα γλωσσολογικά χαρακτηριστικά. Οι Gravanis et al, υιοθέτησαν γενικά χαρακτηριστικά όπως *η μέτρηση λέξεων, άρθρων, προθέσεων* αλλά και *ψυχολογικά (θετικά, αρνητικά συναισθήματα), σχετικότητας (χρόνο, ρήματα κίνησης, πολυπλοκότητας (σύντομες προτάσεις, εύρος λέξεων)* και άλλα συνδυαστικά χαρακτηριστικά (Gravanis et al., 2019).

Για την εύρεση των γλωσσολογικών χαρακτηριστικών υιοθετούνται ειδικές μέθοδοι ανάλυσης κειμένου (με πιο γνωστή τη LIWC) αλλά και τεχνικές επεξεργασίας φυσικής γλώσσας (NLP) (π.χ. NELA, NLTK).

Η LIWC (*Linguistic inquiry and word count*) είναι ένα λογισμικό ανάλυσης κειμένου το οποίο υπολογίζει το ποσοστό των λέξεων σε ένα δεδομένο κείμενο. Το συγκεκριμένο ποσοστό λέξεων αντιστοιχίζεται σε μία ή παραπάνω από 80 κατηγορίες (γλωσσολογικές,

ψυχολογικές, κοινωνιολογικές) που υπάρχουν μέσα στο λεξιλόγιο του λογισμικού (Ahmad et al., 2020; Barrón-Cedeño et al., 2019; Bonet-Jover et al., 2021; Gôlo et al., 2021; Gravanis et al., 2019; B. D. Horne et al., 2020; Huang & Chen, 2020; Isaakidou et al., 2021; Kausar et al., 2020; Khanday et al., 2021; Meel & Vishwakarma, 2020; Reis et al., 2019; Shu, Wang, et al., 2019; Shu, Cui, et al., 2019; Sotirakou et al., 2021). Ο χρήστης μπορεί επιπλέον να δημιουργήσει το δικό του λεξιλόγιο. Χωρίζεται σε **4 κύριες κατηγορίες**:

Συνοπτικές μετρήσεις: οι συγκεκριμένες μετρήσεις προσδιορίζουν τη συχνότητα των λέξεων η οποία αποτυπώνει σκέψη, οπτική και τιμιότητα συγγραφέα. Οι μετρήσεις είναι αποτέλεσμα εμπειρικών ερευνών χρόνων και είναι αλγόριθμοι οι οποίοι έχουν προέλευση από πολλαπλές μετρικές της LIWC. Οι αριθμοί είναι καθιερωμένα μεγέθη τα οποία έχουν μετατραπεί σε αριθμητικά χαρακτηριστικά (1-99) (Boyd et al., 2022).

- Αναλυτική σκέψη (Ψυχολογική διάσταση): η αναλυτική σκέψη είναι μία μεταβλητή η οποία βασίζεται σε διαφορετικές κατηγορίες λειτουργικών λέξεων. Αρχικά παρουσιάστηκε ως ένα Κατηγορικός-Δυναμικός Δείκτης ή CDI. Η Αναλυτική Σκέψη καταγράφει τον βαθμό στον οποίο οι άνθρωποι χρησιμοποιούν λέξεις που υποδηλώνουν τυπικά, λογικά και ιεραρχικά μοτίβα σκέψης. Για παράδειγμα στη μελέτη των Jordan, K. N., Sterling, J., Pennebaker, J. W, Boyd, R. για τη διερεύνηση των τάσεων στην πολιτική και στον πολιτισμό μέσω της γλώσσας των πολιτικών ηγετών επισημαίνεται ότι οι διαφορετικές κατηγορίες των λειτουργικών λέξεων αντανακλούν αναλυτική σκέψη αλλά και επιρροή (π.χ. υψηλός αριθμός άρθρων προθέσεων, πιο αναλυτική σκέψη, αφηγηματική και εμπειρική) (B. D. Horne & Adali, 2017).
- Επιρροή: Η επιρροή αναφέρεται στη σχετική κοινωνική θέση, την εμπιστοσύνη ή την ηγεσία που επιδεικνύουν οι άνθρωποι μέσω της γραφής ή της ομιλίας τους. Ο αλγόριθμος επιρροής αναπτύχθηκε με βάση τα αποτελέσματα μιας σειράς μελετών όπου οι άνθρωποι αλληλεπιδρούσαν μεταξύ τους (π.χ. Kacewicz et al., 2013). Η εξουσία γενικά αλλά και η ανάγκη για εξουσία, αντανακλά την προσοχή ή την επίγνωση των ανθρώπων για τη σχετική θέση σε ένα κοινωνικό περιβάλλον. Στη μελέτη των Kacewicz et al, αναφέρεται ότι άνθρωποι με υψηλότερη κοινωνική θέση χρησιμοποιούν λιγότερο αντωνυμίες πρώτου προσώπου, και περισσότερο αντωνυμίες δευτέρου ενικού προσώπου και πρώτου πληθυντικού (Kacewicz et al., 2014).
- Αυθεντικότητα: Όταν οι άνθρωποι αποκαλύπτονται με «αυθεντικό» ή ειλικρινή τρόπο, τείνουν να μιλούν πιο αυθόρμητα και δεν αυτοπεριορίζονται ή δεν

φιλτράρουν αυτά που λένε. Ο αλγόριθμος για την Αυθεντικότητα προήλθε αρχικά από μια σειρά μελετών όπου οι άνθρωποι παρακινήθηκαν να είναι ειλικρινείς ή απατηλοί (Newman et al., 2003) καθώς και από μια περίληψη μελετών εξαπάτησης που δημοσιεύθηκαν τα επόμενα χρόνια (Pennebaker, 2011). Ωστόσο, με τα χρόνια το μέτρο Αυθεντικότητας έχει λιγότερο να κάνει με την «εξαπάτηση» με την παραδοσιακή έννοια και αντιθέτως είναι περισσότερο μια αντανάκλαση του βαθμού στο οποίο ένα άτομο αυτο-ελέγχεται ή όχι. Παραδείγματα κειμένων με χαμηλή βαθμολογία Αυθεντικότητας περιλαμβάνουν προετοιμασμένα κείμενα (δηλαδή ομιλίες που γράφτηκαν νωρίτερα) και κείμενα όπου ένα άτομο είναι κοινωνικά επιφυλακτικό.

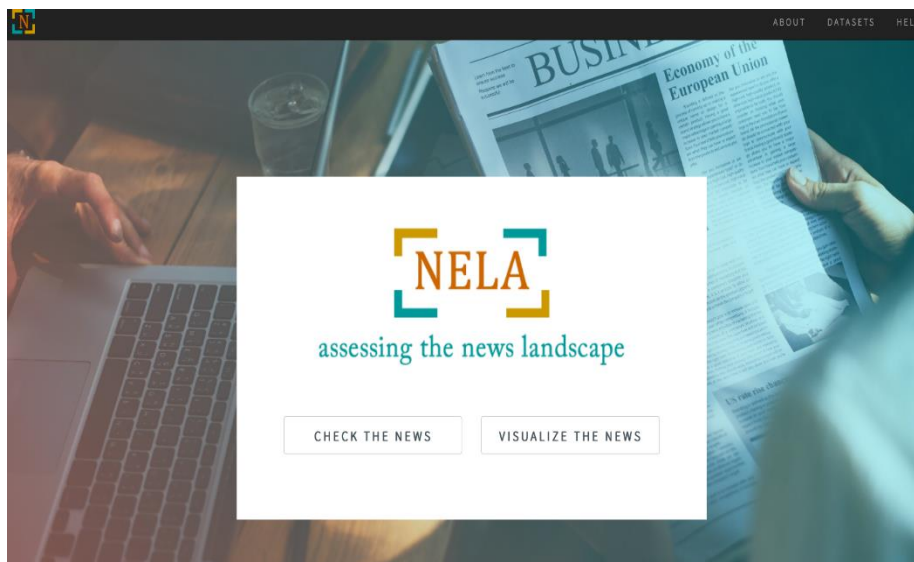
- **Συναισθηματικός τόνος:** Αν και η *L/WC* περιλαμβάνει διαστάσεις θετικού και αρνητικού τόνου, η μεταβλητή τόνου τοποθετεί τις δύο διαστάσεις σε μια ενιαία συνοπτική μεταβλητή. Ο αλγόριθμος είναι κατασκευασμένος έτσι ώστε όσο μεγαλύτερος είναι ο αριθμός, τόσο πιο θετικός είναι ο τόνος. Οι αριθμοί κάτω από το 50 υποδηλώνουν έναν πιο αρνητικό συναισθηματικό τόνο. Αντίστοιχη μελέτη για τις επιπτώσεις στο συναίσθημα αποτελεί και η μελέτη των Monzani, D., Vergani, L., Pizzoli, S. κ.α. (Monzani et al., 2021). Η συγκεκριμένη μελέτη αναλύει την επιρροή του Covid-19 με βάση την αποτύπωση του ως προς τη χρήση των λέξεων των ανθρώπων αναλύοντας γλωσσολογικά και ψυχολογικά χαρακτηριστικά από πληροφορίες δημοσιεύσεων των χρηστών του twitter.
- Λέξεις ανά πρόταση
- Μεγάλες λέξεις: με πάνω από 7 γράμματα
- Λέξεις λεξιλογίου
- **Γλωσσολογικά χαρακτηριστικά:** περιλαμβάνουν τον τρόπο λειτουργίας ενός κειμένου (μέσος όρος αριθμού λέξης ανά πρόταση, ρυθμός λανθασμένης ομιλίας). Μέσα σε αυτή την κατηγορία συμπεριλαμβάνονται οι αρνήσεις, και οι συχνότητες Part of Speech (ουσιαστικά, επίθετα, ρήματα), απρόσωπες αντωνυμίες, προσωπικές αντωνυμίες.
- **Ψυχολογικά χαρακτηριστικά:** στοχεύουν στο συναίσθημα, σε γνωστικές και κοινωνικές διαδικασίες. Περιλαμβάνουν:
 - **Χαρακτηριστικά επιρροής:** θετικά/αρνητικά συναισθήματα, άγχος θυμός, λύπη, βρισιές
 - **Κοινωνικά χαρακτηριστικά:** οικογένεια, φίλοι, γυναικείες αναφορές (γυναίκα, μητέρα), αντρικές αναφορές (αγόρι, πατέρας)

- **Γνωστικά χαρακτηριστικά:** διορατικότητα, επιρροή, διαφορετικότητα, επιφυλακτικότητα, βεβαιότητα, διαφοροποίηση

Εκτεταμένο λεξιλόγιο:

- Αντιληπτικά χαρακτηριστικά: βλέπω, ακούω, αισθάνομαι
- Πολιτισμικά: πολιτική, εθνικότητα, τεχνολογία
- Βιολογικά: σώμα, υγεία, σεξουαλικότητα, κατανάλωση
- Παρορμήσεις: δεσμός, επιτυχία, εξουσία
- Χρονικός προσανατολισμός: παρελθόν, παρόν, μέλλον
- Σχετικότητα: Δράση, διάστημα, ώρα
- Προσωπικοί προβληματισμοί: εργασία, σπίτι, λεφτά
- Χαρακτηριστικά συζήτησης: στο ίντερνετ (lol), συγκατάθεσης (ναι, αμέ..), έλλειψη ευφράδειας
- Συμπληρώματα (εννοώ, ξέρεις)

Η φυσική επεξεργασία γλώσσας (NLP) χρησιμοποιείται συμπληρωματικά με τους αλγόριθμους της μηχανικής μάθησης για εντοπισμό και κατηγοριοποίηση των ψευδών ειδήσεων. Το *NELA (News Landscape)* (B. Horne, 2020) είναι ένα ανάλογο εργαλείο το οποίο αποτελείται από επιμέρους αλγοριθμικά μοντέλα για την αξιοπιστία των ειδήσεων αλλά και για την οπτικοποίηση τους (π.χ. πολυπλοκότητα και γραφή, συναίσθημα, μεροληψία, ψυχολογικά-LIWC, συμμετοχικότητα) (Εικόνα 37).



Εικόνα 37. NELA Toolkit.

Αποτελείται από ανεξάρτητα μοντέλα που λειτουργούν σε επίπεδο άρθρων: πρόβλεψη αξιοπιστίας, πρόβλεψη πολιτικής αμεροληψίας, πρόβλεψη αντικειμενικότητας κειμένου, πρόβλεψη ενδιαφέροντος κοινότητας Reddit. Επίσης, σε επίπεδο πηγής: πρόβλεψη αξιοπιστίας, πρόβλεψη πολιτικής αμεροληψίας και οπτικοποίηση χαρακτηριστικών βάση περιεχομένου (B. D. Horne et al., 2018). Έχει υιοθετήσει αρκετά στοιχεία από τη LIWC (ψυχολογολογικά, με εξειδικευμένο θέμα-topic specific). Τα χαρακτηριστικά που χρησιμοποιεί το NELA είναι:

- **Επίπεδο γραφής και πολυπλοκότητας**
 - SMOG Grade readability
 - Clickbait
 - Flench-Kincaid
 - Λεκτική πολυπλοκότητα (Lexical diversity)
- **Συναισθηματικά χαρακτηριστικά και μετρήσεις**
 - Vader Sentiment
 - Θετικές/αρνητικές λέξεις
 - Λέξεις θυμού/οργής
- **Μεροληπτικά χαρακτηριστικά** (π.χ. λέξεις μεροληψίας, hedges, λέξεις γνώμης)
- **Ψυχολογικά χαρακτηριστικά (υιοθέτηση από LIWC)** (π.χ. γνωστικές λέξεις, κοινωνικές λέξεις)
- **Συμμετοχικά χαρακτηριστικά (engagement)** (π.χ. αριθμός διαμοιρασμών, σχολίων, αντιδράσεων)
- **Ηθικά χαρακτηριστικά** (βασισμένα στη θεωρία ηθικών θεμελίων (π.χ. αγνότητα, προδοσία, υποταγή κλπ.)
- **Γραμματικά & συντακτικά χαρακτηριστικά** (Μέρη λόγου-Part of Speech) (π.χ. μέτρηση επιθέτων, ουσιαστικών κ.λπ.)
- **Εξειδικευμένου θέματος (υιοθέτηση από LIWC)** (π.χ. θρησκευτικές λέξεις, λέξεις κίνησης, εργασιακές λέξεις, ψυχαγωγικές κ.λπ.)

Τα χαρακτηριστικά του NELA έχουν υιοθετηθεί για τον εντοπισμό των προπαγανδιστικών άρθρων σε μοντέλα εντοπισμού ψευδών ειδήσεων όπως το Propgy (Barrón-Cedeño et al., 2019) και το ProSoul (Kausar et al., 2020).

Η *NLTK (Natural language toolkit)* (*NLTK :: Natural Language Toolkit*, 2022) είναι μία πλατφόρμα δημιουργίας προγραμμάτων στη γλώσσα Python με σκοπό τη λειτουργία με δεδομένα ανθρώπινης γλώσσας. Περιλαμβάνει πάνω από 50 βάσεις δεδομένων και λεκτικές πηγές όπως το WordNet (*WordNet*, 2010), μαζί με τεχνικές επεξεργασίας κειμένου για

κατηγοριοποίηση, λεκτική ανάλυση (tokenization), λημματοποίηση, tagging, διαχωρισμό προτάσεων και σημασιολογική ερμηνεία, συναρτήσεις wrappers για βιβλιοθήκες της Φυσικής επεξεργασίας γλώσσας (NLP). Η NLTK έχει χρησιμοποιηθεί ως προς την λεκτική ανάλυση και στην τρέχουσα βιβλιογραφία (Barrón-Cedeño et al., 2019; Braşoveanu & Andonie, 2021; Cremisini et al., 2019; Hunt et al., 2022; Isaakidou et al., 2021; Kar, 2020; Kausar et al., 2020; Khanday et al., 2021; Reis et al., 2019; Shim et al., 2021; Silva et al., 2020; Sotirakou et al., 2021; Taskin et al., 2022; Tundis et al., 2021; Varshney & Vishwakarma, 2021; Vishwakarma et al., 2019). Το SpaCy (*SpaCy · Industrial-Strength Natural Language Processing in Python*, n.d.) είναι ένα ακόμα εργαλείο NLP το οποίο έχει κοινά χαρακτηριστικά με την NLTK δεδομένου ότι χρησιμοποιείται στη λεκτική ανάλυση κειμένου (Du et al., 2021).

Το *Open Korea text* είναι μία τεχνική NLP για την ανάλυση και την εξαγωγή μορφημάτων η οποία χρησιμοποιήθηκε από τον Shim (Shim et al., 2021). Εκτός από την ανάλυση και εξαγωγή μορφημάτων⁹ χρησιμοποιείται και για εξαγωγή φράσεων αλλά και μία σειρά διαδικασιών, την NLP (π.χ. λημματοποίηση) (Hohyon Ryu, 2014).

Εκτός από τα πακέτα Φυσικής επεξεργασίας Γλώσσας χρησιμοποιούνται και ειδικές μέθοδοι όπως ο υπολογισμός διανυσμάτων (Count vectorizer). Η μέθοδος του μετρητή διανυσμάτων αφορά την ανάλυση κειμένων όπου αναζητούνται λέξεις που εμφανίζονται συχνά στα κείμενα καθώς και τη μετατροπή τους σε διανύσματα. Στα διανύσματα εμπεριέχεται ο αριθμός των συχνοτήτων των λέξεων σε κάθε κείμενο ή η συχνότητα εμφάνισής τους. Η συγκεκριμένη μέθοδος συμπεριλαμβάνεται στο λογισμικό πακέτο Φυσικής επεξεργασίας Γλώσσας Scikit και χρησιμοποιείται για την εξαγωγή θεμάτων σε ένα κείμενο (π.χ. μέσω του αλγόριθμου NMF-Non Negative Matrix Factorization) (*CountVectorizer*, n.d.; Ksieniewicz et al., 2019).

Στο λεκτικό επίπεδο της γλώσσας συμπεριλαμβάνονται το επίπεδο χαρακτήρων και λέξεων όπως ο αριθμός των μοναδικών λέξεων και οι συχνότητές τους στο κείμενο (Shu et al., 2017). Πολλά και διαφορετικά χαρακτηριστικά έχουν χρησιμοποιηθεί από τους συγγραφείς της βιβλιογραφίας όπως οι αριθμοί των λέξεων (σε τίτλο και σώμα) (π.χ. 1000 λέξεις στο κείμενο και 6 στον τίτλο), πρώτο πρόσωπο αντωνυμιών, δεύτερο πρόσωπο αντωνυμιών, δεικτικές αντωνυμίες, τροπικά ρήματα, συγκεκριμένες λέξεις κλειδιά για τον εντοπισμό των ψευδών ειδήσεων (π.χ. αναφορά, μελέτη, διακόπτω), ειδικές λέξεις κλειδιά (π.χ. αβέβαιος, περίπου, συνεργάτες) (αποφυγή πρακτορείων ειδήσεων ως προς τη χρήση τους σε τίτλο και σώμα), αν ο τίτλος και το σώμα περιλαμβάνουν συντομογραφίες (π.χ. b4,

⁹ Στη γλωσσολογία μόρφημα είναι λέξεις όπως καλ-, -ος.

jk-just kidding, irl) (Vosoughi et al., 2017), μετρήσεις στίξης, προτάσεις, αριθμός προτάσεων (Meel & Vishwakarma, 2020), υπέρνυμα (=γενικής σημασίας λέξεις), υποώνυμα (=λέξεις που εξειδικεύονται), επικαλυπτόμενες λέξεις (π.χ. ανάμεσα σε τίτλο και σώμα), το μέγιστο μήκος δύο κειμένων σε σχέση με τον αριθμό των λέξεων που υπάρχουν στο κείμενο (*longest common overlap*), τροπικά ρήματα, τρίτο πρόσωπο αντωνυμιών, ρήματα κίνησης, ρήματα ενεστωτικά, ρήματα παρελθοντικά, ρήματα μελλοντικά, πλεονασμοί, γερούνδιο, μέσος όρος ανά στίξη, *hashtags*, ποσοστό *stopwords* (π.χ. το, είναι), ποικιλία ουσιαστικών, ποικιλία ρημάτων, μέσος όρος μήκους πρότασης ή πρότασης με ουσιαστικό, ο χρόνος, η διαφορετικότητα περιεχομένου λέξεων, δηλαδή ο συνολικός αριθμός διαφορετικών λέξεων στο περιεχόμενο ή ο συνολικός αριθμός των λέξεων περιεχομένου ή των όρων, παύση, αβεβαιότητα, τυπογραφικά λάθη, υπολογισμός *hashtag* σε ένα κείμενο (π.χ. *#FollowFriday*) (Abonizio et al., 2020; Ahmad et al., 2020; Braşoveanu & Andonie, 2021; Gravanis et al., 2019; Kasseropoulos & Tjortjis, 2021; Meel & Vishwakarma, 2020; A. C. Pandey & Tikkiwal, 2021; Reis et al., 2019; Saikh et al., 2019; Sheikhi, 2021; Shu, Wang, et al., 2019; Silva et al., 2020; Tundis et al., 2021).

Τα συγκεκριμένα χαρακτηριστικά έχουν ξαναχρησιμοποιηθεί και σε προηγούμενες βιβλιογραφίες για τον εντοπισμό των προφίλ απάτης (Newman et al., 2003; L. Zhou et al., 2003). Αξίζει να επισημανθεί ότι το τρίτο ενικό, το τρίτο πληθυντικό πρόσωπο και τα επιρρήματα τρόπου (π.χ. θα μπορούσα, θα ήθελα, μπορεί) χρησιμοποιούνται σε μεγάλο βαθμό στις ψευδείς ειδήσεις σύμφωνα με τους Kasseropoulos και Tjortjis (Kasseropoulos & Tjortjis, 2021).

Ως προς τον αριθμό προτάσεων οι πραγματικές ειδήσεις αποτελούνται από περισσότερες προτάσεις σε σχέση με τις ψευδείς ειδήσεις. Πιο συγκεκριμένα, σε μελέτη του Del Vicario οι πραγματικές ειδήσεις είχαν 84 προτάσεις ενώ οι ψευδείς 69 (Meel & Vishwakarma, 2020; Vicario et al., 2019).

Σύμφωνα με μελέτη των Yang et al οι ψευδείς ειδήσεις έχουν λιγότερες αντωνυμίες πρώτου προσώπου. Ένας ψεύτης για παράδειγμα δεν θα πει ότι “Εγώ δεν πήρα το βιβλίο σου” αλλά “Αυτό δεν θα το έκανε κάποιος με ακεραιότητα» (Meel & Vishwakarma, 2020; Y. Yang et al., 2022). Επιπλέον, η χρήση του πρώτου προσώπου είναι πιο συχνή σε αληθή άρθρα. Αυτό συμβαίνει γιατί οι συγγραφείς που θέλουν να εξαπατήσουν τους αναγνώστες τείνουν να διαχωρίζουν τους εαυτούς τους από την πληροφορία που διαδίδουν (Kasseropoulos & Tjortjis, 2021; L. Zhou et al., 2003).

Τα ρήματα αναφέρονται στη μελέτη των Gravanis et al (ρήματα κίνησης, ρήματα ενεστωτικά, παρελθοντικά, μελλοντικά, τροπικά) και υιοθετήθηκαν από προηγούμενες

μελέτες εντοπισμού της εξαπάτησης (Newman et al., 2003; L. Zhou et al., 2004) αλλά και από άρθρα της τρέχουσας βιβλιογραφίας (Ahmad et al., 2020; Braşoveanu & Andonie, 2021; Kausar et al., 2020; Reis et al., 2019). Το ρήμα είναι μία λέξη που αποτελεί το κέντρο (σε σχέση με τη γραμματική) ενός κατηγορήματος (υποκείμενο πρότασης) και εκφράζει την πράξη, τη συχνότητα αλλά και τον τρόπο ύπαρξης. Οι Zhou et al τα χρησιμοποίησαν για να ερευνηθεί η ποσότητα των ρημάτων στα παραπλανητικά κείμενα (π.χ. διαπράττω, δείχνω, παράγω). (L. Zhou et al., 2004). Ενώ σε μελέτες του Πανεπιστημίου «Southern Methodist» βρέθηκε ότι χαρακτηριστικά όπως τα ρήματα κίνησης (π.χ. περπατώ, κινούμαι, πηγαίνω), οι άνθρωποι που λένε ψέματα τα χρησιμοποιούν περισσότερο (χαμηλή γνωστική πολυπλοκότητα) (Gravanis et al., 2019; Newman et al., 2003). Τα ενεστωτικά, παρελθοντικά, και μελλοντικά ρήματα αποτελούν μέρος της κατηγορίας της LIWC που αφορά την «τοποθεσία στο χρόνο» (π.χ. πηγαίνω, είναι, είμαι, περπατούσε, ήταν, είχε, θέλω, ίσως, μπορώ να) (Gravanis et al., 2019; Newman et al., 2003).

Το χαρακτηριστικό του χρόνου συμπεριλαμβάνεται στην κατηγορία «εξειδίκευση θέματος» της LIWC όπου βρίσκεται το θέμα συζήτησης στο κείμενο μετρώντας τις κατηγοριοποιημένες λέξεις με χρήση του λεξικού της LIWC και αναφέρεται στις έρευνες των Gravanis και Kausar (Gravanis et al., 2019; Kausar et al., 2020).

Η αβεβαιότητα σχετίζεται με τα επιρρήματα και τις αντωνυμίες τρίτου προσώπου αλλά και γενικούς όρους (π.χ. όλοι) όπου ενισχύουν την αβεβαιότητα γιατί οι εξαπατητές είναι λιγότερο ξεκάθαροι χρησιμοποιώντας αντιθετική και ακατανόητη δομή προτάσεων (Gravanis et al., 2019; Silva et al., 2020; L. Zhou et al., 2004). Ενώ σε σχέση με τη διαφορετικότητα του περιεχομένου των λέξεων, ο Zhou επισημαίνει ότι οι εξαπατητές δεν συμπεριλαμβάνουν εξειδικευμένο περιεχόμενο αλλά και ποικιλία περιεχομένου. Δηλαδή υπάρχει έλλειψη στην περιγραφή ως προς την διαφορετικότητα των γεγονότων (Gravanis et al., 2019; L. Zhou et al., 2004).

Τα τυπογραφικά λάθη αναφέρονται στον συνολικό αριθμό λέξεων που είναι λανθασμένα ειπωμένες σε σχέση με τον συνολικό αριθμό λέξεων σε ένα κείμενο. Στις ψευδείς ειδήσεις υπάρχουν περισσότερα λάθη προφοράς από ότι στις αληθείς. Οι εξαπατητές γενικά χρησιμοποιούν ανεπίσημη γλώσσα μέσω τυπογραφικών λαθών (π.χ. 36% περισσότερα λάθη στις ψευδείς ειδήσεις και μόνο 3% στις αληθείς) (Gravanis et al., 2019; Silva et al., 2020).

Οι Meel και Viswakarma αναφέρουν στη μελέτη τους και γλωσσολογικά χαρακτηριστικά τα οποία αφορούν δημοσιεύσεις φημών σε κοινωνικά δίκτυα (twitter), τα οποία έχουν χρησιμοποιηθεί και σε άλλες έρευνες (Vosoughi et al., 2017) όπως την *αναλογία*

tweet που περιλαμβάνουν αρνήσεις (π.χ. η Sunil Tripathi ΔΕΝ είναι ύποπτη για τον Μαραθώνιο της Βοστώνης the Boston Marathon suspect), τον μέσο όρο τυπικότητας και πολιτισμού στα *tweet* δηλαδή την ύπαρξη χυδαίων λέξεων (*vulgarity*), συντομογραφιών (*abbreviations*) (π.χ. b4, jk-just kidding, irl) (Sheikhi, 2021), *emoticons*, την αναλογία *tweet* που περιλαμβάνουν γνώμη και εσωτερικότητα (π.χ. λέξεις ξέρω, θεωρώ, σκέφτομαι) και την αναλογία έμμεσων και επιφυλακτικών *tweet* με υιοθέτηση στα τελευταία δύο χαρακτηριστικά των προσεγγίσεων της LIWC (Meel & Vishwakarma, 2020; Vosoughi et al., 2017).

Σε επίπεδο πολυπλοκότητας η ποικιλία λεξιλογίου διαμορφώνει το αν μια είδηση είναι ψευδής ή όχι σε ένα κείμενο όπως οι μοναδικές λέξεις που υπάρχουν σε αυτό (*type token ratio*) (Abonizio et al., 2020; Barrón-Cedeño et al., 2019; B. D. Horne et al., 2020; Kausar et al., 2020; Silva et al., 2020; Sotirakou et al., 2021), η αναλογία μεγάλων γραμμάτων στο κείμενο, μέσος όρος μεγέθους λέξης κειμένου (π.χ. Όσο περισσότερες οι λέξεις τόσο μεγαλύτερη η πολυπλοκότητα) (B. D. Horne & Adali, 2017) και μέσο μήκος πολυπλοκότητας λέξεων *tweet* (*average word complexity*) (π.χ. το «Υπάρχει άλλη μία βόμβα στο Harvard» έχει 6 λέξεις και 7, 4, 3, 5, 3, 7 χαρακτήρες και ο μέσος όρος είναι περίπου 5 (4,8) (Vosoughi et al., 2017), μέσος όρος λέξης ανά πρόταση (π.χ. 3943 λέξεις σε ψευδείς ειδήσεις ενώ 4360 σε αληθείς) (Y. Yang et al., 2022), μεγάλες λέξεις (π.χ. λέξεις με πάνω από 6 γράμματα) (Gravanis et al., 2019; Kasseropoulos & Tjortjais, 2021).

Αλλά και χαρακτηριστικά όπως οι συλλαβές ανά λέξη, το μέγεθος της λέξης, η μέτρηση προτάσεων (π.χ. στους εξαπατητές οι προτάσεις είναι περισσότερες σε σχέση με τα ειλικρινή άτομα), η μέτρηση των λέξεων (Abonizio et al., 2020; Tiantian Qin et al., 2005), η ποικιλία των λειτουργικών λέξεων (δηλαδή το σύνολο αντωνυμιών πρώτου, δευτέρου, τρίτου προσώπου) (Kasseropoulos & Tjortjais, 2021; Kausar et al., 2020), αριθμοί χαρακτήρων, πολύπλοκες λέξεις, δηλαδή λέξεις με πάνω από 3 ή περισσότερες συλλαβές αλλά όχι ουσιαστικά ονομάτων, συνήθης φρασεολογία και σύνθετες λέξεις (π.χ. η λέξη «ενδιαφέρον» αν και έχει πάνω από 3 συλλαβές δε σημαίνει ότι είναι και δύσκολη) (Barrón-Cedeño et al., 2019; Bonet-Jover et al., 2021; “Gunning Fog Index,” 2022; Pérez-Rosas et al., 2018), λέξεις μεγάλες σε μήκος δηλαδή προτάσεις που περιλαμβάνουν παραπάνω λέξεις προκειμένου να διατυπώσουν κάτι σαφές¹⁰ (Barrón-Cedeño et al., 2019; B. D. Horne et al., 2018; Kausar et al., 2020), μήκος πρότασης (στα αποτελέσματα της έρευνας του Khanday το μήκος των προτάσεων των προπαγανδιστικών κειμένων ήταν μεγαλύτερο σε σχέση με τα μη

¹⁰ δεξ και (*Examples of Wordy Sentences - and How to Correct Wordiness*, 2019)

προπαγανδιστικά) (B. D. Horne et al., 2020; Khanday et al., 2021), *τύπους λέξεων, αριθμός συλλαβών, αριθμός παραγράφων* (Bonet-Jover et al., 2021), *το σύνολο αριθμών σε τίτλο ή κείμενο* (Y. Liu et al., 2019; Saikh et al., 2019), *οι χαρακτήρες σε τίτλους ή κείμενο (μήκος)* (Y. Liu et al., 2019), *αν ξεκινούν ο τίτλος και το σώμα του κειμένου με αριθμό* (Sheikhi, 2021) *η αναλογία ουσιαστικών με ρήματα* (π.χ. στις αληθείς ειδήσεις η αναλογία των ουσιαστικών με τα ρήματα είναι υψηλότερη -4.27- σε σχέση με τις ψευδείς -3.73) (Marquardt, 2019; Sotirakou et al., 2021), *οι χαρακτήρες σε μικρές ή μεγάλες παραγράφους* (Y. Liu et al., 2019), *συχνές λέξεις* (Meel & Vishwakarma, 2020), *αντιθετικές λέξεις στον τίτλο* (Borges et al., 2019), *ο αριθμός ομιλούμενων λέξεων* (Silva et al., 2020), *ο μέσος όρος κάθε παραγράφου* (Y. Liu et al., 2019), *το ποσοστό των συνηθισμένων λέξεων (stopwords)* (π.χ. ένας, το) με τη χρήση λεξικών όπως η LIWC (Ahmad et al., 2020), *τη δημιουργία λιστών λεξιλογίου διαφορετικών όρων που εμφανίζονται στον τίτλο και στο κείμενο αλλά και μετρικές* (Abonizio et al., 2020; Ahmad et al., 2020; Barrón-Cedeño et al., 2019; Bonet-Jover et al., 2021; Borges et al., 2019; Gravanis et al., 2019; B. D. Horne et al., 2020; Kasseropoulos & Tjortjjs, 2021; Kausar et al., 2020; Y. Liu et al., 2019; Meel & Vishwakarma, 2020; A. C. Pandey & Tikkiwal, 2021; Saikh et al., 2019; Silva et al., 2020; Sotirakou et al., 2021). Οι λίστες λεξιλογίου διαφορετικών όρων χρησιμοποιήθηκαν από τους Pandey και Tikkiwal προκειμένου να εξάγουν προεκπαιδευμένες αναπαραστάσεις λέξεων (word embeddings) από τη γνωσιακή βάση (A. C. Pandey & Tikkiwal, 2021).

Η αναλογία μοναδικών λέξεων σε ένα έγγραφο υπολογίζεται από τον αριθμό των μοναδικών λέξεων σε σχέση με το συνολικό αριθμό λέξεων σε ένα έγγραφο. Η αναλογία μοναδικών λέξεων επισημαίνει τη λεκτική διαφοροποίηση του λεξιλογίου σε ένα έγγραφο αλλά και πλούσιο λεξιλόγιο (B. D. Horne & Adali, 2017; Kausar et al., 2020). Εάν υπάρχει χαμηλή αναλογία μοναδικών λέξεων τότε το έγγραφο έχει περισσότερες επαναλήψεις λέξεων ενώ η υψηλή αναλογία ότι το έγγραφο έχει περισσότερη λεκτική διαφοροποίηση. Σύμφωνα με τον Abonizio ο βαθμός της αναλογίας των μοναδικών λέξεων στις ψευδείς ειδήσεις είναι χαμηλότερος σε σχέση με τις αληθείς (Abonizio et al., 2020; Barrón-Cedeño et al., 2019; B. D. Horne et al., 2020; B. D. Horne & Adali, 2017; Kausar et al., 2020; Silva et al., 2020; Sotirakou et al., 2021).

Οι μεγάλες σε μήκος λέξεις υιοθετήθηκαν από τους Gravanis και Kasseropoulos και Tjortjjs. Η πολυπλοκότητα στα κείμενα των εξαπατητών είναι μικρότερη καθώς χρησιμοποιούν μικρότερες προτάσεις και αδυνατούν να δημιουργήσουν πολύπλοκα μηνύματα (Burgoon et al., 2003; Gravanis et al., 2019; Kasseropoulos & Tjortjjs, 2021).

Οι χαρακτήρες σε τίτλους ή κείμενο (μήκος) αφορούν μη αξιόπιστα άρθρα των οποίων οι τίτλοι τείνουν να είναι πιο μεγάλοι και να έχουν περισσότερα θαυμαστικά από ότι ερωτηματικά γιατί τα «γεγονότα» σε αυτούς τους τίτλους είναι υπερβολικά και οι ψεύτικοι αριθμοί επισημαίνουν τη σημασία τους (π.χ. «Η ρίζα της πικραλίδας σκοτώνει 98% τα καρκινικά κύτταρα μέσα σε 48 ώρες!») (Y. Liu et al., 2019). Ενώ οι αντιθετικές λέξεις σύμφωνα με τον Borges βασίζονται στην παρουσία αντιθετικών λέξεων στον τίτλο και βρίσκονται στη λίστα ενός δεδομένου λεξιλογίου (π.χ. αρνούμαι, αμφιβάλλω, απάτη, εκθέτω) (Borges et al., 2019).

Η εξαγωγή του ποσοστού των συνηθισμένων λέξεων (stopwords) έγινε από την LIWC. Οι συνηθισμένες λέξεις στη μηχανική μάθηση συνήθως απαλείφονται ή φιλτράρονται στην προεπεξεργασία γιατί δεν αποτελούν σημαντική πληροφόρηση. Στο πεδίο έρευνας και εντοπισμού των ψευδών ειδήσεων (όπως στην περίπτωση της μελέτης των Meel & Vishwakarma) χρησιμοποιούνται ως χαρακτηριστικά για την εκπαίδευση των αλγοριθμικών μοντέλων (Ahmad et al., 2020; Meel & Vishwakarma, 2020).

Λέξεις που περιλαμβάνουν πολλές συλλαβές είναι δυσκολότερο να διαβαστούν σε σχέση με λέξεις που έχουν λιγότερες συλλαβές. Το συγκεκριμένο χαρακτηριστικό όπως και οι μεγάλες σε μήκος λέξεις, οι πολύπλοκες λέξεις υιοθετήθηκαν από τον Pérez-Rosas ως χαρακτηριστικά που προσδιορίζουν την κατανόηση του κειμένου. Επιπλέον χρησιμοποιούνται στις μετρικές αναγνωσιμότητας (Flench-Kincaid Level Grade, Gunning Fog Index, Flesch Reading Ease) (Bonet-Jover et al., 2021; Pérez-Rosas et al., 2018).

Οι τύποι μετρικών πολυπλοκότητας που χρησιμοποιήθηκαν στην παρούσα βιβλιογραφία είναι η στατιστική μέθοδος Honoré (Honoré, 1979) και η μέτρηση Yule (Yule K means). Στη μέτρηση Honoré υπολογίζονται οι λέξεις, οι τύποι των λέξεων, λέξεις που υπάρχουν μόνο μία φορά (Harax legomena) σε όλο το κείμενο ή και το έργο ή λέξεις που υπάρχουν δύο φορές στο κείμενο (Harax dislegomena ή Michea's M and Sichel's S). Η στατιστική μέθοδος Honoré είναι ουσιαστικά ένα ευρετήριο πολυπληθούς λεξιλογίου το οποίο χρησιμοποιείται στη στυλομετρική ανάλυση κειμένου. Τα κείμενα που έχουν πλούσιο λεξιλόγιο έχουν υψηλότερο αριθμό σε λέξεις harax legomena (π.χ. η χρήση της μεθόδου σε ασθενείς με alzheimer) (Barrón-Cedeño et al., 2019; Tanaka-Ishii & Aihara, 2015).

Η μέτρηση Yule αφορά συνδυασμούς τύπων που υπάρχουν στο κείμενο με διαφορετικές συχνότητες και λέξεις. Όσο πιο πολλές λέξεις επαναλαμβάνονται τόσο θα μεγαλώνει η μέτρηση K του Yule και το λεξιλόγιο σε αυτήν τη περίπτωση δε θα είναι τόσο πλούσιο. Για παράδειγμα εάν ένα κείμενο έχει 10 λέξεις, εάν η κάθε μία είναι διαφορετική τότε θα έχουμε $1 \times 1 \times 10 = 10$ (υψηλή διαφορετικότητα) ενώ αν η κάθε μία από τις 10 είναι

ταυτόσημη τότε θα έχουμε $10 \times 10 \times 1 = 100$ (χαμηλή διαφορετικότητα). Η αρχική χρήση της μετρικής από τον Yule ήταν για τον προσδιορισμό του συγγραφέα (ταυτοποίηση) υποθέτοντας ότι θα διαφέρει από κείμενα τα οποία έχουν γραφτεί από διαφορετικούς συγγραφείς (Tanaka-Ishii & Aihara, 2015). Οι συγκεκριμένες μετρικές έχουν υιοθετηθεί από τον Barrón-Cedeño (Barrón-Cedeño et al., 2019) στο μοντέλο Proppy ως χαρακτηριστικά πολυπλοκότητας και διαφορετικότητας για τη μέτρηση της προπαγάνδας στις ψευδείς ειδήσεις.

Εκτός από τις μετρικές πολυπλοκότητας χρησιμοποιούνται και μετρικές αναγνωσιμότητας όπως τα Flesch-Kincaid level grade, Flesch Reading Ease, τα ευρετήρια Coleman-Liau (Coleman-Liau index), Gunning Fog (Gunning Fog Index) και ARI (Automated Readability Index).

Το Flesch-Kincaid level grade (Barrón-Cedeño et al., 2019; Bonet-Jover et al., 2021; Kausar et al., 2020; Meel & Vishwakarma, 2020; Sotirakou et al., 2021) και το Flesch Reading Ease (Barrón-Cedeño et al., 2019; Bonet-Jover et al., 2021; Meel & Vishwakarma, 2020; Papadopoulou et al., 2019; Sotirakou et al., 2021) χρησιμοποιούνται για την κατανόηση του κειμένου στην εκπαίδευση των Η.Π.Α. Το Flesch-Kincaid level grade μετρά την κατανόηση ενός κειμένου και δείχνει τον αριθμό της απαιτούμενης εκπαίδευσης που χρειάζεται για να κατανοηθεί το κείμενο. Ως προς την κατανόηση του κειμένου μετρά τον σύνολο των λέξεων, των προτάσεων και των συλλαβών ενός κειμένου. Όσο πιο χαμηλός είναι ο βαθμός τόσο πιο εύκολα κατανοητό είναι ένα κείμενο (π.χ. -3.40) (*THE FLESCH GRADE LEVEL READABILITY FORMULA*, n.d.). Αντίθετα το Flesch Reading Ease μετρά σε κλίμακα 1-100, με το 100 να δηλώνει την ευκολία ανάγνωσης και κατανόησης του κειμένου. Έτσι όσο πιο υψηλό είναι το σκορ ενός κειμένου τόσο πιο εύκολη είναι η ανάγνωση του. Η μετρική υπολογίζει τον μέσο όρο του μήκους της πρότασης (σύνολο λέξεων σε σχέση με αριθμό των προτάσεων) αλλά και τον μέσο όρο των συλλαβών (αριθμός συλλαβών σε σχέση με αριθμό λέξεων) (π.χ. βαθμοί από 90-100 το κείμενο είναι κατανοητό από 5η δημοτικού) (*THE FLESCH READING EASE READABILITY FORMULA*, n.d.).

Το Ευρετήριο Gunning Fog (*Gunning Fog Index*, n.d.) προϋποθέτει τον αναγκαίο αριθμό χρόνων τυπικής εκπαίδευσης για κατανόηση κειμένου. Υπολογίζεται από τον μέσο όρο του αριθμού των λέξεων ανά πρόταση, τον αριθμό των μεγάλων λέξεων ανά λέξη και τον αριθμό των πολύπλοκων λέξεων (π.χ. λέξεις με πάνω από 3 συλλαβές) που υπάρχουν στο κείμενο. Για παράδειγμα αν ο δείκτης του ευρετηρίου δείξει πάνω από 12 τότε ο αναγνώστης θα έχει δυσκολία στην ανάγνωση του κειμένου γιατί το πραγματικό σκορ ανάγνωσης βάση του ευρετηρίου είναι 7 με 8 (*THE GUNNING FOG READABILITY FORMULA*, n.d.). Ο σκοπός της

συγκεκριμένης μετρικής είναι να αναδείξει ότι οι σύντομες προτάσεις σε απλά αγγλικά έχουν καλύτερα αποτελέσματα σε σχέση με μεγάλες σε μέγεθος και πολύπλοκες προτάσεις (Barrón-Cedeño et al., 2019; Bonet-Jover et al., 2021; Meel & Vishwakarma, 2020; Sotirakou et al., 2021).

Το ευρετήριο SMOG (Abonizio et al., 2020; Kausar et al., 2020) προϋποθέτει και αυτό τον τυπικό χρόνο εκπαίδευσης και ουσιαστικά είναι μια παραλλαγή του ευρετηρίου Gunning Fog μόνο που εδώ υπολογίζονται οι πολυσύλλαβες λέξεις. Για παράδειγμα, εάν η βαθμολογία είναι 7,4 υποδηλώνει ότι το κείμενο είναι κατανοητό από έναν μέσο μαθητή στην 7η δημοτικού. Για σύνολο πολυσυλλαβικών λέξεων από 1-6 ο βαθμός επιπέδου με βάση τη μετρική ορίζεται σε 5 ενώ για σύνολο 7-12 ο βαθμός επιπέδου ορίζεται σε 6 (*THE SMOG READABILITY FORMULA, a Simple Measure of Gobbledygook*, n.d.).

Ο δείκτης αυτοματοποιημένης αναγνωσιμότητας ARI (Automated Readability Index) (Abonizio et al., 2020; Bonet-Jover et al., 2021; Ibrishimova & Li, 2020; Meel & Vishwakarma, 2020; Sotirakou et al., 2021) είναι ένα τεστ αναγνωσιμότητας σχεδιασμένο για την αξιολόγηση της κατανόησης του κειμένου. Υπολογίζεται από τον αριθμό γραμμάτων ανά λέξη και τον αριθμό λέξεων ανά πρόταση. Το ευρετήριο ARI προσδιορίζει έναν αριθμό που προσεγγίζει την ηλικία που απαιτείται για την κατανόηση του κειμένου. Για παράδειγμα το επίπεδο 1 αντιστοιχεί σε ηλικίες 6 έως 8 ετών. Το επίπεδο ανάγνωσης 8 αντιστοιχεί στο τυπικό επίπεδο ανάγνωσης ενός παιδιού 14 ετών στις ΗΠΑ (*THE AUTOMATED READABILITY INDEX (ARI) - LEARN HOW TO CALCULATE THE AUTOMATED READABILITY INDEX*, n.d.).

Το ευρετήριο Coleman Liau (Papadopoulou et al., 2019) χρησιμοποιείται για την αναγνωσιμότητα των κειμένων και βασίζεται σε αντίθεση με τον δείκτη ARI σε χαρακτήρες αντί για συλλαβές ανά λέξη. Υπολογίζεται από τον μέσο αριθμό γραμμάτων ανά λέξη (100 λέξεις) και τον μέσο αριθμό προτάσεων ανά λέξη (100 λέξεις). Η μετρική χρησιμοποιήθηκε για αυτοματοποιημένο υπολογισμό των κειμένων γιατί οι υπολογιστές δυσκολεύονταν να υπολογίσουν λέξεις και προτάσεις (*THE COLEMAN-LIAU READABILITY FORMULA (Also Known as The Coleman-Liau Index)*, n.d.).

Στην παρούσα βιβλιογραφία οι μετρικές αναγνωσιμότητας έχουν υιοθετηθεί για την κατανόηση των κειμένων των ψευδών ειδήσεων (Abonizio et al., 2020; Barrón-Cedeño et al., 2019; Bonet-Jover et al., 2021; B. D. Horne et al., 2020; Ibrishimova & Li, 2020; Kausar et al., 2020; Meel & Vishwakarma, 2020; Papadopoulou et al., 2019; Reis et al., 2019; Sotirakou et al., 2021).

4.3 Στυλομετρικά Χαρακτηριστικά

Η συγκεκριμένη οπτική επικεντρώνεται στην αξιολόγηση της γραφής του συγγραφέα εστιάζοντας σε εκείνους που έχουν πρόθεση να αποπροσανατολίζουν το κοινό. Οι εκδότες τείνουν να δημοσιεύουν άρθρα τα οποία επηρεάζουν το κοινό ενώ ορισμένες φορές αλλοιώνουν ή οδηγούν σε παραποιημένη πληροφόρηση. Οι ψευδείς ειδήσεις έχουν είτε εντυπωσιακούς τίτλους (π.χ. τίτλους με έντονη συναισθηματική φόρτιση) είτε χρησιμοποιούν κεφαλαία γράμματα και λιγότερες κοινές λέξεις (π.χ. και, το κ.λ.π) (Abonizio et al., 2020; B. D. Horne et al., 2020; Sheikhi, 2021; Sotirakou et al., 2021) (Εικόνα 38).

BREAKING BOMBHELL: NYPD Blows Whistle on New Hillary Emails: Money Laundering, Sex Crimes with Children, Child Exploitation, Pay to Play, Perjury

Posted on November 2, 2016 by admin



Εικόνα 38. Κεφαλαία γράμματα σε τίτλο.

Τα στυλομετρικά χαρακτηριστικά (χαρακτηριστικά γραφής) υιοθετούν πιο εξελιγμένες τεχνικές NLP και χωρίζονται σε: *συντακτικά χαρακτηριστικά* (οπτική αναγνώριση χαρακτήρων-OCR, ακολουθίες λέξεων-unigrams, bigrams, POS tagging-Spacy, νεολογισμούς-Out Of Vocabulary), *χαρακτηριστικά λεξικών* (LIWC, Bias lexicon, MPQA Subjectivity lexicon) για την καταμέτρηση χαρακτηριστικών υποκειμενικότητας (δηλώσεων υπεκφυγής, κατηγορημάτων) και *μεροληψίας*, *σημασιολογικά χαρακτηριστικά* (βαθμός τοξικότητας κειμένου, αναγνώριση ονομαστικών οντοτήτων-NER, αποδόμηση κειμένου, βαθμοί σύγκρισης προφίλ) αλλά και *χαρακτηριστικά αναπαραστάσεων (μη παρατηρήσιμα)* μέσω εξειδικευμένων τεχνικών όπως: Bag of Words, TF-IDF, Doc2Vec, Word2Vec, FastText αλλά και αλγοριθμικών μοντέλων (BERT) (Ahmad et al., 2020; Barrón-Cedeño et al., 2019; Borges et al., 2019; Du et al., 2021; Elhadad et al., 2020; Goldani et al., 2021; Gravanis et al., 2019; B. D. Horne et al., 2020; Huang & Chen, 2020; Kasseropoulos & Tjortjjs, 2021; Khamday, 2021; Y. Liu et al., 2019; Meel & Vishwakarma, 2020; Reis et al., 2019; Saikh et al., 2019; Shu, Wang, et

al., 2019; Shu, Cui, et al., 2019; Sotirakou et al., 2021; Tundis et al., 2021; Varshney & Vishwakarma, 2021).

Επιπλέον, βοηθούν στο να διαχωριστούν οι νόμιμοι χρήστες από τους ψεύτικους (anomalous accounts) αλλά και στον εντοπισμό ενός παραπλανητικού άρθρου (Hangloo & Arora, 2021) (Εικόνα 39).



Εικόνα 39. Στιγμιότυπο ελληνικού παραπλανητικού άρθρου.

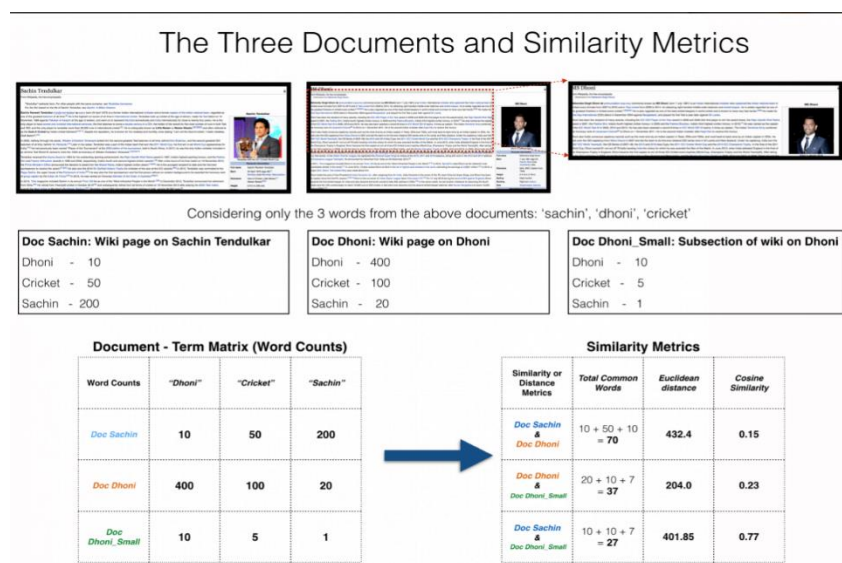
Σε *συντακτικό επίπεδο* εκτός από την εμφάνιση κεφαλαίων γραμμάτων στον τίτλο διερευνώνται και *συχνότητες λειτουργικών λέξεων και φράσεων* (όπως ακολουθίες λέξεων από n αντικείμενα- n grams (Y. Liu et al., 2019), ή *συχνότητες δύο λέξεων, τριών, τεσσάρων κ.λ.π.* (bigrams, trigrams, fourgrams, fivegrams) (Ahmad et al., 2020; Barrón-Cedeño et al., 2019; Elhadad et al., 2020; Huang & Chen, 2020; Kasseropoulos & Tjortjis, 2021; Khanday et al., 2021; Meel & Vishwakarma, 2020; Shu, Wang, et al., 2019), *συχνότητες στίξης* (όπως τελείες, ερωτηματικά, αποσιωπητικά) (Ahmad et al., 2020; Kasseropoulos & Tjortjis, 2021; Y. Liu et al., 2019; Meel & Vishwakarma, 2020), αλλά και η *παρουσία έλλειψης δηλαδή σειρές από τελείες που δείχνουν την εσκεμμένη παράλειψη λέξης, ειδικά σύμβολα* (όπως - ή , .) (Sheikhi, 2021), *μετρικές όπως η μέτρηση ομοιότητας δύο κειμένων (cosine similarity metric)* (Du et al., 2021; Saikh et al., 2019; Varshney & Vishwakarma, 2021), *τυπογραφικά λάθη* (Gravanis et al., 2019), *αλγόριθμους για οπτική αναγνώριση χαρακτήρων* όπως ο αλγόριθμος αναγνώρισης κειμένου πολλών χαρακτήρων και συμβόλων (mixed code text normalization) με χρήση κανόνων (Tundis et al., 2021), *η επισήμανση συγκεκριμένων λέξεων μέσω POS tagging* (ρημάτων, ουσιαστικών, επιθέτων) από ειδικά λογισμικά για τη Φυσική Επεξεργασία Γλώσσας (NLP) όπως NELA, Spacy Package αλλά και το Stanford NLP Parser tool (Abonizio et

al., 2020; Barrón-Cedeño et al., 2019; Bonet-Jover et al., 2021; Elhadad et al., 2020; Huang & Chen, 2020; Posadas-Durán et al., 2019; Reis et al., 2019; Saikh et al., 2019; Shu, Cui, et al., 2019; Silva et al., 2020; Sotirakou et al., 2021).

Οι συχνότητες λειτουργικών λέξεων και φράσεων αφορούν το σύνολο συχνά χρησιμοποιημένων λέξεων από το περιεχόμενο του άρθρου, το προφίλ δημιουργού, τις περιγραφές θεμάτων (J. Zhang et al., 2019) δηλαδή άρθρα, προθέσεις, σύνδεσμοι αλλά και συχνότητες λέξεων μίας λέξης (unigrams), δύο λέξεων (bigrams), τριών (trigrams) κ.λ.π. Οι Barrón-Cedeño, Elhadad και Meel και Vishwakarma επισημαίνουν τη συχνότητα μίας λέξης, δύο λέξεων και τριών στα κείμενα των ψευδών ειδήσεων. Οι συγκεκριμένες συχνότητες αποτελούν χαρακτηριστικά τα οποία έχουν βγει από αλγόριθμους όπως ο TF-IDF αλλά και το μοντέλο Bag of Words τα οποία θα αναλυθούν στις επόμενες ενότητες (Barrón-Cedeño et al., 2019; Elhadad et al., 2020; Meel & Vishwakarma, 2020).

Η χρήση της στίξης βοηθά στη διαφοροποίηση του αληθούς κειμένου από το ψευδές (Meel & Vishwakarma, 2020; Pérez-Rosas et al., 2017). Το συγκεκριμένο χαρακτηριστικό υιοθετήθηκε και από τον Horne (B. D. Horne et al., 2020). Η Pérez-Rosas συμπεριέλαβε τα κόμματα, τα θαυμαστικά, τις τελείες και τα ερωτηματικά στη μελέτη της, ενώ η έλλειψη η οποία αφορά σειρές από τελείες που δείχνουν την εσκεμμένη παράλειψη λέξης σε ένα κείμενο, υιοθετήθηκε ως χαρακτηριστικό από τον Sheikhi (Sheikhi, 2021).

Η μετρική *cosine similarity* (Prabhakaran, 2018) (Εικόνα 40) μετρά την ομοιότητα μεταξύ δύο κειμένων (ή οντοτήτων) ανεξαρτήτως μεγέθους και έχει υιοθετηθεί από τους Varshney και Vishwakarma και Saikh (Borges et al., 2019; Meel & Vishwakarma, 2020; Saikh et al., 2019; Varshney & Vishwakarma, 2021).



Εικόνα 40. Μετρική Cosine Similarity.

Χρησιμοποιείται για αναπαράσταση των διανυσμάτων του σώματος και του τίτλου των κειμένων (Borges et al., 2019; Saikh et al., 2019), με χρήση αλγόριθμων βαθιάς μάθησης όπως του Universal Sentence Encoder (USE) (Saikh et al., 2019) ή απλών αλγορίθμων όπως ο TF-IDF. Σύμφωνα με τον Thotta η μετρική ομοιότητας μετρά την ομοιότητα μεταξύ δύο διαφορετικών κειμένων (ή οντοτήτων) ανεξαρτήτου μεγέθους και πιο συγκεκριμένα μετρά το συνημίτονο της γωνίας μεταξύ δύο διανυσμάτων τα οποία προβάλλονται σε ένα πολυδιάστατο χώρο (άξονας χ και ψ). Ο Thotta στη μελέτη του μέτρησε την ομοιότητα ανάμεσα στο κείμενο και στον τίτλο. Η συγκεκριμένη μετρική αποτελεί χαρακτηριστικό εισόδου σε Νευρωνικά Δίκτυα (π.χ. DNN) (Meel & Vishwakarma, 2020; Thota et al., 2018). Οι Varshney και Vishwakarma μέτρησαν την ομοιότητα του τίτλου και του ερωτήματος, τον βαθμό ομοιότητας ανάμεσα σε τίτλο και κείμενο κάθε παραγράφου, αλλά και τον μέσο όρο για 8 παραγράφους παίρνοντας μία συγκεκριμένη τιμή (π.χ. >0.45) προκειμένου να κρατήσουν τα URL που είναι αξιόπιστα (Varshney & Vishwakarma, 2021).

Ο Borges υπολόγισε και τη μετρική ομοιότητας *soft cosine* ανάμεσα σε αναπαραστάσεις κειμένου υπολογίζοντας τη συχνότητα του τίτλου και του σώματος ή του τίτλου και των δύο πρώτων προτάσεων του κειμένου. Η απλή μετρική *soft cosine similarity* σε αντίθεση με την μετρική ομοιότητας παίρνει ως δεδομένο ότι τα χαρακτηριστικά είναι όμοια ανάμεσα σε ένα σύνολο χαρακτηριστικών. Η Charlet και Damnati επισημαίνουν ότι ακόμα και αν δεν υπάρχουν κοινές λέξεις στα κείμενα αυτά μπορεί να είναι σημασιολογικά όμοια εφόσον οι λέξεις σχετίζονται σημασιολογικά (Borges et al., 2019; Charlet & Damnati, 2017; *Gensim*, n.d.).

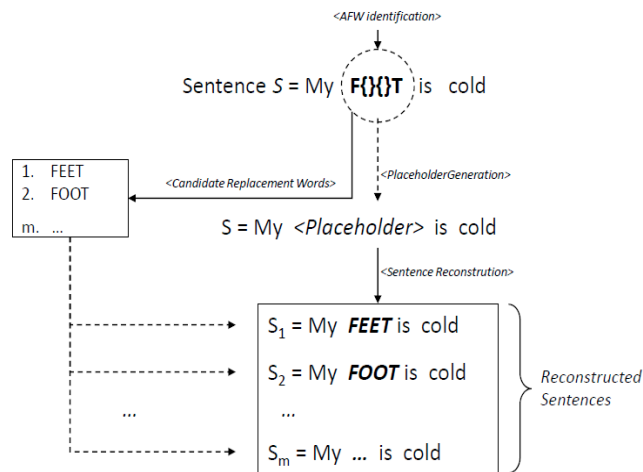
Ο Borges χρησιμοποίησε και τη μετρική ομοιότητας *περιλήψεων* η οποία προσδιορίζει την ποιότητα μίας περιλήψης που δημιουργήθηκε από υπολογιστή, συγκρίνοντας τη με περιλήψεις που δημιουργήθηκαν από ανθρώπους (Rouge Score-Recall-Oriented Understudy), με υπολογισμό των συχνοτήτων λέξεων, ngrams, ζεύγη λέξεων (π.χ. μετρήσεις ομοιότητας ανάμεσα σε επικεφαλίδα και σύνολο προτάσεων από το σώμα κειμένου και της επικεφαλίδας με τις δύο πρώτες προτάσεις του σώματος) αλλά και την *αυτόματη αξιολόγηση μηχανικής μετάφρασης* (*Bleu Score-Bilingual Evaluation Understudy*), όπου συγκρίνεται η αυτόματη μετάφραση ενός κειμένου από υπολογιστή σε σχέση με την μετάφραση που δημιουργείται από ανθρώπους. Ο υπολογισμός της μετρικής *Bleu* γίνεται μέσω της μέτρησης του αριθμού των συχνοτήτων μιας λέξης (unigrams), μίας επιλεγμένης μετάφρασης, οι οποίες βρίσκονται σε μεταφράσεις αναφοράς (μεταφράσεις ανθρώπων) και τον διαχωρισμό της με βάση το σύνολο των λέξεων στην επιλεγμένη μετάφραση (π.χ. μετρήσεις ομοιότητας ανάμεσα σε επικεφαλίδα και σύνολο προτάσεων από το σώμα

κειμένου και της επικεφαλίδας με τις δύο πρώτες προτάσεις του σώματος) (Borges et al., 2019; C.-Y. Lin, 2004; Papineni et al., 2001).

Ο αλγόριθμος *mixed code text normalization* (Εικόνα 41) χρησιμοποιήθηκε από τον Tundis για την αναγνώριση προπαγανδιστικού κειμένου πολλών χαρακτήρων και συμβόλων (πολυκωδικοποιημένου) μέσω κανόνων και βασίζεται στην Οπτική αναγνώριση χαρακτήρων (Optical Character Recognition, OCR). Εκτός από την αναγνώριση κάνει εντοπισμό και και ανάλυση χαρακτήρων-συμβόλων από κείμενο με μεικτή κωδικοποίηση.

Ανήκει στην κατηγορία «ειδικών χαρακτήρων για αλφάβητα» όπου οι χαρακτήρες εντοπίζονται χρησιμοποιώντας μοναδικούς ειδικούς χαρακτήρες ή πολλαπλούς ειδικούς χαρακτήρες. Η αντικατάσταση γίνεται με βάση τον κώδικα που χρησιμοποιείται (προκαθορισμένοι κανόνες). Περιλαμβάνει τα εξής στάδια: α) τμηματοποίηση κειμένου (συντακτική λειτουργία) όπου γίνεται διαχωρισμός της λέξης από διαφορετικά υποσύνολα συμβόλων, β) μεταμόρφωση χαρακτήρων η οποία στοχεύει στην άντληση γραμμάτων της αλφάβητου από κάθε υποσύνολο συμβόλων (κανόνας 1= αλφαβητικοί χαρακτήρες π.χ. α, β..., κανόνας 2=μη αλφαβητικοί π.χ. !,?, @), γ) επιλογή λέξης όπου δημιουργούνται και επιλέγονται υπάρχουσες λέξεις σύμφωνα με το λεξιλόγιο (Αγγλικό) στη βάση των παραγόμενων γραμμάτων (επιλογή λέξεων προερχόμενων από τα αγγλικά) και δ) αναδόμηση πρότασης που στοχεύει στην αντικατάσταση του αρχικού κειμένου μεικτής κωδικοποίησης με τις επιλεγμένες λέξεις, προκειμένου να επιτευχθεί μία κατανοητή πρόταση κειμένου.

Ο συγκεκριμένος αλγόριθμος παράγει εν δυνάμει προτάσεις σε φυσική γλώσσα βασιζόμενος σε «ειδικούς χαρακτήρες για αλφάβητα» και έχει χρησιμοποιηθεί σε αναγνώριση της ρητορικής μίσους (Santosh & Aravind, 2019; Tundis et al., 2021). Οι μετέπειτα αναδομημένες προτάσεις θα χρησιμοποιηθούν σε αλγόριθμους μηχανικής μάθησης για τον εντοπισμό και τη διάκριση των προπαγανδιστικών άρθρων.



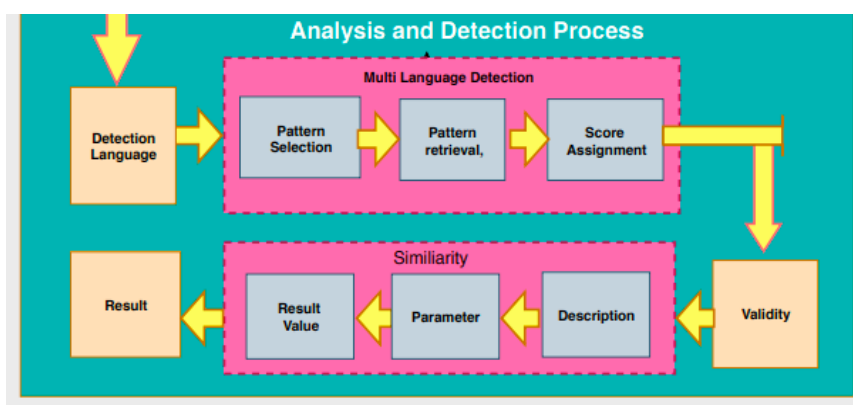
Εικόνα 41. Παράδειγμα αναγνώρισης πρότασης από αλγόριθμο.

Ο Yuliani αναφέρει τέσσερα κριτήρια για τον προσδιορισμό των hoaxes: την αλυσιδωτή γραμματοσειρά (π.χ. «Διάδωσε αυτό σε όποιον ξέρεις. Αλλιώς κάτι δυσάρεστο θα συμβεί»), τη μη συμπίληψη της ημέρας του γεγονότος ή την πραγματική ώρα ή τον μη προσδιορισμό τους (π.χ. δηλώσεις όπως «χθες» ή «δημοσιεύτηκε από» δεν είναι ξεκάθαρες), τη μη ύπαρξη ημερομηνίας λήξης ως προς την πληροφόρηση (με τη έννοια ότι παρόλο που η ημερομηνία υπάρχει και δεν αποδεικνύει τίποτα, θα έχει μία προ-αναμενόμενη επιρροή) και την μη ύπαρξη παραπομπής σε συγκεκριμένη οργάνωση ως πηγή πληροφόρησης ή την περίληψη της οργάνωσης αλλά χωρίς διασυνδεδεμένα δεδομένα.

Πιο συγκεκριμένα, για τον εντοπισμό των πολυγλωσσικών hoaxes (Μαλαισιανά, Ινδονησιακά, Αγγλικά) ανέκτησε αρχικά γλωσσικά μοτίβα με τη χρήση εργαλείων όπως το Google Translate API και την τεχνική τμηματοποίησης δηλαδή την εξαγωγή σημαντικών λέξεων από τη λίστα μορφοτύπου XML. Σύμφωνα με τον ίδιο, σε γλώσσες όπως τα κινέζικα, βιετναμέζικα, ινδικά υπάρχει γλωσσική ιδιομορφία και δεν μπορούν εύκολα να ταυτοποιηθούν (π.χ. λέξεις όπως *Tiêhḡ*, *hanzi*). Μετέπειτα μέσω καλών μοτίβων για θεματικούς κανόνες αλλά και για το σώμα κειμένου (χρήση API) έγινε εισαγωγή τυχαίας επιλογής κειμένων (εισαγωγή σε API) και επεστράφησαν τα αποτελέσματα με τον κώδικα της γλώσσας και τον βαθμό (Tuan Vu et al., 2015; Yuliani et al., 2019).

Στην ανάθεση της βαθμολογίας οι κανόνες δημιουργούνται με βάση τα μοτίβα των spam (κανόνες Spam assassin). Υπάρχουν δύο κανόνες στον spam Assassin: ο κανόνας Θέματος και ο κανόνας Σώματος. Ο κάθε κανόνας προσθέτει ή αφαιρεί βαθμούς από τη βαθμολογία των spam. Αν το μήνυμα έχει υψηλό βαθμό τότε τακτοποιείται ως spam (Tran et al., 2006; Tuan Vu et al., 2015; Yuliani et al., 2019). Για τον εντοπισμό της αξιοπιστίας των hoax ακολουθήθηκαν τα παρακάτω βήματα: έγινε σύγκριση των λέξεων στη βάση

δεδομένων hoax, αφαιρέθηκαν οι κοινές λέξεις (π.χ. το, στο, επίσης), πραγματοποιήθηκε έλεγχος της γλώσσας (μέσω του API), λημματοποίηση (επιστροφή λέξης στην αρχική της μορφή), εύρεση λέξεων στη βάση δεδομένων hoax και στο τελικό στάδιο ο Yuliani χρησιμοποίησε αλγόριθμους ομοιότητας συμβολοσειρών για την αξιολόγησή τους (Similarity algorithm, Levenshtein algorithm, Damerau-Levenshtein). Στα αποτελέσματα, παρουσιάζεται η ομοιότητα των λέξεων σε ποσοστό (100%) αλλά και ο χρόνος που χρειάζεται για την επεξεργασία (λεπτά) (Yuliani et al., 2019) (Εικόνα 42).

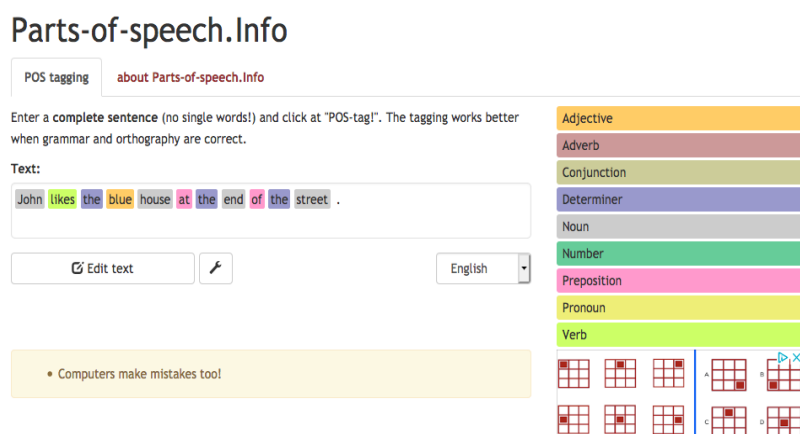


Εικόνα 42. Διαδικασία εντοπισμού hoax.

Το *PoS tagging* έχει χρησιμοποιηθεί για τον εντοπισμό των ψευδών ειδήσεων για να μελετηθεί η διαφορετικότητα εξειδικευμένων επισημάνσεων (π.χ. ρήματα, επίθετα, προσθέσεις¹¹) (Abonizio et al., 2020; Barrón-Cedeño et al., 2019; Bonet-Jover et al., 2021; Braşoveanu & Andonie, 2021; Elhadad et al., 2020; Gravanis et al., 2019; B. D. Horne et al., 2020; Huang & Chen, 2020; Khanday et al., 2021; Meel & Vishwakarma, 2020; Posadas-Durán et al., 2019; Reis et al., 2019; Saikh et al., 2019; Shu, Wang, et al., 2019; Silva et al., 2020; Sotirakou et al., 2021, 2021; Souza Freire et al., 2021) (Εικόνα 43). Το POS tagging αναφέρεται ως επισήμανση του είδους της λέξης μέσω λογισμικού το οποίο διαβάζει το κείμενο και επισημαίνει μέρη του λόγου σε κάθε λέξη (όπως ουσιαστικά, δεικτικά επίθετα, προσωπικές αντωνυμίες) και άλλες λεκτικές μονάδες (όπως επισημάνσεις POS ngrams δηλαδή Unigrams, bigrams, ngrams, trigrams, fourgrams, fivegrams) (Meel & Vishwakarma, 2020; Posadas-Durán et al., 2019). Παραδείγματα λογισμικού που χρησιμοποιούνται είναι το Spacy package και το Stanford Tagger (*SpaCy · Industrial-Strength Natural Language Processing in Python*, n.d.; *The Stanford Natural Language Processing Group*, n.d.). Η αναλογία εξειδικευμένων

¹¹ Πρόσθεση (πρόθεση, μετάθεση)= άκλιτα μέρη του λόγου τα οποία συμπεριλαμβάνουν και χωρο-χρονικές έννοιες ή σημασιολογικές π.χ. στο, κάτω από, προς το, πριν (χωρο-χρονικές), για (σημασιολογικές).

επισημάνσεων προτάθηκε από τον Abonizio για την εξαγωγή συγκεκριμένων επισημάνσεων λέξεων (π.χ. ρημάτων). Επιπλέον, οι υπερθετικοί βαθμοί των λέξεων και των επιρρημάτων τρόπου αλλά και των ουσιαστικών, προσωπικών αντωνυμιών, δεικτικών επιθέτων, επιρρημάτων αποτελούν βασικά χαρακτηριστικά των ψευδών ειδήσεων (Abonizio et al., 2020; Sotirakou et al., 2021).



Εικόνα 43. Pos Tagging.

Στα κοινωνικά μέσα δικτύωσης οι δημοσιεύσεις περιέχουν και νεολογισμούς, αργκό, λανθασμένα ειπωμένες λέξεις (*Out of vocabulary words, OOV*) (Abonizio et al., 2020; Borges et al., 2019; Du et al., 2021) όπου με τη χρήση αλγορίθμων βαθιάς μάθησης μπορούν να εξαχθούν σημασιολογικές αναπαραστάσεις (Du et al., 2021). Οι λέξεις εκτός λεξιλογίου αφορούν ουσιαστικά ασυνήθιστες λέξεις (νεολογισμούς, αργκό) και χρησιμοποιούν ένα λεξικό λέξεων για μια δεδομένη γλώσσα προκειμένου να υπολογίσουν το σύνολο των λέξεων που δεν υπάρχουν σε αυτό και των συχνοτήτων τους στο κείμενο (Abonizio et al., 2020). Για παράδειγμα μόνο για λέξεις με επισήμανση ως επίθετο, επίρρημα, ουσιαστικό. Σύμφωνα με τους Horne και Adali στα clickbait γίνεται χρήση περισσότερων λέξεων αργκό (π.χ. Lol, brb) (B. D. Horne & Adali, 2017).

