ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΕΥΦΥΗΣ ΑΝΑΓΝΩΡΙΣΗ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ

ΛΥΜΠΕΡΙΔΗΣ ΑΠΟΣΤΟΛΟΣ

ΚΥΡΙΑΚΙΔΗΣ ΠΑΝΑΓΙΩΤΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΕΥΘΥΝΟΣ

ΚΟΛΟΜΒΑΤΣΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ
Επίκουρος Καθηγητής

Λαμία 2023

ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΕΥΦΥΗΣ ΑΝΑΓΝΩΡΙΣΗ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ

ΛΥΜΠΕΡΙΔΗΣ ΑΠΟΣΤΟΛΟΣ

ΚΥΡΙΑΚΙΔΗΣ ΠΑΝΑΓΙΩΤΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΕΥΘΥΝΟΣ

ΚΟΛΟΜΒΑΤΣΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ
Επίκουρος Καθηγητής

Λαμία 2023

UNIVERSITY OF
THESSALY

SCHOOL OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE & TELECOMMUNICATIONS

SMART RECOGNITION OF FAKE NEWS

LYMPERIDIS APOSTOLOS

KYRIAKIDIS PANAGIOTIS

FINAL THESIS

ADVISOR

KOLOMVATSOS KONSTANTINOS
Assistant Professor

Lamia 2023

«Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις [(1)], που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

*1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάσθηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.*

*2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.*

3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια

4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία: 5/10/2023

Οι Δηλ..

Παναγιώτης Κυριακίδης *Απόστολος Λυμπερίδης*

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση
του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων
σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.»

Αγγλικοί όροι

Bias = προκατάληψη

Classification = Ταξινόμηση

Corpus = Μια μεγάλη συλλογή γραπτών

Data Frame = Πλαίσιο δεδομένων

Dataset = Σύνολο δεδομένων

Embedding = Ενσωμάτωση

Epoch = Ένα πλήρες πέρασμα του συνόλου δεδομένων εκπαίδευσης με έναν αλγόριθμο.

Feature = Χαρακτηριστικό

Inverse Document Frequency (IDF) = Διαδικασία μέτρησης μοναδικών λέξεων σε μια συλλογή εγγράφων

Machine Learning = Μηχανική Μάθηση

N-gram = μια συλλογή από n διαδοχικά στοιχεία σε ένα έγγραφο κειμένου που μπορεί να περιλαμβάνει λέξεις, αριθμούς, σύμβολα και σημεία στίξης.

Natural Language Processing (NLP) = Κλάδος της τεχνητής νοημοσύνης που διδάσκει στους υπολογιστές να κατανοούν και να ερμηνεύουν ανθρώπινη γλώσσα.

Neural Networks = Νευρωνικά δίκτυα

Word Embeddings = Ενσωματώσεις λέξεων

(Pre)Process = (Προ)Επεξεργασία

Parameter = Παράμετρος

Percentile = Εκατοστημόριο

Regularize / Regularization = Κανονικοποιώ / Κανονικοποίηση ή συστηματοποίηση

Stopwords = Ενδιάμεσες λέξεις που δεν έχουν ουσιαστικό νόημα στις προτάσεις.

