

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

«ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΙΣΗ ΚΑΙ ΣΥΓΚΡΙΤΙΚΗ ΑΝΑΛΥΣΗ ΤΕΧΝΙΚΩΝ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΣΤΗΝ ΑΝΙΧΝΕΥΣΗ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ »

ΝΑΝΟΣ ΜΑΡΙΟΣ

Π2012051

Επιτηρητής: Άννα Σωτηροπούλου

Σεπτέμβριος 2024

Επιβλέπουσα

Άννα Σωτηροπούλου, ΕΔΙΠ
Ιόνιο Πανεπιστήμιο, Τμήμα Πληροφορικής

Τριμελής Επιτροπή

Άννα Σωτηροπούλου, ΕΔΙΠ
Ιόνιο Πανεπιστήμιο, Τμήμα Πληροφορικής

Ανδρέας Καναβός, Αναπληρωτής Καθηγητής
Ιόνιο Πανεπιστήμιο, Τμήμα Πληροφορικής

Θεμιστοκλής Έξαρχος, Αναπληρωτής Καθηγητής
Ιόνιο Πανεπιστήμιο, Τμήμα Πληροφορικής

ΕΙΣΑΓΩΓΗ

ΠΕΡΙΛΗΨΗ

Ο πολλαπλασιασμός και η διάδοση ψευδών ειδήσεων στα ψηφιακά μέσα αποτελεί ένα καθημερινό κοινωνικό πρόβλημα. Τα νέα διαδίδονται πλέον με μεγάλες ταχύτητες και ποσότητες μέσα από τα κοινωνικά δίκτυα και άλλες ψηφιακές πλατφόρμες δυσκολεύοντας ολοένα και περισσότερο την ανίχνευση τους. Σε αυτή τη πτυχιακή εργασία στοχεύω να κάνω μια εκτεταμένη ανασκόπηση και συγκριτική ανάλυση αναμεσά στις τεχνικές εξόρυξης δεδομένων που χρησιμοποιούνται για την αναγνώριση τέτοιου είδους ειδήσεων. Θα παρουσιάσω διάφορες μεθόδους από τεχνικές ταξινόμησης όπως decision trees, support vector machines, τεχνικές clustering όπως K-means μέχρι και τα πιο προηγμένα Machine Learning, Deep Learning και Neural Networking μοντέλα. Επιπλέον, θα δούμε μοντέλα επεξεργασίας φυσικής γλώσσας όπως το BERT. Τέλος, μέσα από δοκιμές και πειράματα που πραγματοποιήθηκαν, θα δούμε τις αδυναμίες και τα πλεονεκτήματα διάφορων μοντέλων και θα συγκρίνουμε τις επιδόσεις τους μεταξύ τους. Η πτυχιακή αυτή έχει ως στόχο να συμβάλει στη καταπολέμηση του προβλήματος, κάνοντας μια λεπτομερειακή ανασκόπηση των τεχνικών που χρησιμοποιούνται και συγκρίνοντας τους με σκοπό την ανάδειξη των καλύτερων μεθόδων.

Περιεχόμενα

ΠΕΡΙΛΗΨΗ	4
Περιεχόμενα.....	5
Κεφάλαιο 1: Εισαγωγή.....	7
1.1 Ερευνητικό Υπόβαθρο	7
1.2 Εγχώρια και διεθνής έρευνα	8
1.2.1 Εγχώρια Έρευνα	8
1.2.2 Διεθνής Έρευνα	9
1.3 Δομή της πτυχιακής.....	10
Κεφάλαιο 2: Θεμελιώδης Γνώσεις.....	11
2.1 Fake News: Ορισμός και χαρακτηριστικά.....	11
2.1.1 Κατηγορίες των Fake News	11
2.1.2 Η επίδραση των Fake News	12
2.1.3 Η επιρροή των Fake News στη ψυχολογία και στη κοινωνία	13
2.2 Τεχνικές Εξόρυξης Δεδομένων	14
2.2.1 Τεχνικές Κατηγοριοποίησης Κειμένου.....	14
2.2.2 Τεχνικές ομαδοποίησης	15
2.2.3 Κανόνες Συσχέτισης	15
2.3 Μηχανική Μάθηση.....	16
2.3.1 Βαθιά Μάθηση.....	17
2.4 Επεξεργασία Φυσικής Γλώσσας (NLP)	17
2.4.1 Αναπαράσταση κειμένου.....	17
2.4.2 Προηγμένα NLP Μοντέλα - BERT.....	19
2.5 Υβριδικά Μοντέλα	19
2.5.1 NLP και Ανάλυση Δικτύου.....	19
2.5.2 Υβριδικά Μοντέλα Μηχανικής Μάθησης.....	20
2.5.3 Υβριδικά Μοντέλα με Πολυμέσα	20
2.5.4 Προκλήσεις και Μελλοντική Ανάπτυξη.....	21
Κεφάλαιο 3: Αρχιτεκτονική Συστημάτων Ανίχνευσης και Μεθοδολογία	22
3.1 Αρχιτεκτονική συστήματος	23
3.2 Τεχνικές Εξαγωγής Χαρακτηριστικών.....	28

3.2.1 Εξαγωγή Χαρακτηριστικών από Πολυμέσα	29
3.2.2 Εξαγωγή Χαρακτηριστικών από Δίκτυα	30
3.3 Εφαρμογές σε πραγματικό χρόνο	31
Κεφάλαιο 4: Πειραματικοί Μέθοδοι και Αποτελέσματα	33
4.1 Εισαγωγή.....	33
4.2 Πειραματική Διαδικασία	33
4.2.1 Σύνολο Δεδομένων	33
4.2.2 Προεπεξεργασία και εξαγωγή χαρακτηριστικών.....	34
4.2.3 Cross-Validation	35
4.2.4 Μετρικοί Δείκτες Αξιολόγησης	35
4.3 Πειραματική Διαδικασία	35
4.3.1 Προετοιμασία Δεδομένων.....	35
4.3.2 Εκπαίδευση Μοντέλων	36
4.3.3 Διαχείριση Υπολογιστικών Πόρων	36
4.4 Ανάλυση αποτελεσμάτων.....	37
4.4.1 Traditional Models Performance	37
4.4.2 BERT	39
4.4.3 Συγκριτική ανάλυση	39
4.5 Συγκριτική ανάλυση με υπάρχουσα βιβλιογραφία	41
4.5.1 Μοντελα Μηχανικής Μάθησης	41
4.5.2 Μοντέλα Βαθιάς Μάθησης.....	43
Κεφάλαιο 5: Συμπεράσματα και Μελλοντική Έρευνα	45
5.1 Περίληψη Αποτελεσμάτων	45
5.2 Συνεισφορές.....	45
5.3 Μελλοντική Έρευνα	46
5.4 Τελικά Συμπεράσματα.....	46
Βιβλιογραφία	47

Κεφάλαιο 1: Εισαγωγή

1.1 Ερευνητικό Υπόβαθρο

Τα τελευταία χρόνια έχει αλλάξει ολοκληρωτικά ο τρόπος που οι άνθρωποι «καταναλώνουν» ειδήσεις. Η πληροφορία σήμερα μεταδίδεται με αστραπιαίες ταχύτητες και η συνεχής εξέλιξη των μέσων κοινωνικής δικτύωσης συμβάλει δραστικά στη μετάδοση των ειδήσεων. Χάρη στη ψηφιοποίηση, έχει παρατηρηθεί τεράστια αύξηση στον αριθμό των δημοσιογραφικών άρθρων, και στις ιστοσελίδες ειδήσεων που συνεχώς αυξάνονται. Ωστόσο η ψηφιοποίηση έχει βοηθήσει σημαντικά και στη διατήρηση και εξάπλωση των Fake News. Ο όρος Fake News αναφέρεται στη διάδοση ψευδών ειδήσεων και παραπλανητικού υλικού μεταμφιεσμένου σαν έγκυρες ειδήσεις. Καλύπτει ένα πολύ μεγάλο φάσμα, από εντελώς φτιαχτές ιστορίες μέχρι παραποιημένα γεγονότα, με σκοπό κυρίως να παραπλανήσει τους αναγνώστες-ακροατές.

Το φαινόμενο αυτό μπορεί να έχει καταστροφικές επιπτώσεις τόσο σε ατομικό επίπεδο όσο και σε συλλογικό και κοινωνικό επίπεδο. Οι ειδήσεις έχουν τη δύναμη να επηρεάσουν τις απόψεις, τα πιστεύω και τη συμπεριφορά των ανθρώπων και κατ' επέκταση, τις κοινωνικές αλληλεπιδράσεις και διαδικασίες. Πολλές φορές μάλιστα, έχει παρατηρηθεί πως κοινωνικές και δημοκρατικές διαδικασίες όπως ψηφοφορίες έχουν επηρεαστεί από δεδομένα που ήρθαν στο φως μέσω ειδήσεων τις περιόδους που λαμβάνανε μέρος. Η επιρροή αυτή ενισχύεται φυσικά ακόμα περισσότερο όταν μπαίνουν στο παιχνίδι της διάδοσης ειδήσεων τα μέσα κοινωνικής δικτύωσης (social media). Οι εταιρίες που έχουν τα μέσα αυτά, χρησιμοποιούν αλγορίθμους που καθορίζουν τη προτεραιότητα και το περιεχόμενο που εμφανίζεται στο κάθε χρήστη. Οι αλγόριθμοι αυτοί δίνουν προτεραιότητα σε υλικό ανάλογο με τις αλληλεπιδράσεις που προκάλεσε, δηλαδή ανάλογο με τον αριθμό σε σχόλια, likes και shares. Είναι πλέον γνωστό πως όσο πιο προκλητική και ενοχλητική είναι μια πληροφορία τόσο περισσότερη είναι η αλληλεπίδραση που υπάρχει γύρω της, με αποτέλεσμα να αποκτά τεράστια δύναμη στη προτεραιότητα εμφάνισης της. Έτσι λοιπόν, μέσω αυτών των αλγορίθμων, ενισχύεται σημαντικά η διάδοση παραπλανητικού υλικού και δυσκολεύει ακόμα πιο πολύ στον περιορισμό του όταν το υλικό έχει γίνει εύρη διαδεδομένο γνωστό και ως “viral”.

Για την αντιμετώπιση της μάστιγας των Fake News, διάφορες τεχνικές αναγνώρισης παραπλανητικού και ψευδούς υλικού έχουν αναπτυχθεί από την επιστημονική κοινότητα. Αξιοποιώντας τεχνολογίες εξόρυξης δεδομένων, έχουν καταφέρει να αναλύσουν τεράστιο όγκο δεδομένων ειδήσεων με σκοπό να αναγνωρίσουν και να καταγράψουν τις τάσεις και τα χαρακτηριστικά που υποδεικνύουν την ύπαρξη ψευδών ειδήσεων. Η μεμονωμένη αλλά και συνδυαστική χρήση του Natural Language Process μαζί με Machine Learning έχει αποδειχθεί σημαντικό εργαλείο και έχει συμβάλει στη ακρίβεια των συστημάτων αναγνώρισης. Το NLP συνεισφέρει στην εξέταση και αναγνώριση της γραπτής πληροφορίας, βρίσκοντας γλωσσικά στοιχεία ύποπτα ως παραπλανητικό υλικό ενώ οι ML αλγόριθμοι συνεισφέρουν στη κατηγοριοποίηση αυτού του υλικού με βάση τα χαρακτηριστικά τους. Τα τελευταία χρόνια, κυρίως μετρά τη περίοδο του COVID-19, έχουν αναπτυχθεί και οπλιστεί ακόμη πιο ανεπτυγμένα μοντέλα Deep Learning και Neural Networks τα οποία έχουν δείξει πολύ ενθαρρυντικά αποτελέσματα στη δυνατότητα τους να αναγνωρίζουν με επιτυχία περίπλοκα μοτίβα και νοήματα μέσα από συμπραζόμενα.

Με τη χρήση υβριδικών μοντέλων που συνδυάζουν τα δυνατά σημεία του NLP, ML, DL και Neural networks, οι ειδικοί έχουν καταφέρει να αναπτύξουν ισχυρούς αλγορίθμους που μπορούν να χρησιμοποιηθούν σε διάφορες μορφές ειδήσεων. Στη πτυχιακή αυτή θα δούμε όλες αυτές τις τεχνικές καθώς και τις διαφορές που έχουν στις επιδόσεις μεταξύ τους.

1.2 Εγχώρια και διεθνής έρευνα

1.2.1 Εγχώρια Έρευνα

Στην Ελλάδα, το θέμα των Fake News έχει απασχολήσει αρκετά τα τελευταία χρόνια. Συγκεκριμένα τη τελευταία δεκαετία, έρευνες έχουν δείξει μεγάλη αύξηση στις παραπλανητικές ειδήσεις τόσο τηλεοπτικών όσο και διαδικτυακών μέσων. Οι οικονομικές δυσκολίες και πολιτικές ανευθυνότητες έχουν συμβάλει στην αύξηση των ειδήσεων αυτών σημαντικά. Τη τελευταία πενταετία έχουν γίνει κάποια σημαντικά βήματα στην ανάπτυξη NLP αλγορίθμων προσαρμοσμένων στην ελληνική γλώσσα και κουλτούρα. Μια ερευνητική προσέγγιση που απέδωσε καρπούς και φάνηκε υποσχόμενη ήταν η εκπαίδευση του προηγμένου μοντέλου μηχανικής μάθησης, BERT, στην ελληνική γλώσσα. Για να το καταφέρουν αυτό, οι ειδικοί συλλέξαν και δημιούργησαν ένα μεγάλο σύνολο δεδομένων από κείμενα στα ελληνικά, προσαρμοσμένα ειδικά για το σκοπό αυτό, κατασκευάζοντας έτσι το GREEK-BERT μοντέλο. Το συγκεκριμένο μοντέλο δοκιμαστικέ στην αναγνώριση ψευδών ειδήσεων στα social media σημειώνοντας σημαντική επιτυχία,

Πίνακας 1: Επίδοση μοντέλου GREEK-BERT

<i>Model</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
<i>GREEK-BERT News Model</i>	0.930	0.930	0.930	0.930
<i>GREEK-BERT FakeTweets Model</i>	0.910	0.900	0.900	0.900

1.2.2 Διεθνής Έρευνα

Σε παγκόσμιο επίπεδο, γίνεται πολύ μεγάλη προσπάθεια για την αντιμετώπιση του φαινομένου. Διάφοροι κρατικοί φορείς και επιστημονικά ινστιτούτα χρηματοδοτούν και υποστηρίζουν την ερευνά στην ανάπτυξη εργαλείων για την ανίχνευση των Fake News. Οι ΗΠΑ, που αντιμετωπίζουν ένα εκτεταμένο πρόβλημα παραπλανητικών νέων, έχουν κινητοποιηθεί από αυτό, καθώς επηρεάζει με εμφανές τρόπο τη ζωή και τις δημοκρατικές διαδικασίες. Οι έρευνες εστιάζουν κυρίως σε ανάπτυξη υβριδικών μοντέλων που συνδυάζουν αλγορίθμους μηχανικής μάθησης (ML) με αλγορίθμους επεξεργασίας φυσικής γλώσσας (NLP) για την αναγνώριση ψεύτικων ειδήσεων. Παράλληλα επιχειρούν να συνδυάσουν Νευρωνικά δίκτυα όπως συνεκτικά νευρωνικά δίκτυα (CNNs) και επαναλαμβανόμενα νευρωνικά δίκτυα (RNNs), για την αναγνώριση μοτίβων. Τέλος καταβάλλεται μεγάλη προσπάθεια εφαρμογής τέτοιων μοντέλων στην ανίχνευση Fake News κοινοποιήσεων στα social media καθώς εκεί έχει παρατηρηθεί η μεγαλύτερη εμφάνιση τους.

Σε ευρωπαϊκό επίπεδο, η επιστημονική κοινότητα έχει στραφεί στην ανάπτυξη αλγορίθμων που μπορούν να εκπαιδευτούν και να λειτουργήσουν σε διαφορετικές γλώσσες. Στην Ευρώπη, εμφανίζονται πολλές γλώσσες και κουλτούρες και αυτό επισημαίνει την ανάγκη για μοντέλα που μπορούν να προσαρμοστούν. Γι' το λόγο αυτό, η Ευρωπαϊκή Ένωση παρείχε χρηματοδότηση, με τη πρωτοβουλία του Horizon 2020, για την ανάπτυξη πολυγλώσσων NLP μοντέλων αναγνώρισης ψευδών ειδήσεων. Σημαντική πρόοδος έχει παρουσιάσει και η Κίνα συνδυάζοντας τεχνίτη νοημοσύνη με συμβατικές δημοσιογραφικές μεθόδους, ενώ στην Ινδία η ερευνά επικεντρώνεται κυρίως γύρω από το Deep Learning για τον έλεγχο εκτεταμένου όγκου δεδομένων.

Τέλος, έχει σημειωθεί μεγάλη πρόοδος μέσω διεθνής συνεργασίας. Το Global Disinformation Index έχει φέρει κοντά επιστήμονες από διάφορα έθνη με σκοπό την ανάπτυξη μιας βάσης μοντέλου που θα μπορεί να προσαρμοστεί σε διάφορες ανάγκες που έχει το κάθε έθνος.

Η μεγάλη κινητικότητα που παρατηρούμε στη παγκόσμια επιστημονική κοινότητα και τα έθνη για την αντιμετώπιση του φαινομένου, αποδεικνύει την επικινδυνότητα του προβλήματος. Η ανάπτυξη υβριδικών μοντέλων που μπορούν να προσαρμοστούν σε διαφορετικές μορφές ειδήσεων αλλά και διαφορετικές γλώσσες και κουλτούρες αποτελούν σημαντική προτεραιότητα.

1.3 Δομή της πτυχιακής

Σε αυτή την ερευνά στοχεύω να πραγματοποιήσω μια ανάλυση όλων των τεχνικών ανίχνευσης Fake News. Για το σκοπό αυτό έχει γίνει ανασκόπηση της υπάρχουσας σχετικής βιβλιογραφίας από το 2015 μέχρι το 2024. Με αυτό το τρόπο θα γίνει μια εκτίμηση της τωρινής κατάστασης στο θέμα της αναγνώρισης ψευδών ειδήσεων και θα παρουσιαστεί μια περιεκτική σύγκριση των προτεινομένων μοντέλων. Παρακάτω ακολουθεί μια σύντομη περιγραφή της δομής της παρούσας εργασίας.

Στην εργασία αυτή θα εξετάσω την αναγνώριση και κατηγοριοποίηση των παραπλανητικών ειδήσεων τόσο από θεωρητική όσο και από πρακτική σκοπιά. Συνολικά πέντε κεφάλαια θα καλύψουν την ερευνά. Στο κεφάλαιο της εισαγωγής είδαμε σύντομα πως η διάδοση των ειδήσεων μπορεί να επηρεάσει τις ζωές μας και πως αυτό το φαινόμενο αποτελεί μεγάλο κίνδυνο σε ατομικό και συλλογικό επίπεδο.