Επιπλέον κάποιοι συγγραφείς υπολόγισαν τα εισαγωγικά (Abonizio et al., 2020; B. D. Horne et al., 2020), την αναλογία τους στο μέγεθος του κειμένου (Abonizio et al., 2020) και τη στίξη (B. D. Horne et al., 2020; Kasseropoulos & Tjortjis, 2021; Reis et al., 2019). Το χαρακτηριστικό των εισαγωγικών προτάθηκε από τον Abonizio και έχει χρησιμοποιηθεί και στην μελέτη των Horne και Adali ως χαρακτηριστικό των τίτλων των ψευδών ειδήσεων (B. D. Horne & Adali, 2017). Ενώ για να βρεθούν τα χαρακτηριστικά των clickbait οι Liu και Saikh υπολόγισαν το ποσοστό των αριθμών στους τίτλους (π.χ. «Eating strawberries helps you loose weight!19 facts about fruit that will astonish you») (Y. Liu et al., 2019; Saikh et al., 2019), τη χρήση προστακτικών λέξεων (π.χ. πρέπει, ποτέ), την ύπαρξη λέξεων ή φράσεων σε

αναδημοσιευμένο κείμενο *clickbait* (π.χ. «κάντε κλικ για να το αγοράσετε», «αναδημοσίευσε το (αυτό στην οικογένειά σου)», *την ύπαρξη εξωτερικών συνδέσμων* (π.χ. "http://") σε αναδημοσιευμένο κείμενο *clickbait*. Ενώ, για την αξιοπιστία των γεγονότων, ο Liu υπολόγισε τις *παρενθέσεις* (π.χ. []) επισημαίνοντας ότι τα αξιόπιστα άρθρα αναφέρουν την βιβλιογραφία (υιοθέτηση επαγγελματικού στυλ) αλλά και τις *συχνότητες λέξεων* (π.χ. «ιατρικοί ερευνητές», «αναφορά», «μελέτη») (Y. Liu et al., 2019). Για τις συχνότητες λέξεων υιοθετήθηκε συγκεκριμένο λεξιλόγιο από το λεξικό Wiktionary για τον υπολογισμό της προπαγάνδας (π.χ. βοηθητικά επιρρήματα), το οποίο αναλύεται στην επόμενη υποενότητα μαζί με λεξικά μεροληψίας και υποκειμενικότητας.

4.3.1 Χαρακτηριστικά Λεξικών (Μεροληψία και Υποκειμενικότητα)

Η χρήση ανοιχτών λεξικών (Wiktionary) ή λογισμικών με χαρακτηριστικά λεξικού (LIWC) βοηθά στον εντοπισμό συγκεκριμένων λέξεων που χρησιμοποιούνται στις ψευδείς ειδήσεις (π.χ. μεροληπτικών λέξεων-bias detection, λέξεων υποκειμενικότητας-subjectivity).

Το Wiktionary ("Wiktionary," 2022) είναι ένα πολύγλωσσο ανοιχτό διαδικτυακό λεξικό το οποίο περιλαμβάνει λέξεις ή φράσεις σε όλες τις φυσικές γλώσσες και την τεχνητή νοημοσύνη. Ο Barrón-Cedeño το χρησιμοποίησε για να μετρήσει τη συχνότητα συγκεκριμένων λέξεων στα προπαγανδιστικά κείμενα (π.χ. βοηθητικά ρήματα, επιρρήματα τρόπου, δράσης, σύγκρισης, υπερθετικών βαθμών) (Barrón-Cedeño et al., 2019).

Η LIWC, όπως αναφέρθηκε και σε προηγούμενα κεφάλαια, χρησιμοποιείται ευρέως ως εφαρμογή ανάλυσης κειμένου με σκοπό τη μελέτη γνωστικών, συναισθηματικών, δομικών στοιχείων ενός κειμένου τα οποία εμφανίζονται στην ομιλία και στο γραπτό λόγο ενός ατόμου, αλλά και ως λεξικό ψυχολογικών κατηγοριών. Περιλαμβάνει στατιστικές μετρήσεις, ρητορικές ερωτήσεις, προστακτικές εντολές, προσωπικές αντωνυμίες, συναισθηματικές λέξεις, χαρακτηριστικά υποκειμενικότητας (π.χ. σημάνσεις βεβαιότητας, προσθήκες, συνδέσμους, παραθέσεις, λέξεις άγχους).

Στην παρούσα βιβλιογραφία οι συγγραφείς τη χρησιμοποίησαν για τη μέτρηση συγκεκριμένων χαρακτηριστικών όπως *πρώτου ενικού και δεύτερου ενικού προσώπου, ρημάτων όρασης και ακοής* (π.χ. βλέπω, ακούω), *σεξουαλικών λέξεων* (π.χ. γκέι), *αρνήσεων* (π.χ. δεν μπορώ), *λέξεων με ισχυρή υποκειμενικότητα* (π.χ. αντισημίτης) αλλά και *αδύναμη* (π.χ. εξτρεμιστής) (Barrón-Cedeño et al., 2019), *το ποσοστό των επιθέτων, των ρημάτων, των προθέσεων* (Ahmad et al., 2020) (Ahmad et al., 2020; Barrón-Cedeño et al., 2019; Gôlo et al., 2021; B. D. Horne et al., 2020; Huang & Chen, 2020; Kausar et al., 2020; Khanday et al., 2021; Meel & Vishwakarma, 2020; Reis et al., 2019; Shu, Cui, et al., 2019).

Η υποκειμενικότητα ενός κειμένου αφορά τον τόνο του συγγραφέα καθώς και τις προσωπικές του απόψεις που εκφράζονται στο κείμενο για αυτό και θεωρείται σημαντικό χαρακτηριστικό ενός ποιοτικού κειμένου (Sotirakou et al., 2021). Η υιοθέτηση λεξικών υποκειμενικότητας όπως το *MPQA Subjectivity Lexicon* (*MPQA Resources*, n.d.) (Reis et al., 2019; Sotirakou et al., 2021) αλλά και λεξικών μεροληψίας όπως το *Bias lexicon* (Barrón-Cedeño et al., 2019; B. D. Horne et al., 2020; Kausar et al., 2020) αποτελούν κύρια χαρακτηριστικά για την αναγνώριση της υποκειμενικότητας.

Το *MPQA Subjectivity Lexicon* περιλαμβάνει 8000 στοιχεία υποκειμενικότητας δηλαδή λέξεις και φράσεις που δηλώνουν ιδιωτικές καταστάσεις αλλά και προσωπικές αντιλήψεις, υποκειμενικές και αντικειμενικές χρήσεις (Wilson et al., 2005). Μέσω του ταξινομικού αλγόριθμου υποκειμενικότητας διαπιστώνεται ο βαθμός αδύναμης¹² ή δυνατής υποκειμενικότητας¹³ σε ένα κείμενο (π.χ. αισθάνομαι, κοιτώ, σκέφτομαι, βαθιά, εξ' ολοκλήρου, πρακτικά) με επισήμανση των προτάσεων ως υποκειμενικές ή αντικειμενικές (χρήση *MPQA corpus*). Τα σκανδαλοθηρικά άρθρα (*clickbait*) για παράδειγμα υιοθετούν περισσότερα υποκειμενικά στοιχεία από ότι η προπαγάνδα και τα *hoaxes* (Volkova et al., 2017).

Το *Bias lexicon* περιλαμβάνει λέξεις μεροληψίας βασισμένες στην πολιτική της Wikipedia NPOV (Neutral Point of View)¹⁴ ("Category," 2022) αλλά και ρήματα βεβαιότητας, *factives* δηλαδή ρήματα, ουσιαστικά ή επίθετα που προϋποθέτουν την αλήθεια μίας ενσωματωμένης πρότασης (π.χ. μπορεί), *hedges* δηλαδή λέξεις που υποδηλώνουν την μείωση της δέσμευσης (π.χ. ίσως), *predicates* (κατηγορήματα) τα οποία αναφέρονται σε ιδιότητες και λέξεις ή φράσεις που δηλώνουν τις συγκεκριμένες ιδιότητες (λογική), *implicatives* δηλαδή σχέσεις που υπονοούνται από τις επόμενες λέξεις ή φράσεις (π.χ. η λέξη «φόνος» έχει άμεση σχέση με την «εκτέλεση» γιατί δεν μπορεί να υπάρξει φόνος χωρίς εκτέλεση) (Barrón-Cedeño et al., 2019; B. D. Horne et al., 2020; Kausar et al., 2020; Recasens et al., 2013).

Τα *hedges* αποφεύγουν τις τολμηρές προβλέψεις ή δηλώσεις. Οι λέξεις συμπεριλαμβάνονται στη λίστα υπεκφυγών του Hyland (Hyland, 2015). Τα κατηγορήματα αναφέρονται σε ιδιότητες και λέξεις ή φράσεις που δηλώνουν συγκεκριμένες ιδιότητες με βάση την κατηγοριοποίηση Hooper. Στη συγκεκριμένη κατηγοριοποίηση τα κατηγορήματα

¹² αδύναμα υποκείμενα=λέξεις που έχουν συγκεκριμένες υποκείμενες χρήσεις χαρακτηρίζονται ως αδύναμα υποκείμενα

¹³ ισχυρά υποκείμενα =λέξεις που είναι υποκείμενες σε περισσότερα πλαίσια δεξ (Wilson et al., 2005)

¹⁴ neutral point of view=παρουσίαση απόψεων στα άρθρα χωρίς μεροληψία

διακρίνονται σε 4 τάξεις: assertives (=διαβεβαίωση) (π.χ. απαντώ, διαβεβαιώνω, ξεκαθαρίζω), non assertives (π.χ. πιθανώς, ίσως), true factive, (π.χ. μετανιώνω, ξεχνώ, διασκεδάζω), semi factive (π.χ. εξερευνώ, γνωρίζω, μαθαίνω). Το συγκεκριμένο χαρακτηριστικό υιοθετήθηκε από τη μελέτη της Recasens για τον εντοπισμό της μεροληπτικής γλώσσας (Recasens et al., 2013). Τα ρήματα που δηλώνουν διαβεβαίωση (assertives) αναφέρονται ως χαρακτηριστικό στις μελέτες των Horne και Barrón-Cedeño (Barrón-Cedeño et al., 2019; B. D. Horne et al., 2020).

Τα implicatives περιλαμβάνουν ρήματα τα οποία υπονοούν την αλήθεια ή όχι της συμπληρωματικής πρότασης εξαρτώντας τη από το θετικό ή αρνητικό χαρακτήρα του κύριου κατηγορήματος (λέξη ή φράση) (π.χ. Αφού **δολοφόνησε** τρεις αστυνομικούς, οι αποικίες ανακοίνωσαν τον Κέλλυ ως ένα καταζητούμενο φυγόδικο. Αφού αυτός **σκότωσε** τρεις αστυνομικούς, οι αποικίες ανακοίνωσαν τον Κέλλυ ως ένα καταζητούμενο φυγόδικο) (B. D. Horne et al., 2020).

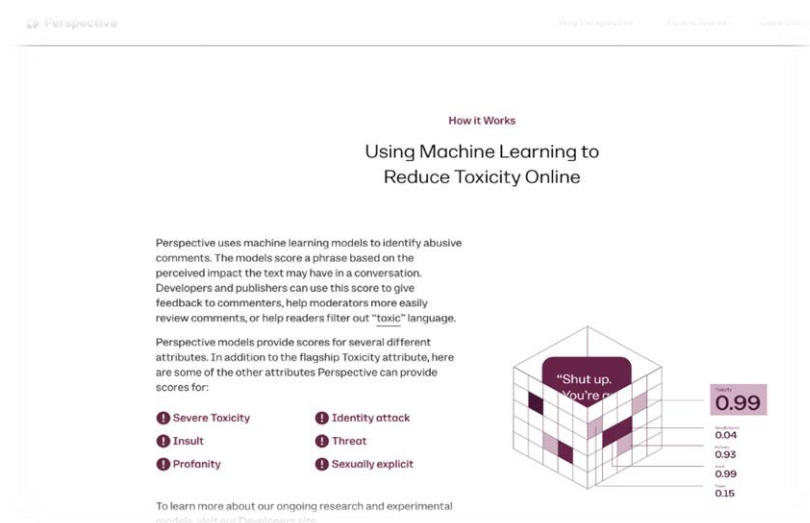
Τα factives είναι ρήματα τα οποία προϋποθέτουν την αλήθεια της συμπληρωματικής πρότασης. Στην πρόταση «Αυτός **διαπίστωσε** ότι η καταπίεση των μαύρων ανθρώπων ήταν περισσότερο ένα αποτέλεσμα οικονομικής εκμετάλλευσης από ότι οτιδήποτε εκ φύσεως ρατσιστικό», το **διαπίστωσε** προϋποθέτει την αλήθεια της καταπίεσης των μαύρων ατόμων. Στην πρόταση «Η θέση του ήταν ότι η καταπίεση των μαύρων ανθρώπων ήταν περισσότερο ένα αποτέλεσμα οικονομικής εκμετάλλευσης από ότι οτιδήποτε εκ φύσεως ρατσιστικό» παρουσιάζει μία πρόταση σαν τη **θέση** κάποιου (B. D. Horne et al., 2020).

Εκτός από τα λεξικά που ειπώθηκαν παραπάνω, συγγραφείς όπως ο Barrón-Cedeño του Propgy αλλά και οι Horne και Kausar υιοθέτησε την *υποκειμενικότητα σε ένα κείμενο*. Ο Barrón-Cedeño υιοθέτησε την κλίμακα του *NELA* όπου το κείμενο βαθμολογείται από το 36 (ελάχιστη μεροληψία) έως το 100 (μέγιστη μεροληψία) (Barrón-Cedeño et al., 2019). Η υποκειμενικότητα σε ένα κείμενο βασίζεται σε επίθετα ή επιρρήματα που ενισχύουν την υποκειμενικότητα σε μία πρόταση ή δήλωση (π.χ. αντισημίτης) (ισχυρά υποκείμενα) αλλά και σε όρους μονόπλευρους (αδύναμα) οι οποίοι ανήκουν σε αντιπαρατιθέμενα θέματα (π.χ. θρησκεία, τρομοκρατία-εξτρεμιστής) (Recasens et al., 2013;). Ενω οι Horne και Kausar υιοθέτησαν λέξεις θετικής και αρνητικής γνώμης (π.χ. διαπρέπω, τρομερό) για τη μέτρηση της μεροληψίας (B. D. Horne et al., 2020; Kausar et al., 2020).

4.3.2 Σημασιολογικά Χαρακτηριστικά

Τα σημασιολογικά χαρακτηριστικά βασίζονται στις σημασιολογικές προσεγγίσεις ενός κειμένου και εξηγούν τη σημαντικότητα ορισμένων χαρακτηριστικών όπως ο *βαθμός*

τοξικότητας ενός κειμένου, η βαθμολογία αξιοπιστίας ενός προφίλ χρήστη, την αποσαφήνιση λέξεων, την αποδόμηση ενός κειμένου σε μεμονωμένες λέξεις ή φράσεις, την ονομαστική αναγνώριση οντοτήτων μέσω ειδικών πακέτων λογισμικού όπως το Polyglot Package (Polyglot-NER) (Abonizio et al., 2020; ΑΙ-Ρφου, n.d.), αλλά και τις διαδικασίες επίλυσης αναφορών (anaphora resolution) (Abonizio et al., 2020; Bonet-Jover et al., 2021; Reis et al., 2019; Shu, Cui, et al., 2019). Ο Reis χρησιμοποιώντας την εφαρμογή Google Perspective API (Perspective API, n.d.) μέτρησε τον βαθμό τοξικότητας των ψευδών ειδήσεων στην πολιτική (Reis et al., 2019) (Εικόνα 44).



Εικόνα 44. Google Perspective API.

Η συγκεκριμένη εφαρμογή χρησιμοποιεί μοντέλα μηχανικής μάθησης για να ποσοτικοποιήσει το πόσο ένα κείμενο (ή σχόλιο) μπορεί να χαρακτηριστεί ως τοξικό. Τα μοντέλα βαθμολογούν μία φράση βασιζόμενα στην απήχηση που έχει το κείμενο σε μία συζήτηση. Κατά αυτόν τον τρόπο είναι πολύ πιο εύκολη η ανατροφοδότηση και η αξιολόγηση των σχολίων.

Εκτός από τη βαθμολόγηση τοξικότητας συμπεριλαμβάνονται και μετρήσεις μέγιστης τοξικότητας, προσβολής, επίθεση ταυτότητας, απειλής, αισχρολογίας, (σεξουαλικής παρενόχλησης-πειραματικό στάδιο). Για παράδειγμα η φράση «είσαι ηλίθιος!» μπορεί να βαθμολογηθεί με 0.8 (0, 1 τιμές) γιατί η πλειονότητα των χρηστών τη θεωρεί τοξική (Perspective | Developers, n.d.). Επιπλέον, ο βαθμός σύγκρισης των προφίλ αλλά και των προσωπικών εμπειριών των χρηστών (πραγματικές επισκέψεις), μειώνουν την επιρροή παραπλανητικών χρηστών οι οποίοι δεν έχουν κάνει ποτέ επίσκεψη (Congroy et al., 2015).

Η αποσαφήνιση λέξης δηλαδή η αναγνώριση της έννοιας μίας διφορούμενης λέξης σε μία πρόταση, λειτουργεί συμπληρωματικά με το PoS Tagging κι αυτό γιατί θεωρείται σημαντικό για την αποδόμηση του κειμένου. Για παράδειγμα η λέξη fine (=πρόστιμο) ως ουσιαστικό (τιμωρία) και fine (=λεπτός) ως επίθετο (καλό). Η αποδόμηση κειμένου σε έννοιες, δηλαδή μεμονωμένες λέξεις ή εκφράσεις πολλών λέξεων είναι βασικό στοιχείο της σημασιολογικής ανάλυσης του κειμένου. Οι μέθοδοι που χρησιμοποιούνται είναι στατιστικές (μέτρηση λέξεων και τοποθεσία λέξης) και κανόνες γλωσσολογικοί (POS tagging, κατάτμηση κειμένου) (Cambria et al., 2017).

Η αναγνώριση των οντοτήτων (NER) (Εικόνα 45) είναι ένας άλλος τρόπος εξαγωγής πληροφοριών, ο οποίος βασίζεται στον εντοπισμό και στην ταξινόμηση ονομαστικών οντοτήτων σε προ-υπάρχουσες κατηγορίες που αναφέρονται σε ένα μη δομημένο κείμενο (Abonizio et al., 2020; Bonet-Jover et al., 2021; Braşoveanu & Andonie, 2021; Cambria et al., 2017; Reis et al., 2019; Saikh et al., 2019). Οι κατηγορίες αφορούν: ονόματα προσώπων, οργανώσεις, τοποθεσίες (Al-Rfou et al., 2014). Το συγκεκριμένο χαρακτηριστικό υιοθετήθηκε από τον Abonizio υπολογίζοντας τις ονομαστικές οντότητες που έχουν αναγνωριστεί και την αναλογία τους συνολικά στο κείμενο (Abonizio et al., 2020). Τα λογισμικά που χρησιμοποιήθηκαν στη συγκεκριμένη βιβλιογραφία είναι τα: Polyglot-NER (Al-Rfou, n.d.), Spacy, Stanford NER tagger (*The Stanford Natural Language Processing Group*, n.d.) (Εικόνα 46).

Stony Brook Entity Extractor

Please enter the text that you need to be processed:

Stony Brook University is located in Long Island, NY. It is administered by John Smith.

Language: English Extra Recall: 0.00 Debug Tokenize Extract

Results

Legend

- Person
- Organization
- Location

Text

Εικόνα 45. Στιγμιότυπο από αναγνώριση ονομαστικών οντοτήτων.

Stony Brook University is located in Long Island , NY .
It is administered by John Smith .

Εικόνα 46. Στιγμιότυπο εξαγωγής ονομαστικών οντοτήτων (NER).

Για τον Bonet-Jover η χρήση σημασιολογικών χαρακτηριστικών όπως οι *ονομαστικές οντότητες* βελτιώνει τον εντοπισμό απαντήσεων σε ερωτήσεις όπως «Πού; Ποιός; Πότε;» (5W1H-Who, Where, When, Why, How) (π.χ. Where was Pythagoras born? Samos Pythagoras was born [AM_LOC on the island of Samos]) (Bonet-Jover et al., 2021).

Οι Braşoveanu & Antonie εξήγαγαν *χαρακτηριστικά οντοτήτων* κυρίως *τριπλές σχέσεων αναφορών οντοτήτων από τη γνωσιακή βάση* της DBpedia με συνδέσμους που προσδιορίζουν την ιδιότητα των οντοτήτων (π.χ. εάν ο Τραμπ είχε αναφέρει τον Ομπάμα σε ένα κείμενο-dbr: Donald_Trump mentions dbr:Barack_Obama- τότε dbo:orderInOffice or dbo:President) (Braşoveanu & Andonie, 2021).

Οι *διαδικασίες επίλυσης αναφορών* ανατρέχουν σε ένα προηγούμενο αντικείμενο (point back reference). Το συγκεκριμένο αντικείμενο προσδιορίζεται ως «αναφορά», ενώ η οντότητα αναφοράς «προηγούμενο». Είδη ανάλογων αναφορών είναι οι αντωνυμικές (αναφορικές αντωνυμίες), αναφορές επιθέτων, και αναφορικές εκφράσεις (φράση με μοναδικό ουσιαστικό) (Cambria et al., 2017). Για παράδειγμα, στην πρόταση «John traveled around France twice. They were both wonderful.» με το τρίτο πληθυντικό πρόσωπο (they) δεν μπορεί να καθοριστεί αν αναφέρεται στο πρόσωπο John ή στη χώρα (Γαλλία). Για την επίλυση έχουν υιοθετηθεί συντακτικοί, γραμματικοί κανόνες και πιο πρόσφατα έρευνες έχουν υιοθετήσει γενετικούς αλγόριθμους (Cambria et al., 2017; Reis et al., 2019).

Συμπληρωματικά, ως προς το ζήτημα επίλυσης αναφορών σύμφωνα με τον Cambria θα πρέπει να λαμβάνονται υπόψιν ορισμένοι περιορισμοί όπως αναφορές ενικού-πληθυντικού αριθμού, συμφωνία φύλου (αρσενικό-θηλυκό), αλλά και σημασιολογική συμβατότητα (η προηγούμενη πρόταση και εκείνη της αναφοράς να έχουν την ίδια σημασιολογική ερμηνεία) (Cambria et al., 2017).

Εκτός από τις ονομαστικές οντότητες οι σημασιολογικοί ρόλοι του κειμένου μπορούν να βελτιστοποιήσουν τον εντοπισμό των απαντήσεων στο μοντέλο 5W1H κυρίως όταν οι ερωτήσεις δεν αφορούν ονομαστικές οντότητες. Οι σημασιολογικοί ρόλοι αποτελούν σημαντική πληροφόρηση αφού εξάγονται από κείμενα φυσικής γλώσσας. Για κάθε κατηγορία σε μία πρόταση οι σημασιολογικοί ρόλοι ανακαλύπτουν όλα τα μέρη προσδιορίζοντας τους ρόλους τους (π.χ. ασθενής, μουσικό όργανο) αλλά και τα συμπληρώματα (π.χ. τοπικά, χρονικά, τρόπου). Οι σημασιολογικοί ρόλοι απαντούν στο «Ποιος, Που, Πότε, Γιατί» και βοηθούν στην εξαγωγή απαντήσεων (όπως στο παράδειγμα με το που γεννήθηκε ο Πυθαγόρας) (Bonet-Jover et al., 2021; Moreda et al., 2011).

4.3.3 Χαρακτηριστικά Ιστοτόπου Ειδήσεων (Domain Specific)

Τα συγκεκριμένα χαρακτηριστικά αφορούν την *ύπαρξη εξωτερικών συνδέσμων, αριθμούς παραγράφων, μήκος παραγράφων, ανομοιότητα μεταξύ ιστοτόπων, μετρικές για την αξιοπιστία του ιστοτόπου (bot score), εξαγωγή χαρακτηριστικών όπως η τοποθεσία, η χώρα, το γεωγραφικό μήκος, πλάτος αλλά και ο αριθμός γράφων και το μέσο μήκος τους* (Kasseropoulos & Tjortjis, 2021; Reis et al., 2019; Shu et al., 2017; Shu, Wang, et al., 2019).

Το χαρακτηριστικό των εξωτερικών συνδέσμων αναφέρεται από τους Kasseropoulos και Tjortjis ως χαρακτηριστικό εντοπισμού των ψευδών ειδήσεων και έχει υιοθετηθεί και από τη Shu. Η Shu αναφέρει ότι αυτοματοποιημένες εφαρμογές όπως το Bullshit Detector ερευνούν τους συνδέσμους σε μία ιστοσελίδα για αναφορές σε μη αξιόπιστες πηγές ελέγχοντας τις πηγές με μία διαμορφωμένη λίστα από ιστοτόπους (λίστα μη αξιόπιστων πηγών π.χ. πηγές που είναι γνωστές για τις θεωρίες συνωμοσίας, πηγές που προωθούν τον ρατσισμό, τον μισογυνισμό, την ομοφοβία και άλλα είδη διάκρισης) (Kasseropoulos & Tjortjis, 2021; Shu, Mahudeswaran, et al., 2019).

Το μέσο μήκος γράφων (Kasseropoulos & Tjortjis, 2021; Shu et al., 2017; Shu, Wang, et al., 2019) αναφέρεται στον Potthast και έχει υιοθετηθεί σε προηγούμενες έρευνες όπως του Ciampaglia για την επίλυση του προβλήματος ελέγχου γεγονότων προτείνοντας μία μετρική για την αξιολόγηση μίας δήλωσης μέσω της ανάλυσης του μήκους των μονοπατιών ανάμεσα σε υπό αμφισβήτηση γεγονότα (Ciampaglia et al., 2015; Potthast et al., 2017). Μία δήλωση είναι αληθής αν υπάρχει ως άκρη (κατηγορημα) ενός γράφου ή εάν υπάρχει ένα κοντινό μονοπάτι το οποίο συνδέει το υποκείμενο με το αντικείμενο του σε ένα γράφο. Αν δεν είναι αληθής δεν θα υπάρχουν ούτε άκρες ούτε κοντινά μονοπάτια που θα συνδέουν το υποκείμενο με το αντικείμενο (Ciampaglia et al., 2015).

Η *ανομοιότητα ιστοτόπων* σύμφωνα με τον Reis αποτελεί βασικό χαρακτηριστικό για τον εντοπισμό των ψευδών ειδήσεων κι αυτό γιατί μη αξιόπιστοι ιστότοποι προσπαθούν να μεταμφιεστούν σε αξιόπιστους ιστοτόπους ειδήσεων με τη χρησιμοποίηση αναγνωρίσιμων εκδοτών. Μέσω του Alexa ranking προσδιορίστηκε η ανομοιότητα και ακολούθησε σύγκριση με τον ιστότοπο της βάσης δεδομένων (Reis et al., 2019). Η ανομοιότητα μετρά τον ελάχιστο αριθμό τροποποιήσιμων λειτουργιών δηλαδή εισαγωγή, διαγραφή, αποκατάσταση προκειμένου να τροποποιηθεί μία συμβολοσειρά σε μία άλλη. Το συγκεκριμένο χαρακτηριστικό υιοθετήθηκε ως χαρακτηριστικό αξιοπιστίας και εμπιστοσύνης των ιστοτόπων από τον Reis ο οποίος χρησιμοποίησε μετρικές απασχόλησης χρηστών από τις σελίδες του Facebook, οι οποίες δημοσιεύουν άρθρα ειδήσεων κάνοντας συσχέτιση με τη θέση τους στην κατάταξη της Alexa (Reis et al., 2019). Οι πηγές χαμηλής αξιοπιστίας

δημοσιεύουν 100 άρθρα ανά εβδομάδα, με αποτέλεσμα το περιεχόμενό τους να αποκτά μεγαλύτερη δημοσιότητα (virality) μέσω και των αναδημοσιεύσεων στο twitter (30 tweet ανά άρθρο ανά εβδομάδα) (Shao et al., 2018).

Επιπλέον με τη βαθμολογία bot (*Botometer by OSoMe*, n.d.) μετρήθηκε η χαμηλή αξιοπιστία του ιστοτόπου (Shao et al., 2018). Τα «bot» συντελούν στην αναπαραγωγή χαμηλής αξιοπιστίας περιεχομένου πηγών μέσω δραστηριότητας διαμοιρασμού (share), με συνέπεια την επιρροή των χρηστών αλλά και τη γρήγορη διάδοση των ψευδών ειδήσεων, ακόμα και στα πρώτα στάδια της διάχυσης πριν αποκτήσει έκταση (Reis et al., 2019; Shao et al., 2018).

Για τον εντοπισμό τοποθεσίας του ιστοτόπου (Reis et al., 2019) πραγματοποιήθηκε εξαγωγή της διεύθυνσης IP της τοποθεσίας ιστοτόπου μέσω του εργαλείου trace route (εύρεση διαδρομής). Για την εξαγωγή χαρακτηριστικών τοποθεσίας (εξαγωγή γεωγραφικού πλάτους, εξαγωγή γεωγραφικού μήκους, εξαγωγή πόλης, χώρα κ.λ.π.) χρησιμοποιήθηκε το Ipstack API (*Ipstack - Free IP Geolocation API*, n.d.). Το συγκεκριμένο χαρακτηριστικό υιοθετήθηκε από τον Reis για περαιτέρω επεξεργασία των πληροφοριών του ιστοτόπου αλλά και γιατί πολλές πόλεις έχουν γίνει γνωστές λόγω των κατοίκων, οι οποίοι δημιουργούν και διαδίδουν ψευδείς ειδήσεις (“The City Getting Rich from Fake News,” 2016).

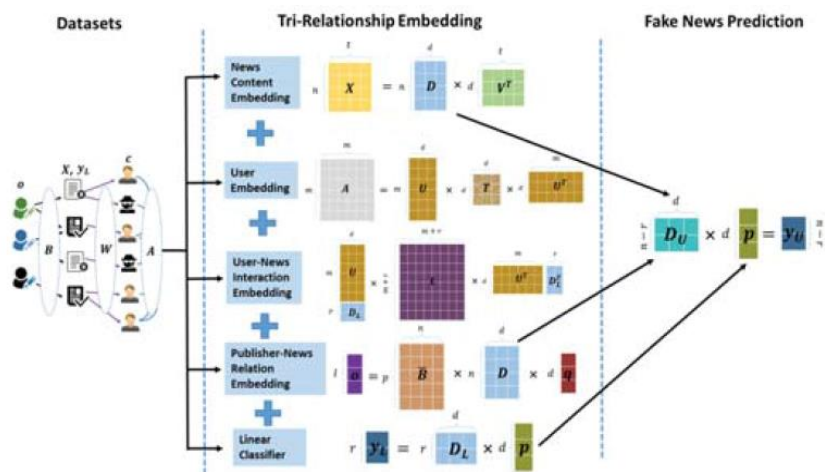
4.3.4 Αναπαράσταση Κειμένου (Μη Παρατηρήσιμα Χαρακτηριστικά)

Με τον όρο *μη παρατηρήσιμα χαρακτηριστικά* εννοούμε όλα εκείνα τα χαρακτηριστικά τα οποία δεν προκύπτουν από επιλογή του ερευνητή αλλά από αναπαραστάσεις κειμένου μοντέλων σε ένα διανυσματικό χώρο (latent space). Ο διανυσματικός χώρος είναι ένας πίνακας παραγοντοποίησης εξαγόμενων χαρακτηριστικών από παρατηρήσιμα χαρακτηριστικά (κείμενο). Αντίστοιχα παραδείγματα θεωρούνται τα μοντέλα ανάλυσης κειμένου τα οποία χρησιμοποιούνται για την ανάλυση των ψευδών ειδήσεων. Οι αναπαραστάσεις κειμένου αποτελούνται από τρεις κατηγορίες: επίπεδο λέξεων, επίπεδο πρότασης και επίπεδο εγγράφου.

Σε επίπεδο λέξεων το μοντέλο *Bag of Words* (“Bag-of-Words Model,” 2022) είναι ένα μοντέλο απλής αναπαράστασης του κειμένου που παρουσιάζει τη συχνότητα των λέξεων σε ένα κείμενο, αγνοώντας τη δομή και τη σειρά των λέξεων μέσα σε αυτό (γι’ αυτό και ο όρος bag=τσάντα) διατηρώντας όμως την πολλαπλότητα. Πιο συγκεκριμένα, το Bag of words αναπαριστά κάθε είδος κειμένου (π.χ. πρόταση, παράγραφο, άρθρο, έγγραφο) ως διάνυσμα όπου κάθε διάσταση αναπαριστά έναν όρο. Περιλαμβάνει ένα λεξιλόγιο γνωστών λέξεων και μία μέτρηση για την παρουσία γνωστών λέξεων (Brownlee, 2017). Ο Al-Ahmad αναφέρει

χαρακτηριστικά ότι στο μοντέλο Bag of Words κάθε είδος κειμένου όπως πρόταση, άρθρο, έγγραφο και παράγραφος χρησιμοποιείται ως ομάδα λέξεων ανεξαρτήτως συντακτικής και σημασιολογικής εξάρτησης (Al-Ahmad et al., 2021).

Αρκετοί συγγραφείς της τρέχουσας βιβλιογραφίας υιοθέτησαν το συγκεκριμένο μοντέλο για την αναπαράσταση της συχνότητας των λέξεων σε ένα κείμενο (Al-Ahmad et al., 2021; Borges et al., 2019; Elhadad et al., 2020; Gôlo et al., 2021; Kaliyar et al., 2020; Khanday et al., 2021; Y. Liu et al., 2019; Meel & Vishwakarma, 2020; Posadas-Durán et al., 2019; Shu, Wang, et al., 2019; Silva et al., 2020; Sotirakou et al., 2021; Vishwakarma et al., 2019; Q. Zhang et al., 2019). Η Shu προσδιόρισε μέσω πολλαπλών πινάκων χαρακτηριστικών Bag of words τις σχέσεις ανάμεσα στους χρήστες του διαδικτύου, τις ειδήσεις και τους εκδότες για τη διάδοση των ψευδών ειδήσεων μέσα σε ένα κοινωνικό πλαίσιο (Shu, Wang, et al., 2019) (Εικόνα 47).



Εικόνα 47. Αναπαράσταση Bag of words (TriFN framework).

Ο Khanday εξήγαγε συχνότητες δύο λέξεων (bigrams) και τριών (trigrams) για περισσότερη πληροφόρηση από το κείμενο συνδυάζοντας χαρακτηριστικά και από το μοντέλο TF-IDF για τον εντοπισμό της προπαγάνδας σε ένα κείμενο (Khanday et al., 2021). Ο Vishwakarma εξήγαγε τις δέκα συχνότερες λέξεις-κλειδιά (π.χ. hoax, hoax fools, false news, fake, fake news, fake death, rumors, false, death hoaxes, falsely) που υπήρχαν στον τίτλο και τις συμπεριέλαβε στον αλγόριθμο για τον εντοπισμό των ψευδών ειδήσεων (Vishwakarma et al., 2019). Ενώ ο Posadas-Durán χρησιμοποίησε το μοντέλο για την επιλογή συγκεκριμένων λέξεων οι οποίες θα βοηθούσαν να εντοπιστούν οι ψευδείς ειδήσεις (Posadas-Durán et al., 2019).

Στο μοντέλο *TF-IDF* οι τιμές των διανυσμάτων προσδιορίζονται από τη συχνότητα της εμφάνισης ενός συγκεκριμένου όρου σε ένα κείμενο (Term frequency) αλλά και στο πόσο κοινή (ή μη κοινή) είναι μια λέξη στο σώμα του κειμένου (Inverse document frequency)(Abdelminaam et al., 2021; Al-Ahmad et al., 2021, p.; Barrón-Cedeño et al., 2019; Bonet-Jover et al., 2021; Borges et al., 2019; Du et al., 2021; Elhadad et al., 2020; Gôlo et al., 2021; Gravanis et al., 2019; Hunt et al., 2022; Isaakidou et al., 2021; Kaliyar et al., 2020; Kar, 2020; Kausar et al., 2020; Khanday et al., 2021; Meel & Vishwakarma, 2020; Saikh et al., 2019; Silva et al., 2020; Taskin et al., 2022; Tundis et al., 2021; Yuliani et al., 2019).

Η συγκεκριμένη τεχνική χρησιμοποιείται στην ανάκτηση πληροφοριών, στην μηχανική μάθηση, στη φυσική επεξεργασία γλώσσας και στην εξαγωγή λέξεων-κλειδιών. Ουσιαστικά στην τεχνική *TF-IDF* η συχνότητα όρου μάς δίνει πληροφορίες για το πόσο συχνά εμφανίζεται ένας όρος σε ένα κείμενο ενώ η αντίστροφη συχνότητα όρου για τη σπανιότητα ενός όρου σε μία συλλογή από κείμενα. Ο πολλαπλασιασμός των τιμών που θα βρεθούν δίνει την τιμή *TF-IDF*. Όσο πιο υψηλή η τιμή *TF-IDF* (πάνω από το 0) τόσο πιο σχετικός (ή σημαντικός) είναι ο όρος ενώ όσο η τιμή πλησιάζει το 0 τόσο λιγότερο σχετικός είναι ο όρος (Simha, n.d.; Stecanella, 2019) (Εικόνα 48).

Word	TF		IDF	TF*IDF	
	A	B		A	B
The	1/7	1/7	$\log(2/2) = 0$	0	0
Car	1/7	0	$\log(2/1) = 0.3$	0.043	0
Truck	0	1/7	$\log(2/1) = 0.3$	0	0.043
Is	1/7	1/7	$\log(2/2) = 0$	0	0
Driven	1/7	1/7	$\log(2/2) = 0$	0	0
On	1/7	1/7	$\log(2/2) = 0$	0	0
The	1/7	1/7	$\log(2/2) = 0$	0	0
Road	1/7	0	$\log(2/1) = 0.3$	0.043	0
Highway	0	1/7	$\log(2/1) = 0.3$	0	0.043

Εικόνα 48. Στιγμιότυπο μετρήσεων *TF-IDF*.

Οι συχνότητες λέξεων που χρησιμοποιήθηκαν στην μέθοδο *TF-IDF* από τους συγγραφείς είναι μίας λέξης (*unigrams*), δύο λέξεων (*bigrams*) (π.χ. *C, CO, OM), τριών (*trigrams*) (**C, * CO, COM) (Abdelminaam et al., 2021; Barrón-Cedeño et al., 2019; Elhadad et al., 2020; Kaliyar et al., 2020; Kausar et al., 2020) τεσσάρων (*fourgrams*) (Abdelminaam et al., 2021), τόσο σε επίπεδο χαρακτήρων (Al-Ahmad et al., 2021; Tundis et al., 2021), όσο και σε επίπεδο λέξεων (Elhadad et al., 2020; Tundis et al., 2021; Yuliani et al., 2019).

Σύμφωνα με τους Ahmed και Gravanis λέξεις όπως το «το» και το «τότε», οι οποίες συχνά υπάρχουν σε ένα κείμενο εάν χρησιμοποιήσουμε την απλή συχνότητα των όρων, αυτές θα υπερισχύσουν στον υπολογισμό της συχνότητας, ενώ με την χρήση της αντίστροφης συχνότητας μειώνεται η επίδραση τους (Ahmed et al., 2018; Gravanis et al., 2019; Meel & Vishwakarma, 2020)

Ο Nouh προσθέτει ότι οι λέξεις ngrams αλλά και οι συχνότητες των όρων έχουν χρησιμοποιηθεί για την επίλυση προβλημάτων όπως οι ομιλίες μίσους και τα εξτρεμιστικά κείμενα και αποδείχτηκαν επιτυχείς (Nouh et al., 2019; Tundis et al., 2021).

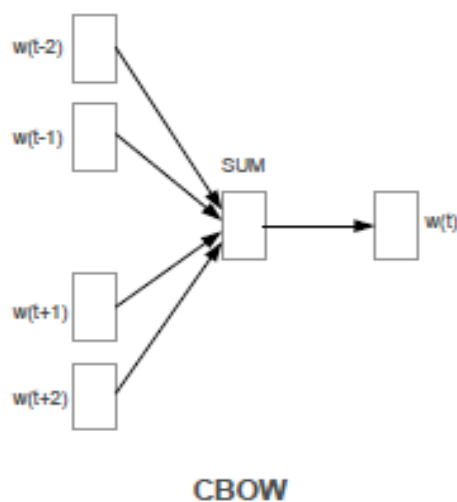
Η καταγραφή της απλής συχνότητας των όρων (TF) έχει κοινά στοιχεία με τη μέθοδο Bag of Words (Al-Ahmad et al., 2021; Elhadad et al., 2020; Khanday et al., 2021; Ozbay & Alatas, 2020; Tundis et al., 2021). Σύμφωνα με τον Al-Ahmad η απλή συχνότητα των όρων μετρά τη συχνότητα συγκεκριμένου όρου στο κείμενο όπως και η μέθοδος Bag of Words. Ο συγγραφέας υπολόγισε την εύρεση της συχνότητας των λέξεων κλειδιών στο έγγραφο διαιρώντας τη με τον συνολικό αριθμό των λέξεων κλειδιών σε όλο το κείμενο με σκοπό την εύρεση των κυριότερων χαρακτηριστικών ψευδούς πληροφόρησης για τον Covid19 (Al-Ahmad et al., 2021). Σύμφωνα με τον Ahmed εάν μία λέξη υπάρχει σε ένα κείμενο θα αναπαρασταθεί με την τιμή 1 εάν όχι με το 0 (Ahmed et al., 2018; Gravanis et al., 2019).

Το μοντέλο *Word2Vec* αποτελεί μία δημοφιλή μέθοδο η οποία χρησιμοποιείται στη φυσική επεξεργασία γλώσσας και ευρύτερα για τον εντοπισμό των ψευδών ειδήσεων σε ένα κείμενο (*Gensim*, n.d.; *Google Code Archive - Long-Term Storage for Google Code Project Hosting.*, n.d.). Το *Word2vec* χρησιμοποιεί ενσωματωμένες λέξεις σε ένα χώρο μεγάλου μεγέθους και τις αναπαριστά ως διανύσματα πραγματικών αριθμών. Τα διανύσματα που είναι κοντά σε απόσταση, αναπαριστούν λέξεις με την ίδια σημασία (Kausar et al., 2020). Οι λέξεις αναπαρίστανται κοντά στο σώμα του κειμένου σε έναν διανυσματικό χώρο χρησιμοποιώντας τη μέθοδο ομοιότητας (cosine similarity). Πιο συγκεκριμένα, ο αλγόριθμος *Word2vec* μαθαίνει συσχετισμούς λέξεων από ένα σώμα κειμένου χρησιμοποιώντας ένα απλό μοντέλο νευρωνικού δικτύου (Borges et al., 2019; Braşoveanu & Andonie, 2021; Gravanis et al., 2019; Huang & Chen, 2020; Kaliyar et al., 2020; Kasseropoulos & Tjortjis, 2021; Kausar et al., 2020; Meel & Vishwakarma, 2020; S. Qian et al., 2021; Shim et al., 2021; Silva et al., 2020; Taskin et al., 2022; Tundis et al., 2021; Q. Zhang et al., 2019).

Περιλαμβάνει δύο επιμέρους μοντέλα το Continuous Bag of Words (CBoW) και το Continuous Skipgram. Ο Shim χρησιμοποίησε το συγκεκριμένο μοντέλο παράλληλα με το μοντέλο *Link2vec* για να διαφοροποιήσει τα αποτελέσματα έρευνας των ψευδών ειδήσεων στο διαδίκτυο από τις αληθείς ειδήσεις (Shim et al., 2021). Ενώ οι Huang και Chen για να

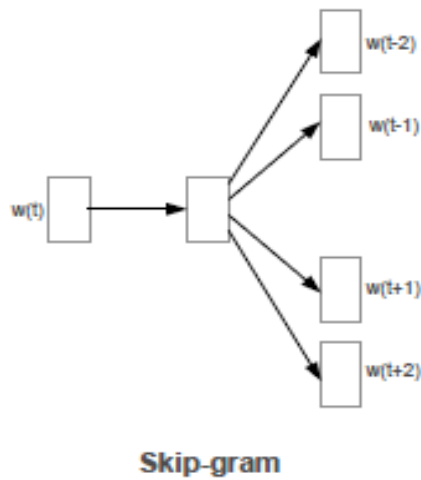
μετρήσουν τις ομοιότητες και τις διαφορές των λέξεων ανάμεσα στις ψευδείς και τις αληθείς ειδήσεις (Huang & Chen, 2020).

Στο μοντέλο Continuous Bag of Words προσδιορίζεται η λέξη «στόχος» από τις συναφείς λέξεις (λέξεις πλαισίου). Οι λέξεις πλαισίου είναι λέξεις πριν και μετά την μεσαία λέξη. Επειδή η σειρά των λέξεων δεν λαμβάνεται υπόψιν για αυτό και το μοντέλο ονομάζεται Continuous bag of words (Εικόνα 49) (Mikolov et al., 2018). Το συγκεκριμένο υπομοντέλο του Word2vec υιοθετήθηκε από τον Kausar ως συμπληρωματική μέθοδος εξαγωγής χαρακτηριστικών για τον εντοπισμό της προπαγάνδας στη γλώσσα Urdu (ProSoul) (Kausar et al., 2020).



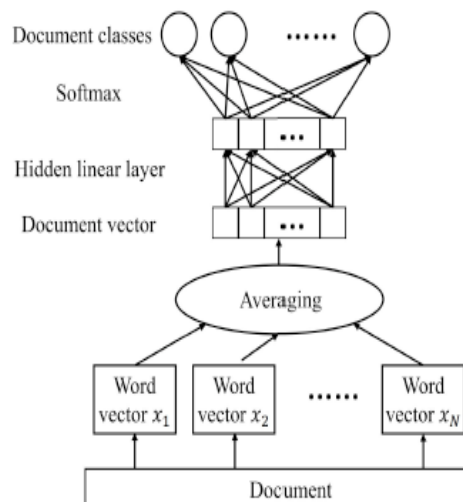
Εικόνα 49. Αρχιτεκτονική μοντέλου CBoW.

Το μοντέλο Skipgram αντίθετα παίρνει ως δεδομένη την τωρινή λέξη και προβλέπει τις παραπλήσιες λέξεις (Mikolov et al., 2013) (Εικόνα 50) (Shim et al., 2021; Silva et al., 2020; Taskin et al., 2022). Υιοθετήθηκε από τον Taskin για την αναπαράσταση μικρών μηνυμάτων κειμένων από το twitter (π.χ. κάτω από 50 λέξεις). Τα χαρακτηριστικά (πίνακας) χρησιμοποιήθηκαν σε αλγόριθμους βαθιάς μάθησης (Taskin et al., 2022). Στην ίδια οπτική και ο Silva υιοθέτησε την συγκεκριμένη υπο-μέθοδο για την αναπαράσταση λέξεων σε διανύσματα τα οποία θα βοηθήσουν περαιτέρω τους παραδοσιακούς αλγόριθμους (π.χ. SVM) στον εντοπισμό ψευδών ειδήσεων στα πορτογαλικά (Silva et al., 2020).

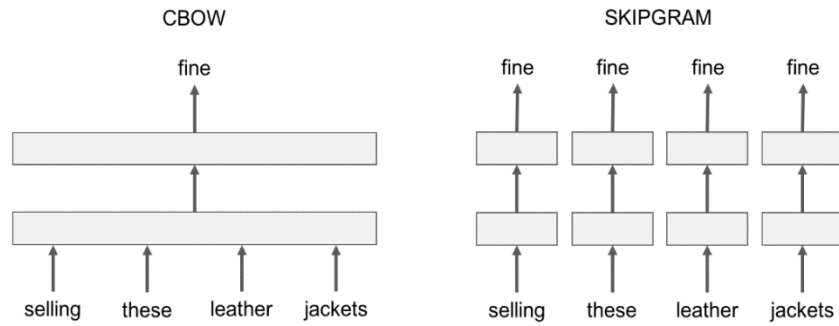


Εικόνα 50. Μοντέλο Skipgram.

Το Fast Text (*FastText*, n.d.) (Εικόνα 51) είναι ένα ανοιχτό λογισμικό για αναπαράσταση κειμένου (απόδοση κατηγοριών) το οποίο δανείζεται στοιχεία από το BoW και από το Word2vec (CBoW) (Εικόνα 52). Οι αναπαραστάσεις του Fast Text βρίσκουν κρυφές πληροφορίες για τη γλώσσα, τη σημασιολογία και τις αναλογίες των λέξεων (Brazoaneanu & Andonie, 2021; Y. Liu et al., 2019; Silva et al., 2020).



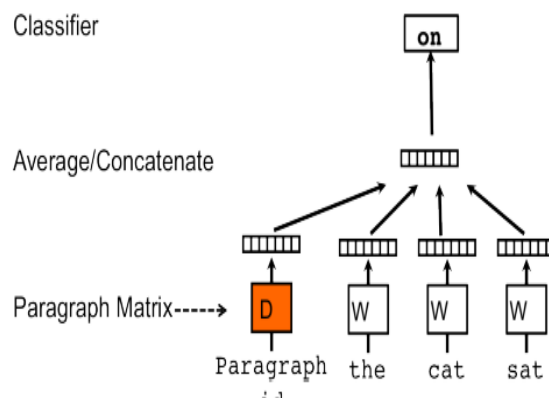
Εικόνα 51. Αρχιτεκτονική Fast Text.



Εικόνα 52. Υπομοντέλα Fast Text.

Το μοντέλο *Doc2vec* αποτελεί επέκταση του *Word2Vec* καθώς έχει προστεθεί μία ακόμα παράμετρος που λαμβάνεται υπόψη, η παράμετρος της παραγράφου. Όπως στο μοντέλο *Word2vec* έτσι και στο συγκεκριμένο μοντέλο η αριθμητική αναπαράσταση του εγγράφου γίνεται σε διανύσματα και το διάνυσμα παραγράφου προβλέπει την επόμενη λέξη στην πρόταση. Το διάνυσμα παραγράφου δημιουργεί αναπαραστάσεις από συχνότητες δεδομένων διαφορετικού μεγέθους. Το *Doc2vec* εφαρμόζεται σε κείμενα με προτάσεις, παραγράφους, και έγγραφα (Cremisini et al., 2019; A. C. Pandey & Tikkiwal, 2021; Shu, Cui, et al., 2019; Tundis et al., 2021).

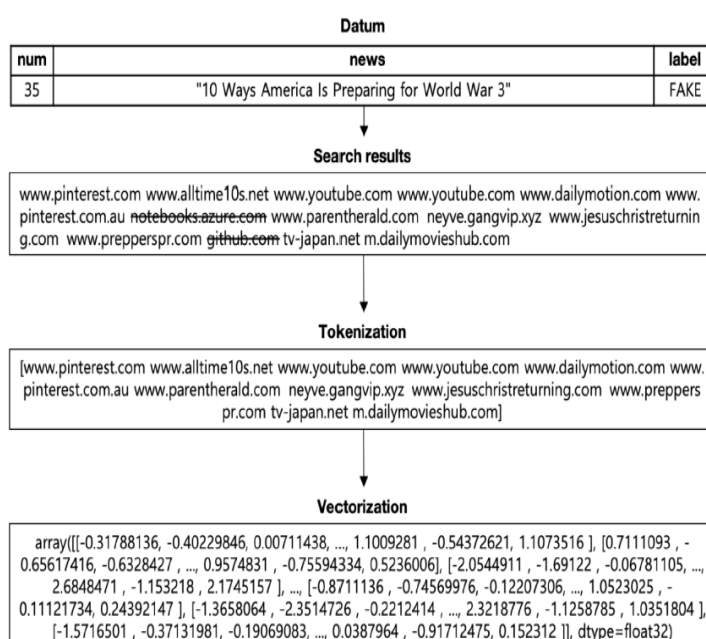
Το *Doc2vec* (Εικόνα 53) σύμφωνα με την Cremisini αποτελεί μία από τις διαδεδομένες τεχνικές κατηγοριοποίησης κειμένου λόγω των πολύ καλών αποτελεσμάτων αλλά και της σημασιολογικής ομοιότητας διαφορετικών συντακτικά κειμένων σε σχέση με την τεχνική *Bag of Words* (Cremisini et al., 2019).



Εικόνα 53. Μοντέλο Doc2Vec.

Το *Link2vec* είναι επέκταση του μοντέλου *Word2vec* με χαρακτηριστική διαφορά την μετατροπή των συνδέσμων σε διανύσματα και όχι των λέξεων ενός κειμένου ή των εγγράφων όπως τα μοντέλα *Word2vec* και *Doc2vec*. Το *Link2vec* αναγνωρίζει συνδέσμους ως λέξεις και

συνδυασμούς συνδέσμων σαν πρόταση και τους μετατρέπει σε διανύσματα (Εικόνα 54). Επιπλέον, η εκμάθηση συνδέσμου γίνεται αυτόματα στα ερευνητικά αποτελέσματα. Ο Shim ο οποίος δημιούργησε το συγκεκριμένο μοντέλο αναφέρει ότι ο μεγάλος αριθμός συλλογής συνδέσμων δημιουργεί «θόρυβο» (δηλαδή πληθώρα μη σχετικών συνδέσμων) ενώ ο πολύ μικρός αριθμός δεν αντανακλά αρκετά στοιχεία επομένως και είναι απαραίτητο να τεθεί συγκεκριμένος αριθμός συνδέσμων (π.χ. 15 κορυφαίοι σύνδεσμοι). Το μοντέλο Link2vec μαθαίνει ουσιαστικά μοτίβα συνδέσμων (από αποτελέσματα αναζήτησης ειδήσεων στον ιστό) και εφαρμόζει ένα μοντέλο εντοπισμού ψευδών ειδήσεων (Shim et al., 2021).



Εικόνα 54. Μετατροπή συνδέσμων στο μοντέλο Link2vec.

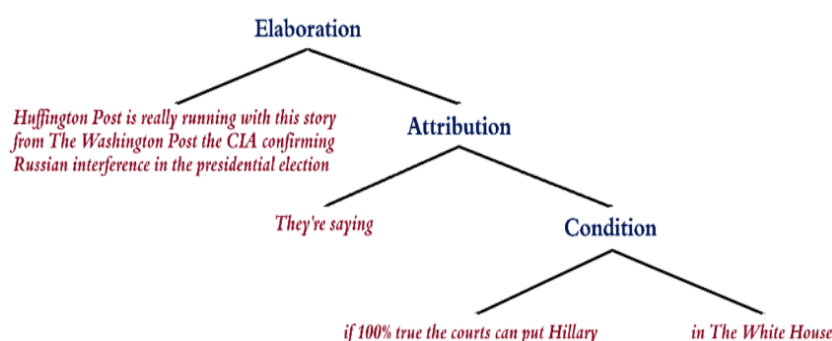
Σε επίπεδο γραφής η *Θεωρία Ρητορικής Δομής (Rhetorical Structure Theory)* (*Rhetorical Structure Theory*, n.d.-a) αποτυπώνει τις συχνότητες των ρητορικών σχέσεων ανάμεσα σε διαφορετικές μονάδες κειμένου και περιγράφει την δενδρική δομή για κάθε ιστορία¹⁵ (Εικόνα 55) (Isaakidou et al., 2021; Shu, Cui, et al., 2019; Shu, Wang, et al., 2019). Το κείμενο σύμφωνα με τη Rubin, αναλύεται σε ένα σύνολο από ρητορικές σχέσεις οι οποίες συνδέονται με ιεραρχικό τρόπο με πιο σημαντικές μονάδες κειμένου, οι οποίες ηγούνται του ιεραρχικού δέντρου (Rubin et al., 2015). Η έμφαση δίνεται σε ορισμένα τμήματα κειμένου

¹⁵ Η θεωρία ρητορικής δομής χρησιμοποιείται στην ανάλυση κειμένου και βασίζεται στην οπτική ότι ένα κείμενο δεν είναι μόνο μια σειρά από προτάσεις αλλά αποτελείται από ομάδες προτάσεων που σχετίζονται η μία με την άλλη (δες *Rhetorical Structure Theory*, n.d.-a) (Mann & Thompson, 1988). Επιπλέον, παρέχεται online μέσω του RST tool (ανάλυση διαγραμμάτων RST) και webtool και δίνεται η δυνατότητα εισαγωγής στη demo μορφή (*RstWeb - Homepage*, n.d.)

προκειμένου να αποδώσουν τη βασική ιδέα, τα οποία προσδιορίζονται μέσω της ανάλυσης ρητορικών σχέσεων σε περισσότερο (πυρήνας) και λιγότερο σημαντικά τμήματα (δορυφόρος).

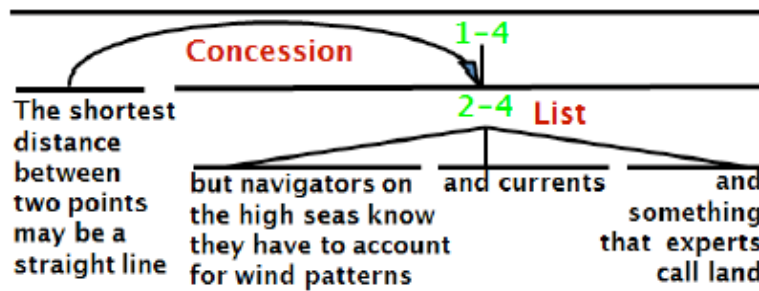
Η θεωρία διαχωρίζει τις σχέσεις σε: *πυρήνες* όπου βρίσκονται οι βασικές πληροφορίες, και *δορυφόρους* δηλαδή τις πρόσθετες πληροφορίες (υποτακτικές-παρατακτικές σχέσεις) αλλά και *σε πολυπυρηνικές*. Οι ρητορικές σχέσεις μπορούν επίσης να κατηγοριοποιηθούν σε: θεματικές-περιεχομένου οι οποίες αφορούν την αναγνώριση του αναγνώστη της υπό ερώτησης σχέσης όπως συνθήκη, ανάπτυξη, αξιολόγηση, ερμηνεία, μέσο, μη ηθελημένη αιτία, μη ηθελημένο αποτέλεσμα, διαφορετική περίπτωση, σκοπός, επίλυση, απόλυτη, εκτός αν, ηθελημένη αιτία, ηθελημένο αποτέλεσμα και τις παραστατικές που δηλώνουν παρακίνηση στον αναγνώστη (επιθυμία για δράση) όπως παραχώρηση, αντίθεση, υπόβαθρο, απόδειξη, ικανότητα, δικαιολογία, κίνητρο, προετοιμασία, επαναδιατύπωση, περίληψη.

Οι κατηγορίες των πολυπυρηνικών σχέσεων είναι μία πρόσθετη κατηγορία οι οποίες δεν έχουν μία ξεκάθαρη επιλογή ενός δορυφόρου, για παράδειγμα η σχέση αντίθεση¹⁶, και αποτελούνται από σχέσεις όπως: αντίθεση, σύνδεση, αποσύνδεση, ένωση, λίστα (σύνολο από γεγονότα), πολυπυρηνική επαναδιατύπωση, συχνότητα (*Rhetorical Structure Theory*, n.d.-b) (Εικόνα 55, Εικόνα 56). Η θεωρία ρητορικής δομής ουσιαστικά χαρτογραφεί συχνότητες ρητορικών σχέσεων σε ένα διανυσματικό μοντέλο χώρου (*Vector Space Model*) (*Jiyfeng / DPLP: A RST Parser with a Trained Model*, n.d.). Η συγκεκριμένη θεωρία χρησιμοποιήθηκε από τον Shu για τον εντοπισμό των ψευδών ειδήσεων (Shu, Wang, et al., 2019).



Εικόνα 55. Θεωρία ρητορικής δομής.

¹⁶ Δες και ορισμό αντίθεσης (*Rhetorical Structure Theory*, n.d.-b)

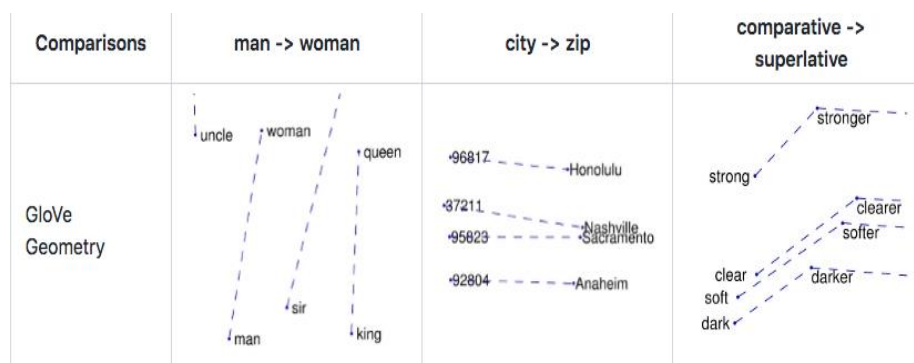


Εικόνα 56. Θεωρία ρητορικής δομής-RST Webtool.

Το *Glove* αποτελεί ακόμη μία γνωστή μέθοδος που χρησιμοποιείται για την εξαγωγή αναπαραστάσεων. Το *Glove* όπως και τα προηγούμενα μοντέλα (*Word2vec*, *Doc2vec*) μετρά τη συχνότητα λέξεων και την εγγύτητά τους σε ένα κείμενο (π.χ. tweet).

Ο Pennington ο οποίος είναι και ο δημιουργός του μοντέλου, αναφέρει χαρακτηριστικά ότι στόχος του μοντέλου είναι η αναλογία της συχνότητας των λέξεων να μπορεί να αποδώσει κάποια ερμηνεία (Pennington et al., 2014). Ο συγγραφέας δίνει ως παράδειγμα τις συχνότητες στοχευμένων λέξεων όπως ο πάγος και ο ατμός, όπου και οι δύο συσχετίζονται με το «νερό», αλλά και τη μη σχετική λέξη «μόδα» η οποία επίσης σχετίζεται με το νερό αλλά λιγότερο συχνά.

Το *Glove* χρησιμοποιεί τη μετρική cosine similarity (Euclidean distance) ως μέθοδο γλωσσολογικής ή σημασιολογικής ομοιότητας των αντιστοιχισμένων λέξεων. Η συγκεκριμένη μετρική μπορεί να αποκαλύψει σχετικές λέξεις εκτός λεξιλογίου της ανθρώπινης γλώσσας. Οι ομοιότητες των μετρικών που προκύπτουν από τη μετρική cosine similarity μπορούν να παράγουν γραμμικές αναπαραστάσεις σχετικότητας λέξεων (Kaliyar et al., 2020; Pennington et al., 2014) (Εικόνα 57).



Εικόνα 57. Παράδειγμα γραμμικής αναπαράστασης *Glove*.

Επιπλέον το μοντέλο *Glove* επιλύει προβλήματα όπως η διαφοροποίηση των λέξεων «άντρας», «γυναίκα» προσδιορίζοντας το νόημα από τις συμφραζόμενες λέξεις (Pennington

et al., 2014). Το συγκεκριμένο μοντέλο υιοθετήθηκε από τους Goldani, Kasseropoulos και Tjortjis, Du, Borges, Gravanis, Abdelminaam, Braşoveanu και Andonie, Kaliyar, Meel και Viswakarma, Bonet-Jover (Abdelminaam et al., 2021; Bonet-Jover et al., 2021; Borges et al., 2019; Braşoveanu & Andonie, 2021; Du et al., 2021; Goldani et al., 2021; Gravanis et al., 2019; Kaliyar et al., 2020; Kasseropoulos & Tjortjis, 2021; Meel & Vishwakarma, 2020).

Σύμφωνα με τον Abdelminaam, στο Glove τα διανύσματα δημιουργούνται από ενσωματώσεις παγκόσμιων στατιστικών μετρικών συχνότητας λέξεων που υπάρχουν σε βάσεις δεδομένων μαζί με μετρικές εγγύτητας των λέξεων π.χ. σε ένα tweet (Abdelminaam et al., 2021). Οι συγγραφείς το υιοθέτησαν με αλγόριθμους βαθιάς μάθησης (Abdelminaam et al., 2021; Borges et al., 2019; Du et al., 2021; Kaliyar et al., 2020) ενώ ο Gravanis με απλούς αλγόριθμους μηχανικής μάθησης (Gravanis et al., 2019).

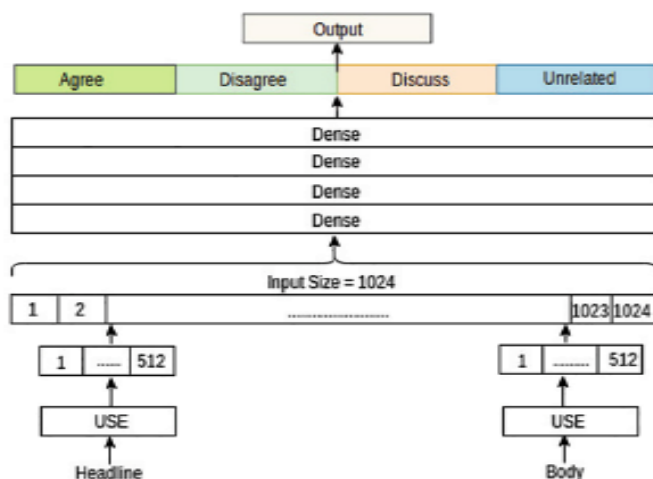
Το μοντέλο *Ngram* είναι απλό αλγοριθμικό μοντέλο το οποίο εξάγει λέξεις με υψηλή συχνότητα όπως unigrams (συχνότητα μιας λέξης), bigrams (συχνότητα δύο λέξεων), trigrams (συχνότητα τριών λέξεων) (Borges et al., 2019; Huang & Chen, 2020; Hunt et al., 2022; Kausar et al., 2020; Posadas-Durán et al., 2019; Tundis et al., 2021). Η απλότητα και η επεκτασιμότητά του σύμφωνα με τους Huang και Chen το έκανε ιδιαίτερα δημοφιλές ως προς τη χρήση σε πιθανότητες, θεωρία επικοινωνίας, φυσική επεξεργασία γλώσσας και στη συμπύεση δεδομένων (Huang & Chen, 2020). Οι συγγραφείς υιοθέτησαν το μοντέλο για τον εντοπισμό των διαφορών ανάμεσα σε λέξεις και προτάσεις που εμφανίζονται στις αληθείς και ψευδείς ειδήσεις. Για τον εντοπισμό των διαφορών μέτρησαν την συχνότητα μίας λέξης (unigrams) (π.χ. ufo, http) ή δύο λέξεων (bigrams) (π.χ. brain, force).

Οι Hunt και Huang και Chen μέτρησαν *συχνότητες δύο λέξεων (bigrams) και τριών (trigrams) (word ngrams)*. Οι Huang και Chen μάλιστα διαπίστωσαν διαφορές ως προς την συχνότητα των ngrams στις ψευδείς και τις αληθείς ειδήσεις (π.χ. συχνότερη χρήση bigrams στις ψευδείς ειδήσεις) (Huang & Chen, 2020).

Οι Posadas-Durán και Hunt μέτρησαν *συχνότητες τριών* (Hunt et al., 2022; Posadas-Durán et al., 2019) και *τεσσάρων γραμμάτων* (χαρακτήρες γραμμάτων n) (Posadas-Durán et al., 2019). Ο Posadas-Durán για την αναπαράσταση του στυλ γραφής του συγγραφέα ενώ ο Hunt για βελτίωση της πρόβλεψης (Hunt et al., 2022; Posadas-Durán et al., 2019).

Το *USE (TensorFlow Hub, n.d.)* (Εικόνα 58) αποτελεί άλλο ένα μοντέλο αναπαράστασης λέξεων, τίτλου και σώματος (προτάσεων είτε εγγράφων) το οποίο υιοθετήθηκε από τον Saikh (Saikh et al., 2019). Σύμφωνα με τον Saikh το συγκεκριμένο μοντέλο περιλαμβάνει δύο υπομοντέλα το Transformer και το DAN (Deep Average Network).

Ο συγγραφέας χρησιμοποίησε το *Transformer* γιατί το συγκεκριμένο μοντέλο αποδίδει καλύτερα σε σχέση με το DAN ως προς τη μεταφορική μάθηση¹⁷. Το μοντέλο *Transformer* κωδικοποιεί αναπαραστάσεις λέξεων σε μία πρόταση λαμβάνοντας υπόψη και τη σειρά και την ταυτότητα των άλλων λέξεων, χρησιμοποιώντας έναν μηχανισμό προσοχής για να υπολογίσει το περιεχόμενο. Από τον μέσο όρο αναπαραστάσεων περιεχομένου των λέξεων υπολογίζεται η ενσωμάτωση των προτάσεων (Cer et al., 2018).



Εικόνα 58. Αρχιτεκτονική Μοντέλου USE.

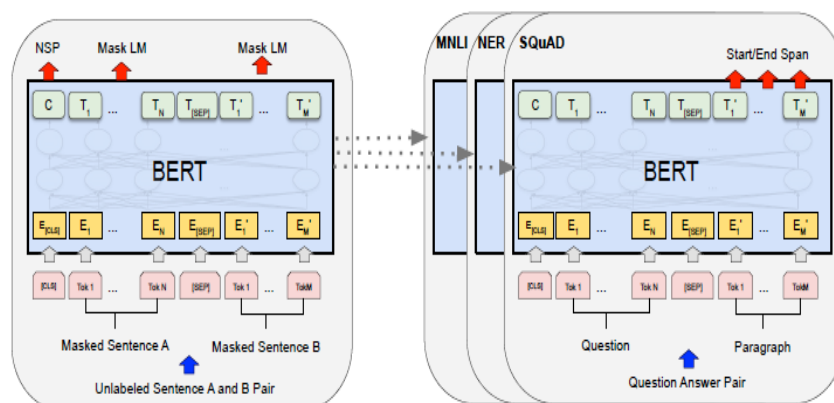
Ο Saikh χρησιμοποίησε τον *τίτλο και το σώμα* ως χαρακτηριστικά του μοντέλου USE όπου οι αναπαραστάσεις του θα χρησιμοποιηθούν ως χαρακτηριστικό του συνελκτικού νευρωνικού δικτύου (Feed-forward network) (Saikh et al., 2019).

Οι *ενσωματώσεις Tensor* είναι ένα μοντέλο που παρέχει συνοπτικές ενσωματώσεις ισχυρισμών που χρησιμοποιούνται για να δημιουργήσουν γράφους ισχυρισμών ως προς το ποιες κατηγορίες διαδίδονται (Q. Zhang et al., 2019). Ο Zhang υιοθέτησε το συγκεκριμένο μοντέλο ως μοντέλο σύγκρισης με το μοντέλο Bayes βαθιάς μάθησης.

Το *μοντέλο BERT* (Google-Research / Bert, 2020) (Εικόνα 59) χρησιμοποιεί προεκπαιδευμένες (γλωσσολογικές) αναπαραστάσεις διπλής κατεύθυνσης (Kausar et al., 2020; Tundis et al., 2021; Ying et al., 2021). Περιλαμβάνει πολυεπίπεδη κωδικοποίηση με ενσωμάτωση μίας ή δύο προτάσεων μαζί. Έχει την ικανότητα να εμβαθύνει εάν επεκταθεί σε 12 επίπεδα (Aggarwal & Sadana, 2019) και μπορεί να αποτυπώσει τη σημασία μίας λέξης με

¹⁷ μεταφορική μάθηση: η υιοθέτηση όσο το δυνατόν περισσότερης γνώσης από ένα προϋπάρχον μοντέλο μηχανικής μάθησης σε ένα καινούργιο το οποίο έχει σχεδιαστεί για την ίδια εργασία (π.χ. η μεταφορά γενικών χαρακτηριστικών ως προς το πόσα αντικείμενα κατηγοριοποιήθηκαν και η προσθήκη εξειδικευμένης γνώσης στο καινούργιο μοντέλο) δεξ (Seldon, 2021)

βάση της θέσης της και την σειρά των λέξεων σε μία πρόταση. Σχεδιάστηκε για να προεκπαιδεύει αναπαραστάσεις διπλής κατεύθυνσης σε βαθύτερο επίπεδο από μη κατηγοριοποιημένο κείμενο, μαθαίνοντας τη σημασιολογία τους δεξιά και αριστερά σε όλα τα επίπεδα. Χρησιμοποιεί ένα μοντέλο κάλυψης γλώσσας (mask). Το συγκεκριμένο μοντέλο καλύπτει κάποιες λέξεις από τα δεδομένα εισόδου και στόχος του είναι να προβλέψει την προέλευση της ταυτότητας της λέξης βασιζόμενο μόνο στα συμφραζόμενα (Devlin et al., 2019; Kausar et al., 2020; Ying et al., 2021). Έχει χρησιμοποιηθεί από τον Ying για την αναπαράσταση πολλαπλών δημοσιεύσεων (κείμενο και εικόνα) (Ying et al., 2021).



Εικόνα 59. Μοντέλο BERT (προεκπαίδευση και προσαρμογή).

4.4 Χαρακτηριστικά Αλγορίθμων Μηχανικής Μάθησης

Η μηχανική μάθηση είναι η επιστήμη που εστιάζεται στο σχεδιασμό και την ανάπτυξη διαδικασιών των αλγορίθμων που θα επιτρέψει την κατηγοριοποίηση, ομαδοποίηση ή πρόβλεψη μεγάλου αριθμού δεδομένων που συλλέχθηκαν από εντοπιστές ή σε μία βάση δεδομένων (Taskin et al., 2022). Οι προσεγγίσεις μηχανικής μάθησης βασίζονται σε αλγόριθμους οι οποίοι χωρίζονται σε επιβλεπόμενης μάθησης (παραδοσιακοί), πολλαπλής εκμάθησης, μη επιβλεπόμενης μάθησης, βαθιάς μάθησης, μεταεπιστημονικούς ή εξελικτικούς αλγόριθμους καθώς και νευρωνικά δίκτυα.