Token / Tokenization / Tokenizer = Διακριτικό / Διακριτοποίηση / Διακριτοποιητής

Training = Εκπαίδευση

Units = Μονάδες

Unsupervised = Χωρίς Επίβλεψη

Vector / Vectorizer = Διάνυσμα / Διανυσματοποιητής

ΠΕΡΙΛΗΨΗ

Στην ψηφιακή εποχή, ο πολλαπλασιασμός των ψευδών ειδήσεων έχει αναδειχθεί ως μια σημαντική κοινωνική πρόκληση, καθιστώντας αναγκαία την ανάπτυξη ισχυρών μοντέλων ανίχνευσής τους. Αυτή η πτυχιακή εμβαθύνει στον περίπλοκο τομέα της ανίχνευσης ψευδών ειδήσεων χρησιμοποιώντας ένα ποικίλο σύνολο συνόλων δεδομένων και τεχνικών μηχανικής εκμάθησης. Τα σύνολα δεδομένων περιλαμβάνουν τα Fake News Corpus, WELFake και LIAR, παρέχοντας ένα ολοκληρωμένο πεδίο δοκιμών για την αναλυτική μας ικανότητα. Εκπαιδεύτηκε και αξιολογήθηκε μια εξαντλητική σειρά μοντέλων μηχανικής μάθησης, όπως το Logistic Regression, ο Passive Aggressive Classifier, το Random Forest, τα Decision Trees, το Polynomial Naive Bayes, τα Support Vector Machines, το BERT, το FastText, τα CNN, τα LSTM και υβριδικά μοντέλα CNN+GRU. Στην επιδίωξη αποτελεσματικής ανίχνευσης ψεύτικων ειδήσεων, αξιοποιήθηκαν μια σειρά τεχνικών επεξεργασίας φυσικής γλώσσας (NLP), συμπεριλαμβανομένων παραδοσιακών μεθόδων όπως το Bag-of-Words και η αντίστροφη συχνότητα εγγράφων (IDF), παράλληλα με σύγχρονες προσεγγίσεις όπως οι ενσωματώσεις λέξεων (word embeddings). Κάθε μοντέλο αξιολογήθηκε αυστηρά σε πολλαπλά σύνολα δεδομένων για να διακριθεί η ικανότητά τους για γενίκευση μεταξύ συνόλων δεδομένων. Τα ευρήματα αποκάλυψαν ότι ενώ τα μεμονωμένα μοντέλα πέτυχαν αξιέπαινη ακρίβεια στα αντίστοιχα σύνολα δεδομένων εκπαίδευσης, η γενίκευση μεταξύ συνόλων δεδομένων παρέμεινε μια τρομερή πρόκληση. Ένα από τα μοντέλα που χρησιμοποιήθηκαν μας έδωσε την μεγαλύτερη μέση ακρίβεια με τιμή σχεδόν 71% όταν χρησιμοποιήθηκε στην πρόβλεψη όλων των συνόλων δεδομένων αλλά παρουσίασε παρόλα αυτά μείωση της ακρίβειας στα σύνολα δεδομένων που ήταν διαφορετικά από το αρχικό που εκπαιδεύτηκε. Αυτά τα αποτελέσματα υπογραμμίζουν τη δυσκολία της ανίχνευσης ψεύτικων ειδήσεων, υπογραμμίζοντας την ανάγκη για καινοτόμες τεχνικές και μοντέλα ικανά να υπερβαίνουν τις αποχρώσεις που αφορούν συγκεκριμένα δεδομένα. Η διατριβή ολοκληρώνεται με μια πρόσκληση για μελλοντική έρευνα για τη διερεύνηση νέων οδών για τον μετριασμό των προκλήσεων που θέτει η γενίκευση μεταξύ συνόλων δεδομένων, προσφέροντας πολύτιμες γνώσεις τόσο για τον εντοπισμό ψεύτικων ειδήσεων όσο και για τη μηχανική μάθηση γενικότερα.

ABSTRACT

In the digital age, the proliferation of fake news has emerged as a major societal challenge, necessitating the development of robust fake news detection models. This thesis delves into the complex field of fake news detection using a diverse set of datasets and machine learning techniques. The datasets include the Fake News Corpus, WELFake and LIAR, providing a comprehensive testing ground for our analytical capability. An exhaustive set of machine learning models including Logistic Regression, Passive Aggressive Classifier, Random Forest, Decision Trees, Polynomial Naive Bayes, Support Vector Machines, BERT, FastText, CNNs, LSTMs and hybrid CNN+GRU models. In epitomizing effective fake news detection, a range of natural language processing (NLP) techniques were leveraged, including traditional methods such as Bag-of-Words and Inverse Document Frequency (IDF), alongside modern approaches such as word embeddings. Each model was rigorously evaluated on multiple data sets to distinguish between them for general data. The findings revealed that while individual models achieved commendable accuracy on their respective training datasets, generalization across datasets remained a formidable challenge. One of the models gave us the highest average accuracy with a value of nearly 71% when used to predict all data sets but still showed a decrease in accuracy on data sets that were different from the original trained. These results highlight the complexities of counterfeit detection, highlighting the need for innovative techniques and models capable of overcoming data-specific nuances. The thesis concludes with a call for future research to explore new avenues to mitigate the challenges posed by generalization across datasets, offering valuable insights for both fake news detection and machine learning in general.