Στο δεύτερο κεφάλαιο, «Θεμελιώδης Γνώσεις», θα δούμε από θεωρητική σκοπιά τους τρόπους που εφαρμόζονται για την αντιμετώπιση του φαινομένου. Αρχικά θα δοθεί ένας ορισμός καθώς και κάποιες μορφές Fake News για την καλύτερη κατανόηση του προβλήματος. Στη συνέχεια θα παρουσιάσω τις διάφορες τεχνικές ανά κατηγορία: εξόρυξη δεδομένων, μηχανική μάθηση, επεξεργασία φυσικής γλώσσας και υβριδικές εκδοχές.

Στο τρίτο κεφάλαιο, «Αρχιτεκτονική Συστημάτων Ανίχνευσης και Μεθοδολογία», θα περιγράψω την αρχιτεκτονική ενός συστήματος ανίχνευσης. Εκεί θα δούμε το σχεδιασμό ενός τέτοιου συστήματος βήμα προς βήμα και θα δώσουμε έμφαση στις τεχνικές αποκοπής και προεπεξεργασίας πληροφοριών και δεδομένων.

«Πειραματικοί Μέθοδοι και Αποτελέσματα», το τέταρτο κεφάλαιο της πτυχιακής, περιγράφει το πως μπορεί να στηθεί ένα πείραμα εκπαίδευσης και αξιολόγησης μοντέλων αναγνώρισης, παρουσιάζοντας τα πειράματα που πραγματοποιήθηκαν από το συντάκτη αλλά και από τις υπάρχον έρευνες με σκοπό τη συγκριτική ανάλυση των αποτελεσμάτων. Θα αναφέρω τη διαδικασία που πρέπει να ακολουθήσει κάποιος για να διεξάγει ένα πείραμα τέτοιου τύπου καθώς και τι σημαίνουν τα διάφορα αποτελέσματα που μπορούμε να πάρουμε από την αξιολόγηση.

Τέλος, στο πέμπτο κεφάλαιο «Συμπεράσματα και Μελλοντική Έρευνα» γίνεται μια περίληψη των δεδομένων που συλλέχθηκαν για την εργασία, των περιορισμών που υπάρχουν αυτή τη στιγμή αλλά και ιδέες για μελλοντική ερευνά.

Κεφάλαιο 2: Θεμελιώδης Γνώσεις

2.1 Fake News: Ορισμός και χαρακτηριστικά

Δεν έχει υπάρξει μέχρι στιγμής κάποιος παγκόσμιος ορισμός για το τι είναι ‘Fake News’. Θα προσπαθήσω να δώσω έναν ορισμό για να καταλάβουμε καλύτερα την έννοια του. Με τον ορό ‘Fake News’ αναφερόμαστε στην μετάδοση παραπλανητικού υλικού το οποίο παρουσιάζεται σαν έγκυρες ειδήσεις. Εμφανίζεται σε διάφορες μορφές, από εξ ολοκλήρου φτιαχτές ιστορίες έως παραποιημένα γεγονότα, συνήθως με σκοπό να παραπλανήσει το κοινό. Μπορεί να έχει γραπτή (τίτλος ή άρθρο), οπτικοακουστική μορφή (ηχητικό υλικό, φωτογραφίες, βίντεο κλπ.) ή ένα συνδυασμό αυτών όπως συνηθίζουμε να παρατηρούμε.

2.1.1 Κατηγορίες των Fake News

Η υπάρχουσα λογοτεχνία γύρω από το θέμα των ‘Fake News’ συνήθως συνδυάζει τον ορό αυτόν με παραπλανητικές ειδήσεις (deceptive news), ψευδείς ειδήσεις (false news), παραπληροφόρηση (disinformation & misinformation), cherry-picking, δόλωμα (clickbait) και φήμες (rumors). Στις πρώτες δυο κατηγορίες, το περιεχόμενο είναι εξ’ ολοκλήρου ή μερικώς ψεύτικο και έχει σκοπό να παραπλανήσει, για οικονομικούς, πολιτικούς ή κοινωνικούς σκοπούς. Η παραπληροφόρηση περιλαμβάνει παραποιημένες ειδήσεις με τη μορφή άρθρων ή οπτικοακουστικού υλικού με σκοπό και εκεί την παραπλάνηση ενώ η σάτιρα, αν και δεν έχει σκοπό να παραπλανήσει, συνήθως εκλαμβάνεται λανθασμένα ως γεγονότα. Το cherry-picking αποτελεί μια μορφή παραπληροφόρησης όπου οι πληροφορίες είναι τροποποιημένες ή μειωμένες με σκοπό να υποστηρίξουν μια συγκεκριμένη ιδεολογία, αφήνοντας κρυφές πληροφορίες που μπορεί να αντικρούσουν τα συμφέροντα αυτά. Το Clickbait είναι συνήθως ένας τίτλος ειδήσεων με μορφή συνδέσμου. Σε αυτή τη περίπτωση γίνεται προσπάθεια να παραπλανηθεί ο χρήστης και να πατήσει το σύνδεσμο ώστε να μεταφερθεί σε μια ύποπτη σελίδα. Τέλος οι φήμες είναι μια μορφή παραπληροφόρησης καθώς συνήθως αποτελούνται από ανακρίβειες.

Βασισμένοι σε αυτές τις κατηγορίες μπορούμε να κρίνουμε το περιεχόμενο με βάσει τρία κριτήρια: (i) Αυθεντικότητα (authenticity), (ii) Πρόθεση (intention), και (iii) το ενδεχόμενο του περιεχομένου να είναι ειδήσεις. Παρακάτω είναι μια περίληψη των κατηγοριών και κριτηρίων αυτών.

Πίνακας 2. Σύγκριση μεταξύ εννοιών που σχετίζονται με ψευδείς ειδήσεις

CONCEPT	AUTHENTICITY	INTENTION	NEWS
DECEPTIVE NEWS	Μη-έγκυρη	Παραπλάνηση	Ναι
FALSE NEWS	Μη-έγκυρη	Παραπλάνηση	Ναι
SATIRE NEWS	(Μη-ορισμένη)	Διασκέδαση	Ναι
DISINFORMATION	Μη-έγκυρη	Παραπλάνηση	(Μη-ορισμένη)
MISINFORMATION	Μη-έγκυρη	Ίσος Παραπλάνηση	(Μη-ορισμένη)
CHERRY-PICKING	Μερικώς έγκυρη	Παραπλάνηση	(Μη-ορισμένη)
CLICKBAIT	(Μη-ορισμένη)	Παραπλάνηση	(Μη-ορισμένη)
RUMOR	Μη-έγκυρη	Ίσος Παραπλάνηση	(Μη-ορισμένη)

2.1.2 Η επίδραση των Fake News

Οι ψευδείς ειδήσεις έχουν τεράστιες και καταστροφικές επιπτώσεις τόσο μεμονωμένα σε άτομα, όσο και συλλογικά στην κοινωνία. Μπορούν να βλάψουν πολλές πτυχές της κοινωνίας συμπεριλαμβανομένου την εμπιστοσύνη των ανθρώπων στη πολιτεία, τη δημοσία υγεία και τη κατάσταση της οικονομίας

Σχετικά με τη πολιτική, η διάδοση ψευδούς πληροφοριών έχει αποδειχθεί πως μπορεί να φέρει επικίνδυνες επιπτώσεις στην κοινωνία. Κατά τη διάρκεια των τελευταίων εκλογών στις ΗΠΑ η προπαγάνδα έπαιξε μεγάλο ρόλο τόσο στις εκλογές όσο και στην αντίδραση του κόσμου μετά το αποτέλεσμα. Μια από τις σημαντικότερες επιπτώσεις ήταν αλώςτε η επιδρομή των υποστηρικτών του Τραμπ στο Καπιτώλιο καθώς πίστεψαν πως το αποτέλεσμα των εκλογών είχε αλλοιωθεί. Έρευνες έχουν δείξει πως η παραπληροφόρηση οδηγεί σε δυσπιστία στη πολιτική ηγεσία και τις δημοκρατικές διαδικασίες.

Η διατάραξη της δημοσίας υγείας είναι ένα ακόμα πλήγμα. Κατά τη διάρκεια της επιδημίας COVID-19, παρατηρήθηκε τεράστια κινητικότητα ψεύτικων και παραπλανητικών στοιχείων σχετικά με τις θεραπείες και τα εμβόλια με αποτέλεσμα να δημιουργηθεί μεγάλη σύγχυση ανάμεσα στους ανθρώπους. Πολλοί αγνόησαν τα μέτρα και τους κανόνες που έθεσε ο παγκόσμιος οργανισμός υγείας με αποτέλεσμα σε πολλές περιπτώσεις να δημιουργηθούν κρίσεις. Μάλιστα η περίοδος αυτή αποτέλεσε σημείο εκκίνησης για πολλές έρευνες σχετικά με την ανίχνευση ψεύτικων στοιχείων σχετικών με το COVID-19.

Επιπτώσεις στην οικονομία έχουν επίσης παρατηρηθεί. Η διάδοση ανακριβή στοιχείων έχει οδηγήσει μετοχές εταιριών στην άνοδο και άλλων στην κάθοδο. Ένα παράδειγμα είναι η κατάρρευση των μετοχών της Metro Bank UK κατά 11% λόγω διάδοσης ψεύτικων στοιχείων σχετικά με την οικονομική της κατάσταση στα μέσα κοινωνικής δικτύωσης. Ένα άλλο παράδειγμα είναι η αύξηση της μετοχής του Twitter (X) όταν κυκλοφορήσαν φήμες για εξαγορά δισεκατομμυρίων.

Ένα ακόμα φαινόμενο που έχει παρατηρηθεί είναι ο κοινωνικός διαχωρισμός των ανθρώπων δημιουργώντας ‘echo chambers’. Κατά αυτό το φαινόμενο, οι χρήστες εκτίθενται μόνο σε υλικό που ταιριάζει με τις δικές τους απόψεις και ενδιαφέρονται. Με αυτό το τρόπο εντείνεται ο κοινωνικός διαχωρισμός και οι διακρίσεις. Επιπλέον, η καθημερινή εμφάνιση των Fake News έχει δημιουργήσει δυσπιστία κατά όλων των μέσων, με αποτέλεσμα να χάνουν την αξιοπιστία τους ακόμα και φορείς που δημοσιεύουν έγκυρες πληροφορίες.

Αυτές οι επιπτώσεις δείχνουν ποσό επικίνδυνη μπορεί να γίνει η διασπορά των Fake News και της παραπληροφόρησης.

2.1.3 Η επιρροή των Fake News στη ψυχολογία και στη κοινωνία

Ένα σημαντικό βήμα για την ανάπτυξη μοντέλων αντιμετώπισης των ‘Fake News’ είναι η κατανόηση των ψυχολογικών και κοινωνικών δυναμικών πίσω από τη διάδοση τέτοιου είδους ειδήσεων. Έρευνες έχουν δείξει ότι προκαταλήψεις όπως αυτή της προσωπικής επιβεβαίωσης παίζουν σημαντικό ρόλο στη διάδοση ψευδών ειδήσεων. Άτομα είναι ευάλωτα στο να αποδεχθούν και να μοιράσουν ψευδείς πληροφορίες που συμβαδίζουν με τις προσωπικές τους απόψεις. Επίσης άρθρα ψεύτικων ειδήσεων συνήθως κάνουν επίκληση στο συναίσθημα των αναγνωστών με σκοπό να προκαλέσει μεγαλύτερες αντιδράσεις, ωθώντας το κοινό να αλληλεπιδράσει με την είδηση (share, like, comment).

Η κοινωνική εικόνα συνεισφέρει επίσης στο πρόβλημα. Δεν πρέπει να παραβλέψουμε τη δύναμη που έχουν οι ‘influencers’, τα social networks και οι ομάδες ομότιμων στη διαμόρφωση απόψεων. Όταν μια «έμπιστη» πηγή μέσα στο κοινωνικό δίκτυο ενός ανθρώπου μοιράζεται μια πληροφορία ή άποψη, είναι ευκολότερο να υιοθετηθεί και να προωθηθεί. Τέλος η ανωνυμία του διαδικτύου είναι ακόμα ένας λόγος που συμβάλει στη διασπορά της παραπληροφόρησης. Οι χρήστες νιώθοντας ότι δεν θα έρθουν αντιμέτωποι με τις συνέπειες των πράξεων τους, μοιράζουν ή ακόμα και δημιουργούν ψεύτικες πληροφορίες, κάνοντας έτσι την ανίχνευση δυσκολότερη.

Εν κατακλείδι, οι ψευδείς ειδήσεις είναι ένα πολυδιάστατο φαινόμενο με πολλαπλές επιπτώσεις στη κοινωνία. Η ικανότητα να μεταπείσουν τα πιστεύω και τις απόψεις των ανθρώπων είναι ένας μεγάλος κίνδυνος καθώς μπορεί να χρησιμοποιηθεί για τη διάδοση μίσους και επικίνδυνων ιδεολογιών

Στη συνέχεια θα δούμε τις τεχνικές με τις οποίες μπορούμε να αναγνωρίσουμε και να κατηγοριοποιήσουμε ειδήσεις με βάση το περιεχόμενό τους και την αξιοπιστία τους.

2.2 Τεχνικές Εξόρυξης Δεδομένων

Η επιστήμη της εξόρυξης δεδομένων (data mining) αποτελεί τη ‘ραχοκοκαλιά’ από τα περισσότερα συστήματα ανίχνευσης. Η εξόρυξη δεδομένων μας επιτρέπει να επιλέξουμε πληροφορίες ανάμεσα από τεράστιο όγκο δεδομένων αναγνωρίζοντας μοτίβα και τάσεις. Αυτό είναι εξαιρετικά χρήσιμο καθώς υπάρχει πληθώρα ειδήσεων προς αναγνώριση. Σε αυτή την ενότητα θα δούμε κάποιες από τις σημαντικότερες τεχνικές που εφαρμόζονται στην ανίχνευση των Fake News

2.2.1 Τεχνικές Κατηγοριοποίησης Κειμένου

Η κατηγοριοποίηση (classification) είναι μια από τις πιο παραδοσιακές και πολυχρησιμοποιημένες τεχνικές εξόρυξης δεδομένων που εφαρμόζεται στην αναγνώριση παραπλανητικού υλικού. Με αυτό το τρόπο κατηγοριοποιούμε τα δεδομένα μας σε προκαθορισμένες κλάσεις ανάλογα με κάποια χαρακτηριστικά. Αυτά τα χαρακτηριστικά συνήθως είναι ετικέτες που φέρουν τα σετ δεδομένων σχετικά με τα άρθρα που περιλαμβάνουν και συνήθως είναι ‘real’ ή ‘fake’. Μοντέλα που έχουν εκπαιδευτεί με αυτό το τρόπο μπορούν να κρίνουν άρθρα που δεν έχουν δει ξανά.

Μια από τις πιο διαδομένες classification τεχνικές είναι το Δέντρο Αποφάσεων (*Decision Tree*). Ένα μοντέλο δέντρου παίρνει αποφάσεις ανάλογες με τις πληροφορίες εισόδου που δέχτηκε. Αυτό οδηγεί σε ένα δενδροτό σχήμα όπου κάθε κόμβος αναπαριστά ένα χαρακτηριστικό, κάθε κλαδί μια απόφαση, και κάθε φύλο αναπαριστά ένα αποτέλεσμα (real ή fake). Ωστόσο αυτή η τεχνική είναι επιρρεπής σε overfitting, όταν έχει να κάνει με περιπλοκά δεδομένα. Overfitting συμβαίνει όταν ένας αλγόριθμος ταιριάζει σχεδόν ή εντελώς με τα δεδομένα της εκπαίδευσης του με αποτέλεσμα να μην μπορεί να κάνει έγκυρες προβλέψεις.

Ένας τρόπος να παρακάμψουμε αυτό το πρόβλημα είναι η τεχνική του Random Forests. Αυτή η τεχνική συνδυάζει πολλά Decision Trees για να βελτιώσει τις προβλέψεις. Αυτό είναι εφικτό καθώς τα Random Forests εκμεταλλεύονται το μέσο όρο των προβλέψεων των διαφόρων δέντρων μετριάζοντας έτσι το κίνδυνο υπερβολικής προσαρμογής. Το πετυχαίνει αυτό λόγω του νόμου των μεγάλων αριθμών και με τον τρόπο αυτό μπορούν να παρέχουν ισχυρές προβλέψεις. Με τη κατάλληλη μορφή τυχαιότητας μπορούμε να κάνουμε ταξινομήσεις με περισσότερη ακρίβεια και παλινδρόμηση. Χρησιμοποιώντας εκτιμήσεις out-of-bag μπορούν να μετατρέψουν σε βάσιμες τιμές τις κατά τα άλλα θεωρητικές τιμές των αποτελεσμάτων.

2.2.2 Τεχνικές ομαδοποίησης

Η ομαδοποίηση (clustering) είναι μια μη εποπτευομένη τεχνική εξόρυξης δεδομένων. Η χρησιμότητά της αφορά την ομαδοποίηση δεδομένων πάνω στις ομοιότητες τους. Σε αντίθεση με τις τεχνικές classification, η ομαδοποίηση δεν απαιτεί τα δεδομένα να είναι επισημασμένα, το οποίο τη καθιστά ένα χρήσιμο εργαλείο στο εντοπισμό μοτίβων.

Το K-Means Clustering είναι ένας από τους παλαιότερους και πιο συχνά χρησιμοποιούμενους τύπους αλγορίθμων ομαδοποίησης και λειτουργεί με βάση διανυσματική κβαντοποίηση. Υπάρχει ένα σημείο στο διάστημα που επιλέγεται ως αρχή και, στη συνέχεια, σχεδιάζονται διανύσματα από την αρχή σε όλα τα σημεία στο σύνολο δεδομένων. Η συστάδες ομαδοποιούνται με βάση την ομοιότητα τους. Με αυτό το τρόπο αποτελεί ένα χρήσιμο εργαλείο για την αναγνώριση των Fake News γιατί μπορεί να αναγνωρίσει και να ομαδοποιήσει άρθρα που φαίνεται να έχουν ίδιο θέμα και ομοιότητες με ψευδή άρθρα.

Μια άλλη τεχνική, είναι η Ιεραρχική συσταδοποίηση (hierarchical clustering). Η συγκεκριμένη τεχνική, σε αντίθεση με τις υπόλοιπες τεχνικές ομαδοποίησης, δεν απαιτεί τον προκαθορισμό του αριθμού των συστάδων. Η τεχνική αυτή δημιουργεί ένα δενδρόγραμμα σύμφωνα με την ιεραρχική σχέση μεταξύ των διαφόρων συστάδων και έτσι διευκολύνει τον εντοπισμό συσχετισμών ανάμεσα στα δεδομένα.

Οι τεχνικές αυτές αποτελούν ένα πολύ σημαντικό εργαλείο επειδή μας δείχνουν ομοιότητες ανάμεσα σε άρθρα και μας βοηθούν να αναγνωρίσουμε μοτίβα που ακολουθούν τα παρόμοια άρθρα μεταξύ τους.