Οι συγκεκριμένες τεχνικές έχουν υιοθετηθεί για περισσότερη ακρίβεια και καλύτερη κατηγοριοποίηση των ειδήσεων από τους συγγραφείς της σχετικής βιβλιογραφίας και έχουν χρησιμοποιηθεί σε πολλαπλά επιστημονικά πεδία όπως η επεξεργασία εικόνας και η οπτική αναγνώριση προσώπων και αντικειμένων, για την αξιολόγηση του ρίσκου και των αλγοριθμικών συναλλαγών, για την φαρμακευτική έρευνα αλλά και διάγνωση των όγκων,

την αυτοκίνηση και την κατασκευαστική βιομηχανία αλλά για την αναγνώριση ομιλίας στη φυσική επεξεργασία γλώσσας (Tundis et al., 2021).

4.4.1 Αλγόριθμοι Επιβλεπόμενης Μάθησης - Παραδοσιακοί (Supervised Learning)

Η συγκεκριμένη κατηγορία αλγορίθμων εκπαιδεύεται σε κατηγοριοποιημένα δεδομένα εκπαίδευσης (δεδομένα και ετικέτες δεδομένων) και με βάση τα συγκεκριμένα δεδομένα οι αλγόριθμοι προβλέπουν τα δεδομένα εξόδου (ο αλγόριθμος μαθαίνει ένα δίκτυο σχέσεων π.χ. οπτική αναγνώριση χαρακτήρων) (Taskin et al., 2022). Στην συγκεκριμένη κατηγορία ανήκουν οι αλγόριθμοι kNN, Decision Tree, Logistic Regression, Linear Regression, SVM, κ.ά.

Ο αλγόριθμος κοντινότερων γειτόνων ή *kNN* (1.6. *Nearest Neighbors*, n.d.) έχει υιοθετηθεί από αρκετούς συγγραφείς της βιβλιογραφίας τους όπως οι Kaliyar, Meel και Viswakarma, Reis, Taskin, Gravanis, Abonizio κ.ά. σε χαρακτηριστικά περιεχομένου, συναισθήματος, υβριδικά (δηλαδή συναισθηματικά και περιεχομένου) (Abdelminaam et al., 2021; Abonizio et al., 2020; Ahmad et al., 2020; Al-Ahmad et al., 2021; Cremisini et al., 2019; Elhadad et al., 2020; Goldani et al., 2021; Gravanis et al., 2019; Hunt et al., 2022; Kaliyar et al., 2020; Y. Liu et al., 2019; Mahabub, 2020; Meel & Vishwakarma, 2020; Reis et al., 2019; Sheikhi, 2021; Taskin et al., 2022; Varshney & Vishwakarma, 2021). Ο kNN είναι ένας αλγόριθμος ο οποίος υπολογίζει τους πλησιέστερους γείτονες με τιμή K και ταξινομεί το αποτέλεσμα με βάση τη μέγιστη ψήφο των γειτόνων (τιμή K). Δηλαδή εάν η τιμή K είναι 1 τότε το νέο σημείο δεδομένων αποδίδεται σε τάξη με την κοντινότερη απόσταση. Η απόσταση υπολογίζεται βάση της Ευκλείδειας απόστασης (π.χ. $k=2$) (Ahmad et al., 2020). Οι Taskin και Bonet-Jover υιοθέτησαν την υπο-τάξη του αλγορίθμου kNN (K neighbors classifier). Η συγκεκριμένη μέθοδος περιλαμβάνει τρεις αποστάσεις εκτός από την Ευκλείδεια, τις Manhattan και Minkowski. Εάν η απόσταση p είναι 1 χρησιμοποιείται η λειτουργία της απόστασης Manhattan, εάν είναι 2, η Ευκλείδεια απόσταση και πάνω από 3, η Minkowski (Bonet-Jover et al., 2021; Taskin et al., 2022).

Πιο συγκεκριμένα, σε δεδομένα εκπαίδευσης τα οποία περιλαμβάνουν παραδείγματα ειδήσεων ο συγκεκριμένος αλγόριθμος μπορεί να βοηθήσει στον διαχωρισμό των κοντινότερων ειδήσεων στο αν είναι ψευδείς ή όχι. Οι κοντινότερες ειδήσεις αντιπαραβάλλονται με τα δεδομένα εκπαίδευσης με την εύρεση ενός κοντινότερου σημείου και αντιπαραβολής του με τα δεδομένα εκπαίδευσης. Το κοντινότερο αρχείο K χαρακτηρίζεται ως ψευδές (Mahabub, 2020).

Οι Ozbay και Alatas και Meel και Vishwakarma αναφέρουν τον αλγόριθμο *lbc* ως παρόμοιο με τον kNN. Ο αλγόριθμος *lbc* έχει επηρεαστεί από τον kNN με τη μόνη διαφορά ότι ο συγκεκριμένος αλγόριθμος δημιουργεί πρόβλεψη σε ένα δείγμα και όχι ταξινόμηση. Ο αλγόριθμος υπολογίζει την απόσταση για εύρεση κοντινών σημείων στα δεδομένα εκπαίδευσης για κάθε δείγμα και χρησιμοποιεί επιλεγμένα σημεία για υπολογισμό. Στόχος του είναι να παρέχει ανάλυση σε ποιες τάξεις σχέσεων πρέπει να εκπαιδευτεί (Meel & Vishwakarma, 2020; Ozbay & Alatas, 2020).

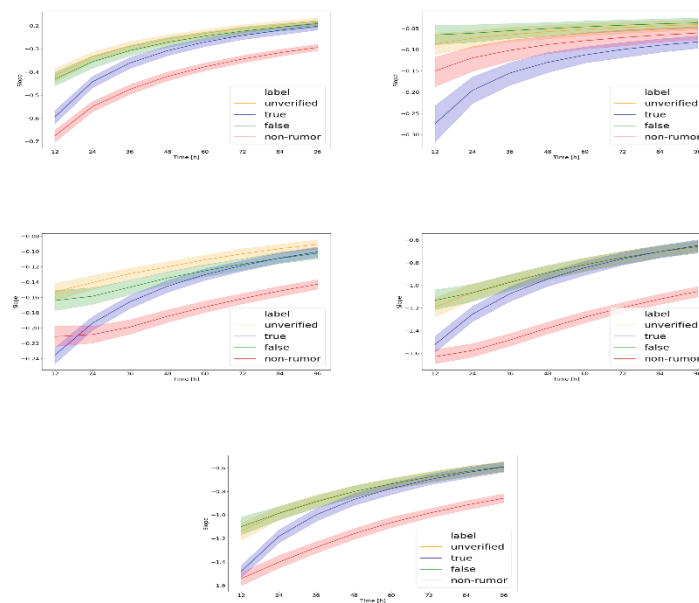
Οι Meel και Vishwakarma αναφέρουν και τον *K*Star* ως άλλον έναν αλγόριθμο ταξινόμησης. Ο *K*Star* θεωρείται «τεμπέλικος» αλγόριθμος διότι αργεί να παράξει άμεσα αποτελέσματα. Είναι ένας αλγόριθμος ομοιοτήτων ο οποίος παράγει μία ετικέτα τάξης με βάση την ομοιότητα του ερωτήματος σε ένα κοντινότερο γείτονα ενός συνόλου δεδομένων.

Η *λογιστική παλινδρόμηση (Logistic Regression) (1.1. Linear Models, n.d.)* κατηγοριοποιεί τα δεδομένα σε δυαδικές ή πολλαπλές τάξεις (Ahmad et al., 2020) μέσω της πρόβλεψης ενός γεγονότος, αποτελέσματος ή παρατήρησης (π.χ. 0, 1, λάθος, σωστό) (*What Is Logistic Regression?, n.d.*). Ο συγγραφέας τη χρησιμοποίησε για την ταξινόμηση των ψευδών και αληθών ειδήσεων επισημαίνοντας ότι η κατηγοριοποίηση τους θεωρείται δυαδικό πρόβλημα. Επιπλέον, ο αλγόριθμος υπολογίζει την πιθανότητα να συμβεί ένα γεγονός (π.χ. ψήφος ή μη ψήφος) βάση μίας δεδομένης βάσης δεδομένων από ανεξάρτητες τιμές (0,1).

Σύμφωνα με τον Abdelminaam ο συγκεκριμένος αλγόριθμος χρησιμοποιεί τη σιγμοειδή συνάρτηση (σιγμοειδή καμπύλη) για τη μετατροπή των δεδομένων σε πιθανοτικές τιμές. Σε περίπτωση που ταξινομηθούν τα δεδομένα σε δύο τάξεις, εάν το εύρος της πιθανότητας είναι 0.5, τα δεδομένα θα είναι στην τάξη 0 ενώ εάν το εύρος της πιθανότητας είναι πάνω από 0.5 τότε τα δεδομένα θα είναι στη δεύτερη τάξη (τάξη 1) (Abdelminaam et al., 2021). Ο συγκεκριμένος αλγόριθμος υιοθετήθηκε και από άλλους συγγραφείς της παρούσας βιβλιογραφίας όπως τους Bonet-Jover, Shim, Braşoveanu & Andonie, Silva, Kausar et al, Meel και Vishwakarma κ.α. (Abdelminaam et al., 2021; Ahmad et al., 2020; Bonet-Jover et al., 2021; Braşoveanu & Andonie, 2021; Du et al., 2021; Elhadad et al., 2020; Girgis et al., 2018; Ibrishimova & Li, 2020; Kausar et al., 2020; Mahabub, 2020; Meel & Vishwakarma, 2020; Posadas-Durán et al., 2019; Shim et al., 2021; Shu, Wang, et al., 2019; Silva et al., 2020; Tundis et al., 2021; Varshney & Vishwakarma, 2021; Y. Yu et al., 2020). Σύμφωνα με τον Kausar η λογιστική παλινδρόμηση στοχεύει στη μεγιστοποίηση της ποιότητας των προβλέψεων για να χαρτογραφήσει τα χαρακτηριστικά του κειμένου και τις κατηγορίες του (Kausar et al., 2020). Χρησιμοποιήθηκε από τον Tacchini για τον εντοπισμό των hoaxes σε δημοσιεύσεις στο

Facebook λόγω του μεγάλου όγκου των δεδομένων και των χαρακτηριστικών (Abdelminaam et al., 2021; Tacchini et al., 2017).

Η γραμμική παλινδρόμηση (*Linear Regression*) είναι ένας αλγόριθμος ο οποίος υιοθετήθηκε από τους Meel και Vishwakarma, Goldani, Weiss, Varshney και Vishwakarma με χαρακτηριστικά περιεχομένου, συναισθηματικά και υβριδικά (δηλαδή περιεχόμενο και συναίσθημα) (Goldani et al., 2021; Meel & Vishwakarma, 2020; Varshney & Vishwakarma, 2021; Weiss et al., 2019). Ο Weiss χρησιμοποίησε τη γραμμική παλινδρόμηση για την εύρεση της ταχύτητας καθώς και τη διαφορά ως προς την ταχύτητα μετάδοσης των φημών (Weiss et al., 2019) (Εικόνα 60).



Εικόνα 60. Γραμμική παρουσίαση χαρακτηριστικών (δημοσιεύσεων, σχολίων, αλληλεπιδράσεων).

Ο Barrón-Cedeño χρησιμοποίησε την *εντροπία* για τον διαχωρισμό των δεδομένων σε 4 τάξεις: έμπιστα, hoax, σάτιρα, προπαγάνδα (Barrón-Cedeño et al., 2019). Η εντροπία είναι ένας αλγόριθμος ταξινομητής που επιλέγει εκείνο το μοντέλο που έχει τη μεγαλύτερη εντροπία. Χρησιμοποιείται στην ταξινόμηση κειμένου επειδή οι λέξεις δεν είναι ανεξάρτητες μεταβλητές (Vryniotis, 2013). Η αρχή της μέγιστης εντροπίας είναι εκείνη που ικανοποιεί τα όρια της προηγούμενης γνώσης (*Maximum Entropy*, 2015).

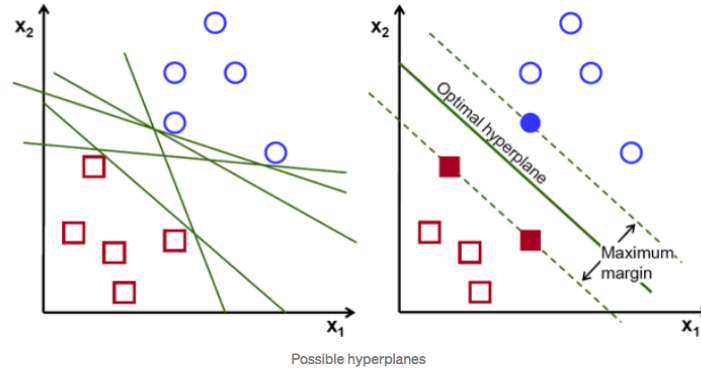
Ο *SVM* (ή *μηχανή διανυσματικής υποστήριξης*) είναι από τους πιο πολυχρησιμοποιημένους και διαδεδομένους αλγόριθμους της μηχανικής μάθησης. Ο SVM χωρίζει τα δεδομένα σε δύο άξονες (δύο υποχώρους) και υπολογίζει το βέλτιστο σημείο (όριο απόφασης) ανάλογα με το σύνολο των χαρακτηριστικών για την κατηγοριοποίηση των σημείων των δεδομένων δηλαδή προσδιορίζει μία γραμμή που διαχωρίζει τα σημεία

δεδομένων δύο τάξεων με μέγιστο όριο (hyperplane) (Gandhi, 2018b; Taskin et al., 2022). Χρησιμοποιείται για την ταξινόμηση του κειμένου (π.χ. προπαγάνδα, μη προπαγάνδα) (Khanday et al., 2021).

Σύμφωνα με τον Zhang, ο SVM (1.4. *Support Vector Machines*, n.d.) είχε την καλύτερη απόδοση ως προς τον εντοπισμό της παραπληροφόρησης σε αντίστοιχη διοργάνωση για τις ψευδείς ειδήσεις (*RumourEval: Determining Rumour Veracity and Support for Rumours < SemEval-2017 Task 8*, n.d.; Q. Zhang et al., 2019). Ο Mahabub επισημαίνει ότι η τεχνική του SVM βασίζεται στην απομόνωση των ψευδών και αληθών ειδήσεων μέσω ενός επιλεγμένου ορίου διαμορφωμένου σε συγκεκριμένες συνθήκες. Τον χρησιμοποίησε για την εύρεση της ομοιότητας ανάμεσα στα δεδομένα εκπαίδευσης και σε πλησιέστερες ειδήσεις (ομοιότητα με τεχνική kNN) (Mahabub, 2020; A. Pandey et al., 2014).

Ο Abdelminaam προσθέτει ότι ο συγκεκριμένος αλγόριθμος χρησιμοποιεί μία μη γραμμική κατηγοριοποίηση με την άμεση μετατροπή των δεδομένων εισόδου σε χώρο μέγιστων διαστάσεων. Υιοθετείται συχνά με γλωσσολογικά και οπτικά χαρακτηριστικά (Abdelminaam et al., 2021).

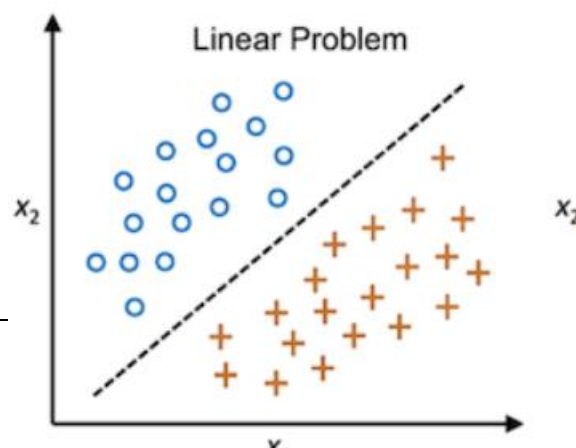
Οι Yu και Hunt αναφέρουν και αυτοί την χρήση του SVM ως προς τον εντοπισμό των καναλιών παραπληροφόρησης σε εταιρείες αλλά και τον εντοπισμό των ψεύτικων δημοσιεύσεων στο twitter αντίστοιχα, δεδομένου των πολύ καλών αποτελεσμάτων του. Την ίδια οπτική συμμερίζονται και οι υπόλοιποι συγγραφείς που υιοθέτησαν τον συγκεκριμένο αλγόριθμο (Abdelminaam et al., 2021; Abonizio et al., 2020; Ahmad et al., 2020; Barrón-Cedeño et al., 2019; Bonet-Jover et al., 2021; Braşoveanu & Andonie, 2021; Du et al., 2021; Girgis et al., 2018; Goldani et al., 2021; Gravanis et al., 2019; Hunt et al., 2022; Kar, 2020; Kasseropoulos & Tjortjis, 2021; Khanday et al., 2021; Y. Liu et al., 2019; Mahabub, 2020; Meel & Vishwakarma, 2020; Ozbay & Alatas, 2020; Papadopoulou et al., 2019; Posadas-Durán et al., 2019; Reis et al., 2019; Saikh et al., 2019; Sheikhi, 2021; Shim et al., 2021; Silva et al., 2020; Souza Freire et al., 2021; Taskin et al., 2022; Tundis et al., 2021; Varshney & Vishwakarma, 2021; Vishwakarma et al., 2019; Y. Yu et al., 2020; Q. Zhang et al., 2019) (Εικόνα 61).



Εικόνα 61. Στιγμιότυπο από SVM.

Ο γραμμικός SVM (*Sklearn.Svm.LinearSVC*, n.d.) είναι μία υποκατηγορία του αλγόριθμου SVM (Abonizio et al., 2020; Ahmad et al., 2020; Cremisini et al., 2019; Elhadad et al., 2020; Khanday et al., 2021; Posadas-Durán et al., 2019; Varshney & Vishwakarma, 2021). Η λειτουργία του είναι πανομοιότυπη με του SVM και χρησιμοποιείται σε δεδομένα με γραμμικό διαχωρισμό (

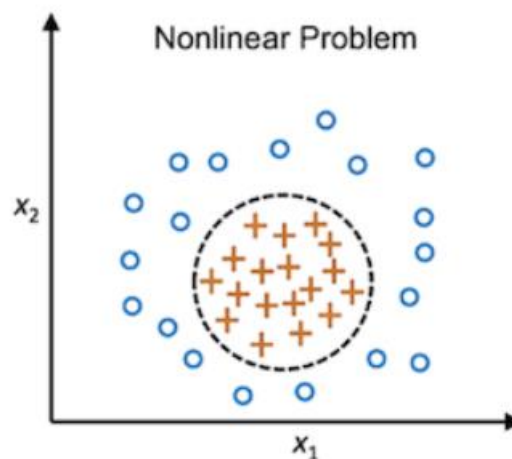
Εικόνα 62). Αυτό σημαίνει ότι η βάση δεδομένων μπορεί να διαχωριστεί σε δύο τάξεις χρησιμοποιώντας μια ευθεία γραμμή (Purale, 2019; *Support Vector Machine (SVM) Algorithm - Javatpoint*, n.d.). Ο Abonizio τον χρησιμοποίησε για την αξιολόγηση του γραμμικού διαχωρισμού των δειγμάτων (Abonizio et al., 2020). Ο Ahmad υιοθετώντας τη συγκεκριμένη υποκατηγορία του SVM από έρευνα του Pérez-Rosas (Pérez-Rosas et al., 2017) επισημαίνει ότι ο γραμμικός SVM χρησιμοποιείται σε δεδομένα που δεν μπορούν εύκολα να διαχωριστούν ή σε δεδομένα που είναι πολυδιάστατα (Ahmad et al., 2020).



Εικόνα 62. Στιγμιότυπο Linear kernel SVM.

Οι Qian και Ying χρησιμοποίησαν και εκείνοι άλλοι μία υποκατηγορία του αλγόριθμου SVM τον *SVM-TS (time series) (LIBSVM -- A Library for Support Vector Machines, n.d.)*. Ο συγκεκριμένος αλγόριθμος βασίζεται στη χρονική ακολουθία των γεγονότων (π.χ. δημοσιεύσεις σε κοινωνικά δίκτυα ανά ώρα) και χρησιμοποιεί κοινούς κανόνες και τον γραμμικό ταξινομητή SVM για εντοπισμό των ψευδών ειδήσεων (S. Qian et al., 2021). Οι συγγραφείς υιοθέτησαν τον αλγόριθμο σε χαρακτηριστικά κοινωνικού περιεχομένου (S. Qian et al., 2021; Ying et al., 2021). Η χρονική ακολουθία των γεγονότων προσδιορίστηκε από χαρακτηριστικά χρηστών, διάχυσης και περιεχομένου σε συγκεκριμένες χρονικές στιγμές και γεγονότα (π.χ. διάρκεια χρονικής ακολουθίας σε σχέση με αριθμό λεπτών, ωρών, ημερών και τον δείκτη ώρας) (Ma et al., 2015; S. Qian et al., 2021; Ying et al., 2021).

Οι Vishwakarma και Reis χρησιμοποίησαν τον *SVM-RBF (Radial Basis Function)* μαζί με άλλους αλγόριθμους όπως τυχαία δάση, δέντρα απόφασης κ.α. για τον εντοπισμό των ψευδών ειδήσεων (Reis et al., 2019; Vishwakarma et al., 2019). Ο συγκεκριμένος αλγόριθμος χρησιμοποιείται σε δεδομένα με μη γραμμικό διαχωρισμό (Kumar, 2020) (Εικόνα 63).



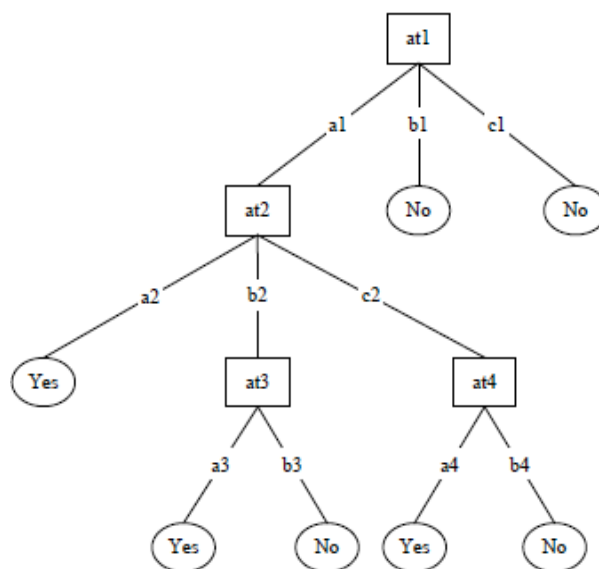
Εικόνα 63. Στιγμιότυπο από SVM-RBF.

Το δέντρο απόφασης (*Decision Tree*) (1.10. *Decision Trees*, n.d.) είναι ένας ακόμη αλγόριθμος που υιοθετήθηκε από αρκετούς συγγραφείς της βιβλιογραφίας εξαιτίας της πολύ καλής απόδοσής του (Abdelminaam et al., 2021; Alhazbi, 2020; Altunbey Ozbay & Alatas, 2019; Bonet-Jover et al., 2021; Braşoveanu & Andonie, 2021; Elhadad et al., 2020;

Goldani et al., 2021; Gravanis et al., 2019; Hunt et al., 2022; Isaakidou et al., 2021; Kaliyar et al., 2020; Kasseropoulos & Tjortjis, 2021; Y. Liu et al., 2019; Mahabub, 2020; Meel & Vishwakarma, 2020; Ozbay & Alatas, 2020; Shu, Wang, et al., 2019; Shu, Cui, et al., 2019; Silva et al., 2020; Sotirakou et al., 2021; Varshney & Vishwakarma, 2021; Y. Yu et al., 2020).

Σύμφωνα με τον Abdelminaam, το δέντρο απόφασης περιλαμβάνει δεσμούς και κλαδιά όπου οι δοκιμές για κάθε χαρακτηριστικό αναπαριστώνται ως δεσμοί, τα αποτελέσματα σαν κλαδιά και οι τάξεις σαν δεσμοί δέντρων. Στόχος είναι η κατηγοριοποίηση μίας στοχευμένης μεταβλητής μέσω εκμάθησης απλών κανόνων απόφασης (χαρακτηριστικά δεδομένων) (Abdelminaam et al., 2021). Ο Alhazbi προσθέτει ότι το κάθε μονοπάτι από τη ρίζα στους δεσμούς φύλλων περιλαμβάνει πολλαπλούς εσωτερικούς δεσμούς που αναπαριστούν τα χαρακτηριστικά για να παρθεί μία απόφαση (Alhazbi, 2020; Kotsiantis, 2007) (Εικόνα 64).

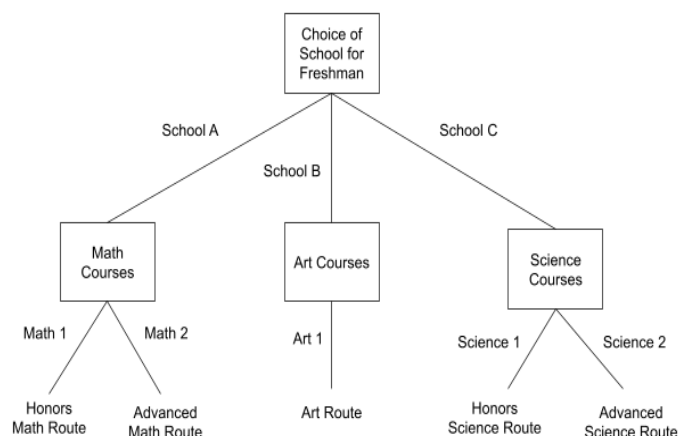
Ενώ οι Ozbay και Alatas επισημαίνουν ότι τα δέντρα απόφασης παίρνουν μία προκαθορισμένη στοχευμένη τιμή. Ως προς τη δομή έχουν στρατηγική από πάνω προς τα κάτω. Η συγκεκριμένη δομή χρησιμοποιείται για τον διαχωρισμό μεγάλων δεδομένων σε μικρές ομάδες εφαρμόζοντας κανόνες απόφασης (Ozbay & Alatas, 2020).



Εικόνα 64. Στιγμιότυπο από δέντρο απόφασης.

Το δέντρο απόφασης J-48 υιοθετήθηκε από τους Meel & Vishwakarma, Ozbay και Alatas (Meel & Vishwakarma, 2020; Ozbay & Alatas, 2020). Σύμφωνα με τους Ozbay και Alatas ο J-48 είναι ένας στατιστικός αλγόριθμος σε λογική δέντρου απόφασης (βάση αλγορίθμων

ID3 και C4,5)¹⁸. Λειτουργεί και αυτός με δεσμούς, με τη λογική χρήσης ενός δεσμού από κάθε δέντρο και σειρών διακλάδωσης (child row) από την πρώτη σειρά. Συγκαταλέγεται στους πιο γρήγορους και ακριβείς ταξινομικούς αλγόριθμους (Ozbay & Alatas, 2020) (Εικόνα 65).



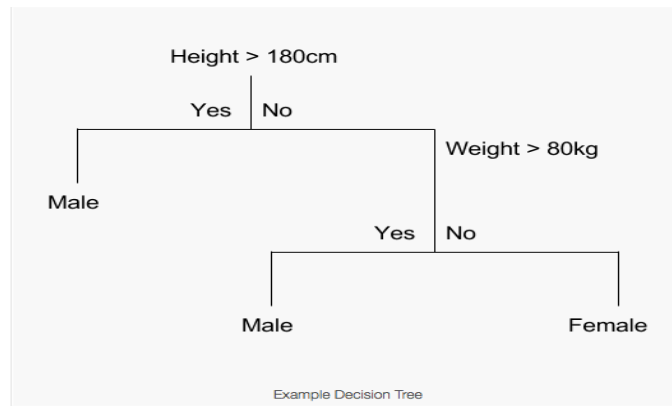
Εικόνα 65. Δέντρο απόφασης J-48.

Ο αλγόριθμος *CART* είναι μία ακόμη υποκατηγορία του δέντρου απόφασης. Ο συγκεκριμένος αλγόριθμος ταξινόμησης είναι υπό μορφή δέντρου δύο τιμών. Κάθε κλαδί αντιπροσωπεύει μια τιμή (x) και ένα σημείο διαμοιρασμού της τιμής ενώ τα φύλλα είναι η εξωτερική μεταβλητή (y) πρόβλεψης¹⁹. Ο *CART* διαφοροποιεί τα χαρακτηριστικά με τη μέθοδο τυχαίου υποχώρου επιλέγοντας εκείνα που είναι προβληματικά σε ένα δεδομένο μέγεθος. Ο *Ksieniewicz* υιοθέτησε το συγκεκριμένο αλγόριθμο με χαρακτηριστικά από λέξεις κλειδιά και περιεχομένου (*Ksieniewicz et al.*, 2019) (Εικόνα 66). Ενώ η *Isaakidou* χρησιμοποίησε τον αλγόριθμο για τη μέτρηση των ψευδών ειδήσεων σε άρθρα υιοθετώντας τις παραμέτρους (κριτήριο, best, random για διαχωρισμό δεδομένων, μέγιστο βάθος δέντρου²⁰) (*Isaakidou et al.*, 2021).

¹⁸ Δες και (*Khanna*, 2021)

¹⁹ Δες και (*Brownlee*, 2016).

²⁰ μέγιστο βάθος δέντρου: αριθμός δεσμών δέντρου που επεκτείνοντας τα θα μείνουν μόνο φύλλα καθαρά, ή όταν όλα τα φύλλα που περιέχουν να έχουν λιγότερα δείγματα.

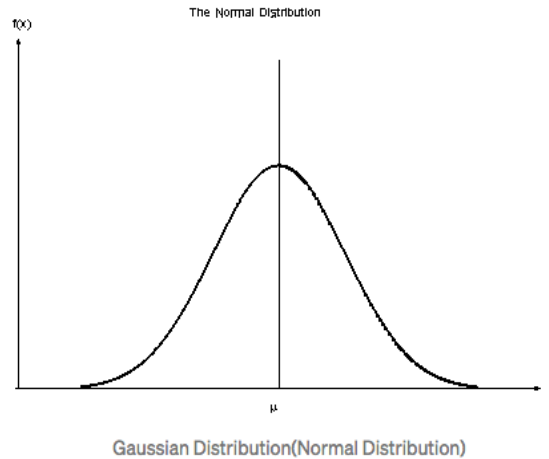


Εικόνα 66. Αλγόριθμος CART.

Ο *Naïve Bayes* είναι ένας απλός ταξινομητής, που βασίζεται στο θεώρημα Bayes, ο οποίος εντοπίζει τις ψευδείς ειδήσεις μέσω του κανόνα απόφασης Maximum Posterior. Το συγκεκριμένο θεώρημα υποθέτει την ανεξαρτησία των χαρακτηριστικών (Abdelminaam et al., 2021). Προϋποθέτει την ύπαρξη ενός φίλτρου για τον έλεγχο των ψευδών ειδήσεων.

Σύμφωνα με τον Mahabub ο συγκεκριμένος ταξινομητής ήταν από τους πρώτους που υιοθετήθηκαν για τον εντοπισμό των ψευδών ειδήσεων. Ο *Naïve Bayes* ελέγχει την πιθανότητα των λέξεων στο προπαρασκευαστικό σύνολο στις παρακείμενες ειδήσεις και με το αποτέλεσμα που έλαβε μπορεί να μετακινήσει τις ψευδείς ειδήσεις σε φάκελο (Mahabub, 2020). Αξίζει να επισημανθεί ότι αποδίδει καλύτερα με συναισθηματικά χαρακτηριστικά (Varshney & Vishwakarma, 2021)

Ο συγκεκριμένος αλγόριθμος ουσιαστικά αποτελεί έναν βασικό αλγόριθμο σε επίπεδο ακριβείας αλλά και μία από τις πρώτες προσπάθειες ταξινόμησης των ψευδών ειδήσεων γι' αυτό και έχει υιοθετηθεί και από τους συγγραφείς της παρούσας βιβλιογραφίας (Abdelminaam et al., 2021; Gravanis et al., 2019; Mahabub, 2020; Meel & Vishwakarma, 2020; Ozbay & Alatas, 2020; Reis et al., 2019; Sheikhi, 2021; Shu, Wang, et al., 2019; Shu, Cui, et al., 2019; Silva et al., 2020; Varshney & Vishwakarma, 2021; Y. Yu et al., 2020) (Εικόνα 67). Αποτελείται από 3 υπο-αλγόριθμους: τον Multinomial *Naïve Bayes*, τον Bernoulli *Naïve Bayes*, και τον Gaussian *Naïve Bayes* (Gandhi, 2018a).



Εικόνα 67. Στιγμιότυπο από Gaussian Naive Bayes.

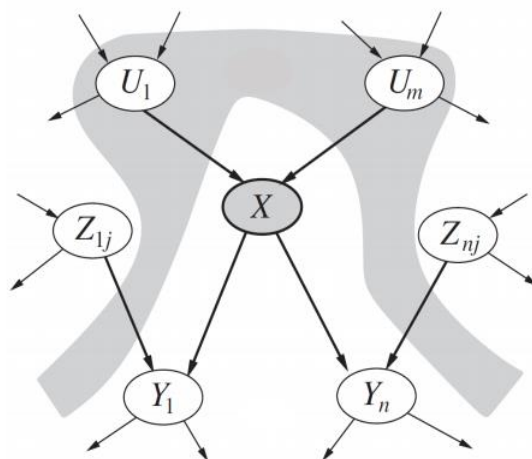
Ο *Multinomial Naïve Bayes* (1.9. *Naive Bayes*, n.d.-a) χρησιμοποιεί τον κλασικό αλγόριθμο Bayes για κατηγοριοποίηση του κειμένου μέσω υπολογισμού συχνοτήτων (Bonet-Jover et al., 2021; Braşoveanu & Andonie, 2021; Elhadad et al., 2020; Kaliyar et al., 2020; Khanday et al., 2021; Mahabub, 2020; Tundis et al., 2021). Σύμφωνα με τον Tundis ο πολυονιμικός Naïve Bayes χρησιμοποιεί την πολυονιμική κατανομή για κάθε χαρακτηριστικό και είναι ένας από τους δημοφιλείς αλγόριθμους για τον εντοπισμό των ψευδών ειδήσεων μαζί με τον SVM (Tundis et al., 2021). Ο υπολογισμός γίνεται με βάση τον αριθμό των δημοσιεύσεων (tweet) ανά ειδήσεις σε κάθε τάξη προς τον συνολικό αριθμό των δημοσιεύσεων ή ειδήσεων και στη συνέχεια υπολογίζεται ο αριθμός των λέξεων κάθε τάξης καθώς και οι λέξεις σε μία πρόταση (Khanday et al., 2021).

Ο *Bernoulli Naive Bayes* (*Sklearn.Naive_bayes.BernoulliNB*, n.d.) (Εικόνα 68) έχει ως βάση το θεώρημα Bernoulli με το οποίο μπορεί να υπάρχουν πολλαπλά χαρακτηριστικά με δυαδικές μεταβλητές (0, 1) (Boolean) (SHARMA, 2020). Ο συγκεκριμένος αλγόριθμος έχει καλύτερη απόδοση σε μικρότερα κείμενα (π.χ. συχνότητα λέξης σε ένα κείμενο) (1.9. *Naive Bayes*, n.d.-b). Οι Elhadad και Bonet-Jover υιοθέτησαν τον αλγόριθμο για τον εντοπισμό των ψευδών ειδήσεων στον Covid-19 και στα παραδοσιακά μέσα ενημέρωσης (Bonet-Jover et al., 2021; Elhadad et al., 2020).

$$P(x_i | y) = P(i | y)x_i + (1 - P(i | y))(1 - x_i)$$

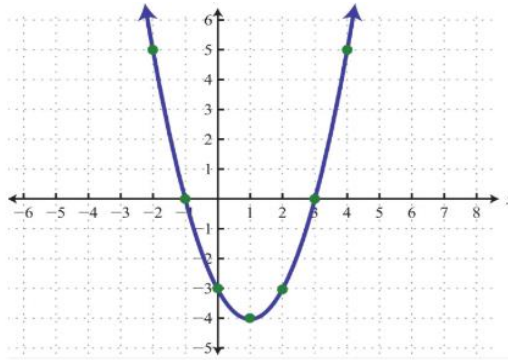
Εικόνα 68. Μαθηματικός τύπος Bernoulli Naive Bayes.

Το *δίκτυο Bayes* (Εικόνα 67) είναι ένας αλγόριθμος ταξινόμησης που χρησιμοποιεί διαφορετικές ερευνητικές μεθόδους και διάφορα ποιοτικά κριτήρια. Σύμφωνα με τους Ozbay και Alatas ο αλγόριθμος δικτύου Bayes είναι ένα πιθανοτικό μοντέλο γράφου με χρήση για την αναπαράσταση εξαρτώμενων συνθηκών ανάμεσα σε σύνολα από μεταβλητές (Langley et al., 1992; Ozbay & Alatas, 2020). Οι συγγραφείς υιοθέτησαν τον αλγόριθμο για τον εντοπισμό των ψευδών ειδήσεων στα κοινωνικά μέσα (Ozbay & Alatas, 2020).



Εικόνα 69. Δίκτυο Bayes.

Ο αλγόριθμος *Stochastic Gradient Descent (SGD)* (1.1. Linear Models, n.d.) είναι ένας αλγόριθμος ο οποίος χρησιμοποιεί την επαναληπτική μέθοδο για να βελτιστοποιήσει μία στοχευμένη λειτουργία. Χρησιμοποιεί τυχαία επιλεγμένα δείγματα για την αξιολόγηση των μεταβλητών (στοχαστικό μοντέλο) (Ozbay & Alatas, 2020). Ο συγκεκριμένος αλγόριθμος υιοθετήθηκε και από προηγούμενες έρευνες όπως του Ruder ο οποίος αναφέρει ότι ο αλγόριθμος είναι πιο γρήγορος, μπορεί να χρησιμοποιηθεί και για διαδικτυακή εκμάθηση και εκτελεί συχνές αναβαθμίσεις με υψηλή ποικιλομορφία (Ruder, 2017) (Εικόνα 70) (Bonet-Jover et al., 2021; Braşoveanu & Andonie, 2021; Goldani et al., 2021; Meel & Vishwakarma, 2020; Ozbay & Alatas, 2020). Χρησιμοποιείται όταν ο αριθμός των δειγμάτων είναι πολύ μεγάλος. Ουσιαστικά αποτελεί μία τεχνική βελτιστοποίησης η οποία χρησιμοποιείται για την εκπαίδευση ενός μοντέλου (δες και παράδειγμα). Αξίζει να επισημανθεί ότι ο συγκεκριμένος αλγόριθμος αποτελεί τη βάση των νευρωνικών δικτύων (Srinivasan, 2019).



Εικόνα 70. Αλγόριθμος Stochastic Gradient Descent (SGD).

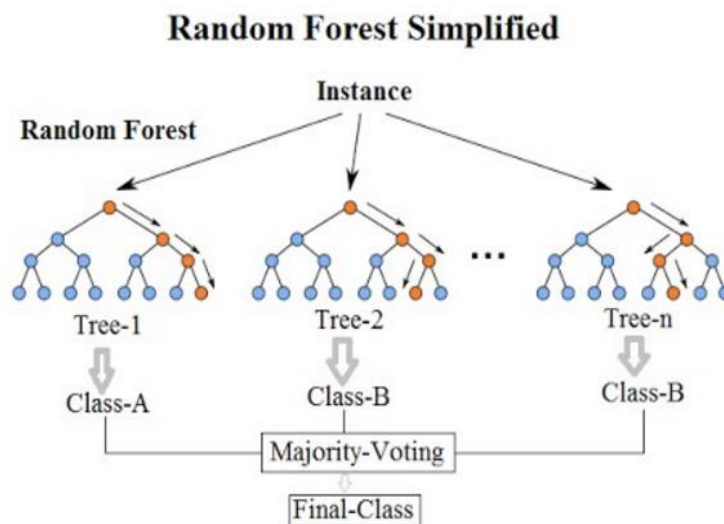
4.4.2 Αλγόριθμοι Πολλαπλής Εκμάθησης (Ensemble Learning)

Οι αλγόριθμοι πολλαπλής εκμάθησης αποτελούν συνδυασμό πολλών αδύναμων ταξινομητών με σκοπό τη δημιουργία ενός πιο δυνατού ταξινομητή με βελτιωμένη απόδοση. Σύμφωνα με τον Gravanis οι αλγόριθμοι επιβλεπόμενης μάθησης αποτελούν ένα συνδυασμό από αδύναμους ταξινομικούς αλγόριθμους οι οποίοι αν συνδυαστούν δίνουν ένα δυνατό μοντέλο με περισσότερη ακρίβεια στην απόδοση (Gravanis et al., 2019; Rokach, 2010).

Το δάσος τυχαίας απόφασης (Random Forest) είναι ένας πιο εξελιγμένος αλγόριθμος των δέντρων απόφασης. Περιλαμβάνει ένα μεγάλο αριθμό δέντρων που δουλεύουν αυτόνομα για να βρουν το αποτέλεσμα μίας τάξης με τη μέθοδο της υψηλής πλειοψηφίας (Ahmad et al., 2020). Το δάσος τυχαίας απόφασης υιοθετήθηκε από αρκετούς συγγραφείς της τρέχουσας βιβλιογραφίας (Abdelminaam et al., 2021; Abonizio et al., 2020; Ahmad et al., 2020; Alhazbi, 2020; Braşoveanu & Andonie, 2021; Cremisini et al., 2019; B. D. Horne et al., 2020; Hunt et al., 2022; Kaliyar et al., 2020; Kasseropoulos & Tjortjis, 2021; Y. Liu et al., 2019; Mahabub, 2020; Meel & Vishwakarma, 2020; Posadas-Durán et al., 2019; Reis et al., 2019; Sheikhi, 2021; Shu, Wang, et al., 2019; Shu, Cui, et al., 2019; Silva et al., 2020; Sotirakou et al., 2021; Souza Freire et al., 2021; Taskin et al., 2022; Varshney & Vishwakarma, 2021).

Ο συγκεκριμένος αλγόριθμος παρέχει ευελιξία, και περισσότερη ακρίβεια σε σχέση με τα απλά δέντρα απόφασης. Στόχος του αλγόριθμου είναι η κατηγοριοποίηση μιας στοχευμένης μεταβλητής μέσω εκμάθησης απλών κανόνων απόφασης (χαρακτηριστικά δεδομένων) από περισσότερα του ενός δέντρου (Abdelminaam et al., 2021). Ο Alhazbi αναφέρει ότι το δάσος τυχαίας απόφασης είναι μία επέκταση της τεχνικής bagging που όμως διαφέρει στο ότι χρησιμοποιεί τυχαία υποσύνολα χαρακτηριστικών για κάθε αυτόνομο δέντρο προκειμένου να μειώσει τη συσχέτιση ανάμεσα στα χαρακτηριστικά (Alhazbi, 2020; Breiman, 2001).

Τα δάση τυχαίας απόφασης δημιουργούν διαφορετικά δέντρα επιλέγοντας διαφορετικά υποσύνολα. Στη συνέχεια τα δέντρα απόφασης κάνουν μία εκτίμηση και στην κατηγοριοποίηση επιλέγουν την τιμή με την υψηλότερη ψήφο με βάση τις εκτιμήσεις (Taskin et al., 2022). Στην έρευνα της Sotirakou το δάσος τυχαίας απόφασης αποτέλεσε τον αλγόριθμο με τη μέγιστη ακρίβεια στα αποτελέσματα²¹ (Sotirakou et al., 2021) (Εικόνα 71).



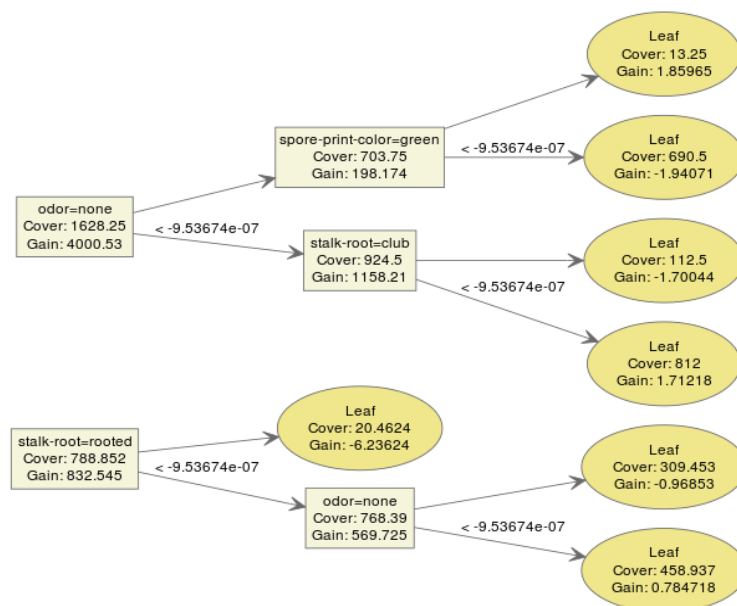
Εικόνα 71. Δάσος Τυχαίας Απόφασης.

Ο αλγόριθμος *Bagging* (ή *bootstrap aggregating*) (1.11. *Ensemble Methods*, n.d.-a) είναι ένας μετα-αλγόριθμος στατιστικής ταξινόμησης για τη βελτίωση της ακρίβειας των αλγορίθμων (Ahmad et al., 2020; Gravanis et al., 2019; Ozbay & Alatas, 2020; Silva et al., 2020). Οι Ozbay και Alatas αναφέρουν ότι ο συγκεκριμένος αλγόριθμος προτάθηκε από τον Breiman. Ο Breiman στην μελέτη του διατύπωσε ότι ο αλγόριθμος Bagging αποτελείται από πολλαπλές μεθόδους ενός αλγορίθμου πρόβλεψης, οι οποίες χρησιμοποιούνται για να δώσουν μία συνολική πρόβλεψη (Breiman, 1996). Οι Ozbay και Alatas επισημαίνουν ότι κατά τη διάρκεια της εκπαίδευσης πραγματοποιείται αντικατάσταση του δείγματος. Τα δεδομένα εκπαίδευσης αντικαθίστανται από n δείγματα (Bagging). Κάθε επιλεγμένο δείγμα επιστρέφεται στα δεδομένα εκπαίδευσης. Με βάση την τυχαιότητα των εκπαιδευμένων ομάδων οι βασικοί αλγόριθμοι εκπαίδευσης εκπαιδεύονται για να υπάρξει ανομοιομορφία. Κατά αυτόν τρόπο επιτυγχάνεται η συλλογική επιτυχία (Ozbay & Alatas, 2020).

²¹ Δες και (Yamini, 2021).

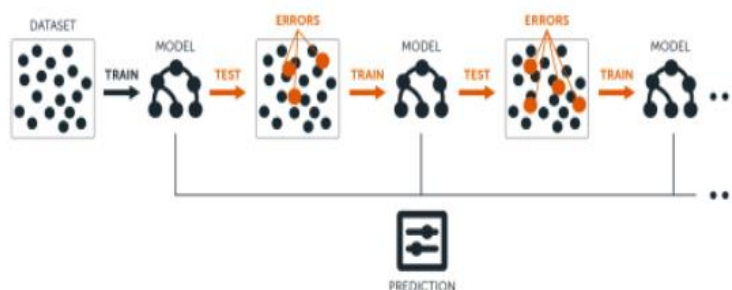
Ο Ahmad επιπροσθέτως, αναφέρει ότι ο συγκεκριμένος αλγόριθμος επιλέγει μία κατηγορία δεδομένων με βάση την πλειοψηφική μέθοδο βασισμένη σε αριθμό από δέντρα M τα οποία μειώνουν την διαφορετικότητα και τα δεδομένα κάθε δέντρου επιλέγονται με τη χρήση τυχαίας δειγματοληψίας, αντικαθιστώντας όλη τη βάση δεδομένων (Ahmad et al., 2020). Ο Gravanis μάλιστα αναφέρει ότι ο αλγόριθμος Bagging (ή μέθοδος Bagging) είναι από τους πλέον γνωστούς αλγόριθμους και εκείνος με την πιο μεγάλη απόδοση (Gravanis et al., 2019).

Ο *XGBoost* είναι ένας αλγόριθμος που βασίζεται σε δέντρα απόφασης και χρησιμοποιεί ένα διαβαθμισμένο πλαίσιο ενίσχυσης (Gradient Boosted Trees) (*An Introduction to XGBoost R Package*, n.d.; *Extreme Gradient Boosting*, 2014/2022). Η ενίσχυση βελτιώνει την απόδοση ταξινόμησης και μειώνει τα σφάλματα. Ουσιαστικά είναι ένας συνδυασμός πολλών αδύναμων μοντέλων μάθησης για τη δημιουργία ενός πιο ανθεκτικού μοντέλου μάθησης (Abonizio et al., 2020; Ahmad et al., 2020; Elhadad et al., 2020; Hunt et al., 2022; Reis et al., 2019; Sheikhi, 2021; Souza Freire et al., 2021). Ο Sheikhi τον χρησιμοποίησε για τον εντοπισμό των ψευδών ειδήσεων συνδυαστικά με τον αλγόριθμο βελτιστοποίησης whale optimization (Sheikhi, 2021). Ο Abonizio επισημαίνει την επιτυχία αλλά και την ευρεία υιοθέτηση του αλγόριθμου σε προηγούμενες έρευνες (H. Chen et al., 2016). Ενώ ο Souza Freire τον χρησιμοποίησε ως έναν από τους αλγόριθμους του πληθοποριστικού μοντέλου για τον εντοπισμό των ψευδών ειδήσεων στα κοινωνικά δίκτυα (π.χ. μετρηση δημοσιεύσεων tweet) (Souza Freire et al., 2021) (Εικόνα 72).



Εικόνα 72. Αλγόριθμος XGBoost.

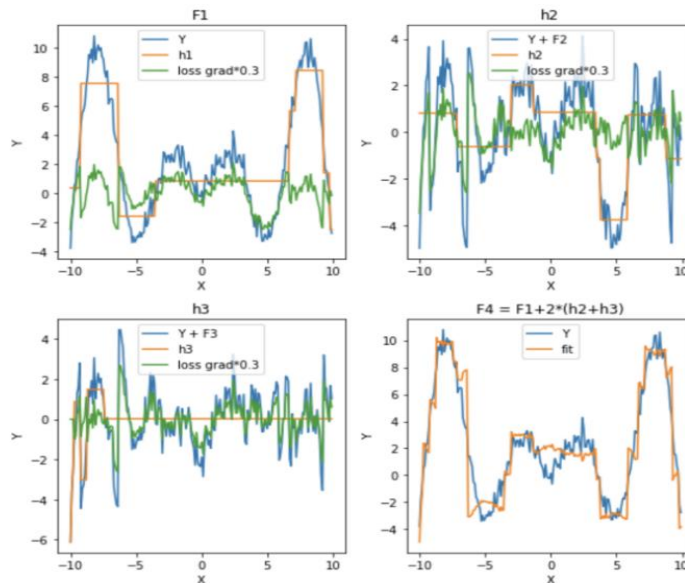
Ο *Gradient Boosting* είναι ένας αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται στην ταξινόμηση και την παλινδρόμηση με ταχύτητα στην πρόβλεψη και ακρίβεια ο οποίος υιοθετήθηκε από τους συγγραφείς της παρούσας βιβλιογραφίας (Alhazbi, 2020; Elhadad et al., 2020; Kasseropoulos & Tjortjis, 2021; Mahabub, 2020; Ozbay & Alatas, 2020). Ο αλγόριθμος προσθέτει μοντέλα σε κάθε συχνότητα, ξεκινώντας από ένα αδύναμο μοντέλο και ενισχύοντας (boosting) διαδοχικά την απόδοση, δημιουργώντας νέα βελτιωμένα δέντρα από τα λάθη των προηγούμενων (1.11. *Ensemble Methods*, n.d.-b; Greenwell, n.d.) (Εικόνα 72). Η βελτιστοποίηση στην απόδοση του μοντέλου πραγματοποιείται με τη χρήση του αλγόριθμου gradient descent (Alhazbi, 2020; Friedman, 2002).



Εικόνα 73. Παράδειγμα λειτουργίας Gradient Boosting.

Ο αλγόριθμος *boosting* (*Sklearn.Ensemble.GradientBoostingClassifier*, n.d.) εκπαιδεύει αδύναμα μοντέλα σε δυνατά μέσω της πλειοψηφικής μεθόδου και σύμφωνα με τους συγγραφείς έχει αποδώσει πολύ καλά αποτελέσματα ανάλογα των κλασικών αλγορίθμων όπως η μηχανή διανυσματικής υποστήριξης (SVM) και το δάσος τυχαίας απόφασης (Random Forest) (Ahmad et al., 2020; Posadas-Durán et al., 2019). Στη συγκεκριμένη μέθοδο οι αδύναμοι ταξινομητές κατηγοριοποιούν τα σημεία δεδομένων σε ακολουθίες. Κάθε δέντρο ακολουθίας που δημιουργείται σε κάθε κύκλο μαθαίνει να μειώνει τα λάθη από τον προηγούμενο κύκλο βελτιώνοντας την ακρίβεια σε σημεία δεδομένων που ήταν μη σωστά κατηγοριοποιημένα σε προηγούμενους κύκλους (Ahmad et al., 2020; Dos Santos et al., 2009) (Εικόνα 74)²².

²² Δες και (Z. Zhang, 2019)



Εικόνα 74. Αλγόριθμος Gradient Boosting.

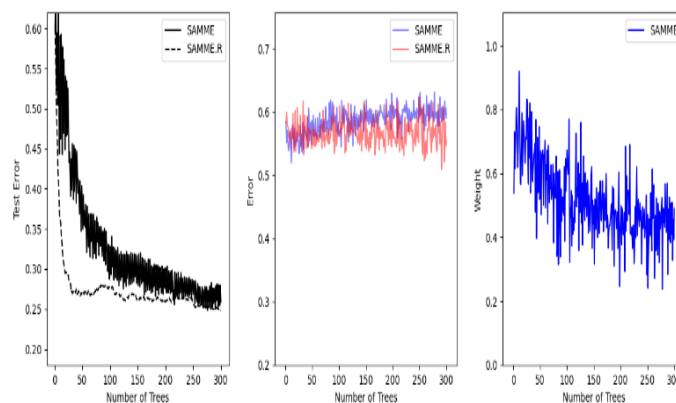
Ο αλγόριθμος *Adaboost* (Multi-Class AdaBoosted Decision Trees, n.d.) είναι ένας μετα-αλγόριθμος στατιστικής ταξινόμησης όπου και εδώ συνδυάζονται πολλοί αδύναμοι ταξινομητές για να δημιουργηθεί ένας ταξινομητής (π.χ. δέντρο) (Ahmad et al., 2020; Alhazbi, 2020; Gravanis et al., 2019; Hunt et al., 2022; Mahabub, 2020; Silva et al., 2020). Σύμφωνα με τον Gravanis ο αλγόριθμος *adaboost* επικεντρώνεται σε μοτίβα που δύσκολα κατηγοριοποιούνται ενώ η μέθοδος *bagging* είναι εκείνη με την καλύτερη απόδοση (Gravanis et al., 2019).

Ενώ ο Mahabub αναφέρει ότι ο *Adaboost* αποτελεί έναν μετα-αλγόριθμο πρόβλεψης ο οποίος μετά την εισαγωγή των δεδομένων σε έναν ταξινομητή προσθέτει επιπλέον αντίγραφα ταξινομητών σε αντίστοιχη βάση δεδομένων όπου υπάρχουν παραδείγματα ομαδοποιημένα με στόχο ότι οι επόμενοι αλγόριθμοι θα εστιαστούν σε προβληματικές περιπτώσεις (Hastie et al., 2009; Mahabub, 2020) (Εικόνα 75).

Επιπροσθέτως, ο Alhazbi αναφέρει ότι στον *adaboost* συνδυάζονται μία σειρά από αδύναμους ταξινομητές όπου αρχικά σε όλα τα χαρακτηριστικά των δεδομένων επιλέγονται ισόποσα βάρη και αφού γίνει εκπαίδευση ενός αδύναμου ταξινομητή, τα μη σωστά κατηγοριοποιημένα χαρακτηριστικά παίρνουν υψηλότερα βάρη για να μπορούν να είναι ορατά και να τα επιλέξει για εκπαίδευση ο επόμενος αλγόριθμος (πχ. βάρος ανάλογο τιμής λάθους όπως ποσοστό μη κατηγοριοποιημένων χαρακτηριστικών).

Η διαφορά του *adaboost* και του *gradient boosting* είναι ο τρόπος που ο καθένας από τους δύο αντιμετωπίζει τις αδυναμίες από τους προηγούμενους αλγόριθμους δηλαδή ο *adaboost* επιλέγει ένα υψηλό βάρος για κάθε μη σωστά κατηγοριοποιημένο χαρακτηριστικό

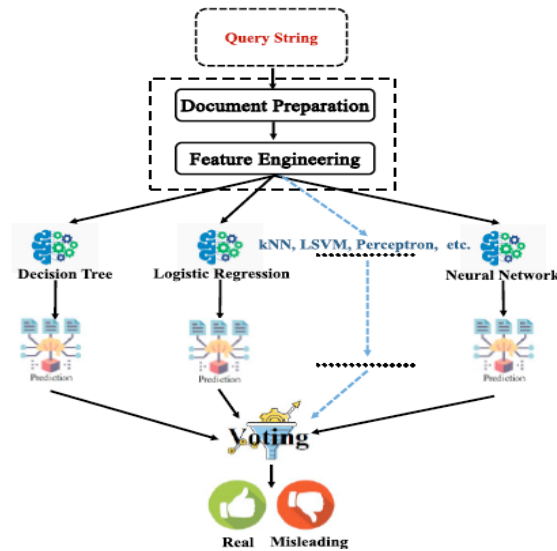
των δεδομένων εκπαίδευσης ενώ ο Gradient boosting τον gradient descent όπως ειπώθηκε και παραπάνω (Alhazbi, 2020).



Εικόνα 75. Στιγμιότυπο από Adaboost.

Ο αλγόριθμος *ensemble voting* (1.11. Ensemble Methods, n.d.-a) είναι ένας αλγόριθμος ταξινόμησης που επιτρέπει τον συνδυασμό δύο ή περισσότερων μοντέλων εκμάθησης. Κάθε μοντέλο προβλέπει ένα αποτέλεσμα (ψήφος) υπέρ μιας τάξης που το μοντέλο έχει προβλέψει. Το τελικό αποτέλεσμα βασίζεται στην πλειοψηφική ψήφο για μια συγκεκριμένη τάξη (Ahmad et al., 2020; Elhadad et al., 2020; Ksieniewicz et al., 2019; Mahabub, 2020). Ο Elhadad χρησιμοποίησε τη συγκεκριμένη μέθοδο για τον εντοπισμό των ψευδών ειδήσεων για τον Covid19, που αποτελείται από αλγόριθμους (kNN, Decision Tree, Neural Network) οι οποίοι εκπαιδεύονται παράλληλα και αποτελούν το μοντέλο πρόβλεψης. Στη συνέχεια στο μοντέλο εισάγονται σειρές ερωτημάτων. Η κατηγοριοποίηση γίνεται με δύο ή τρεις αλγόριθμους οι οποίοι προβλέπουν την τάξη ως Πραγματική (Real) ή Παραπληροφόρηση (Misleading) (Elhadad et al., 2020) (Εικόνα 76). Σύμφωνα με τον Ahmad η συγκεκριμένη μέθοδος είναι πιο απλή ως προς την εφαρμογή και παράγει αποτελέσματα τα οποία συνεισφέρουν στην πρόβλεψη μέσω της πλειοψηφικής ψήφου (majority voting) (Ahmad et al., 2020).

Η *πλειοψηφική ψήφος (Majority voting)* (1.11. Ensemble Methods, n.d.-a) δημιουργεί ένα ανταγωνιστικό σύστημα ταξινόμησης για μοντέλα βαθιάς μάθησης. Η επιλογή της τάξης βασίζεται στην απόφαση των ταξινομητών (π.χ. δέντρα απόφασης). Η απόφαση των ταξινομητών υιοθετεί δύο κανόνες τον κανόνα που βασίζεται στην ψήφο και τον κανόνα της ενσωμάτωσης όπου εμπλουτίζεται με τα βάρη (τιμές χαρακτηριστικών με βάση τη σημασία) (Ksieniewicz et al., 2019).



Εικόνα 76. Ensemble Voting.

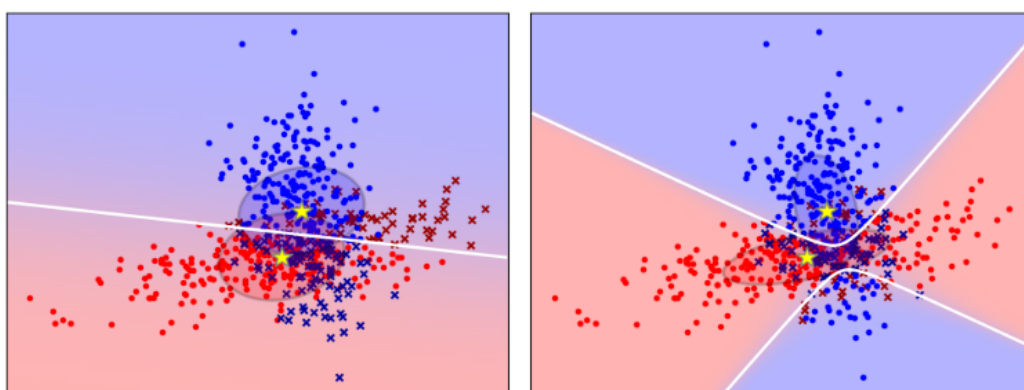
4.4.3 Αλγόριθμοι Μη Επιβλεπόμενης Μηχανικής Μάθησης (Unsupervised Learning)

Σύμφωνα με τον Taskin οι μη επιβλεπόμενοι αλγόριθμοι μηχανικής μάθησης εκπαιδεύονται με μη κατηγοριοποιημένα δεδομένα ως δεδομένα εισόδου και προσδιορίζουν σε πόσες ομάδες ο αλγόριθμος μπορεί να κατανείμει τα δεδομένα. Οι αλγόριθμοι μαθαίνοντας δομές και σχέσεις από τα δεδομένα, τα διαχωρίζουν σε ομάδες και ο αριθμός των ομάδων μπαίνει στον αλγόριθμο (Goodfellow et al., 2017; Taskin et al., 2022).

Η Γραμμική ανάλυση διάκρισης (*Linear Discriminant Analysis*) (1.2. *Linear and Quadratic Discriminant Analysis*, n.d.) προσδιορίζει ένα μέγιστο όριο (hyperplane) προκειμένου να ομαδοποιήσει τα δεδομένα με βάση την ομοιότητά τους (Taskin et al., 2022). Ο Taskin χρησιμοποίησε τον αλγόριθμο για τον εντοπισμό των ψευδών ειδήσεων στις τούρκικες ειδήσεις ενώ η Cremisini για τον εντοπισμό της παραπληροφόρησης για την Ουκρανική κρίση (2014-2015) (Cremisini et al., 2019; Taskin et al., 2022). Για την κατηγοριοποίηση των δεδομένων χρησιμοποιούνται δύο κανόνες: ο μέγιστος κανόνας πιθανότητας οποίος υποθέτει ότι εάν κάθε τάξη μπορεί να εμφανιστεί με ίση πιθανότητα τότε μπορεί να βρεθεί η πιθανότητα (x) μιας τάξης j (μέσω του τύπου $j = \text{argmax}_i f_i(x)$) και ο κανόνας Bayes που υποθέτει ότι εάν οι πιθανότητες των προηγούμενων τάξεων είναι γνωστές (π) τότε μπορεί να εντοπιστεί η πιθανότητα x σε μια τάξη j (μέσω του τύπου $j = \text{argmax}_i \pi_i f_i(x)$) (Xiaozhou, 2022).

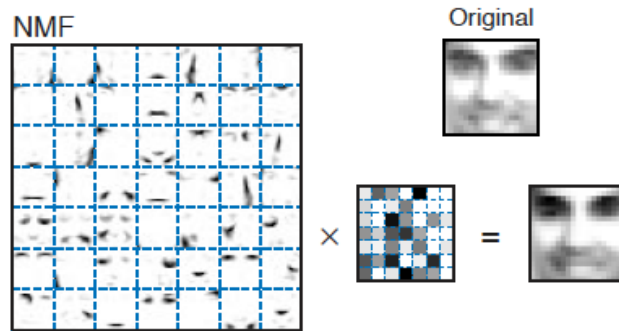
Η *ανάλυση τετραγωνικής διακρίσεως δεδομένων (quadratic discriminant analysis, QDA)* (Cremisini et al., 2019) είναι ένας άλλος τύπος μη γραμμικής ανάλυσης που επιτρέπει το μη γραμμικό διαχωρισμό των δεδομένων και χρησιμοποιείται σε περιπτώσεις όπου υπάρχει εκ των προτέρων γνώση ότι στις ανεξάρτητες τάξεις υπάρχουν διαφοροποιήσεις (*Linear, Quadratic, and Regularized Discriminant Analysis*, 2018) .

Οι δύο αλγόριθμοι που αναφέρθηκαν παραπάνω διαφέρουν στο ότι η LDA μαθαίνει δεδομένα μόνο με γραμμικό τρόπο (γραμμικό όριο) ενώ η QDA με τετραπλό τρόπο (τετραπλό όριο) (Εικόνα 77). Επιπλέον δεν υπάρχουν παράμετροι και μπορούν να υπολογιστούν εύκολα (1.2. *Linear and Quadratic Discriminant Analysis*, n.d.) .



Εικόνα 77. Γραμμική ανάλυση διάκρισης (LDA) και γραμμική ανάλυση τετραγωνικής διακρίσεως (QDA).

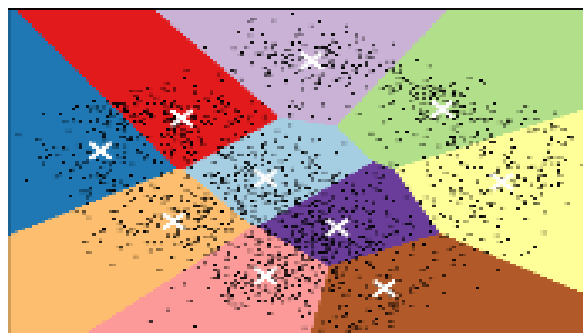
Στη *μη αρνητική παραγοντοποίηση μήτρας (NMF)* (2.5. *Decomposing Signals in Components (Matrix Factorization Problems)*, n.d.) γίνεται εκμάθηση αναπαραστάσεων των μερών των προσώπων και των σημασιολογικών χαρακτηριστικών του κειμένου. Χρησιμοποιήθηκε από τους Shu και Taskin παράλληλα με αλγόριθμους όπως η γραμμική ανάλυση δεδομένων (LDA) και άλλους παραδοσιακούς αλγόριθμους για τον εντοπισμό των ψευδών ειδήσεων. Σύμφωνα με το Taskin σκοπός του NMF είναι να προσεγγίσει ένα μη αρνητικό πίνακα διαχωρίζοντάς τον σε δύο υποπίνακες (με μη αρνητικές τιμές) (Εικόνα 78) (Lee & Seung, 1999; Shu, Wang, et al., 2019; Taskin et al., 2022).



Εικόνα 78. Μη αρνητική παραγοντοποίηση μήτρας (NMF).

Ο *K-means* (Εικόνα 79) (2.3. *Clustering*, n.d.) είναι ένας αλγόριθμος ο οποίος παίρνει σαν υπερπαραμέτρο των αριθμό των εξωτερικών ομάδων (συστάδων). Ο αλγόριθμος επιλέγει ένα τυχαίο κεντρικό σημείο για κάθε δεδομένο στοιχείο των ομάδων και τα στοιχεία που είναι κοντά στο κεντρικό σημείο συμπεριλαμβάνονται στις ομάδες (clusters). Ο Taskin χρησιμοποίησε τον αλγόριθμο και προσδιόρισε τον αριθμό των ομάδων (παραμέτρος) σε δύο για την αναπαράσταση των ψευδών και μη ψευδών κατηγοριών (Taskin et al., 2022).

Ο συγκεκριμένος αλγόριθμος χρησιμοποιείται για την ομαδοποίηση των κειμένων και αποτελείται από τον αριθμό των ομάδων (clusters) K , την αρχική κεντροειδή ομάδα (initial cluster centroids), τη διαίρεση κάθε κειμένου σε ομόκεντρες ομάδες (partitioning of each document to similar centroids), και τον βαθμό ομοιότητας (similarity score). Ο αλγόριθμος διαχωρίζει τη συλλογή των κειμένων σε υποσύνολα ομάδων και χρησιμοποιεί το βαθμό ομοιότητας για να προσδιορίσει για κάθε έγγραφο ένα κεντρικό σημείο (βάση πίνακα δεδομένων). Η αναπαράσταση των κειμένων γίνεται υπό μορφή διανυσμάτων με βάση την βαρύνουσα σημασία κάθε όρου (Abualigah et al., 2018).



Εικόνα 79. Αλγόριθμος K means (οπτικοποίηση).

Ο αλγόριθμος ομοιότητας (similarity algorithm) υιοθετήθηκε από τον Yuliani για τον εντοπισμό της εξαπάτησης στις ειδήσεις (hoax). Ο συγκεκριμένος αλγόριθμος

κατηγοριοποιεί το κείμενο με βάση την ομοιότητα των μοτίβων του (συμβολοσειρών) (Εικόνα 80). Ο αλγόριθμος έχει τρεις υπομεθόδους, την περιγραφή, την εύρεση παραμέτρων και την παραγωγή αποτελεσμάτων από την ομοιότητα των δεδομένων λέξεων hoax. Τα αποτελέσματα της ομοιότητας παράγονται σε ποσοστό και υπολογίζεται ο χρόνος εντοπισμού (δευτερόλεπτα) των hoax (Vuković et al., 2009; Yuliani et al., 2019).

Example #1 similar_text() argument swapping example

```
<?php
```

```
$sim = similar_text('bafoobar', 'barfoo', $perc);
```

```
echo "similarity: $sim ($perc %)\n";
```

```
$sim = similar_text('barfoo', 'bafoobar', $perc);
```

```
echo "similarity: $sim ($perc %)\n";
```

Εικόνα 80. Στιγμιότυπο από αλγόριθμο ομοιότητας (similarity algorithm).

Ο ίδιος συγγραφέας υιοθέτησε και άλλους τρεις αλγορίθμους τους *Levenshstein algorithm*, *Damerau-Levenshtein*, και *Smith waterman (database search algorithm)* για την επιλογή των λέξεων hoax. Ο αλγόριθμος *Levenshstein* μετρά τις διαφορές μεταξύ δύο συχνοτήτων (λέξεων) (ελάχιστος αριθμός αλλαγής μίας λέξης με μία άλλη). Για παράδειγμα η απόσταση μεταξύ της λέξης gily και geely είναι 2 (“The Levenshtein Algorithm,” 2017) (Εικόνα 81).

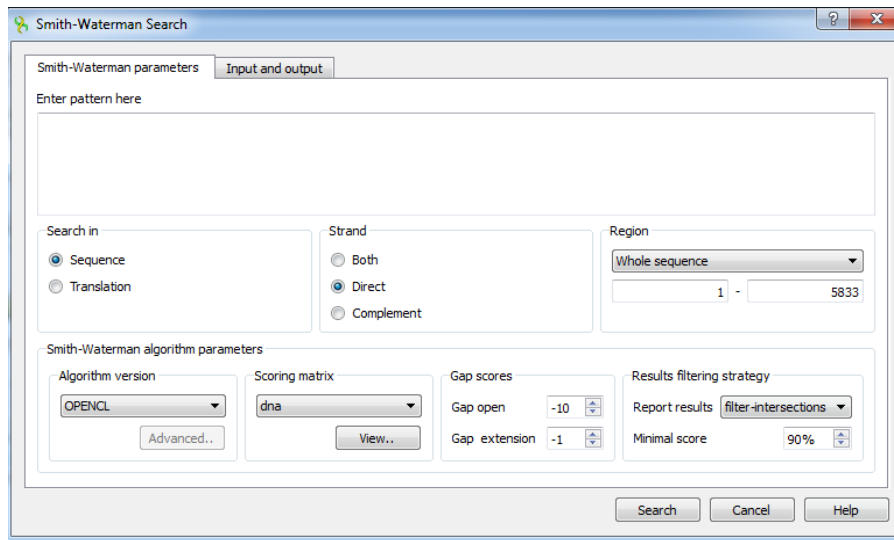
G	I		L	Y
G	E	E	L	Y

G		I	L	Y
G	E	E	L	Y

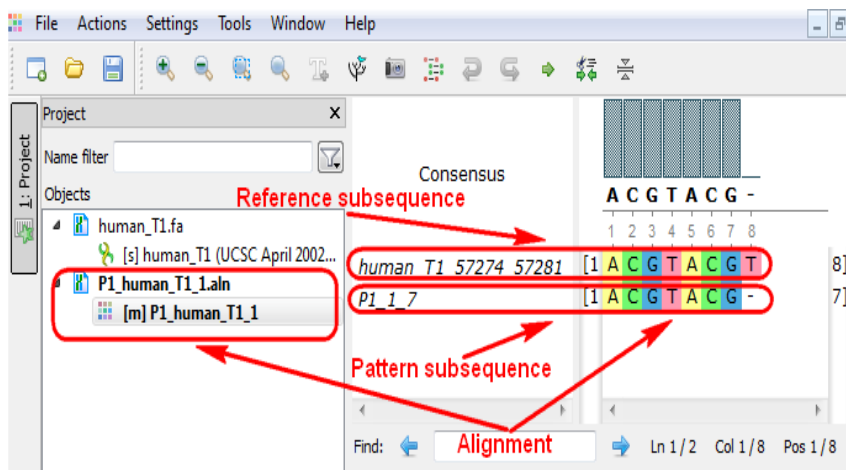
Εικόνα 81. Παράδειγμα διαφορών λέξεων.