2.2.3 Κανόνες Συσχέτισης

Η εκμάθηση κανόνων συσχέτισης (Association Rule Learning) είναι μια μη επιβλεπόμενη μέθοδος μάθησης και εξόρυξης δεδομένων. Η τεχνική αυτή χρησιμεύει στην αναγνώριση μοτίβων και σχέσεων ανάμεσα σε στοιχεία ενός μεγάλου σετ δεδομένων. Με αυτό το τρόπο μπορούμε να φτιάξουμε κανόνες για το πως συνδέονται τα δεδομένα μεταξύ τους. Έτσι στη συγκεκριμένη περίπτωση, χρησιμεύει για την αναγνώριση σχέσεων άλλα και χαρακτηριστικών των άρθρων και ειδικά αυτών που είναι σηματοδεδειμένα ως Fake News.

Ένας βασικός αλγόριθμος συσχέτισης είναι ο Apriori. Ο αλγόριθμος αυτός έχει χρησιμοποιηθεί στην ανίχνευση ως υποστηρικτική τεχνική σε μοντέλα μηχανικής και βαθιάς μάθησης. Συγκεκριμένα χρησιμεύει στην ανάλυση δεδομένων που κοινοποιήθηκαν σε social networks. Έρευνες έχουν δείξει πως μπορεί να βοηθήσει στην ανίχνευση παραποιημένων πληροφοριών στα κοινωνικά μέσα ωστόσο όχι με την επιτυχία που έχουν τα μοντέλα μηχανικής και βαθιάς μάθησης. Παρόλα αυτά αποτελεί ένα εργαλείο στην καταπολέμηση του φαινομένου αυτού.

2.3 Μηχανική Μάθηση

Η Μηχανική Μάθηση (Machine Learning) είναι ένα από τα σημαντικότερα εργαλεία που διαθέτουμε για την ανίχνευση Fake News. Αποτελείται από τεχνικές που χρησιμοποιούνται στην εξόρυξη δεδομένων, οι οποίες μπορούν να αναλύσουν τεράστια σετ από δεδομένα, να ανακαλύψουν συσχετισμούς και μοτίβα μεταξύ αυτών και να κάνουν προβλέψεις σχετικά με τις αναλύσεις αυτές.

Οι αλγόριθμοι αυτοί είναι ένα βασικό συστατικό για το στήσιμο αποτελεσματικών μοντέλων ανίχνευσης, και όπως θα δούμε στη συνέχεια, πολύ αποτελεσματικοί. Οι αλγόριθμοι αυτοί εκπαιδεύονται σε προσαρμοσμένα και οργανωμένα σετ από έγκυρες και ψεύτικες ειδήσεις και έπειτα είναι ικανά να αναλύσουν και να αναγνωρίσουν μοτίβα ψεύτικων ειδήσεων σε πραγματικό χρόνο με ελάχιστη ανθρώπινη παρεμβολή. Σε αυτό το κεφάλαιο παρουσιάζονται μερικοί από τους σημαντικότερους αλγορίθμους μηχανικής μάθησης που χρησιμεύουν στην ανίχνευση των Fake News.

Η εποπτευόμενη μηχανική μάθηση είναι η πιο διαδεδομένη εκδοχή της μηχανικής μάθησης, ιδιαίτερα στο θέμα της κατηγοριοποίησης ειδήσεων και αναγνώρισης παραπλανητικών άρθρων. Σε αυτή τη κατηγορία, τα μοντέλα εκπαιδεύονται σε προκαθορισμένα σετ δεδομένων. Δηλαδή τα δεδομένα φέρουν ετικέτες με τα χαρακτηριστικά τους και τις σωστές απαντήσεις. Για παράδειγμα ένα σετ δεδομένων με άρθρα ειδήσεων όπου κάθε άρθρο θα έχει μια ετικέτα ανάλογη με το αν είναι έγκυρο ή ψευδής. Με αυτό το τρόπο το μοντέλο εκπαιδεύεται και στη συνέχεια μπορεί να κάνει έγκυρη πρόβλεψη σε άρθρα χωρίς ετικέτες.

Τα **Διανυσματικά Μηχανήματα Υποστήριξης** ή **Support Vector Machines (SVMs)** είναι μια πολύ ισχυρή Machine Learning τεχνική κατηγοριοποίησης. Τα SVM's βρίσκουν το υπερεπίπεδο (hyperlane) που διαχωρίζει με καλύτερο τρόπο τα δεδομένα σε διαφορετικές κατηγορίες. Αυτό τα καθιστά αποτελεσματικά όταν δουλεύουν σε χώρους πολλαπλών διαστάσεων και ξεχωρίζουν για την εργασία τους σε δυαδική ταξινόμηση, γεγονός που τα καθιστά καταλληλά για τη διάκριση μεταξύ ψευδούς και γνήσιου ειδησεογραφικού υλικού.

Ένας άλλος αλγόριθμος στη κατηγορία αυτή είναι ο **Naïve Bayes**, ο οποίος βασίζεται στο κανόνα του Bayes και τη θεωρία πιθανοτήτων. Αν και είναι ένας απλός αλγόριθμος δεν σημαίνει όμως πως δεν είναι αποτελεσματικός στην ταξινόμηση ειδήσεων. Ο αλγόριθμος αυτός αναλύει ένα άρθρο ειδήσεων και το κατηγοριοποιεί ανάλογα με τις λέξεις που αυτό περιέχει. Υπολογίζει δηλαδή τη πιθανότητα κάθε λέξης να εμφανιστεί σε ένα αναληθές ή έγκυρο άρθρο ειδήσεων. Είναι ένας γρήγορος και απλός τρόπος για τη διαδικασία αυτή που όμως δεν είναι τόσο αποτελεσματικός όσο οι πιο ανεπτυγμένοι αλγόριθμοι ανίχνευσης.

Μια άλλη τεχνική που επίσης χρησιμοποιείται τακτικά στην αντιμετώπιση του φαινομένου είναι η **Λογιστική Παλινδρόμηση (Logistic Regression)**. Ο αλγόριθμος αυτός μπορεί να λύσει προβλήματα δυαδικής ταξινόμησης όπου το αποτέλεσμα-έξοδος είναι μια από δυο τιμές (αναληθές ή έγκυρη είδηση).

2.3.1 Βαθιά Μάθηση

Η Βαθιά Μάθηση γνωστή ως Deep Learning, είναι μια επέκταση της μηχανικής μάθησης. Ο κλάδος αυτός έχει κυριαρχήσει τα τελευταία χρόνια στη ανάλυση δεδομένων καθώς μπορεί να αναγνωρίσει πολύπλοκα μοτίβα και να κατανοήσει έννοιες μέσα από την ανάλυση κείμενων.

Τα **Συνεκτικά Νευρωνικά Δίκτυα** ή **Convolutional Neural Networks (CNNs)** και τα **Επαναλαμβανόμενα Νευρωνικά Δίκτυα** ή **Recurrent Neural Networks (RNNs)** είναι Deep Learning μοντέλα τα οποία χρησιμοποιούνται στην ανίχνευση τα τελευταία χρόνια. Τα CNN παλιότερα χρησιμοποιούνταν για την επεξεργασία εικόνας και πλέον έχουν προσαρμοστεί στη κατηγοριοποίηση κειμένου. Τα μοντέλα αυτά έχουν αποδειχθεί ικανά στο να βρουν μοτίβα σε ακολουθίες λέξεων. Από την άλλη πλευρά, τα RNN είναι σχεδιασμένα να χειρίζονται διαδοχικά δεδομένα. Αυτό έχει συνεισφέρει σημαντικά στην ανάλυση και ταξινόμηση ειδήσεων με βάση το νόημα. Αυτές οι τεχνικές έχουν φέρει σημαντικά αποτελέσματα και δείχνουν πολύ υποσχόμενες στο να εντοπίζουν διαφοροποιημένα μοτίβα ψευδών ειδήσεων. Ωστόσο για να καταφέρουν να το κάνουν προϋποθέτουν μεγάλο αριθμό από σετ δεδομένων για να εκπαιδευτούν και επιπλέον υπολογιστικούς πόρους.

2.4 Επεξεργασία Φυσικής Γλώσσας (NLP)

Το NLP είναι ένας κλάδος της τεχνίτης νοημοσύνης που ασχολείται με την επεξεργασία φυσικής γλώσσας και την αλληλεπίδραση της με τον υπολογιστή. Αποτελεί ένα από τα σημαντικότερη εργαλεία για την ανάλυση των δεδομένων που έχουν εξαχθεί καθώς έχει την ικανότητα να αναλύσει γλωσσικά μοτίβα και δομές που παρατηρούνται συχνά στα παραπλανητικά άρθρα ειδήσεων. Τα μοντέλα αυτά μπορούν να κατανοήσουν και να ερμηνεύσουν την ανθρώπινη γλώσσα με τεράστια επιτυχία φέρνοντας έτσι εκπληκτικά αποτελέσματα στην ανίχνευση των Fake News.

2.4.1 Αναπαράσταση κειμένου

Η διαδικασία με την οποία μετατρέπουμε τα δεδομένα κειμένου σε μορφή τέτοια ώστε να είναι πλήρως κατανοητή και γρήγορη στη προσπέλαση από τον υπολογιστή ονομάζεται τεχνική αναπαράστασης κειμένου ή αλλιώς **text representation**. Ο τρόπος και οι τεχνικές που θα χρησιμοποιήσουμε για να το πετύχουμε αυτό επηρεάζει καθοριστικά το μοντέλο ανίχνευσης και την αποτελεσματικότητα που θα έχει.

Μια απλή άλλα πολυσύχναστη μορφή αναπαράστασης είναι το μοντέλο **Bag of Words (BoW)**. Σε αυτό το μοντέλο, τα κείμενα αναπαρίστανται ως πολυώνυμα λέξεων χωρίς να δίνεται σημασία στη σειρά που εμφανίζονται οι λέξεις. Έτσι η γραμματική και η σημασία κάθε λέξης χάνεται, όμως μένει η συχνότητα κάθε λέξης. Για κάθε άρθρο, παράγεται ένα λεξιλόγιο το οποίο περιέχει όλες τις μοναδικές λέξεις που εμφανίζονται σε αυτό. Έπειτα μετρίεται ο αριθμός που εμφανίζεται η κάθε λέξη στο κείμενο και τέλος δημιουργείται ένα διάνυσμα στο οποίο κάθε θέση του αντιστοιχεί μια λέξη του λεξιλογίου και ο αριθμός εμφάνισής της. Όπως είναι αναμενόμενο, το μοντέλο αυτό έχει μειονέκτημα ότι αγνοεί τη σειρά εμφάνισής των λέξεων και έτσι χάνεται το νόημα το οποίο πολλές φορές είναι σημαντικό στην ανίχνευση ψευδών ειδήσεων.

Για να καλυφθεί το μειονέκτημα αυτό, καθώς επεξεργαζόμαστε την αναπαράσταση ενός κειμένου, χρησιμοποιείται το μοντέλο **Term Frequency-Inverse Document Frequency (TF-IDF)**. Η τεχνική αυτή χρησιμοποιείται ευρέως πλέον στην ανίχνευση ψεύτικων ειδήσεων. Η αναπαράσταση αυτή αποδίδει βαρύτητα στις λέξεις ανάλογα τόσο με την εμφάνιση τους στο κείμενο που αναλύεται όσο και με την εμφάνιση τους στο σύνολο των κείμενων. Η βασική αρχή του είναι ότι δίνει μεγαλύτερη σημασία σε λέξεις που εμφανίζονται συχνά στο κείμενο αλλά σπάνια σε άλλα κείμενα.

Το TF-IDF αποτελείται από δύο βασικά συστατικά:

TF (Term Frequency): Ο τύπος αυτός μετρά ποσό συχνά εμφανίζεται μια λέξη στο έγγραφο.

$$TF(t, d) = \frac{\text{Αριθμός εμφανίσεων της λέξης } t \text{ στο έγγραφο } d}{\text{Συνολικός αριθμός λέξεων στο έγγραφο } d}$$

IDF (Inverse Document Frequency): Ο τύπος αυτός μειώνει τη σημασία των λέξεων που εμφανίζονται συχνά σε πολλά έγγραφα (π.χ., κοινές λέξεις όπως "το", "και").

$$IDF(t, D) = \log \left(\frac{\text{Συνολικός αριθμός εγγράφων}}{\text{Αριθμός εγγράφων όπου εμφανίζεται η λέξη } t} \right)$$

Ο συνδυασμός αυτών των δύο στοιχείων (**TF** και **IDF**) παράγει ένα τελικό βάρος για κάθε λέξη στο έγγραφο. Οι λέξεις που εμφανίζονται συχνά στο έγγραφο αλλά όχι σε πολλά άλλα έγγραφα, παίρνουν μεγαλύτερο βάρος και θεωρούνται πιο σημαντικές για το συγκεκριμένο κείμενο.

Μια εξίσου διαδεδομένη τεχνική αναπαράστασης είναι τα **Word Embeddings**. Αποτελούν τεχνική όπου λέξεις που έχουν παρόμοια σημασία τοποθετούνται κοντά η μια με την άλλη. Σε αντίθεση με τα δυο μοντέλα που είδαμε προηγούμενος, BoW και TF-IDF, τα Word Embeddings προσπαθούν να λαμβάνουν υπόψη τη σημασιολογική σύνδεση και τις σχέσεις μεταξύ των λέξεων. Μερικές από τις πιο δημοφιλείς τεχνικές που περιλαμβάνουν τα Word Embeddings είναι το **Word2Vec**, το μοντέλο **GloVe**, και το **FastText**.

Για την ανίχνευση Fake News, τα Word Embeddings μοντέλα χρησιμοποιούνται ώστε να μετατρέψουν τις λέξεις σε διανύσματα και στη συνέχεια τροφοδοτούν αυτά τα διανύσματα στα μοντέλα Μηχανικής ή Βαθιάς Μάθησης. Έτσι τα μοντέλα μάθησης μπορούν να βρουν και να κατανοούν τη σημασία των λέξεων όπως συμφραζόμενα και υπονοούμενα.

Τέλος, πρόσφατα έχει αναπτυχθεί και μια πιο εξελιγμένη μορφή αναπαράστασης κειμένου σε σύγκριση με τα στατικά Word Embeddings των Word2Vec και GloVe. Σε αυτό το μοντέλο, μια λέξη έχει διαφορετική αναπαράσταση ως διάνυσμα ανάλογα με τη σημασία της στο κείμενο και το συμφραζόμενο της σε αυτό. Αυτές οι μορφές ονομάζονται **contextualized word embeddings** και παρατηρούνται στα πιο ανεπτυγμένα μοντέλα όπως το BERT που θα δούμε στη συνέχεια.

2.4.2 Προηγμένα NLP Μοντέλα - BERT

Τα τελευταία χρόνια, η ανίχνευση ειδήσεων έχει ενισχυθεί σημαντικά με την ανάπτυξη πολλών προηγμένων μοντέλων επεξεργασίας φυσικής γλώσσας. Τα μοντέλα αυτά χρησιμοποιούν βαθιά μάθηση και προεκπαιδευμένα γλωσσικά μοντέλα και καταφέρνουν έτσι μεγάλες επιδόσεις στην αντιμετώπιση των Fake News.

Το **BERT** (Bidirectional Encoder Representations from Transformers), ένα υπερσύγχρονο τελευταίας λέξης μοντέλο που αναπτύχθηκε από τη Google AI το 2018, είναι ίσος το πιο διαδεδομένο προηγμένο μοντέλο. Το συγκεκριμένο μοντέλο κάνει χρήση της αρχιτεκτονικής των transformers και contextualized word embeddings, που όπως αναφέρθηκε παραπάνω, είναι αναπαραστάσεις λέξεων που λαμβάνουν κατά νου τα συμφραζόμενα και τις έννοιες. Το βασικότερο ίσος χαρακτηριστικό του BERT που το κάνει να ξεχωρίζει από τα προηγούμενα μοντέλα είναι η αμφίδρομη εκπαίδευση (**bidirectional training**). Πιο συγκεκριμένα, το μοντέλο επεξεργάζεται το κείμενο προς και τις δυο κατευθύνσεις και έτσι μπορεί να κατανοεί με μεγαλύτερη ακρίβεια τις έννοιες και τις σχέσεις μεταξύ των λέξεων μιας πρότασης μέσα στο κείμενο. Επιπλέον το BERT προεκπαιδύεται σε μεγάλα σετ δεδομένων και έπειτα μπορεί να προσαρμοστεί σε συγκεκριμένες διεργασίες όπως η ανίχνευση των ειδήσεων.

2.5 Υβριδικά Μοντέλα

Τα υβριδικά μοντέλα συνδυάζουν διαφορετικές τεχνικές για να δημιουργήσουν ένα πιο αποτελεσματικό σύστημα ανίχνευσης. Σε γενικές γραμμές, κανένα μοντέλο από μόνο του δεν είναι αρκετό για να αντιμετωπίσει κάθε είδους δυσκολία που εμφανίζεται σε τέτοιες διαδικασίες. Έχει αποδειχθεί μέσα από πειράματα πως ο συνδυασμός των μοντέλων μπορεί να είναι πολύ πιο αποδοτικός καθώς μπορούν να προσεγγίσουν από πολλές πλευρές την πολυπλοκότητα του προβλήματος των Fake News.

Ο πιο διαδεδομένος συνδυασμός είναι αυτός που χρησιμοποιεί NLP τεχνικές μαζί με Machine Learning ή Deep Learning αλγορίθμους ή με αλγορίθμους ανάλυσης δικτύου. Με αυτή τη τεχνική, τα διάφορα μοντέλα εκπαιδεύονται ξεχωριστά και τα αποτελέσματά τους συνδυάζονται για να παραχθεί μια τελική πρόβλεψη. Τα τελευταία χρόνια έχει δημιουργηθεί τεράστιο ενδιαφέρον γύρω από τέτοιου είδους μοντέλα καθώς έχουν δείξει τρομερές προοπτικές.

2.5.1 NLP και Ανάλυση Δικτύου

Μία από τις πιο αποτελεσματικές μεθόδους ανίχνευσης ψευδών ειδήσεων συνδυάζει την ανάλυση φυσικής γλώσσας (NLP) με την ανάλυση δικτύων. Το NLP χρησιμοποιείται για την ανάλυση του κειμένου των ειδήσεων, εντοπίζοντας χαρακτηριστικά όπως συναισθηματικά φορτισμένη γλώσσα, υπερβολές και ασυνέπειες που συχνά συνδέονται με ψευδείς ειδήσεις. Από την άλλη, η ανάλυση δικτύων εξετάζει πώς διαδίδονται οι ειδήσεις στα κοινωνικά δίκτυα, μελετώντας τη συμπεριφορά των χρηστών και τη δομή των συνδέσεων τους, ώστε να εντοπίσει ύποπτα μοτίβα διάδοσης.

Ο συνδυασμός αυτών των δύο μεθόδων προσφέρει μια πιο ολοκληρωμένη προσέγγιση στην ανίχνευση ψευδών ειδήσεων. Ένα άρθρο που μπορεί να θεωρηθεί ύποπτο βάσει του κειμένου του, επιβεβαιώνεται εάν διαπιστωθεί ότι διαδίδεται γρήγορα μέσω bots ή σε κλειστές ομάδες με παρόμοιες απόψεις. Αυτή η υβριδική προσέγγιση βελτιώνει σημαντικά την ακρίβεια των συστημάτων ανίχνευσης, αφού λαμβάνει υπόψη τόσο το περιεχόμενο όσο και το πλαίσιο διάδοσης της είδησης.

2.5.2 Υβριδικά Μοντέλα Μηχανικής Μάθησης

Τα υβριδικά μοντέλα μηχανικής μάθησης συνδυάζουν διαφορετικές τεχνικές για να βελτιώσουν την ανίχνευση ανακριβών ειδήσεων. Αντί να βασίζονται μόνο σε έναν τρόπο ανάλυσης, χρησιμοποιούν συνδυασμούς μεθόδων, όπως η ανάλυση κειμένου με deep learning και τα παραδοσιακά μοντέλα όπως το Random Forest. Έτσι, ενώ ένα σύστημα μπορεί να εντοπίζει ύποπτες λέξεις ή φράσεις στο κείμενο, ένας άλλος αλγόριθμος παίρνει την τελική απόφαση, προσφέροντας πιο σίγουρα αποτελέσματα.

Αυτός ο συνδυασμός γίνεται ακόμη πιο αποτελεσματικός όταν λαμβάνονται υπόψη και άλλοι παράγοντες, όπως ο τρόπος με τον οποίο κυκλοφορεί μια είδηση στα social media. Αν αντιληφθούμε μια είδηση να εξαπλώνεται αστραπιαία μέσω bots ή μέσα σε συγκεκριμένες ομάδες, αυτό μπορεί να είναι ένδειξη παραπληροφόρησης. Συνδυάζοντας την ανάλυση του κειμένου με τη μελέτη του πώς μοιράζεται μια είδηση, τα υβριδικά μοντέλα έχουν μεγαλύτερη ακρίβεια και μπορούν να ξεχωρίσουν τις ψευδείς ειδήσεις πιο εύκολα.

2.5.3 Υβριδικά Μοντέλα με Πολυμέσα

Οι ψευδείς ειδήσεις δεν περιορίζονται πια μόνο στο κείμενο. Συνοδεύονται από εικόνες, βίντεο και ήχο, κάτι που κάνει την ανίχνευσή τους πιο περίπλοκη. Για να αντιμετωπιστεί αυτό, τα συστήματα ανίχνευσης έχουν αρχίσει να συνδυάζουν την ανάλυση κειμένου με την ανάλυση πολυμέσων. Για παράδειγμα, μια εικόνα μπορεί να έχει αλλοιωθεί και τα σημάδια της επεξεργασίας να φανούν σε λεπτομέρειες όπως το φως ή τα pixels. Στα βίντεο, ειδικά με τα deepfakes, η ανίχνευση είναι πιο δύσκολη, καθώς χρειάζεται να εξεταστούν προσεκτικά η κίνηση του προσώπου και ο συγχρονισμός του ήχου.

Όταν αυτά τα συστήματα συνδυάζουν την ανάλυση πολυμέσων με τις κλασικές μεθόδους αναγνώρισης κειμένου και διάδοσης στα κοινωνικά δίκτυα, η αποτελεσματικότητά τους αυξάνεται σημαντικά. Μια είδηση μπορεί να μοιάζει ύποπτη λόγω του κειμένου, αλλά αν συνοδεύεται από παραποιημένα πολυμέσα, το σύστημα μπορεί να την εντοπίσει πιο εύκολα. Καθώς η παραπληροφόρηση με πολυμέσα αυξάνεται, αυτά τα υβριδικά συστήματα θα παίζουν ακόμα πιο σημαντικό ρόλο στην ανίχνευση αναληθών ειδήσεων.

2.5.4 Προκλήσεις και Μελλοντική Ανάπτυξη

Οι υβριδικές μέθοδοι για την ανίχνευση ψευδών ειδήσεων προσφέρουν πολλά πλεονεκτήματα, αλλά συνοδεύονται και από δυσκολίες. Ένα από τα μεγαλύτερα προβλήματα είναι το γεγονός ότι χρειάζονται πολλοί υπολογιστικοί πόροι για να επεξεργαστούν τεράστιες ποσότητες δεδομένων, όπως κείμενα, εικόνες και βίντεο. Αυτό καθιστά δύσκολη την ευρεία εφαρμογή τους.

Επιπλέον, υπάρχει η πρόκληση του πώς θα συνδυαστούν όλα αυτά τα δεδομένα με έναν τρόπο που να λειτουργούν ομαλά μαζί. Κάθε τύπος δεδομένων, είτε κείμενο είτε εικόνα, απαιτεί διαφορετική προσέγγιση για να επεξεργαστεί σωστά, και δεν είναι πάντα εύκολο να τα κάνεις να "συνεργαστούν". Αν αυτό δεν γίνει σωστά, η ακρίβεια του συστήματος μπορεί να μειωθεί.

Ένα ακόμα ζήτημα είναι ότι τα Fake News συνεχώς αλλάζουν και εξελίσσονται. Οι δημιουργοί τους γίνονται πιο έξυπνοι και ανακαλύπτουν νέους τρόπους για να παραπλανούν το κοινό. Τα συστήματα ανίχνευσης πρέπει διαρκώς να προσαρμόζονται σε αυτές τις νέες τακτικές, κάτι που σημαίνει πως απαιτείται συνεχής έρευνα και ανάπτυξη για να παραμένουν αποτελεσματικά.

Στο μέλλον, η πρόκληση είναι να κάνουμε αυτά τα συστήματα πιο αποδοτικά, ώστε να μην απαιτούν τόσους πόρους. Ίσως με την ανάπτυξη νέων τεχνολογιών ή τη χρήση του cloud να επιτευχθεί αυτό. Ταυτόχρονα, τα συστήματα θα πρέπει να είναι ικανά να παρακολουθούν τη διάδοση των Fake News σε διαφορετικές πλατφόρμες, από τα social media μέχρι τις ειδησεογραφικές ιστοσελίδες.

Εν τέλει, ενώ οι υβριδικές μέθοδοι έχουν τεράστια προοπτική, η αντιμετώπιση αυτών των δυσκολιών είναι απαραίτητη για να γίνουν πραγματικά αποτελεσματικές. Με τις σωστές βελτιώσεις, αυτά τα συστήματα μπορούν να γίνουν πολύ πιο ισχυρά και ευέλικτα απέναντι στη συνεχή απειλή της παραπληροφόρησης.

Κεφάλαιο 3: Αρχιτεκτονική Συστημάτων Ανίχνευσης και Μεθοδολογία

Σε αυτό το κεφάλαιο θα αναλύουμε την αρχιτεκτονική και τις μεθόδους που χρησιμοποιούνται για τη δημιουργία ενός συστήματος ανίχνευσης ειδήσεων. Δεδομένου ότι αυτή η διαδικασία είναι σύνθετη, χρειάζεται ένα σύστημα που συνδυάζει τεχνικές για την επεξεργασία δεδομένων, την εξαγωγή χαρακτηριστικών και την ανάλυση σε πραγματικό χρόνο. Σκοπός μου είναι να παρουσιάσω τα βασικά στοιχεία ενός τέτοιου συστήματος, εξηγώντας τον ρόλο κάθε σταδίου.

Ένα σύστημα ανίχνευσης αποτελείται συνήθως από διάφορα επίπεδα, όπως η συλλογή και προετοιμασία δεδομένων, η εξαγωγή χαρακτηριστικών και η εκπαίδευση των μοντέλων. Υπάρχουν πολλές τεχνικές που κάνουν την ίδια δουλειά και επιλέγονται βάσει της ικανότητάς τους να διαχειρίζονται διαφορετικούς τύπους δεδομένων (κείμενα, εικόνες, βίντεο) και να καλύπτουν την ανάγκη για ακρίβεια και κλιμάκωση.

Αρχικά, παρουσιάζεται η συνολική αρχιτεκτονική του συστήματος, προχωρώντας στις τεχνικές εξαγωγής χαρακτηριστικών και στη συνέχεια στην εκπαίδευση και βελτιστοποίηση των μοντέλων μηχανικής μάθησης. Τέλος, θα δούμε πως λειτουργεί η ανάλυση σε πραγματικό χρόνο, που είναι κρίσιμη για την άμεση ανίχνευση και αντιμετώπιση των ψευδών ειδήσεων καθώς εξαπλώνονται στις ψηφιακές πλατφόρμες.

3.1 Αρχιτεκτονική συστήματος

Η αρχιτεκτονική ενός συστήματος ανίχνευσης Fake News έχει σχεδιαστεί για να διαχειρίζεται αποτελεσματικά τις σύνθετες εργασίες συλλογής, επεξεργασίας, ανάλυσης και ταξινόμησης δεδομένων. Μια καλά δομημένη αρχιτεκτονική διασφαλίζει ότι αυτές οι διαδικασίες πραγματοποιούνται συντονισμένα και με δυνατότητα κλιμάκωσης, επιτρέποντας στο σύστημα να λειτουργεί αποτελεσματικά σε περιβάλλοντα όπου οι ψευδείς ειδήσεις διαδίδονται γρήγορα και απρόβλεπτα.

- **Επίπεδο Συλλογής Δεδομένων:** Το πρώτο επίπεδο της αρχιτεκτονικής είναι υπεύθυνο για τη συλλογή δεδομένων από διάφορες πηγές, όπως πλατφόρμες κοινωνικών δικτύων, ειδησεογραφικά sites, δημόσια datasets και υποβολές χρηστών. Τα δεδομένα μπορεί να είναι σε διάφορες μορφές όπως κείμενο, εικόνες, βίντεο και μεταδεδομένα. Συχνά χρησιμοποιούνται APIs και εργαλεία scraping για τη συλλογή δεδομένων σε πραγματικό χρόνο, εξασφαλίζοντας ότι το σύστημα ενημερώνεται συνεχώς.
- **Επίπεδο Προεπεξεργασίας:** Μόλις συλλεχθούν τα δεδομένα, περνούν στο στάδιο της προεπεξεργασίας, όπου καθαρίζονται και μετασχηματίζονται. Τα κείμενα χωρίζονται σε λέξεις, κανονικοποιούνται (π.χ., μετατρέπονται σε πεζά, αφαιρούνται οι λέξεις που δεν προσφέρουν πληροφορίες), και μετατρέπονται σε διανύσματα, ώστε να είναι κατάλληλα για περαιτέρω ανάλυση. Οι εικόνες και τα βίντεο επεξεργάζονται για την εξαγωγή χαρακτηριστικών, όπως μοτίβα pixels ή keyframes. Σε αυτό το επίπεδο, διαχειρίζεται επίσης ελλιπή δεδομένα, μείωση θορύβου και ενοποίηση δεδομένων από διαφορετικές πηγές, εξασφαλίζοντας ένα καθαρό και συνεπές σετ δεδομένων για την ανάλυση.
- **Επίπεδο Εξαγωγής Χαρακτηριστικών:** Αφού ολοκληρωθεί η προεπεξεργασία, τα δεδομένα προχωρούν στο στάδιο της εξαγωγής χαρακτηριστικών, όπου εντοπίζονται και εξάγονται σημαντικά στοιχεία. Στην περίπτωση των κειμένων, μπορεί να περιλαμβάνει για παράδειγμα την εξαγωγή embeddings ή συντακτικών μοτίβων. Για το πολυμεσικό περιεχόμενο, η εξαγωγή χαρακτηριστικών μπορεί να περιλαμβάνει την ανάλυση οπτικών στοιχείων, ηχητικών μοτίβων ή την ανίχνευση ανωμαλιών στα καρέ των βίντεο. Η διαδικασία εξαγωγής χαρακτηριστικών είναι κρίσιμη για τη μείωση της πολυπλοκότητας των δεδομένων, διατηρώντας τα πιο σημαντικά στοιχεία για τη χρήση από τα μοντέλα μηχανικής μάθησης.
- **Επίπεδο Εκπαίδευσης και Πρόβλεψης Μοντέλων:** Αυτό το επίπεδο είναι η ουσία του συστήματος ανίχνευσης ψευδών ειδήσεων, όπου εκπαιδεύονται τα μοντέλα μηχανικής μάθησης και εφαρμόζονται για την ταξινόμηση του περιεχομένου ως αληθές ή ψευδές. Τα μοντέλα εκπαιδεύονται με ετικετοποιημένα δεδομένα, χρησιμοποιώντας τα χαρακτηριστικά που εξήχθησαν στο προηγούμενο στάδιο ως είσοδο. Η αρχιτεκτονική μπορεί να υποστηρίζει διάφορα μοντέλα, από παραδοσιακούς αλγορίθμους όπως SVMs και Random Forests, μέχρι πιο σύγχρονα μοντέλα βαθιάς μάθησης όπως τα CNNs και RNNs. Το σύστημα μπορεί επίσης να ενσωματώνει υβριδικές προσεγγίσεις για βελτίωση της ακρίβειας. Αφού εκπαιδευτούν, αυτά τα μοντέλα χρησιμοποιούνται για την πρόβλεψη νέου, άγνωστου περιεχομένου σε πραγματικό χρόνο.

- **Αξιολόγηση Μοντέλων:** Η αξιολόγηση των μοντέλων είναι ένα κρίσιμο βήμα στην ανάπτυξη και την εφαρμογή των συστημάτων ανίχνευσης αναληθών ειδήσεων. Περιλαμβάνει την αξιολόγηση της απόδοσης του μοντέλου, διασφαλίζοντας ότι μπορεί να διαχωρίζει σωστά τις ψευδείς από τις αληθείς ειδήσεις. Η σωστή αξιολόγηση βοηθά στην κατανόηση των δυνατών και αδύνατων σημείων του μοντέλου και καθοδηγεί περαιτέρω βελτιώσεις. Παρουσιάζονται οι βασικές μετρικές αξιολόγησης, με εξηγήσεις για τη χρήση και τους τύπους τους.

Accuracy

Η ακρίβεια (accuracy) είναι η πιο απλή μετρική αξιολόγησης και μετρά το ποσοστό των σωστών προβλέψεων (τόσο θετικών όσο και αρνητικών) σε σχέση με το συνολικό αριθμό των δεδομένων. Αν και η ακρίβεια δίνει μια γενική εικόνα της απόδοσης του μοντέλου, μπορεί να είναι παραπλανητική σε περιπτώσεις όπου τα δεδομένα είναι ανισομερώς κατανομημένα (π.χ., αν υπάρχουν περισσότερες αληθινές ειδήσεις από ψευδείς).

- **Τύπος:**

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Number of Instances}}$$

- **Χρήση:** Η ακρίβεια είναι χρήσιμη σε ισορροπημένα σύνολα δεδομένων, όπου οι κατηγορίες είναι περίπου ίσες σε αριθμό.

Precision

Η ακρίβεια (Precision) μετρά το ποσοστό των σωστά αναγνωρισμένων θετικών περιπτώσεων (αληθώς θετικά) σε σχέση με όλες τις περιπτώσεις που το μοντέλο προέβλεψε (αληθώς θετικά και ψευδώς θετικά). Στην ανίχνευση ψευδών ειδήσεων, η ακρίβεια μας δείχνει πόσα από τα άρθρα που το μοντέλο χαρακτήρισε ως ψευδή είναι πράγματι ψευδή.

- **Τύπος:**

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- **Χρήση:** Η ακρίβεια είναι κρίσιμη όταν ο αριθμός των ψευδώς θετικών είναι υψηλό, όπως σε περιπτώσεις όπου η λανθασμένη επισήμανση αληθινών ειδήσεων ως ψευδών μπορεί να οδηγήσει σε αδικαιολόγητη λογοκρισία ή απώλεια αξιοπιστίας.

Recall (Sensitivity or True Positive Rate)

Η ανάκληση (Recall), γνωστή και ως ευαισθησία, μετρά το ποσοστό των πραγματικών θετικών περιπτώσεων (αληθώς θετικά) που αναγνωρίστηκαν σωστά από το μοντέλο. Είναι ο λόγος των αληθώς θετικών προς το άθροισμα των αληθώς θετικών και των ψευδώς αρνητικών. Με πιο απλά λόγια, αφορά το ποσοστό των πραγματικών θετικών περιπτώσεων που εντοπίστηκαν από το μοντέλο. Η ανάκληση είναι ιδιαίτερα σημαντική στην ανίχνευση ψευδών ειδήσεων, καθώς δείχνει πόσο καλά μπορεί το μοντέλο να εντοπίσει τις ψευδείς ειδήσεις χωρίς να χάσει καμία.

- **Τυπος:**

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- **Χρήση:** Η ανάκληση είναι κρίσιμη όταν το κόστος των ψευδώς αρνητικών είναι υψηλό, όπως σε περιπτώσεις όπου η αποτυχία ανίχνευσης ψευδών ειδήσεων θα μπορούσε να επιτρέψει την ανεξέλεγκτη διάδοση παραπληροφόρησης.

F1-Score

Η βαθμολογία F1 είναι ο αρμονικός μέσος της ακρίβειας (precision) και της ανάκλησης (recall), παρέχοντας έναν ενιαίο δείκτη που εξισορροπεί και τα δύο. Είναι ιδιαίτερα χρήσιμη όταν υπάρχει ανισομερής κατανομή των κατηγοριών ή όταν και η ακρίβεια και η ανάκληση είναι εξίσου σημαντικές, καθώς προσφέρει μια πιο ολοκληρωμένη εικόνα της απόδοσης του μοντέλου σε σχέση με τη χρήση μόνο της ακρίβειας ή της ανάκλησης.

- **Τυπος:**

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Χρήση:** Η βαθμολογία F1 είναι ιδανική σε περιπτώσεις όπου πρέπει να ισορροπήσεις ανάμεσα στην ακρίβεια και την ανάκληση, ειδικά όταν έχεις ανισομερή κατανομή δεδομένων.

Confusion Matrix

Ο πίνακας σύγχυσης (Confusion Matrix) είναι ένας πίνακας που παρέχει αναλυτική εικόνα της απόδοσης του μοντέλου σε όλες τις κατηγορίες, παρουσιάζοντας τον αριθμό των αληθώς θετικών, ψευδώς θετικών, αληθώς αρνητικών και ψευδώς αρνητικών προβλέψεων. Δεν αποτελεί από μόνος του μετρική, αλλά είναι ένα εργαλείο για την οπτικοποίηση και τον υπολογισμό διαφόρων μετρικών, όπως η ακρίβεια (accuracy), η ακρίβεια (precision), η ανάκληση (recall) και η βαθμολογία F1.

- **Πίνακας σύγχυσης:**

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

- **Χρήση:** Ο πίνακας σύγχυσης είναι χρήσιμος για την κατανόηση των συγκεκριμένων λαθών που κάνει το μοντέλο, βοηθώντας έτσι στη βελτίωσή του για καλύτερη απόδοση.