Ο αλγόριθμος *Damerau-Levenshtein* (ή απόσταση *Damerau-Levenshtein*) παίρνει ως δεδομένο ότι η απόσταση ανάμεσα σε 2 λέξεις σχετίζεται με τον ελάχιστο αριθμό εργασιών που πρέπει να πραγματοποιηθούν (εισαγωγή, διαγραφή, αποκατάσταση) (προστίθενται δηλαδή υποκατάσταση από διπλανά σύμβολα). Ανάλογο παράδειγμα είναι η μετατροπή συμβολοσειράς ASCII σε UTF-8.

Ο *Smith waterman* (*Smith-Waterman Search - Unipro UGENE Online User Manual v. 1.15.0 - WIKI, n.d.*) είναι ο αλγόριθμος που χρησιμοποιείται για την εύρεση της υψηλότερης βαθμολογίας με βάση έναν παραγόμενο πίνακα. Ο συγκεκριμένος αλγόριθμος δημιουργήθηκε από τον Waterman για την μέτρηση της απόστασης ανάμεσα σε δύο συχνότητες βιολογικές, όπως η συχνότητα αμινοξέως, και περιλαμβάνει πολλαπλές διαγραφές, εισαγωγές και μεμονωμένες τροποποιήσεις, αλλά και υπολογισμό για περισσότερες από δύο συχνότητες²³ (Εικόνα 82, Εικόνα 83) (Yuliani et al., 2019).



Εικόνα 82. Αλγόριθμος Smith-Waterman.



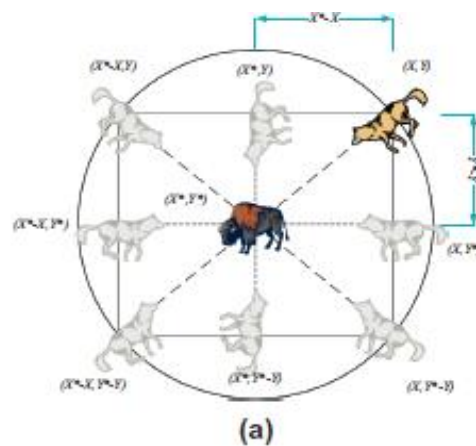
Εικόνα 83. Παράδειγμα μοτίβων συχνότητας Smith Waterman.

²³ Δες και (Waterman et al., 1976)

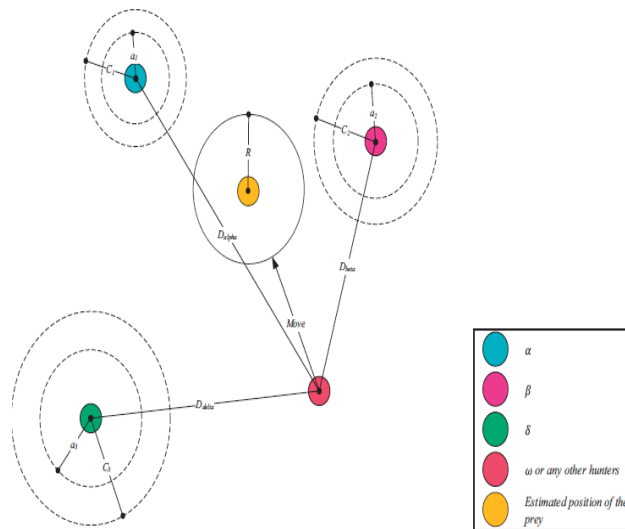
4.4.4 Μεταευριστικοί Αλγόριθμοι-Εξελικτικοί (Metaheuristics)

Οι μετα-ευριστικοί αλγόριθμοι (ή εξελικτικοί) έχουν επηρεαστεί από τη φύση και από το ζωικό βασίλειο και ο κύριος στόχος τους είναι να δώσουν λύση σε προβλήματα βελτιστοποίησης τα οποία βασίζονται σε ελλιπή δεδομένα (INFORMS, n.d.). Επιπλέον, έχουν καλύτερα αποτελέσματα σε σχέση με τους παραδοσιακούς αλγόριθμους (ChandraPandey et al., 2018; A. C. Pandey & Rajroot, 2019).

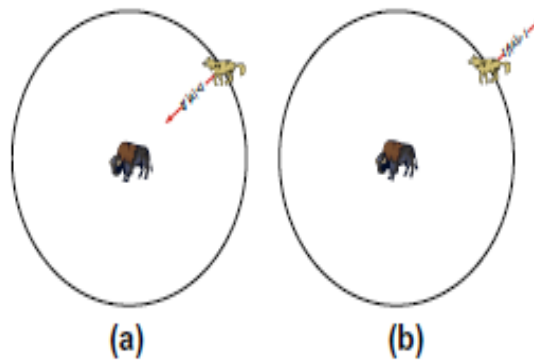
Ο αλγόριθμος *Grey Wolf optimization* είναι ένας αλγόριθμος βασιζόμενος στον ζωικό πληθυσμό (Mirjalili, n.d.-a). Έχει επηρεαστεί από τις κυνηγετικές αξίες και τις κοινωνικές σχέσεις των γκρίζων λύκων (Εικόνα 84, Εικόνα 85). Ο αλγόριθμος διαχωρίζει τις υποψήφιες λύσεις σε 4 ομάδες (α , β , δ , ω) και επιλέγει τους 3 καλύτερους λύκους. Μετέπειτα γίνεται η κατηγοριοποίηση των 3 καλύτερων (α , β , δ) (Εικόνα 86). Οι υπόλοιποι λύκοι (ω) ακολουθούν τους α , β , δ για την εύρεση βέλτιστων λύσεων. Η διαδικασία τερματίζεται έως ότου βρεθεί η καλύτερη τιμή α (Mirjalili et al., 2014; Ozbay & Alatas, 2020). Ο αλγόριθμος υιοθετήθηκε από τους συγγραφείς παίρνοντας ως βάση ότι ο εντοπισμός των ψευδών ειδήσεων είναι θέμα βελτιστοποίησης (Ozbay & Alatas, 2020; A. C. Pandey & Tikkiwal, 2021; Sheikhi, 2021).



Εικόνα 84. Αναβάθμιση θέσης γκρίζου λύκου βάση θέσης θηράματος (Grey wolf optim. algorithm).



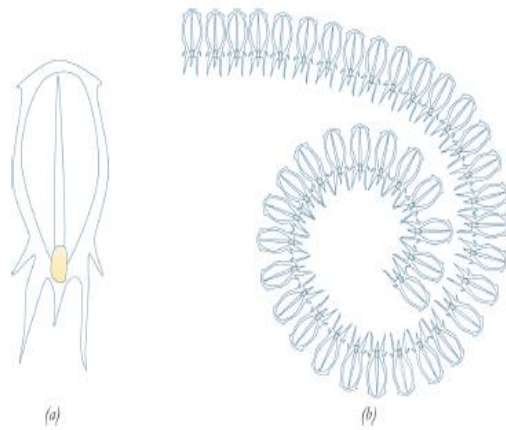
Εικόνα 85. Αναβάθμιση θέσης πράκτορα βάση υποψήφιων λύσεων α , β , δ .



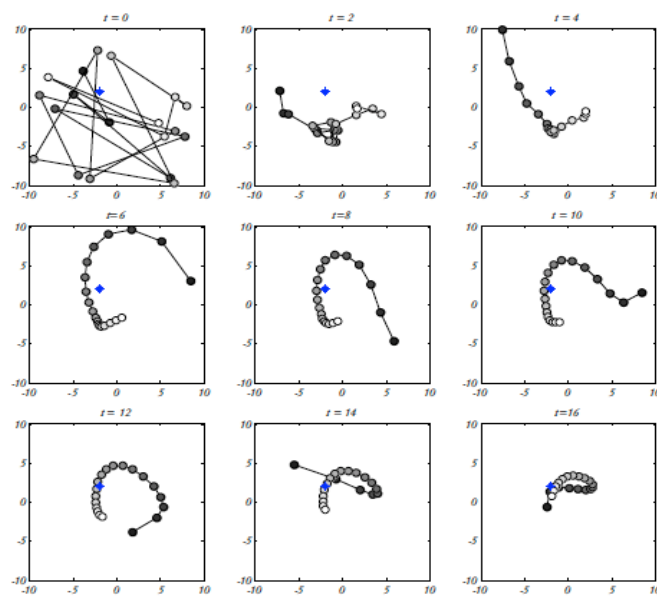
Εικόνα 86. Επίθεση και εύρεση θηράματος από αλγόριθμο.

Ο αλγόριθμος *Salp swarm optimization* (Mirjalili, n.d.-c) (Εικόνα 87) είναι ένα αλγόριθμος όπου βασίζεται στον θαλάσσιο πληθυσμό. Ο πληθυσμός μιμείται την κίνηση του σκουληκιού και την εύρεση τροφής από σάλπες²⁴ και περιλαμβάνει ένα σύνολο από n διαστάσεις τυχαίων λύσεων. Για κάθε διάσταση οι θέσεις αρχηγού-ακολουθών αναβαθμίζονται (για $i=1$) (Εικόνα 88). Οι καλύτερες τιμές υπολογίζονται για κάθε υποψήφια λύση και επιλέγεται η σάλπα με την καλύτερη τιμή (Al-Ahmad et al., 2021, p.; Mirjalili et al., 2017; Ozbay & Alatas, 2020). Ο Al-Ahmad χρησιμοποίησε τον δυαδικό αλγόριθμο σάλπας (Binary Salp Swarm) ακολουθώντας τα βήματα του πρωτότυπου.

²⁴ Οι σάλπες ανήκουν στην οικογένεια των Salpidae και έχουν διαφανές σώμα σε σχήμα κυλινδρικό. Οι ιστοί τους είναι όμοιοι με τις μέδουσες και κινούνται όπως αυτές. Δες και (Mirjalili et al., 2017).



Εικόνα 87. Σάλπα.

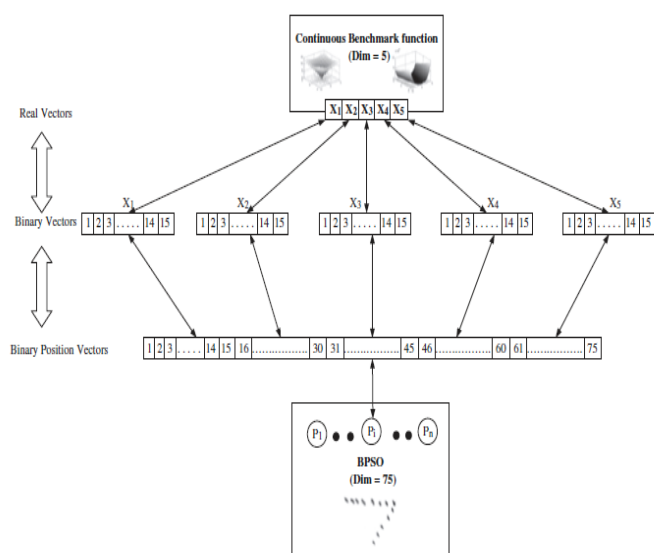


Εικόνα 88. Αλγόριθμος Salp Swarm Optimization.

Ο Al-Ahmad χρησιμοποίησε και τον *δυναμικό αλγόριθμο σμήνους (Binary Particle Swarm)* (Mirjalili, n.d.-b) όπου πηγή έμπνευσης είναι τα σμήνη των πουλιών που ψάχνουν για τροφή. Στον συγκεκριμένο αλγόριθμο οι λύσεις εκφράζονται ως σωματίδια και η εύρεση τροφής πραγματοποιείται από δύο παράγοντες: τον *rbest* δηλαδή την εύρεση λύσης με βάση την καλύτερη εμπειρία από προηγούμενα σωματίδια και τον *gbest* όπου αφορά την καλύτερη λύση (0,1) σε όλο το πλήθος (Εικόνα 89). Η ταχύτητα και η θέση των σωματιδίων αυξάνεται σε κάθε επανάληψη.

Σύμφωνα με τον Al-Ahmad αλγόριθμοι όπως ο αλγόριθμος σάλπας και ο αλγόριθμος σμήνους έχουν επιτυχία γιατί παρέχουν προσαρμοστικότητα σε περιβαλλοντικές αλλαγές και αναβαθμίζουν την συμπεριφορά τους. Αυτό ονομάζεται και αυτο-οργάνωση μέσω της οποίας αναβαθμίζονται συνεχώς σε κάθε χρονικό διάστημα, και κατανεμημένο έλεγχο,

έχοντας ευκολία στη μέτρηση. Ουσιαστικά μπορούν να ανταπεξέλθουν σε αυξημένο αριθμό πρακτόρων χωρίς αλλαγή στην σχεδίαση του αλγορίθμου και διαθέτουν μεγάλη ευελιξία (Al-Ahmad et al., 2021).



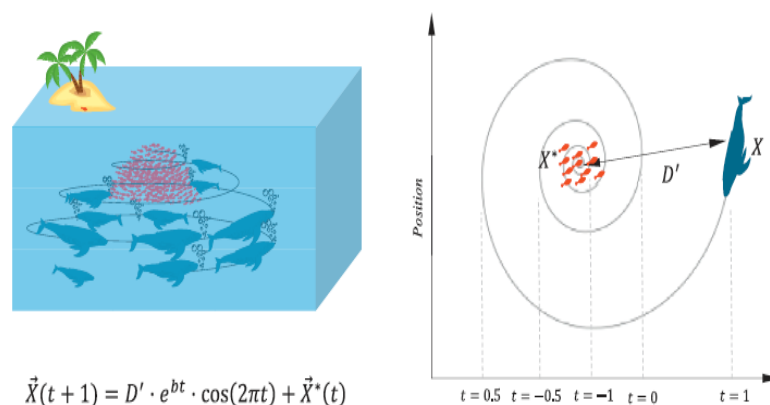
Εικόνα 89. Μετατροπή του αλγορίθμου σμήνους σε δυαδικό.

Ο αλγόριθμος *Whale optimization* (Mirjalili, n.d.-d) (Εικόνα 90) είναι ένας αλγόριθμος βασισμένος στον κυνηγετικό τρόπο των μεγάπτρων φαλαινών και χρησιμοποιείται για επίλυση διαφορετικών προβλημάτων βελτιστοποίησης. Περιλαμβάνει τρεις φάσεις:

- **Περικύκλωση δολώματος (Encircling prey):** αναβάθμιση διανυσματικών θέσεων ερευνητικών πρακτόρων με βάση την θέση του πρόσφατου καλύτερου πράκτορα
- **Επίθεση με φυσαλίδες (Bubble-net attacking):** επαναπροσδιορισμός λύσης για εύρεση βέλτιστης (εάν η πιθανότητα είναι κάτω από 0.5 τότε αναλαμβάνει ο μηχανισμός περικύκλωσης μέσω αλλιώς χρήση μηχανισμού σπιδάλ για αναβάθμιση θέσης)
- **Αναζήτηση δολώματος (Search for prey):** ο ερευνητικός πράκτορας ψάχνει για δόλωμα τυχαία

Ο μηχανισμός περικύκλωσης δολώματος (encircling prey) αποτελείται από δύο υποφάσεις, τον μηχανισμό περικύκλωσης δολώματος όπου η τιμή μειώνεται (από 2 σε 0) και τον μηχανισμό σπιδάλ όπου οι πράκτορες κινούνται σε ελικοειδή μορφή προς το δόλωμα. Στην έρευνα δολώματος (search for prey) η τυχαία κίνηση του πράκτορα βοηθά στην καλύτερη διερεύνηση της υπό έρευνας περιοχής και στην αποφυγή μπλοκαρίσματος του αλγορίθμου σε τοπικό μέγιστο (Mafarja & Mirjalili, 2017; A. C. Pandey & Tikkiwal, 2021;

Sheikhi, 2021). Οι συγγραφείς χρησιμοποίησαν τον αλγόριθμο για τη βελτιστοποίηση του νευρωνικού δικτύου MLP προκειμένου να εντοπιστεί η σχέση ανάμεσα σε δύο κείμενα ως προς τον ισχυρισμό αλλά και για τη βελτιστοποίηση του αλγόριθμου XGBoost ως προς τον εντοπισμό των ψευδών ειδήσεων.



Εικόνα 90. Αλγόριθμος Whale Optimization.

Ο αλγόριθμος *Improved whale optimization* είναι ο αλγόριθμος που βελτιστοποιεί τον *Whale optimization* ως προς την ταχύτητα κίνησης αλλά και ως προς την παραγωγή καλύτερων αποτελεσμάτων. Η επιλογή των μεγάπτερων φαλαινών πραγματοποιείται μέσω της επιλογής *tournament* και *roulette wheel*. Η επιλογή *tournament* γίνεται σε μονή επανάληψη για την εύρεση πρακτόρων (μεγάπτερες φάλαινες) ενώ στον *roulette wheel* ο έλεγχος της ανταλλαγής επιτυγχάνεται ανάμεσα στην διερεύνηση και την παραγωγικότητα. Στην επιλογή *tournament* οι ερευνητές πράκτορες επιλέγονται τυχαία, υπολογίζεται για κάθε πράκτορα ερευνητή η καταλληλότητα, κατηγοριοποιούνται σύμφωνα με αυτή και επιλέγεται ο ατομικός ερευνητής με τη καλύτερη τιμή (νικητής) (Mafarja & Mirjalili, 2017; A. C. Pandey & Tikkiwal, 2021).

4.4.5 Αλγόριθμοι Βαθιάς Μάθησης (Deep Learning)

Οι αλγόριθμοι, και τα μοντέλα βαθιάς μάθησης περιλαμβάνουν λεπτομερή εκμάθηση πολλαπλών πληροφοριών από κείμενο και εικόνα. Στους αλγόριθμους βαθιάς μάθησης υπολογίζονται τα απλά νευρωνικά δίκτυα (*ANN*), ο αλγόριθμος μακροχρόνιας μνήμης (*LSTM*) αλλά και υποκατηγορίες του όπως ο *Bi-LSTM*, τα εμπροσθοβαρή νευρωνικά δίκτυα (*FFNN*), ο *MLP*, *GRU*, *Vanilla*, τα επαλαμβανόμενα νευρωνικά δίκτυα (*RNN*), τα συνελκτικά νευρωνικά δίκτυα (*CNN*), μοντέλα όπως το *FNDeep*.

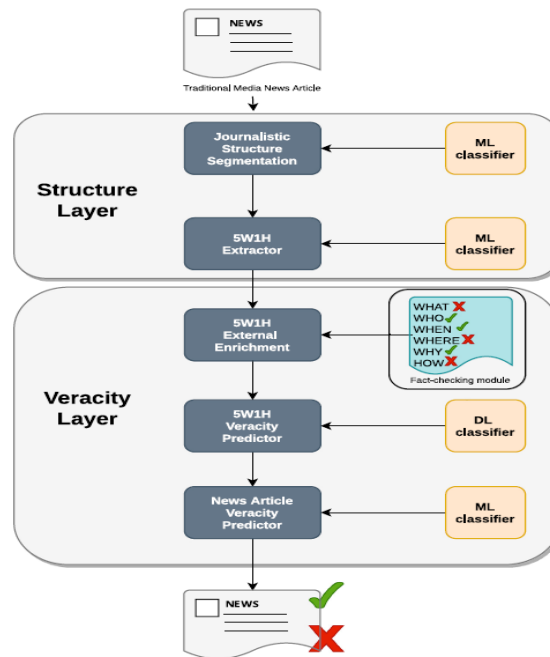
Ο Bonet-Jover δημιούργησε ένα μοντέλο εντοπισμού ψευδών ειδήσεων βαθιάς μάθησης για τον Covid19 (FNDEEP) (Εικόνα 91) με δύο επίπεδα: το επίπεδο δομής και το επίπεδο αξιοπιστίας.

Στο επίπεδο δομής η πληροφορία δομείται σύμφωνα με δύο επίπεδα αναπαράστασης. Σε πρώτη φάση διαχωρίζονται οι ειδήσεις σε μέρη: τίτλο, υπότιτλο, πρώτη παράγραφος, σώμα, συμπέρασμα και στη συνέχεια εξάγεται το βασικό περιεχόμενο με βάση τα ερωτήματα 5W1H τα οποία αφορούν τα γεγονότα και τις συνθήκες (What), τα άτομα εμπλοκής (Who), την τοποθεσία των γεγονότων (Where), την ώρα και τον χρόνο του γεγονότος (When), την αιτία του γεγονότος (Why) και τον τρόπο εξέλιξης των γεγονότων (How).

Στο επίπεδο αξιοπιστίας πραγματοποιείται ο προσδιορισμός και η πρόβλεψη της αξιοπιστίας των τμημάτων των ειδήσεων. Μετέπειτα το μοντέλο λαμβάνει στοιχεία από τα ερωτήματα 5W1H όπου εμπλουτίζονται οι πληροφορίες από εξωτερική πηγή ελέγχου γεγονότων (π.χ. IFCN²⁵). Στη συνέχεια ακολουθεί η πρόβλεψη αξιοπιστίας με χρήση κατηγοριοποιημένων χαρακτηριστικών από το κείμενο και τη γνώση ελέγχου γεγονότων (fact checking knowledge) όπου γίνεται η κατηγοριοποίηση κάθε στοιχείου μέσω αλγορίθμων βαθιάς μάθησης (αλγόριθμοι μακροχρόνιας μνήμης και εμπροσθοβαρούς νευρωνικού δικτύου).

Στο τέλος το μοντέλο προβλέπει την αξιοπιστία όλων των ειδήσεων με τη χρήση κλασικών αλγορίθμων μηχανικής μάθησης (π.χ. γραμμική παλινδρόμηση, δέντρο απόφασης, πολυονιμικός Naïve Bayes) (Bonet-Jover et al., 2021).

²⁵ Παγκόσμιο δίκτυο που αποτελεί μονάδα του Ινστιτούτου Poynter και που ενώνει τους ελεγκτές γεγονότων παγκοσμίως. Δες (“International Fact-Checking Network,” n.d.)

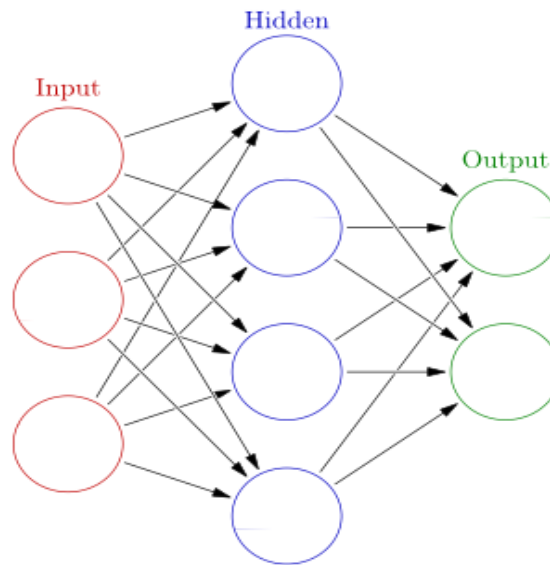


Εικόνα 91. FNDEEP.

Στα απλά νευρωνικά δίκτυα (ANN) (1.17. Neural Network Models (Supervised), n.d.) η εκμάθηση των αναπαραστάσεων βασίζεται σε προηγούμενες πληροφορίες. Η πληροφορία τροποποιείται με βάση τα δεδομένα των μοντέλων (Mahabub, 2020). Σύμφωνα με τον Patidar το νευρωνικό δίκτυο έχει επηρεαστεί από τα βιολογικά νευρωνικά δίκτυα και αποτελείται από ομάδες διασυνδεδεμένων νευρώνων οι οποίοι επεξεργάζονται την πληροφορία. Η δομή του επηρεάζεται από την εσωτερική και εξωτερική πληροφόρηση κατά την εκπαίδευση (Patidar et al., 2013).

Αποτελείται από τρία επίπεδα: το επίπεδο εισόδου, το κρυφό επίπεδο και επίπεδο εξόδου. Οι δεσμοί στο επίπεδο εισόδου συνδέονται με συγκεκριμένο αριθμό δεσμών στο κρυφό επίπεδο. Οι δεσμοί από το κρυφό επίπεδο συνδέονται με τη σειρά τους είτε με δεσμούς του επόμενου κρυφού επιπέδου είτε με το εξωτερικό επίπεδο (Annasaheb & Verma, 2016).

Τα νευρωνικά δίκτυα χρησιμοποιήθηκαν από τους συγγραφείς για ακρίβεια στις προβλέψεις (Elhadad et al., 2020; Mahabub, 2020; Shim et al., 2021) αλλά και σε προηγούμενες μελέτες για εκμάθηση μοτίβων των μηνυμάτων spam (Patidar et al., 2013). Σύμφωνα με τον Poddar το τεχνητό νευρωνικό δίκτυο δε μαθαίνει μόνο λέξεις κλειδιά αλλά ενσωματώνει όλες τις λέξεις στο δίκτυο βοηθώντας σε βελτίωση των αποτελεσμάτων (Poddar et al., 2019) (Εικόνα 92).

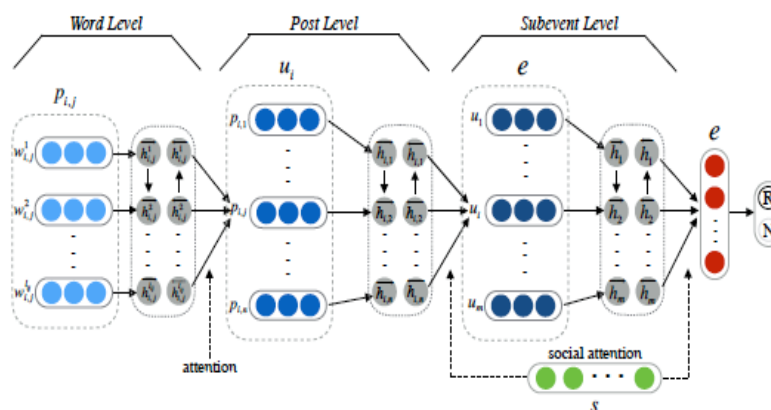


Εικόνα 92. Νευρωνικό δίκτυο.

Το *HPA-BLSTM* (Εικόνα 93) είναι ένα νευρωνικό δίκτυο που μαθαίνει την αναπαράσταση μέσω ενός δικτύου ιεραρχικού μηχανισμού προσοχής σε λεκτικό επίπεδο, επίπεδο δημοσίευσης και επίπεδο υπο-γεγονότος (H. Guo et al., 2018; Shu, Cui, et al., 2019). Σύμφωνα με τον Guo και τα τρία επίπεδα χρησιμοποιούν τον αλγόριθμο Bi-LSTM, στο λεκτικό επίπεδο περιλαμβάνεται και το επίπεδο προσοχής (*attention layer*), ενώ στα επίπεδα δημοσίευσης και υπογεγονότος το επίπεδο προσοχής κοινωνικών χαρακτηριστικών.

Στο λεκτικό επίπεδο κάθε λέξη ενσωματώνεται ως διάνυσμα και σε κάθε χρόνο τροφοδοτείται ο Bi-LSTM με τη χρήση ενός διανύσματος λέξης. Μέσω του μηχανισμού προσοχής δίνεται βαρύτητα σε σημαντικές λέξεις και βελτιώνεται η αναπαράσταση. Στη συνέχεια υπολογίζονται τα βάρη κάθε λέξης μέσω της λειτουργίας *softmax* και η αναπαράσταση για κάθε δημοσίευση και υπογεγονός.

Στο επίπεδο δημοσίευσης και στο επίπεδο γεγονότος ακολουθείται η ίδια διαδικασία προσθέτοντας και τα κοινωνικά χαρακτηριστικά υποβοηθώντας, κατά αυτόν τον τρόπο, τον μηχανισμό προσοχής να εστιαστεί στον εντοπισμό αξιόλογων δημοσιεύσεων (H. Guo et al., 2018).



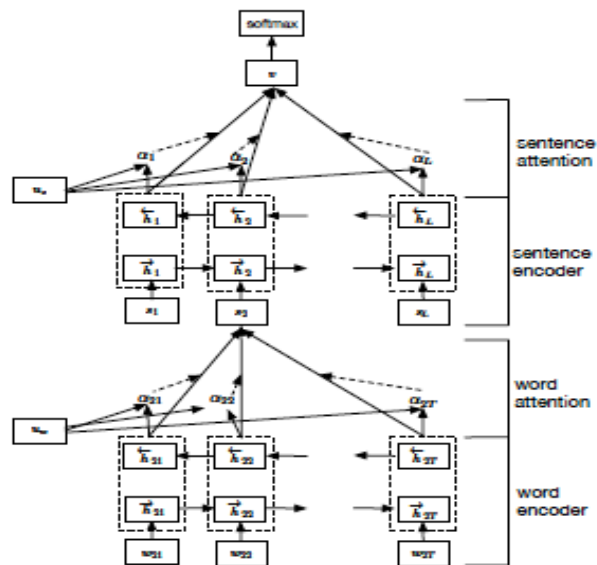
Εικόνα 93. HPA-BLSTM.

Το νευρωνικό δίκτυο ιεραρχικού μηχανισμού προσοχής (HAN) (TextClassifier, n.d.) (Εικόνα 94) κωδικοποιεί το περιεχόμενο των ειδήσεων σε διανυσματική αναπαράσταση, παράγει έναν πίνακα λέξεων, εξάγει τις σημαντικές λέξεις για την πρόταση και τις αναπαριστά σε διάνυσμα (sentence vector), κωδικοποιεί το διάνυσμα κειμένου από διανυσματικές προτάσεις και μετρά τη σημαντικότητα κάθε πρότασης (sentence-level context vector) (Shu, Cui, et al., 2019).

Το συγκεκριμένο δίκτυο σύμφωνα με τον Yang αποτελείται από τον κωδικοποιητή συχνότητας λέξης (word sequence encoder), το επίπεδο προσοχής λέξης (word attention layer), τον κωδικοποιητή πρότασης (sentence level encoder), και το επίπεδο προσοχής πρότασης (sentence attention layer). Ο κωδικοποιητής συχνότητας λέξης περιλαμβάνει τον αλγόριθμο GRU ο οποίος εντοπίζει την κατάσταση των συχνοτήτων χρησιμοποιώντας διαφορετικές μονάδες μνήμης. Περιλαμβάνει δύο στοιχεία την πύλη διαγραφής και την πύλη αναβάθμισης οι οποίες ελέγχουν την πληροφορία και πώς αναβαθμίζεται σε κάθε κατάσταση (π.χ. εάν η τιμή είναι 0 στην πύλη διαγραφής τότε ξεχνά την προηγούμενη κατάσταση).

Το επίπεδο προσοχής αναπαριστά το πρωτογενές έγγραφο σε διανυσματική αναπαράσταση όπου μέσω του ταξινομητή αλγόριθμου γίνεται η κατηγοριοποίηση του κειμένου. Στη συνέχεια στον κωδικοποιητή λέξης ο αλγόριθμος GRU συγκεντρώνει την πληροφορία από τις λέξεις και από τις δύο κατευθύνσεις (αναγνώριση πρότασης μπροστά και πίσω). Μέσω του μηχανισμού προσοχής εξάγονται οι λέξεις που έχουν σημασία σε σχέση με την πρόταση και ενσωματώνονται οι αναπαραστάσεις τους (διανυσματικές).

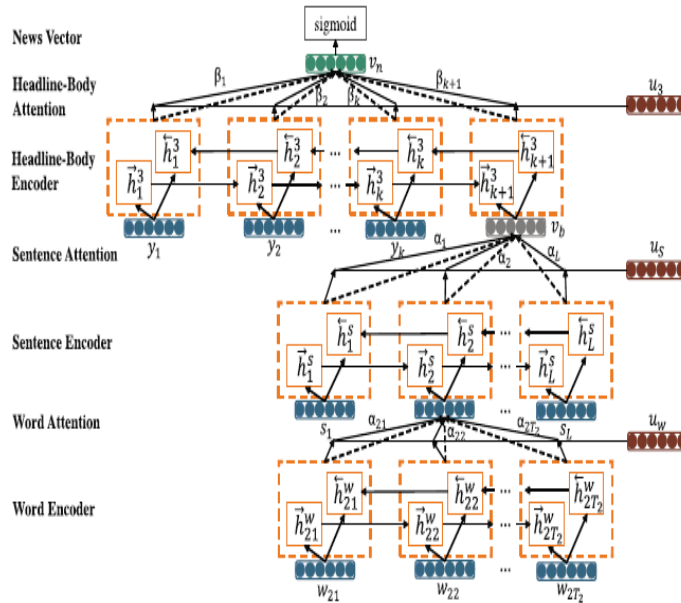
Ο κωδικοποιητής πρότασης μετατρέπει τις αναπαραστάσεις των εγγράφων σε διανύσματα μέσω του αλγόριθμου GRU και μέσω του μηχανισμού προσοχής δίνεται έμφαση στις προτάσεις που παρέχουν αρκετά στοιχεία για την κατηγοριοποίηση του εγγράφου (Z. Yang et al., 2016).



Εικόνα 94. Hierarchical Attention Network (HAN).

Το δίκτυο *3HAN* (Singhanía et al., n.d.) (Εικόνα 95) είναι ένα ακόμη δίκτυο το οποίο έχουν υιοθετήσει προηγούμενες μελέτες για τον εντοπισμό των ψευδών ειδήσεων (Meel & Vishwakarma, 2020). Αποτελείται από τρία επίπεδα ιεραρχικού δικτύου προσοχής (πρόταση, τίτλος, λέξη) όπου η εκμάθηση της πληροφορίας πραγματοποιείται από κάτω προς τα πάνω (bottom-up). Στο συγκεκριμένο δίκτυο η προσοχή εστιάζεται σε διαφορετικά μέρη του άρθρου βασιζόμενα στη σημασία τους.

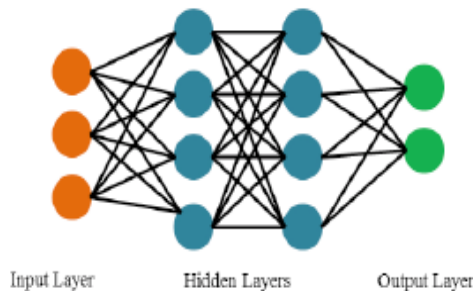
Σύμφωνα με την Singhanía έχει επηρεαστεί από το δίκτυο HAN και αποτελείται από τον κωδικοποιητή συχνότητας λέξης, από το επίπεδο προσοχής λέξης (επίπεδο 1), τον κωδικοποιητή πρότασης, το επίπεδο ελέγχου πρότασης (επίπεδο 2) και στο τρίτο επίπεδο προστέθηκε ο κωδικοποιητής τίτλου-σώματος και το επίπεδο προσοχής τίτλου-σώματος (Singhanía et al., 2017). Η διαδικασία των κωδικοποιητών (λέξης-προσοχής, πρότασης-προσοχής) είναι παρόμοια με τη μόνη διαφορά στον κωδικοποιητή τίτλου ο οποίος προσθέτει λέξεις από τον τίτλο και δίνει έμφαση (προσοχή) σε κάθε μία από αυτές χρησιμοποιώντας και σε αυτό το επίπεδο τον αλγόριθμο διπλής κατεύθυνσης GRU. Ο συγκεκριμένος αλγόριθμος δικτύου είναι γρήγορος και παρέχει πολύ μεγάλη ακρίβεια (πάνω από 95%) στον εντοπισμό των ψευδών ειδήσεων.



Εικόνα 95. 3HAN.

Το βαθύ νευρωνικό δίκτυο DNN (Dense Neural Network) (Εικόνα 96) αποτελείται από τρία επίπεδα: το επίπεδο εισόδου, το κρυφό επίπεδο, και το επίπεδο εξόδου. Το επίπεδο εισόδου επιτρέπει την εισαγωγή συχνοτήτων λέξεων. Το συγκεκριμένο δίκτυο χρησιμοποιήθηκε σε προηγούμενες έρευνες (Thota et al., 2018) για την αξιολόγηση των ψευδών ειδήσεων με βάση τις απόψεις (stance) σε έναν ισχυρισμό (συμφωνώ, διαφωνώ, συζητώ) (Meel & Vishwakarma, 2020).

Dense Neural Network (DNN)

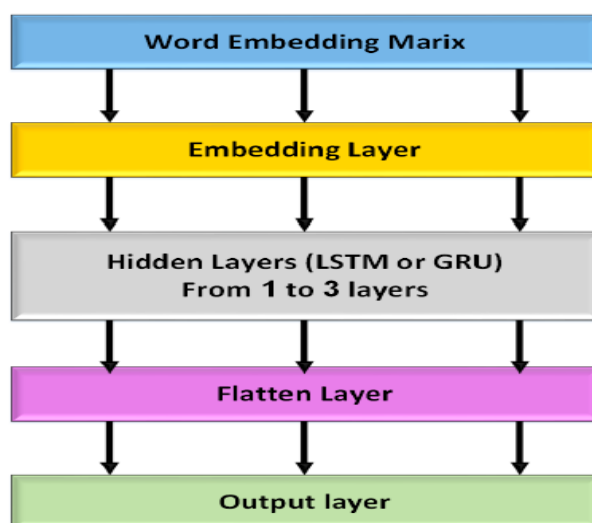


Εικόνα 96. Dense Neural Network.

Το νευρωνικό δίκτυο βαθιάς μάθησης (Εικόνα 97) (Deep Neural Network) (Team, n.d.-i, n.d.-g, n.d.-f, n.d.-a, n.d.-h) αποτελείται από πέντε επίπεδα: το επίπεδο ενσωμάτωσης

(*embedding layer*) όπου εισάγεται ο πίνακας με τις τιμές των λέξεων, τα κρυφά επίπεδα με τους αλγόριθμους μακροχρόνιας μνήμης (*LSTM*) ή επαναλαμβανόμενες μονάδες πύλης (*Gated Recurrent Unit, GRU*), το ενιαίο επίπεδο και το επίπεδο εξόδου (Abdelminaam et al., 2021; Manaswi, 2018; Meel & Vishwakarma, 2020).

Το επίπεδο ενσωμάτωσης προσδιορίζει το λεξιλόγιο της βάσης εκπαίδευσης (διάσταση εισόδου), το μέγεθος του διανυσματικού χώρου για την ενσωμάτωση των λέξεων (διάσταση εξόδου) και το μήκος των συχνοτήτων εισόδου με χρήση της βιβλιοθήκης προγραμματισμού Keras (Gullì & Pal, 2017)²⁶. Το κρυφό επίπεδο χρησιμοποιεί για κάθε αλγόριθμο (*LSTM, GRU*) διαφορετικό αριθμό κρυφών επιπέδων (ένα, δύο, τρία) και διαφορετικούς νευρώνες για κάθε επίπεδο. Το επίπεδο εξόδου κατηγοριοποιεί τα tweet σε αληθινά και ψεύτικα (Abdelminaam et al., 2021; Wanto et al., 2017; Z. Zhang, 2018). Το συγκεκριμένο δίκτυο χρησιμοποιήθηκε για βαθέα μάθηση από τον Agarap (Agarap, 2019) όπου ο συγγραφέας χρησιμοποίησε το δίκτυο μαζί με την τεχνική Glove για τη μέτρηση των λέξεων σε ψεύτικα tweet.



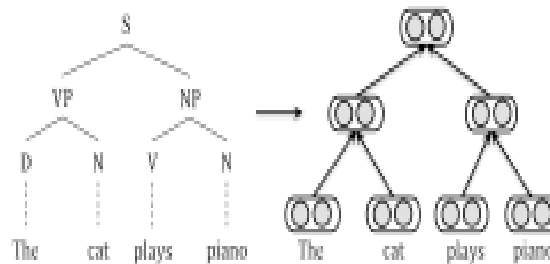
Εικόνα 97. Νευρωνικό δίκτυο βαθιάς μάθησης (Deep Neural Network).

Το αναδρομικό νευρωνικό δίκτυο (*RvNN*) (Rumor_RvNN, 2018) είναι ένα είδος νευρωνικού δικτύου δεντρικής δομής (Εικόνα 98). Η πρώτη έκδοση χρησιμοποίησε δυαδικά δέντρα ανάλυσης (σύνταξης) προτάσεων όπου η αναπαράσταση η οποία σχετίζεται με κάθε

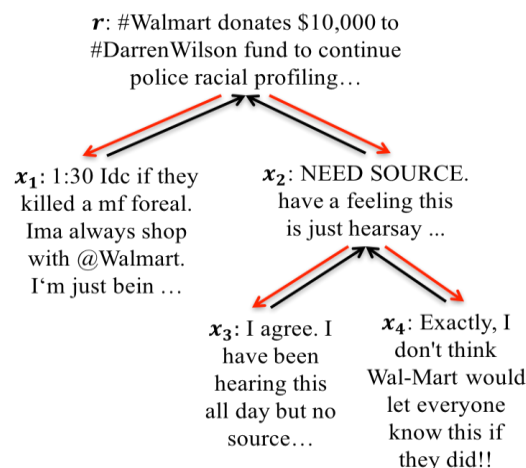
²⁶ Η βιβλιοθήκη προγραμματισμού Keras μαζί με το Tensorflow αποτελούν βασικά στοιχεία για την μηχανική μάθηση αφού το πακέτο περιλαμβάνει ροές διαχείρισης εργασιών μηχανικής μάθησης, ευελιξία και μπορεί να λειτουργήσει σε μεγάλους επεξεργαστές. Χρησιμοποιείται από τη NASA, το CERN και άλλα παγκοσμίως γνωστά ιδρύματα κυρίως για έρευνα σε βαθέα μηχανική μάθηση. Δες (Keras: The Python Deep Learning API, n.d.)

δέσμη δέντρου υπολογίζεται από τα παράγωγά του (children). Τα φύλλα των δεσμών είναι οι λέξεις των ενσωματώσεων των προτάσεων ενώ οι δεσμοί χωρίς φύλλα αναπαριστούν τους υποδεσμούς (Ma et al., 2018) .

Σύμφωνα με τους Meel και Viswakarma το νευρωνικό δίκτυο RvNN υιοθετήθηκε από τους Ma και Davis και Proctor για εντοπισμό των ψευδών ειδήσεων (Davis & Proctor, 2017; Ma et al., 2018). Το συγκεκριμένο δίκτυο αποτελεί και αυτό ένα δίκτυο βαθιάς μάθησης. Ο Ma επέκτεινε το δίκτυο με δύο διαφορετικά υπομοντέλα από πάνω προς τα κάτω (top-down) και κάτω προς τα πάνω (bottom-up). Στο μοντέλο Bottom-up δημιουργούνται τα διανύσματα για κάθε υποδέντρο με επαναληπτική επίσκεψη κάθε δέσμης φύλλων από κάτω προς τα πάνω (Εικόνα 99) (Ma et al., 2018).



Εικόνα 98. Αλγόριθμος δικτύου RvNN.

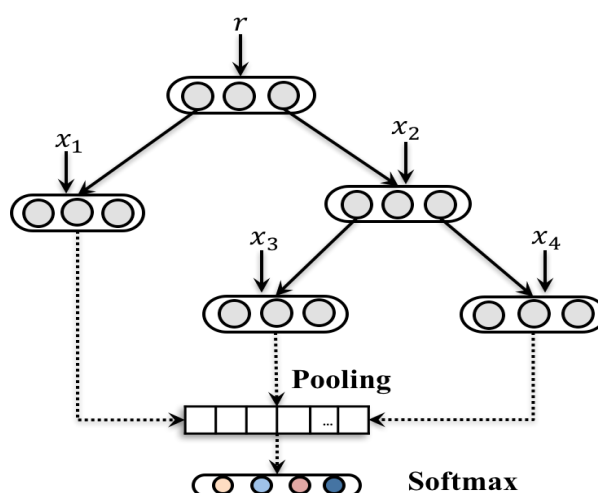


Εικόνα 99. Top-Down, Bottom-Up RvNN.

Ο αλγόριθμος δικτύου δενδρικής δομής (TRNN) (Rumor_RvNN/Model/Main_TD_RvNN.Py, n.d.) (Εικόνα 100) αποτελεί μια υποκατηγορία του

αλγόριθμοι RvNN που εκπαιδεύεται σε επιλεγμένα χαρακτηριστικά από περιεχόμενο απαντήσεων με το να παρακολουθεί τη μη συνεχή δομή διάδοσης τους (Q. Zhang et al., 2019). Ο συγγραφέας επέλεξε τη δομή από πάνω προς τα κάτω (top-down) για την αναπαράσταση των tweet καθώς είναι καλύτερη από την δομή από κάτω προς τα πάνω (bottom-up). Η συγκεκριμένη υποκατηγορία ενισχύει τη δομή του αλγόριθμου για να αποτυπώσει πολύπλοκα μοτίβα διάδοσης προκειμένου να κατηγοριοποιήσει τους ισχυρισμούς.

Ουσιαστικά ο TRNN (Εικόνα 100) μοντελοποιεί τον τρόπο ροής της πληροφόρησης από την πηγή δημοσίευσης στον τρέχοντα δεσμό. Η διαφορά του με την υποκατηγορία (bottom-up) έγκειται στο ότι η αναπαράσταση υπολογίζεται συνδυάζοντας τα δεδομένα εισόδου και τον κύριο δεσμό (parent node) και όχι τον υποδεσμό (child node). Η διαδικασία εξελίσσεται (συνελκτικά) από τον πρωτότυπο δεσμό έως τους υποδεσμούς μέχρι όλοι οι δεσμοί (φύλλων) να ενεργοποιηθούν (Ma et al., 2018).

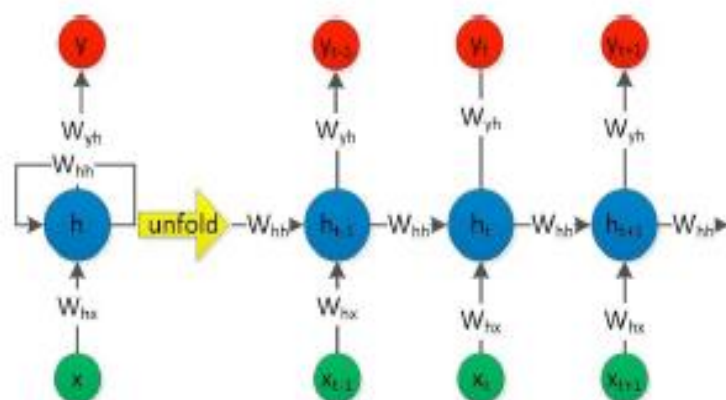


Εικόνα 100. Αλγόριθμος δικτύου TRNN.

Το επαναλαμβανόμενο νευρωνικό δίκτυο (RNN) (GitHub - LuisPB7/Fnc-Msc, n.d.; Team, n.d.-j) (Εικόνα 101) κωδικοποιεί τη συχνότητα συμβόλων σε μία σταθερή διανυσματική αναπαράσταση και αποκωδικοποιεί την αναπαράσταση σε μία άλλη συχνότητα συμβόλων. Επιπλέον μαθαίνει τη συνεχή αναπαράσταση μίας πρότασης σε ένα διανυσματικό χώρο βοηθώντας κατά αυτόν τον τρόπο τη διατήρηση της σημασιολογικής και συντακτικής δομής της πρότασης (Abdelminaam et al., 2021; Borges et al., 2019; Du et al., 2021; Girgis et al., 2018; Kar, 2020; Meel & Vishwakarma, 2020; Shu, Cui, et al., 2019; Taskin et al., 2022).

Ο Kar αναφέρει ότι το RNN αν και μπορεί να μοντελοποιήσει τη συχνότητα των δεδομένων και να μειώσει την πολυπλοκότητά τους δεν μπορεί να επεξεργαστεί μεγάλες συχνότητες (Kar, 2020). Σύμφωνα με τον Du το δίκτυο RNN περιλαμβάνει επίπεδα του αλγόριθμου LSTM και επίπεδα προσοχής αντί επιπέδων συνελκτικών δικτύων (convolutional) και επιπέδων μέγιστης εισροής (max pooling layer) (σε σχέση με το δίκτυο CNN) (Du et al., 2021).

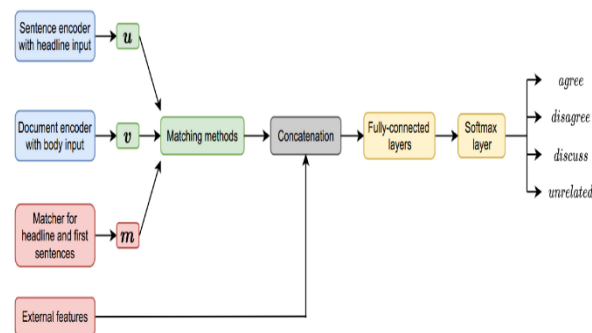
Ο Taskin ο οποίος υιοθέτησε το συγκεκριμένο δίκτυο επισημαίνει ότι η μνήμη του RNN είναι μικρή και ανεπαρκής αφού δεν μπορεί να γίνει η επεξεργασία των μεγάλων προτάσεων μέχρι το τέλος (ή κειμένων), χάνοντας κατά αυτόν τον τρόπο σημαντικές πληροφορίες (Taskin et al., 2022). Η Shu χρησιμοποίησε το RNN για την κωδικοποίηση των λέξεων (σχολίων κοινωνικών μέσων) ως μέρος του αποκωδικοποιητή λέξης επιλέγοντας και εκείνη τους αλγόριθμους δικτύων LSTM και GRU (Shu, Cui, et al., 2019). Ενώ ο Borges το χρησιμοποίησε συμπληρωματικά με νευρωνικά δίκτυα διπλής κατεύθυνσης (BiRNN) για περαιτέρω λεπτομερή εκμάθηση του κωδικοποιητή πρότασης ως προς τις ειδήσεις (Borges et al., 2019)



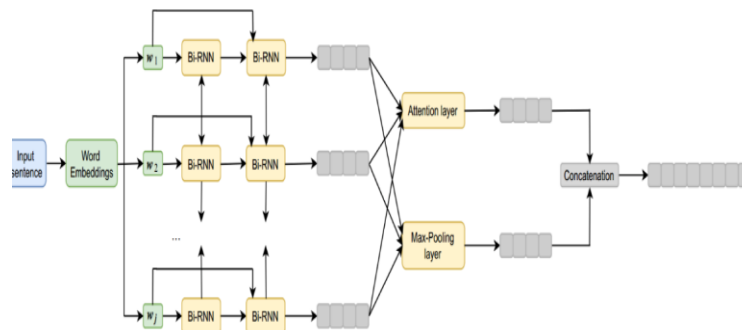
Εικόνα 101. Αλγόριθμος δικτύου RNN.

Το δίκτυο *BiRNN* (υποκατηγορία του δικτύου RNN) (Team, n.d.-c; *Tf.Keras.Sequential / TensorFlow v2.10.0*, n.d.) (Εικόνα 102, Εικόνα 103) είναι ένα μοντέλο νευρωνικού δικτύου διπλής κατεύθυνσης. Περιλαμβάνει έναν κωδικοποιητή προτάσεων για τη δημιουργία αναπαραστάσεων εισερχόμενων προτάσεων με βάση τη συχνότητα λέξεων στις προτάσεις. Στη συνέχεια ο *BiRNN* επεξεργάζεται τον παραγόμενο πίνακα. Μέσω των δύο πρόσθετων δικτύων διπλής κατεύθυνσης βελτιστοποιείται η απόδοση της πρότασης. Με την ίδια μέθοδο παράγεται και η κωδικοποίηση του κειμένου (Borges et al., 2019; Taskin et al., 2022).

Σύμφωνα με τον Taskin αν και το δίκτυο BiRNN είναι διπλής κατεύθυνσης (ένα επίπεδο κινείται προς τα εμπρός και το άλλο προς τα πίσω), η επεξεργασία της πληροφορίας γίνεται αντίστροφα (προς τα πίσω) (Taskin et al., 2022). Ο Borges χρησιμοποίησε το δίκτυο BiRNN συμπληρωματικά με το δίκτυο RNN για την κωδικοποίηση των προτάσεων στις ψευδείς ειδήσεις. Η συγκεκριμένη οπτική μπορεί να εισάγει γνώση από τη δομή του εγγράφου και παράλληλα προβλέπει την σχέση μεταξύ των προτάσεων και των λέξεων (Borges et al., 2019).



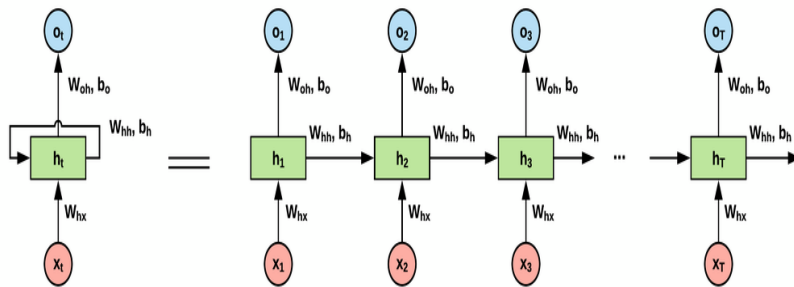
Εικόνα 102. Αλγόριθμος δικτύου BiRNN.



Εικόνα 103. Αλγόριθμος δικτύου BiRNN.

Το δίκτυο *Vanilla RNN* ήταν το πρώτο ανατροφοδοτούμενο δίκτυο που χρησιμοποιήθηκε το 1980 για την ταξινόμηση της πληροφορίας σύμφωνα με τον Girgis (Girgis et al., 2018). Το συγκεκριμένο δίκτυο είναι μονοεπίπεδο και ανατροφοδοτούμενο. Ο συγγραφέας το χρησιμοποίησε σε συνδυασμό με το δίκτυο μακροχρόνιας μνήμης LSTM και τη μονάδα ανατροφοδοτούμενης πύλης GRU. Οι συγκεκριμένοι αλγόριθμοι δικτύου βοηθούν σε περαιτέρω επίλυση των προβλημάτων που δημιουργεί το δίκτυο *Vanilla RNN* (ροή πληροφορίας) (Εικόνα 104)²⁷.

²⁷ Δες και (User, n.d.)



Εικόνα 104. Αλγόριθμος δικτύου Vanilla RNN.

Η μονάδα ανατροφοδοτούμενης πύλης (*Gated Recurrent Unit, GRU*) (Team, n.d.-g) (Εικόνα 105) είναι ένας πολυεπίπεδος αλγόριθμος δικτύου το οποίο παίρνει ως δεδομένο ότι οι δημοσιεύσεις είναι χρονικές ακολουθίες με διαφορετικό χρόνο (S. Qian et al., 2021). Ο αλγόριθμος δικτύου GRU αποτελείται από την πύλη διαγραφής και την πύλη αναβάθμισης (Kar, 2020).

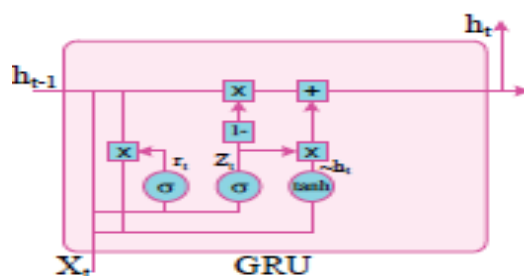
Οι συγγραφείς Kar και Shu τον υιοθέτησαν ως πρόσθετο αλγόριθμο για το δίκτυο RNN με σκοπό την επεξεργασία των δεδομένων μεγάλων συχνοτήτων επειδή το συγκεκριμένο δίκτυο χρησιμοποιεί λιγότερους πόρους μνήμης και ο RNN δυσκολεύεται να τα επεξεργαστεί (Kar, 2020; Shu, Cui, et al., 2019). Ο GRU χρησιμοποιήθηκε από την Shu μαζί με το δίκτυο RNN για την κωδικοποίηση των συχνοτήτων λέξεων υιοθετώντας την οπτική του Yang αλλά και στην κωδικοποίηση των προτάσεων των ειδήσεων στο μοντέλο DEFEND (Shu, Cui, et al., 2019; Z. Yang et al., 2016).

Στις μελέτες των Taskin και Girgis διαπιστώθηκε ότι ο αλγόριθμος δικτύου GRU ήταν πιο αποτελεσματικός από ότι οι υπόλοιποι αλγόριθμοι βαθιάς μάθησης ως προς τον εντοπισμό ψευδών ειδήσεων (Taskin et al., 2022). Ο Girgis επιπροσθέτως αναφέρει ότι οι ερευνητές χρησιμοποιούν τον αλγόριθμο GRU από το 2014 λόγω προβλημάτων του αλγορίθμου Vanilla (φιλτράρισμα ροής πληροφοριών) (Girgis et al., 2018).

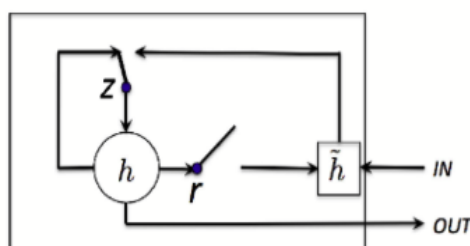
Σύμφωνα με τον Abdelminaam ο αλγόριθμος GRU (μετεξέλιξη του δικτύου RNN) έχει την ίδια δομή με το μοντέλο RNN με τη διαφορά ότι ο GRU αναβαθμίζει την πύλη (αναβάθμιση «κρυφής κατάστασης»). Ο GRU χρησιμοποιεί την πύλη διαγραφής και αναβαθμίζει την πύλη επιλέγοντας εάν η πληροφορία στην προηγούμενη «κρυφή κατάσταση» είναι χρήσιμη, τότε την κρατά και αφαιρεί την μη χρήσιμη πληροφορία (Abdelminaam et al., 2021; Kingma & Ba, 2017).

Οι συγγραφείς της τρέχουσας βιβλιογραφίας χρησιμοποίησαν τον GRU παράλληλα με τον αλγόριθμο δικτύου RNN και τον LSTM για βέλτιστα αποτελέσματα στον εντοπισμό των ψευδών ειδήσεων (Abdelminaam et al., 2021; Borges et al., 2019; Braşoveanu & Andonie, 2021; Girgis et al., 2018; Kar, 2020; Meel & Vishwakarma, 2020; S. Qian et al., 2021; Shu, Cui, et al., 2019; Taskin et al., 2022; Ying et al., 2021).

Εκτός από τον αλγόριθμο μονής ροής GRU, οι Taskin και Borges χρησιμοποίησαν και τον αλγόριθμο διπλής κατεύθυνσης BiGRU (Borges et al., 2019; Braşoveanu & Andonie, 2021; Taskin et al., 2022). Ο συγκεκριμένος αλγόριθμος έχει κοινά στοιχεία με τον αντίστοιχο αλγόριθμο διπλής κατεύθυνσης BiRNN αφού και εδώ η επεξεργασία των πληροφοριών πραγματοποιείται σε διπλή κατεύθυνση με δύο αλγορίθμους GRU με μόνη διαφορά ότι ο BiGRU περιλαμβάνει την πύλη εισόδου και την πύλη διαγραφής (*Papers with Code - BiGRU Explained*, n.d.; Team, n.d.-c) (Εικόνα 106).



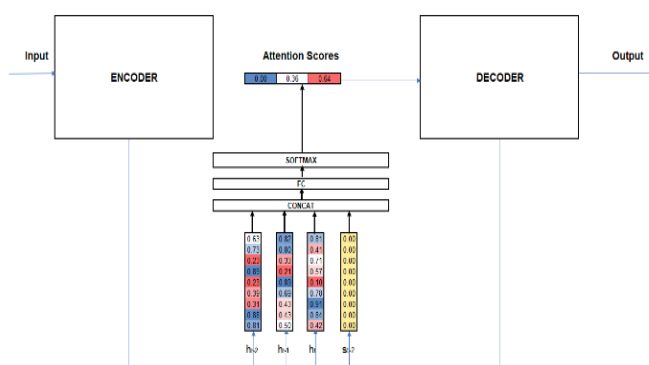
Εικόνα 105. Αλγόριθμος δικτύου GRU.



Εικόνα 106. Αλγόριθμος δικτύου διπλής κατεύθυνσης (BiGRU).

Ενώ οι Braşoveanu και Andonie μέσω του αλγορίθμου προσοχής GRU (GRU attention) (Team, n.d.-b) πέτυχαν καλύτερα αποτελέσματα με τη χρήση των ενσωματώσεων Glove εξαγοντας χαρακτηριστικά σχέσεων (relation extraction) (Braşoveanu & Andonie, 2021). Σύμφωνα με την Irie ο μηχανισμός προσοχής προσδιορίζεται ως ένα επίπεδο του οποίου τα δεδομένα εισόδου σε ένα χρονικό διάστημα t είναι η έξοδος του προηγούμενου επιπέδου

ανά χρόνο. Ο συγκεκριμένος μηχανισμός μαθαίνει τη σχετικότητα των δεδομένων εισόδου για κάθε πρόβλεψη (Irie et al., 2016) ²⁸ (Εικόνα 107).



Εικόνα 107. Μηχανισμός προσοχής GRU.

Το δίκτυο μακράς βραχύχρονης μνήμης (*LSTM*) (Olah, 2015; Team, n.d.-i) (Εικόνα 108) μαθαίνει πληροφορίες για μεγάλες χρονικές περιόδους (Abdelminaam et al., 2021; Bonet-Jover et al., 2021; Borges et al., 2019; Braşoveanu & Andonie, 2021; Girgis et al., 2018; Huang & Chen, 2020; Kaliyar et al., 2020; Kar, 2020; Kasseropoulos & Tjortjjs, 2021; Meel & Vishwakarma, 2020; Taskin et al., 2022; Q. Zhang et al., 2019). Σύμφωνα με τον Girgis ο αλγόριθμος δικτύου LSTM έχει παρεμφερή λειτουργία με τον RNN με τη διαφορά να εναπόκειται στη δομική του λειτουργία. Ο LSTM υπολογίζει την «κρυφή κατάσταση» (hidden state) μέσω τριών πυλών-μηχανισμών: εισόδου, εξόδου και λήθης (forget gate) αλλά και μία πρόσθετη μονάδα μνήμης η οποία μπορεί να κρατά την πληροφορία για μεγάλο χρονικό διάστημα (Girgis et al., 2018). Για τον Abdelminaam η διαφορά ανάμεσα στην LSTM και τον RNN βασίζεται στο ότι η LSTM έχει τέσσερα επίπεδα που αλληλοεπιδρούν ενώ το RNN μόνο ένα και είναι πιο αξιόπιστο από το απλό RNN (Abdelminaam et al., 2021; Gers, 1999; Hochreiter & Schmidhuber, 1997).

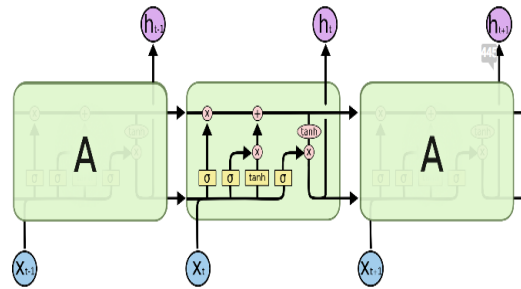
Η LSTM γενικότερα αποτελεί ένα πιο εξελιγμένο δίκτυο σε σχέση με το RNN. Οι Huang και Chen χρησιμοποίησαν δύο εκδοχές του μοντέλου LSTM, την *LSTM* για ενσωματώσεις λέξεων αλλά και την *LSTM* για το μέγεθος των προτάσεων αφού έχουν αρκετή σχετικότητα μεταξύ τους και μπορούν να κρατήσουν πληροφορίες στο χρόνο (Huang & Chen, 2020).

Οι Braşoveanu & Andonie υιοθέτησαν την LSTM με σημασιολογικά χαρακτηριστικά για την βελτίωση του εντοπισμού των ψευδών ειδήσεων (Braşoveanu & Andonie, 2021). Ο Bonet-Jover χρησιμοποίησε ένα μοντέλο κατηγοριοποίησης συχνότητας συνελκτικό της

²⁸ Δες και (Panda, 2020)

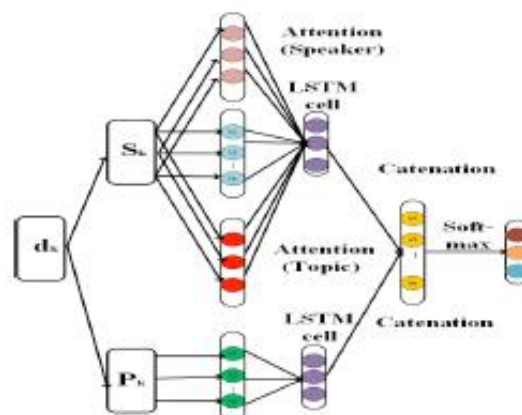
LSTM για την πρόβλεψη των ψευδών ειδήσεων σε πανδημικό μοντέλο (5W1H) (Bonet-Jover et al., 2021).

Ο Zhang μέσω της LSTM αποτύπωσε τη σημασιολογική ποικιλομορφία της πληροφορίας των απαντήσεων σε ισχυρισμούς διότι η LSTM δίνει μεγαλύτερη έμφαση σε πρόσφατα δεδομένα εισόδου (έμφαση σε τελευταία απάντηση) (Q. Zhang et al., 2019).



Εικόνα 108. Αλγόριθμος δικτύου LSTM.

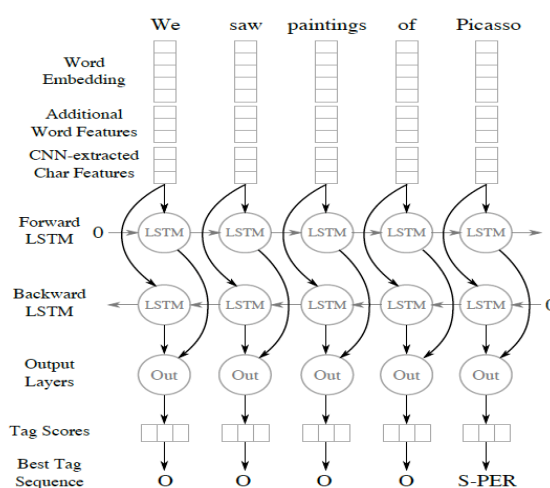
Ο Goldani σύγκρινε την LSTM με μηχανισμό προσοχής με τον αλγόριθμο δικτύου κάλυψης (capsule network) που ο ίδιος δημιούργησε για τον εντοπισμό των ψευδών ειδήσεων. Η LSTM με μηχανισμό προσοχής (Εικόνα 109) υιοθετήθηκε και από τον Long για τον προσδιορισμό της σημαντικότητας των διαφορετικών λέξεων σε ένα κείμενο και αποτελεί και αυτός έναν από τους κορυφαίους αλγόριθμους που χρησιμοποιούνται εξαιτίας της πολύ καλής του απόδοσης (Goldani et al., 2021; Long et al., 2017). Ο Long χρησιμοποίησε δύο LSTM με μηχανισμό προσοχής για τα προφίλ των ομιλητών και τη θεματική πληροφόρηση των άρθρων (Long et al., 2017).



Εικόνα 109. LSTM με μηχανισμό προσοχής.

Ο αλγόριθμος LSTM διπλής κατεύθυνσης (BiLSTM) (Team, n.d.-c) (Εικόνα 110) είναι ένα επαναλαμβανόμενο νευρωνικό δίκτυο όπου η χρήση των πληροφοριών γίνεται αμφίδρομα (Ahmad et al., 2020; Borges et al., 2019; Braşoveanu & Andonie, 2021; Girgis et al., 2018; Taskin et al., 2022; Q. Zhang et al., 2019). Ο Ahmad χρησιμοποίησε την LSTM διπλής κατεύθυνσης υιοθετώντας την οπτική του Wang.

Ο Wang στην προσέγγισή του υιοθέτησε τον LSTM μαζί με το δίκτυο CNN για την κωδικοποίηση των αναπαραστάσεων κειμένου (W. Y. Wang, 2017)²⁹. Ο LSTM διπλής κατεύθυνσης αποτελείται από δύο κρυφά επίπεδα της LSTM (εμπρόσθιο και προς τα πίσω) και δύο μονάδες μνήμης (Graves & Schmidhuber, 2005). Σύμφωνα με τον Zhang η BiLSTM αποτυπώνει σύντομες σημασιολογικές εξαρτήσεις από προηγούμενους χρόνους και μελλοντικά χρονικά βήματα μέσω αμφίδρομης κατάστασης (εμπρόσθια, προς τα πίσω) (Q. Zhang et al., 2019).



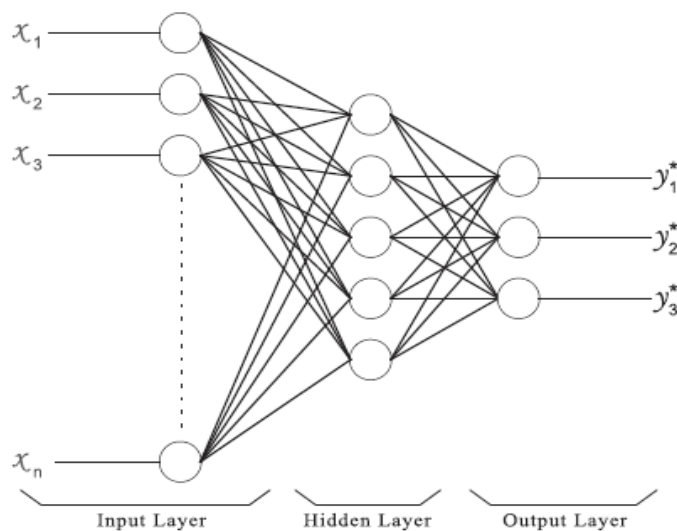
Εικόνα 110. Αλγόριθμος δικτύου BiLSTM.

Ο *MultiLayer Perceptron* (πολυεπίπεδη μηχανή) (1.17. *Neural Network Models (Supervised)*, n.d.) (Εικόνα 111) αποτελεί έναν απλό τύπο νευρωνικού δικτύου με πολυεπίπεδη κατηγοριοποίηση αναπαραστάσεων που προέρχονται από άλλα νευρωνικά δίκτυα όπως BiLSTM, CNN (Ahmad et al., 2020; Borges et al., 2019; Cremisini et al., 2019; Elhadad et al., 2020; Hunt et al., 2022; Mahabub, 2020; Meel & Vishwakarma, 2020; Ozbay & Alatas, 2020; A. C. Pandey & Tikkiwal, 2021; Saikh et al., 2019; Sheikhi, 2021; Q. Zhang et al., 2019).

²⁹ Δες και (Zvornicanin, 2022).

Ο MLP αποτελείται από τρία επίπεδα το επίπεδο εισόδου, το επίπεδο εξόδου, και το κρυφό επίπεδο. Το κρυφό επίπεδο είναι εκείνο που μαθαίνει πολύπλοκα χαρακτηριστικά δεδομένων. Μία παραλλαγή του MLP αποτελούν και τα εμπροσθοβαρή δίκτυα (FFNN) (Ahmad et al., 2020; A. C. Pandey & Tikkiwal, 2021). Οι συγγραφείς χρησιμοποίησαν τον MLP μαζί με μετα-αλγόριθμους (π.χ. βελτιωμένος αλγόριθμος φάλαινας-improved whale optimization) για να βελτιώσουν την απόδοση των αποτελεσμάτων.

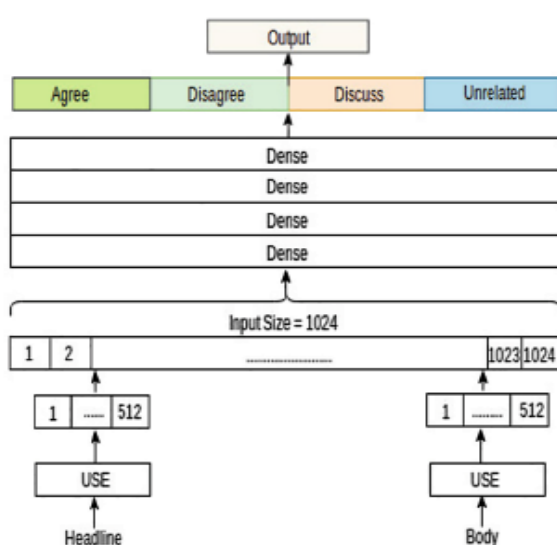
Ο Zhang υιοθέτησε τον MLP για την εύρεση της πολυπλοκότητας των σχέσεων ως προς τους ισχυρισμούς και τις απαντήσεις τους (Q. Zhang et al., 2019). Σύμφωνα με τον Borges αν και ο MLP χρησιμοποιείται στη φυσική επεξεργασία γλώσσας υπάρχουν προβλήματα ως προς τα δεδομένα μεγάλου μεγέθους γι' αυτό χρησιμοποιούνται αλγόριθμοι δικτύου όπως ο CNN (Borges et al., 2019). Ο Orhan αντίθετα διατυπώνει ότι ο MLP (MLPNN) αποτελεί έναν επιτυχημένο αλγόριθμο και ιδιαίτερα γνωστό αφού είναι γρήγορος, εύκολος ως προς την εφαρμογή και έχει ελάχιστες απαιτήσεις σε δεδομένα (Orhan et al., 2011).



Εικόνα 111. Αλγόριθμος δικτύου MLP.

Το νευρωνικό δίκτυο (εμπρόσθιας τροφοδότησης) (FFNN) (Team, n.d.-e; *Tf.Keras.Sequential | TensorFlow v2.10.0*, n.d.) αποτελείται από τέσσερα επίπεδα με το softmax επίπεδο να υπολογίζει την κατηγορία με υψηλό βαθμό πιθανότητας (Bonet-Jover et al., 2021; A. C. Pandey & Tikkiwal, 2021; Saikh et al., 2019). Ο συγκεκριμένος αλγόριθμος έχει κοινά χαρακτηριστικά με τον MLP και αποτελεί μια πιο εξειδικευμένη μορφή του (Saikh et al., 2019). Σύμφωνα με τον συγγραφέα οι νευρώνες του FFNN συνδέονται μονοεπίπεδα και σε μονή κατεύθυνση. Οι σύνδεσμοι αποτυπώνονται υπό μορφή βαρών που είναι πραγματικοί αριθμοί με κλίμακα -1, 1 (Féraud & Clérot, 2002; A. C. Pandey & Tikkiwal, 2021).