Area Under the ROC Curve (AUC-ROC)

Η καμπύλη ROC (Receiver Operating Characteristic) απεικονίζει τον ρυθμό αληθών θετικών (ανάκληση) σε σχέση με τον ρυθμό ψευδών θετικών (1-ειδικότητα) για διάφορες τιμές ορίων. Το AUC (Area Under the Curve) μετρά τη συνολική απόδοση σε όλα τα όρια. Το AUC-ROC είναι ιδιαίτερα χρήσιμο σε δυαδικές ταξινομήσεις, όπως η ανίχνευση ψεύτικων ειδήσεων, καθώς δείχνει την ισορροπία μεταξύ ευαισθησίας και ειδικότητας.

- **Χρήση:** Το AUC-ROC χρησιμοποιείται για την αξιολόγηση της ικανότητας του μοντέλου να διακρίνει μεταξύ των θετικών και αρνητικών κατηγοριών, με υψηλότερο AUC να υποδεικνύει καλύτερη απόδοση.

Log Loss

Το Log loss, γνωστό και ως λογιστική απώλεια ή cross-entropy loss, μετρά την αβεβαιότητα στις προβλέψεις του μοντέλου, λαμβάνοντας υπόψη την πιθανότητα που αποδίδεται στην πραγματική κατηγορία.

- **Τυπος:**

$$\text{Log Loss} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

Όπου N είναι ο αριθμός των παραδειγμάτων, y_i είναι η πραγματική ετικέτα, και p_i είναι η προβλεπόμενη

πιθανότητα για τη θετική κατηγορία..

- **Χρήση:** Το Log loss είναι ιδιαίτερα χρήσιμο για μοντέλα που παράγουν πιθανότητες αντί για κατηγορηματικές ετικέτες, προσφέροντας μια πιο λεπτομερή εικόνα της βεβαιότητας των προβλέψεων.

- **Επίπεδο Επεξεργασίας σε Πραγματικό Χρόνο:** Το τελευταίο επίπεδο της αρχιτεκτονικής είναι το πλαίσιο επεξεργασίας σε πραγματικό χρόνο, το οποίο είναι υπεύθυνο για την εφαρμογή των εκπαιδευμένων μοντέλων σε ένα ζωντανό περιβάλλον. Αυτό το πλαίσιο διασφαλίζει ότι το σύστημα μπορεί να επεξεργάζεται γρήγορα τα εισερχόμενα δεδομένα, να εφαρμόζει τα μοντέλα ανίχνευσης και να παράγει αποτελέσματα επισημαίνοντας πιθανώς ανακριβείς ειδήσεις, εικόνες ή βίντεο. Το επίπεδο επεξεργασίας σε πραγματικό χρόνο πρέπει να είναι βελτιστοποιημένο για ταχύτητα και αποδοτικότητα, χρησιμοποιώντας τεχνολογίες όπως πλατφόρμες ροής δεδομένων (π.χ., Apache Kafka, Apache Flink) για τη διαχείριση της ροής των δεδομένων και την εξασφάλιση έγκαιρων απαντήσεων.
- **Μηχανισμός Ενσωμάτωσης και Ανατροφοδότησης:** Εκτός από τα κύρια επίπεδα, η αρχιτεκτονική μπορεί να περιλαμβάνει σημεία ενσωμάτωσης με εξωτερικά συστήματα (π.χ., βάσεις δεδομένων ελέγχου γεγονότων, μηχανισμούς ανατροφοδότησης χρηστών) και βρόχους ανατροφοδότησης για συνεχή μάθηση. Καθώς το σύστημα λειτουργεί, μπορεί να συλλέγει ανατροφοδότηση σχετικά με τις προβλέψεις του, η οποία μπορεί να χρησιμοποιηθεί για την επανεκπαίδευση των μοντέλων και τη βελτίωση της ακρίβειάς τους με την πάροδο του χρόνου.

Συνοψίζοντας, η αρχιτεκτονική του συστήματος για την ανίχνευση ψευδών ειδήσεων είναι ένα πολυεπίπεδο πλαίσιο, σχεδιασμένο να διαχειρίζεται την πολύπλοκη και δυναμική φύση της παραπληροφόρησης. Κάθε επίπεδο διαδραματίζει κρίσιμο ρόλο στη διασφάλιση της αποδοτικής και αποτελεσματικής λειτουργίας του συστήματος, από τη συλλογή δεδομένων έως την ανάλυση σε πραγματικό χρόνο.

3.2 Τεχνικές Εξαγωγής Χαρακτηριστικών

Η εξαγωγή χαρακτηριστικών είναι ένα αξιοσημείωτο βήμα στην ανάπτυξη ενός συστήματος ανίχνευσης ψευδών ειδήσεων. Ο κύριος στόχος είναι να μετατρέψουμε τα ακατέργαστα δεδομένα σε δομημένες πληροφορίες που μπορούν να επεξεργαστούν τα μοντέλα μηχανικής μάθησης. Αυτό είναι απαραίτητο για να μπορέσει το σύστημα να διακρίνει με ακρίβεια τις ψευδείς από τις αληθινές ειδήσεις. Οι τεχνικές εξαγωγής διαφέρουν ανάλογα με τον τύπο των δεδομένων (κειμένο, πολυμέσα ή δικτυακές πληροφορίες) και ο κάθε ένας από αυτούς απαιτεί διαφορετική προσέγγιση.

Μία από τις πιο πρωταρχικές μεθόδους για κείμενα είναι το **Bag of Words (BoW)**. Αυτή η μέθοδος μετατρέπει το κείμενο σε έναν πίνακα συχνοτήτων λέξεων, όπου κάθε σειρά αντιπροσωπεύει ένα κείμενο και κάθε στήλη μία λέξη.

Ένα απλό παράδειγμα είναι το εξής: Αν έχουμε τις φράσεις "Τα μήλα είναι γλυκά," "Μου αρέσουν τα μήλα," "Τα μήλα και τα πορτοκάλια είναι φρούτα," και "Μου αρέσουν τα γλυκά φρούτα," φτιάχνουμε έναν πίνακα όπου κάθε σειρά αντιστοιχεί σε μια από αυτές τις προτάσεις και κάθε στήλη σε μια λέξη από το λεξιλόγιο. Για παράδειγμα, στην πρώτη φράση, οι λέξεις "μήλα," "είναι," και "γλυκά" εμφανίζονται μία φορά, ενώ οι υπόλοιπες λέξεις δεν εμφανίζονται καθόλου. Έτσι, κάθε πρόταση μπορεί να αναπαρασταθεί με αριθμούς, κάτι που βοηθά στην παρακάτω ανάλυση.

Πίνακας 3: Πίνακας **BoW**

	μήλα	είναι	γλυκά	εγώ	αρέσουν	και	Πορτοκάλια	φρούτα
Doc 1	1	1	1	0	0	0	0	0
Doc 2	1	0	0	1	1	0	0	0
Doc 3	1	1	0	0	0	1	1	1
Doc 4	0	0	1	1	1	0	0	1

Παρόλο που το BoW είναι απλό, δεν καταφέρνει να αποτυπώσει το πλήρες πλαίσιο ή τη σημασία των λέξεων. Για αυτόν τον λόγο, συνδυάζεται όπως είδαμε με άλλες μεθόδους, όπως το TF-IDF, που προσαρμόζει τη σημασία μιας λέξης ανάλογα με το πόσο συχνά εμφανίζεται σε όλα τα κείμενα.

Στο προηγούμενο παράδειγμα, μπορούμε να υπολογίσουμε το σκορ **TF-IDF** για τη λέξη "μήλα" στις παρακάτω τέσσερις προτάσεις: "Τα μήλα είναι γλυκά," "Μου αρέσουν τα μήλα," "Τα μήλα και τα πορτοκάλια είναι φρούτα," και "Μου αρέσουν τα γλυκά φρούτα." Πρώτα, υπολογίζουμε τη Συχνότητα Όρων (TF) για τη λέξη "μήλα" σε κάθε πρόταση. Στην πρώτη πρόταση, "Τα μήλα είναι γλυκά," η λέξη "μήλα" εμφανίζεται μία φορά σε τρεις λέξεις, άρα το TF είναι $1/3$ ή 0.33 . Στη δεύτερη πρόταση, "Μου αρέσουν τα μήλα," η λέξη "μήλα" εμφανίζεται και πάλι μία φορά σε τρεις λέξεις, δίνοντας το ίδιο TF, 0.33 . Στην τρίτη πρόταση, "Τα μήλα και τα πορτοκάλια είναι φρούτα," η λέξη "μήλα" εμφανίζεται μία φορά σε πέντε λέξεις, δίνοντας TF $1/5$

ή 0.20. Στην τέταρτη πρόταση, "Μου αρέσουν τα γλυκά φρούτα," η λέξη "μήλα" δεν εμφανίζεται, οπότε το TF είναι 0.

Στη συνέχεια, υπολογίζουμε την Αντίστροφη Συχνότητα Εγγράφου (IDF) για τη λέξη "μήλα." Επειδή η λέξη "μήλα" εμφανίζεται σε τρεις από τις τέσσερις προτάσεις, η IDF υπολογίζεται ως $\log(4/3) \approx 0.18$. Τέλος, πολλαπλασιάζουμε το TF με την IDF για να βρούμε το σκορ TF-IDF για τη λέξη "μήλα" σε κάθε πρόταση. Για την πρώτη πρόταση, το σκορ TF-IDF είναι $0.33 * 0.18 \approx 0.06$. Για τη δεύτερη πρόταση, το σκορ είναι επίσης $0.33 * 0.18 \approx 0.06$. Για την τρίτη πρόταση, το σκορ είναι $0.20 * 0.18 \approx 0.04$. Στην τέταρτη πρόταση, αφού η λέξη "μήλα" δεν εμφανίζεται, το σκορ TF-IDF είναι 0.

Αυτά τα σκορ δείχνουν ότι, ενώ η λέξη "μήλα" είναι σημαντική στις πρώτες τρεις προτάσεις, δεν είναι μια μοναδική λέξη στο σύνολο των προτάσεων, κάτι που αντανακλάται στα μέτρια σκορ TF-IDF.

Μια άλλη τεχνική είναι το **Word2Vec**, το οποίο δημιουργεί πιο σύνθετες αναπαραστάσεις των λέξεων, λαμβάνοντας υπόψη τις σχέσεις τους με τις υπόλοιπες λέξεις σε ένα κείμενο. Το μοντέλο μαθαίνει τις συσχετίσεις μεταξύ των λέξεων, κάτι που το κάνει εξαιρετικά αποτελεσματικό στο να κατανοεί τη σημασία των λέξεων σε ένα συγκεκριμένο πλαίσιο.

Το **GloVe** είναι μια παρόμοια τεχνική που βασίζεται στη συχνότητα εμφάνισης ζευγαριών λέξεων. Αυτό το μοντέλο αποτυπώνει τις σχέσεις ανάμεσα στις λέξεις με βάση τα στατιστικά τους και είναι ιδιαίτερα χρήσιμο σε περιπτώσεις όπου η κατανόηση της σημασίας εξαρτάται από το γενικότερο πλαίσιο.

Το **BERT** είναι ένα πιο εξελιγμένο μοντέλο που χρησιμοποιεί δύο κύριες μεθόδους: προβλέπει λέξεις που λείπουν από προτάσεις και μαθαίνει τις σχέσεις ανάμεσα σε προτάσεις. Μετά την αρχική εκπαίδευση, το BERT μπορεί να προσαρμοστεί σε διάφορες εργασίες, όπως η ανίχνευση ψευδών ειδήσεων, και να προσφέρει πολύ ακριβή αποτελέσματα. Το **FakeBERT** είναι μια ειδική προσαρμογή του BERT για την ανίχνευση αναληθών δημοσιευμάτων. Συνδυάζει το BERT με τεχνικές CNNs, επιτρέποντάς του να ανιχνεύει τοπικά μοτίβα που μπορεί να αποκαλύψουν παραπληροφόρηση. Αυτός ο συνδυασμός το καθιστά ιδιαίτερα αποτελεσματικό και πολλές φορές ξεπερνά άλλα μοντέλα στην ακρίβεια.

3.2.1 Εξαγωγή Χαρακτηριστικών από Πολυμέσα

Στην περίπτωση των πολυμέσων, η εξαγωγή χαρακτηριστικών είναι πιο πολύπλοκη λόγω της φύσης των οπτικών και ακουστικών πληροφοριών. Για τις εικόνες, συνήθως χρησιμοποιούνται τα CNNs. Τα CNNs εφαρμόζουν μια σειρά από φίλτρα συνελίξεων στην είσοδο της εικόνας, ανιχνεύοντας άκρα, υφές και άλλα οπτικά χαρακτηριστικά. Αυτά τα φίλτρα "γλιστρούν" πάνω στην εικόνα για να δημιουργήσουν χάρτες χαρακτηριστικών που αναδεικνύουν διαφορετικές πτυχές του οπτικού περιεχομένου. Οι στρώσεις pooling μειώνουν τη διαστατικότητα αυτών των χαρτών χαρακτηριστικών, διατηρώντας τα πιο σημαντικά χαρακτηριστικά και απορρίπτοντας τις περιττές πληροφορίες. Οι τελικές στρώσεις του CNN, που συνήθως είναι πλήρως συνδεδεμένες, επεξεργάζονται αυτά τα χαρακτηριστικά για να εκτελέσουν εργασίες όπως η ταξινόμηση εικόνων ή η ανίχνευση ανωμαλιών.

Για την ανάλυση βίντεο, η εξαγωγή χαρακτηριστικών περιλαμβάνει τόσο χωρικές όσο και χρονικές διαστάσεις. Τα βίντεο διασπώνται σε μεμονωμένα καρέ, με τα CNN να εξάγουν χωρικά χαρακτηριστικά από κάθε καρέ. Για να αποτυπωθεί η χρονική αλληλουχία αυτών των καρέ, χρησιμοποιούνται τα Επαναλαμβανόμενα Νευρωνικά Δίκτυα (RNNs) ή τα Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (LSTMs). Τα LSTMs είναι ιδανικά για αυτή την εργασία, καθώς μπορούν να διατηρούν πληροφορίες για μεγάλες ακολουθίες, κάτι που είναι απαραίτητο για την ανίχνευση μικρών αλλαγών ή ασυνεπειών στα βίντεο, οι οποίες μπορεί να υποδηλώνουν παραποίηση. Ο συνδυασμός CNNs για την ανάλυση του χώρου και LSTMs για την ανάλυση του χρόνου επιτρέπει στο σύστημα να ανιχνεύει με μεγαλύτερη ακρίβεια ψευδείς ειδήσεις σε περιεχόμενο βίντεο.

3.2.2 Εξαγωγή Χαρακτηριστικών από Δίκτυα

Η εξαγωγή χαρακτηριστικών από δίκτυα είναι κρίσιμη για την κατανόηση του τρόπου με τον οποίο οι ψευδείς δημοσιεύσεις διαδίδονται στα κοινωνικά δίκτυα και άλλες διαδικτυακές πλατφόρμες. Η ανάλυση αυτή επικεντρώνεται στις σχέσεις και τις αλληλεπιδράσεις μεταξύ των χρηστών και στο πώς η πληροφορία διαχέεται μέσα από αυτά τα δίκτυα.

Τα μέτρα κεντρικότητας χρησιμοποιούνται για την αναγνώριση των πιο επιδραστικών κόμβων σε ένα δίκτυο. Για παράδειγμα, η βαθμολογία κεντρικότητας μετρά τον αριθμό των άμεσων συνδέσεων που έχει ένας κόμβος, υποδεικνύοντας την επιρροή του. Η κεντρικότητα διαμεσολάβησης μετρά σε ποιο βαθμό ένας κόμβος λειτουργεί ως γέφυρα μεταξύ άλλων κόμβων, ενώ η κεντρικότητα εγγύτητας εντοπίζει κόμβους που μπορούν να διαδώσουν γρήγορα πληροφορίες στο δίκτυο. Αυτές οι μετρήσεις βοηθούν στην αναγνώριση των κύριων παραγόντων ή "κόμβων" που μπορεί να παίζουν σημαντικό ρόλο στη διάδοση ψευδών πληροφοριών.

Οι αλγόριθμοι ανίχνευσης κοινοτήτων, όπως η μέθοδος Louvain ή η βελτιστοποίηση αρθρωτότητας, χρησιμοποιούνται για τον εντοπισμό ομάδων μέσα στο δίκτυο. Αυτές οι ομάδες συχνά αντιπροσωπεύουν "ηχώθαλάμους," όπου οι πληροφορίες μοιράζονται ανάμεσα σε άτομα με παρόμοιες απόψεις. Η κατανόηση αυτών των κοινοτήτων είναι κρίσιμη για την ανίχνευση της αρχικής διάδοσης των ψευδών ειδήσεων και τον εντοπισμό πιθανών στόχων για παρέμβαση.

Επίσης, αναλύονται τα πρότυπα διάδοσης και η ταχύτητα για να κατανοηθεί η δυναμική της διάδοσης των ψευδών ειδήσεων. Τεχνικές όπως η ανάλυση χρονοσειρών μπορούν να παρακολουθήσουν πόσο γρήγορα διαδίδεται η πληροφορία μέσα σε ένα δίκτυο και να εντοπίσουν πρότυπα που είναι τυπικά της παραπληροφόρησης. Για παράδειγμα, οι αναληθείς ειδήσεις συχνά διαδίδονται γρήγορα και φτάνουν σε μεγάλο κοινό σε σύντομο χρονικό διάστημα, σε αντίθεση με τις αληθινές ειδήσεις, οι οποίες μπορεί να διαδίδονται πιο αργά και σταδιακά.

Με την ενσωμάτωση αυτών των διαφορετικών τεχνικών εξαγωγής χαρακτηριστικών, το σύστημα ανίχνευσης μπορεί να επεξεργαστεί και να αναλύσει σύνθετα σύνολα δεδομένων, παρέχοντας μια ισχυρή βάση για τα επόμενα στάδια ταξινόμησης και ανάλυσης.

3.3 Εφαρμογές σε πραγματικό χρόνο

Η ανίχνευση πλαστών ειδήσεων σε πραγματικό χρόνο είναι ζωτικής σημασίας, δεδομένης της ταχύτητας με την οποία μπορεί να διαδοθεί η παραπληροφόρηση στα ψηφιακά μέσα. Ένα καλά οργανωμένο σύστημα μπορεί να αναλύει δεδομένα καθώς αυτά εισέρχονται και να παρέχει άμεσα ειδοποιήσεις για ενδεχόμενα ψευδή περιεχόμενα, ώστε να αποτρέψει τη διάδοση παραπληροφόρησης.

Εισαγωγή Ροών Δεδομένων

Η διαδικασία ξεκινά με τη συλλογή δεδομένων από διάφορες πηγές, όπως κοινωνικά δίκτυα, ειδησεογραφικούς ιστότοπους και φόρουμ. Για να διαχειριστούν τους μεγάλους όγκους δεδομένων που εισέρχονται συνεχώς, χρησιμοποιούνται τεχνολογίες όπως το **Apache Kafka** και το **Apache Flink**. Αυτά τα εργαλεία βοηθούν στο να φιλτράρονται τα άσχετα δεδομένα και να τυποποιούνται οι πληροφορίες, ώστε να είναι έτοιμες για ανάλυση.

Εξαγωγή Χαρακτηριστικών σε Πραγματικό Χρόνο

Μετά τη συλλογή των δεδομένων, το σύστημα περνά στη διαδικασία εξαγωγής χαρακτηριστικών. Αυτό περιλαμβάνει τη μετατροπή των κειμένων σε μορφές που μπορούν να κατανοηθούν από τους αλγορίθμους, όπως η τοκενιστικοποίηση και η διανυσματοποίηση, και την ανάλυση εικόνων ή βίντεο με τη χρήση τεχνολογιών όπως τα **CNNs**. Τα προηγμένα μοντέλα όπως το **BERT** μπορούν επίσης να χρησιμοποιηθούν σε αυτή τη φάση, αλλά είναι σημαντικό να επιτευχθεί μια ισορροπία ανάμεσα στην ταχύτητα και τις υπολογιστικές απαιτήσεις.

Εφαρμογή Μοντέλου σε Πραγματικό Χρόνο

Αφού ολοκληρωθεί η εξαγωγή των χαρακτηριστικών, τα μοντέλα μηχανικής μάθησης αναλαμβάνουν να αναλύσουν και να ταξινομήσουν το περιεχόμενο ως αληθές ή ψευδές. Για να ανταποκριθεί το σύστημα σε πραγματικό χρόνο, τα μοντέλα πρέπει να είναι ελαφριά και γρήγορα. Συχνά χρησιμοποιούνται μικρότερες και πιο βελτιστοποιημένες εκδόσεις μεγαλύτερων μοντέλων, ώστε να διασφαλιστεί η αποδοτικότητα. Η χρήση μικρότερων εφαρμογών επιτρέπει την εύκολη κλιμάκωση του συστήματος και την αναβάθμισή του, όταν χρειάζεται.

Ειδοποιήσεις και Ανατροφοδότηση

Μετά την ανάλυση των δεδομένων, το σύστημα στέλνει ειδοποιήσεις στους διαχειριστές της πλατφόρμας ή στους χρήστες, ανάλογα με τη ρύθμιση. Μπορεί επίσης να ενεργοποιήσει αυτοματοποιημένες ενέργειες, όπως η σήμανση μιας ανάρτησης ή η μείωση της προβολής της στο κοινό. Ο μηχανισμός ανατροφοδότησης συλλέγει δεδομένα σχετικά με την απόδοση του συστήματος, που χρησιμοποιούνται για συνεχή βελτίωση και εκπαίδευση των μοντέλων.

Κλιμάκωση και Αποδοτικότητα

Για να διαχειριστεί τις αυξανόμενες απαιτήσεις, το σύστημα πρέπει να είναι σχεδιασμένο ώστε να κλιμακώνεται εύκολα. Τεχνολογίες όπως το **Docker** και το **Kubernetes** επιτρέπουν την εύκολη ανάπτυξη και την επέκταση των δυνατοτήτων του συστήματος. Οι υπηρεσίες cloud προσφέρουν την απαραίτητη υπολογιστική ισχύ για να εξασφαλίσουν ότι το σύστημα παραμένει αποδοτικό, ακόμη και σε περιόδους υψηλής ζήτησης.

Ασφάλεια και Απόρρητο

Η επεξεργασία δεδομένων σε πραγματικό χρόνο απαιτεί αυστηρά μέτρα ασφάλειας και προστασίας προσωπικών δεδομένων. Το σύστημα πρέπει να διασφαλίζει ότι τα δεδομένα που συλλέγονται είναι κρυπτογραφημένα και ότι οι ευαίσθητες πληροφορίες ανωνυμοποιούνται. Παράλληλα, πρέπει να εφαρμόζονται έλεγχοι πρόσβασης ώστε να συμμορφώνονται με τους κανονισμούς, όπως το **GDPR**.

Κεφάλαιο 4: Πειραματικοί Μέθοδοι και Αποτελέσματα

4.1 Εισαγωγή

Αυτό το κεφάλαιο παρουσιάζει τη διαδικασία των πειραμάτων, τις μεθοδολογίες και τα αποτελέσματα που προέκυψαν από την εφαρμογή διαφόρων μοντέλων μηχανικής μάθησης στην ανίχνευση πλαστών δημοσιευμάτων. Ο κύριος στόχος των πειραμάτων ήταν η αξιολόγηση της απόδοσης τόσο των παραδοσιακών μοντέλων όσο και των μοντέλων βαθιάς μάθησης, όπως το BERT, σε ένα μεγάλο σύνολο δεδομένων με ψευδείς ειδήσεις μέσω μιας σειράς πειραμάτων που περιλάμβαναν διασταυρούμενη επικύρωση (cross-validation) και κλασικές διαχωρίσεις εκπαίδευσης-δοκιμής.

4.2 Πειραματική Διαδικασία

Για τις ανάγκες της εργασίας στήθηκαν τέσσερα διαφορετικά πειράματα όπου εξετάζεται η απόδοση μοντέλων ανίχνευσης με διαφορετικά text representations όπως για παράδειγμα Glove embeddings, TF_IDF vectorization, Bag of Words και το BertTokenizer για το μοντέλο Bert. Μέσα από αυτές τις δοκιμές μπορούμε να καταλάβουμε καλύτερα τις δυνατότητες κάθε μοντέλου και πώς μπορούμε να τις αξιοποιήσουμε. Τα προεπεξεργασμένα δεδομένα τροφοδοτήθηκαν σε διάφορα μοντέλα μηχανικής μάθησης για την εκπαίδευση και αξιολόγηση τους. Τα μοντέλα που δοκιμάστηκαν είναι τα Support Vector Machine (SVM), Logistic Regression, Random Forest, Naive Bayes και BERT (Bidirectional Encoder Representations from Transformers).

4.2.1 Σύνολο Δεδομένων

Το σύνολο δεδομένων που χρησιμοποιήθηκε σε αυτή τη μελέτη είναι το **WELFake dataset**, το οποίο περιλαμβάνει 72.134 ειδησεογραφικά άρθρα, εκ των οποίων 35.028 είναι αληθινά και 37.106 ψευδείς ειδήσεις. Το σύνολο δεδομένων δημιουργήθηκε από τη συγχώνευση τεσσάρων δημοφιλών συνόλων ειδήσεων (Kaggle, McIntire, Reuters, BuzzFeed Political), ώστε να αποφευχθεί η υπερπροσαρμογή των ταξινομητών και να παρασχεθούν περισσότερα δεδομένα για καλύτερη εκπαίδευση των μοντέλων μηχανικής μάθησης.

Το σετ δεδομένων περιλαμβάνει τέσσερις στήλες: **Αριθμό Σειράς**, **Τίτλο** (επικεφαλίδα του άρθρου), **Κείμενο** (περιεχόμενο της είδησης), και **Ετικέτα** (0 για ψευδείς και 1 για αληθινές ειδήσεις).

Επέλεξα αυτό το σύνολο δεδομένων λόγω της πληρότητάς του όσον αφορά τα διάφορα είδη ειδησεογραφικού περιεχομένου και της ισορροπημένης κατανομής αληθινών και ψευδών ειδήσεων. Το σύνολο δεδομένων χωρίστηκε σε δύο μέρη, με το 70% να χρησιμοποιείται για την εκπαίδευση των μοντέλων και το υπόλοιπο 30% για τη δοκιμή.

4.2.2 Προεπεξεργασία και εξαγωγή χαρακτηριστικών

Η προεπεξεργασία είναι ένα κρίσιμο βήμα για την προετοιμασία των δεδομένων για την εκπαίδευση των μοντέλων. Για τις δοκιμές που έγιναν εφαρμόστηκαν διαφορετικοί τρόποι προεπεξεργασίας.

- **Καθαρισμός Κειμένου:** Αρχικά όλα τα κείμενα μετατράπηκαν σε πεζά για να διασφαλιστεί η ομοιομορφία. Αφαιρέθηκαν τα σημεία στίξης και οι άχρηστες λέξεις (stopwords) μέσω της βιβλιοθήκης NLTK.
- **GloVe:** Για την περίπτωση με το text representation Glove, χρησιμοποίησα pretrained embeddings για την αναπαράσταση των λέξεων. Επέλεξα το Wikipedia 2014 + Gigaword 5 (6B tokens, 400K vocab, uncased, 300d vectors) του GloVe. Το συγκεκριμένο περιλαμβάνει τα embeddings για 400.000 λέξεις. Στη συνέχεια, χρησιμοποίησα το GloVe για να αντιστοιχίσω λέξεις από το σύνολο δεδομένων σε αυτά τα προ εκπαιδευμένα embeddings. Αναλυτικότερα, δημιουργήθηκαν αρχεία όπως τα **cooccurrence.bin** και **cooccurrence.shuf.bin**. Αυτά τα αρχεία περιέχουν τους συνδυασμούς λέξεων και τις αντίστοιχες συχνότητες εμφάνισής τους στα δεδομένα που έδωσα στο GloVe. Το αρχείο **cooccurrence.bin** είναι το αποτέλεσμα της πρώτης φάσης εκπαίδευσης του GloVe, ενώ το **cooccurrence.shuf.bin** είναι το ίδιο αρχείο αλλά με τυχαία αναδιάταξη των δεδομένων για καλύτερη εκπαίδευση. Τέλος, δημιούργησα το αρχείο **vectors.txt** το οποίο περιέχει τα εκπαιδευμένα word embeddings σε μορφή κειμένου. Αυτό σημαίνει ότι για κάθε λέξη του λεξιλογίου υπάρχει μια αντίστοιχη σειρά από αριθμούς (διαστάσεων 300) που αντιπροσωπεύουν την αριθμητική αναπαράστασή της. Το **vectors.bin** είναι το δυαδικό αρχείο που περιέχει τα ίδια embeddings, αλλά σε συμπιεσμένη μορφή για ταχύτερη φόρτωση. Αυτά τα αρχεία ήταν κρίσιμα για να χρησιμοποιήσω το GloVe. Με τα **vectors.txt** και **vocab.txt**, φλορτωσα τα embeddings στις διάφορες αναπαραστάσεις κειμένου και τα χρησιμοποίησα για να εκπαιδεύσω τα μοντέλα ανίχνευσης ψευδών ειδήσεων.
- **Bag of Words (BoW):** Για την περίπτωση του Bag of Words (BoW), έγινε η μετατροπή των κειμένων σε διανύσματα χρησιμοποιώντας τη μέθοδο όπου το κείμενο μετατρέπεται σε αριθμητική αναπαράσταση βασισμένη στη συχνότητα που κάθε λέξη εμφανίζεται στα δεδομένα.
- **Διανυσματοποίηση TF-IDF:** Ανάλογα στήθηκε και η εκπαίδευση με TF-IDF vectorization. Στη περίπτωση αυτή, για τη μεταποίηση του κειμένου σε διανύσματα χαρακτηριστικών, χρησιμοποιείται ο TfidfVectorizer της βιβλιοθήκης **sklearn**.
- **Bert tokenizer:** Το προηγμένο μοντέλο BERT χρησιμοποιεί τον προκαθορισμένο tokenizer του για την κωδικοποίηση των κειμένων και δημιουργεί ένα DataLoader για την εκπαίδευση και τη δοκιμή του μοντέλου. Στη συνέχεια, εκπαιδεύει το μοντέλο BERT με τη χρήση της συνάρτησης απώλειας CrossEntropyLoss και του βελτιστοποιητή AdamW.
- **Κωδικοποίηση Ετικετών:** Οι ετικέτες (ψευδείς ή αληθινές ειδήσεις) μετατράπηκαν σε δυαδική μορφή, όπου το 0 αντιπροσωπεύει τις ψευδείς ειδήσεις και το 1 τις αληθινές.

4.2.3 Cross-Validation

Για να διασφαλιστεί η εγκυρότητα και η αξιοπιστία των αποτελεσμάτων εφαρμόστηκε μια διασταυρωμένη επικύρωση. Αυτή η μέθοδος χωρίζει τα δεδομένα εκπαίδευσης σε διάφορα κομμάτια και εκπαιδεύει το μοντέλο σε διαφορετικά υποσύνολα κάθε φορά. Αυτό βοηθά στο να αποφύγουμε τη υπερβολική προσαρμογή (overfitting) του μοντέλου, κάνοντάς το πιο ικανό να γενικεύει τα αποτελέσματά του. Στη συγκεκριμένη περίπτωση, χρησιμοποιήσα διασταυρούμενη επικύρωση 5 διασταυρώσεων (5-fold cross-validation).

4.2.4 Μετρικοί Δείκτες Αξιολόγησης

Για να αξιολογηθεί η απόδοση των μοντέλων, χρησιμοποιήθηκαν οι εξής μετρικές:

- **Ακρίβεια (Accuracy):** Το ποσοστό των σωστά ταξινομημένων περιπτώσεων σε σχέση με το σύνολο των περιπτώσεων.
- **Ακρίβεια Επαναλήψεων (Precision):** Το ποσοστό των πραγματικά θετικών περιπτώσεων από τις περιπτώσεις που ταξινομήθηκαν ως θετικές.
- **Ανάκληση (Recall):** Το ποσοστό των πραγματικών θετικών περιπτώσεων που εντοπίστηκαν από το μοντέλο.
- **F1 Score:** Ο αρμονικός μέσος της ακρίβειας και της ανάκλησης, που δίνει μια ισορροπημένη μέτρηση της απόδοσης.

4.3 Πειραματική Διαδικασία

Αυτή η ενότητα περιγράφει τη διαδικασία των πειραμάτων, από την προετοιμασία των δεδομένων, την εκπαίδευση των μοντέλων, τη διασταυρούμενη επικύρωση, έως την τελική αξιολόγηση. Τα βήματα εκτελέστηκαν συστηματικά για να διασφαλιστεί η ακρίβεια και η αξιοπιστία των αποτελεσμάτων.

4.3.1 Προετοιμασία Δεδομένων

Η προετοιμασία των δεδομένων περιλάμβανε μερικά βασικά βήματα για να διασφαλιστεί ότι το σύνολο δεδομένων ήταν έτοιμο για εκπαίδευση και δοκιμή. Αυτά τα βήματα ακολουθήθηκαν σε όλα τα πειράματα που διεξάχθηκαν.

1. **Φόρτωση Συνόλου Δεδομένων:** Το σετ δεδομένων WELFake φορτώθηκε με τη χρήση της βιβλιοθήκης **Pandas** και ελέγχθηκε για τυχόν κενές τιμές, οι οποίες αφαιρέθηκαν για να διατηρηθεί η ακρίβεια των δεδομένων.
2. **Προεπεξεργασία Κειμένου:** Τα κείμενα μετατράπηκαν σε πεζά γράμματα, αφαιρέθηκαν σημεία στίξης και άχρηστες λέξεις (stopwords) χρησιμοποιώντας τη βιβλιοθήκη **NLTK**. Αυτό βοήθησε στη δημιουργία καθαρού και τυποποιημένου κειμένου για τα μοντέλα.
3. **Αφαίρεση χαρακτηριστικών:** Πραγματοποιήθηκε αφαίρεση των χαρακτηρισμών με τους τρόπους που αναφέρθηκαν παραπάνω.
4. **Διαχωρισμός Δεδομένων:** Το σετ χωρίζεται σε test και train (70-30%).

4.3.2 Εκπαίδευση Μοντέλων

Η διαδικασία εκπαίδευσης διέφερε ελαφρώς μεταξύ των παραδοσιακών μοντέλων μηχανικής μάθησης και του μοντέλου BERT λόγω των διαφορετικών αναγκών τους. Κάθε εκπαιδευμένο μοντέλο αποθηκεύτηκε για μελλοντική χρήση.

- **Παραδοσιακά Μοντέλα Μηχανικής Μάθησης:** Μοντέλα όπως το SVM, η Λογιστική Παλινδρόμηση, το Random Forest, και το Naive Bayes εκπαιδεύτηκαν χρησιμοποιώντας Bow, GloVe και διανυσματοποίηση TF-IDF. Η διασταυρούμενη επικύρωση εφαρμόστηκε για να εξασφαλίσει ότι τα αποτελέσματα δεν ήταν τυχαία.
 - **Διασταυρούμενη Επικύρωση:** Χρησιμοποιήθηκε η τεχνική 5-fold cross-validation, όπου τα δεδομένα εκπαίδευσης χωρίζονται σε πέντε μέρη. Το μοντέλο εκπαιδεύεται σε τέσσερα μέρη και επικυρώνεται στο πέμπτο, με τη διαδικασία να επαναλαμβάνεται για καλύτερη γενίκευση.
- **Μοντέλο BERT:** Το μοντέλο BERT απαιτούσε διαφορετική προσέγγιση. Τα δεδομένα μετατράπηκαν σε διανύσματα με τη βοήθεια του BERT tokenizer, και το μοντέλο εκπαιδεύτηκε με batch size 32 για τρεις εποχές (epochs).
 - **Εκπαίδευση BERT:** Η εκπαίδευση του BERT έγινε με την υποστήριξη GPU, που επέτρεψε γρήγορη προσαρμογή του μοντέλου και ελαχιστοποίηση της απώλειας κατά τη διάρκεια της εκπαίδευσης.

4.3.3 Διαχείριση Υπολογιστικών Πόρων

Λόγω του μεγάλου μεγέθους των δεδομένων και της πολυπλοκότητας του μοντέλου BERT, η σωστή χρήση των υπολογιστικών πόρων ήταν κρίσιμη.

- **Χρήση GPU:** Η εκπαίδευση του BERT έγινε με τη χρήση GPU για μείωση του χρόνου εκπαίδευσης και αποτελεσματική διαχείριση της υπολογιστικής επιβάρυνσης.
- **Διαχείριση Μνήμης:** Χρησιμοποιήθηκαν mini-batches για καλύτερη διαχείριση της μνήμης κατά τη διάρκεια της εκπαίδευσης του BERT.
- **Αξιοποίηση CPU:** Για τα παραδοσιακά μοντέλα, χρησιμοποιήθηκαν τεχνικές παράλληλης επεξεργασίας με τη χρήση όλων των διαθέσιμων πυρήνων, βελτιώνοντας την ταχύτητα εκτέλεσης και την απόδοση.

4.4 Ανάλυση αποτελεσμάτων

Σε αυτό το σημείο της εργασίας θα δούμε αναλυτικά τα αποτελέσματα των μοντέλων και θα προσπαθήσουμε να αξιολογήσουμε το κάθε μοντέλο στη κάθε περίπτωση αναπαράστασης κειμένου. Τα αποτελέσματα επίσης θα παρουσιαστούν σε μορφή πινάκων όπου θα φαίνεται το Accuracy, Precision, Recall, και F1 σκορ.