Ο FFNN περιλαμβάνει ένα επίπεδο εισόδου, ένα κρυφό επίπεδο και το επίπεδο εξόδου. Ο Saikh υιοθέτησε τέσσερα επίπεδα του FFNN (Εικόνα 112) αφού ήταν πιο αποτελεσματικά στην κωδικοποίηση των αναπαραστάσεων των τίτλων και του κειμένου. Για την τελική πρόβλεψη τα δεδομένα εξόδου δίδονται στο επίπεδο softmax για την πρόβλεψη της τάξης με την υψηλότερη βαθμολογία (Saikh et al., 2019). Ο Bonet-Jover χρησιμοποίησε τον αλγόριθμο (χρήση δύο επιπέδων) για την πρόβλεψη αξιοπιστίας και τον έλεγχο της πληροφορίας στο μοντέλο 5W1H.



Εικόνα 112. Αλγόριθμος δικτύου FFNN.

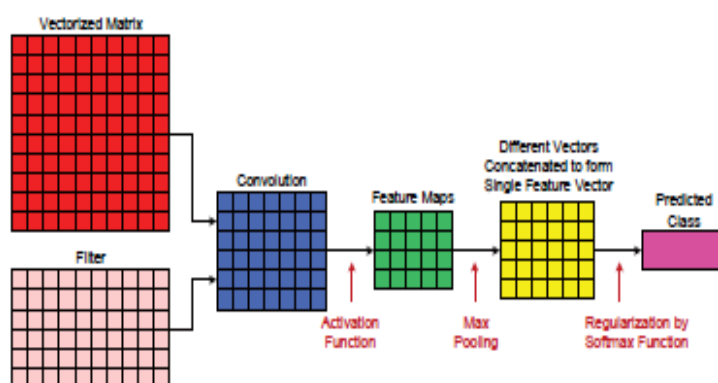
Ο αλγόριθμος δικτύου CNN (Team, n.d.-d) (*Convolutional Neural Network*) (Εικόνα 113) είναι ένα συνελκτικό νευρωνικό δίκτυο που έχει επηρεαστεί από τα συνδεδεμένα στοιχεία του ανθρώπινου εγκεφάλου (Ahmad et al., 2020; Braşoveanu & Andonie, 2021; Du et al., 2021; Girgis et al., 2018; Huang & Chen, 2020; Kaliyar et al., 2020; Kar, 2020; Kasseropoulos & Tjortjjs, 2021; Kausar et al., 2020; Meel & Vishwakarma, 2020; S. Qian et al., 2021; Tundis et al., 2021; Ying et al., 2021; Q. Zhang et al., 2019).

Σύμφωνα με τον Du ο CNN αποτελείται από ένα συνελκτικό επίπεδο, το επίπεδο μέγιστης εισροής (max pooling), ένα επίπεδο πλήρους σύνδεσης και το επίπεδο softmax για την ταξινόμηση. Η τροφοδότηση του CNN πραγματοποιείται με ενσωματώσεις λέξεων (π.χ. μοντέλο Glove) (Du et al., 2021). Ο Qian επισημαίνει ότι ο CNN μαθαίνει αναπαραστάσεις με τη μετατροπή παρεμφερών δημοσιεύσεων σε συχνότητες συγκεκριμένου μεγέθους (S. Qian et al., 2021; F. Yu et al., 2017). Ο Yu επιπροσθέτως αναφέρει ότι η αρχιτεκτονική του CNN βοηθά στη μοντελοποίηση σημασιολογικών χαρακτηριστικών και στη βελτίωση σε τεχνικές

αναγνώρισης ομιλίας, σημασιολογική τμηματοποίηση εικόνας κ.α. (Abdel-Hamid et al., 2012; F. Yu et al., 2017; H. Zhao et al., 2017).

Οι Huang και Chen υιοθέτησαν τον αλγόριθμο δικτύου CNN μαζί με αλγόριθμους όπως ο LSTM, σε συνδυασμό με αλγόριθμους γλωσσολογικούς όπως η LIWC αλλά και με μοντέλα όπως το ngram για τον διαχωρισμό των προτάσεων και των λέξεων σε αληθείς ή ψευδείς (Huang & Chen, 2020).

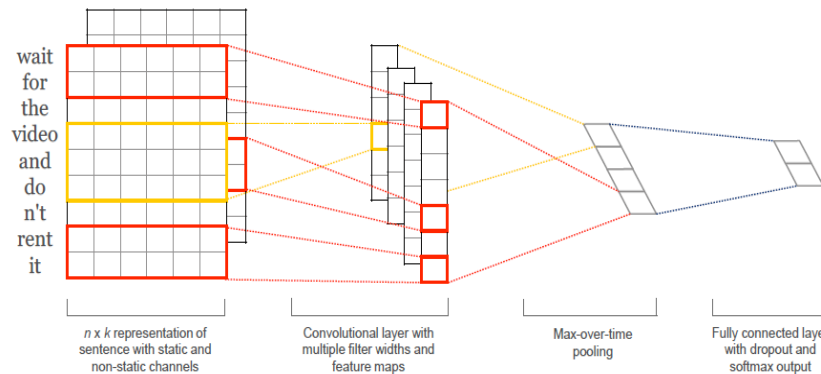
Ο Zhang χρησιμοποίησε τον CNN για τον εντοπισμό της αξιοπιστίας των ψευδών ειδήσεων καθώς είναι αρκετά ισχυρό ως προς την κατανόηση της φυσικής επεξεργασίας της γλώσσας (Y.-C. Chen et al., 2017; Dauphin et al., 2017; dos Santos & Gatti, 2014; Y. Kim, 2014; Y. Yang et al., 2022). Ο συγγραφέας αποτύπωσε χαρακτηριστικά όμοια με τα ngrams (=συχνότητες λέξεων) σε διαφορετικές χρονικές περιόδους (Y.-C. Chen et al., 2017; Q. Zhang et al., 2019). Οι Kasseropoulos και Tjortjis, Tundis, Kaliyar και Girgis επέλεξαν τον CNN για την εντοπισμό των ψευδών ειδήσεων διαπιστώνοντας πολύ καλά αποτελέσματα απόδοσης, (Girgis et al., 2018; Kaliyar et al., 2020; Kasseropoulos & Tjortjis, 2021; Tundis et al., 2021; Y. Zhang & Wallace, 2016) αλλά και ταχύτητας (Girgis et al., 2018). Ο Kar προσθέτει ότι η δομή του CNN έχει αρκετά πλεονεκτήματα αφού μπορεί να εντοπίσει τα σημαντικά χαρακτηριστικά χωρίς ανθρώπινη επίβλεψη και με υψηλή απόδοση (Kar, 2020; Patil et al., 2017).



Εικόνα 113. Αλγόριθμος δικτύου CNN.

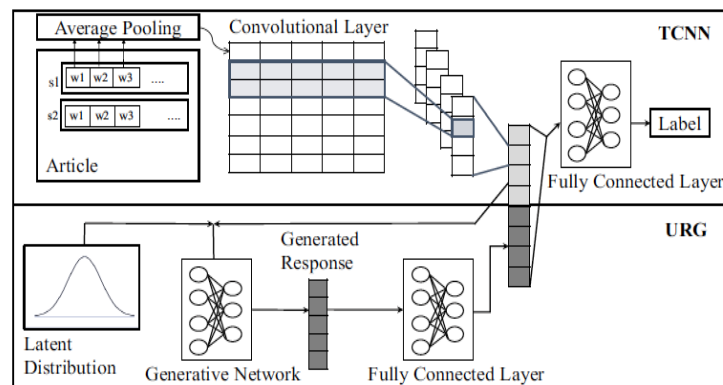
Ο *συνελκτικός αλγόριθμος κειμένου (Text-CNN) (Cnn-Text-Classification-Tf, n.d.)* (Εικόνα 114) είναι ένας αλγόριθμος τεχνητού νευρωνικού δικτύου για τη μοντελοποίηση περιεχομένου ειδήσεων το οποίο επικεντρώνεται στην πολυπλοκότητα των χαρακτηριστικών του κειμένου μέσω πολλαπλών φίλτρων (Braşoneanu & Andonie, 2021; Shu, Cui, et al., 2019). Ο συγκεκριμένος CNN είναι πιο απλός από το πρωτότυπο μοντέλο και περιλαμβάνει ένα μόνο επίπεδο συνελκτικού δικτύου και διαφορετικές διαστάσεις επιπέδων (Y. Kim, 2014). Οι Braşoneanu και Andonie προσθέτουν ότι αν και ο συγκεκριμένος αλγόριθμος έχει κοινή δομή

με τον CNN εντούτοις στα αποτελέσματα δεν πήγε πολύ καλά σε σχέση με άλλους αλγόριθμους βαθιάς μάθησης (π.χ. BiLSTM, GRU) (Βραζονεαυ & Andonie, 2021). Οι Kim και Shu υιοθέτησαν τον αλγόριθμο μαζί με διανύσματα λέξεων λόγω των πολύ καλών αποτελεσμάτων σε πολλαπλές μετρήσεις (Y. Kim, 2014; Shu, Cui, et al., 2019).



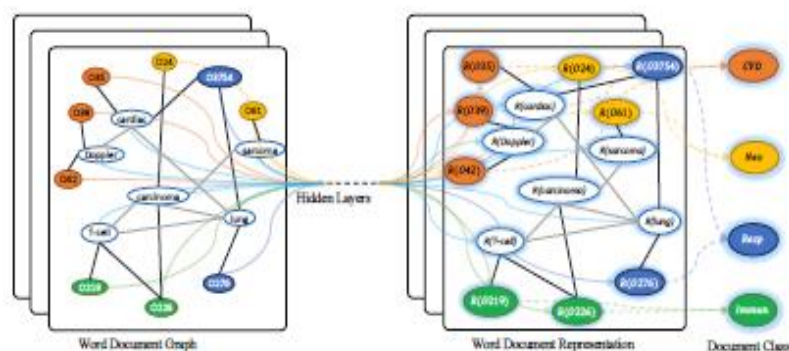
Εικόνα 114. Συνελκτικός αλγόριθμος δικτύου Text-CNN.

Ο αλγόριθμος TCNN-URG (Εικόνα 115) είναι ένας αλγόριθμος νευρωνικού δικτύου δύο επιπέδων. Στο πρώτο επίπεδο (TCNN) γίνεται η εκμάθηση των αναπαραστάσεων (αναπαράσταση άρθρου σε δύο επίπεδα και κατηγοριοποίηση) και το δεύτερο επίπεδο περιλαμβάνει έναν αυτόματο κωδικοποιητή (URG) για την εξαγωγή χαρακτηριστικών σχολιασμού των χρηστών (F. Qian et al., 2018; Shu, Cui, et al., 2019). Η Qian μάλιστα επισημαίνει ότι το πρώτο επίπεδο (TCNN) σχεδιάστηκε για να μετατρέπει τις λεκτικές πληροφορίες σε προτάσεις και στη συνέχεια να τις επεξεργάζεται μέσω του συνελκτικού νευρωνικού δικτύου προκειμένου να αποτυπώσει τη σημασιολογική ερμηνεία από μεγάλα κείμενα (F. Qian et al., 2018).



Εικόνα 115. Αλγόριθμος δικτύου TCNN-URG.

Ο συνελκτικός αλγόριθμος γράφου κειμένου (*Text-Graph Convolutional Network, Text-GCN*) (Text_gcn, 2019) (Εικόνα 116) χρησιμοποιεί τον γράφο του συνελκτικού νευρωνικού δικτύου (CNN) για εκμάθηση ενσωματώσεων λέξεων και εγγράφων (S. Qian et al., 2021; Ying et al., 2021). Ο Yao υιοθέτησε τον αλγόριθμο ο οποίος περιλάμβανε ενσωματώσεις διανυσμάτων λέξεων και εγγράφων από δεσμούς έτσι ώστε να μοντελοποιηθούν οι συχνότητες των λέξεων. Ο αριθμός των δεσμών σε ένα γράφο κειμένου είναι ο αριθμός εγγράφων μαζί με τις μοναδικές λέξεις του κειμένου (Yao et al., 2019). Σύμφωνα με τον Qian το συνελκτικό δίκτυο γράφου έχει χρησιμοποιηθεί και σε άλλες έρευνες (Kipf & Welling, 2017; Marcheggiani et al., 2018) για την αναπαράσταση κειμένου, και την κωδικοποίηση προτάσεων όμως δεν είναι κατάλληλο για πολυμοντελικές δημοσιεύσεις κοινωνικών μέσων (S. Qian et al., 2021).

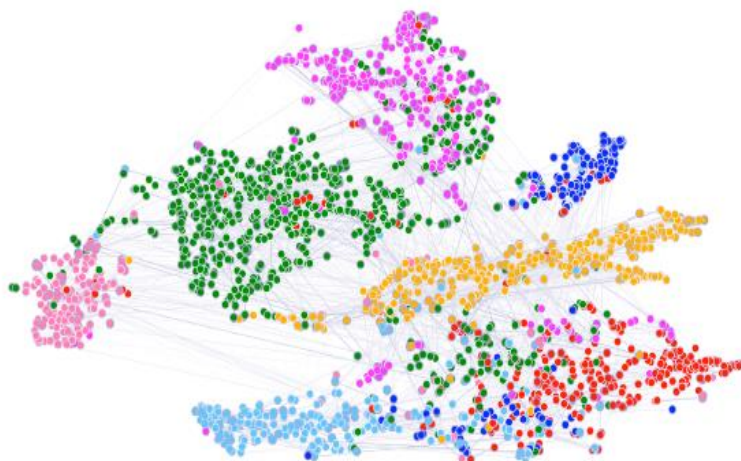


Εικόνα 116. Συνελκτικός αλγόριθμος γράφου κειμένου TEXT-GCN.

Ο αλγόριθμος δικτύου γράφου προσοχής (*Graph Attention Network*) (Veličković et al., n.d.) (Εικόνα 117) βασίζεται σε πολλαπλούς γράφους δεσμών όπου το βάρος δίνεται στους γειτονικούς δεσμούς (neighbor nodes) (Cohen et al., 2020; S. Qian et al., 2021). Σύμφωνα με τον Veličković ο συγκεκριμένος αλγόριθμος περιλαμβάνει τα εξής χαρακτηριστικά: αποτελεσματική λειτουργία ως προς τα ζεύγη δεσμού-γειτόνων, εφαρμόζεται σε δεσμούς γράφων με διαφορετικές αποστάσεις (προσδιορισμών βαρών σε γείτονες). Επιπλέον, ως δεδομένα εισόδου παίρνει χαρακτηριστικά δεσμών και εξάγει ένα καινούργιο σύνολο από χαρακτηριστικά (διαφορετικών αριθμητικών στοιχείων) (Veličković et al., 2018).

Η Qian προσθέτει ότι ο συγκεκριμένος αλγόριθμος βελτιώνει άλλους αλγόριθμους όπως ο συνελκτικός αλγόριθμος δικτύου (GCN) με την εφαρμογή της οπτικής προσοχής προκειμένου να ενσωματώσει χαρακτηριστικά από δεσμούς γειτόνων μέσω διαχωρισμού (S. Qian et al., 2021). Ο Cohen χρησιμοποίησε τον αλγόριθμο για τον εντοπισμό των πηγών

παραπληροφόρησης. Ο συγκεκριμένος αλγόριθμος επιταχύνει την εκμάθηση και έχει πολύ καλή απόδοση καθότι δεν επικεντρώνεται μόνο στο συναίσθημα των δημοσιεύσεων αλλά και σε σχέσεις ανάμεσα σε φίλους που συνδέονται μέσω συνδέσμων (π.χ. θετικά ή αρνητικά σχόλια / αντιδράσεις φίλων) (Cohen et al., 2020).



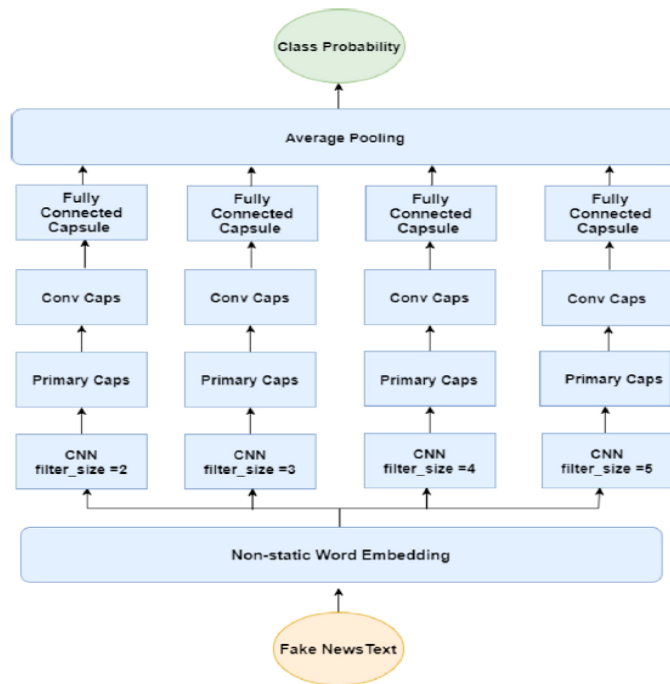
Εικόνα 117. Αλγόριθμος δικτύου προσοχής Graph Attention Network (GAT)³⁰.

Ο Goldani χρησιμοποίησε τον αλγόριθμο νευρωνικού δικτύου κάψουλας (*capsule neural network*) (Choi & Kim, n.d.; W. Zhao et al., n.d.) (Εικόνα 118) για την εκμάθηση πιο λεπτομερών και εκτεταμένων αναπαραστάσεων κειμένου διαφορετικών επιπέδων η λέξεων σύμφωνα με το μέγεθος του κειμένου (Brazoveanu & Andonie, 2021; Goldani et al., 2021).

Τα χαρακτηριστικά που εξάγονται ενσωματώνονται σε κάψουλες-ομάδες νευρώνων. Για τον εντοπισμό των ψευδών ειδήσεων ο Goldani υιοθέτησε αρκετά παράλληλα δίκτυα κάψουλας. Τα συγκεκριμένα δίκτυα περιλαμβάνουν το συνελκτικό επίπεδο ngram, το πρώτο επίπεδο κάψουλας, το πρώτο συνελκτικό δίκτυο κάψουλας, ένα εμπροσθοβαρές επίπεδο δικτύου (FFNN) και στο τελικό στάδιο το επίπεδο εισροής. Το συγκεκριμένο μοντέλο δηλαδή μπορεί να μάθει αναπαραστάσεις κειμένου με διαφορετικά ngrams (αριθμούς συχνοτήτων λέξεων) ανάλογα με το μέγεθος του κειμένου (Goldani et al., 2021).

Οι Brazoveanu και Andonie υιοθέτησαν το συγκεκριμένο δίκτυο από παλαιότερα αντίστοιχα μοντέλα (Fentaw & Kim, 2019; J. Kim et al., 2020) μαζί με τον αλγόριθμο διπλής κατεύθυνσης BiGRU για τον εντοπισμό των ψευδών ειδήσεων με πολύ καλά αποτελέσματα σε σχέση με παραδοσιακούς αλγόριθμους (πχ. SVM) (Brazoveanu & Andonie, 2021).

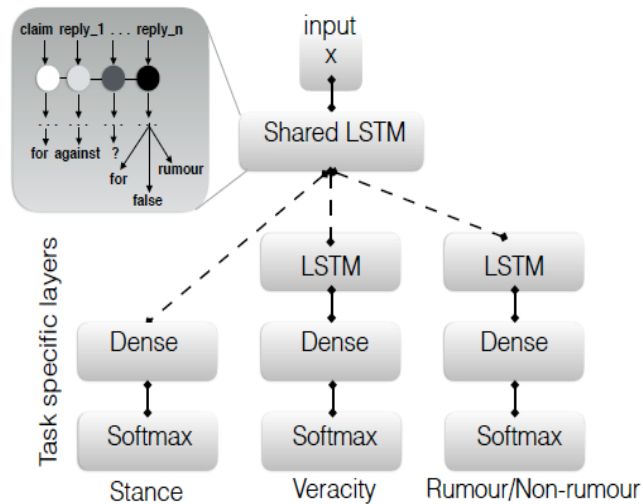
³⁰ Στην εικόνα τα χρώματα δεσμών προσδιορίζουν τις τάξεις ενώ το πάχος άκρης, τους ενσωματωμένους συντελεστές δεσμών προσοχής.



Εικόνα 118. Αλγόριθμος δικτύου κάψουλας (CapsNet).

Ο αλγόριθμος *Multitask Learning* (Kochkina, n.d.) (Εικόνα 119) είναι ένα μοντέλο που βελτιστοποιεί τη σχέση ανάμεσα σε δύο εργασίες ως προς τον εντοπισμό της αξιοπιστίας, τον εντοπισμό διαφορετικών δημοσιεύσεων οι οποίες συμφωνούν, προβληματίζονται, αρνούνται και σχολιάζουν έναν ισχυρισμό (stance) και την πρόβλεψη της αξιοπιστίας (Kochkina et al., 2018; Q. Zhang et al., 2019). Ο Zhang τον χρησιμοποίησε τον αλγόριθμο συγκρίνοντας το με άλλους γνωστούς αλγόριθμους (CNN, SVM) χρησιμοποιώντας περιεχόμενο ισχυρισμών και απαντήσεις χρηστών με το μοντέλο βαθιάς μάθησης Bayes.

Σύμφωνα με την Kochkina ο αλγόριθμος πολυεργασιών εκπαίδευσης (multitask learning) αναφέρεται ως κοινή εκμάθηση αρκετών εργασιών που σχετίζονται με μια αναπαράσταση. Η αποτελεσματικότητα του αλγόριθμου έγκειται στην αύξηση του μεγέθους των δεδομένων της βάσης δεδομένων εκπαίδευσης (προσθήκη επιπλέον βάσεων), αφού το μοντέλο μαθαίνει κοινές αναπαραστάσεις για πολλαπλές εργασίες αποφεύγοντας την υπερεκπαίδευση (Kochkina et al., 2018).



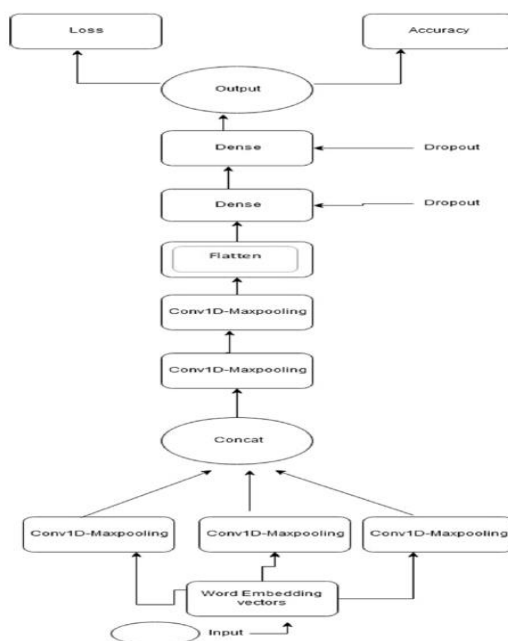
Εικόνα 119. Αλγόριθμος Multitask Learning.

Το μοντέλο *FNDNET* (Εικόνα 120) είναι ένα μοντέλο αυτόματης εκμάθησης επιλεγμένων χαρακτηριστικών μέσω πολλαπλών κρυφών επιπέδων (Kaliyar et al., 2020). Το μοντέλο αποτελείται από το συνελκτικό επίπεδο (convolutional layer), το επίπεδο μέγιστης εισροής (max pooling layer), το επίπεδο λείανσης (flatten layer), το πλήρες συνδεδεμένο επίπεδο (dense layer), την τεχνική κανονικοποίησης (Dropout), και τη λειτουργία ενεργοποίησης RELU. Το συνελκτικό επίπεδο αποτελείται από φίλτρα για περιορισμένη εισαγωγή δεδομένων. Τα συγκεκριμένα φίλτρα δρουν ως εκμαθητές όταν εντοπίζουν ένα συγκεκριμένο είδος χαρακτηριστικού στην είσοδο (Kaliyar et al., 2020; Zhong et al., 2019).

Ο Kaliyar χρησιμοποίησε τρία παράλληλα συνελκτικά δίκτυα με διαφορετικά μεγέθη τα οποία προσδιορίζουν τον αριθμό των λέξεων (υποβοήθηση με άγνωστες λέξεις σε διαφορετικά μοντέλα ενσωματώσεων). Το *επίπεδο μέγιστης εισροής* περνά την αξιόπιστη πληροφορία από το προηγούμενο στο επόμενο επίπεδο. Ο συγγραφέας ενσωμάτωσε τρία επίπεδα μέγιστης εισροής τα οποία διαχωρίζουν την έξοδο από τα συνελκτικά επίπεδα. Το *επίπεδο λείανσης* μετατρέπει τα χαρακτηριστικά από τον συσσωρευτή και τα οπτικοποιεί σε μία μοναδική στήλη. Το *πλήρες συνδεδεμένο επίπεδο* περιλαμβάνει μία γραμμική λειτουργία όπου κάθε είσοδος συνδέεται με κάθε έξοδο μέσω βαρών. Επιπλέον, υιοθέτησε δύο συνδεδεμένα επίπεδα με το πρώτο να παίρνει τα δεδομένα εξόδου από το επίπεδο λείανσης και να τα περνά στο πρώτο επίπεδο και μετά στο δεύτερο το οποίο προβλέπει το τελικό αποτέλεσμα.

Σύμφωνα με τον Zong, η τεχνική κανονικοποίησης μειώνει την πολυπλοκότητα του μοντέλου και αποτρέπει την υπερμάθηση (Zhong et al., 2019). Ενώ η λειτουργία Relu

απομακρύνει τις αρνητικές τιμές δίνοντάς τους την τιμή 0. Αποτελεί την πιο αποτελεσματική λειτουργία στην βαθέα μάθηση λόγω των αποτελεσμάτων της.



Εικόνα 120. Μοντέλο FNDNET.

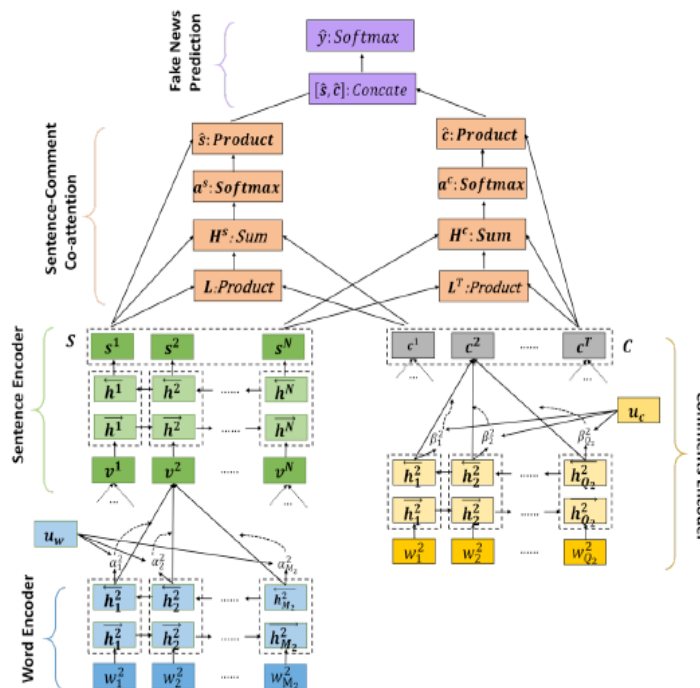
Το μοντέλο *DEFEND* (Shu et al., n.d.) (Εικόνα 121) είναι ένα μοντέλο εντοπισμού ψευδών ειδήσεων το οποίο κωδικοποιεί τις ειδήσεων, τα σχόλια των χρηστών και δίνει έμφαση στη σημασιολογική ομοιότητα των προτάσεων και των σχολίων. Παράλληλα πραγματοποιεί ταυτόχρονη εκμάθηση προτάσεων και σχολίων (Shu, Cui, et al., 2019). Αποτελείται από τον κωδικοποιητή των ειδήσεων, τον κωδικοποιητή χρηστών και την πρόταση προσοχής σχολίων.

Ο κωδικοποιητής ειδήσεων υιοθετεί ένα ιεραρχικό νευρωνικό δίκτυο για τη μοντελοποίηση των αναπαραστάσεων των λέξεων και των προτάσεων ακολουθώντας την οπτική του Chen (H. Chen et al., 2016). Αποτελείται από τον κωδικοποιητή λέξης όπου γίνεται η εκμάθηση των αναπαραστάσεων μέσω του αλγόριθμου δικτύου RNN και του GRU για την αποτύπωση των πληροφοριών σε πιο μεγάλο χρονικό διάστημα, εξοικονόμηση περισσότερης μνήμης (Z. Yang et al., 2016) και την αποτύπωση της συμπληρωματικής πληροφόρησης μέσω μοντελοποίησης της GRU σε διπλή κατεύθυνση (Bahdanau et al., 2016). Στον κωδικοποιητή πρότασης μέσω της χρήσης του RNN και του BiGRU κωδικοποιείται κάθε πρόταση ειδήσεων μαζί με την συμπληρωματική πληροφόρηση για την εκμάθηση των αναπαραστάσεων.

Στο κωδικοποιητή σχολίων χρηστών (π.χ. αντιδράσεις συναισθηματικές) υιοθετείται η σημασιολογική προσέγγιση για τον εντοπισμό των ψευδών ειδήσεων. Εξάγονται σχόλια μικρής έκτασης από τα κοινωνικά μέσα μέσω του RNN για την εκμάθηση των μη παρατηρήσιμων χαρακτηριστικών επαναχρησιμοποιώντας και τον αλγόριθμο διπλής κατεύθυνσης.

Στον μηχανισμό προσοχής επιλέγονται προτάσεις από τις ειδήσεις και σχόλια χρηστών τα οποία εξηγούν τον λόγο που ένα μέρος της είδησης είναι ψευδές. Με τον μηχανισμό προσοχής το βάρος εστιάζεται στη σημασιολογική σχέση των προτάσεων και των σχολίων και στην περαιτέρω παράλληλη εκμάθηση των βαρών του μηχανισμού προσοχής των προτάσεων και των σχολίων (Lu et al., 2017).

Στο τελευταίο επίπεδο (μέγιστης εισροής, softmax) ενσωματώνονται τα στοιχεία από τις αναπαραστάσεις και τα σχόλια των προηγούμενων επιπέδων και γίνεται η πρόβλεψη μέσω υπολογισμού της πιθανότητας μία είδηση να είναι ψευδής (1) ή αληθής (0).



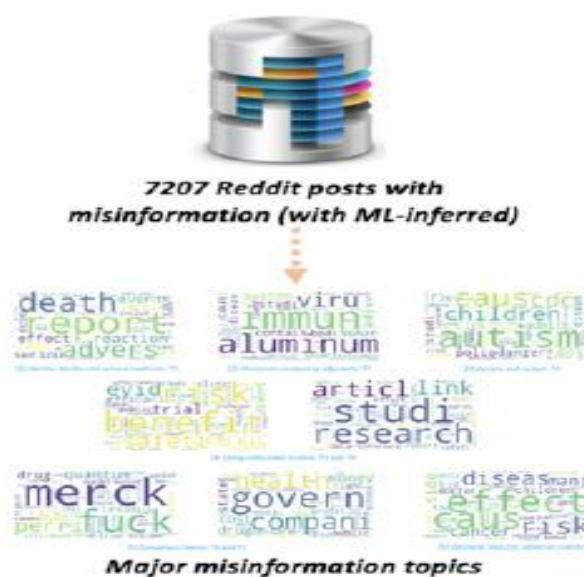
Εικόνα 121. Μοντέλο DEFEND.

4.5 Θεματικά Χαρακτηριστικά

Σύμφωνα με τον Du οι αλγόριθμοι μηχανικής και βαθιάς μάθησης δεν έχουν την δυνατότητα να κατηγοριοποιούν τα είδη της παραπληροφόρησης παρά μόνο να προσδιορίζουν τις δημοσιεύσεις παραπληροφόρησης σε κοινωνικά δίκτυα όπως το Reddit (Du et al., 2021). Για αυτό τον λόγο ο συγγραφέας υιοθέτησε ένα θεματικό μοντέλο (Bitern

topic model) προκειμένου να συγκεκριμενοποιήσει και να οπτικοποιήσει τα βασικά θέματα της παραπληροφόρησης στις δημοσιεύσεις.

Τα θεματικά μοντέλα είναι στατιστικά μοντέλα τα οποία έχουν ως σκοπό τη δημιουργία ομάδων αφηρημένων θεμάτων που εμφανίζονται σε μία συλλογή από κείμενα (π.χ. “General vaccine adverse events”, “The HPV vaccine is unnecessary & unsafe”, “Conspiracy Theory”, “Citing unfounded studies”, “Vaccine deaths and serious reactions”). Το αλγοριθμικό μοντέλο χρησιμοποιείται αφού οι αλγόριθμοι μηχανικής μάθησης προσδιορίσουν την παραπληροφόρηση. Στη συνέχεια το μοντέλο εξάγει τη συχνότητα κάθε θέματος βασιζόμενη στη συχνότητα των λέξεων (Biterm) (μοτίβα) (π.χ. εξαγωγή δύο λέξεων μεμονωμένων από το ίδιο θέμα) εντός μίας συλλογής κειμένων (Χιαοηui, 2013). Για την οπτικοποίηση των θεμάτων χρησιμοποιήθηκαν λέξεις νέφους όπου το μέγεθος κάθε λέξης είναι σχετικό με την πιθανότητα εμφάνισής της σε δημοσιεύσεις που σχετίζονται με το θέμα (Du et al., 2021)³¹ (Εικόνα 122).



Εικόνα 122. Θεματικό Μοντέλο (Biterm Topic Model).

4.6 Συναισθηματικά Χαρακτηριστικά

Συναισθηματικά χαρακτηριστικά όπως θυμός, άγχος, επιρροή, βρισιές αλλά και μετρικές συναισθήματος (ουδέτερο, θετικό, αρνητικό) ήταν μερικά από τα χαρακτηριστικά που χρησιμοποιήθηκαν από τους συγγραφείς για τον εντοπισμό των ψευδών ειδήσεων πάντα με χρήση εξειδικευμένων τεχνικών μηχανικής μάθησης και NLP (LIWC, NELA).

³¹ Δες και (“Generating Word Cloud in Python,” 2018)

Οι Horne, Kausar, και Sotirakou χρησιμοποίησαν ειδικά λεξικά συναισθήματος όπως τη LIWC αλλά και το *NRC Affect Lexicon* (S. Mohammad, 2020; Pennebaker, n.d.) για να εντοπίσουν τα χαρακτηριστικά των ψευδών ειδήσεων και ένα από αυτά ήταν και ο *θυμός* (B. D. Horne et al., 2020; Kausar et al., 2020; Sotirakou et al., 2021). Σύμφωνα με την Sotirakou οι ψευδείς ειδήσεις αποτυπώνουν *περισσότερη αρνητικότητα* από ότι οι αληθείς και το περιεχόμενο παραπλάνησης *αποτυπώνεται με θυμό* προκειμένου να εξοργίσει το κοινό (Sotirakou et al., 2021). Εκτός από το συναίσθημα του θυμού η Sotirakou μέτρησε και τα *συναίσθημα φόβου, χαράς και λύπης* που αποτυπώνονται στα κείμενα βασιζόμενη σε θεωρίες συναισθήματος (S. M. Mohammad, 2017) αλλά και *συναίσθημα επιρροής* (κτητικότητα, έξαψη, κυριαρχία) υιοθετώντας την οπτική του Russell (Russell, 2003) με χρήση του λεξικού NRC Vader (S. Mohammad, n.d.).

Συναίσθημα επιρροής μέτρησαν και οι Horne και Barrón-Cedeño μέσω των λεξικών Vader και LIWC μέσα στο κείμενο (Barrón-Cedeño et al., 2019; B. D. Horne et al., 2020). Η διαφορά ανάμεσα στο συναίσθημα και την επιρροή προσδιορίζεται στον Shouse. Το συναίσθημα αφορά την εκδήλωση ενός συναισθήματος το οποίο μπορεί να είναι αληθινό ή ψεύτικο και είναι προσωπικό και βιωματικό αφού κάθε άνθρωπος μπορεί να εξηγήσει (κατηγοριοποιήσει) τα συναίσθημα του εκτός από τα παιδιά (έλλειψη γλώσσας και βιωμάτων). Ενώ η επιρροή είναι μία ασυνείδητη εμπειρία σχετιζόμενη με την ένταση και η έννοια της είναι πιο αφηρημένη γιατί δεν μπορεί να κατανοηθεί από τη γλώσσα αλλά και γιατί η επιρροή υπάρχει εκτός συνείδησης (π.χ. εκφράσεις μύων προσώπου, αναπνευστικού συστήματος, φωνές) (Shouse, 2005).

Το *συναίσθημα της λύπης* εντοπίστηκε και από τους Kausar και τη Sotirakou με τη χρήση των λεξικών LIWC, NELA, NRC (Kausar et al., 2020; S. M. Mohammad, 2017; Sotirakou et al., 2021). Οι *βρισιές* είναι άλλο ένα χαρακτηριστικό των ψευδών ειδήσεων το οποίο εντοπίστηκε από τους Horne, Kausar, Sotirakou και Sheikhi (π.χ. στο διάολο, χάσου από δω, γαμώτο). Οι βρισιές υιοθετήθηκαν από τον Sheikhi ως χαρακτηριστικό για την κατηγοριοποίηση των τίτλων των ειδήσεων αφού τα επαγγελματικά πρακτορεία ειδήσεων αποφεύγουν αυτές τις λέξεις στους τίτλους (B. D. Horne et al., 2020; Kausar et al., 2020; Sheikhi, 2021; Sotirakou et al., 2021; Tausczik & Pennebaker, 2010). Ενώ ο *βαθμός ευτυχίας* εντοπίστηκε από τον Barrón-Cedeño ως ένα ακόμη δομικό χαρακτηριστικό των ψευδών ειδήσεων (Barrón-Cedeño et al., 2019). Το συγκεκριμένο χαρακτηριστικό σύμφωνα με τον συγγραφέα έχει υιοθετηθεί και από τον Horne (B. Horne et al., 2018).

Ένα ακόμη χαρακτηριστικό των ψευδών ειδήσεων είναι και οι *αντιθέσεις συναισθήματος (από υψηλό σε χαμηλό) (Polarity)* (Abonizio et al., 2020; Ahmad et al., 2020;

Borges et al., 2019; Braşoveanu & Andonie, 2021; Ozbay & Alatas, 2020; Papadopoulou et al., 2019; Reis et al., 2019; Saikh et al., 2019). Οι Braşoveanu και Andonie ενσωμάτωσαν αντιθέσεις συναισθήματος σε επίπεδο πρότασης τις οποίες υπολόγισαν μέσω της βιβλιοθήκης ανοιχτού κώδικα TextBlob (*TextBlob: Simplified Text Processing — TextBlob 0.16.0 Documentation*, n.d.) (Εικόνα 123).

```
>>> testimonial.sentiment
Sentiment(polarity=0.3916666666666666, subjectivity=0.4357142857142857)
>>> testimonial.sentiment.polarity
0.3916666666666666
```

Εικόνα 123. Αντιθέσεις συναισθήματος (Polarity)- TextBlob.

Ο Borges, βασίστηκε στην παρουσία λέξεων με *υψηλή ένταση συναισθήματος σε επικεφαλίδες, σώμα κειμένου και πρώτες δύο προτάσεις* (π.χ. fake=ψευδές, fraud=απάτη, hoax=ισχυρισμός, false=λάθος) (*FakeNewsChallenge / Fnc-1-Baseline*, 2017a; *FakeNewsChallenge / Fnc-1-Baseline*, 2017b) αλλά και σε *βαθμούς αντίθεσης από το σώμα και τους τίτλους με βάση λεξικό αντιθέσεων συναισθήματος (Vader Polarity lexicon)*(Εικόνα 124) (Hutto, 2014) (Borges et al., 2019).

Word	Polarity
The	0
movie	0
was	0
really	0
good	1
.	0

Εικόνα 124. Βαθμός αντίθεσης συναισθήματος (Polarity).

Ο Abonizio χρησιμοποίησε το λογισμικό NLP Polyglot (*AboSamoor / Polyglot: Multilingual Text (NLP) Processing Toolkit*, 2020)³² για να μετρήσει τον βαθμό αντίθεσης στο συναίσθημα στις ψευδείς ειδήσεις σε πολυγλωσσικό περιβάλλον (Abonizio et al., 2020; Al-Rfou' et al., 2013). Ο Saikh προσδιόρισε την παρουσία αρνήσεων όπως «όχι, ποτέ, αρνούμαι» ως χαρακτηριστικό αντίθεσης σε μια πρόταση (Saikh et al., 2019) με τη βοήθεια της

³² Δες και (*Welcome to Polyglot's Documentation! — Polyglot 16.07.04 Documentation*, n.d.)

επισήμανσης των ονομαστικών οντοτήτων και παρουσίας λέξεων κλειδιών σε προτάσεις (Saikh et al., 2017).

Οι Ahmad, Papadopoulou, και Cohen *μέτρησαν τον αριθμό των θετικών και αρνητικών λέξεων συναισθήματος που υπάρχουν σε ένα κείμενο* (Ahmad et al., 2020; ή βίντεο (Papadopoulou et al., 2019) μέσω του λεξικού της LIWC και μετρήσεων συναισθηματικής ανάλυσης από το λογισμικό Textblob.

Για τη *μέτρηση της πολικότητας* οι Barrón-Cedeño, Meel & Viswakarma, Horne, Yu, Ibrishimova & Li, Papadopoulou, Braşoveanu & Andonie, Reis, εστίαστηκαν σε *θετικά συναισθήματα* (π.χ. -1 αρνητικό, +1 πολύ θετικό) (Hutto, 2014) υιοθετώντας μετρικές από LIWC και NELA συνδυαστικά με λεξικά όπως το *Vader* το οποίο είναι πιο ακριβές στη συναισθηματική μέτρηση αλλά και μέσω της εφαρμογής Google NLP API (*Natural Language API Basics*, n.d.) (Ahmad et al., 2020; Barrón-Cedeño et al., 2019; Braşoveanu & Andonie, 2021; Cohen et al., 2020; B. D. Horne et al., 2020; Ibrishimova & Li, 2020; Meel & Vishwakarma, 2020; Papadopoulou et al., 2019; Reis et al., 2019; Y. Yu et al., 2020). Πιο συγκεκριμένα το *λεξικό Vader* περιλαμβάνει ειδικό λεξικό συναισθήματος και χρησιμοποιείται για ανάλυση των συναισθημάτων στα κοινωνικά δίκτυα. Επιπλέον ενσωματώνει λέξεις και emoticons από τη LIWC και οι μετρικές του συναισθήματος είναι θετικό, αρνητικό, ουδέτερο.

Το αρνητικό συναίσθημα μετρήθηκε από τους Barrón-Cedeño, Meel & Viswakarma, Horne, Kausar, Yu, Ibrishimova & Li, Papadopoulou, Braşoveanu & Andonie, Cohen, Reis (Ahmad et al., 2020; Barrón-Cedeño et al., 2019; Braşoveanu & Andonie, 2021; Cohen et al., 2020; B. D. Horne et al., 2020; Ibrishimova & Li, 2020; Kausar et al., 2020; Meel & Vishwakarma, 2020; Papadopoulou et al., 2019; Reis et al., 2019; Y. Yu et al., 2020) και το ουδέτερο από τους Yu και Ibrishimova και Li (Ibrishimova & Li, 2020; Y. Yu et al., 2020).

Συγγραφείς όπως οι Varshney & Viswakarma *μέτρησαν τον συνολικό βαθμό συναισθήματος σε όλες τις παραγράφους (θετικό, αρνητικό, ουδέτερο) αλλά και αναλογίες θετικού και αρνητικού συναισθήματος σε σχέση με το θετικό συναίσθημα* (Varshney & Vishwakarma, 2021).

Οι *βαθμοί υποκειμενικότητας και συναισθήματος* μετρήθηκαν από τους Barrón-Cedeño, Papadopoulou, Reis, και Braşoveanu και Andonie μέσω του TextBlob API αφού δίνουν τη δυνατότητα να εντοπιστεί με περισσότερη ακρίβεια η υποκειμενικότητα ενός κειμένου (π.χ. 0=πολύ αντικειμενικό, 1=πολύ υποκειμενικό) (Barrón-Cedeño et al., 2019; Braşoveanu & Andonie, 2021; Papadopoulou et al., 2019; Reis et al., 2019).

Η Sotirakou αντίθετα μέτρησε το συναίσθημα με τη χρήση λεξικών έντονου συναισθήματος (*NRC emotion intensity lexicon*) αλλά και επιρροής (*Affect-NRC VAD lexicon*) (Sotirakou et al., 2021).

Το λεξικό *NRC Emotion Intensity lexicon* (πρώην *NRC Affect Intensity Lexicon*) (*NRC Affect Intensity Lexicon*, n.d.) (Εικόνα 125) περιλαμβάνει κλίμακα χειρίστου, βέλτιστου (π.χ. βαθμός 1=λέξη υψηλού συναισθήματος, βαθμός 0=λέξη κατώτατου συναισθήματος) η οποία προκύπτει από λίστα αγγλικών λέξεων με βαθμούς έντασης οκτώ συναισθημάτων (*anger*=θυμός, *anticipation*=προσμονή, *disgust*=αηδία, *fear*=φόβος, *joy*=χαρά, *sadness*=λύπη, *surprise*=έκπληξη, *trust*=εμπιστοσύνη). Σύμφωνα με την συγγραφέα οι ψευδείς ειδήσεις έχουν περισσότερη αρνητικότητα και θυμό από ότι οι αληθείς.

Το λεξικό *NRC Vader* (*Saif | VAD Lexicon*, 2022) (Εικόνα 126) περιλαμβάνει περισσότερες από 2000 λέξεις αγγλικές και βαθμολογίες για την έλξη (θετικό), την αποστροφή (*valence*) (αρνητικό), την έξαψη (*arousal*), την επιβολή (*dominance*) (π.χ. συμπόσιο-θετική, νευρικός περισσότερη έξαψη από τεμπέλης) (Sotirakou et al., 2021). Στα συγκεκριμένα λεξικά, η κλίμακα προσδιορίζεται μέσω πληθοπορισμού όπου κάθε ζεύγος λέξεων (χειρότερης-καλύτερης κλίμακας) βαθμολογείται από 6 άτομα (*Sentiment and Emotion Lexicons*, n.d.).

Word	Anger	Word	Fear	Word	Joy	Word	Sadness
<i>outraged</i>	0.964	<i>horror</i>	0.923	<i>sohappy</i>	0.868	<i>sad</i>	0.844
<i>brutality</i>	0.959	<i>horrified</i>	0.922	<i>superb</i>	0.864	<i>suffering</i>	0.844
<i>satanic</i>	0.828	<i>hellish</i>	0.828	<i>cheered</i>	0.773	<i>guilt</i>	0.750
<i>hate</i>	0.828	<i>grenade</i>	0.828	<i>positivity</i>	0.773	<i>incest</i>	0.750
<i>violence</i>	0.742	<i>strangle</i>	0.750	<i>merrychristmas</i>	0.712	<i>accursed</i>	0.697
<i>molestation</i>	0.742	<i>tragedies</i>	0.750	<i>bestfeeling</i>	0.712	<i>widow</i>	0.697
<i>volatility</i>	0.687	<i>anguish</i>	0.703	<i>complement</i>	0.647	<i>infertility</i>	0.641
<i>eradication</i>	0.685	<i>grisly</i>	0.703	<i>affection</i>	0.647	<i>drown</i>	0.641
<i>cheat</i>	0.630	<i>cuthroat</i>	0.664	<i>exalted</i>	0.591	<i>crumbling</i>	0.594
<i>agitated</i>	0.630	<i>pandemic</i>	0.664	<i>woot</i>	0.588	<i>deportation</i>	0.594
<i>defiant</i>	0.578	<i>smuggler</i>	0.625	<i>money</i>	0.531	<i>isolated</i>	0.547
<i>coup</i>	0.578	<i>pestilence</i>	0.625	<i>rainbow</i>	0.531	<i>unkind</i>	0.547
<i>overbearing</i>	0.547	<i>convict</i>	0.594	<i>health</i>	0.493	<i>chronic</i>	0.500
<i>deceive</i>	0.547	<i>rot</i>	0.594	<i>liberty</i>	0.486	<i>injurious</i>	0.500
<i>unleash</i>	0.515	<i>turbulence</i>	0.562	<i>present</i>	0.441	<i>memorials</i>	0.453
<i>bile</i>	0.515	<i>grave</i>	0.562	<i>tender</i>	0.441	<i>surrender</i>	0.453
<i>suspicious</i>	0.484	<i>failing</i>	0.531	<i>warms</i>	0.391	<i>beggar</i>	0.422
<i>oust</i>	0.484	<i>stressed</i>	0.531	<i>gesture</i>	0.387	<i>difficulties</i>	0.421
<i>ultimatum</i>	0.439	<i>disgusting</i>	0.484	<i>healing</i>	0.328	<i>perpetrator</i>	0.359
<i>deleterious</i>	0.438	<i>hallucination</i>	0.484	<i>tribulation</i>	0.328	<i>hindering</i>	0.359

Εικόνα 125. Μετρήσεις λέξεων επιρροής-λεξικό NRC Affect.

Entries with Highest and Lowest Scores in the VAD Lexicon

Dimension	Word	Score [↑]	Word	Score [↓]
valence	<i>love</i>	1.000	<i>toxic</i>	0.008
	<i>happy</i>	1.000	<i>nightmare</i>	0.005
	<i>happily</i>	1.000	<i>shit</i>	0.000
arousal	<i>abduction</i>	0.990	<i>mellow</i>	0.069
	<i>exorcism</i>	0.980	<i>siesta</i>	0.046
	<i>homicide</i>	0.973	<i>napping</i>	0.046
dominance	<i>powerful</i>	0.991	<i>empty</i>	0.081
	<i>leadership</i>	0.983	<i>frail</i>	0.069
	<i>success</i>	0.981	<i>weak</i>	0.045

Εικόνα 126. Μετρήσεις λεξικού NRC-Vader.

Ψυχολογικά χαρακτηριστικά τα οποία ενσωματώνονται στη μέθοδο LIWC όπως γνωστικές (αντιληπτικές) λέξεις υιοθετήθηκαν από τους Barrón-Cedeño, Kausar, Reis, Shu, Meel και Viswakarma, Gravanis, Gôlo, Sotirakou (Barrón-Cedeño et al., 2019; Gôlo et al., 2021; Gravanis et al., 2019; Kausar et al., 2020; Meel & Vishwakarma, 2020; Reis et al., 2019; Shu, Cui, et al., 2019; Sotirakou et al., 2021). Γνωστικές λέξεις όπως αλλά, όχι, εάν, ξέρω υιοθετήθηκαν από τους Barrón-Cedeño και Kausar αλλά και λέξεις επιρροής, διαφωνίας (π.χ. θα ήθελα, θα μπορούσα), διστακτικότητας (π.χ. ίσως υποθέτω), βεβαιότητας (π.χ. πάντα, ποτέ), κοινωνικές (π.χ. μιλώ, φίλος), επιρροής συναισθήματος (π.χ. χαρούμενος, άσχημος, πικραμένος), αίσθησης (π.χ. ήχοι, όσφρηση, όραση, φυσικών αισθήσεων, οπτικές λεπτομέρειες), αντίληψης (π.χ. σκέφτομαι, υπολογίζω) (Gravanis et al., 2019; Newman et al., 2003; L. Zhou et al., 2004) αλλά και λύπης (επιρροή) (Gôlo et al., 2021). Ενώ οι Kausar, Gravanis, Gôlo, Shu μέτρησαν και τα θετικά και αρνητικά συναισθήματα (Gôlo et al., 2021; Kausar et al., 2020; Shu, Cui, et al., 2019).

Ένα ακόμη χαρακτηριστικό το οποίο υιοθετήθηκε από την παρούσα βιβλιογραφία είναι η ηθικότητα όπου έχει ως βάση της τη θεωρία ηθικών θεμελίων (*Moral foundation theory*) (*Moral Foundations Theory | MoralFoundations.Org*, n.d.). Η συγκεκριμένη θεωρία έχει επηρεαστεί από την Κοινωνική Ψυχολογία και αφορά τη θεμελίωση των ηθικών αξιών για κάθε πολιτισμό. Αποτελείται από πέντε ηθικά θεμέλια (αξίες): Φροντίδα/πόνος (*Care/harm*), δικαιοσύνη/απιστία (*Fairness/cheating*), εμπιστοσύνη/προδοσία (*Loyalty/betrayal*), εξουσία/υποταγή (*Authority/subversion*), αγιοσύνη-αγνότητα/δυσμένεια (*Sanctity/degradation*) (Barrón-Cedeño et al., 2019; B. D. Horne et al., 2020; Kausar et al., 2020). Οι δημιουργοί της θεωρίας (Atari et al., 2020; Graham et al., 2013; Haidt & Graham, 2007) προσθέτουν και ένα επιπλέον θεμέλιο της ελευθερίας/καταπίεσης (*Liberty/oppression*).

Η *θεωρία ηθικών θεμελίων* ουσιαστικά είναι ένα λεξικό ηθικών λέξεων και αδυναμιών το οποίο χρησιμοποιείται συμπληρωματικά με τη LIWC. Το *πρώτο ηθικό θεμέλιο* φροντίδα/πόνος σχετίζεται με την εξελικτική θεωρία με συστήματα στοργικότητας αλλά και συναισθήματα ενσυναίσθησης (ή μη) του πόνου των άλλων. Σχετίζεται με αξίες όπως καλοσύνη, ευγένεια, φροντίδα. Το *δεύτερο ηθικό θεμέλιο* δικαιοσύνη/απιστία αφορά την εξελικτική διαδικασία μίας πραγματιστικής προσέγγισης και αφορά ιδέες όπως η δικαιοσύνη (ισορροπία), τα δικαιώματα και η αυτονομία.

Το *τρίτο ηθικό θεμέλιο* εμπιστοσύνη/προδοσία σχετίζεται με θέσπιση συμμαχιών μέσα στα πλαίσια της φυλετικής εξέλιξης και καταδεικνύει αξίες όπως ο πατριωτισμός και η αυτοθυσία σε μια ομάδα («ένας για όλους και όλοι για έναν»). Το *τέταρτο θεμέλιο* εξουσία/υποταγή δημιουργήθηκε από ιστορικές κοινωνικές αλληλεπιδράσεις. Υπονοεί ηγετικές αξίες και αξίες συντροφικότητας συμπεριλαμβανομένου του σεβασμού στην εξουσία και τις παραδόσεις. Το *πέμπτο ηθικό θεμέλιο* αγιοσύνη-αγνότητα/δυσμένεια δημιουργήθηκε από την ψυχολογία της αποστροφής και της μόλυνσης. Υπονοεί θρησκευτικές αντιλήψεις και έναν πιο ευγενή (λιγότερο σαρκικό) τρόπο ζωής.

Το συμπληρωματικό θεμέλιο *ελευθερία/καταπίεση* αφορά αισθήματα αποστροφής και αντίδρασης σε αυτούς που εξουσιάζουν και περιορίζουν την ελευθερία (π.χ. bullying, οι καταπιεστές ωθούν τον κόσμο σε αλληλεγγύη και κοινή δράση ενάντια στον καταπιεστή).

Ο Kausar ήταν ένας από τους συγγραφείς που υιοθέτησαν τα χαρακτηριστικά της συγκεκριμένης θεωρίας όπως το πόνος, απιστία, αγνότητα, εμπιστοσύνη συμπληρωματικά με τα υπόλοιπα χαρακτηριστικά (γλωσσολογικά, χαρακτηριστικά γραφής) (Kausar et al., 2020).

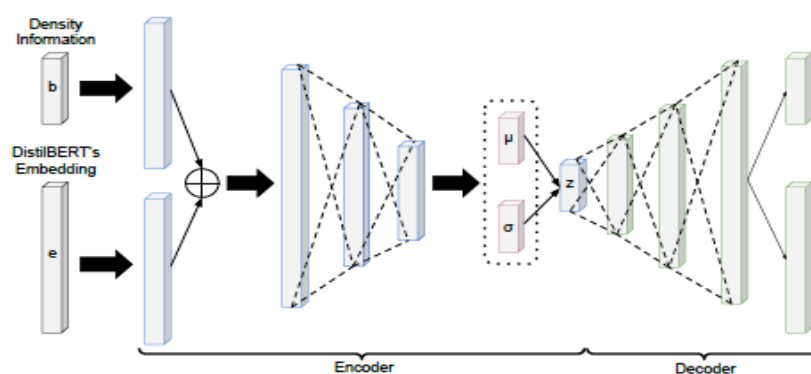
4.7 Υβριδικά Χαρακτηριστικά

Οι Gôlo, Isaakidou, Yuliani, Shu, και Baeth και Aktas υιοθέτησαν υβριδικά μοντέλα τα οποία έχουν την δυνατότητα να μαθαίνουν πολλαπλές αναπαραστάσεις και είναι περισσότερο αποτελεσματικές από ότι οι μονομοντελικές (Baeth & Aktas, 2019; Gôlo et al., 2021; Isaakidou et al., 2021; S. Qian et al., 2021; Shu, Cui, et al., 2019; Ying et al., 2021). Μάλιστα σύμφωνα με τον Gôlo οι πολυμοντελικές αναπαραστάσεις συνεισφέρουν ως προς την απόδοση των αλγορίθμων μηχανικής μάθησης (Gôlo et al., 2021).

Οι συγγραφείς Gôlo, Ying, Qian υιοθέτησαν τον *πολυμοντελικό αυτοματοποιημένο κωδικοποιητή (Multimodal Variational AutoEncoder) (MVAE, n.d.)* ο οποίος χρησιμοποιεί σημασιολογικά και συντακτικά χαρακτηριστικά (DBERTML embeddings) μαζί με την πυκνότητα της πληροφορίας (density information) ως οπτική μοντελοποίηση προκειμένου

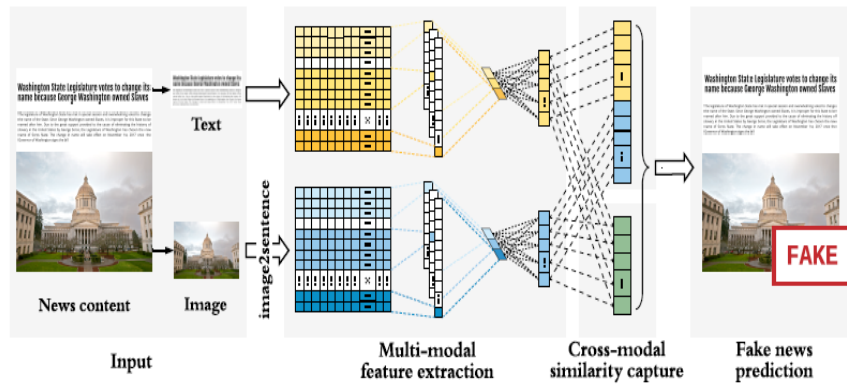
να αποτυπώσει τη χωρική κατανομή των θεμάτων (πολιτική, θρησκεία, επιστήμη). Για την οπτικοποίηση της πληροφορίας (κειμένων) χρησιμοποιείται ένας αλγόριθμος ομαδοποίησης. Μετά τη μοντελοποίηση θα γίνει η εκμάθηση των αναπαραστάσεων από ένα απλό νευρωνικό δίκτυο (FeedForward Neural Network) το οποίο κωδικοποιεί και αποκωδικοποιεί την πληροφορία.

Στο τελικό στάδιο ο MVAE αποτυπώνει μία καινούργια αναπαράσταση δεδομένων η οποία μετέπειτα θα εφαρμοστεί σε έναν αλγόριθμο (μονής κατηγορίας-One class learning) ο οποίος θα εντοπίσει κατά πόσο η είδηση θεωρείται ψευδής (Gòlo et al., 2021; Khattar et al., 2019; S. Qian et al., 2021; Ying et al., 2021) (Εικόνα 127).



Εικόνα 127. Πολυμοντελικός κωδικοποιητής.

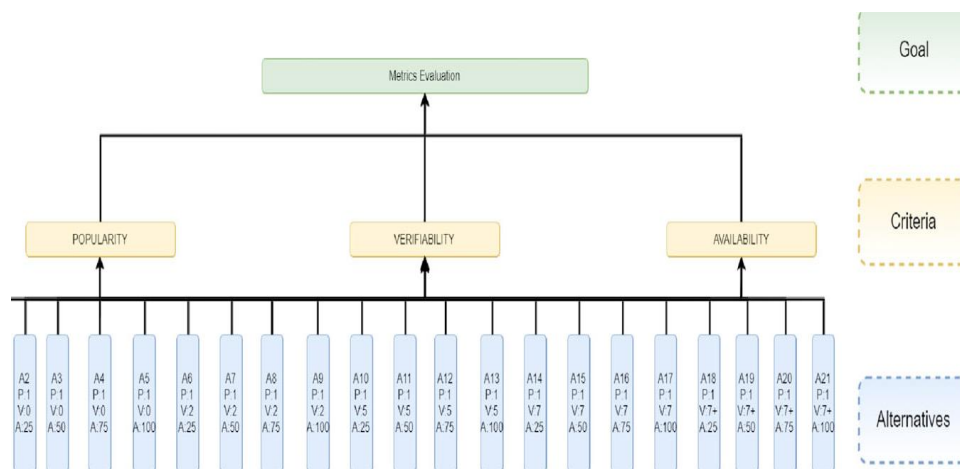
Το *Safe* (SAFE, n.d.) είναι ένα μοντέλο νευρωνικού δικτύου που χρησιμοποιεί πολυμοντελική πληροφόρηση για τον εντοπισμό των ψευδών ειδήσεων (S. Qian et al., 2021; Ying et al., 2021). Η εξαγωγή των χαρακτηριστικών των ειδήσεων γίνεται από κείμενο και εικόνα μέσω επέκτασης του νευρωνικού δικτύου Text-CNN. Η συγκεκριμένη επέκταση περιλαμβάνει ένα επιπλέον επίπεδο για την εξαγωγή χαρακτηριστικών από το άρθρο. Στη συνέχεια εξάγονται και τα χαρακτηριστικά εικόνας μέσω του Text-CNN και διερευνώνται οι σχέσεις μεταξύ τους (π.χ. αναγνώριση λάθους κειμένου, εικόνας και ανομοιοτήτων τους) (Εικόνα 128). Σύμφωνα με τη Zhou οι δημιουργοί των ψευδών ειδήσεων χρησιμοποιούν μη σχετικές εικόνες για ψευδείς δηλώσεις έτσι ώστε να προσελκύσουν το κοινό (X. Zhou et al., 2020).



Εικόνα 128. Μοντέλο SAFE.

Οι Baeth και Aktas υιοθέτησαν την μέθοδο της ασαφούς λογικής (*fuzzy logic*), μία εύκολη υπολογιστικά μαθηματική μέθοδο η οποία έχει εφαρμοστεί για την υιοθέτηση απόφασης με πολλαπλά κριτήρια. Το κύριο στοιχείο της ασαφούς λογικής περιλαμβάνει την χρήση συνεχών και διαφορετικών αποτελεσμάτων για την κατάταξή τους χρησιμοποιώντας διαφορετικές εναλλακτικές (Baeth & Aktas, 2019).

Οι συγγραφείς υιοθέτησαν την συγκεκριμένη μέθοδο χρησιμοποιώντας μετρικές χρηστών αλλά και μετρικές περιεχομένου. Πιο συγκεκριμένα υιοθέτησαν την ιστορική χαρτογράφηση αλλά και αξιολόγηση των πληροφοριών για τον εντοπισμό της «μολυσμένης πηγής» (παραβίαση κανόνων copyright). Το μοντέλο περιλαμβάνει 3 ιεραρχικά επίπεδα: το επίπεδο κορυφής αναπαριστά τον σκοπό (γενικός στόχος) (*goal*), το κριτήριο (*criterion*) (μεσαίο επίπεδο) το οποίο επηρεάζει τον σκοπό, και τις εναλλακτικές (*alternatives*) (τελευταίο επίπεδο) οι οποίες αναπαριστούν όλες τις πιθανές εναλλακτικές (Εικόνα 129).



Εικόνα 129. Μοντέλο ασαφούς λογικής (*fuzzy logic*).

Στο μεσαίο επίπεδο δημιουργούνται πίνακες με ζευγάρια για την σύγκριση και τον υπολογισμό της σημαντικότητας των βαρών σε κάθε κριτήριο απόφασης (Baeth & Aktas, 2019). Η κλίμακα που χρησιμοποιείται έχει υιοθετηθεί από τον Saaty (κλίμακα 1-9) όπου το 1 αναπαριστά την ίση σημασία ενώ το 9 την μέγιστη και μετέπειτα υπολογίζεται ο μέσος όρος για κάθε κριτήριο (π.χ. ίσης σημασίας 1 1 1, λίγο πιο σημαντικό 2/3, 1, 3/2, περισσότερο σημαντικό 3/2/, 2, 5/2) (Saaty, 1987). Τα ζεύγη χρησιμοποιούν τριαδικούς ασαφής αριθμούς οι οποίοι αναπαριστώνται από πραγματικούς αριθμούς (π.χ. 1, 1, 1 ή 2/3,1, 3/2) (Εικόνα 130).

	Popularity	Availability	Verifiability
Popularity	1	(1.5, 2, 2.5)	(1.5, 2, 2.5)
Availability	(0.4, 0.5, 0.67)	1	(0.4, 0.5, 0.67)
Verifiability	(0.4, 0.5, 0.67)	(0.67, 1, 1.5)	1

Εικόνα 130. Κλίμακα μεσαίου επιπέδου.

Η μέθοδος για να συνθέσει τις προτεραιότητες των κριτηρίων χρησιμοποιεί τους τριαδικούς αριθμούς για να εκφράσει τις ασαφείς σχέσεις ή τα γεγονότα ως ένα πρόσθετο επίπεδο. Τα κριτήρια αποτελούν *μετρήσεις αξιοπιστίας χρηστών* (δημοφιλία, διαθεσιμότητα, αξιοπιστία) *από δημοσιεύσεις σε twitter* αλλά και *λειτουργίες κοινωνικής ανατροφοδότησης* (λειτουργία «μου αρέσει», παρουσία φίλων που πάτησαν like η σχολίασαν) (θετική ανατροφοδότηση) ενώ η *μή ύπαρξη αλληλεπίδρασης, αρνητική ανατροφοδότηση*. Οι μετρήσεις αναπαριστώνται υπό μορφή γράφου (χρήση λογισμικών Komadu, WorkflowSim) όπου εισάγονται σε έναν αλγόριθμο εντοπισμού ψευδών ειδήσεων.

Στο τελευταίο επίπεδο (εναλλακτικές) πραγματοποιείται συνδυασμός εναλλακτικών βαθμών με τα βάρη των κριτηρίων (δημοφιλία, διαθεσιμότητα, αξιοπιστία) για να γίνει παραγωγή της συνολικής βαθμολόγησης για κάθε εναλλακτική (Εικόνα 131).

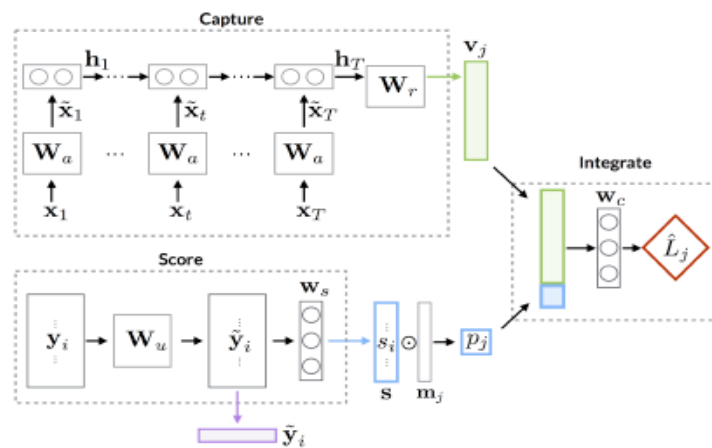
Alternatives	Popularity	Availability	Verifiability
A1	1	0	0
A2	0.89	0.11	0
A3	0.81	0.19	0
A4	0.73	0.27	0
A5	0.67	0.33	0
A6	0.76	0.09	0.15
A7	0.7	0.17	0.14
A8	0.64	0.23	0.13
A9	0.6	0.29	0.12
A10	0.66	0.08	0.26
A11	0.61	0.15	0.24
A12	0.57	0.21	0.22
A13	0.53	0.26	0.21
A14	0.58	0.07	0.35
A15	0.55	0.13	0.32
A16	0.51	0.19	0.30
A17	0.48	0.23	0.28
A18	0.52	0.06	0.41
A19	0.49	0.12	0.39
A20	0.46	0.17	0.37
A21	0.44	0.21	0.35

Εικόνα 131. Μετρικές βαρών για κάθε εναλλακτική.

Ο Shu χρησιμοποίησε ένα υβριδικό μοντέλο βαθιάς μάθησης (CSI) (*CSI-Code*, n.d.) ως μέτρο σύγκρισης με το Defend για τον εντοπισμό των ψευδών ειδήσεων. Το συγκεκριμένο μοντέλο χρησιμοποιεί πληροφόρηση από κείμενο, πηγή και απάντηση. Η αναπαράσταση γίνεται μέσω του νευρωνικού δικτύου μακροχρόνιας μνήμης (LSTM) με τη χρήση των αναπαραστάσεων (Doc2vec) από το κείμενο. Το μοντέλο CSI (Εικόνα 132) περιλαμβάνει την ταξινόμηση (Integrate-classify), σύλληψη (Capture) νέων αναπαραστάσεων, τη βαθμολόγηση (Score-users) και την ενσωμάτωση (integrate).

Στη «σύλληψη» πραγματοποιείται η μοντελοποίηση της αναπαράστασης των ειδήσεων μέσω δικτύου μακροχρόνιας μνήμης με ενσωμάτωση από το Doc2vec των περιεχομένων ειδήσεων και των σχολίων των χρηστών. Στη βαθμολόγηση διευκρινίζεται το χαρακτηριστικό της πηγής (χρήστης) μέσω ενός γράφου βαρών. Η άκρη ενός γράφου προσδιορίζει τον αριθμό των άρθρων που δύο χρήστες έχουν διαμοιράσει.

Στην «ενσωμάτωση» (integrate) συγκρίνονται οι αναπαραστάσεις άρθρων με τη βαθμολόγηση των χρηστών. Παράλληλα γίνεται εκμάθηση πληροφοριών από τον χρήστη και το άρθρο μέσω του αλγόριθμου ταξινόμησης (Meel & Vishwakarma, 2020; Ruchansky et al., 2017; Shu, Cui, et al., 2019). Σύμφωνα με τους Meel και Vishwakarma, το συγκεκριμένο αλγοριθμικό μοντέλο έχει χρησιμοποιηθεί για τον εντοπισμό εξαπάτησης αφού ενσωματώνει νέες τεχνολογίες (Meel & Vishwakarma, 2020; Ruchansky et al., 2017).



Εικόνα 132. Μοντέλο CSI.

Οι Shim και Viswakarma υιοθέτησαν την οπτική της λευκής λίστας. Το συγκεκριμένο μοντέλο υιοθετεί χαρακτηριστικά κειμένου και κοινωνικού περιεχομένου της πληροφορίας παράλληλα και προσδιορίζει πόσοι αξιόπιστοι σύνδεσμοι συλλέγονται από τα αποτελέσματα του διαδικτύου (π.χ. 15 κορυφαίοι σύνδεσμοι σε λευκή λίστα) (Shim et al., 2021; Vishwakarma et al., 2019).

Ο Qian υιοθέτησε ένα πολυμοντελικό συνελικτικό δίκτυο γράφου γνώσης (*Knowledge aware multimodal graph convolutional network*) (Εικόνα 133) το οποίο μοντελοποιεί δημοσιεύσεις ως γράφους. Αποτελείται από 4 στοιχεία: την εξαγωγή σημασιολογικών αναπαράστασεων οντοτήτων (knowledge conceptualization), το συνελικτικό δίκτυο υπό μορφή γράφου όπου γίνεται εκμάθηση των τοπολογικών χαρακτηριστικών (δομική αποτύπωση πληροφοριών), την αναπαράσταση λέξεων δημοσιευμένου κειμένου μαζί με έναν μηχανισμό προσοχής για εύρεση σημασιολογικών ομοιοτήτων κειμένου και εικόνων, και από τον εντοπιστή των ψευδών ειδήσεων για την αναγνώριση των ψευδών δημοσιεύσεων.

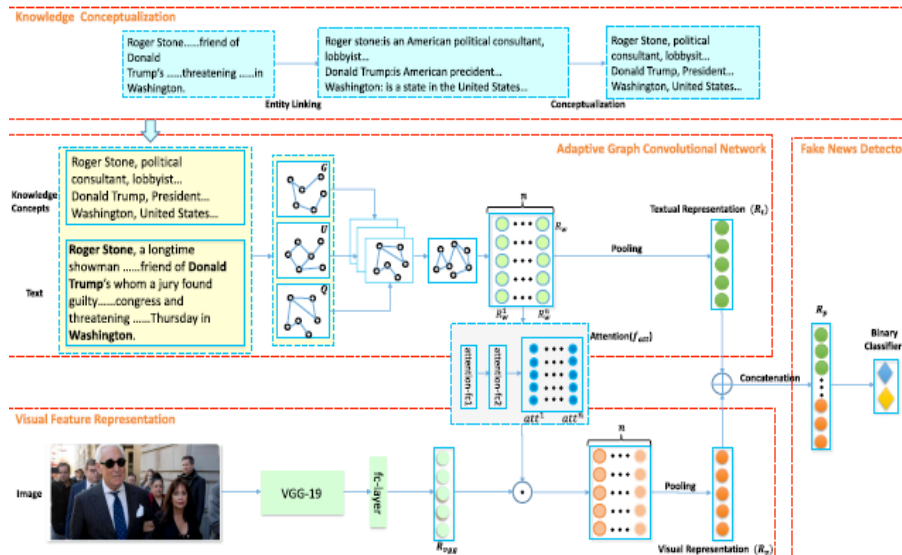
Η εξαγωγή των σημασιολογικών αναπαραστάσεων των οντοτήτων από έναν γράφο γνώσης είναι σημαντική γιατί παρέχει πρόσθετες πληροφορίες οι οποίες βοηθούν στον εντοπισμό των ψευδών ειδήσεων. Το συγκεκριμένο μοντέλο στοχεύει στην ανάκτηση συναφούς γνώσης για κάθε δημοσίευση από γνωσιακούς γράφους. Περιλαμβάνει τις μεθόδους ενοποίησης οντοτήτων (π.χ. Rel-Norm, Link Detector, EDEL και STEL) για την ενοποίηση αβέβαιων αναφορών οντοτήτων σε ένα κείμενο με τις σωστές οντότητες στις οποίες αναφέρονται (L. Chen et al., 2018; Kolitsas et al., 2018; Le & Titov, 2018; Milne & Witten, 2008). Με τη βοήθεια του γνωσιακού γράφου ανακτώνται σύνολα από γενικές έννοιες για κάθε οντότητα (παράδειγμα που ακολουθεί). Για μια δεδομένη δημοσίευση ο

γράφος γνώσης λειτουργεί ως οδηγός για κάθε οντότητα και για κάθε σύνολο ιδεών στόχος είναι η παραγωγή γνώσης για κάθε δημοσίευση.