4.4.1 Traditional Models Performance

Αποτελέσματα με BoW (Bag-of-Words)

Τα αποτελέσματα από τη χρήση της μεθόδου Bag-of-Words δείχνουν ξεκάθαρα ότι το **SVM** ήταν το πιο αποδοτικό μοντέλο, καταφέρνοντας να φτάσει σε ακρίβεια **0.934**, με precision **0.935** και F1 score **0.934**. Αυτή η απόδοση δείχνει ότι το SVM μπόρεσε να αναλύσει αποτελεσματικά τα δεδομένα και να ξεχωρίσει τις ψευδείς ειδήσεις με ακρίβεια, διαχειριζόμενο σωστά την πολυπλοκότητα των κειμένων.

Το **Logistic Regression** ακολούθησε στενά με ακρίβεια **0.937** και F1 score **0.938**. Παρόλο που υπολείπεται λίγο σε σχέση με το SVM, παραμένει σταθερά αξιόπιστο στη διάκριση των πραγματικών από τις ψευδείς ειδήσεις, διατηρώντας καλή ισορροπία ανάμεσα στο precision και το recall.

Το **Random Forest** είχε επίσης ισχυρή απόδοση με ακρίβεια **0.930**, αλλά ξεχώρισε περισσότερο στο recall με **0.944**, κάτι που δείχνει ότι το μοντέλο αυτό είναι καλό στο να εντοπίζει τις πραγματικά ψευδείς ειδήσεις, ακόμα και αν αυτό σημαίνει ότι θυσιάζει λίγο την ακρίβεια για να είναι πιο ευαίσθητο.

Τέλος, το **Naive Bayes** δεν μπόρεσε να συμβαδίσει με τα υπόλοιπα μοντέλα, καθώς είχε ακρίβεια **0.8** και F1 score **0.86**. Όπως αναμενόταν, αυτό οφείλεται στο γεγονός ότι το Naive Bayes κάνει απλοϊκές υποθέσεις για την ανεξαρτησία των χαρακτηριστικών, οι οποίες δεν ισχύουν πάντα στα πραγματικά δεδομένα.

Αυτά τα αποτελέσματα δείχνουν ότι, ενώ τα παραδοσιακά μοντέλα όπως το **SVM** και το **Logistic Regression** είναι πιο ικανά στο να αναγνωρίζουν την πολυπλοκότητα των κειμένων, το **Random Forest** ξεχωρίζει όταν η προτεραιότητα είναι να ανιχνεύσει όλες τις πιθανές περιπτώσεις ψευδών ειδήσεων. Από την άλλη πλευρά, το **Naive Bayes**, αν και γρήγορο και απλό, δεν αποδίδει το ίδιο καλά σε πιο σύνθετα σύνολα δεδομένων.

Πίνακας 4: Αξιολόγηση μοντέλων με BoW

Model	Accuracy	Precision	Recall	F1 Score
SVM	0.9344	0.9353	0.9344	0.9348
Logistic Regression	0.9373	0.9322	0.9442	0.9382
Random Forest	0.9301	0.9189	0.9446	0.9316
Naive Bayes	0.8674	0.8683	0.8686	0.8684

Αποτελέσματα με GloVe Embeddings

Η χρήση των GloVe embeddings είχε σημαντική επίδραση στην απόδοση των μοντέλων, επιτρέποντας τους να εκμεταλλευτούν πιο πλούσιες αναπαραστάσεις λέξεων και να κατανοήσουν καλύτερα τις σχέσεις και τις δομές του κειμένου. Το **SVM** ξεχώρισε και εδώ με την υψηλότερη ακρίβεια στο **(0.9343)** και πολύ καλό F1 score στο **(0.9348)**. Αυτά τα αποτελέσματα δείχνουν ότι το **SVM** αξιοποίησε στο έπακρο την πληροφορία που προσφέρουν τα GloVe embeddings, κάτι που του επέτρεψε να διαχωρίσει τις ψευδείς ειδήσεις από τις πραγματικές με μεγαλύτερη ακρίβεια.

Παράλληλα, το **Random Forest** είχε αξιόλογη απόδοση με υψηλή ανάκληση στο **(0.9301)**. Η υψηλή ανάκληση σημαίνει ότι το **Random Forest** ήταν πολύ καλό στο να αναγνωρίζει τα αληθινά ψευδείς ειδήσεις, αν και αυτό μπορεί να συνοδεύεται από κάποιες απώλειες στο **precision**. Αυτό το αποτέλεσμα ήταν αναμενόμενο, δεδομένου ότι το **Random Forest** είναι γνωστό για την ικανότητά του να επεξεργάζεται μεγάλα σύνολα δεδομένων και να αξιοποιεί διαφορετικά χαρακτηριστικά εισόδου.

Το **Logistic Regression**, αν και δεν έφτασε τα επίπεδα του **SVM**, διατήρησε μια σταθερή απόδοση με ακρίβεια **(0.9314)** και **F1 score (0.9320)**. Αυτό δείχνει ότι είναι ακόμα ένα αξιόπιστο μοντέλο για ταξινόμηση, ειδικά όταν χρησιμοποιεί αναπαραστάσεις δεδομένων υψηλής ποιότητας όπως τα GloVe embeddings.

Συνολικά, τα αποτελέσματα αυτά επιβεβαιώνουν ότι τα GloVe embeddings παρέχουν μια σημαντική βελτίωση στις επιδόσεις των μοντέλων, βοηθώντας τα να κατανοήσουν καλύτερα τις πολύπλοκες γλωσσικές σχέσεις και να κάνουν πιο ακριβείς προβλέψεις.

Πίνακας 5: Αξιολόγηση μοντέλων με GloVe

<i>Model</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
<i>SVM</i>	0.9343	0.9352	0.9344	0.9348
<i>Logistic Regression</i>	0.9314	0.9316	0.9324	0.9320
<i>Random Forest</i>	0.9301	0.9189	0.9445	0.9316

Αποτελέσματα με TF-IDF

Η χρήση των TF-IDF αναπαραστάσεων οδήγησε σε εξαιρετικές επιδόσεις από τα παραδοσιακά μοντέλα, με το **SVM** να καταγράφει πάλι υψηλή ακρίβεια στο **(0.9513)** και πολύ καλό **F1 score** στο **(0.9518)**. Το μοντέλο αυτό εκμεταλλεύτηκε πλήρως τη δυνατότητα του TF-IDF να εντοπίζει σημαντικές λέξεις και φράσεις που σχετίζονται με την ψευδή πληροφορία, προσφέροντας έτσι ακριβείς και αξιόπιστες προβλέψεις.

Το **Random Forest** επίσης είχε αξιόλογη απόδοση με μεγαλύτερη ακρίβεια **(0.9538)** και υψηλό **Recall** στο **(0.9644)**, δείχνοντας την ικανότητά του να εντοπίζει μεγάλο αριθμό θετικών περιπτώσεων (ψευδών ειδήσεων) χωρίς να θυσιάζει την ακρίβεια.

Αυτό το μοντέλο φαίνεται να επωφελείται από την ισχυρή δόμηση των δεδομένων που προσφέρει το TF-IDF, κάνοντάς το ιδιαίτερα αποτελεσματικό στην ανίχνευση της παραπληροφόρησης.

Το **Logistic Regression** παρόλο που είχε ελαφρώς χαμηλότερη απόδοση σε σύγκριση με τα άλλα δύο μοντέλα, διατήρησε μια αξιοσημείωτη ακρίβεια στο (**0.9463**) και **F1 score (0.9470)**, γεγονός που αποδεικνύει ότι είναι ακόμα ένα αξιόπιστο εργαλείο για ταξινόμηση κειμένων.

Το **Naive Bayes**, όπως και σε άλλες αναπαραστάσεις, σημείωσε τη χαμηλότερη απόδοση με ακρίβεια **0.8436** και **F1 score (0.8475)**. Αυτό ήταν αναμενόμενο, καθώς το μοντέλο αυτό έχει περιορισμούς όταν πρόκειται για πολύπλοκες σχέσεις μεταξύ των λέξεων, που συνήθως ενισχύονται από τις TF-IDF αναπαραστάσεις.

Πίνακας 6: Αξιολόγηση μοντέλων με TF-IDF

<i>Model</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
SVM	0.9513	0.9486	0.9551	0.9518
Logistic Regression	0.9463	0.9414	0.9527	0.9470
Random Forest	0.9538	0.9450	0.9644	0.9546
Naive Bayes	0.8436	0.8323	0.8633	0.8475

4.4.2 BERT

Το BERT σημείωσε εξαιρετική απόδοση με ακρίβεια (**0.9782**) και **F1 score (0.9789)**. Αυτό το αποτέλεσμα ήταν αναμενόμενο, καθώς το BERT χρησιμοποιεί προχωρημένες τεχνικές επεξεργασίας φυσικής γλώσσας που του επιτρέπουν να κατανοεί βαθύτερα τη σημασία και το πλαίσιο των λέξεων μέσα στο κείμενο. Η πολύ υψηλή τιμή **recall (0.991)** δείχνει ότι το BERT είναι εξαιρετικό στην ανίχνευση των πραγματικά θετικών περιπτώσεων, κάνοντάς το ιδανικό για την αναγνώριση ψευδών ειδήσεων. Τα αποτελέσματα αυτά υποδεικνύουν ότι το BERT είναι το πιο ακριβές μοντέλο, αν και απαιτεί περισσότερους υπολογιστικούς πόρους για να λειτουργήσει.

4.4.3 Συγκριτική ανάλυση

Τα αποτελέσματα της ανάλυσης επιβεβαιώνουν αρκετές προσδοκίες σχετικά με την απόδοση των διαφορετικών μοντέλων στην ανίχνευση ψευδών ειδήσεων. Παρατηρήσαμε ότι τα πιο προηγμένα μοντέλα, όπως το **BERT**, έχουν σαφές πλεονέκτημα σε σχέση με τα παραδοσιακά μοντέλα, και αυτό οφείλεται κυρίως στην ικανότητά τους να κατανοούν σε βάθος το πλαίσιο και τις σύνθετες δομές του κειμένου.

Το **BERT** σημείωσε τα υψηλότερα αποτελέσματα σε όλες τις μετρικές, κάτι που ήταν αναμενόμενο λόγω της εξελιγμένης αρχιτεκτονικής του. Τα παραδοσιακά μοντέλα, όπως το **SVM** και το **Random Forest**, είχαν πολύ καλή απόδοση όταν χρησιμοποιήθηκαν με το **TF-IDF** και

GloVe embeddings. Αυτά τα μοντέλα είναι γνωστά για την ικανότητά τους να εκμεταλλεύονται την ποιότητα των χαρακτηριστικών εισόδου. Με τη χρήση των **TF-IDF** και **GloVe**, τα μοντέλα κατάφεραν να αναλύσουν με ακρίβεια τις λέξεις που σχετίζονται με την ψευδή πληροφορία, βελτιώνοντας έτσι τις προβλέψεις τους. Το **SVM** ειδικότερα είναι γνωστό για την ικανότητά του να διαχωρίζει τις κατηγορίες σε υψηλές διαστάσεις, κάτι που αποδείχθηκε χρήσιμο σε αυτό το πρόβλημα ταξινόμησης. Τα χαμηλότερα αποτελέσματα του **Naive Bayes** δεν αποτελούν επίσης έκπληξη, καθώς αυτό το μοντέλο βασίζεται στην υπόθεση ότι όλα τα χαρακτηριστικά είναι ανεξάρτητα μεταξύ τους. Αυτή η προσέγγιση λειτουργεί καλά για απλά προβλήματα, αλλά δεν μπορεί να κατανοήσει τις περίπλοκες σχέσεις και αλληλεπιδράσεις μεταξύ των λέξεων που υπάρχουν στα δεδομένα ψευδών ειδήσεων. Για παράδειγμα, ο τρόπος με τον οποίο χρησιμοποιούνται οι λέξεις μπορεί να έχει σημαντική σημασία, και το Naive Bayes δυσκολεύεται να το αντιληφθεί αυτό, οδηγώντας σε περισσότερα λάθη στην πρόβλεψη.

Πίνακας 7: Συνολική αξιολόγηση μοντέλων

<i>Model</i>	<i>Representation</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
<i>SVM</i>	Bag-of-Words	0.9344	0.9353	0.9344	0.9348
<i>Logistic Regression</i>	Bag-of-Words	0.9315	0.9316	0.9324	0.9320
<i>Random Forest</i>	Bag-of-Words	0.9301	0.9189	0.9446	0.9316
<i>Naive Bayes</i>	Bag-of-Words	0.8900	0.8800	0.9000	0.8900
<i>SVM</i>	GloVe	0.9344	0.9353	0.9344	0.9348
<i>Logistic Regression</i>	GloVe	0.9315	0.9316	0.9324	0.9320
<i>Random Forest</i>	GloVe	0.9301	0.9189	0.9446	0.9316
<i>Naive Bayes</i>	GloVe	0.8900	0.8800	0.9000	0.8900
<i>SVM</i>	TF-IDF	0.9514	0.9487	0.9551	0.9512
<i>Logistic Regression</i>	TF-IDF	0.9464	0.9414	0.9527	0.9470
<i>Random Forest</i>	TF-IDF	0.9539	0.9450	0.9644	0.9546
<i>Naive Bayes</i>	TF-IDF	0.8436	0.8323	0.8633	0.8476
<i>BERT</i>	-	0.9782	0.9669	0.9912	0.9789

4.5 Συγκριτική ανάλυση με υπάρχουσα βιβλιογραφία

4.5.1 Μοντέλα Μηχανικής Μάθησης

Οι πίνακες που παρατίθενται παρακάτω συγκρίνουν την απόδοση διάφορων μοντέλων μηχανικής μάθησης σε ένα ενοποιημένο σύνολο δεδομένων, χρησιμοποιώντας διαφορετικές μεθόδους εξαγωγής χαρακτηριστικών όπως το Bag of Words, GloVe και Word2Vec Vectorizer και το TF-IDF. Τα αποτελέσματα αυτά τοποθετούνται σε σύγκριση με τα πλέον σύγχρονα μοντέλα που έχουν αναπτυχθεί τα τελευταία χρόνια.

Πίνακας 8: Απόδοση διαφορετικών μεθόδων μηχανικής μάθησης με χρήση του Bag of Words

<i>Model</i>	<i>Accuracy</i>	<i>Recall</i>	<i>Precision</i>	<i>F1-Score</i>	<i>AUC-ROC</i>
<i>Logistic Regression</i>	0.904	0.892	0.902	0.897	0.889
<i>Bernoulli NB</i>	0.881	0.858	0.891	0.875	0.880
<i>Multinomial NB</i>	0.890	0.882	0.891	0.886	0.889
<i>Gaussian NB</i>	0.888	0.862	0.904	0.882	0.887
<i>Random Forest</i>	0.892	0.886	0.891	0.888	0.889
<i>Decision Tree</i>	0.889	0.886	0.887	0.886	0.887
<i>XGBoost</i>	0.920	0.889	0.950	0.916	0.921
<i>SVM (linear kernel)</i>	0.909	0.886	0.924	0.904	0.908

Σύμφωνα με τα παραπάνω αποτελέσματα, το **XGBoost** και το **SVM (linear kernel)** είχαν την καλύτερη απόδοση μεταξύ των παραδοσιακών μοντέλων με τη χρήση του Bag of Words, με τα υψηλότερα ποσοστά ακρίβειας και ισορροπημένο F1-Score, γεγονός που δείχνει ότι αυτά τα μοντέλα κατάφεραν να κατανοήσουν καλύτερα τα δεδομένα και να δώσουν ακριβείς προβλέψεις.

Πίνακας 9: Απόδοση διαφορετικών μεθόδων μηχανικής μάθησης με χρήση του TF-IDFTF-IDF Feature

<i>Model</i>	<i>Accuracy</i>	<i>Recall</i>	<i>Precision</i>	<i>F1-Score</i>	<i>AUC-ROC</i>
<i>Logistic Regression</i>	0.913	0.887	0.931	0.908	0.856
<i>Bernoulli NB</i>	0.877	0.860	0.883	0.871	0.876
<i>Multinomial NB</i>	0.856	0.848	0.855	0.851	0.856
<i>Gaussian NB</i>	0.888	0.862	0.904	0.882	0.887
<i>Random Forest</i>	0.873	0.873	0.868	0.870	0.856
<i>Decision Tree</i>	0.889	0.886	0.887	0.886	0.856
<i>XGBoost</i>	0.920	0.889	0.950	0.916	0.922
<i>SVM (linear kernel)</i>	0.923	0.888	0.949	0.917	0.921

Τα αποτελέσματα με τη χρήση του TF-IDF δείχνουν ότι το **SVM** με **TF-IDF** ξεχωρίζει, επιτυγχάνοντας την υψηλότερη ακρίβεια και F1-Score, κάτι ανάλογο με τα αποτελέσματα των δοκιμών που έγιναν για την πτυχιακή.

Πίνακας 10: Σύγκριση μεθόδων μηχανικής μάθησης με σύγχρονα μοντέλα

<i>Model/Year</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
<i>SVM (2015)</i>	0.650	0.640	0.650	0.630
<i>CPCV model (2016)</i>	0.840	0.918	0.933	0.853
<i>LSVM (2017)</i>	0.920	—	—	—
<i>LFD model (2020)</i>	0.860	—	—	—
<i>Naïve Bayes (2020)</i>	0.920	0.960	0.970	0.930
<i>Randomized DT (2020)</i>	0.910	0.920	0.900	0.902
<i>KNN model (2021)</i>	0.920	0.930	0.900	0.910
<i>Naïve Bayes (2021)</i>	0.600	0.590	0.600	0.590
<i>Naïve Bayes (2021)</i>	0.860	0.860	0.860	0.860
<i>Logistic Regression (2021)</i>	0.850	0.840	0.870	0.850
<i>SVM + TF-IDF(2021)</i>	0.923	0.949	0.888	0.917

Η ανάλυση των αποτελεσμάτων από τα μοντέλα μηχανικής μάθησης που εφαρμόστηκαν σε άλλες έρευνες αναδεικνύει ενδιαφέρουσες πληροφορίες για το πόσο αποτελεσματικές είναι οι διάφορες προσεγγίσεις στην κατηγοριοποίηση κειμένου. Το **SVM** με **TF-IDF** αναδείχθηκε και εκεί ως το κορυφαίο μοντέλο, με την υψηλότερη ακρίβεια (**0.923**) και ισορροπημένο F1-Score (**0.917**). Αυτό σημαίνει ότι το μοντέλο δεν κάνει απλώς σωστές προβλέψεις, αλλά καταφέρνει να διατηρεί μια καλή ισορροπία ανάμεσα στην ακρίβεια και την ανάκληση, κάτι που είναι σημαντικό όταν θέλουμε να ελαχιστοποιήσουμε τόσο τα ψευδώς θετικά όσο και τα ψευδώς αρνητικά αποτελέσματα.

Συγκρίνοντας τις τεχνικές εξαγωγής χαρακτηριστικών, φάνηκε και εδώ ότι η επιλογή της μεθόδου μπορεί να επηρεάσει σημαντικά την απόδοση του μοντέλου. Τα μοντέλα που χρησιμοποίησαν **TF-IDF** είχαν συστηματικά καλύτερες επιδόσεις σε σχέση με αυτά που βασίστηκαν στο **Bag of Words** κάτι που παρατηρήθηκε και στα δικά μου πειράματα.