Ο συγγραφέας υιοθέτησε αρχικά το STEL για τις αβέβαιες αναφορές των οντοτήτων οι οποίες αποκτήθηκαν από προϋπάρχοντα γράφο (όπως π.χ. Yago, Probase) (*Home / Yago Project*, n.d.; *Max-Planck-Institut Für Informatik: YAGO*, n.d.; Suchanek et al., 2008; Wu et al., 2012) (π.χ. “Thank you @realDonaldTrump for appointing me Chairman of the Minnesota Trump campaign! We will make Minnesota great again!,” σύνολο οντοτήτων $T = \{\text{realDonaldTrump, Minnesota}\}$, εξαγωγή οντοτήτων $\text{CrealDonaldTrump} = (\text{Donald Trump, American president, politician, Republican})$, $\text{CMinnesota} = (\text{state, the United States})$).

Το συνελικτικό δίκτυο υπό μορφή γράφου (Kirf & Welling, 2017) αποτελείται από ένα πολυεπίπεδο μοντέλο νευρωνικού δικτύου (GCN) που υπολογίζει έναν γράφο που έχει προσδιοριστεί από πριν. Επιπλέον προσπαθεί να αποτυπώσει την πληροφορία περιεχομένου όπου μία λέξη του περιεχομένου υπολογίζεται σαν ένας δεσμός γράφου. Για πιο δυναμική παρουσίαση χρησιμοποιήθηκε μία ευέλικτη δομή γράφου (Adaptive Graph Convolutional Network) που αντικατέστησε τον GCN και αποτελείται από 3 υπογράφους: τον παγκόσμιο γράφο, τον παραμετροποιημένο γράφο και τον ανεξάρτητο γράφο. Οι συγκεκριμένοι γράφοι βελτιώνουν την ευελιξία του μοντέλου.

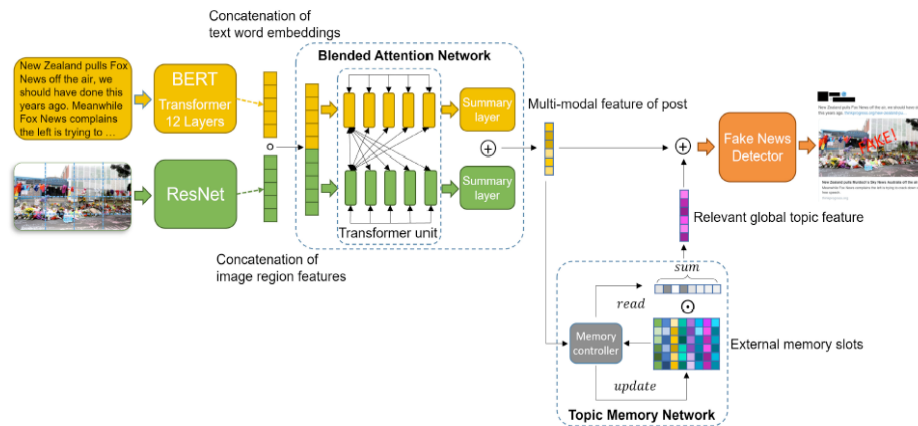
Η αναπαράσταση των οπτικών χαρακτηριστικών και των χαρακτηριστικών κειμένου γίνεται μέσω του δικτύου VGG-19 για την εύρεση των σημασιολογικών σχέσεων ανάμεσα στο κείμενο και την εικόνα (εναπόθεση βαρών στα οπτικά χαρακτηριστικά κάθε εικόνας η οποία είναι σημασιολογικά όμοια με μία λέξη του κειμένου αφού κάθε λέξη του κειμένου θεωρείται ότι σχετίζεται με τις εννοιολογικές αποτυπώσεις μίας εικόνας). Στο τέλος ο αλγόριθμος εντοπισμού παίρνει ως δεδομένου εισόδου την πολυμοντελική αναπαράσταση και στοχεύει στο αν η δημοσίευση είναι ψευδής ή όχι (S. Qian et al., 2021). Η εφαρμογή των αλγορίθμων πραγματοποιήθηκε στη βιβλιοθήκη βαθιάς μάθησης Pytorch (Paszke et al., 2017; *PyTorch*, n.d.).



Εικόνα 133. Πολυμοντελικό συνελκτικό δίκτυο γράφου γνώσης (KMAGCN).

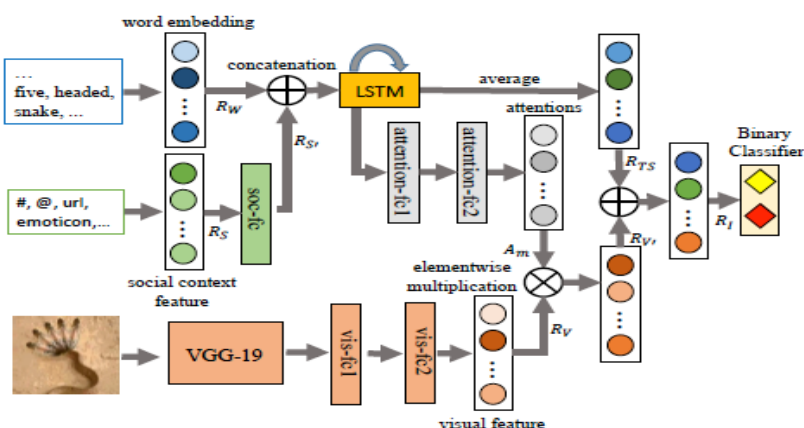
Ο Ying χρησιμοποίησε ένα τοπικό πολυμοντελικό θεματικό δίκτυο μνήμης (MTMN) για να βελτιστοποιήσει τον εντοπισμό των ψευδών ειδήσεων. Το συγκεκριμένο δίκτυο αποτελείται από ένα δίκτυο κωδικοποίησης κειμένου και εικόνας (text & image encoding network), από ένα πολυμοντελικό μοντέλο αναπαράστασεων (Blended attention module), από ένα τοπικό δίκτυο μνήμης (Topic memory network), και ένα δίκτυο εντοπισμού ψευδών ειδήσεων.

Η κωδικοποίηση των λέξεων κειμένου γίνεται από το μοντέλο BERT (Hofev, 2018) και της εικόνας από το ResNet 50 (Boesch, 2022). Στη συνέχεια αναζητούνται εσωτερικές και εξωτερικές σχέσεις ανάμεσα στο κείμενο (τμήματα κειμένου) και την εικόνα (εξαγωγή χαρακτηριστικών περιοχής). Το θεματικό τοπικό δίκτυο μαθαίνει αναπαραστάσεις κοινών θεμάτων και παγκόσμια χαρακτηριστικά από μη παρατηρήσιμα χαρακτηριστικά (αναπαραστάσεις πολυμοντελικών δημοσιεύσεων) και στο τελικό δίκτυο εντοπίζονται οι ψευδείς ειδήσεις μέσω πολυμοντελικών αναπαραστάσεων (Ying et al., 2021) (Εικόνα 134).



Εικόνα 134. Τοπικό πολυμοντελικό θεματικό δίκτυο μνήμης (MTMN).

Οι Qian και Ying χρησιμοποίησαν το πολυμοντελικό δίκτυο μηχανισμού προσοχής *AttRNN* ως μέτρο σύγκρισης με τα δικά τους μοντέλα. Το συγκεκριμένο δίκτυο συνδυάζει κείμενο, οπτική και κοινωνική πληροφόρηση. Επιπλέον, ενσωματώνει εικόνες σε κοινωνικά χαρακτηριστικά και χαρακτηριστικά κειμένου τα οποία προέρχονται από από το δίκτυο μακροχρόνιας μνήμης Long Short Term Memory (LSTM). Αποτελείται από ένα δίκτυο επαναλαμβανόμενο το οποίο ενσωματώνει περιεχόμενο κειμένου και κοινωνικού περιεχομένου, ένα υποδίκτυο οπτικοποίησης εικόνας (ResNet 50) και ένα νευρωνικό δίκτυο προσοχής το οποίο κωδικοποιεί τα οπτικά χαρακτηριστικά (S. Qian et al., 2021; Ying et al., 2021) (Εικόνα 135).



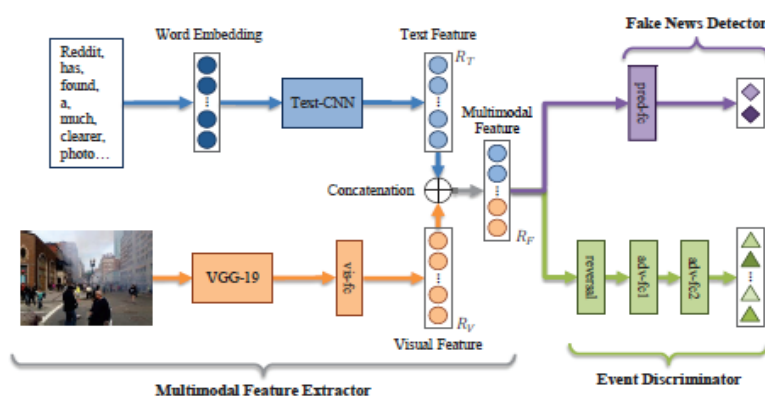
Εικόνα 135. Πολυμοντελικό δίκτυο μηχανισμού προσοχής AttRNN.

Επιπλέον, χρησιμοποίησαν το *EANN* (Event Adversarial Neural Network) (Y. Wang et al., n.d.) με σκοπό την εκμάθηση πολυμοντελικών χαρακτηριστικών σταθερών γεγονότων για κάθε δημοσίευση μέσω ενός αντιθετικού δικτύου (εξαγωγή συνδυαστικών χαρακτηριστικών εμπάθυνσης από κείμενο και εικόνα). Το συγκεκριμένο μοντέλο ενσωματώνει τρία βασικά

χαρακτηριστικά: τον πολυμοντελικό εξαγωγέα χαρακτηριστικών, τον εντοπιστή ψευδών ειδήσεων και τον διαχωριστή γεγονότων.

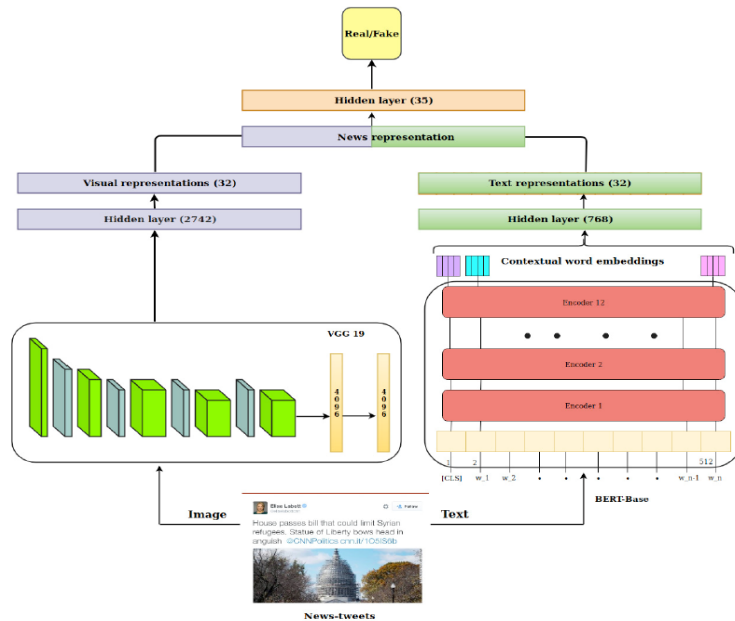
Ο πολυμοντελικός εξαγωγέας περιλαμβάνει εξαγωγείς κειμένου και εικόνας για διαφορετικά δεδομένα εισόδου. Αφού έχει γίνει η εκμάθηση των μη παρατηρήσιμων χαρακτηριστικών (αναπαραστάσεων) μετέπειτα ενσωματώνονται για να δημιουργήσουν μία τελική αναπαράσταση. Ο εντοπιστής ψευδών ειδήσεων παίρνει ως δεδομένο εισόδου την αναπαράσταση για να προβλέψει αν οι δημοσιεύσεις (post) είναι ψευδείς ή αληθείς. Ο περιγραφέας γεγονότος προσδιορίζει την κατηγοριοποίηση του γεγονότος βασιζόμενος στην αναπαράσταση (Εικόνα 136) (S. Qian et al., 2021; Y. Wang et al., 2018; Ying et al., 2021).

Το Spotfake και Spotfake+ χρησιμοποιήθηκαν από τον Qian για σύγκριση με τον γνωσιακό γράφο συνελκτικού νευρωνικού δικτύου (KMAGCN). Τα συγκεκριμένα πολυμοντέλα εκπαιδεύονται με το BERT και το XLNet³³ αντίστοιχα για εκμάθηση αναπαραστάσεων κειμένου και για εκμάθηση χαρακτηριστικών εικόνας (VGG-19). Το Spotfake+ (Singhal, n.d.) αποτελεί εξελιγμένη έκδοση του Spotfake όπου η εξαγωγή των χαρακτηριστικών γίνεται από προεκπαιδευμένα μοντέλα γλώσσας (Transformers) και εικόνας (image models). Τα χαρακτηριστικά των διανυσμάτων εισάγονται σε ένα πλήρες συνδεδεμένο επίπεδο για κατηγοριοποίηση (Εικόνα 137, Εικόνα 138) (S. Qian et al., 2021; Singhal et al., 2020).

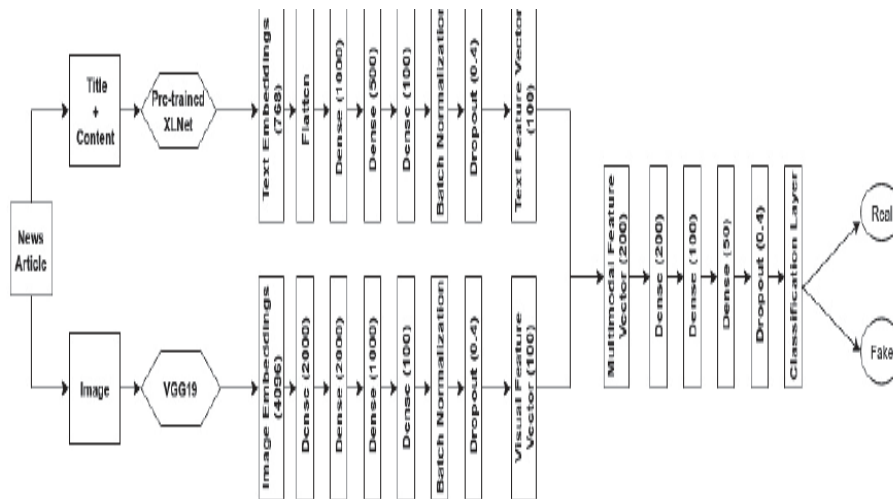


Εικόνα 136. Μοντέλο EANN.

³³ Η XLNet αποτελεί επέκταση του μετατροπέα XL (Transformer-XL) όπου η εκμάθηση συμφραζόμενων πληροφοριών πραγματοποιείται αμφίδρομα (αμφίδρομο πλαίσιο). Επιπλέον ενσωματώνει έναν μηχανισμό επανάληψης τμημάτων αλλά και σχετικό σχήμα κωδικοποίησης του Transformer-XL στην προεκπαίδευση, το οποίο βελτιώνει την απόδοση κυρίως για κείμενα μεγαλύτερης ακολουθίας. (Z. Yang et al., 2020)



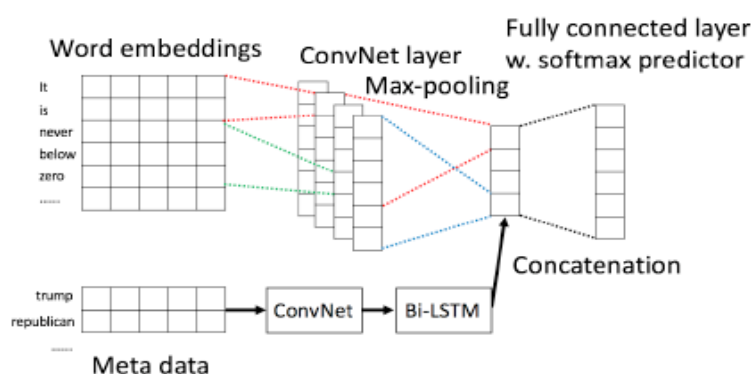
Εικόνα 137. Μοντέλο Spotfake.



Εικόνα 138. Μοντέλο Spotfake+.

Ο αλγόριθμος *Hybrid CNN* (Εικόνα 139) είναι ένα υβριδικός αλγόριθμος ο οποίος ενσωματώνει κείμενο και μεταδεδομένα (Goldani et al., 2021; W. Y. Wang, 2017). Έχει την ίδια δομή με τον απλό CNN (μονοεπίπεδο), τον αλγόριθμο ενσωμάτωσης BiLSTM (συνδυασμός μεταδεδομένων και ενσωματώσεων λέξεων) με ένα πλήρες συνδεδεμένο επίπεδο (πρόβλεψη softmax). Ο Wang το υιοθέτησε για την αποτύπωση των διανυσμάτων των μεταδεδομένων με πολύ καλά αποτελέσματα (W. Y. Wang, 2017). Ενω ο Goldani το χρησιμοποίησε για να το συγκρίνει με άλλους δημοφιλείς αλγόριθμους (LSTM προσοχής)

αλλά και με τον αλγόριθμο δικτύου που δημιούργησαν (αλγόριθμος δικτύου κάψουλας) (Goldani et al., 2021).



Εικόνα 139. Μοντέλο Hybrid CNN.

4.8 Οπτικά Χαρακτηριστικά

Το οπτικό περιεχόμενο (εικόνες και βίντεο) σε αρκετές περιπτώσεις αυξάνει την αξιοπιστία των ειδήσεων για αυτό και πολλοί εκδότες χρησιμοποιούν προβοκατόρικο οπτικό περιεχόμενο, για να προσελκύσουν, να εντυπωσιάσουν αλλά και να παραπλανήσουν τους αναγνώστες. Ο Viswakarma υιοθέτησε τον αλγόριθμο εντοπισμού κειμένου από εικόνα-Άκρως Σταθερές Ακραίες Περιοχές (Maximally Stable Extremal Region) (Vishwakarma et al., 2019).

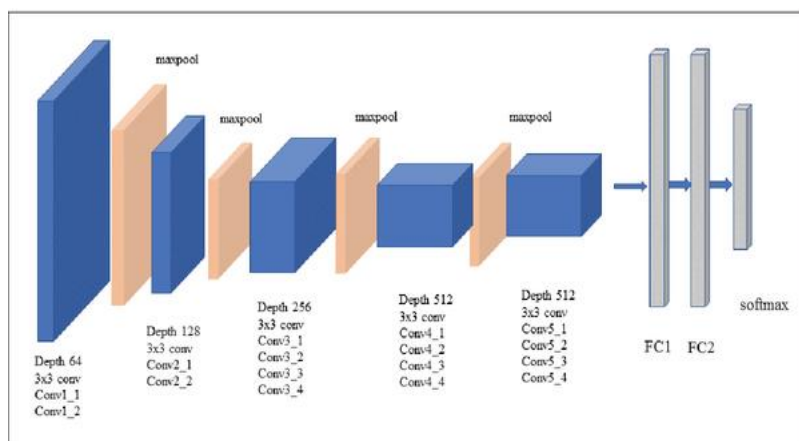
Ο αλγόριθμος Άκρως Σταθερές Ακραίες Περιοχές (Maximally Stable Extremal Region-MSER) ("Maximally Stable Extremal Regions," 2022)³⁴ χρησιμοποιείται συμπληρωματικά με την οπτική αναγνώρισης χαρακτήρων (*Optical Character Recognition*) (Tundis et al., 2021) και υιοθετήθηκε από τον Viswakarma για την εξαγωγή κειμένου από εικόνα υιοθετώντας την οπτική του Chidiac για τον εντοπισμό της περιοχής του κειμένου. Η συγκεκριμένη μέθοδος χωρίζεται σε τρία στάδια.

Στο πρώτο στάδιο ο αλγόριθμος εντοπίζει την τοποθεσία του κείμενου από οποιαδήποτε τοποθεσία και σε οποιοδήποτε μέγεθος και γραμματοσειρά. Στο δεύτερο στάδιο αναγνωρίζει καλύτερα τα γράμματα (π.χ. άκρες), και στο τρίτο εφαρμόζεται ο εντοπιστής πλάτους όπου εντοπίζεται το μέγεθος των χαρακτήρων (οριοθέτηση των pixel της εικόνας). Στο τέλος φιλτράρεται η περιοχή που δεν περιλαμβάνει χαρακτήρες κειμένου.

³⁴ Ο συγκεκριμένος αλγόριθμος δημιουργήθηκε αρχικά από τον Matas (Matas et al., 2004).

Πιο συγκεκριμένα, ο Chidiac αναφέρει ότι οι συγκεκριμένες περιοχές εντοπίζονται όταν εισάγεται η εικόνα και περιλαμβάνουν στοιχεία του κειμένου τα οποία οριοθετούνται ως λευκά pixels. Επειδή η εικόνα όμως έχει μετατραπεί σε δυαδικά ψηφία δεν μπορεί να αποκαλύψει τα ακριβή όρια του κειμένου γι' αυτό η εικόνα βελτιώνεται ποιοτικά. Μετέπειτα, εντοπίζονται οι άκρες και εισέρχονται στον εντοπιστή πλάτους όπου οι γραμμές, το πλάτος γραμμών και τα συνδεδεμένα στοιχεία φιλτράρονται. Ο συγκεκριμένος αλγόριθμος θεωρείται από τους καλύτερους στον εντοπισμό κειμένου καθώς αντέχει στις αλλαγές όπως η αλλαγή φωτεινότητας (Chidiac et al., 2016; Vishwakarma et al., 2019).

Ο Qian υιοθέτησε το *νευρωνικό δίκτυο VGG-19* (Simonyan, n.d.; *Visual Geometry Group - University of Oxford*, n.d.) με 19 επίπεδα βάθους το οποίο κατηγοριοποιεί εικόνες σε πάνω από 1000 κατηγορίες αντικειμένων (πχ. ποντίκι, πληκτρολ. μολύβι, ζώα) προκειμένου να προσδιορίσει συγκεκριμένα μοτίβα των φημών (S. Qian et al., 2021) (Εικόνα 140)³⁵. Το συγκεκριμένο δίκτυο δημιουργήθηκε από τους Simonyan και Zisserman με στόχο την εις βάθος αναγνώριση των εικόνων μέσω ενός συνελκτικού δικτύου 19 επιπέδων. Η εικόνα περνάει από ένα σύνολο συνελκτικών επιπέδων όπου χρησιμοποιούνται φίλτρα (3X3 διαστάσεων) για την αποτύπωση της κατεύθυνσης (κέντρο, δεξιά, αριστερά, πάνω, κάτω). Μετέπειτα ακολουθούν τρία επίπεδα εκ των οποίων το τρίτο πραγματοποιεί την κατηγοριοποίηση (Simonyan & Zisserman, 2015).



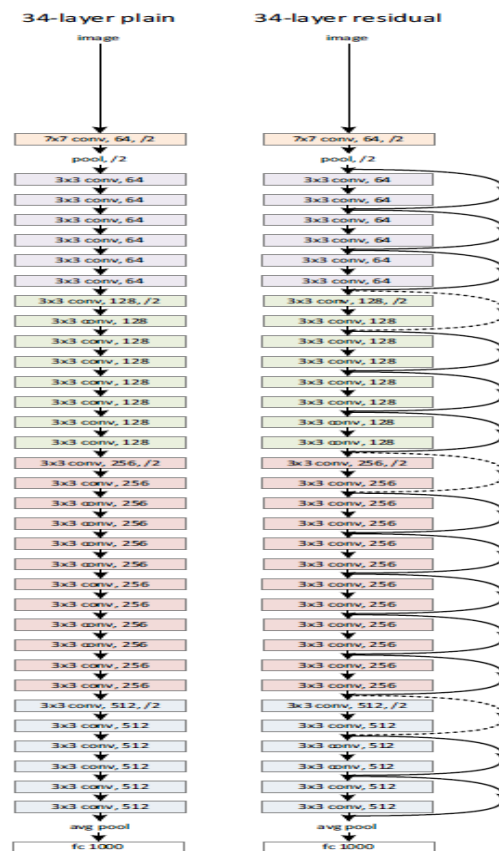
Εικόνα 140. Νευρωνικό δίκτυο VGG-19.

Το *νευρωνικό δίκτυο βαθιάς μάθησης ResNet 50* (Microsoft/Resnet-50 · Hugging Face, n.d.)³⁶ (Εικόνα 141) το οποίο είναι η μετεξέλιξη του ResNet 34 χρησιμοποιήθηκε από τον Ying για να εξάγει χαρακτηριστικά περιοχής για κάθε εικόνα (Ying et al., 2021). Το συγκεκριμένο δίκτυο αποτελείται από 50 επίπεδα και επιλύει σύνθετα προβλήματα,

³⁵ Δες και (Sec, 2021).

³⁶ Δες και (Boesch, 2022)

ελαχιστοποιεί τα σφάλματα, είναι πιο ακριβές, και παρέχει βέλτιστη απόδοση (He et al., 2015).



Εικόνα 141. ResNet 50.

Οι Meel και Vishwakarma και Zhang υιοθέτησαν τα ακόλουθα χαρακτηριστικά που αφορούν εικόνες: τον αριθμό των προσώπων, την ανάλυση εικόνας, τα οπτικά χαρακτηριστικά εικόνας όπως ο βαθμός ευκρίνειας, ο βαθμός διαφορετικότητας, ο βαθμός συνοχής, ο βαθμός ομοιότητας (ομαδοποίησης) (Jin, Cao, Zhang, et al., 2017). Επιπλέον, ανέφεραν και τη χρήση στατιστικών χαρακτηριστικών εικόνας όπως η αναλογία εικόνας, η αναλογία πολυμέσου, η αναλογία θερμικής εικόνας (Meel & Vishwakarma, 2020). Στη μελέτη του Yang διαπιστώθηκε ότι οι αληθείς ειδήσεις είχαν περισσότερα πρόσωπα (π.χ. 0.366 πρόσωπα κατά μέσο όρο στις αληθείς και 0.299 στις ψευδείς) ενώ οι ψευδείς ειδήσεις είχαν άσχετες εικόνες όπως ζώα και σκηνές. Επιπλέον η εικόνα στις αληθείς ειδήσεις έχει καλύτερη ανάλυση (455X277) από ότι στις ψευδείς (355X228) (Y. Yang et al., 2022).

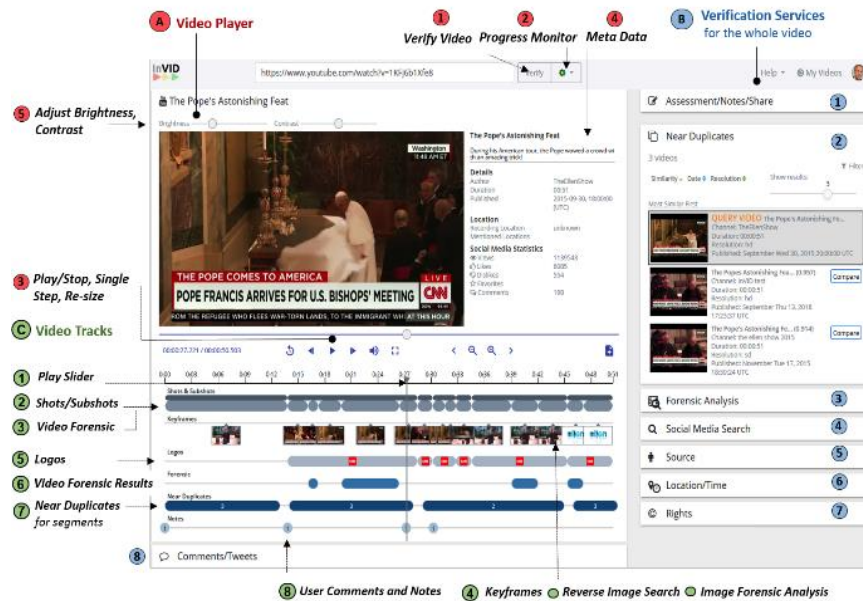
Ενώ ο Borges χρησιμοποίησε τη μετρική Cider (T. Y. Lin et al., n.d.) για την αξιολόγηση της εικόνας. Η συγκεκριμένη μετρική αντιστοιχίζει την πρόταση με περιγραφές εικόνων γραμμένες από ανθρώπους. Ουσιαστικά μετρά την ομοιότητα ανάμεσα στην πρόταση που

έχει δημιουργηθεί και στο σύνολο των πραγματικών προτάσεων που έχουν γραφτεί από ανθρώπους (Borges et al., 2019; Vedantam et al., 2015). Χρησιμοποιεί τον αλγόριθμο Tf-idf για τη συχνότητα ενός ngram σε μια πρόταση αναφοράς (X. Chen et al., 2015).

Σε επίπεδο βίντεο η πλατφόρμα αξιολόγησης γνώσης *InVid* (*In Video Veritas*) (*InVID Project - Video Verification*, n.d.) (Εικόνα 142) παρέχει δυνατότητες όπως αυθεντικοποίηση βίντεο και διαμοιρασμένων αρχείων στα μέσα κοινωνικής δικτύωσης, ελέγχου και αξιολόγησης της αξιοπιστίας των αρχείων βίντεο καθώς και εντοπισμού της απάτης (Meel & Vishwakarma, 2020; Papadopoulou et al., 2019). Η συγκεκριμένη διαδικτυακή πλατφόρμα βασίστηκε στο *Verification Handbook* (Silverman et al., n.d.) με την καθοδήγηση των συνεργατών του *Invid*, του Γαλλικού τύπου (*Agence France Presse*) και της *Deutsche Welle* (*News and Current Affairs from Germany and around the World – DW*, n.d.; *The News Hub*, 2012).

Το διαδικτυακό εργαλείο πιστοποίησης βίντεο περιλαμβάνει τον προβολέα βίντεο το οποίο παρουσιάζει το υπό έρευνα υλικό παρέχοντας διακοπή και επανάληψη τμηματικά καθώς και διαχείριση φωτεινότητας και αντίθεσης. Για την αξιολόγηση του βίντεο παρέχονται υπηρεσίες πιστοποίησης στους χρήστες όπως:

- Έλεγχος για προηγούμενη χρήση του βίντεο
 - ειδική αντίστροφη αναζήτηση στη συλλογή των βίντεο *InVid* και προσεκτική εξέταση του ερωτήματος και των πανομοιότυπων βίντεο μέσω παράλληλης αναπαραγωγής
 - αντίστροφη αναζήτηση παρέχεται και διαδικτυακά μέσω παρεχόμενων εικονιδίων βίντεο από το YouTube
- έλεγχος προέλευσης βίντεο και δικαιώματα
 - μέσω εντοπισμού λογοτύπου βίντεο
 - χρήση δικαιωμάτων διαχείρισης βίντεο από εφαρμογή
- έλεγχος πληροφόρησης περιεχομένου βίντεο
 - από ανάλυση περιεχομένου κοινωνικών μέσων
 - αξιολόγηση τοποθεσίας, χρόνου και άλλων μεταδεδομένων του βίντεο αλλά και ιστορικού δεδομένων για τον καιρό
- έλεγχος νομιμότητας βίντεο μέσω
 - προηγμένων φίλτρων για βίντεο
 - παροχής τμηματικής εξέτασης βίντεο στην επιφάνεια εργασίας του χρήστη



Εικόνα 142. Πλατφόρμα InVid.

Η Παπαδούλου χρησιμοποίησε την συγκεκριμένη πλατφόρμα για την αξιολόγηση των ψεύτικων βίντεο σε κοινωνικά μέσα δικτύωσης και εξαγωγή χαρακτηριστικών όπως την καταμέτρηση θέασης καναλιού, την καταμέτρηση σχολιασμών καναλιού, την καταμέτρηση εγγραφών ατόμων, τη διάρκεια των βίντεο. Ως προς την καταμέτρηση σχολιασμών σύμφωνα με την έρευνα διαπιστώθηκε ότι τα ψεύτικα βίντεο έχουν περισσότερα σχόλια από ότι τα αληθινά. Αυτό συμβαίνει λόγω της ύπαρξης διπλότυπων ψεύτικων βίντεο αλλά και συνδέσμων από χρήστες οι οποίοι επαναδημοσιεύουν τα ψεύτικα ακόμα και αν κάποια από αυτά είναι αρκετά παλιά (δεκαετίας) (π.χ. 4.991.366 σχόλια για τα ψεύτικα βίντεο και 433.139 για τα αληθινά στο YouTube). Για την καταμέτρηση εγγραφών ατόμων βρέθηκε ότι ο αριθμός εγγραφών σε αληθινά βίντεο είναι μεγαλύτερος ενώ σε ψεύτικα μικρότερος (π.χ. 349 χρήστες για τα αληθινά βίντεο ενώ για τα ψεύτικα μόνο 98). Ενώ για τη διάρκεια των βίντεο, ότι τα ψεύτικα βίντεο έχουν μικρότερη διάρκεια (π.χ. 92 λεπτά) ενώ τα αληθινά μεγαλύτερη (π.χ. 149 λεπτά) (Boididou et al., 2018; Papadopoulou et al., 2017, 2019)

Επιπλέον, από την περιγραφή του βίντεο εξήγαγε το εύρος κειμένου του βίντεο, τον αριθμό των λέξεων, θαυμαστικά ή ερωτηματικά (Boolean), το 1ο και 3ο πρόσωπο, τον αριθμό κεφαλαίων χαρακτήρων, την ύπαρξη θετικών ή αρνητικών λέξεων, την ύπαρξη λέξεων αργκό, αν έχει σύμβολο άνω κάτω τελείας «:», και τον αριθμό των ερωτηματικών και θαυμαστικών. Αξίζει να επισημανθεί ότι τα λεκτικά χαρακτηριστικά των βίντεο (1^ο, 3^ο πρόσωπο, λέξεις αργκό, σύμβολα κλ.π.) ψεύτικων και μη δεν παρουσίασαν ιδιαίτερες διαφοροποιήσεις (Papadopoulou et al., 2019).

4.9 Κοινωνικό Περιεχόμενο

Τα χαρακτηριστικά κοινωνικού περιεχομένου περιλαμβάνουν *χαρακτηριστικά χρηστών στα μέσα κοινωνικής δικτύωσης* (Facebook, Reddit, Twitter) (π.χ. προφίλ χρήστη, φίλοι χρήστη, ακόλουθοι, άτομα που ακολουθεί, ημερομηνία δημιουργίας λογαριασμού), καθώς και *χαρακτηριστικά ανάλυσης κοινωνικών δικτύων* (δικτύων φίλων, επιδημιολογικών μοντέλων), *δικτύων διάχυσης* (propagation), *χαρακτηριστικών αλληλεπίδρασης χρηστών* (*shares, likes, dislike*), *χαρακτηριστικά δημοσιεύσεων, σχολίων, χωροχρονικά χαρακτηριστικά* (Alhazbi, 2020; Baeth & Aktas, 2019; Cohen et al., 2020; Hunt et al., 2022; Meel & Vishwakarma, 2020; Shu, Cui, et al., 2019; Weiss et al., 2019; Q. Zhang et al., 2019).

Τα *χαρακτηριστικά χρηστών* που χρησιμοποιήθηκαν στην παρούσα βιβλιογραφία είναι το *προφίλ χρήστη* (Shu, Cui, et al., 2019), η *ταυτότητα χρήστη, όνομα εμφάνισης χρήστη* (*user display name*), *όνομα χρήστη (στην οθόνη), περιγραφή προφίλ χρήστη, κείμενο tweet* (*δημοσίευση χρήστη*) (Cohen et al., 2020; Shu, Wang, et al., 2019; Taskin et al., 2022), *ημέρα και ώρα δημιουργίας tweet, γλώσσα δημοσίευσης tweet* (Weiss et al., 2019), *αναδημοσιεύτηκε* (Cohen et al., 2020), *αριθμός αγαπημένων δημοσιεύσεων* (Baeth & Aktas, 2019), *αριθμός ακολούθων* (Baeth & Aktas, 2019; Cohen et al., 2020; Hunt et al., 2022; Meel & Vishwakarma, 2020; Shu, Cui, et al., 2019; Taskin et al., 2022; Weiss et al., 2019), *αριθμός ατόμων που ακολουθεί* (Baeth & Aktas, 2019; Cohen et al., 2020; Meel & Vishwakarma, 2020; Shu, Cui, et al., 2019; Weiss et al., 2019), ενώ ο Weiss υπολόγισε τον *αριθμό των λιστών που εμφανίζονται οι χρήστες, το μήκος της περιγραφής του προφίλ και το μήκος ονόματος χρήστη* (Weiss et al., 2019).

Σύμφωνα με τον Shu οι δημοσιεύσεις εκφράζουν τα συναισθήματα ή τις γνώμες των χρηστών για τις ψευδείς ειδήσεις όπως σκέψεις, συναισθηματικές αντιδράσεις αλλά και δημοσιεύσεις από ομάδα αντιεμβολιαστών (@pharmthesheep) (instagram) (Shu et al., 2017, 2018; Shu, Wang, et al., 2019). Σχετικά με την ώρα δημοσίευσης (ώρα δημιουργίας) στο twitter αναφέρεται σε δημοσιεύσεις κατά τη διάρκεια της ημέρας. Σύμφωνα με τον Alhazbi τα τρολ τουιτάρουν συνηθισμένες ώρες (ώρες γραφείου) σε σχέση με τους συνηθισμένους χρήστες οι οποίοι δημοσιεύουν απόγευμα και μεσάνυχτα (66%) (Alhazbi, 2020).

Ο αριθμός των ακολούθων σύμφωνα με έρευνα του Cohen λειτουργεί σαν δείκτης απήχησης της παραπληροφόρησης ενώ σύμφωνα με μελέτη του AlRubaian οι ακόλουθοι όπως και οι φίλοι των χρηστών συνδέονται με τη φήμη του χρήστη (AlRubaian et al., 2015; Cohen et al., 2020). Ο Ghenaï στην έρευνα του για το twitter διαπίστωσε αυξημένο αριθμό ακολούθων σε περιβάλλοντα φημών (Ghenai, 2019; Ghenaï & Mejona, 2018). Επιπλέον όσον

αφορά τον αριθμό των ακολούθων στην έρευνα του Taskin διαπιστώθηκε ότι οι χρήστες που έχουν υψηλό βαθμό δραστηριότητας ακολουθούν περισσότερο χρήστες που διαδίδουν ψευδείς ειδήσεις από ότι αληθείς (π.χ. 120 χρήστες ακολουθούν 304 λογαριασμούς ψευδών ειδήσεων) (Taskin et al., 2022).

Εκτός από τα παραπάνω χαρακτηριστικά οι Weiss και Cohen υιοθέτησαν και τον *γεωεντοπισμό* δηλαδή αν ο χρήστης έχει ενεργό τον γεωεντοπισμό στο λογαριασμό του αλλά και τον *αριθμό των δημοσιεύσεων* που έχει ο χρήστης του ως στοιχείο για νωρίτερο εντοπισμό των φημών και της διάδοσης των ψευδών ειδήσεων (Cohen et al., 2020; Weiss et al., 2019). Ο Cohen υιοθετώντας το *μοντέλο Ghennai* αναφέρει ότι οι χρήστες που διέδιδαν φήμες χρησιμοποιούσαν τον γεωεντοπισμό για τη διάδοσή τους με τις τοποθεσίες να βρίσκονται σε δυσπρόσιτες περιοχές (Cohen et al., 2020). Οι Weiss, Meel & Vishwakarma και Hunt υπολόγισαν και τον αριθμό των χρηστών με πιστοποιημένους λογαριασμούς (Hunt et al., 2022; Meel & Vishwakarma, 2020; Weiss et al., 2019). Σύμφωνα με τον Indu η αυθεντικότητα του ενός λογαριασμού βασίζεται στο αν ο χρήστης έχει πιστοποιημένο λογαριασμό (π.χ. Twitter) (Indu & Thampi, 2019; Meel & Vishwakarma, 2020) ή αν όχι (π.χ. οι χρήστες που έχουν πιστοποιημένο λογαριασμό έχουν βαθμό 1.0) (Castillo et al., 2011; Meel & Vishwakarma, 2020). Οι Meel και Vishwakarma υπολόγισαν επίσης το φύλο, την εικόνα και την τοποθεσία χρήστη (Meel & Vishwakarma, 2020).

Ο *αριθμός των μηνυμάτων tweet των χρηστών* (Meel & Vishwakarma, 2020; Shu, Wang, et al., 2019; Souza Freire et al., 2021), ο *αριθμός παρελθοντικών tweet*, η *ημερομηνία ή ώρα δημιουργίας λογαριασμού*, η *κατάσταση του λογαριασμού*, η *συμπεριφορά χρήστη*, το *κατά πόσο είναι ολοκληρωμένο το προφίλ του*, ο *αριθμός των καθημερινών tweet* αλλά και ο *αριθμός των φίλων* υιοθετήθηκαν ως χαρακτηριστικά των χρηστών από τους Meel και Vishwakarma, Souza Freire, Shu, Alhazbi (Alhazbi, 2020; Meel & Vishwakarma, 2020; Shu, Cui, et al., 2019; Souza Freire et al., 2021).

Για τον αριθμό tweet ή μηνυμάτων που δημοσιεύονται από χρήστη ο Indu αναφέρει σε δεδομένα από έρευνα του για τη διάδοση των φημών, ότι ο χρήστης με ταυτότητα ID: 25549 μετέδωσε 105 χιλιάδες tweets ψεύτικων φημών (Indu & Thampi, 2019; Meel & Vishwakarma, 2020).

Οι Alhazbi και Cohen χρησιμοποίησαν τον *αριθμό των hashtags*, και των *URL* ως ένα ακόμη χαρακτηριστικό των χρηστών που διαδίδουν ψευδείς ειδήσεις (Alhazbi, 2020; Cohen et al., 2020). Οι χρήστες χρησιμοποιούν hashtags προκειμένου να δημιουργήσουν ή να ακολουθήσουν μία συζήτηση χρησιμοποιώντας μια λέξη ή φράση μαζί με ένα σύμβολο (#). Σύμφωνα με τον Alhazbi το συγκεκριμένο χαρακτηριστικό υιοθετείται από πολιτικά τρολ για

διάδοση προπαγάνδας και αναζωπύρωση αμφιλεγόμενων θεμάτων (π.χ. #BlackLives matter). Για παράδειγμα 32% των ρωσικών τρολ εμπεριέχουν στα tweet τους τα hashtags. Τα τρολ χρησιμοποιούν εκτός από τα hashtags και τα URL σε άλλα tweet ή εξωτερικούς συνδέσμους σχετικούς με τις πολιτικές τοποθετήσεις τους (π.χ. 53% των tweet των Ρώσικων τρολ εμπεριείχαν συνδέσμους URL) (Alhazbi, 2020).

Ο αριθμός αναδημοσιεύσεων υιοθετήθηκε από τους Alhazbi, Hunt, Cohen και Souza Freire ως ένα ακόμη χαρακτηριστικό των χρηστών (Alhazbi, 2020; Cohen et al., 2020; Hunt et al., 2022; Souza Freire et al., 2021). Οι Hunt και Souza Freire το χρησιμοποίησαν ως χαρακτηριστικό της παραπληροφόρησης στο twitter για την εκπαίδευση των αλγορίθμων ενώ ο Alhazbi επισημαίνει τις αναδημοσιεύσεις των τρολ στο twitter προκειμένου να μεγιστοποιήσουν την αποτελεσματικότητά τους επαναδημοσιεύοντας άλλα tweet (από άλλα τρολ) σχετικά με γεγονότα πολιτικά (Alhazbi, 2020) αλλά και από τρολ που παρουσιάζονται ως δημοσιογράφοι των ΗΠΑ (π.χ. «#ThanksObama We're FINALLY evicting Obama. Now Donald Trump will bring back jobs for the lazy ass Obamacare recipients», δεξιό τρολ, “#ClintonBodyCount if anyone else had her rap sheet - they'd be on death row”, αριστερό τρολ) (Caucaso, n.d.).

Οι ίδιοι συγγραφείς υιοθέτησαν ως χαρακτηριστικό τον αριθμό προτιμήσεων «μου αρέσει» (π.χ. Και μετά ουρλιάζουν για το #NΔ_παιδεραστές “Ελεύθερος και αυτός με τον νόμο Παρασκευόπουλου κύριε @tsiaras_kostas”-72 επισημανσεις μου αρέσει από ελληνικό λογαριασμο Twitter, φήμη για μετανάστευση λόγω τυφώνα Χάρβευ (ΗΠΑ) -210 μου αρέσει ανά tweet) (Hunt et al., 2022; Souza Freire et al., 2021) ενώ ο Alhazbi τον αριθμό των απαντήσεων, και το ποσοστό των tweet για το σαββατοκύριακο (Alhazbi, 2020; Hunt et al., 2022; Souza Freire et al., 2021). Ως προς τον αριθμό των απαντήσεων σύμφωνα με τον Alhazbi τα τρολ προωθούν δικούς τους λογαριασμούς (ίδιας πολιτικής ομάδας) (twitter) ή άλλων τρολ για διάδοση προπαγάνδας (π.χ. μέσω της ανάλυσης των tweet βρέθηκε ότι 20% tweet των ρώσικων τρολ ήταν απαντήσεις σε αναδημοσιεύσεις) (Alhazbi, 2020; Caucaso, n.d.). Οι έρευνες των Keller και Alhazbi διαπιστώνουν λιγότερα tweet των τρολ το σαββατοκύριακο. Το ίδιο επισημαίνει ο Alhazbi για τα ρωσικά τρολ και τα τρολ της Σαουδικής Αραβίας τα οποία δημοσιεύουν λιγότερα tweet τα σαββατοκύριακα σε σχέση με τους κανονικούς χρήστες (με βάση τη συμπεριφορά των χρηστών στις εκλογές της Ν. Κορέας το 2012) (Alhazbi, 2020). Ο Cohen πήρε ως ένα ακόμη χαρακτηριστικό για τον εντοπισμό των χρηστών που διαδίδουν ψευδείς ειδήσεις την ημέρα της εβδομάδας που γράφτηκε το tweet (Cohen et al., 2020).

Οι Baeth και Aktas χρησιμοποίησαν μετρικές χρηστών όπως η αξιοπιστία, η φήμη, η διαθεσιμότητα, η κοινωνική επίδραση, το κύρος και την επίδραση του αντικειμένου του

χρήστη για τον εντοπισμό της παραπληροφόρησης στα κοινωνικά δίκτυα (Baeth & Aktas, 2019).

Η *μετρική αξιοπιστίας* αφορά τον μέσο όρο του αριθμού των ανεξάρτητων κοινωνικών ιστοσελίδων που χρησιμοποιήθηκαν για να αξιολογήσουν την τιμή για παράδειγμα η αξιολόγηση των χαρακτηριστικών του ιστορικού των πληροφοριών του χρήστη (Baeth & Aktas, 2019; Meel & Vishwakarma, 2020).

Η *μετρική της φήμης* αφορά το κύρος του χρήστη (Baeth & Aktas, 2019). Ως παράδειγμα οι συγγραφείς αναφέρουν ότι εάν ένας χρήστης έχει υψηλή δημοτικότητα τότε θα έχει πολλούς δεσμούς (nodes) με άλλους χρήστες (η συγκεκριμένη μετρική χρησιμοποιείται στο Facebook).

Η *μετρική της διαθεσιμότητας* σχετίζεται με τις πληροφορίες των χρηστών οι οποίες αποτελούν βάση αξιοπιστίας. Για παράδειγμα αν ένας χρήστης παίρνει πληροφορίες από αξιόπιστο χρήστη τότε δίνει αξιοπιστία στην ίδια την πληροφορία (Baeth & Aktas, 2019).

Οι *μετρικές της κοινωνικής επίδρασης και του κύρους* αφορούν τις μετρήσεις αλληλεπιδράσεων των χρηστών δηλαδή tweets και απαντήσεις σε χρήστες που δεν είναι άμεσα συνδεδεμένοι μεταξύ τους (π.χ. υψηλός αριθμός αναδημοσιεύσεων, απαντήσεων).

Η *μετρική της επίδρασης* της δημοσίευσης του χρήστη σχετίζεται με έναν μη γνωστό χρήστη ο οποίος δημοσιεύει αξιόλογες πληροφορίες και ο οποίος μπορεί να έχει περισσότερους ακολούθους στο Twitter (Baeth & Aktas, 2019).

Οι Meel και Viswakarma εκτός από τα χαρακτηριστικά των χρηστών παραθέτουν και άλλες βιβλιογραφικές έρευνες που αφορούν συγκεκριμένα χαρακτηριστικά όπως η *αυθεντικότητα* στην επικοινωνία των χρηστών (δηλαδή τον αριθμό των πρωτότυπων tweet αλλά και των φορών που κάποιος χρήστης αναδημοσίευσε tweet άλλων χρηστών) αλλά και την *αναλογία των ακολούθων και εκείνων που ακολουθεί ο χρήστης* (Meel & Vishwakarma, 2020). Τα συγκεκριμένα χαρακτηριστικά υιοθετήθηκαν από τον Vosoughi για την πρόβλεψη των φημών στο Twitter. Για την αυθεντικότητα επισήμανε ότι όσο πιο μεγάλη η αναλογία των πρωτότυπων tweet τόσο πιο αυθεντικός είναι ο χρήστης ενώ αν επαναλαμβάνει ότι λένε οι άλλοι τότε υπάρχει χαμηλή αναλογία (Vosoughi et al., 2017).

Ο Cohen προσθέτει επιπλέον και χαρακτηριστικά όπως αν ο χρήστης έχει *αναφορές, hashtags, την χώρα προέλευσης* αλλά και την *ηλικία του λογαριασμού του χρήστη* (Cohen et al., 2020) ενώ ο Shu την *ηλικία εγγραφής του χρήστη* (Shu, Cui, et al., 2019). Το χαρακτηριστικό των αναφορών υιοθετήθηκε από τον Ghenaί ως χαρακτηριστικό αξιοπιστίας (π.χ. @CNN) (Cohen et al., 2020; Ghenaί & Mejova, 2017). Τα hashtags αποτελούν ετικέτες

περιεχομένου οι οποίες έχουν δημιουργηθεί από χρήστες. Σύμφωνα με τον Cohen τα hashtags συντελούν στη παραπληροφόρηση (π.χ. #coronavirus5G) (Cohen et al., 2020; *Coronavirus5G*, n.d.). Ενώ ως προς τη χώρα προέλευσης ο Ghenai αναφέρει ότι οι αποφάσεις μιας χώρας για παράδειγμα στον τομέα υγείας όπως η διακοπή του εμβολίου για τον HPV στην Ιαπωνία μπορεί να επιφέρει φήμες και συζητήσεις σε πολλές χώρες ανά τον κόσμο για την ασφάλεια του εμβολίου (Cohen et al., 2020; Ghenai & Mejona, 2017).

Η Papadopoulou επεσήμανε και την *αξιοπιστία των σχολίων* ως ένα ακόμη χαρακτηριστικό κοινωνικού περιεχομένου. Χαρακτηριστικά όπως το *μέγεθος κειμένου (σχολίου)*, ο *αριθμός των λέξεων*, το *αν περιλαμβάνει θαυμαστικό, ερωτηματικό, τον αριθμό των θαυμαστικών και ερωτηματικών αντίστοιχα*, *χαρακτηριστικά emoticon (χαρούμενο, θλιμμένο)*, *ύπαρξη πρώτου, δευτέρου, τρίτου προσώπου σε ουσιαστικό*, καθώς και του *αριθμού κεφαλαίων χαρακτήρων*. Επιπλέον, πρόσθεσε και την *εμφάνιση λέξεων με θετικά και αρνητικά συναισθήματα στο σχόλιο*, *λέξεων αργκό*, *αν έχει σύμβολο «:» ή «παρακαλώ» (Boolean)* καθώς και το *σκορ αναγνωσιμότητας του σχολίου* (Papadopoulou et al., 2019).

Οι Shu, Sotirakou, Reis, καθώς και άλλοι συγγραφείς της τρέχουσας βιβλιογραφίας συμπεριέλαβαν και τις *αντιδράσεις των χρηστών στα κοινωνικά μέσα (engagement)* (π.χ. likes, shares, αλληλεπιδράσεις) (Baeth & Aktas, 2019; Cohen et al., 2020; Meel & Vishwakarma, 2020; Reis et al., 2019; Shu, Wang, et al., 2019; Sotirakou et al., 2021; Taskin et al., 2022). Ο *αριθμός διαμοιρασμένων δημοσιεύσεων* (Baeth & Aktas, 2019; Cohen et al., 2020; Reis et al., 2019; Shu, Wang, et al., 2019; Sotirakou et al., 2021; Taskin et al., 2022) μπορεί να συνδεθεί με την αληθοφάνεια του περιεχομένου των ειδήσεων σύμφωνα με την Sotirakou (Sotirakou et al., 2021). Ο *αριθμός αντιδράσεων «μου αρέσει» (likes)* (Cohen et al., 2020; Reis et al., 2019; Sotirakou et al., 2021), *αγάπης (love)* που αναπαριστούν την αποδοχή και την εμπάθεια (Cohen et al., 2020; Sotirakou et al., 2021) αλλά και *αντιδράσεις έκπληξης, γέλιο, λύπης* καθώς και *μη αποδοχή της δημοσίευσης (θυμωμένος)* είναι μερικά από τα χαρακτηριστικά που υιοθετήθηκαν από τη Sotirakou (Sotirakou et al., 2021).

Οι Sotirakou και Reis υπολόγισαν τον *αριθμό των σχολίων των χρηστών*. Πιο συγκεκριμένα, ο Reis μέτρησε τον αριθμό των σχολίων σε διάφορα χρονικά διαστήματα από την ώρα δημοσίευσης τους στο Facebook (δευτερόλεπτα) (π.χ. 900, 1800, 2700) (Reis et al., 2019; Sotirakou et al., 2021). Ενώ η Sotirakou υπολόγισε τον συνολικό αριθμό αλληλεπιδράσεων, και τον βαθμό υπεραπόδοσης (Sotirakou et al., 2021). Η αυτόματη μετρική υπεραπόδοσης (overperforming score) υπολογίζεται μέσω της πλατφόρμας Crowdtangle (CrowdTangle, n.d.) και βασίζεται στην απόδοση των συναφών δημοσιεύσεων από την ίδια σελίδα σε παρόμοιες χρονικές περιόδους (π.χ. ένα βίντεο των New York Times

που δημοσιεύτηκε στα τελευταία 15 λεπτά με υψηλό βαθμό υπεραπόδοσης σημαίνει ότι το βίντεο είχε περισσότερες αλληλεπιδράσεις από το βίντεο των NYT που δημοσιεύτηκε τα προηγούμενα 15 λεπτά) (*CrowdTangle Codebook*, n.d.).

Ο Reis υπολόγισε τις *μετρικές απασχόλησης* (π.χ. σελίδες Facebook με υπολογισμό των «μιλάνε γι' αυτό», υπολογισμός «θαυμαστών-οπαδών») υιοθετώντας το χαρακτηριστικό από προηγούμενες έρευνες (Ebrahimi et al., 2017; Reis et al., 2019; Vosoughi et al., 2018) ενώ ο Zhang υπολόγισε την *αξιοπιστία των ισχυρισμών με βάση τη στάση των χρηστών-stance* (π.χ. υποστηρικτική, ουδέτερη, αντίθετη) (Q. Zhang et al., 2019).

Οι Meel και Vishwakarma, Baeth και Aktas και Weiss υπολόγισαν τα *χαρακτηριστικά μηνυμάτων ή δημοσιεύσεων των χρηστών* όπως το *μήκος κειμένου* (Weiss et al., 2019), *αν στο μήνυμα υπάρχει URL, επισήμανση (μήνυμα @)* (Meel & Vishwakarma, 2020; Weiss et al., 2019), *αριθμός επισημάνσεων* (Meel & Vishwakarma, 2020), *μήνυμα hashtag* (Meel & Vishwakarma, 2020; Weiss et al., 2019), *αριθμός hashtag, μήνυμα πηγής, η ημερομηνία δημιουργίας του μηνύματος, ο αριθμός δημοσιεύσεων, η ώρα δημοσίευσης* (Meel & Vishwakarma, 2020), *την κατάσταση δημοσίευσης από την στιγμή που δημιουργήθηκε* (Baeth & Aktas, 2019), *αν έχει αναδημοσιευτεί* (Baeth & Aktas, 2019; Meel & Vishwakarma, 2020), *τη μέτρηση αναδημοσιεύσεων, αν συμπεριλαμβάνει emoticon* (Meel & Vishwakarma, 2020), *αν έχει «μου αρέσει» καθώς και αντίστοιχο αριθμό, αν εμπεριέχει δημοσίευση, απάντηση, «μου αρέσει» φίλων, σχόλια φίλων* (Baeth & Aktas, 2019).

Το χαρακτηριστικό που αναφέρεται στο αν υπάρχει στο μήνυμα URL αναφέρεται ως χαρακτηριστικό της παραπληροφόρησης και των φημών σε έρευνες των Weiss και Meel και Vishwakarma. Σύμφωνα με τον Weiss οι πληροφορίες που δεν είναι φήμες συνήθως έχουν url (Meel & Vishwakarma, 2020; Weiss et al., 2019). Ενώ η επισήμανση είναι μία ένδειξη εάν έχει αναφέρει κι άλλους χρήστες στο μήνυμα δηλαδή εάν αναφέρει κάποιον χρήστη αποτελεί μη αξιόπιστη πληροφόρηση (Aphiwongsophon & Chongstitvatana, 2018; Castillo et al., 2011; Meel & Vishwakarma, 2020; Weiss et al., 2019). Τα hashtags αποτελούν σύντομη επεξήγηση του μηνύματος που ακολουθεί. Στην έρευνα των Indu και Trampi ήταν έντονη η παρουσία επαναλαμβανόμενων hashtag σε tweet φημών (π.χ. #followFriday) (Aphiwongsophon & Chongstitvatana, 2018; Castillo et al., 2011; Indu & Thampi, 2019; Meel & Vishwakarma, 2020; Weiss et al., 2019). Σύμφωνα με έρευνα του Castillo τα tweet που συμπεριλαμβάνουν emoticon αποτελούν μη αξιόπιστες πηγές (Castillo et al., 2011; Meel & Vishwakarma, 2020).

Τα χαρακτηριστικά συμμετοχικότητας (engagement) των χρηστών στα κοινωνικά δίκτυα όπως *αριθμοί δημοσιεύσεων (tweet), αριθμοί αναδημοσιεύσεων (retweet), αριθμοί «μου αρέσει»*, αλλά και *διαμοιρασμένες δημοσιεύσεις* υιοθετήθηκαν από συγγραφείς της

παρούσας βιβλιογραφίας (Baeth & Aktas, 2019; Cohen et al., 2020; Reis et al., 2019; Shu, Wang, et al., 2019; Sotirakou et al., 2021; Taskin et al., 2022) όπως ο Taskin ο οποίος τα ενσωμάτωσε σε αλγόριθμους (όπως το Τυχαίο Δάσος-Random Forest) για τον εντοπισμό των ψευδών ειδήσεων (Taskin et al., 2022). Ο αριθμός «μου αρέσει» αποτελεί ένα ακόμη χαρακτηριστικό που υιοθετήθηκε από τους συγγραφείς της παρούσας βιβλιογραφίας (Cohen et al., 2020; Reis et al., 2019; Sotirakou et al., 2021). Σύμφωνα με την Sotirakou ο αριθμός των like στο Facebook έχει χαρακτηριστεί ως ένδειξη hoax. Αλλά και στην μελέτη του Tacchini διαπιστώθηκε ότι οι δημοσιεύσεις ισχυρισμών (hoax) έχουν περισσότερα likes (π.χ. 204.5 αριθμός like σε δημοσιεύσεις hoax σε σχέση με 84 που δεν ήταν hoax) (Sotirakou et al., 2021; Tacchini et al., 2017).

Ο βαθμός υπεραπόδοσης είναι μία αυτόματη μετρική (overperforming score) η οποία υπολογίζεται μέσω της πλατφόρμας Crowdtangle (Sotirakou et al., 2021) και βασίζεται στην απόδοση των συναφών δημοσιεύσεων από την ίδια σελίδα σε παρόμοιες χρονικές περιόδους (π.χ. Overperforming ≥ 1.0 x, Underperforming ≤ -1.0 x). Για παράδειγμα ένα βίντεο των New York Times που δημοσιεύτηκε στα τελευταία 15 λεπτά με υψηλό βαθμό υπεραπόδοσης σημαίνει ότι το βίντεο είχε περισσότερες αλληλεπιδράσεις από το βίντεο των NYT που δημοσιεύτηκε το προηγούμενο 15λεπτο (*CrowdTangle Codebook*, n.d.; *How Do You Calculate Overperforming Scores?*, n.d.).

Οι Weiss, Meel και Viswakarma, Souza Freire μέτρησαν τα *χαρακτηριστικά διάδοσης* που υιοθετούνται από *δέντρα αναπαραγωγής (διάδοσης)* και κυρίως από μηνύματα αναδημοσιεύσεων του Twitter (retweet) (Meel & Vishwakarma, 2020; Souza Freire et al., 2021; Weiss et al., 2019).

Ο Weiss υιοθέτησε *χαρακτηριστικά όπως ώρα, μοναδική δημοσίευση, αναδημοσίευση και του μέσου όρου βάθους του κλαδιού (δέντρο αναπαραγωγής)* σε επίπεδο συμπεριφορικής διάδοσης των ισχυρισμών στο Twitter. Σύμφωνα με τον συγγραφέα οι πληροφορίες που δεν εμπίπτουν σε φήμες έχουν περισσότερες αλληλεπιδράσεις και αναδημοσιεύσεις ενώ οι αληθείς φήμες παράγουν λιγότερες συζητήσεις (Weiss et al., 2019). Σύμφωνα με τον Kwon τα γεγονότα που αφορούν φήμες εμφανίζουν επαναλαμβανόμενες κορυφώσεις ανά χρονικά διαστήματα και αυτό οδηγεί σε ένα κυκλικό φαινόμενο ενώ τα γεγονότα που δεν είναι φήμες εμφανίζουν μεγάλη κορύφωση αρχικά αλλά κατά το πέρασμα του χρόνου μειώνονται (Kwon et al., 2017; Weiss et al., 2019).

Οι Meel και Viswakarma στη μελέτη τους αναφέρουν ως πρόσθετα χαρακτηριστικά τον *βαθμό της πηγής (ρίζας) του δέντρου διάδοσης, το σύνολο των αριθμών tweet (μεγάλη υποδιακλάδωση ρίζας), το μέγιστο αλλά και τον μέσο όρος βαθμού του κόμβου (εκτός ρίζας),*

το βάθος του δέντρου διάδοσης, το μέγιστο μέγεθος επιπέδου (δέντρο διάδοσης) καθώς και το μέγιστο μέγεθος υποδιακλάδωσης. Τα συγκεκριμένα χαρακτηριστικά αναφέρονται στις μελέτες των Castillo και Jin για την αξιοπιστία της πληροφορίας στο Twitter και για τον εντοπισμό των φημών (Castillo et al., 2011; Jin, Cao, Guo, et al., 2017; Meel & Vishwakarma, 2020). Το βάθος του δέντρου διάδοσης αναφέρεται στη μελέτη των Meel & Vishwakarma και υιοθετήθηκε παλιότερα από τον Castillo (π.χ. 0= άδειο δέντρο, 1=μόνο αρχικά tweet, 2= μόνο αναδημοσιεύσεις από τη ρίζα (και ο μέσος όρος ανά δεσμό) (Castillo et al., 2011; Meel & Vishwakarma, 2020).

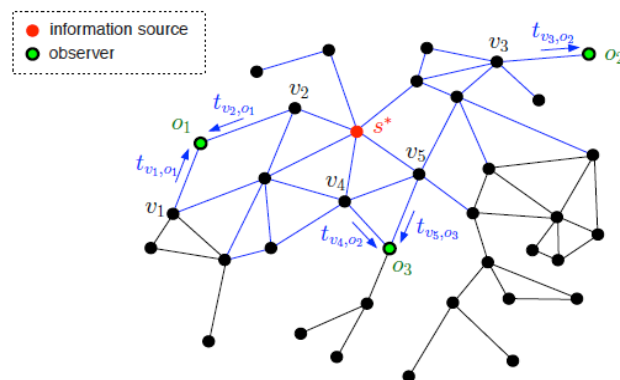
Αξίζει να επισημανθεί ότι σύμφωνα με τους Meel και Vishwakarma τα χαρακτηριστικά διάδοσης και η δυναμική μετάδοσης έχουν κοινά στοιχεία με *επιδημιολογικά μοντέλα* (π.χ. Sardana, SIR, κ.α.) τα οποία παίρνουν την διάδοση της παραπληροφόρησης ως ιογενή (virus). Στα συγκεκριμένα μοντέλα οι χρήστες που υιοθετούν ή αναμεταδίδουν τις ψευδείς ειδήσεις αναφέρονται ως μολυσμένοι ενώ εκείνοι που μόνο τις άκουσαν αλλά δεν τις διαδίδουν ως «ύποπτοι» (Meel & Vishwakarma, 2020).

Η ανάλυση κοινωνικών δικτύων όπως το *δίκτυο φίλων* (Meel & Vishwakarma, 2020; Taskin et al., 2022), το *δίκτυο δυναμικών σχέσεων* (Meel & Vishwakarma, 2020), και το *δίκτυο διάχυσης* (Meel & Vishwakarma, 2020; Paluch et al., 2020; Taskin et al., 2022) αποτελούν βασικά χαρακτηριστικά της εν λόγω ανάλυσης και επηρεάζονται από τις σχέσεις των χρηστών-φίλων, τις σχέσεις των χρηστών με άλλους χρήστες, καθώς και την διάχυση των πληροφοριών από τους χρήστες στα κοινωνικά δίκτυα. Η συγκεκριμένη ανάλυση έχει επηρεαστεί από τις κοινωνιολογικές επιστήμες και τη θεωρία δικτύων.

Το *δίκτυο δυναμικών σχέσεων* αναφέρεται στη μελέτη των Meel και Vishwakarma και υιοθετήθηκε από τους Ishida και Kuraya. Στο συγκεκριμένο δίκτυο δημιουργείται ένα δίκτυο αξιολόγησης της αξιοπιστίας. Σκοπός του είναι: α) η δυναμικότητα δηλαδή η αξιολόγηση του κάθε δεσμού εξαρτάται από την αξιολόγηση των συνδεδεμένων δεσμών. Αυτό αλλάζει και όταν αλλάζει η αξιοπιστία άλλων δεσμών και ακόμα και όταν προστίθεται ένας νέος δεσμός ή διαγράφεται υπάρχον δεσμός. Η σχετικότητα (β) δηλαδή η αξιοπιστία κάθε δεσμού αλληλεπιδρά με την αξιοπιστία και αξιολόγηση άλλων δεσμών και άρα η αξιοπιστία δεν μπορεί να βασίζεται μόνο στο περιεχόμενο του δεσμού στόχου και γ) αμοιβαιότητα δηλαδή η αξιολόγηση της αξιοπιστίας γίνεται κοινά με την αξιολόγηση της αξιοπιστίας των συνδεδεμένων δεσμών (περισσότερο από δύο δεσμοί χρειάζονται και η ψευδής αξιολόγηση θα μειώσει την αξιολόγηση του δεσμού) (Ishida & Kuraya, 2018; Meel & Vishwakarma, 2020).

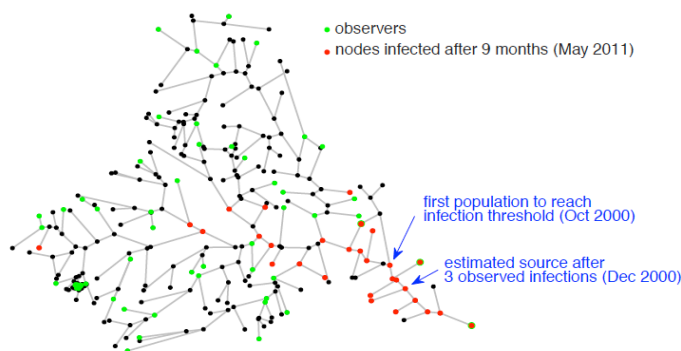
Το δίκτυο διάχυσης³⁷ δανείζεται από επιδημιολογικά μοντέλα ιατρικής, όπου όπως αναφέρθηκε και παραπάνω, η παραπληροφόρηση αναγνωρίζεται ως ένας «ιός» ο οποίος πρέπει να καταπολεμηθεί από την πηγή και που οι παραπληροφορημένοι χρήστες θεωρούνται «μολυσμένοι» (infected). Το συγκεκριμένο μοντέλο δικτύου είναι ένας γράφος που περιλαμβάνει tweet χρηστών ως δεσμούς και εκτιμώμενα μονοπάτια διάδοσης αλλά και κοινωνικές σχέσεις (ακολούθηση είδησης ενός δημιουργού από ένα άλλον) ως άκρες (Monti et al., 2019). Για την Shu το δίκτυο διάχυσης είναι σημαντικό γιατί μαθαίνει αναπαραστάσεις δομής και μοτίβων χρόνου και βοηθάει στον εντοπισμό των ψευδών ειδήσεων. Ο εντοπισμός της πηγής και των μονοπατιών της διάδοσης ανάμεσα στους χρήστες μειώνει το πρόβλημα (Meel & Vishwakarma, 2020; S. Yang et al., 2019).

Ο Paluch υιοθέτησε τον αλγόριθμο εντοπισμού πηγής (*Pinto Thiran Vetterli -PTVA Li*) για τον εντοπισμό της πηγής παραπληροφόρησης (Paluch et al., 2020). Ο συγκεκριμένος αλγόριθμος παίρνει ως δεδομένο το δίκτυο δεσμών (συνδέσεων) και το σύνολο από εντοπιστές (δεσμούς) που αναφέρουν την ώρα μόλυνσης. Στόχος του αλγόριθμου είναι η εύρεση της αρχικής πηγής μόλυνσης (το σημείο 0) (Εικόνα 143, Εικόνα 144).



Εικόνα 143. Αλγόριθμος εντοπισμού πηγής (παρατηρητές).

³⁷ δεξ και (Systems Innovation, n.d.)



Εικόνα 144. Αλγόριθμος εντοπισμού πηγής (εντοπιστές).

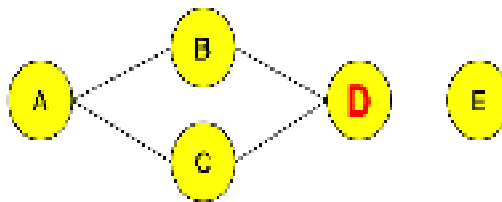
Ο ίδιος συγγραφέας υιοθέτησε τη μέθοδο συλλογικής κεντρικότητας, την ενδιάμεση κεντρικότητα, τον υψηλό βαθμό κάλυψης, το μέσο K , την υψηλή διαφορετικότητα των παρατηρητών, την τυχαία επιλογή εντοπιστών, και την μετρική ομοιότητας ανάμεσα στο σύνολο των εντοπιστών.

Η μέθοδος συλλογικής κεντρικότητας (*collective betweenness*) προσδιορίζει ένα σύνολο από εντοπιστές (S), τον αριθμό κοντινών μονοπατιών ανάμεσα σε δεσμούς (i) , (j) και το σύνολο του αριθμού των κοντινών μονοπατιών ανάμεσά τους (σ_{ij}). Στη συγκεκριμένη μέθοδο κάθε κοντινό μονοπάτι υπολογίζεται μία φορά.

Η ενδιάμεση κεντρικότητα (*betweenness centrality*) είναι ένας αλγόριθμος που μετρά τους δεσμούς με την υψηλότερη κεντρικότητα σχέσεων. Ουσιαστικά προσδιορίζει το πλησιέστερο «μονοπάτι»-γέφυρα και υπολογίζει πόσες φορές κάθε δεσμός συμπίπτει πάνω σε αυτό (Disney, 2020). Αν υπάρχει υψηλή κεντρικότητα τόσο ο συγκεκριμένος δεσμός (ο οποίος «ασκεί εξουσία») επηρεάζει διαφορετικές ομάδες σε ένα δίκτυο (Εικόνα 145). Για παράδειγμα αν έχουμε τους δεσμούς A B C D E οι δεσμοί B και C μπορούν να διαγραφούν γιατί θα εξακολουθήσει να υπάρχει μονοπάτι που οδηγεί στον E όμως ο D δεν μπορεί γιατί συμπίπτει πάνω σε μονοπάτι-γέφυρα που οδηγεί στον δεσμό E (EMBL-EBI, n.d.) (Εικόνα 146).



Εικόνα 145. Υψηλή κεντρικότητα δεσμών σε δίκτυο.



Εικόνα 146. Παράδειγμα κεντρικότητας.

Ο υψηλός βαθμός κάλυψης είναι ένας αλγόριθμος ο οποίος επιλέγει ένα σύνολο από εντοπιστές που έχουν μέγιστο αριθμό μοναδικών γειτόνων. Οι δεσμοί που έχουν ένα εντοπιστή σαν γείτονα είναι «καλυπτόμενοι» και η αναλογία επικαλυπτόμενων δεσμών ονομάζεται «βαθμός κάλυψης». Ο μέσος K ως μέθοδος αφορά την ελαχιστοποίηση του συνόλου των αποστάσεων ανάμεσα στους δεσμούς και στους κοντινούς εντοπιστές.

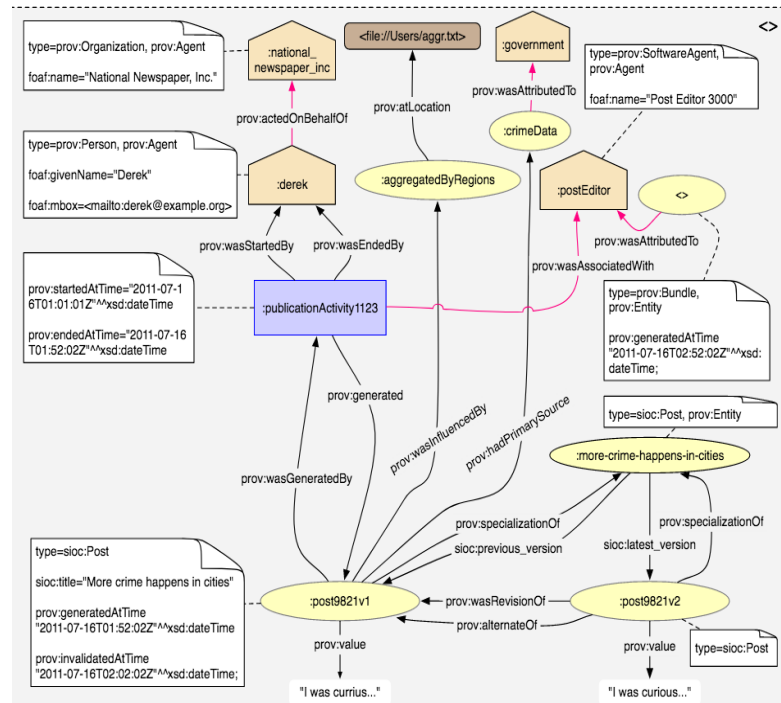
Η υψηλή διαφορετικότητα των παρατηρητών βασίστηκε στη μέθοδο κάλυψης *Spinelli* (Spinelli et al., 2016) και ερευνά το σύνολο των εντοπιστών που μεγιστοποιεί την ιδιότητα των δεσμών³⁸. Αν για παράδειγμα ένας δεσμός βρίσκεται σε ένα κοντινό μονοπάτι απόστασης (L) ανάμεσα σε δύο εντοπιστές ονομάζεται «μοναδικό μονοπάτι κάλυψης» (Paluch et al., 2020).

Οι Meel και Viswakarma και Taskin επισημαίνουν και τη χρήση γράφων δικτύων (π.χ. Twitter, Facebook) (*Graph-Tool: Efficient Network Analysis with Python*, n.d.) για την αναπαράσταση των σχέσεων και των αλληλεπιδράσεων μεταξύ των χρηστών όπως οι φιλικές σχέσεις και η εξαγωγή χαρακτηριστικών χρόνου, διάχυσης κ.λ.π. (Meel & Vishwakarma, 2020;

³⁸ Πόσο κοντά βρίσκεται ένα σύνολο δεσμών σε ένα μονοπάτι απόστασης L ανάμεσα σε 2 εντοπιστές σε ένα σύνολο δεσμών S .

Taskin et al., 2022). Το συγκεκριμένο χαρακτηριστικό υιοθετήθηκε και από προηγούμενες έρευνες ως μέθοδος αξιοπιστίας της πληροφόρησης (Bovet & Makse, 2019; Galán-García et al., 2015; Monti et al., 2019; Vosoughi et al., 2017). Οι Bovet & Makse τον χρησιμοποίησαν για την επιρροή των ψευδών ειδήσεων κατά τη διάρκεια των εκλογών του 2016 για την ανάδειξη των 100 κορυφαίων μεταδοτών ειδήσεων σε ψευδείς ειδήσεις (άκρα δεξιά, άκρα αριστερά, κέντρο, αριστερά, δεξιά) ανάλογα με την επιρροή (Bovet & Makse, 2019; Meel & Vishwakarma, 2020). Εκτός από τον γράφο δικτύου οι Meel και Vishwakarma προσθέτουν και το γράφο αναδημοσιεύσεων από το κοινωνικό δίκτυο ο οποίος χρησιμοποιήθηκε και από άλλες έρευνες (Bovet & Makse, 2019; Ma et al., 2017; Meel & Vishwakarma, 2020; Rath et al., 2017).

Οι Baeth και Aktas χρησιμοποίησαν τον γράφο του ιστορικού των πληροφοριών των χρηστών στα κοινωνικά δίκτυα (οντολογία PROV-O) (PROV-DM: The PROV Data Model, n.d.) για την αναπαράσταση μεταδεδομένων υπό μορφή XML. Ο συγκεκριμένος γράφος αφορά μετρικές συμμετεχόντων χρηστών και αντιδράσεις που δημιουργήθηκαν με την αλλαγή κατάστασης της οντότητας στόχου (π.χ. εάν υπάρχει θετική ανατροφοδότηση για κάθε χρήστη) (Baeth & Aktas, 2019) (Εικόνα 147).



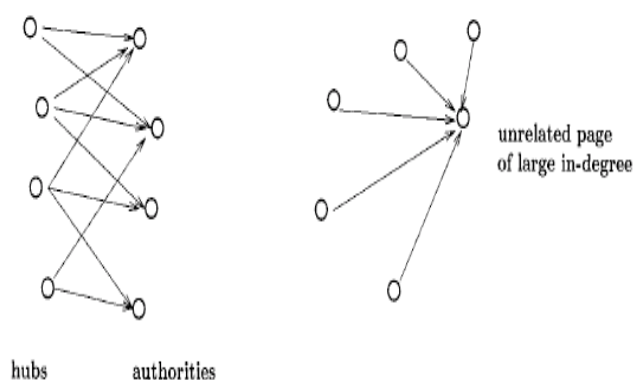
Εικόνα 147. Παράδειγμα οντολογίας PROV-O.

Η οντολογία PROV-O αποτελείται από την δραστηριότητα, τον πράκτορα, και την οντότητα. Η δραστηριότητα περιλαμβάνει την υποστήριξη ή αποδοχή μιας δήλωσης μιας

οντότητας ή περιεχομένου, την παρατήρηση, την κριτική, την επισήμανση, τη δημόσια διανομή κειμένου ή γραφικού υλικού, την ακολούθηση ή παρακολούθηση της κίνησης ή της εξέλιξης κάποιου, την αναπαραγωγή, τη μετάδοση, τη διάχυση (*PROV-O: The PROV Ontology*, n.d.). Ο *πράκτορας* αποτελείται από άτομο που ακολουθεί ένα άλλο άτομο και εγγράφεται στη ροή των tweet του (ακόλουθος-follower) και ένα άτομο που παρακολουθείται σε κοινωνική ιστοσελίδα δικτύωσης ή εφαρμογή (ακολουθεί άλλο άτομο-follower). Η *οντότητα* από το ιστολόγιο ή το μήνυμα στην ώρα κατάστασης δημιουργίας του από τον δημιουργό (πρωτότυπο tweet) αλλά και από την αναθεώρηση και τροποποίηση υπό την εμφάνιση των καινούργιων στοιχείων (π.χ. απαντήσεις, αναθεωρημένο tweet) (Baeth & Aktas, 2017).

Οι Taskin και Shim χρησιμοποίησαν τον *αλγόριθμο PageRank* (Gephi org, n.d.) ως μέθοδο ανάλυσης των κοινωνικών δικτύων για την κατάταξη των ιστοσελίδων. Ο συγκεκριμένος αλγόριθμος παίρνει ως δεδομένο την πιθανότητα ενός τυχαίου χρήστη να επισκεφτεί μια ιστοσελίδα (π.χ. υψηλή βαθμολογία ιστοσελίδας αν πολλές σελίδες οδηγούν σε αυτή) (Shim et al., 2021; Taskin et al., 2022).

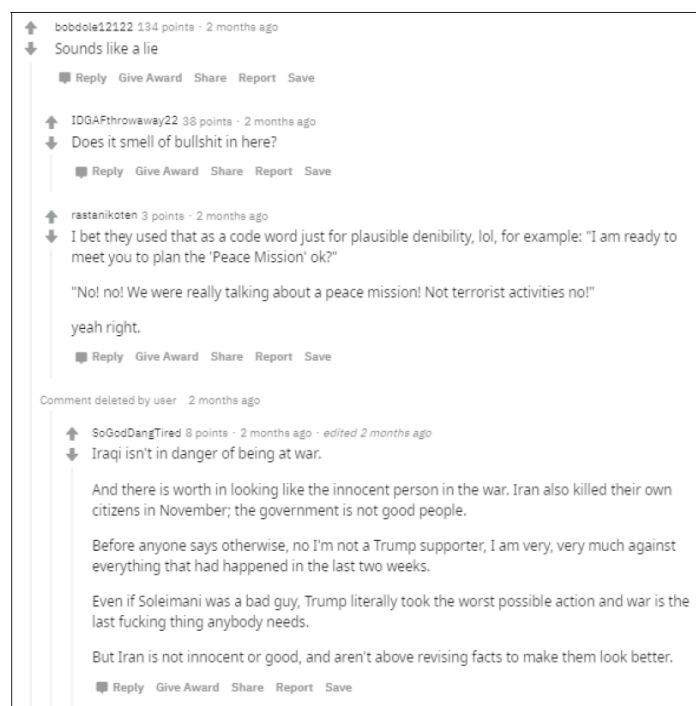
Ο *αλγόριθμος ανάλυσης συνδέσμων (Hits) (Hyperlink Induced Topic Search)* (Gephi org, n.d.) υιοθετήθηκε από τον Taskin. Ο αλγόριθμος Hits βαθμολογεί ιστοσελίδες και αξιολογεί τη σχέση ανάμεσα στους δεσμούς ενός γράφου (μικρός υπογράφος). Περιλαμβάνει την αξιόπιστη πληροφορία δεσμού (authority score), και τον κόμβο-κέντρο (hub) δηλαδή την αξία της διασυνδεδεμένης πληροφορίας με άλλους συνδεδεμένους δεσμούς (A. Kim, 2019) (π.χ. ένας καλός κόμβος είναι ιστοσελίδα που υποδεικνύει αξιόπιστη πληροφόρηση) (Taskin et al., 2022) (Εικόνα 148).



Εικόνα 148. Αλγόριθμος Hits.

Ο Cohen για αντιμετωπίσει την παραπληροφόρηση στα κοινωνικά δίκτυα Reddit και Facebook βασίστηκε στο μοντέλο πολλών πρακτόρων (Sardana). Το *μοντέλο Sardana* αποτελεί ένα σύστημα πολλών πρακτόρων βασισμένο στην σύσταση μηνύματος

εμπιστοσύνης. Το μοντέλο φιλτράρει μηνύματα για ένα συγκεκριμένο πράκτορα βασισμένο στην συμβουλή άλλων πρακτόρων³⁹ από το δίκτυο. Στόχος του μοντέλου είναι να δώσει στους χρήστες όσα μηνύματα είναι αξιόλογα και να φιλτράρει εκείνα που είναι μη αξιόπιστα. Μέσω του συνόλου των συμβούλων που απάντησαν και σχολίασαν στο μήνυμα (αξιολόγηση αναξιοπιστίας), την αξιοπιστία τους, τις απαντήσεις στο μήνυμα (δηλαδή τα σχόλια και τις ψήφους) καθώς και τις παρελθοντικές βαθμολογίες τους (πράκτορα και συμβούλων) παράγεται μια τριπλέτα παρατήρησης (βαθμολόγηση με τιμές 0, 1 για καλό/κακό μήνυμα, ομοιότητα, βαθμολόγηση αναξιοπιστίας, ομοιότητα αναξιοπιστίας, αξιοπιστία). Στο τελικό στάδιο το μήνυμα θα εμφανιστεί μέσα από συνδυασμό του να υπάρχει παραπληροφόρηση, την σφοδρότητα (κάποια τιμή αν παρερμηνευθεί το μήνυμα), και την ανοχή του χρήστη (Cohen et al., 2020) (Εικόνα 149).



Εικόνα 149. Εμφάνιση μηνυμάτων μέσω μοντέλου Sardana.

Το μοντέλο ανεξάρτητου δικτύου ροής (*Independent Cascade model*) είναι ένα μοντέλο εντοπισμού πηγής το οποίο αποτελείται από ένα σύνολο δεσμών-πηγών όπου η κάθε πηγή-δεσμός λαμβάνει μια πληροφορία και ενεργοποιείται (Amoruso et al., 2020; Meel & Vishwakarma, 2020). Αφού ενεργοποιηθεί ο δεσμός μετέπειτα προσπαθεί να επηρεάσει

³⁹ Στο συγκεκριμένο μοντέλο ως “πράκτορες” θεωρούνται οι χρήστες που δημοσιεύουν μια είδηση ενώ οι “σύμβουλοι” είναι οι άλλοι χρήστες που σχολίασαν ή έδωσαν θετική ή αρνητική ψήφο.

και τους μη ενεργούς γειτονικούς δεσμούς. Όταν κανείς δεσμός δεν παραμένει ενεργοποιημένος τερματίζεται η διαδικασία. Η πιθανότητα του αν ο δεσμός ενεργοποιηθεί εξαρτάται από την πιθανότητα επιρροής του δεσμού σε άλλον δεσμό (Nguyen et al., 2012)⁴⁰.

Το μοντέλο πιθανοτικής μόλυνσης (*Susceptible infected model*) (*The SIR Model for Spread of Disease - The Differential Equation Model | Mathematical Association of America*, n.d.) υιοθετήθηκε από τον Paluch για τη διάχυση της πληροφορίας στα κοινωνικά δίκτυα και αναφέρεται και στη μελέτη των Meel και Vishwakarma (Meel & Vishwakarma, 2020; Paluch et al., 2020). Το συγκεκριμένο μοντέλο (επιδημιολογικό) βασίζεται στην πιθανότητα ανά χρονικό διάστημα ένας μολυσμένος δεσμός να μεταφέρει τη μόλυνση σε ένα μη μολυσμένο γείτονα. Τα χαρακτηριστικά του μοντέλου σχετίζονται με τον βαθμό μόλυνσης (infection rate) και την αναλογία ανάμεσα στη σταθερή απόκλιση και το μέσο όρο του αριθμού των χρονικών ρυθμών που απαιτούνται για τη μετάδοση της μόλυνσης (transmission variance (ξ)) (Bailey, 1975; Paluch et al., 2020; Spinelli et al., 2016).

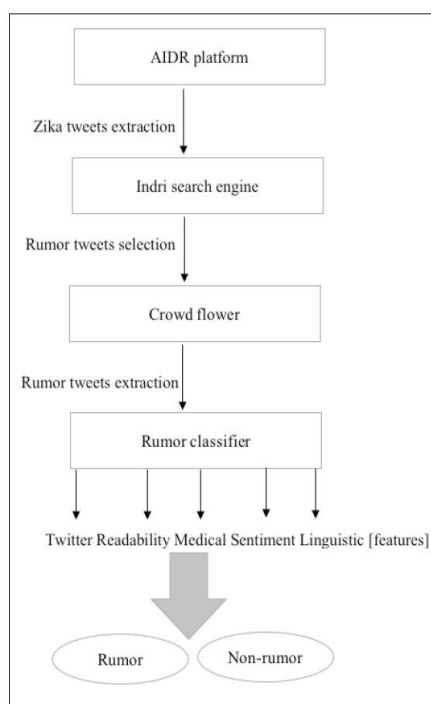
Οι Meel & Vishwakarma αναφέρουν στη μελέτη τους και μια σειρά παρεμφερών μοντέλων διάδοσης πληροφοριών τα οποία έχουν χρησιμοποιηθεί σε προηγούμενες έρευνες (Dhar et al., 2016; Dong et al., 2017; Han et al., 2017; Y. Liu et al., 2016; Luo et al., 2013; Nguyen et al., 2012; Shah & Zaman, 2011; Sharma et al., 2019; Turenne, 2018; L. Zhao et al., 2012) για τον εντοπισμό της παραπληροφόρησης (π.χ. Stochastic model) και άλλα επιδημιολογικά μοντέλα εντοπισμού πηγής (π.χ. SI Model, SIS model, SIRS, SEI, SEIR) (Meel & Vishwakarma, 2020).

Εκτός από το μοντέλο πολλών πρακτόρων ο Cohen υιοθέτησε το μοντέλο *Ghenai* και το μοντέλο *Parmentier* για τον εντοπισμό των ψευδών ειδήσεων στο Twitter. Το μοντέλο *Ghenai* πρωτοχρησιμοποιήθηκε από τον Ghenai με αποτελεσματικότητα ως προς τον εντοπισμό ψεύτικων θεραπειών για τον καρκίνο αλλά και σε ψεύτικες φήμες για τον ιό Ζίκα (Ghenai, 2019; Ghenai & Mejova, 2017, 2018).

Στο συγκεκριμένο μοντέλο γίνεται απομόνωση της παραπληροφόρησης μέσω γνώσης ειδικών, πληθοπορισμού και αλγορίθμων μηχανικής μάθησης. Η μελέτη έγινε από τον Ghenai προκειμένου να ταυτοποιηθούν οι χρήστες που προωθούσαν τις ψεύτικες θεραπείες. Ο αλγόριθμος του μοντέλου *Ghenai* αποτελείται από υποσύνολα χρηστών καθώς και χρήστες που αναδημοσιεύουν tweet ή retweet. Μέσω των συγκεκριμένων αναδημοσιεύσεων υπολογίζεται η αξιοπιστία του χρήστη (Εικόνα 150).

⁴⁰ Δες και (Kundu, 2013).

Για την βαθμολογία των tweet ο Cohen υιοθέτησε το μοντέλο *Parmentier*. Το μοντέλο *Parmentier* αποτελεί ένα πολυμοντελικό σύστημα το οποίο συνδυάζει την εμπιστοσύνη και την φήμη σε μία βαθμολογία. Αναπτύχθηκε για την μέτρηση της εμπιστοσύνης των χρηστών και του περιεχομένου που παράγουν (Parmentier & Cohen, 2020). Το μοντέλο βαθμολογεί τα tweet ανάλογα με τις απαντήσεις σε μορφή δένδρου. Η αξιολόγηση των δημοσίευσης γίνεται με βάση την μετρική Pearson Correlation Coefficient. Η συγκεκριμένη μετρική παίρνει ως χαρακτηριστικό την ομοιότητα μεταξύ του πρόσφατου χρήστη και του υποσυνόλου χρηστών δηλαδή αξιολόγηση σε ένα tweet με 1 (αν ο χρήστης επιδράσει θετικά-μου αρέσει), αξιολόγηση 0.5 (αν ο χρήστης επαναδημοσιεύσει το tweet) και αξιολόγηση -1 (αν ο χρήστης απαντήσει στο tweet και το συναίσθημα πέσει κάτω του ουδέτερου). Ο συγγραφέας πήρε ως δεδομένο τα tweets που είχαν αρκετές απαντήσεις υποθέτοντας ότι για να φτάσει στο συγκεκριμένο επίπεδο θα πρέπει το δέντρο απαντήσεων να έχει δύο υποδιακλαδώσεις (Cohen et al., 2020) (βλέπε ξανά Εικόνα 149).



Εικόνα 150. Μοντέλο Ghenaï για τον εντοπισμό των φημών.

Για το δίκτυο Reddit ο Cohen υιοθέτησε το μοντέλο *Sardana* με τα εξής χαρακτηριστικά: *είδος μηνύματος, βαθμολογία δικτύου μιας δημοσίευσης ή σχολιασμού* (π.χ. θετικότητα ή αρνητικότητα της αντίδρασης της κοινότητας σε ένα μήνυμα, σύγκριση και με αριθμό σχολίων), *τον αριθμό σχολίων* δηλαδή το ποσοστό προσοχής που λαμβάνει το

μήνυμα και τη σύγκριση του αριθμού των σχολίων με τον αριθμό ακολούθων από subreddit ή τον αριθμό σχολίων σε υψηλής βαθμολογίας δημοσιεύσεις σε αυτό.

Αλλά και χαρακτηριστικά όπως η υψηλότερη βαθμολογία σε σχόλιο που ακολουθείται με συναίσθημα δυσπιστίας, αν ήταν εσκεμμένο το μήνυμα δηλαδή αν ο συγγραφέας κατά λάθος ή εσκεμμένα προσπάθησε να διαδώσει παραποιημένες πληροφορίες (π.χ. αν ο συγγραφέας απολογηθεί σε επόμενα σχόλια συνιστά ένδειξη ατυχούς συμβάντος, αν αρνηθεί την μη εγκυρότητα με φτωχά επιχειρήματα υπάρχει ένδειξη σκοπιμότητας), τη βαθμολογία του συγγραφέα (κάρμα) για παρελθοντικές δημοσιεύσεις ή σχόλια (δείκτης αξιοπιστίας συγγραφέα), αν ο συγγραφέας είναι επανειλημμένα παραβάτης (αξιοπιστία συγγραφέα), αν είναι εύκολα ταυτοποιήσιμο, αν παρέχει υποστηρικτικά στοιχεία για την αξιοπιστία της δημοσίευσης, μη επαγγελματική χρήση της γλώσσας (αξιοπιστία δημοσίευσης).

Το συναίσθημα δυσπιστίας αποτελεί ένδειξη δυσπιστίας της κοινότητας σε ένα μήνυμα. Η υψηλή βαθμολογία συγκρίνεται με έναν αριθμό σχολίων ή τη βαθμολογία του δικτύου. Η εύκολη ταυτοποίηση της αξιοπιστίας του μηνύματος επηρεάζει την υψηλή βαθμολογία σε έναν επόμενο σχολιασμό με το αίσθημα μη εμπιστοσύνης εφόσον λίγα μέλη της κοινότητας θα μπορούσαν να πιστοποιήσουν το μήνυμα.

Στο Facebook ο Cohen μέτρησε την αξιοπιστία παίρνοντας δυαδικές συνεχόμενες τιμές 0, 1, για την μέτρηση της αξιοπιστίας του κατόχου της ιστοσελίδας, τις δημοσιεύσεις σε άλλες σελίδες (π.χ. αν η δημοσίευση του κατόχου της σελίδας μαζί με τον διαμοιρασμό του συνδέσμου επηρέασε την κοινότητα), τις δημοσιεύσεις σε ομάδες (αξιοπιστία κατόχου σελίδας ομάδας). Για την αξιοπιστία του συγγραφέα τον σύνδεσμο της πηγής (π.χ. εάν η δημοσίευση δημιουργήθηκε διαμοιράζοντας έναν σύνδεσμο από άλλη ιστοσελίδα). Για την βαθμολόγηση της δημοσίευσης τον αριθμό των αγαπημένων (hearts) αλλά και τον αριθμό «μου αρέσει» (π.χ. τελευταίες 100 δημοσιεύσεις χρήστη) και για την ομοιότητα πράκτορα-χρήστη, το ιστορικό βαθμολόγησης (Cohen et al., 2020).

Οι Meel και Vishwakarma τέλος, αλλά και άλλοι συγγραφείς της τρέχουσας βιβλιογραφίας αναφέρουν και τα χωροχρονικά χαρακτηριστικά ως χαρακτηριστικά των ψευδών ειδήσεων τα οποία έχουν υιοθετηθεί και σε προηγούμενες έρευνες αλλά και από την τρέχουσα βιβλιογραφία (Gravanis et al., 2019; B. D. Horne et al., 2020; Meel & Vishwakarma, 2020; Papadopoulou et al., 2019; Reis et al., 2019). Τα χωροχρονικά χαρακτηριστικά αφορούν πληροφορίες σχετικά με τοποθεσίες ή τη χωρική κατανομή ατόμων ή αντικειμένων ή πληροφορίες για το πότε έγινε το γεγονός ή περιγράφουν με ακρίβεια τη συχνότητα των γεγονότων (Gravanis et al., 2019; L. Zhou et al., 2004). Χαρακτηριστικά όπως

η χρονική κατανομή (π.χ. η υψηλή συχνότητα ροών σχεδόν πανομοιότυπων ψεύτικων βίντεο) (Meel & Vishwakarma, 2020; Papadopoulou et al., 2019; Vishwakarma et al., 2019), χρονικά χαρακτηριστικά φημών (*Periodic External Shock*) (Kwon et al., n.d.; Meel & Vishwakarma, 2020), η αναλογία δημοσιεύσεων χρηστών (*user engagement*) (Reis et al., 2019), η χρονική σήμανση ειδήσεων (Meel & Vishwakarma, 2020; Shu, Mahudeswaran, et al., 2019), η χρονική σήμανση απαντήσεων (Ruchansky et al., 2017; Q. Zhang et al., 2019), καθώς και η αναλογία νέων χρηστών (Meel & Vishwakarma, 2020; Vosoughi et al., 2017) υιοθετήθηκαν για τον εντοπισμό των ψευδών ειδήσεων.

Η χρονική κατανομή αφορά τη διακύμανση των σχεδόν ταυτόσημων βίντεο με βάση τη χρονολογία τους. Η Papadopoulou αναφέρει ότι η χρονική κλίμακα μεταξύ της πρώτης δημοσίευσης ενός βίντεο και του πιο πρόσφατου σχεδόν πανομοιότυπου βίντεο (διπλότυπου) φτάνει τα 10 χρόνια. Τα αληθινά σχεδόν πανομοιότυπα βίντεο είναι λιγότερα σε δημοσιεύσεις μετά από 10 μέρες σε σχέση με τα ψεύτικα τα οποία δημοσιεύονται σε υψηλότερη κλίμακα και για περισσότερο χρονικό διάστημα (Papadopoulou et al., 2019).

Τα χρονικά χαρακτηριστικά φημών αφορούν φήμες οι οποίες έχουν πολλαπλές και περιοδικές κορυφώσεις ενώ εκείνες που δεν είναι έχουν μία μοναδική κορύφωση και βασίζονται στο μοντέλο Περιοδικών Εξωτερικών Σοκ (PES). Το συγκεκριμένο μοντέλο υιοθετήθηκε από τον Kwon έχοντας ως βάση το μοντέλο Spike του Matsubara. Το μοντέλο PES περιγράφει τις φήμες από τον ημερήσιο κύκλο αλλά και από τον εξωτερικό κύκλο σοκ. Το συγκεκριμένο μοντέλο αποτελεί επέκταση του Susceptible Infected Model (Kwon et al., n.d.; Matsubara et al., 2012).

Η χρονική σήμανση ειδήσεων αφορά τη διάδοση ψευδών ειδήσεων στα κοινωνικά μέσα αλλά και τη χρονική αλλαγή θεμάτων τους (πχ. αναβάθμιση συλλογής tweet από Tweet Collector ανά ημέρα) (Meel & Vishwakarma, 2020). Η χρονική σήμανση απαντήσεων αφορά χρονολογική σειρά των απαντήσεων. Σύμφωνα με τον Zhang οι τελευταίες απαντήσεις είναι πιο αξιόπιστες σε σχέση με τις προηγούμενες ως προς την αληθοφάνεια ενός γεγονότος (Q. Zhang et al., 2019). Η αναλογία δημοσιεύσεων των χρηστών αφορά τον υπολογισμό των δραστηριοτήτων των σχολίων του χρήστη με βάση τον ρυθμό που δημοσιεύονται (Reis et al., 2019).

Ο Horne υιοθέτησε χαρακτηριστικά όπως ο *αριθμός τοποθεσιών (στο άρθρο)*, αλλά και τον *αριθμό ημερομηνιών ή ωρών (στο άρθρο)*. Σύμφωνα με τον συγγραφέα οι επαγγελματίες δημοσιογράφοι αναφέρουν και την ώρα και την τοποθεσία ενός γεγονότος ενώ εκδότες ειδήσεων που αναφέρουν αναληθή γεγονότα δεν βάζουν ώρα και τοποθεσία (π.χ. μεγάλα γεγονότα αλλάζουν τον διαχωρισμό ψευδών-αληθών ειδήσεων όπως η αύξηση

ψευδών ειδήσεων στις αμερικάνικες εκλογές του 2016) (Allcott & Gentzkow, 2017; B. D. Horne et al., 2020).

Τέλος ο Zhang υπολόγισε τη χρονολογική σειρά των απαντήσεων. Σύμφωνα με τον συγγραφέα οι τελευταίες απαντήσεις είναι πιο αξιόπιστες σε σχέση με τις προηγούμενες ως προς την αληθοφάνεια ενός γεγονότος (Q. Zhang et al., 2019).

Κεφάλαιο 5. Συζήτηση-συμπεράσματα

Στην παρούσα διπλωματική εργασία μελετήθηκαν οι θεωρητικές προσεγγίσεις του φαινομένου των ψευδών ειδήσεων καθώς και τα χαρακτηριστικά των αλγορίθμων. Αξίζει να επισημανθεί ότι το φαινόμενο των ψευδών ειδήσεων δεν είναι ένα φαινόμενο τωρινό αλλά έχει ιστορικές ρίζες από τις απαρχές της ανθρωπότητας. Η διπλωματική εργασία εστιάστηκε σε βασικά χαρακτηριστικά των αλγορίθμων τα οποία πάρθηκαν έπειτα από τη βιβλιογραφική επισκόπηση και επιλογή των άρθρων. Δε μελετήθηκαν τα email spam αλλά και τα bots καθώς αποτελούν υποπεριπτώσεις των ψευδών ειδήσεων οι οποίες όμως δεν αποτελούν πρωτογενείς επιρροές ως προς τη διάδοση.

Κατά τη βιβλιογραφική επισκόπηση ενδιαφέρον παρουσιάζει ότι η πλειονότητα των άρθρων ταυτίζει την παραπληροφόρηση και τη διασπορά των ψευδών ειδήσεων με την *προπαγάνδα (πολιτική, κρατική ή θρησκευτική)* αλλά και με τα *κοινωνικά μέσα δικτύωσης (Facebook, Twitter)*.

Χαρακτηριστικά παραδείγματα είναι το άρθρο του Alhazbi για τα πολιτικά τρολ τα οποία επηρεάζουν πολιτικά και δημιουργούν πόλωση στην κοινωνία και χειραγώγηση, η χρήση βάσεων δεδομένων από πολιτικές εκλογές (π.χ. αμερικάνικες εκλογές 2016), το άρθρο της Cremisini για την κρίση στην Ουκρανία (προσάρτηση Κριμαίας) (εντοπισμός μεροληψίας), το άρθρο του Kausar για την προπαγάνδα στη γλώσσα Urdu (μοντέλο ProSoul), το μοντέλο Propgy του Barrón-Cedeño. Επιπλέον η εμφάνιση του ιού Covid19 στις αρχές του 2020 και η διασπορά θεωριών συνωμοσίας γύρω από τον ιό έκανε περαιτέρω επιτακτική την ανάγκη η ερευνητική κοινότητα να αντιμετωπίσει το φαινόμενο ως έναν ιό ο οποίος μολύνει και επηρεάζει το κοινωνικό περιβάλλον και γι αυτό θα πρέπει να βρεθεί η πηγή διάδοσης. Τα μοντέλα Sardana, Zika, CoAID-Deer, αλλά και το μοντέλο ανεξάρτητου δικτύου ροής (Independent Cascade model) κινούνται προς αυτήν την κατεύθυνση με αξιόλογα αποτελέσματα.

Σε σχέση με τα χαρακτηριστικά των αλγορίθμων επισημαίνεται η ποικιλομορφία και η διαφορετικότητά τους καθώς στα πρώτα στάδια της έρευνας ήταν δυσδιάκριτες οι διαφορές και η κατηγοριοποίησή τους (π.χ. λεξιλόγιο, σύνταξη, σημασιολογία, θεματικά, υβριδικά). Εξαιρέση αποτελούν τα συναισθηματικά χαρακτηριστικά αλλά και τα ψυχολογικά καθώς και τα χαρακτηριστικά κοινωνικού περιεχομένου.

Συμπερασματικά λοιπόν **τα πιο αξιόλογα χαρακτηριστικά των ψευδών ειδήσεων είναι:** α) ως προς το περιεχόμενο: *η στίξη (π.χ. θαυμαστικά, ερωτηματικά), οι αντιθέσεις των*

συναισθημάτων (*polarity*), οι βρισιές, ο θυμός, η επιρροή, οι βαθμοί υποκειμενικότητας (π.χ. επίθετα, αντωνυμίες, επιρρήματα, λέξεις μεροληψίας) με χρήση λεξικών όπως η LIWC, τα βοηθητικά ρήματα, οι αντωνυμίες πρώτου προσώπου, η χρήση ρημάτων, η αβεβαιότητα, το τρίτο ενικό και τρίτο πληθυντικό πρόσωπο, ο αριθμός αρνητικών ή θετικών λέξεων σε μια δημοσίευση, το ποσοστό των μοναδικών λέξεων (πολυπλοκότητα του λεξιλογίου), το μήκος των προτάσεων, η αναγνωσιμότητα του κειμένου (χρήση μετρικών ανάγνωσης π.χ. *Flesh reading Ease*), η χρήση των συχνοτήτων των λέξεων (μιας λέξης-unigrams, δύο λέξεων-bigrams) με την αξιοποίηση αλγοριθμικών μοντέλων όπως το *Bag of words*, της συχνότητας των όρων σε ένα κείμενο (*Tf-Idf*), του *Glove*, η εύρεση στοχευμένων λέξεων μέσω του αλγορίθμου *CBOW*, η εύρεση λέξεων με συναφές περιεχόμενο (μοντέλο *Skipgram*), σχέσεις ανάμεσα σε προτάσεις (θεωρία ρητορικής δομής), οι συχνότητες των λέξεων-κλειδιών μέσω λογισμικών όπως το *Rake library*, η μέτρηση της ομοιότητας μεταξύ δύο διαφορετικών κειμένων (ή οντοτήτων) ανεξαρτήτου μεγέθους (*cosine similarity*), οι επισημάνσεις *POS*, χρήση αργκό ή νεολογισμών, χρήση υπερθετικών βαθμών και τροπικών επιρρημάτων β) ως προς το κοινωνικό περιεχόμενο: ο αριθμός των ακολούθων (*followers*) σε ένα λογαριασμό χρήστη (κοινωνικό μέσο) αλλά και ο αριθμός των χρηστών που ακολουθεί (*followees*), ο αριθμός πιστοποιημένων χρηστών, αριθμός *hashtag*, αριθμός *URL*, ο αριθμός των αναδημοσιεύσεων ενός *tweet*, ο αριθμός των προτιμήσεων του χρήστη (*likes*), η αξιοπιστία του χρήστη, ο διαμοιρασμός άλλων δημοσιεύσεων (αληθοφάνεια ειδήσεων), τα σχόλια, και το δίκτυο φίλων του χρήστη.

Πρέπει να διευκρινιστεί ότι δεν επιλέχθηκε το χαρακτηριστικό των κεφαλαίων γραμμάτων ως χαρακτηριστικό των ψευδών ειδήσεων καθώς στην ελληνική ειδησεογραφία πολλά άρθρα αληθή χρησιμοποιούν κεφαλαία γράμματα ως ένα μέσο επιρροής και εντυπωσιασμού αλλά και προσέλκυσης αναγνωστών.

Επιπροσθέτως, καθοριστικό ρόλο στην επιτυχία των αλγορίθμων συμβάλλει και ο έλεγχος της πληροφόρησης από παραδοσιακά *site* ελέγχου γεγονότων τα οποία αποτυπώνουν στοιχεία όπου αργότερα βάση αυτών δημιουργούνται οι γνωσιακές βάσεις (*Liar dataset*, *fake news dataset*).

Όσον αφορά τους αλγόριθμους η πλειονότητα των άρθρων εστιάζει στα πολύ καλά αποτελέσματα των παραδοσιακών αλγορίθμων όπως η μηχανή διανυσματικής υποστήριξης (*SVM*) αλλά και των τυχαίων δέντρων απόφασης (*Random Forest*) ως προς τον εντοπισμό των ψευδών ειδήσεων. Επιπλέον, δεδομένων των σύνθετων πλέον στοιχείων της πληροφορίας (εικόνα, κείμενο, βίντεο) η χρήση απλών αλγορίθμων νευρωνικών δικτύων όπως το *ANN* αλλά και πιο σύνθετων (βαθιάς μάθησης) όπως ο *GRU*, ο *RNN*, *CNN*, ο *MLP (Perceptron)*, και

ο LSTM αποδεικνύει ότι ο εντοπισμός της παραπληροφόρησης δεν εμφανίζεται μόνο σε κείμενα ή site αλλά και ότι παραδοσιακοί και κοστοβόροι αλγόριθμοι όπως ο SVM δεν αρκούν για την έγκαιρη πρόβλεψή της.

Ιδιαίτερο ενδιαφέρον ως προς τον εντοπισμό των ψευδών ειδήσεων παρουσιάζουν και οι *εξελικτικοί αλγόριθμοι* οι οποίοι έχουν υιοθετήσει τεχνικές από το ζωικό και το φυσικό βασίλειο όπως ο αλγόριθμος Grey Wolves αλλά και ο Whale, ο αλγόριθμος σάλπας όπως και ο XGBoost οι οποίοι παίζουν καθοριστικό ρόλο για τη βελτίωση της απόδοσης των παραδοσιακών αλγορίθμων αλλά και στον συνδυασμό τους με εμπροσθοβαρή δίκτυα για βελτίωση της πρόβλεψης.

Οι *αλγόριθμοι νευρωνικών δικτύων (CNN, RNN)* πλέον χρησιμοποιούνται ευρύτατα από την ερευνητική κοινότητα λόγω της αποτελεσματικότητάς τους, της ταχύτητας αλλά και της ευελίξιάς τους ως προς την προσαρμογή (πχ. προσθήκη επιπλέον επιπέδων για καλύτερο και έγκυρο αποτέλεσμα). Οι αλγόριθμοι δικτύων έχουν επηρεαστεί από τα νευρωνικά δίκτυα του ανθρώπινου εγκεφάλου (CNN) αλλά και από την επιστήμη της γνωστικής ψυχολογίας (LSTM). Από την παρούσα βιβλιογραφία τείνουν να χρησιμοποιούνται συχνότερα πλέον από τους ερευνητές για τον εντοπισμό των ψευδών ειδήσεων με το *νευρωνικό δίκτυο κάψουλας*, και τα *δίκτυα γράφων* να αποτελούν αξιόλογες και ενδιαφέρουσες προτάσεις κυρίως ως προς την εκτεταμένη και λεπτομερή ανάλυση των δεδομένων.

Αξιόλογες προσπάθειες έχουν γίνει και από Έλληνες ερευνητές για τον εντοπισμό των ψευδών ειδήσεων όπως του εικοσάχρονου Τζέκα ο οποίος δημιούργησε τον αλγόριθμο *fighthoax.com* για το ευρωκοινοβούλιο (2018) που μπορεί να ανιχνεύσει κάθε είδους πληροφορία και να επισημάνει αν είναι ψευδής ή αληθής (*FightHoax*, 2018) αλλά και του Σταύρου Λυγερού όπου εκπαίδευσε αλγορίθμους για τον εντοπισμό των ψευδών ειδήσεων για την πανδημία Covid19 (*Ο Έλληνας αστυνομικός που “εκπαίδευσε” αλγόριθμους για τα fake news της πανδημίας*, n.d.).

Μελλοντικές έρευνες για τον εντοπισμό των ψευδών ειδήσεων μπορούν να επεκταθούν και στα τrol, στα bot όπου και τα δύο συντελούν στη διάδοση του φαινομένου καθώς και σε μια ανερχόμενη νέα έννοια τα «*deepfakes*» τα οποία αφορούν ψεύτικα βίντεο ή εικόνες όπου έχει αντικατασταθεί το πρωτότυπο πρόσωπο με άλλο και τα οποία περιλαμβάνουν την παιδική πορνογραφία, το revenge porn, το bullying και την οικονομική απάτη .

Τέλος, κανείς αλγόριθμος δεν μπορεί από μόνος του να επιλύσει το φαινόμενο των ψευδών ειδήσεων αν δεν υπάρξει σωστή πληροφόρηση του κοινού μέσω πληροφοριακής παιδείας κυρίως από τους επιστήμονες της πληροφόρησης αλλά και της ίδιας της πολιτείας.

Η υιοθέτηση καλών πρακτικών όπως έχουν αποτυπωθεί από την IFLA δηλαδή ο έλεγχος της συγγραφής κειμένου, η κριτική σκέψη, οι υποστηρικτικές πηγές, ο έλεγχος της πηγής (URL, ημερομηνία δημοσίευσης) αλλά και τα ερωτήματα σε ειδικούς της πληροφόρησης αποτελούν πρώτα βήματα ενημέρωσης των πολιτών για τις ψευδείς ειδήσεις.

Βιβλιογραφικές Αναφορές

- ** Twitter Trails: Tool for monitoring the propagation of rumors ***. (n.d.). Retrieved September 10, 2022, from <http://twittertrails.com/>
- 1.1. *Linear Models*. (n.d.). Scikit-Learn. Retrieved October 23, 2022, from https://scikit-learn/stable/modules/linear_model.html
- 1.2. *Linear and Quadratic Discriminant Analysis*. (n.d.). Scikit-Learn. Retrieved September 13, 2022, from https://scikit-learn/stable/modules/lda_qda.html
- 1.4. *Support Vector Machines*. (n.d.). Scikit-Learn. Retrieved November 9, 2022, from <https://scikit-learn/stable/modules/svm.html>
- 1.6. *Nearest Neighbors*. (n.d.). Scikit-Learn. Retrieved November 9, 2022, from <https://scikit-learn/stable/modules/neighbors.html>
- 1.9. *Naive Bayes*. (n.d.-a). Scikit-Learn. Retrieved October 23, 2022, from https://scikit-learn/stable/modules/naive_bayes.html
- 1.9. *Naive Bayes*. (n.d.-b). Scikit-Learn. Retrieved October 23, 2022, from https://scikit-learn/stable/modules/naive_bayes.html
- 1.10. *Decision Trees*. (n.d.). Scikit-Learn. Retrieved November 11, 2022, from <https://scikit-learn/stable/modules/tree.html>
- 1.11. *Ensemble methods*. (n.d.-a). Scikit-Learn. Retrieved October 23, 2022, from <https://scikit-learn/stable/modules/ensemble.html>
- 1.11. *Ensemble methods*. (n.d.-b). Scikit-Learn. Retrieved October 22, 2022, from <https://scikit-learn/stable/modules/ensemble.html>
- 1.17. *Neural network models (supervised)*. (n.d.). Scikit-Learn. Retrieved October 27, 2022, from https://scikit-learn/stable/modules/neural_networks_supervised.html
- 2.3. *Clustering*. (n.d.). Scikit-Learn. Retrieved October 23, 2022, from <https://scikit-learn/stable/modules/clustering.html>

- 2.5. *Decomposing signals in components (matrix factorization problems)*. (n.d.). Scikit-Learn. Retrieved October 23, 2022, from <https://scikit-learn/stable/modules/decomposition.html>
- Abdel-Hamid, O., Abdel-Rahman, M., Jiang, H., & Penn, G. (2012, May). Applying Convolutional Neural Networks concepts to hybrid NN-HMM model for speech recognition. *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference*. <https://doi.org/10.1109/ICASSP.2012.6288864>
- Abdelminaam, D. S., Ismail, F. H., Taha, M., Taha, A., Houssein, E. H., & Nabil, A. (2021). CoAID-DEEP: An Optimized Intelligent Framework for Automated Detecting COVID-19 Misleading Information on Twitter. *IEEE Access*, 9, 27840–27867. <https://doi.org/10.1109/ACCESS.2021.3058066>
- Abonizio, H. Q., de Moraes, J. I., Tavares, G. M., & Barbon Junior, S. (2020). Language-Independent Fake News Detection: English, Portuguese, and Spanish Mutual Features. *Future Internet*, 12(5), 87. <https://doi.org/10.3390/fi12050087>
- aboSamoor / polyglot: Multilingual text (NLP) processing toolkit*. (2020). [Python]. <https://github.com/aboSamoor/polyglot>
- Abualigah, L. M., Khader, A. T., & Hanandeh, E. S. (2018). A combination of objective functions and hybrid Krill herd algorithm for text document clustering analysis. *Engineering Applications of Artificial Intelligence*, 73, 111–125. <https://doi.org/10.1016/j.engappai.2018.05.003>
- Ackland, R., & Gwynn, K. (2020). Truth and the Dynamics of News Diffusion on Twitter. In R. Greifeneder, M. Jaffe, E. Newman, & N. Schwarz, *The Psychology of Fake News* (1st ed., pp. 27–46). Routledge. <https://doi.org/10.4324/9780429295379-4>
- Agarap, A. F. (2019). *Deep Learning using Rectified Linear Units (ReLU)* (arXiv:1803.08375). arXiv. <http://arxiv.org/abs/1803.08375>

- Aggarwal, K., & Sadana, A. (2019). NSIT@NLP4IF-2019: Propaganda Detection from News Articles using Transfer Learning. *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, 143–147. <https://doi.org/10.18653/v1/D19-5021>
- Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake News Detection Using Machine Learning Ensemble Methods. *Complexity*, 2020, 1–11. <https://doi.org/10.1155/2020/8885861>
- Ahmed, H., Traore, I., & Saad, S. (2018). Detecting opinion spams and fake news using text classification. *Security and Privacy*, 1(1), e9. <https://doi.org/10.1002/spy2.9>
- Al-Ahmad, B., Al-Zoubi, A. M., Abu Khurma, R., & Aljarah, I. (2021). An Evolutionary Fake News Detection Method for COVID-19 Pandemic Information. *Symmetry*, 13(6), 1091. <https://doi.org/10.3390/sym13061091>
- Al-Daihani, S. M., & AlAwadhi, S. A. (2015). Exploring academic libraries' use of Twitter: A content analysis. *The Electronic Library*, 33(6), 1002–1015. <https://doi.org/10.1108/EL-05-2014-0084>
- Alhazbi, S. (2020). Behavior-Based Machine Learning Approaches to Identify State-Sponsored Trolls on Twitter. *IEEE Access*, 8, 195132–195141. <https://doi.org/10.1109/ACCESS.2020.3033666>
- Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Al-Rfou, R. (n.d.). *Rami Al-Rfou—Polyglot-NER*. Retrieved September 11, 2022, from <https://sites.google.com/site/rmyeid/projects/polylgot-ner>
- Al-Rfou, R., Kulkarni, V., Perozzi, B., & Skiena, S. (2014). *POLYGLOT-NER: Massive Multilingual Named Entity Recognition* (arXiv:1410.3791). arXiv. <http://arxiv.org/abs/1410.3791>

- Al-Rfou', R., Perozzi, B., & Skiena, S. (2013). Polyglot: Distributed Word Representations for Multilingual NLP. *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, 183–192.
<https://aclanthology.org/W13-3520>
- AlRubaian, M., Al-Qurishi, M., Al-Rakhami, M., Rahman, S. M. M., & Alamri, A. (2015). A Multistage Credibility Analysis Model for Microblogs. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, 1434–1440. <https://doi.org/10.1145/2808797.2810065>
- Altunbey Ozbay, F., & Alatas, B. (2019). A Novel Approach for Detection of Fake News on Social Media Using Metaheuristic Optimization Algorithms. *Elektronika Ir Elektrotechnika*, 25(4), 62–67. <https://doi.org/10.5755/j01.eie.25.4.23972>
- Amazon Mechanical Turk*. (n.d.). Retrieved September 10, 2022, from <https://www.mturk.com/>
- Amoruso, M., Anello, D., Auletta, V., Cerulli, R., Ferraioli, D., & Raiconi, A. (2020). Contrasting the Spread of Misinformation in Online Social Networks. *Journal of Artificial Intelligence Research*, 69, 847–879.
<https://doi.org/10.1613/jair.1.11509>
- An Introduction to XGBoost R package*. (n.d.). Retrieved September 13, 2022, from <https://xgboost.ai/rstats/2016/03/10/xgboost.html>
- Annasaheb, A. B., & Verma, V. K. (2016). Data Mining Classification Techniques: A Recent Survey. *International Journal of Emerging Technologies in Engineering Research*, 4(8), 51–54.
- Aphiwongsophon, S., & Chongstitvatana, P. (2018). Detecting Fake News with Machine Learning Method. *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 528–531.
<https://doi.org/10.1109/ECTICon.2018.8620051>

- Aronson, J. (1995). A Pragmatic View of Thematic Analysis. *The Qualitative Report*.
<https://doi.org/10.46743/2160-3715/1995.2069>
- Atari, M., Graham, J., & Deghani, M. (2020). Foundations of morality in Iran. *Evolution and Human Behavior*, 41(5), 367–384.
<https://doi.org/10.1016/j.evolhumbehav.2020.07.014>
- Attride-Stirling, J. (2001). Thematic networks: An analytic tool for qualitative research. *Qualitative Research*, 1(3), 385–405.
<https://doi.org/10.1177/146879410100100307>
- Baeth, M. J., & Aktas, M. S. (2017). A Large Scale Synthetic Social Provenance Database. *DBKDA 2017: The Ninth International Conference on Advances in Databases, Knowledge, and Data Applications*, 16–22.
- Baeth, M. J., & Aktas, M. S. (2019). Detecting misinformation in social networks using provenance data. *Concurrency and Computation: Practice and Experience*, 31(3), e4793. <https://doi.org/10.1002/cpe.4793>
- Bag-of-words model. (2022). In *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Bag-of-words_model&oldid=1119717306
- Bahdanau, D., Cho, K., & Bengio, Y. (2016). *Neural Machine Translation by Jointly Learning to Align and Translate* (arXiv:1409.0473). arXiv.
<http://arxiv.org/abs/1409.0473>
- Bailey, N. T. J. (1975). *The mathematical theory of infectious diseases and its applications* (2nd ed). Griffin. <https://www.cabdirect.org/cabdirect/abstract/19762902036>
- Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., & Nakov, P. (2018). Predicting Factuality of Reporting and Bias of News Media Sources. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 3528–3539.
<https://doi.org/10.18653/v1/D18-1389>

- Bangani, S. (2021). The fake news wave: Academic libraries' battle against misinformation during COVID-19. *The Journal of Academic Librarianship*, 47(5), 102390. <https://doi.org/10.1016/j.acalib.2021.102390>
- Barrón-Cedeño, A., Jaradat, I., Da San Martino, G., & Nakov, P. (2019). Propopy: Organizing the news based on their propagandistic content. *Information Processing & Management*, 56(5), 1849–1864. <https://doi.org/10.1016/j.ipm.2019.03.005>
- Beautiful Soup Documentation—Beautiful Soup 4.4.0 documentation*. (n.d.). Retrieved September 10, 2022, from <https://beautiful-soup-4.readthedocs.io/en/latest/>
- Boesch, G. (2022, June 29). *Deep Residual Networks (ResNet, ResNet50)—2022 Guide*. Viso.Ai. <https://viso.ai/deep-learning/resnet-residual-neural-network/>
- Boididou, C., Papadopoulos, S., Zampoglou, M., Apostolidis, L., Papadopoulou, O., & Kompatsiaris, Y. (2018). Detection and visualization of misleading content on Twitter. *International Journal of Multimedia Information Retrieval*, 7(1), 71–86. <https://doi.org/10.1007/s13735-017-0143-x>
- Bonet-Jover, A., Piad-Morffis, A., Saquete, E., Martínez-Barco, P., & Ángel García-Cumbreras, M. (2021). Exploiting discourse structure of traditional digital media to enhance automatic fake news detection. *Expert Systems with Applications*, 169, 114340. <https://doi.org/10.1016/j.eswa.2020.114340>
- Borges, L., Martins, B., & Calado, P. (2019). Combining Similarity Features and Deep Representation Learning for Stance Detection in the Context of Checking Fake News. *Journal of Data and Information Quality*, 11(3), 1–26. <https://doi.org/10.1145/3287763>
- Botometer by OSoMe*. (n.d.). Retrieved September 11, 2022, from <https://botometer.iuni.iu.edu>
- Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, 10(1), 7. <https://doi.org/10.1038/s41467-018-07761-2>

- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage Publications.
- Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *LIWC22-PSYCHOMETRICS MANUAL*. The University of Texas at Austin.
<https://www.liwc.app/help/psychometrics-manuals>
- Braşoveanu, A. M. P., & Andonie, R. (2021). Integrating Machine Learning Techniques in Semantic Fake News Detection. *Neural Processing Letters*, 53(5), 3055–3072.
<https://doi.org/10.1007/s11063-020-10365-x>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
<https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. (pp. 57–71). American Psychological Association.
<https://doi.org/10.1037/13620-004>
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140.
<https://doi.org/10.1007/BF00058655>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Brownlee, J. (2016, April 8). *Classification and regression trees for machine learning*. Machine Learning Mastery. <https://machinelearningmastery.com/classification-and-regression-trees-for-machine-learning/>
- Brownlee, J. (2017, October 9). *A Gentle Introduction to the Bag-of-Words Model*. <https://machinelearningmastery.com/gentle-introduction-bag-words-model/>
- Burgoon, J. K., Blair, J. P., Qin, T., & Nunamaker, J. F. (2003). Detecting Deception through Linguistic Analysis. In H. Chen, R. Miranda, D. D. Zeng, C. Demchak, J. Schroeder,

- & T. Madhusudan (Eds.), *Intelligence and Security Informatics* (Vol. 2665, pp. 91–101). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-44853-5_7
- Burns, T., & Berelson, B. (1953). Content Analysis in Communication Research. *The British Journal of Sociology*, 4(1), 88. <https://doi.org/10.2307/587175>
- Cambria, E., Poria, S., Gelbukh, A., & Thelwall, M. (2017). Sentiment Analysis Is a Big Suitcase. *IEEE Intelligent Systems*, 32(6), 74–80. <https://doi.org/10.1109/MIS.2017.4531228>
- Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. *Proceedings of the 20th International Conference on World Wide Web - WWW '11*, 675. <https://doi.org/10.1145/1963405.1963500>
- Category:All Wikipedia neutral point of view disputes. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Category:All_Wikipedia_neutral_point_of_view_disputes&oldid=1066929354
- Caucaso, O. B. e. (n.d.). *Troll Factories: The Internet Research Agency and State-Sponsored Agenda Building*. Media Freedom Resource Centre OBCT. Retrieved November 13, 2022, from <https://www.rcmediafreedom.eu/Publications/Academic-sources/Troll-Factories-The-Internet-Research-Agency-and-State-Sponsored-Agenda-Building>
- Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Strope, B., & Kurzweil, R. (2018). Universal Sentence Encoder for English. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 169–174. <https://doi.org/10.18653/v1/D18-2029>
- ChandraPandey, A., SinghRajpoot, D., & Saraswat, M. (2018). Data Clustering Based on Data Transformation and Hybrid Step Size-Based Cuckoo Search. *2018 Eleventh International Conference on Contemporary Computing (IC3)*, 1–6. <https://doi.org/10.1109/IC3.2018.8530571>

- Charlet, D., & Damnati, G. (2017). SimBow at SemEval-2017 Task 3: Soft-Cosine Semantic Similarity between Questions for Community Question Answering. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 315–319. <https://doi.org/10.18653/v1/S17-2051>
- Chen, H., Sun, M., Tu, C., Lin, Y., & Liu, Z. (2016). Neural Sentiment Classification with User and Product Attention. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1650–1659. <https://doi.org/10.18653/v1/D16-1171>
- Chen, L., Liang, J., Xie, C., & Xiao, Y. (2018). Short Text Entity Linking with Fine-grained Topics. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 457–466. <https://doi.org/10.1145/3269206.3271809>
- Chen, X., Fang, H., Lin, T.-Y., Vedantam, R., Gupta, S., Dollar, P., & Zitnick, C. L. (2015). *Microsoft COCO Captions: Data Collection and Evaluation Server* (arXiv:1504.00325). arXiv. <http://arxiv.org/abs/1504.00325>
- Chen, Y.-C., Liu, Z.-Y., & Kao, H.-Y. (2017). IKM at SemEval-2017 Task 8: Convolutional Neural Networks for stance detection and rumor verification. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 465–469. <https://doi.org/10.18653/v1/S17-2081>
- Chidiac, N.-M., Damien, P., & Yaacoub, C. (2016). A robust algorithm for text extraction from images. *2016 39th International Conference on Telecommunications and Signal Processing (TSP)*, 493–497. <https://doi.org/10.1109/TSP.2016.7760928>
- Choi, S., & Kim, J. (n.d.). *Capsule Networks for Text Classification*. <https://github.com/TeamLab/text-capsule-network>
- Choraś, M., Pawlicki, M., Kozik, R., Demestichas, K., Kosmides, P., & Gupta, M. (2019). SocialTruth Project Approach to Online Disinformation (Fake News) Detection and Mitigation. *Proceedings of the 14th International Conference on Availability, Reliability and Security*, 1–10. <https://doi.org/10.1145/3339252.3341497>

- Chu, H. (2015). Research methods in library and information science: A content analysis. *Library & Information Science Research*, 37(1), 36–41.
<https://doi.org/10.1016/j.lisr.2014.09.003>
- Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F., & Flammini, A. (2015). Computational Fact Checking from Knowledge Networks. *PLOS ONE*, 10(6), e0128193. <https://doi.org/10.1371/journal.pone.0128193>
- ClaimBuster: Automated Live Fact-checking*. (n.d.). Retrieved September 10, 2022, from <https://idir.uta.edu/claimbuster/>
- Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology*, 12(3), 297–298. <https://doi.org/10.1080/17439760.2016.1262613>
- Cnn-text-classification-tf*. (n.d.). <https://github.com/dennybritz/cnn-text-classification-tf>
- Cohen, R., Moffatt, K., Ghenai, A., Yang, A., Corwin, M., Lin, G., Zhao, R., Ji, Y., Parmentier, A., P'ng, J., Tan, W., & Gray, L. (2020). Addressing Misinformation in Online Social Networks: Diverse Platforms and the Potential of Multiagent Trust Modeling. *Information*, 11(11), 539. <https://doi.org/10.3390/info11110539>
- Conroy, N. K., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1), 1–4. <https://doi.org/10.1002/pra2.2015.145052010082>
- Coronavirus5G*. (n.d.). Twitter. Retrieved November 13, 2022, from https://twitter.com/search?src=typed_query&q=coronavirus5G
- CountVectorizer*. (n.d.). Scikit-Learn. Retrieved November 7, 2022, from https://scikit-learn/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html
- Cremsini, A., Aguilar, D., & Finlayson, M. A. (2019). A Challenging Dataset for Bias Detection: The Case of the Crisis in the Ukraine. In R. Thomson, H. Bisgin, C. Dancy, & A. Hyder (Eds.), *Social, Cultural, and Behavioral Modeling* (Vol. 11549,

pp. 173–183). Springer International Publishing. https://doi.org/10.1007/978-3-030-21741-9_18

CrowdTangle. (n.d.). *CrowdTangle | Content Discovery and Social Monitoring Made Easy*.

CrowdTangle. Retrieved September 17, 2022, from

<https://www.crowdtangle.com//>

CrowdTangle Codebook. (n.d.). Retrieved September 17, 2022, from

<https://socialscience.one/crowdtangle-codebook>

CSI-Code. (n.d.). <https://github.com/sungyongs/CSI-Code>

csurfer. (n.d.). *rake-nltk: RAKE short for Rapid Automatic Keyword Extraction algorithm, is a domain independent keyword extraction algorithm which tries to determine key phrases in a body of text by analyzing the frequency of word appearance and its co-occurrence with other words in the text*. [Python; POSIX]. Retrieved September 10, 2022, from <https://csurfer.github.io/rake-nltk>

Daniilides, T. (2019, July 12). *Ellinika Hoaxes-O “σύμβουλος” του Τούρκου ΥΕΘΑ δήλωσε ότι «Συνεχίζουμε με αγορά (βλημάτων) Bastion & Iskander»*; <https://www.ellinikahoaxes.gr/2019/07/12/turkish-advisors-supposed-comments/>

Dauphin, Y. N., Fan, A., Auli, M., & Grangier, D. (2017). *Language Modeling with Gated Convolutional Networks* (arXiv:1612.08083). arXiv. <http://arxiv.org/abs/1612.08083>

Davis, R. A., & Proctor, C. (2017). *Fake News, Real Consequences: Recruiting Neural Networks for the Fight Against Fake News* (pp. 1–8). Stanford University. <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2761239>

DBpedia Spotlight—Shedding light on the web of documents. (n.d.). Retrieved September 10, 2022, from <https://www.dbpedia-spotlight.org/>

- de Beer, D., & Matthee, M. (2021). Approaches to Identify Fake News: A Systematic Literature Review. In T. Antipova (Ed.), *Integrated Science in Digital Age 2020* (Vol. 136, pp. 13–22). Springer International Publishing.
https://doi.org/10.1007/978-3-030-49264-9_2
- De Paor, S., & Heravi, B. (2020). Information literacy and fake news: How the field of librarianship can help combat the epidemic of fake news. *The Journal of Academic Librarianship*, 46(5), 102218.
<https://doi.org/10.1016/j.acalib.2020.102218>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. <http://arxiv.org/abs/1810.04805>
- Dhar, J., Jain, A., & K. Gupta, V. (2016). A mathematical model of news propagation on online social network and a control strategy for rumor spreading. *Social Network Analysis and Mining*, 6(1), 57. <https://doi.org/10.1007/s13278-016-0366-5>
- Digital Wildfires in a Hyperconnected World*. (n.d.). World Economic Forum.
https://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/?doing_wp_cron=1637438286.9953560829162597656250
- Discussion document: Predatory Publishing*. (2019). Committee on Publication Ethics.
<https://doi.org/10.24318/cope.2019.3.6>
- Disney, A. (2020, January 2). *Social network analysis: Understanding centrality measures*. Cambridge Intelligence. <https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/>
- Dobrovolskyi, H., & Keberle, N. (2019). On Convergence of Controlled Snowball Sampling for Scientific Abstracts Collection. In V. Ermolayev, M. C. Suárez-Figueroa, V. Yakovyna, H. C. Mayr, M. Nikitchenko, & A. Spivakovsky (Eds.), *Information and Communication Technologies in Education, Research, and Industrial Applications*

- (Vol. 1007, pp. 18–42). Springer International Publishing.
https://doi.org/10.1007/978-3-030-13929-2_2
- Dobrovolskyi, H., Keberle, N., & Todoriko, O. (2017). Probabilistic Topic Modelling for Controlled Snowball Sampling in Citation Network Collection. In P. Rózewski & C. Lange (Eds.), *Knowledge Engineering and Semantic Web* (Vol. 786, pp. 85–100). Springer International Publishing. https://doi.org/10.1007/978-3-319-69548-8_7
- Dong, S., Deng, Y.-B., & Huang, Y.-C. (2017). SEIR Model of Rumor Spreading in Online Social Network with Varying Total Population Size. *Communications in Theoretical Physics*, 68(4), 545. <https://doi.org/10.1088/0253-6102/68/4/545>
- dos Santos, C., & Gatti, M. (2014). Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 69–78.
<https://aclanthology.org/C14-1008>
- Dos Santos, E. M., Sabourin, R., & Maupin, P. (2009). Overfitting cautious selection of classifier ensembles with genetic algorithms. *Information Fusion*, 10(2), 150–162. <https://doi.org/10.1016/j.inffus.2008.11.003>
- Dovring, K. (1954). Quantitative Semantics in 18th Century Sweden. *Public Opinion Quarterly*, 18(4), 389. <https://doi.org/10.1086/266531>
- Drobnic-Holan, A. (2018, February 18). *PolitiFact - The Principles of the Truth-O-Meter: PolitiFact's methodology for independent fact-checking*. Politifact.
<https://www.politifact.com/article/2018/feb/12/principles-truth-o-meter-politifacts-methodology-i/>
- Du, J., Preston, S., Sun, H., Shegog, R., Cunningham, R., Boom, J., Savas, L., Amith, M., & Tao, C. (2021). Using Machine Learning–Based Approaches for the Detection and Classification of Human Papillomavirus Vaccine Misinformation: Infodemiology

- Study of Reddit Discussions. *Journal of Medical Internet Research*, 23(8), e26478.
<https://doi.org/10.2196/26478>
- Ebrahimi, M., Yazdavar, A. H., & Sheth, A. (2017). Challenges of Sentiment Analysis for Dynamic Events. *IEEE Intelligent Systems*, 32(5), 70–75.
<https://doi.org/10.1109/MIS.2017.3711649>
- Elhadad, M. K., Li, K. F., & Gebali, F. (2020). Detecting Misleading Information on COVID-19. *IEEE Access*, 8, 165201–165215.
<https://doi.org/10.1109/ACCESS.2020.3022867>
- Ellinika Hoaxes*. (n.d.). EH-Hellinika Hoaxes. <https://www.ellinikahoaxes.gr/>
Ellinika Hoaxes-Μεθοδολογία ελέγχου και επεξήγηση χαρακτηρισμών. (n.d.).
<https://www.ellinikahoaxes.gr/transparency-and-methodology/>
- EMBL-EBI. (n.d.). *Betweenness centrality | Network analysis of protein interaction data*. Retrieved September 17, 2022, from
<https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/building-and-analysing-ppins/topological-ppin-analysis/betweenness-centrality/>
- Emergent*. (n.d.). Retrieved September 10, 2022, from <http://www.emergent.info/>
- Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), 189–200.
<https://doi.org/10.1177/0165551512437638>
- Examples of wordy sentences—And how to correct wordiness*. (2019, October 24). Readable. <https://readable.com/blog/examples-of-wordy-sentences-and-how-to-correct-wordiness/>
- EXtreme Gradient Boosting*. (2022). [C++]. Distributed (Deep) Machine Learning Community. <https://github.com/dmlc/xgboost> (Original work published 2014)
- Fact Check Tools*. (n.d.). Retrieved September 10, 2022, from
<https://toolbox.google.com/factcheck/explorer>

Fact Extraction and VERification. (n.d.-a). Retrieved September 10, 2022, from <https://fever.ai/>

Fact Extraction and VERification. (n.d.-b). Retrieved September 10, 2022, from <https://fever.ai/dataset/fever.html>

Factcheck. (n.d.). Retrieved September 9, 2022, from <https://factcheck.snu.ac.kr/>

Faix, A., & Fyn, A. (2020). Framing Fake News: Misinformation and the ACRL Framework. *Portal: Libraries and the Academy*, 20(3), 495–508. <https://doi.org/10.1353/pla.2020.0027>

FakeNewsChallenge / fnc-1-baseline. (2017a). [Python]. <https://github.com/FakeNewsChallenge/fnc-1-baseline>

FakeNewsChallenge / fnc-1-baseline. (2017b, March 2). https://github.com/FakeNewsChallenge/fnc-1-baseline/blob/master/feature_engineering.py

FastText. (n.d.). Retrieved September 12, 2022, from <https://fasttext.cc/index.html>

Fentaw, H. W., & Kim, T.-H. (2019). Design and Investigation of Capsule Networks for Sentence Classification. *Applied Sciences*, 9(11), 2200. <https://doi.org/10.3390/app9112200>

Féraud, R., & Clérot, F. (2002). A methodology to explain neural network classification. *Neural Networks*, 15(2), 237–246. [https://doi.org/10.1016/S0893-6080\(01\)00127-7](https://doi.org/10.1016/S0893-6080(01)00127-7)

FightHoax: Ένας «ελληνικός» ανιχνευτής ψευδών ειδήσεων. (2018, February 10). HuffPost Greece. https://www.huffingtonpost.gr/entry/fighthoax-enas-ellenikos-anichneetes-pseedon-eideseon_gr_5a7ea36ee4b0c6726e138f31

Fiskkit- A better way to discuss news. (n.d.). Fiskkit-Discuss News That Matters and Find out What's True. Retrieved September 10, 2022, from https://fiskkit.com/_OG_URL_

- Franz, D., Marsh, H. E., Chen, J. I., & Teo, A. R. (2019). Using Facebook for Qualitative Research: A Brief Primer. *Journal of Medical Internet Research*, 21(8), e13544. <https://doi.org/10.2196/13544>
- Frenkel, S., Decker, B., & Alba, D. (2020, May 20). *How the “plandemic” movie and its falsehoods spread widely online*. The New York Times. <https://www.nytimes.com/2020/05/20/technology/plandemic-movie-youtube-facebook-coronavirus.html>
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367–378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Full Fact—Full Fact is the UK’s independent fact checking organisation*. (n.d.). Full Fact. Retrieved September 9, 2022, from <https://fullfact.org>
- Galán-García, P., Puerta, J. G. D. L., Gómez, C. L., Santos, I., & Bringas, P. G. (2015). Supervised machine learning for the detection of troll profiles in twitter social network: Application to a real case of cyberbullying. *Logic Journal of IGPL*, jzv048. <https://doi.org/10.1093/jigpal/jzv048>
- Gandhi, R. (2018a, May 17). *Naive Bayes Classifier*. Medium. <https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c>
- Gandhi, R. (2018b, July 5). *Support Vector Machine—Introduction to Machine Learning Algorithms*. Medium. <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- Generating Word Cloud in Python. (2018, May 11). *GeeksforGeeks*. <https://www.geeksforgeeks.org/generating-word-cloud-python/>
- Gensim: Topic modelling for humans*. (n.d.). Retrieved November 7, 2022, from https://radimrehurek.com/gensim/auto_examples/tutorials/run_scm.html
- Georgacopoulos, C., & Mores, G. (2020, July). *How Fake News Affected the 2016 Presidential Election*. LSU.EDU. <https://faculty.lsu.edu/fakenews/elections/sixteen.php>

- Gephi org. (n.d.). *Features*. Gephi: Makes Graph Handy. Retrieved November 14, 2022, from <https://gephi.org/features/>
- Gers, F. A. (1999). Learning to forget: Continual prediction with LSTM. *9th International Conference on Artificial Neural Networks: ICANN '99, 1999*, 850–855. <https://doi.org/10.1049/cp:19991218>
- Ghenai, A. (2019). *Health Misinformation in Search and Social Media*. <https://uwspace.uwaterloo.ca/handle/10012/15268>
- Ghenai, A., & Mejova, Y. (2017). Catching Zika Fever: Application of Crowdsourcing and Machine Learning for Tracking Health Misinformation on Twitter. *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, 518–518. <https://doi.org/10.1109/ICHI.2017.58>
- Ghenai, A., & Mejova, Y. (2018). Fake Cures: User-centric Modeling of Health Misinformation in Social Media. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1–20. <https://doi.org/10.1145/3274327>
- Girgis, S., Amer, E., & Gadallah, M. (2018). Deep Learning Algorithms for Detecting Fake News in Online Text. *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, 93–97. <https://doi.org/10.1109/ICCES.2018.8639198>
- GitHub - LuisPB7/fnc-msc: Deep Learning model to tackle the Fake News Challenge. (n.d.). GitHub. Retrieved November 11, 2022, from <https://github.com/LuisPB7/fnc-msc>
- Goldani, M. H., Momtazi, S., & Safabakhsh, R. (2021). Detecting fake news with capsule neural networks. *Applied Soft Computing*, 101, 106991. <https://doi.org/10.1016/j.asoc.2020.106991>
- Gôlo, M., Caravanti, M., Rossi, R., Rezende, S., Nogueira, B., & Marcacini, R. (2021). Learning Textual Representations from Multiple Modalities to Detect Fake News

- Through One-Class Learning. *Proceedings of the Brazilian Symposium on Multimedia and the Web*, 197–204. <https://doi.org/10.1145/3470482.3479634>
- Goodfellow, I., Bengio, Y., & Courville, A. (2017). *Deep Learning*. Deep Learning. <https://www.deeplearningbook.org/>
- Google Code Archive—Long-term storage for Google Code Project Hosting. (n.d.). Retrieved November 8, 2022, from <https://code.google.com/archive/p/word2vec/>
- Google-research / bert. (2020). <https://github.com/google-research/bert>
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations Theory. In *Advances in Experimental Social Psychology* (Vol. 47, pp. 55–130). Elsevier. <https://doi.org/10.1016/B978-0-12-407236-7.00002-4>
- graph-tool: Efficient network analysis with python. (n.d.). Retrieved November 13, 2022, from <https://graph-tool.skewed.de/>
- Gravanis, G., Vakali, A., Diamantaras, K., & Karadais, P. (2019). Behind the cues: A benchmarking study for fake news detection. *Expert Systems with Applications*, 128, 201–213. <https://doi.org/10.1016/j.eswa.2019.03.036>
- Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5–6), 602–610. <https://doi.org/10.1016/j.neunet.2005.06.042>
- Greenwell, B. B. & B. (n.d.). *Chapter 12 Gradient Boosting | Hands-On Machine Learning with R*. Retrieved September 13, 2022, from <https://bradleyboehmke.github.io/HOML/gbm.html>
- Greifeneder, R., Jaffe, M., Newman, E., & Schwarz, N. (2020). *The Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation* (1st ed.). Routledge. <https://doi.org/10.4324/9780429295379>
- Gulli, A., & Pal, S. (2017). *Deep learning with Keras: Implement neural networks with Keras on Theano and TensorFlow*. Packt Publishing.

Gunning Fog Index. (n.d.). Retrieved September 11, 2022, from <http://gunning-fog-index.com/>

Gunning fog index. (2022). In *Wikipedia*.

https://en.wikipedia.org/w/index.php?title=Gunning_fog_index&oldid=1106198590

Gunther, R., Beck, P. A., & Nisbet, E. C. (2018). *Fake News May Have Contributed to Trump's 2016 Victory*. from

<https://www.documentcloud.org/documents/4429952-Fake-News-May-Have-Contributed-to-Trump-s-2016.html>

Guo, H., Cao, J., Zhang, Y., Guo, J., & Li, J. (2018). Rumor Detection with Hierarchical Social Attention Network. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 943–951.

<https://doi.org/10.1145/3269206.3271709>

Guo, L., & Vargo, C. (2020). “Fake News” and Emerging Online Media Ecosystem: An Integrated Intermedia Agenda-Setting Analysis of the 2016 U.S. Presidential Election. *Communication Research*, 47(2), 178–200.

<https://doi.org/10.1177/0093650218777177>

Habermas, J. (2020). *Teoría de la acción comunicativa: Complementos y estudios previos* (9a ed). Cátedra.

Haidt, J., & Graham, J. (2007). When Morality Opposes Justice: Conservatives Have Moral Intuitions that Liberals may not Recognize. *Social Justice Research*, 20(1), 98–116. <https://doi.org/10.1007/s11211-007-0034-z>

Han, Q., Miao, F., & Fan, W. (2017). Rumor spreading and monitoring deployment in online social networks. *2017 IEEE 17th International Conference on Communication Technology (ICCT)*, 1347–1351.