Από τα επιπλέον μοντέλα που δοκιμάστηκαν στις άλλες έρευνες φάνηκε πόσο πολύ έχει προχωρήσει η μηχανική μάθηση τα τελευταία χρόνια. Το **SVM** με **TF-IDF** που αναπτύχθηκε εδώ ξεπερνά αρκετά προηγούμενα κορυφαία μοντέλα. Τα παλιότερα μοντέλα, όπως το **SVM (2015)** και το **CPCV (2016)**, είχαν χαμηλότερες επιδόσεις, δείχνοντας ότι τα σύγχρονα μοντέλα επωφελούνται από τις νέες τεχνικές και τους πιο εξελιγμένους αλγόριθμους.

Συνολικά, η μελέτη δείχνει ότι οι σύγχρονες μέθοδοι που συνδυάζουν αλγόριθμους όπως το **SVM** με τεχνικές όπως το **TF-IDF** μπορούν να προσφέρουν κορυφαία αποτελέσματα στην κατηγοριοποίηση κειμένου.

4.5.2 Μοντέλα Βαθιάς Μάθησης

Οι παρακάτω πίνακες παρουσιάζουν την απόδοση διαφόρων μοντέλων Βαθιάς Μάθησης σε ένα ενοποιημένο σύνολο δεδομένων, χρησιμοποιώντας ενσωματώσεις λέξεων (**Word2Vec**, **GloVe** και **FastText**), και τα συγκρίνουν με τις επιδόσεις των πιο σύγχρονων προτύπων από προηγούμενες μελέτες.

Πίνακας 11: Απόδοση Μοντέλων Βαθιάς Μάθησης με Χρήση Word2Vec

<i>Model</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>AUC-ROC</i>
<i>LSTM</i>	0.944	0.944	0.948	0.946	0.987
<i>BiLSTM</i>	0.970	0.974	0.968	0.977	0.993
<i>CNN-LSTM</i>	0.969	0.966	0.974	0.970	0.991
<i>CNN-BiLSTM</i>	0.975	0.984	0.970	0.977	0.992

Πίνακας 12: Απόδοση Μοντέλων Βαθιάς Μάθησης με Χρήση GloVe

<i>Model</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>AUC-ROC</i>
<i>LSTM</i>	0.969	0.974	0.967	0.970	0.994
<i>BiLSTM</i>	0.972	0.973	0.974	0.974	0.995
<i>CNN-LSTM</i>	0.971	0.972	0.972	0.972	0.993
<i>CNN-BiLSTM</i>	0.972	0.981	0.965	0.973	0.994

Πίνακας 13: Απόδοση Μοντέλων Βαθιάς Μάθησης με Χρήση FastText

<i>Model</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>AUC-ROC</i>
<i>LSTM</i>	0.969	0.976	0.965	0.971	0.988
<i>BiLSTM</i>	0.968	0.965	0.965	0.965	0.993
<i>CNN-LSTM</i>	0.966	0.967	0.969	0.968	0.992
<i>CNN-BiLSTM</i>	0.972	0.980	0.966	0.973	0.993

Πίνακας 14: Σύγκριση προηγμένων μοντέλων

<i>Model/Year</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>AUC-ROC</i>
<i>CSI model (2017)</i>	0.893	—	—	—	—
<i>CSI model (2017)</i>	0.953	—	—	—	—
<i>CNN-based model (2018)</i>	—	0.922	0.9271	0.921	—
<i>CNN (2019)</i>	0.824	0.880	0.751	0.810	0.873
<i>CNN (2019)</i>	0.940	—	—	—	0.980
<i>MLP (ELMo) (2020)</i>	0.833	—	—	—	—
<i>MLP (ELMo) (2020)</i>	0.790	—	—	—	—
<i>RNN (2020)</i>	0.965	—	—	—	—
<i>CNN (2020)</i>	0.968	—	—	—	—
<i>Hybrid CNN-RNN (2021)</i>	0.600	0.590	0.600	0.590	—
<i>SciBERT (2021)</i>	0.690	0.720	0.690	—	—
<i>Longformer (2021)</i>	0.970	0.970	0.970	0.970	—
<i>CNN-BiLSTM + Word2Vec(2021)</i>	0.975	0.984	0.970	0.977	0.992

Από την ανάλυση των μοντέλων βαθιάς μάθησης, βλέπουμε ότι το **CNN-BiLSTM** ξεχωρίζει σταθερά ως η καλύτερη επιλογή. Το συγκεκριμένο μοντέλο αποδίδει εξαιρετικά σε όλους τους δείκτες, ανεξάρτητα από την ενσωμάτωση λέξεων που χρησιμοποιείται. Για παράδειγμα, με το **Word2Vec**, το **CNN-BiLSTM** έφτασε ακρίβεια (**0.975**) και **F1-score (0.977)**, κάτι που δείχνει την ικανότητά του να χειρίζεται αποτελεσματικά σύνθετες δομές στο κείμενο.

Αυτό το μοτίβο συνεχίζεται και με άλλες ενσωματώσεις λέξεων, όπως το **GloVe** και το **FastText**, όπου το **CNN-BiLSTM** διατηρεί τη θέση του ως το πιο ισχυρό. Η ισορροπία μεταξύ **precision** και **recall** είναι αξιοσημείωτη, πράγμα που σημαίνει ότι το μοντέλο είναι τόσο ακριβές όσο και ικανό να αναγνωρίζει θετικά παραδείγματα χωρίς να δημιουργεί ψευδώς θετικά.

Όταν συγκρίνουμε αυτά τα μοντέλα με παλαιότερες μεθόδους, βλέπουμε μια σαφή πρόοδο στην απόδοση. Το **CNN-BiLSTM** με **Word2Vec** ξεπερνά κατά πολύ τις παλαιότερες προσεγγίσεις, όπως τα μοντέλα **CSI (2017)** και **CNN-based (2018-2019)**, επιτυγχάνοντας καλύτερα αποτελέσματα σε όλους τους βασικούς δείκτες.

Όπως αποδεικνύεται, το **CNN-BiLSTM** είναι η καλύτερη μέθοδος για ανάλυση κειμένου όταν συνδυάζεται με σύγχρονες ενσωματώσεις λέξεων. Αυτό το υβριδικό μοντέλο προσφέρει ένα ισχυρό εργαλείο για την επεξεργασία σύνθετων δεδομένων κειμένου, διασφαλίζοντας ακρίβεια και αξιοπιστία στις προβλέψεις του. Αυτά τα αποτελέσματα υπογραμμίζουν τη σημασία της χρήσης εξελιγμένων αρχιτεκτονικών νευρωνικών δικτύων σε συνδυασμό με προηγμένες τεχνικές ενσωμάτωσης λέξεων για τη βέλτιστη απόδοση στις εργασίες φυσικής γλώσσας.

Κεφάλαιο 5: Συμπεράσματα και Μελλοντική Έρευνα

5.1 Περίληψη Αποτελεσμάτων

Αυτή η διπλωματική εργασία εστίασε στην ανίχνευση ψευδών ειδήσεων, με την ανάπτυξη και αξιολόγηση τόσο παραδοσιακών μοντέλων μηχανικής μάθησης όσο και προηγμένων αρχιτεκτονικών βαθιάς μάθησης. Τα αποτελέσματα έδειξαν ότι ο συνδυασμός εξελιγμένων τεχνικών εξαγωγής χαρακτηριστικών με καλά βελτιστοποιημένα μοντέλα μπορεί να βελτιώσει σημαντικά την ακρίβεια και την αξιοπιστία των συστημάτων ανίχνευσης ψευδών ειδήσεων. Στα παραδοσιακά μοντέλα, όπως το SVM, το Logistic Regression και το Random Forest, χρησιμοποιήθηκαν χαρακτηριστικά όπως το Bag of Words και το TF-IDF. Μεταξύ αυτών, το SVM με TF-IDF ξεχώρισε με την υψηλότερη ακρίβεια και ισορροπημένο F1-score, υπερβαίνοντας τα μοντέλα αιχμής της βιβλιογραφίας και αναδεικνύοντας τη σημασία της εξαγωγής χαρακτηριστικών στην απόδοση των ταξινομητών.

Στον κλάδο της βαθιάς μάθησης, το μοντέλο **CNN-BiLSTM** αποδείχθηκε ιδιαίτερα αποτελεσματικό. Χρησιμοποιώντας διαφορετικές τεχνικές αναπαράστασης, όπως Word2Vec, GloVe και FastText, πέτυχε ανώτερα αποτελέσματα σε όλες τις μετρικές.

Η συγκριτική ανάλυση με την υπάρχουσα βιβλιογραφία έδειξε ότι τα μοντέλα που αναπτύχθηκαν σε αυτήν την εργασία δεν είναι μόνο ανταγωνιστικά, αλλά σε πολλές περιπτώσεις υπερτερούν των τρεχουσών προσεγγίσεων. Αυτό ήταν ιδιαίτερα αληθές για τα μοντέλα **SVM + TF-IDF** και **CNN-BiLSTM**, τα οποία σταθερά ξεπέρασαν τους αντιπάλους τους. Αυτά τα ευρήματα υπογραμμίζουν το δυναμικό των σύγχρονων μοντέλων μηχανικής και βαθιάς μάθησης στην αποτελεσματική αντιμετώπιση της πρόκλησης της ανίχνευσης ψευδών ειδήσεων.

5.2 Συνεισφορές

Αυτή η εργασία κάνει αρκετές σημαντικές συνεισφορές στον τομέα της ανίχνευσης ψευδών ειδήσεων. Πρώτον, ανέπτυξε και δοκίμασε με ακρίβεια τόσο παραδοσιακά μοντέλα μηχανικής μάθησης όσο και προηγμένες αρχιτεκτονικές βαθιάς μάθησης, παρέχοντας πολύτιμες γνώσεις για την αποτελεσματικότητά τους. Η ενσωμάτωση υβριδικών μοντέλων και η σύγκριση με τη βιβλιογραφία διεύρυνε την κατανόηση των πλεονεκτημάτων και των περιορισμών των διάφορων μεθοδολογιών στον τομέα αυτό. Επιπλέον, η εργασία αντιμετώπισε την πρόκληση της βελτιστοποίησης της χρήσης πόρων, διασφαλίζοντας ότι τα μοντέλα είναι όχι μόνο ακριβή αλλά και αποδοτικά για πραγματικές εφαρμογές. Συγκρίνοντας τα αποτελέσματα με την υπάρχουσα βιβλιογραφία, η εργασία κατέδειξε την υπεροχή των συγκεκριμένων μοντέλων **SVM + TF-IDF**, **BERT** και **CNN-BiLSTM**.

5.3 Μελλοντική Έρευνα

Παρά τα σημαντικά βήματα που έγιναν στην ανίχνευση ψευδών ειδήσεων, υπάρχουν ακόμα τομείς που αξίζουν περαιτέρω διερεύνηση. Η μελλοντική έρευνα θα μπορούσε να περιλαμβάνει την εφαρμογή αυτών των μοντέλων σε πιο ποικιλόμορφα και μεγαλύτερα σύνολα δεδομένων για την επιβεβαίωση της γενικευσιμότητάς τους. Μια άλλη ενδιαφέρουσα κατεύθυνση είναι η βελτίωση της επεξηγησιμότητας των μοντέλων βαθιάς μάθησης, ώστε να γίνουν πιο κατανοητά και αξιόπιστα. Επιπλέον, η υλοποίηση αυτών σε πραγματικό χρόνο για την άμεση ανίχνευση ψευδών ειδήσεων στις πλατφόρμες κοινωνικών δικτύων θα μπορούσε να είναι ένα σημαντικό βήμα προς τις πρακτικές εφαρμογές. Η επέκταση τους για να καλύπτουν πολλαπλές γλώσσες θα διεύρυνε ακόμα περισσότερο τις δυνατότητές τους, καθώς η παραπληροφόρηση είναι ένα παγκόσμιο πρόβλημα.

5.4 Τελικά Συμπεράσματα

Ο συνεχής αγώνας κατά της παραπληροφόρησης αναδεικνύει την ανάγκη για αποτελεσματικά και αξιόπιστα μοντέλα ανίχνευσης ψευδών ειδήσεων. Αυτή η εργασία απέδειξε ότι με τον κατάλληλο συνδυασμό τεχνικών εξαγωγής χαρακτηριστικών, αρχιτεκτονικών μοντέλων και βελτιστοποίησης των υπολογιστικών πόρων, μπορούμε να αναπτύξουμε συστήματα που είναι όχι μόνο ακριβή αλλά και κλιμακώσιμα για πραγματικές εφαρμογές. Καθώς το πεδίο συνεχίζει να εξελίσσεται, τα ευρήματα αυτής της έρευνας θα συμβάλλουν στην ανάπτυξη ακόμη πιο εξελιγμένων και αποτελεσματικών λύσεων για την συνεχή καταπολέμηση αυτού του προβλήματος.

Βιβλιογραφία

1. Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). *A Benchmark Study of Machine Learning Models for Online Fake News Detection*. arXiv:1905.04749v2 [cs.CL], 26 Mar 2021.
2. Mridha, M. F., Keya, A. J., Hamid, M. A., Monowar, M. M., & Rahman, M. S. (2021). *A Comprehensive Review on Fake News Detection With Deep Learning*. IEEE Access, 10.1109/ACCESS.2021.3129329.
3. Kaliyar, R., Goswami, A., & Narang, P. (2021). *A Hybrid Model for Effective Fake News Detection with a Novel COVID-19 Dataset*. In Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021) - Volume 2, pages 1066-1072. DOI: 10.5220/0010316010661072.
4. Prokopidis, P., & Piperidis, S. (2020). *A Neural NLP toolkit for Greek*. In 11th Hellenic Conference on Artificial Intelligence (SETN 2020), ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3411408.3411430>.
5. Zhou, X., & Zafarani, R. (2020). *A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities*. ACM Comput. Surv. 53, 5, Article 109 (September 2020), 40 pages. <https://doi.org/10.1145/3395046>.
6. Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). *FNDNet – A deep convolutional neural network for fake news detection*. Cognitive Systems Research, <https://doi.org/10.1016/j.cogsys.2019.12.005>.
7. Zhang, X., & Ghorbani, A. A. (2020). *An overview of online fake news: Characterization, detection, and discussion*. Information Processing and Management, 57, 102025. <https://doi.org/10.1016/j.ipm.2019.03.004>.
8. Jwa, H., Oh, D., Park, K., Kang, J. M., & Lim, H. (2019). *exBAKE: Automatic Fake News Detection Model Based on Bidirectional Encoder Representations from Transformers (BERT)*. Applied Sciences, 9(19), 4062. DOI: 10.3390/app9194062.
9. Aslam, N., Khan, I. U., Alotaibi, F. S., Aldaej, L. A., & Aldubaikil, A. K. (2021). *Fake Detect: A Deep Learning Ensemble Model for Fake News Detection*. Complexity, Volume 2021, Article ID 5557784. <https://doi.org/10.1155/2021/5557784>.
10. Thota, A., Tilak, P., Ahluwalia, S., & Lohia, N. (2018). *Fake News Detection: A Deep Learning Approach*. SMU Data Science Review, Vol. 1: No. 3, Article 10. Available at: <https://scholar.smu.edu/datasciencereview/vol1/iss3/10>.
11. University of Baltimore. (2019). *THE ECONOMIC COST OF BAD ACTORS ON THE INTERNET FAKE NEWS*.
12. Wang, C. (2020). *Fake News and Related Concepts: Definitions and Recent Research Development*. Contemporary Management Research, 16(3), 145-174. DOI: 10.7903/cmr.20677
13. Nasir, J. A., Khan, O. S., & Varlamis, I. (2020). *Fake news detection: A hybrid CNN-RNN based deep learning approach*. Elsevier Ltd. <https://doi.org/10.1016/j.jjime.2020.100007>
14. Essa, E., Omar, K., & Alqahtani, A. (2023). *Fake news detection based on a hybrid BERT and LightGBM models*. Complex & Intelligent Systems, 9, 6581–6592. DOI: 10.1007/s40747-023-01098-0

15. Hamid, A., Sheikh, N., Said, N., Ahmad, K., Gul, A., Hassan, L., & Al-Fuqaha, A. (2020). *Fake News Detection in Social Media using Graph Neural Networks and NLP Techniques: A COVID-19 Use-case*. MediaEval'20. arXiv:2012.07517v1 [cs.CL] 30 Nov 2020
16. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2020). *Fake News Detection on Social Media: A Data Mining Perspective*. Arizona State University, Michigan State University, and Charles River Analytics. <https://doi.org/10.1145/3137597.3137600>
17. Khalil, A., Jarrah, M., & Aldwairi, M. (2024). *Hybrid Neural Network Models for Detecting Fake News Articles*. Human-Centric Intelligent Systems, 4, 136–146. DOI: 10.1007/s44230-023-00055-x
18. Kulkarni, P., Karwande, S., Keskar, R., Kale, P., & Iyer, S. (2021). *Fake News Detection using Machine Learning*. ITM Web of Conferences, 40, 03003. ICACC-2021. <https://doi.org/10.1051/itmconf/20214003003>
19. Fioretos, D. P. (2021). *Fake News Detection with the GREEK-BERT Model with a focus on COVID-19*. BSc Thesis, National and Kapodistrian University of Athens, School of Science, Department of Informatics and Telecommunications.
20. Olan, F., Jayawickrama, U., Arakpogun, E. O., Suklan, J., & Liu, S. (2024). *Fake news on Social Media: the Impact on Society*. Information Systems Frontiers, 26, 443–458. <https://doi.org/10.1007/s10796-022-10242-z>
21. Aïmeur, E., Amri, S., & Brassard, G. (2023). *Fake news, disinformation and misinformation in social media: a review*. Social Network Analysis and Mining, 13, 30. <https://doi.org/10.1007/s13278-023-01028-5>
22. Kaliyar, R. K., Goswami, A., & Narang, P. (2021). *FakeBERT: Fake news detection in social media with a BERT-based deep learning approach*. Multimedia Tools and Applications, 80, 11765–11788. Published by Springer. <https://doi.org/10.1007/s11042-020-08915-6>
23. *Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power*. (2020). Harvard Kennedy School Misinformation Review, 1(4). DOI: <https://doi.org/10.37016/mr-2020-024>
24. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Google AI Language. arXiv:1810.04805v2 [cs.CL].
25. 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>
26. Nikiforos, M. N., Voutos, Y., Drougani, A., Mylonas, P., & Kermanidis, K. L. (2021). *The Modern Greek Language on the Social Web: A Survey of Data Sets and Mining Applications*. Data, 6(52). <https://doi.org/10.3390/data6050052>
27. Chen, W., Zhang, Y., Yeo, C. K., Lau, C. T., & Lee, B. S. (2017). *Unsupervised rumor detection based on users' behaviors using neural networks*. Nanyang Technological University, School of Computer Science and Engineering. <https://doi.org/10.1016/j.patrec.2017.10.014>