<https://doi.org/10.1109/ICCT.2017.8359853>

- Handcock, M. S., & Gile, K. J. (2011). Comment: On the Concept of Snowball Sampling. *Sociological Methodology*, 41(1), 367–371. <https://doi.org/10.1111/j.1467-9531.2011.01243.x>
- Hangloo, S., & Arora, B. (2021). *Fake News Detection Tools and Methods—A Review* (arXiv:2112.11185). arXiv. <http://arxiv.org/abs/2112.11185>
- Hastie, T., Rosset, S., Zhu, J., & Zou, H. (2009). Multi-class AdaBoost. *Statistics and Its Interface*, 2(3), 349–360. <https://doi.org/201308302105>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition* (arXiv:1512.03385). arXiv. <http://arxiv.org/abs/1512.03385>
- Heckathorn, D. D. (2011). Comment: Snowball versus Respondent-Driven Sampling. *Sociological Methodology*, 41(1), 355–366. <https://doi.org/10.1111/j.1467-9531.2011.01244.x>
- Heckathorn, D. D., & Cameron, C. J. (2017). Network Sampling: From Snowball and Multiplicity to Respondent-Driven Sampling. *Annual Review of Sociology*, 43(1), 101–119. <https://doi.org/10.1146/annurev-soc-060116-053556>
- Hill, H. (2021). *A thematic analysis of library association policies on services to persons with disabilities*. 77(6), 1379–1392. <https://doi.org/10.1108/jd-10-2020-0183>
- Hoaxy: How claims spread online*. (n.d.). Retrieved September 10, 2022, from <http://hoaxy.iuni.iu.edu/>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hohyon Ryu, W. (2014). *Open-korean-text*. <https://github.com/open-korean-text/open-korean-text>
- Home*. (n.d.). Plandemic. Retrieved September 19, 2022, from <https://plandemicseries.com/>
- Home | Yago Project*. (n.d.). Retrieved November 12, 2022, from <https://yago-knowledge.org/>

- Honoré, A. (1979). Some Simple Measures of Richness of Vocabulary. *Association for Literary and Linguistic Computing Bulletin*, 7(2), 172–177.
- Horev, R. (2018, November 17). *BERT Explained: State of the art language model for NLP*. Medium. <https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270>
- Horne, B. (2020). *Benjamin Horne/ The NELA Toolkit* (3.0.1.) [Postgre SQL, Python]. <https://github.com/BenjaminDHorne/The-NELA-Toolkit>
- Horne, B. D., & Adali, S. (2017). *This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News* (arXiv:1703.09398). arXiv. <http://arxiv.org/abs/1703.09398>
- Horne, B. D., Dron, W., Khedr, S., & Adali, S. (2018). Assessing the News Landscape: A Multi-Module Toolkit for Evaluating the Credibility of News. *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*, 235–238. <https://doi.org/10.1145/3184558.3186987>
- Horne, B. D., Nørregaard, J., & Adali, S. (2020). Robust Fake News Detection Over Time and Attack. *ACM Transactions on Intelligent Systems and Technology*, 11(1), 1–23. <https://doi.org/10.1145/3363818>
- Horne, B., Khedr, S., & Adali, S. (2018). Sampling the News Producers: A Large News and Feature Data Set for the Study of the Complex Media Landscape. *Proceedings of the International AAAI Conference on Web and Social Media*, 12(1). <https://ojs.aaai.org/index.php/ICWSM/article/view/14982>
- How do you calculate overperforming scores?* (n.d.). Retrieved November 13, 2022, from <http://help.crowdtangle.com/en/articles/2013937-how-do-you-calculate-overperforming-scores>
- Huang, Y.-F., & Chen, P.-H. (2020). Fake news detection using an ensemble learning model based on Self-Adaptive Harmony Search algorithms. *Expert Systems with Applications*, 159, 113584. <https://doi.org/10.1016/j.eswa.2020.113584>

- Hunt, K., Agarwal, P., & Zhuang, J. (2022). Monitoring Misinformation on Twitter During Crisis Events: A Machine Learning Approach. *Risk Analysis*, 42(8), 1728–1748.
<https://doi.org/10.1111/risa.13634>
- Hutto, C. J. (2014). *Cjhutto / vaderSentiment* [Python].
<https://github.com/cjhutto/vaderSentiment#features-and-updates>
- Hyland, K. (2015). Metadiscourse. In K. Tracy, T. Sandel, & C. Ilie (Eds.), *The International Encyclopedia of Language and Social Interaction* (1st ed., pp. 1–11). Wiley.
<https://doi.org/10.1002/9781118611463.wbielsi003>
- Ibrishimova, M. D., & Li, K. F. (2020). A Machine Learning Approach to Fake News Detection Using Knowledge Verification and Natural Language Processing. In L. Barolli, H. Nishino, & H. Miwa (Eds.), *Advances in Intelligent Networking and Collaborative Systems* (Vol. 1035, pp. 223–234). Springer International Publishing. https://doi.org/10.1007/978-3-030-29035-1_22
- Indu, V., & Thampi, S. M. (2019). A nature—Inspired approach based on Forest Fire model for modeling rumor propagation in social networks. *Journal of Network and Computer Applications*, 125, 28–41.
<https://doi.org/10.1016/j.jnca.2018.10.003>
- INFORMS. (n.d.). *Metaheuristics in Optimization: Algorithmic Perspective*. INFORMS. Retrieved September 13, 2022, from <https://www.informs.org/Publications/OR-MS-Tomorrow/Metaheuristics-in-Optimization-Algorithmic-Perspective>
- Institutions (IFLA), I. F. of L. A. and. (2017). *How To Spot Fake News*.
<https://repository.ifla.org/handle/123456789/167>
- International Fact-Checking Network. (n.d.). *Poynter*. Retrieved September 13, 2022, from <https://www.poynter.org/ifcn/>
- InVID Project—Video Verification*. (n.d.). InVID Project. Retrieved September 16, 2022, from <https://www.invid-project.eu/>

- ipstack—Free IP Geolocation API*. (n.d.). Retrieved November 8, 2022, from <https://ipstack.com/>
- Irie, K., Tüske, Z., Alkhouli, T., Schlüter, R., & Ney, H. (2016). LSTM, GRU, Highway and a Bit of Attention: An Empirical Overview for Language Modeling in Speech Recognition. *Interspeech 2016*, 3519–3523. <https://doi.org/10.21437/Interspeech.2016-491>
- Isaakidou, M., Zoulias, E., & Diomidous, M. (2021). Machine Learning to Identify Fake News for COVID-19. In J. Mantas, L. Stoicu-Tivadar, C. Chronaki, A. Hasman, P. Weber, P. Gallos, M. Crişan-Vida, E. Zoulias, & O. S. Chirila (Eds.), *Studies in Health Technology and Informatics*. IOS Press. <https://doi.org/10.3233/SHTI210130>
- Ishida, Y., & Kuraya, S. (2018). Fake News and its Credibility Evaluation by Dynamic Relational Networks: A Bottom up Approach. *Procedia Computer Science*, 126, 2228–2237. <https://doi.org/10.1016/j.procs.2018.07.226>
- Järvelin, K., & Vakkari, P. (2021). LIS research across 50 years: Content analysis of journal articles. *Journal of Documentation*, 78(7), 65–88. <https://doi.org/10.1108/JD-03-2021-0062>
- Jin, Z., Cao, J., Guo, H., Zhang, Y., & Luo, J. (2017). Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs. *Proceedings of the 25th ACM International Conference on Multimedia*, 795–816. <https://doi.org/10.1145/3123266.3123454>
- Jin, Z., Cao, J., Zhang, Y., Zhou, J., & Tian, Q. (2017). Novel Visual and Statistical Image Features for Microblogs News Verification. *IEEE Transactions on Multimedia*, 19(3), 598–608. <https://doi.org/10.1109/TMM.2016.2617078>
- Jiyfeng / DPLP: A RST Parser with a trained model*. (n.d.). <https://github.com/jiyfeng/DPLP>

- Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology, 33*(2), 125–143. <https://doi.org/10.1177/0261927X13502654>
- Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet – A deep convolutional neural network for fake news detection. *Cognitive Systems Research, 61*, 32–44. <https://doi.org/10.1016/j.cogsys.2019.12.005>
- Kar, D. (2020). Spotting misinformation to limit the impact of disruption on society by using machine learning. *2020 IEEE Applied Signal Processing Conference (ASPCON)*, 26–30. <https://doi.org/10.1109/ASPCON49795.2020.9276723>
- Kasseropoulos, D. P., & Tjortjis, C. (2021). An Approach Utilizing Linguistic Features for Fake News Detection. In I. Maglogiannis, J. Macintyre, & L. Iliadis (Eds.), *Artificial Intelligence Applications and Innovations* (Vol. 627, pp. 646–658). Springer International Publishing. https://doi.org/10.1007/978-3-030-79150-6_51
- Kausar, S., Tahir, B., & Mehmood, M. A. (2020). ProSOUL: A Framework to Identify Propaganda From Online Urdu Content. *IEEE Access, 8*, 186039–186054. <https://doi.org/10.1109/ACCESS.2020.3028131>
- Keras: The Python deep learning API*. (n.d.). Retrieved September 14, 2022, from <https://keras.io/>
- Khamday. (2021). Detecting Textual Propaganda Using Machine Learning Techniques. *Baghdad Science Journal*. <https://doi.org/10.21123/bsj.2021.18.1.0199>
- Khan, A., Brohman, K., & Addas, S. (2022). The anatomy of ‘fake news’: Studying false messages as digital objects. *Journal of Information Technology, 37*(2), 122–143. <https://doi.org/10.1177/02683962211037693>
- Khanday, A. M. U., Rabani, S. T., & Khan, Q. R. (2021). Detecting Textual Propaganda Using Machine Learning Techniques. *Baghdad Science Journal, 18*(1), 199–209. <https://doi.org/10.21123/bsj.2021.18.1.0199>

- Khanna, N. (2021, August 18). J48 Classification (C4.5 Algorithm) in a Nutshell. *Medium*.
<https://medium.com/@nilimakhanna1/j48-classification-c4-5-algorithm-in-a-nutshell-24c50d20658e>
- Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019). MVAE: Multimodal Variational Autoencoder for Fake News Detection. *The World Wide Web Conference*, 2915–2921. <https://doi.org/10.1145/3308558.3313552>
- Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical Teacher*, 42(8), 846–854.
<https://doi.org/10.1080/0142159X.2020.1755030>
- Kim, A. (2019, July 26). What is the application of HITS algorithm?
Theburningofrome.Com. <https://www.theburningofrome.com/blog/what-is-the-application-of-hits-algorithm/>
- Kim, J., Jang, S., Park, E., & Choi, S. (2020). Text classification using capsules. *Neurocomputing*, 376, 214–221. <https://doi.org/10.1016/j.neucom.2019.10.033>
- Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–1751. <https://doi.org/10.3115/v1/D14-1181>
- Kingma, D. P., & Ba, J. (2017). *Adam: A Method for Stochastic Optimization* (arXiv:1412.6980). arXiv. <http://arxiv.org/abs/1412.6980>
- Kipf, T. N., & Welling, M. (2017). *Semi-Supervised Classification with Graph Convolutional Networks* (arXiv:1609.02907). arXiv. <http://arxiv.org/abs/1609.02907>
- Kochkina, E., Liakata, M., & Zubiaga, A. (2018). All-in-one: Multi-task Learning for Rumour Verification. *Proceedings of the 27th International Conference on Computational Linguistics*, 3402–3413. <https://aclanthology.org/C18-1288>
- Kolitsas, N., Ganea, O.-E., & Hofmann, T. (2018). *End-to-End Neural Entity Linking* (arXiv:1808.07699). arXiv. <http://arxiv.org/abs/1808.07699>

- Kosciejew, M. (2021). The coronavirus pandemic, libraries and information: A thematic analysis of initial international responses to COVID-19. *Global Knowledge, Memory and Communication*, 70(4/5), 304–324.
<https://doi.org/10.1108/GKMC-04-2020-0041>
- Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, 31(3), 249–268.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed). Sage.
- Ksieniewicz, P., Choraś, M., Kozik, R., & Woźniak, M. (2019). Machine Learning Methods for Fake News Classification. In H. Yin, D. Camacho, P. Tino, A. J. Tallón-Ballesteros, R. Menezes, & R. Allmendinger (Eds.), *Intelligent Data Engineering and Automated Learning – IDEAL 2019* (Vol. 11872, pp. 332–339). Springer International Publishing. https://doi.org/10.1007/978-3-030-33617-2_34
- Kumar, A. (2020, July 28). *SVM RBF Kernel Parameters With Code Examples—DZone AI*. Dzone.Com. <http://dzone.com/articles/using-jsonb-in-postgresql-how-to-effectively-store-1>
- Kundu, S. (2013, July 25). *How to Code Independent Cascade Model of Information Diffusion*. Articles. <http://home.iitj.ac.in/~suman/articles/>
- Kwon, S., Cha, M., & Jung, K. (2017). Rumor Detection over Varying Time Windows. *PLOS ONE*, 12(1), e0168344. <https://doi.org/10.1371/journal.pone.0168344>
- Kwon, S., Chen, W., Cha, M., & Wang, Y. (n.d.). *Prominent Features of Rumor Propagation in Online Social Media*. 6. <https://doi.org/10.1109/ICDM.2013.61>
- Langley, P., Iba, W., & Thompson, K. (1992). An Analysis of Bayesian Classifiers. *Proceedings of the Tenth National Conference on Artificial Intelligence*, 223–228.
- Le, P., & Titov, I. (2018). *Improving Entity Linking by Modeling Latent Relations between Mentions* (arXiv:1804.10637). arXiv. <http://arxiv.org/abs/1804.10637>

- Lecy, J. D., & Beatty, K. E. (2012). *Representative Literature Reviews Using Constrained Snowball Sampling and Citation Network Analysis* (SSRN Scholarly Paper No. 1992601). <https://doi.org/10.2139/ssrn.1992601>
- Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788–791. <https://doi.org/10.1038/44565>
- Leonie, H., & Althuis, J. (Eds.). (2018). *Fake News. A Roadmap*. NATO Strategic Communications Centre of Excellence. <https://stratcomcoe.org/publications/fake-news-a-roadmap/137>
- Libraries and Fake News – Paper for Zagreb Roundtable on Freedom of Access to Information – IFLA*. (n.d.). Retrieved August 30, 2022, from <https://www.ifla.org/publications/libraries-and-fake-news-paper-for-zagreb-roundtable-on-freedom-of-access-to-information/>
- LIBSVM -- A Library for Support Vector Machines*. (n.d.). Retrieved November 9, 2022, from <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- Lim, S. (2020). Academic library guides for tackling fake news: A content analysis. *The Journal of Academic Librarianship*, *46*(5), 102195. <https://doi.org/10.1016/j.acalib.2020.102195>
- Lin, C.-Y. (2004). *ROUGE: A Package for Automatic Evaluation of Summaries*. 74–81. <https://www.microsoft.com/en-us/research/publication/rouge-a-package-for-automatic-evaluation-of-summaries/>
- Lin, T. Y., Fang, H., & Chen, X. (n.d.). *Coco-caption*. <https://github.com/tylin/coco-caption>
- Linear, Quadratic, and Regularized Discriminant Analysis*. (2018, November 30). <https://www.datascienceblog.net/post/machine-learning/linear-discriminant-analysis/>
- Liu, Y., Gao, C., She, X., & Zhang, Z. (2016). A bio-inspired method for locating the diffusion source with limited observers. *2016 IEEE Congress on Evolutionary Computation (CEC)*, 508–514. <https://doi.org/10.1109/CEC.2016.7743836>

- Liu, Y., Yu, K., Wu, X., Qing, L., & Peng, Y. (2019). Analysis and Detection of Health-Related Misinformation on Chinese Social Media. *IEEE Access*, 7, 154480–154489.
<https://doi.org/10.1109/ACCESS.2019.2946624>
- Lofi, C., & Maarry, K. E. (2014). Design Patterns for Hybrid Algorithmic-Crowdsourcing Workflows. *2014 IEEE 16th Conference on Business Informatics*, 1–8.
<https://doi.org/10.1109/CBI.2014.16>
- Long, Y., Lu, Q., Xiang, R., Li, M., & Huang, C.-R. (2017). Fake News Detection Through Multi-Perspective Speaker Profiles. *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 252–256.
<https://aclanthology.org/I17-2043>
- Lu, J., Yang, J., Batra, D., & Parikh, D. (2017). *Hierarchical Question-Image Co-Attention for Visual Question Answering* (arXiv:1606.00061). arXiv.
<http://arxiv.org/abs/1606.00061>
- Lukas Koster. (2011, November 18). *Linked Open Data for Libraries*.
<https://www.slideshare.net/lukask/linked-open-data-for-libraries>
- Luo, W., Tay, W. P., & Leng, M. (2013). Identifying Infection Sources and Regions in Large Networks. *IEEE Transactions on Signal Processing*, 61(11), 2850–2865.
<https://doi.org/10.1109/TSP.2013.2256902>
- Ma, J., Gao, W., Wei, Z., Lu, Y., & Wong, K.-F. (2015). Detect Rumors Using Time Series of Social Context Information on Microblogging Websites. *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 1751–1754. <https://doi.org/10.1145/2806416.2806607>
- Ma, J., Gao, W., & Wong, K.-F. (2017). Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 708–717. <https://doi.org/10.18653/v1/P17-1066>

- Ma, J., Gao, W., & Wong, K.-F. (2018). Rumor Detection on Twitter with Tree-structured Recursive Neural Networks. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1980–1989. <https://doi.org/10.18653/v1/P18-1184>
- Macnamara, J. (2018). 11. Content Analysis. In P. M. Napoli (Ed.), *Mediated Communication* (pp. 191–212). De Gruyter. <https://doi.org/10.1515/9783110481129-012>
- Mafarja, M. M., & Mirjalili, S. (2017). Hybrid Whale Optimization Algorithm with simulated annealing for feature selection. *Neurocomputing*, 260, 302–312. <https://doi.org/10.1016/j.neucom.2017.04.053>
- Mahabub, A. (2020). A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers. *SN Applied Sciences*, 2(4), 525. <https://doi.org/10.1007/s42452-020-2326-y>
- Manaswi, N. K. (2018). CNN in Keras. In N. K. Manaswi, *Deep Learning with Applications Using Python* (pp. 105–114). Apress. https://doi.org/10.1007/978-1-4842-3516-4_8
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text - Interdisciplinary Journal for the Study of Discourse*, 8(3). <https://doi.org/10.1515/text.1.1988.8.3.243>
- Marcheggiani, D., Bastings, J., & Titov, I. (2018). Exploiting Semantics in Neural Machine Translation with Graph Convolutional Networks. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 486–492. <https://doi.org/10.18653/v1/N18-2078>
- Marquardt, D. (2019). Linguistic Indicators in the Identification of Fake News. *Mediatization Studies*, 3, 95. <https://doi.org/10.17951/ms.2019.3.95-114>

- Matas, J., Chum, O., Urban, M., & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10), 761–767. <https://doi.org/10.1016/j.imavis.2004.02.006>
- Matsubara, Y., Sakurai, Y., Prakash, B. A., Li, L., & Faloutsos, C. (2012). Rise and fall patterns of information diffusion: Model and implications. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '12*, 6. <https://doi.org/10.1145/2339530.2339537>
- Maximally stable extremal regions. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Maximally_stable_extremal_regions&oldid=1087119909
- Maximum Entropy*. (2015, June). <http://www.di.fc.ul.pt/~jpn/r/maxent/maxent.html>
- Max-Planck-Institut für Informatik: YAGO*. (n.d.). Retrieved November 12, 2022, from <https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/>
- Media Bias/Fact Check News*. (n.d.). Media Bias/Fact Check. Retrieved September 9, 2022, from <https://mediabiasfactcheck.com/>
- Meel, P., & Vishwakarma, D. K. (2020). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications*, 153, 112986. <https://doi.org/10.1016/j.eswa.2019.112986>
- Merton, R. K. (1975). Thematic Analysis in Science: Notes on Holton's Concept. *Science*, 188(4186), 335–338. <https://doi.org/10.1126/science.188.4186.335>
- Methodology*. (n.d.). Media Bias/Fact Check. Retrieved September 9, 2022, from <https://mediabiasfactcheck.com/methodology/>
- Metodología de Maldito Bulo · Maldita.es—Periodismo para que no te la cuelen*. (2020, May 19). Maldita.es — Periodismo para que no te la cuelen. <https://maldita.es/metodologia-de-maldito-bulo/>

- MeVer—Image Verification Assistant*. (n.d.). Retrieved September 10, 2022, from <https://mever.iti.gr/forensics/>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space* (arXiv:1301.3781). arXiv.
<http://arxiv.org/abs/1301.3781>
- Mikolov, T., Grave, E., Bojanowski, P., Puhersch, C., & Joulin, A. (2018, May). Advances in Pre-Training Distributed Word Representations. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. LREC 2018, Miyazaki, Japan. <https://aclanthology.org/L18-1008>
- Milne, D., & Witten, I. H. (2008). Learning to link with wikipedia. *Proceeding of the 17th ACM Conference on Information and Knowledge Mining - CIKM '08*, 509.
<https://doi.org/10.1145/1458082.1458150>
- Mirjalili, S. (n.d.-a). *Grey Wolf Optimizer*. GWO. Retrieved October 23, 2022, from <https://seyedalimirjalili.com/gwo>
- Mirjalili, S. (n.d.-b). *Particle Swarm Optimization*. Particle Swarm Optimization. Retrieved November 11, 2022, from <https://seyedalimirjalili.com/pso>
- Mirjalili, S. (n.d.-c). *Salp Swarm Algorithm*. SSA. Retrieved October 23, 2022, from <https://seyedalimirjalili.com/ssa>
- Mirjalili, S. (n.d.-d). *Whale optimization algorithm*. WOA. Retrieved October 23, 2022, from <https://seyedalimirjalili.com/woa>
- Mirjalili, S., Gandomi, A. H., Mirjalili, S. Z., Saremi, S., Faris, H., & Mirjalili, S. M. (2017). Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems. *Advances in Engineering Software*, 114, 163–191.
<https://doi.org/10.1016/j.advengsoft.2017.07.002>
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. *Advances in Engineering Software*, 69, 46–61.
<https://doi.org/10.1016/j.advengsoft.2013.12.007>

- Mitra, T., & Gilbert, E. (2015). CREDBANK: A Large-Scale Social Media Corpus With Associated Credibility Annotations. *Proceedings of the International AAAI Conference on Web and Social Media*, 9(1), 258–267.
- Mohammad, S. (n.d.). *Saif | VAD Lexicon*. Retrieved October 27, 2022, from <http://saifmohammad.com/WebPages/nrc-vad.html>
- Mohammad, S. (2020). *NRC Affect Intensity Lexicon*.
<https://saifmohammad.com/WebPages/AffectIntensity.htm>
- Mohammad, S. M. (2017). *Word Affect Intensities* (arXiv:1704.08798). arXiv.
<http://arxiv.org/abs/1704.08798>
- Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M. M. (2019). *Fake News Detection on Social Media using Geometric Deep Learning* (arXiv:1902.06673). arXiv. <http://arxiv.org/abs/1902.06673>
- Monzani, D., Vergani, L., Pizzoli, S. F. M., Marton, G., & Pravettoni, G. (2021). Emotional Tone, Analytical Thinking, and Somatosensory Processes of a Sample of Italian Tweets During the First Phases of the COVID-19 Pandemic: Observational Study. *Journal of Medical Internet Research*, 23(10), e29820.
<https://doi.org/10.2196/29820>
- Moral Foundations Theory | moralfoundations.org*. (n.d.). Retrieved September 15, 2022, from <https://moralfoundations.org/>
- Moreau, L. (n.d.). *ProvToolbox*. Retrieved September 10, 2022, from <http://lucmoreau.github.io/ProvToolbox/>
- Moreda, P., Llorens, H., Saquete, E., & Palomar, M. (2011). Combining semantic information in question answering systems. *Information Processing & Management*, 47(6), 870–885. <https://doi.org/10.1016/j.ipm.2010.03.008>
- Mourão, R. R., & Robertson, C. T. (2019). Fake News as Discursive Integration: An Analysis of Sites That Publish False, Misleading, Hyperpartisan and Sensational

- Information. *Journalism Studies*, 20(14), 2077–2095.
<https://doi.org/10.1080/1461670X.2019.1566871>
- MPQA Resources. (n.d.). Retrieved September 11, 2022, from
http://mpqa.cs.pitt.edu/#subj_lexicon
- Multi-class AdaBoosted Decision Trees. (n.d.). Scikit-Learn. Retrieved September 13, 2022, from https://scikit-learn/stable/auto_examples/ensemble/plot_adaboost_multiclass.html
- MVAE. (n.d.). <https://github.com/dhruvkhattar/MVAE>
- Naeem, S. B., Bhatti, R., & Khan, A. (2021). An exploration of how fake news is taking over social media and putting public health at risk. *Health Information & Libraries Journal*, 38(2), 143–149. <https://doi.org/10.1111/hir.12320>
- Nanni, L., Ghidoni, S., & Brahnam, S. (2017). Handcrafted vs. Non-handcrafted features for computer vision classification. *Pattern Recognition*, 71, 158–172.
<https://doi.org/10.1016/j.patcog.2017.05.025>
- Natural Language API Basics. (n.d.). Google Cloud. Retrieved September 16, 2022, from <https://cloud.google.com/natural-language/docs/basics>
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Sage Publications.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying Words: Predicting Deception from Linguistic Styles. *Personality and Social Psychology Bulletin*, 29(5), 665–675. <https://doi.org/10.1177/0146167203029005010>
- News and current affairs from Germany and around the world – DW. (n.d.). Dw.Com. Retrieved October 29, 2022, from <https://www.dw.com/en/top-stories/s-9097>
- NewsGuard—Combating Misinformation with Trust Ratings for News. (n.d.). NewsGuard. Retrieved September 9, 2022, from <https://www.newsguardtech.com/>
- Nguyen, N. P., Yan, G., Thai, M. T., & Eidenbenz, S. (2012). Containment of misinformation spread in online social networks. *Proceedings of the 3rd Annual ACM Web Science*

- Conference on - WebSci '12*, 213–222.
<https://doi.org/10.1145/2380718.2380746>
- NLTK :: Natural Language Toolkit*. (2022, March 25). <https://www.nltk.org/index.html>
- Nouh, M., Nurse, J. R. C., & Goldsmith, M. (2019). Understanding the Radical Mind: Identifying Signals to Detect Extremist Content on Twitter. *2019 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 98–103.
<https://doi.org/10.1109/ISI.2019.8823548>
- Noy, C. (2008). Sampling Knowledge: The Hermeneutics of Snowball Sampling in Qualitative Research. *International Journal of Social Research Methodology*, 11(4), 327–344. <https://doi.org/10.1080/13645570701401305>
- NRC Affect Intensity Lexicon*. (n.d.). Retrieved September 15, 2022, from <https://saifmohammad.com/WebPages/AffectIntensity.htm>
- Olah, C. (2015, August 27). *Understanding LSTM Networks—Colah's blog*. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Orhan, U., Hekim, M., & Ozer, M. (2011). EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Systems with Applications*, 38(10), 13475–13481.
<https://doi.org/10.1016/j.eswa.2011.04.149>
- Our Fact Check Ratings, Explained*. (n.d.). Snopes.Com. Retrieved September 9, 2022, from <https://www.snopes.com/fact-check-ratings/>
- Ozbay, F. A., & Alatas, B. (2020). Fake news detection within online social media using supervised artificial intelligence algorithms. *Physica A: Statistical Mechanics and Its Applications*, 540, 123174. <https://doi.org/10.1016/j.physa.2019.123174>
- Paluch, R., Gajewski, Ł. G., Hołyst, J. A., & Szymanski, B. K. (2020). Optimizing sensors placement in complex networks for localization of hidden signal source: A review. *Future Generation Computer Systems*, 112, 1070–1092.
<https://doi.org/10.1016/j.future.2020.06.023>

- Panda, A. (2020, December 13). A Visual and Intuitive Guide to LSTM, GRU, and Attention. *The Startup*. <https://medium.com/swlh/a-visual-and-intuitive-guide-to-lstm-gru-and-attention-8350518fa849>
- Pandey, A. C., & Rajpoot, D. S. (2019). Spam review detection using spiral cuckoo search clustering method. *Evolutionary Intelligence*, 12(2), 147–164. <https://doi.org/10.1007/s12065-019-00204-x>
- Pandey, A. C., & Tikkiwal, V. A. (2021). Stance detection using improved whale optimization algorithm. *Complex & Intelligent Systems*, 7(3), 1649–1672. <https://doi.org/10.1007/s40747-021-00294-0>
- Pandey, A., Iyer, A., Pathak, K., Dipti, P., & Hajgude, J. (2014). Email Filtering and Analysis Using Classification Algorithms. *International Journal of Computer Science Issues - IJCSI*, 11(4), 115–124.
- Papadopoulou, O., Zampoglou, M., Papadopoulos, S., & Kompatsiaris, I. (2019). A corpus of debunked and verified user-generated videos. *Online Information Review*, 43(1), 72–88. <https://doi.org/10.1108/OIR-03-2018-0101>
- Papadopoulou, O., Zampoglou, M., Papadopoulos, S., & Kompatsiaris, Y. (2017). Web Video Verification using Contextual Cues. *Proceedings of the 2nd International Workshop on Multimedia Forensics and Security*, 6–10. <https://doi.org/10.1145/3078897.3080535>
- Papers with Code—BiGRU Explained*. (n.d.). Retrieved September 14, 2022, from <https://paperswithcode.com/method/bigru>
- Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2001). BLEU: A method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, 311. <https://doi.org/10.3115/1073083.1073135>
- Parmentier, A., & Cohen, R. (2020). Personalized Multi-Faceted Trust Modeling in Social Networks. In C. Goutte & X. Zhu (Eds.), *Advances in Artificial Intelligence* (Vol.

- 12109, pp. 445–450). Springer International Publishing.
https://doi.org/10.1007/978-3-030-47358-7_46
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & Lerer, A. (2017). *Automatic differentiation in PyTorch*. 4.
<https://openreview.net/forum?id=BJJsrnfCZ>
- Patidar, V., Singh, D., & Singh, A. (2013). A Novel Technique of Email Classification for Spam Detection. *International Journal of Applied Information Systems*, 5(10), 15–19. <https://doi.org/10.5120/ijais13-450976>
- Patil, S., Gune, A., & Nene, M. (2017). Convolutional neural networks for text categorization with latent semantic analysis. *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, 499–503. <https://doi.org/10.1109/ICECDS.2017.8390217>
- Pennebaker. (n.d.). *Welcome to LIWC-22*. Retrieved October 27, 2022, from <https://www.liwc.app/>
- Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us* (1st U.S. ed). Bloomsbury Press.
- Pennington, J., Socher, R., & Manning, C. D. (2014, August). *GloVe: Global Vectors for Word Representation*. <https://nlp.stanford.edu/projects/glove/>
- Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2017). *Automatic Detection of Fake News* (arXiv:1708.07104). arXiv. <http://arxiv.org/abs/1708.07104>
- Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2018). Automatic Detection of Fake News. *Proceedings of the 27th International Conference on Computational Linguistics*, 3391–3401. <https://aclanthology.org/C18-1287>
- Perspective | Developers*. (n.d.). Retrieved September 11, 2022, from <https://developers.perspectiveapi.com/s/about-the-api-score>
- Perspective API*. (n.d.). Using Machine Learning to Reduce Toxicity Online. Retrieved September 11, 2022, from <https://www.perspectiveapi.com/>

- Poddar, K., Amali D., G. B., & Umadevi, K. S. (2019). Comparison of Various Machine Learning Models for Accurate Detection of Fake News. *2019 Innovations in Power and Advanced Computing Technologies (i-PACT)*, 1–5. <https://doi.org/10.1109/i-PACT44901.2019.8960044>
- PolitiFact*. (n.d.). Retrieved September 9, 2022, from <https://www.politifact.com/>
- Portada · Maldita.es—Periodismo para que no te la cuelen*. (n.d.). Retrieved September 9, 2022, from <https://maldita.es/>
- Posadas-Durán, J.-P., Gómez-Adorno, H., Sidorov, G., & Escobar, J. J. M. (2019). Detection of fake news in a new corpus for the Spanish language. *Journal of Intelligent & Fuzzy Systems*, 36(5), 4869–4876. <https://doi.org/10.3233/JIFS-179034>
- Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2017). *A Stylometric Inquiry into Hyperpartisan and Fake News* (arXiv:1702.05638). arXiv. <http://arxiv.org/abs/1702.05638>
- Prabhakaran, S. (2018, October 22). Cosine Similarity - Understanding the math and how it works? (With python). *Machine Learning Plus*. <https://www.machinelearningplus.com/nlp/cosine-similarity/>
- PROV-DM: The PROV Data Model*. (n.d.). Retrieved September 10, 2022, from <https://www.w3.org/TR/prov-dm/#section-overview-components>
- PROV-O: The PROV Ontology*. (n.d.). Retrieved September 17, 2022, from <https://www.w3.org/TR/prov-o/>
- Pupale, R. (2019, February 11). *Support Vector Machines(SVM)—An Overview*. Medium. <https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989>
- PyTorch*. (n.d.). Retrieved November 12, 2022, from <https://www.pytorch.org>
- Qian, F., Gong, C., Sharma, K., & Liu, Y. (2018). Neural User Response Generator: Fake News Detection with Collective User Intelligence. *Proceedings of the Twenty-*

- Seventh International Joint Conference on Artificial Intelligence*, 3834–3840.
<https://doi.org/10.24963/ijcai.2018/533>
- Qian, S., Hu, J., Fang, Q., & Xu, C. (2021). Knowledge-aware Multi-modal Adaptive Graph Convolutional Networks for Fake News Detection. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 17(3), 1–23.
<https://doi.org/10.1145/3451215>
- Rath, B., Gao, W., Ma, J., & Srivastava, J. (2017). From Retweet to Believability: Utilizing Trust to Identify Rumor Spreaders on Twitter. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 179–186. <https://doi.org/10.1145/3110025.3110121>
- Rating Process and Criteria*. (2022, April 8). NewsGuard.
<https://www.newsguardtech.com/ratings/rating-process-criteria/>
- Recasens, M., Danescu-Niculescu-Mizil, C., & Jurafsky, D. (2013). Linguistic Models for Analyzing and Detecting Biased Language. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1650–1659. <https://aclanthology.org/P13-1162>
- Reis, J. C. S., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019). Supervised Learning for Fake News Detection. *IEEE Intelligent Systems*, 34(2), 76–81.
<https://doi.org/10.1109/MIS.2019.2899143>
- Rhetorical Structure Theory*. (n.d.-a). Retrieved September 12, 2022, from
<https://www.sfu.ca/rst/01intro/intro.html>
- Rhetorical Structure Theory*. (n.d.-b). Retrieved September 12, 2022, from
<https://www.sfu.ca/rst/01intro/definitions.html>
- Riffe, D., Lacy, S., Watson, B. R., & Fico, F. (2019). *Analyzing Media Messages: Using Quantitative Content Analysis in Research* (4th ed.). Routledge.
<https://doi.org/10.4324/9780429464287>

- Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1–2), 1–39. <https://doi.org/10.1007/s10462-009-9124-7>
- rstWeb—Homepage*. (n.d.). Retrieved September 12, 2022, from <https://gucorpling.org/rstweb/info/>
- Rubin, V. L., Conroy, N. J., & Chen, Y. (2015). *Towards News Verification: Deception Detection Methods for News Discourse*. HICSS2015. <https://works.bepress.com/victoriarubin/6/>
- Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A Hybrid Deep Model for Fake News Detection. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 797–806. <https://doi.org/10.1145/3132847.3132877>
- Ruder, S. (2017). *An overview of gradient descent optimization algorithms* (arXiv:1609.04747). arXiv. <http://arxiv.org/abs/1609.04747>
- Rumor_RvNN*. (2018). https://github.com/majingCUHK/Rumor_RvNN
- Rumor_RvNN/model/Main_TD_RvNN.py*. (n.d.). https://github.com/majingCUHK/Rumor_RvNN/blob/master/model/Main_TD_RvNN.py
- RumourEval: Determining rumour veracity and support for rumours < SemEval-2017 Task 8*. (n.d.). Retrieved September 12, 2022, from <https://alt.qcri.org/semeval2017/task8/>
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110(1), 145–172. <https://doi.org/10.1037/0033-295X.110.1.145>
- Saaty, R. W. (1987). The analytic hierarchy process—What it is and how it is used. *Mathematical Modelling*, 9(3–5), 161–176. [https://doi.org/10.1016/0270-0255\(87\)90473-8](https://doi.org/10.1016/0270-0255(87)90473-8)
- SAFE*. (n.d.). <https://github.com/Jindi0/SAFE>

- Sahoo, J., Sahu, S. C., & Mohanty, B. (2021). Research on Fake News An Empirical Analysis of Selected Library and Information Science Journals. *DESIDOC Journal of Library & Information Technology*, 41(4), 268–274.
<https://doi.org/10.14429/djlit.41.4.17168>
- Saif / VAD Lexicon. (2022, August). <https://saifmohammad.com/WebPages/nrc-vad.html>
- Saikh, T., Anand, A., Ekbal, A., & Bhattacharyya, P. (2019). A Novel Approach Towards Fake News Detection: Deep Learning Augmented with Textual Entailment Features. In E. Métais, F. Meziane, S. Vadera, V. Sugumaran, & M. Saraee (Eds.), *Natural Language Processing and Information Systems* (Vol. 11608, pp. 345–358). Springer International Publishing. https://doi.org/10.1007/978-3-030-23281-8_30
- Saikh, T., Ghosal, T., Ekbal, A., & Bhattacharyya, P. (2017). Document Level Novelty Detection: Textual Entailment Lends a Helping Hand. *Proceedings of the 14th International Conference on Natural Language Processing (ICON-2017)*, 131–140.
<https://aclanthology.org/W17-7517>
- Santosh, T. Y. S. S., & Aravind, K. V. S. (2019). Hate Speech Detection in Hindi-English Code-Mixed Social Media Text. *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, 310–313.
<https://doi.org/10.1145/3297001.3297048>
- Sciortino, A. (2020). FAKE NEWS AND INFODEMIA AT THE TIME OF COVID-19. *Direito Público*, 17(94).
<https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/4823>
- Scrapy | A Fast and Powerful Scraping and Web Crawling Framework. (n.d.). Retrieved September 10, 2022, from <https://scrapy.org/>
- Sec, D. I. (2021, March 6). VGG-19 Convolutional Neural Network. *All about Machine Learning*. <https://blog.techcraft.org/vgg-19-convolutional-neural-network/>

- Seldon. (2021, June 29). *Transfer Learning for Machine Learning*. Seldon.
<https://www.seldon.io/transfer-learning>
- Sentiment and Emotion Lexicons*. (n.d.). Retrieved September 15, 2022, from
<http://saifmohammad.com/WebPages/lexicons.html>
- Shah, D., & Zaman, T. (2011). Rumors in a Network: Who's the Culprit? *IEEE Transactions on Information Theory*, 57(8), 5163–5181.
<https://doi.org/10.1109/TIT.2011.2158885>
- Shao, C., Ciampaglia, G. L., Varol, O., Yang, K., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, 9(1), 4787. <https://doi.org/10.1038/s41467-018-06930-7>
- Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating Fake News: A Survey on Identification and Mitigation Techniques. *ACM Transactions on Intelligent Systems and Technology*, 10(3), 1–42.
<https://doi.org/10.1145/3305260>
- SHARMA, N. (2020, June 17). Bernoulli Naive Bayes and it's implementation. *Medium*.
<https://medium.com/@nansha3120/bernoulli-naive-bayes-and-its-implementation-cca33ccb8d2e>
- Sheikhi, S. (2021). An effective fake news detection method using WOA-xgbTree algorithm and content-based features. *Applied Soft Computing*, 109, 107559.
<https://doi.org/10.1016/j.asoc.2021.107559>
- Shevtsova, Y., & Shemaieva, H. (2020). Content analysis of European library and information science journals. *Technium Social Sciences Journal*, 8, 161–170.
<https://doi.org/10.47577/tssj.v8i1.540>
- Shim, J.-S., Lee, Y., & Ahn, H. (2021). A link2vec-based fake news detection model using web search results. *Expert Systems with Applications*, 184, 115491.
<https://doi.org/10.1016/j.eswa.2021.115491>

- Shouse, E. (2005). Feeling, Emotion, Affect. *M/C Journal*, 8(6).
<https://doi.org/10.5204/mcj.2443>
- Shu, K., Cui, L., Wang, S., Lee, D., & Liu, H. (n.d.). *DEFEND*. Dropbox. Retrieved October 26, 2022, from
<https://www.dropbox.com/sh/rzczwopo618jyv2/AAA1mI2yvbt6TAfqpcxfjL8va?dl=0>
- Shu, K., Cui, L., Wang, S., Lee, D., & Liu, H. (2019). dEFEND: Explainable Fake News Detection. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 395–405.
<https://doi.org/10.1145/3292500.3330935>
- Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). *FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media* (arXiv:1809.01286v1). arXiv.
<http://arxiv.org/abs/1809.01286>
- Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2019). *FakeNewsNet: A Data Repository with News Content, Social Context and Spatialtemporal Information for Studying Fake News on Social Media* (arXiv:1809.01286). arXiv.
<http://arxiv.org/abs/1809.01286>
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). *Fake News Detection on Social Media: A Data Mining Perspective* (arXiv:1708.01967). arXiv.
<http://arxiv.org/abs/1708.01967>
- Shu, K., Wang, S., & Liu, H. (2019). Beyond News Contents: The Role of Social Context for Fake News Detection. *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 312–320.
<https://doi.org/10.1145/3289600.3290994>

- Silva, R. M., Santos, R. L. S., Almeida, T. A., & Pardo, T. A. S. (2020). Towards automatically filtering fake news in Portuguese. *Expert Systems with Applications*, 146, 113199. <https://doi.org/10.1016/j.eswa.2020.113199>
- Silverman, C., Buttry, S., Wardle, C., Barot, T., Browne, M., Ingram, M., Meier, P., Knight, S., Tsubaki, R., Ayala Iacucci, A., Phillips, T., Stowell, D., Koettl, C., Law, P., De Rosa, A., Trewinnard, T., Durand, S., Bair, M., Bannock, C., ... SH, M. (n.d.). *Verification Handbook: Homepage*. Retrieved October 29, 2022, from <http://verificationhandbook.com>
- Simha, A. (n.d.). *Understanding TF-IDF for Machine Learning*. Capital One. Retrieved September 12, 2022, from <https://www.capitalone.com/tech/machine-learning/understanding-tf-idf/>
- Simonyan, K. (n.d.). *ILSVRC-2014 model (VGG team) with 19 weight layers*. Gist. Retrieved October 29, 2022, from <https://gist.github.com/ksimonyan/3785162f95cd2d5fee77>
- Simonyan, K., & Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition* (arXiv:1409.1556; Version 6). arXiv. <https://doi.org/10.48550/arXiv.1409.1556>
- Singhal, S. (n.d.). *SpotFakePlus*. <https://github.com/shiivangii/SpotFakePlus>
- Singhal, S., Kabra, A., Sharma, M., Shah, R. R., Chakraborty, T., & Kumaraguru, P. (2020). SpotFake+: A Multimodal Framework for Fake News Detection via Transfer Learning (Student Abstract). *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(10), 13915–13916. <https://doi.org/10.1609/aaai.v34i10.7230>
- Singhania, S., Fernandez, N., & Rao, S. (n.d.). *3HAN*. <https://github.com/ni9elf/3HAN>
- Singhania, S., Fernandez, N., & Rao, S. (2017). 3HAN: A Deep Neural Network for Fake News Detection. In D. Liu, S. Xie, Y. Li, D. Zhao, & E.-S. M. El-Alfy (Eds.), *Neural Information Processing* (Vol. 10635, pp. 572–581). Springer International Publishing. https://doi.org/10.1007/978-3-319-70096-0_59

- Sitistas, T., & Pournis, S. (2019, October 7). *Ellinika Hoaxes-Τα αγορασμένα “βραβεία ματαιοδοξίας” του δημάρχου Άργους-Μυκηνών*. Ellinika Hoaxes.
<https://www.ellinikahoaxes.gr/2019/10/07/kamposos-awards/>
- Sklearn.ensemble.GradientBoostingClassifier*. (n.d.). Scikit-Learn. Retrieved November 11, 2022, from <https://scikit-learn/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html>
- Sklearn.naive_bayes.BernoulliNB*. (n.d.). Scikit-Learn. Retrieved November 10, 2022, from https://scikit-learn/stable/modules/generated/sklearn.naive_bayes.BernoulliNB.html
- Sklearn.svm.LinearSVC*. (n.d.). Scikit-Learn. Retrieved November 9, 2022, from <https://scikit-learn/stable/modules/generated/sklearn.svm.LinearSVC.html>
- Smith-Waterman Search—Unipro UGENE Online User Manual v. 1.15.0—WIKI*. (n.d.). Retrieved November 11, 2022, from <https://doc.ugene.net/wiki/display/UUOUM15/Smith-Waterman+Search>
- Snopes.com*. (n.d.). Snopes.Com. Retrieved September 9, 2022, from <https://www.snopes.com/>
- SNU FactCheck Wins Human Technology Award at Asia Future Forum—News—Newsroom—SNU NOW*. (n.d.). Seoul National University. Retrieved September 9, 2022, from https://en.snu.ac.kr/snunow/snu_media/news?md=v&bbsidx=125749
- SNUFactCheck*. (n.d.). Retrieved September 9, 2022, from <https://factcheck.snu.ac.kr/home/about>
- Sotirakou, C., Karampela, A., & Mourlas, C. (2021). Evaluating the Role of News Content and Social Media Interactions for Fake News Detection. In J. Bright, A. Giachanou, V. Spaiser, F. Spezzano, A. George, & A. Pavliuc (Eds.), *Disinformation in Open*

- Online Media* (Vol. 12887, pp. 128–141). Springer International Publishing.
https://doi.org/10.1007/978-3-030-87031-7_9
- Souza Freire, P. M., Matias da Silva, F. R., & Goldschmidt, R. R. (2021). Fake news detection based on explicit and implicit signals of a hybrid crowd: An approach inspired in meta-learning. *Expert Systems with Applications*, 183, 115414.
<https://doi.org/10.1016/j.eswa.2021.115414>
- SpaCy · Industrial-strength Natural Language Processing in Python*. (n.d.). Retrieved September 10, 2022, from <https://spacy.io/>
- Spinelli, B., Celis, L. E., & Thiran, P. (2016). Observer placement for source localization: The effect of budgets and transmission variance. *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 743–751.
<https://doi.org/10.1109/ALLERTON.2016.7852307>
- Srinivasan, A. V. (2019, September 7). *Stochastic Gradient Descent—Clearly Explained !!* Medium. <https://towardsdatascience.com/stochastic-gradient-descent-clearly-explained-53d239905d31>
- Stecanella, B. (2019, May 10). *Understanding TF-ID: A Simple Introduction*. MonkeyLearn Blog. <https://monkeylearn.com/blog/what-is-tf-idf/>
- Stiahnite si Bullshit detector*. (n.d.). Websupport.sk. Retrieved September 10, 2022, from <https://www.websupport.sk/o-nas/bullshit-detector/>
- Suchanek, F. M., Kasneci, G., & Weikum, G. (2008). YAGO: A Large Ontology from Wikipedia and WordNet. *Journal of Web Semantics*, 6(3), 203–217.
<https://doi.org/10.1016/j.websem.2008.06.001>
- Sumpter, R. S. (2001). News about News: John G. Speed and the First Newspaper Content Analysis. *Journalism History*, 27(2), 64–72.
<https://doi.org/10.1080/00947679.2001.12062572>

- Support Vector Machine (SVM) Algorithm*—Javatpoint. (n.d.). Wwww.Javatpoint.Com. Retrieved November 9, 2022, from <https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>
- Systems Innovation (Director). (n.d.). *Network Diffusion & Contagion*. Retrieved November 13, 2022, from <https://www.youtube.com/watch?v=bTXUJQhEqL0>
- Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & de Alfaro, L. (2017). *Some Like it Hoax: Automated Fake News Detection in Social Networks* (arXiv:1704.07506). arXiv. <http://arxiv.org/abs/1704.07506>
- Tagliabue, F., Galassi, L., & Mariani, P. (2020). The “Pandemic” of Disinformation in COVID-19. *SN Comprehensive Clinical Medicine*, 2(9), 1287–1289. <https://doi.org/10.1007/s42399-020-00439-1>
- Tanaka-Ishii, K., & Aihara, S. (2015). Computational Constancy Measures of Texts—Yule’s K and Rényi’s Entropy. *Computational Linguistics*, 41(3), 481–502. https://doi.org/10.1162/COLI_a_00228
- Taskin, S. G., Kucuksille, E. U., & Topal, K. (2022). Detection of Turkish Fake News in Twitter with Machine Learning Algorithms. *Arabian Journal for Science and Engineering*, 47(2), 2359–2379. <https://doi.org/10.1007/s13369-021-06223-0>
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), 24–54. <https://doi.org/10.1177/0261927X09351676>
- Team, K. (n.d.-a). *Keras documentation: Adam*. Retrieved October 27, 2022, from <https://keras.io/api/optimizers/adam/>
- Team, K. (n.d.-b). *Keras documentation: Attention layer*. Retrieved November 12, 2022, from https://keras.io/api/layers/attention_layers/attention/
- Team, K. (n.d.-c). *Keras documentation: Bidirectional layer*. Retrieved October 27, 2022, from https://keras.io/api/layers/recurrent_layers/bidirectional/

Team, K. (n.d.-d). *Keras documentation: Convolution layers*. Retrieved October 27, 2022, from https://keras.io/api/layers/convolution_layers/

Team, K. (n.d.-e). *Keras documentation: Dense layer*. Retrieved October 27, 2022, from https://keras.io/api/layers/core_layers/dense/

Team, K. (n.d.-f). *Keras documentation: Flatten layer*. Retrieved October 27, 2022, from https://keras.io/api/layers/reshaping_layers/flatten/

Team, K. (n.d.-g). *Keras documentation: GRU layer*. Retrieved October 27, 2022, from https://keras.io/api/layers/recurrent_layers/gru/

Team, K. (n.d.-h). *Keras documentation: Layer activation functions*. Retrieved October 27, 2022, from <https://keras.io/api/layers/activations/>

Team, K. (n.d.-i). *Keras documentation: LSTM layer*. Retrieved October 27, 2022, from https://keras.io/api/layers/recurrent_layers/lstm/

Team, K. (n.d.-j). *Keras documentation: SimpleRNN layer*. Retrieved October 27, 2022, from https://keras.io/api/layers/recurrent_layers/simple_rnn/

TensorFlow Hub. (n.d.). Retrieved November 9, 2022, from <https://tfhub.dev/google/collections/universal-sentence-encoder/1>

TextBlob: Simplified Text Processing—TextBlob 0.16.0 documentation. (n.d.). Retrieved September 15, 2022, from <https://textblob.readthedocs.io/en/dev/>

TextClassifier. (n.d.). <https://github.com/richliao/textClassifier>

Text_gcn. (2019). https://github.com/yao8839836/text_gcn

Tf.keras.Sequential | TensorFlow v2.10.0. (n.d.). TensorFlow. Retrieved October 27, 2022, from https://www.tensorflow.org/api_docs/python/tf/keras/Sequential

THE AUTOMATED READABILITY INDEX (ARI)—LEARN HOW TO CALCULATE THE AUTOMATED READABILITY INDEX. (n.d.). Retrieved September 11, 2022, from <https://readabilityformulas.com/automated-readability-index.php>

The city getting rich from fake news. (2016, December 5). *BBC News*. <https://www.bbc.com/news/magazine-38168281>

THE COLEMAN-LIAU READABILITY FORMULA (also known as The Coleman-Liau Index).

(n.d.). Retrieved September 11, 2022, from

<https://readabilityformulas.com/coleman-liau-readability-formula.php>

THE FLESCH GRADE LEVEL READABILITY FORMULA. (n.d.). Retrieved September 11,

2022, from [https://readabilityformulas.com/flesch-grade-level-readability-](https://readabilityformulas.com/flesch-grade-level-readability-formula.php)

[formula.php](https://readabilityformulas.com/flesch-grade-level-readability-formula.php)

THE FLESCH READING EASE READABILITY FORMULA. (n.d.). Retrieved September 11,

2022, from [https://readabilityformulas.com/flesch-reading-ease-readability-](https://readabilityformulas.com/flesch-reading-ease-readability-formula.php)

[formula.php](https://readabilityformulas.com/flesch-reading-ease-readability-formula.php)

THE GUNNING FOG READABILITY FORMULA. (n.d.). Retrieved September 11, 2022, from

<https://readabilityformulas.com/gunning-fog-readability-formula.php>

The Levenshtein Algorithm. (2017, January 25). *Cuelogic An LTI Company.*

<https://www.cuelogic.com/blog/the-levenshtein-algorithm>

The news hub. (2012, January 16). AFP.Com. <https://www.afp.com/en/news-hub>

The Selenium Browser Automation Project. (n.d.). Selenium. Retrieved September 10,

2022, from <https://www.selenium.dev/documentation/>

The SIR Model for Spread of Disease—The Differential Equation Model | Mathematical

Association of America. (n.d.). Retrieved November 14, 2022, from

<https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model>

THE SMOG READABILITY FORMULA, a Simple Measure of Gobbledygook. (n.d.). Retrieved

September 11, 2022, from [https://readabilityformulas.com/smog-readability-](https://readabilityformulas.com/smog-readability-formula.php)

[formula.php](https://readabilityformulas.com/smog-readability-formula.php)

The Stanford Natural Language Processing Group. (n.d.). Retrieved November 7, 2022,

from <https://nlp.stanford.edu/software/tagger.shtml>

- Thota, A., Tilak, P., Ahluwalia, S., & Lohia, N. (2018). Fake News Detection: A Deep Learning Approach. *SMU Data Science Review*, 1(3).
<https://scholar.smu.edu/datasciencereview/vol1/iss3/10>
- Tiantian Qin, Burgoon, J. K., Blair, J. P., & Nunamaker, J. F. (2005). Modality Effects in Deception Detection and Applications in Automatic-Deception-Detection. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 23b–23b. <https://doi.org/10.1109/HICSS.2005.436>
- Tran, Q.-A., Duan, H., & Li, X. (2006). Real-time statistical rules for spam detection. *IJCSNS International Journal of Computer Science and Network Security*, 6(2B), 178–184. http://paper.ijcsns.org/07_book/200602/200602C12.pdf
- Tschiatschek, S., Singla, A., Gomez Rodriguez, M., Merchant, A., & Krause, A. (2018). Fake News Detection in Social Networks via Crowd Signals. *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*, 517–524.
<https://doi.org/10.1145/3184558.3188722>
- Tuan Vu, M., Faculty of Information Technology, Hanoi University, Hanoi, Vietnam, Tran, Q. A., Faculty of Information Technology, Hanoi University, Hanoi, Vietnam, Jiang, F., School of Engineering and IT, University of New South Wales, Canberra, Australia, Tran, V. Q., & Faculty of Information Technology, Hanoi University, Hanoi, Vietnam. (2015). Multilingual Rules for Spam Detection. *Journal of Machine to Machine Communications*, 1(2), 107–122.
<https://doi.org/10.13052/jmmc2246-137X.122>
- Tundis, A., Mukherjee, G., & Mühlhäuser, M. (2021). An Algorithm for the Detection of Hidden Propaganda in Mixed-Code Text over the Internet. *Applied Sciences*, 11(5), 2196. <https://doi.org/10.3390/app11052196>
- Turenne, N. (2018). The rumour spectrum. *PLOS ONE*, 13(1), e0189080.
<https://doi.org/10.1371/journal.pone.0189080>

- User, S. (n.d.). *Vanilla_RNN_for_Classification*. Retrieved September 14, 2022, from <https://www.easy-tensorflow.com/tf-tutorials/recurrent-neural-networks/vanilla-rnn-for-classification>
- Varpio, L., Paradis, E., Uijtdehaage, S., & Young, M. (2020). The Distinctions Between Theory, Theoretical Framework, and Conceptual Framework. *Academic Medicine*, 95(7), 989–994. <https://doi.org/10.1097/ACM.0000000000003075>
- Varshney, D., & Vishwakarma, D. K. (2021). Hoax news-inspector: A real-time prediction of fake news using content resemblance over web search results for authenticating the credibility of news articles. *Journal of Ambient Intelligence and Humanized Computing*, 12(9), 8961–8974. <https://doi.org/10.1007/s12652-020-02698-1>
- Vedantam, R., Zitnick, C. L., & Parikh, D. (2015). *CIDEr: Consensus-based Image Description Evaluation* (arXiv:1411.5726). arXiv. <http://arxiv.org/abs/1411.5726>
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (n.d.). *Graph Attention Networks*. Retrieved October 27, 2022, from <https://petar-v.com/GAT/>
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). *Graph Attention Networks* (arXiv:1710.10903). arXiv. <http://arxiv.org/abs/1710.10903>
- Vicario, M. D., Quattrociocchi, W., Scala, A., & Zollo, F. (2019). Polarization and Fake News: Early Warning of Potential Misinformation Targets. *ACM Transactions on the Web*, 13(2), 1–22. <https://doi.org/10.1145/3316809>
- Vishwakarma, D. K., Varshney, D., & Yadav, A. (2019). Detection and veracity analysis of fake news via scrapping and authenticating the web search. *Cognitive Systems Research*, 58, 217–229. <https://doi.org/10.1016/j.cogsys.2019.07.004>
- Visual Geometry Group—University of Oxford. (n.d.). Very Deep Convolutional Networks for Large-Scale Visual Recognition. Retrieved October 29, 2022, from https://www.robots.ox.ac.uk/~vgg/research/very_deep/

- Volkova, S., Shaffer, K., Jang, J. Y., & Hodas, N. (2017). Separating Facts from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts on Twitter. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 647–653.
<https://doi.org/10.18653/v1/P17-2102>
- Vosoughi, S., Mohsenvand, M. 'Neo,' & Roy, D. (2017). Rumor Gauge: Predicting the Veracity of Rumors on Twitter. *ACM Transactions on Knowledge Discovery from Data*, 11(4), 1–36. <https://doi.org/10.1145/3070644>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Vryniotis, V. (2013, November 20). *Machine Learning Tutorial: The Max Entropy Text Classifier*. <https://blog.datumbox.com/machine-learning-tutorial-the-max-entropy-text-classifier/>
- Vuković, M., Pripuzić, K., & Belani, H. (2009). An Intelligent Automatic Hoax Detection System. In J. D. Velásquez, S. A. Ríos, R. J. Howlett, & L. C. Jain (Eds.), *Knowledge-Based and Intelligent Information and Engineering Systems* (Vol. 5711, pp. 318–325). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04595-0_39
- Wang, S., Wang, H., & Khalil, N. (2018). A Thematic Analysis of Interdisciplinary Journal of Information, Knowledge, and Management (IJIKM). *Interdisciplinary Journal of Information, Knowledge, and Management*, 13, 201–231.
<https://doi.org/10.28945/4095>
- Wang, W. Y. (2017). “Liar, Liar Pants on Fire”: A New Benchmark Dataset for Fake News Detection. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 422–426.
<https://doi.org/10.18653/v1/P17-2067>

- Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., & Gao, J. (2018). EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 849–857. <https://doi.org/10.1145/3219819.3219903>
- Wang, Y., Ma, F., & Jin, Z. (n.d.). *Yaqingwang / EANN-KDD18*. <https://github.com/yaqingwang/EANN-KDD18>
- Wanto, A., Windarto, A. P., Hartama, D., & Parlina, I. (2017). Use of Binary Sigmoid Function And Linear Identity In Artificial Neural Networks For Forecasting Population Density. *IJISTECH (International Journal of Information System and Technology)*, 1(1), 43–54. <https://doi.org/10.30645/ijistech.v1i1.6>
- Waterman, M. S., Smith, T. F., & Beyer, W. A. (1976). Some biological sequence metrics. *Advances in Mathematics*, 20(3), 367–387. [https://doi.org/10.1016/0001-8708\(76\)90202-4](https://doi.org/10.1016/0001-8708(76)90202-4)
- Watson, C. A. (2018). Information Literacy in a Fake/False News World: An Overview of the Characteristics of Fake News and its Historical Development. *International Journal of Legal Information*, 46(2), 93–96. <https://doi.org/10.1017/jli.2018.25>
- Weber, R. P. (1990). Content Analysis. In *Basic Content Analysis* (2nd ed., p. 8). Sage.
- Weiss, F., Espinoza, I., Hurtado, J., & Mendoza, M. (2019). Claim Behavior over Time in Twitter. In G. Meiselwitz (Ed.), *Social Computing and Social Media. Design, Human Behavior and Analytics* (Vol. 11578, pp. 468–479). Springer International Publishing. https://doi.org/10.1007/978-3-030-21902-4_33
- Welcome to polyglot's documentation! —Polyglot 16.07.04 documentation.* (n.d.). Retrieved September 15, 2022, from <https://polyglot.readthedocs.io/en/latest/>
- Wget—GNU Project—Free Software Foundation.* (n.d.). Retrieved September 10, 2022, from <https://www.gnu.org/software/wget/>
- What Is Logistic Regression? Equation, Assumptions, Types, and Best Practices |.* (n.d.). Retrieved November 9, 2022, from

<https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/>

Wiktionary:Main Page. (2022). In *Wikipedia, the free encyclopedia*.

https://en.wiktionary.org/w/index.php?title=Wiktionary:Main_Page&oldid=68627981

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05*, 347–354. <https://doi.org/10.3115/1220575.1220619>

WordNet. (2010). <https://wordnet.princeton.edu/>

Wu, W., Li, H., Wang, H., & Zhu, K. Q. (2012). Probase: A probabilistic taxonomy for text understanding. *Proceedings of the 2012 International Conference on Management of Data - SIGMOD '12*, 481. <https://doi.org/10.1145/2213836.2213891>

Xiaohui, Y. (2013). *Code of Biterm Topic Model*. <https://github.com/xiaohuiyan/BTM>

Xiaozhou, Y. (2022, January 27). *Linear Discriminant Analysis, Explained*. Medium. <https://towardsdatascience.com/linear-discriminant-analysis-explained-f88be6c1e00b>

Yamini. (2021, April 23). Random Forest—Ensemble method. *Geek Culture*. <https://medium.com/geekculture/random-forest-ensemble-method-860aaf4fcd16>

Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., & Liu, H. (2019). Unsupervised Fake News Detection on Social Media: A Generative Approach. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01), 5644–5651. <https://doi.org/10.1609/aaai.v33i01.33015644>

Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P. S. (2022). *TI-CNN: Convolutional Neural Networks for Fake News Detection* (arXiv:1806.00749). arXiv. <http://arxiv.org/abs/1806.00749>

- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2020). *XLNet: Generalized Autoregressive Pretraining for Language Understanding* (arXiv:1906.08237). arXiv. <http://arxiv.org/abs/1906.08237>
- Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical Attention Networks for Document Classification. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1480–1489. <https://doi.org/10.18653/v1/N16-1174>
- Yao, L., Mao, C., & Luo, Y. (2019). Graph Convolutional Networks for Text Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33, 7370–7377. <https://doi.org/10.1609/aaai.v33i01.33017370>
- Ying, L., Yu, H., Wang, J., Ji, Y., & Qian, S. (2021). Fake News Detection via Multi-Modal Topic Memory Network. *IEEE Access*, 9, 132818–132829. <https://doi.org/10.1109/ACCESS.2021.3113981>
- Young, J. C., Boyd, B., Yefimova, K., Wedlake, S., Coward, C., & Hapel, R. (2021). The role of libraries in misinformation programming: A research agenda. *Journal of Librarianship and Information Science*, 53(4), 539–550. <https://doi.org/10.1177/0961000620966650>
- Yu, F., Liu, Q., Wu, S., Wang, L., & Tan, T. (2017). A Convolutional Approach for Misinformation Identification. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 3901–3907. <https://doi.org/10.24963/ijcai.2017/545>
- Yu, Y., Yu, G., Li, T., Man, Q., & Chen, Q. (2020). Quantitative Characterization and Identification of the Company-Related Disinformation Channel Among Media. *IEEE Access*, 8, 29196–29204. <https://doi.org/10.1109/ACCESS.2020.2971727>
- Yuliani, S., Faizal, M., Sahib, S., & Supriadi, Y. (2019). A Framework for Hoax News Detection and Analyzer used Rule-based Methods. *International Journal of*

- Advanced Computer Science and Applications*, 10(10).
<https://doi.org/10.14569/IJACSA.2019.0101055>
- Zhang, J., Dong, B., & Yu, P. S. (2019). *FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network* (arXiv:1805.08751). arXiv.
<http://arxiv.org/abs/1805.08751>
- Zhang, Q., Lipani, A., Liang, S., & Yilmaz, E. (2019). Reply-Aided Detection of Misinformation via Bayesian Deep Learning. *The World Wide Web Conference on - WWW '19*, 2333–2343. <https://doi.org/10.1145/3308558.3313718>
- Zhang, Y., & Wallace, B. (2016). *A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification* (arXiv:1510.03820). arXiv. <http://arxiv.org/abs/1510.03820>
- Zhang, Z. (2018). Improved Adam Optimizer for Deep Neural Networks. *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, 1–2.
<https://doi.org/10.1109/IWQoS.2018.8624183>
- Zhang, Z. (2019, August 7). *Boosting Algorithms Explained*. Medium.
<https://towardsdatascience.com/boosting-algorithms-explained-d38f56ef3f30>
- Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid Scene Parsing Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6230–6239.
<https://doi.org/10.1109/CVPR.2017.660>
- Zhao, L., Wang, J., Chen, Y., Wang, Q., Cheng, J., & Cui, H. (2012). SIHR rumor spreading model in social networks. *Physica A: Statistical Mechanics and Its Applications*, 391(7), 2444–2453. <https://doi.org/10.1016/j.physa.2011.12.008>
- Zhao, W., Hershovich, D., Wang, L., & Qianming, X. (n.d.). *Text Classification with Capsule Network*. https://github.com/andyweizhao/capsule_text_classification
- Zhong, B., Xing, X., Love, P., Wang, X., & Luo, H. (2019). Convolutional neural network: Deep learning-based classification of building quality problems. *Advanced Engineering Informatics*, 40, 46–57. <https://doi.org/10.1016/j.aei.2019.02.009>

- Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. (2004). Automating Linguistics-Based Cues for Detecting Deception in Text-Based Asynchronous Computer-Mediated Communications. *Group Decision and Negotiation*, 13(1), 81–106.
<https://doi.org/10.1023/B:GRUP.0000011944.62889.6f>
- Zhou, L., Twitchell, D., Qin, T., Burgoon, J. K., & Nunamaker, J. F. Jr. (2003). An Exploratory Study into Deception Detection in Text-based Computer-Mediated Communication. *Proceedings of the 36th Annual Hawaii International Conference on System Sciences: 6 - 9 January 2003, Big Island, Hawaii ; Abstracts and CD-ROM of Full Papers ; [HICSS-36]*, 1–10. <https://doi.org/10.1109/HICSS.2003.1173793>
- Zhou, X., Wu, J., & Zafarani, R. (2020). *SAFE: Similarity-Aware Multi-Modal Fake News Detection* (arXiv:2003.04981). arXiv. <http://arxiv.org/abs/2003.04981>
- Zhou, X., & Zafarani, R. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys*, 53(5), 1–40.
<https://doi.org/10.1145/3395046>
- Zuiderwijk, A., Shinde, R., & Jeng, W. (2020). What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. *PLOS ONE*, 15(9), e0239283.
<https://doi.org/10.1371/journal.pone.0239283>
- Zvornicanin, E. (2022, February 5). *Differences Between Bidirectional and Unidirectional LSTM*. <https://www.baeldung.com/cs/bidirectional-vs-unidirectional-lstm>
- Kochkina, E. (n.d.). *Multitask4Veracity*.
<https://github.com/kochkinaelena/Multitask4Veracity>
- Κυριαζή, Ν. (2009). *Η κοινωνιολογική έρευνα: Κριτική επισκόπηση των μεθόδων και των τεχνικών* (2η εκδ). Ελληνικά Γράμματα.
- Ο Έλληνας αστυνομικός που “εκπαίδευσε” αλγόριθμους για τα fake news της πανδημίας. (n.d.). Retrieved September 18, 2022, from <https://www.news247.gr/sunday->

edition/o-ellinas-astynomikos-poy-ekpaideyse-algorithmoy-gia-ta-fake-news-tis-pandimias.9656566.html

Τζάνη, Μ., & Κεχαγιάς, Χ. (2005). *ΣΗΜΕΙΩΣΕΙΣ ΓΙΑ ΤΟ ΜΑΘΗΜΑ «ΜΕΘΟΔΟΛΟΓΙΑ ΕΡΕΥΝΑΣ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ»*. Εθνικό Καποδιστριακό Πανεπιστήμιο Αθηνών.

Παράρτημα Α. Πίνακας Ακρωνυμίων

ABBREVIATIONS	ΑΚΡΩΝΥΜΙΑ
ANN	SIMPLE NEURAL NETWORKS
API	APPLICATION INTERFACE
ATTRNN	RECURRENT NEURAL NETWORK WITH ATTENTION
BERT	BIDIRECTIONAL ENCODING REPRESENTATIONS FROM TRANSFORMERS
BILSTM	BI DIRECTIONAL LONG SHORT TERM MEMORY
BIGRU	BIDIRECTIONAL GATED RECURRENT UNIT
BIRNN	BI DIRECTIONAL RECURRENT NEURAL NETWORK
CART	CLASSIFICATION AND REGRESSION TREES
CNN	CONVOLUTIONAL NEURAL NETWORK
CSI	CAPTURE SCORE INTEGRATE
DAN	DEEP AVERAGE NETWORK
DAX	DIRECT ACYCLIC GRAPH
DBERTML	DISTILBERT MULTILINGUAL MODEL
DEFEND	EXPLAINABLE FAKE NEWS DETECTION
DNN	DENSE NEURAL NETWORK
DNN	DEEP NEURAL NETWORK

DT	DECISION TREE
EANN	EVENT ADVERSARIAL NEURAL NETWORK
FFNN	FEED FORWARD NEURAL NETWORKS
FNDNET	FAKE NEWS DETECTION NETWORK
GAT	GRAPH ATTENTION NETWORK
GCN	GRAPH CONVOLUTIONAL NETWORK
GRU	GATED RECURRENT UNIT
HAN	HIERARCHICAL ATTENTION NETWORK
HITS	HYPERLINK INDUCED TOPIC SEARCH
HPA-BLSTM	HIERARCHICAL POST ATTENTION BIDIRECTIONAL LONG SHORT TERM MEMORY
IFLA	INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATION
INVID	IN VIDEO VERITAS
KMAGCN	KNOWLEDGE AWARE MULTIMODAL GRAPH CONVOLUTIONAL NETWORK
KNN	K NEAREST NEIGHBOUR
LDA	LINEAR DISCRIMINANT ANALYSIS
LIWC	LINGUISTIC INQUIRY AND WORD COUNT
LSTM	LONG SHORT TERM MEMORY
MLP	MULTILAYER PERCEPTRON
MNB	MULTINOMIAL NAIVE BAYES

MPQA	MULTI-PERSPECTIVE QUESTION ANSWERING
MTMN	MULTIMODAL TOPIC MEMORY NETWORK
MVAE	MULTIMODAL VARIATIONAL AUTOENCODER
NELA	NEWS LANDSCAPE
NER	NAMED ENTITY RECOGNITION
NLP	NATURAL LANGUAGE PROCESSING
NLTK	NATURAL LANGUAGE TOOLKIT
NMF	NON NEGATIVE MATRIX FACTORIZATION
NPOV	NEUTRAL POINT OF VIEW
NRC	NATIONAL RESEARCH COUNCIL CANADA
OCR	OPTICAL CHARACTER RECOGNITION
OOV	OUT OF VOCABULARY
POS	PART OF SPEECH
PROV-O	PROVENANCE ONTOLOGY
PTVA	PINTO THIRAN VETTERLI
QDA	QUADRATIC DISCRIMINANT ANALYSIS
RELU	RECTIFIED LINEAR UNIT
RESNET	RESIDUAL NETWORK
RNN	RECURRENT NEURAL NETWORK

RvNN	RECURSIVE NEURAL NETWORK
SAFE	SIMILARITY AWARE FAKE (NEWS DETECTION)
SGD	STOCHASTIC GRADIENT DESCENT
SIR	SUSCEPTIBLE INFECTED RECOVERED
SVM	SUPPORT VECTOR MACHINE
TCNN-URG	TEXT CONVOLUTIONAL NEURAL NETWORK USER RESPONSE GENERATOR
TF-IDF	TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY
TRNN	TOP DOWN RECURSIVE NEURAL NETWORK
URL	UNIFORM RESOURCE LOCATOR
USE	UNIVERSAL SENTENCE ENCODER
VAD	VALENCE AROUSAL DOMINANCE
VGG	VISUAL GEOMETRY GROUP
XLNET	TRANSFORMER (-XL) NETWORK
XML	EXTENSIBLE MARKUP LANGUAGE

Παράρτημα Β. Πίνακας χαρακτηριστικών εντοπισμού ψευδών ειδήσεων

Ο συγκεκριμένος πίνακας περιέχεται στο αρχείο Word με τίτλο “Μέθοδοι Ελέγχου Γεγονότων» το οποίο επισυνάπτεται στην παρούσα διπλωματική εργασία.

Παράρτημα Γ. Αποτελέσματα Web of Science και Scopus

Τα αποτελέσματα από τις ερευνητικές βάσεις δεδομένων Web of Science και Scopus περιέχονται στο αρχείο MSEXcel με τίτλο totally merged το οποίο επισυνάπτεται στην παρούσα διπλωματική εργασία.

Παράρτημα Δ. Συμπληρωματικές πηγές με τη μέθοδο snowball

Ο συγκεκριμένος πίνακας περιέχεται στο αρχείο Word με τίτλο «snowball» το οποίο επισυνάπτεται στην παρούσα διπλωματική εργασία.