Αφιέρωση και ευχαριστίες

Αρχικά, θα ήθελα να ευχαριστήσω την οικογένεια μου η οποία κατά την διάρκεια των σπουδών μου με υποστήριξαν οικονομικά, ηθικά και ψυχολογικά. Ιδιαίτερα, θα ήθελα να ευχαριστήσω τον αδερφό μου ο οποίος μου παρείχε κίνητρο δείχνοντας μου ότι στην ζωή μπορείς να καταφέρεις πολλά πράγματα αρκεί να βάλεις στόχους και να έχεις αρκετή θέληση. Δεν θα μπορούσα να μην αναφερθώ στον επιβλέποντα καθηγητή μου κ. Κωνσταντίνο Κολομβάτσο ο οποίος είναι ένας άνθρωπος που μέσα από τον τρόπο που διδάσκει μου κίνησε το ενδιαφέρον για γνώση και κατέληξα να ενδιαφέρομαι για τον τομέα της μηχανικής μάθησης. Υπήρξε αρωγός για την πτυχιακή εργασία αφού μας παρείχε γνώση, μας καθοδήγησε και η υπομονή του και ο χρόνος που μας παρείχε για τυχόν απορίες και προβλήματα ήταν πολύτιμη ώστε να φέρουμε εις πέρας την πτυχιακή εργασία. Επιπλέον, οι φίλοι μου οι οποίοι αποτελούν ένα αναπόσπαστο κομμάτι της καθημερινότητας μου, είναι εκείνοι που διαρκώς μου παρείχαν αξέχαστες αναμνήσεις, υποστήριξη σε πολλά θέματα που με αφορούσαν και αμέτρητη γνώση λόγω των συζητήσεων που κάναμε, που πολλές φορές κατέληγα να διορθώνω τον εαυτό μου μαθαίνοντας καινούργια πράγματα και βλέποντας τον κόσμο και από τα δικά τους μάτια αλλάζοντας συνεχώς την ιδιοσυγκρασία μου και μετατρέποντας με σε καλύτερο άνθρωπο, τολμώ να πω. Μεταξύ αυτών των φίλων θα ήθελα να ευχαριστήσω ιδιαιτέρως τον καλό μου φίλο και συνεργάτη Παναγιώτη Κυριακίδη του οποίου η παρέα έκανε ευχάριστη όλη την διαδρομή της ολοκλήρωσης της πτυχιακής εργασίας, ενώ η επιμονή το ενδιαφέρον και η ομαδικότητα του συνέδραμαν στην απόκτηση πολύπλευρης γνώσης και πιο αποτελεσματικής προσέγγισης της εργασίας.

Δεν θα μπορούσα να καταφέρω πολλά πράγματα στην ζωή μου χωρίς όλους όσους ανέφερα και είμαι διαρκώς ευγνώμων για την υγεία μου και τα λίγα πράγματα που έχω στην ζωή τα οποία μου παρέχουν συνεχώς κίνητρο αυτοβελτίωσης και μου επιτρέπουν να βάζω και να κυνηγάω υψηλούς στόχους στην ζωή μου.

- Απόστολος

Ευχαριστώ την οικογένεια μου, τους φίλους μου και όλους τους ανθρώπους που με αγαπούν για την στήριξη και την κατανόηση που μου προσέφεραν καθ' όλη την διάρκεια εκπόνησης και συγγραφής αυτής της εργασίας, καθώς και των χρόνων φοίτησής μου. Ευχαριστώ επίσης και τον Απόστολο Λυμπερίδη, με τον οποίο συνεργαστήκαμε για την ολοκλήρωση της εργασίας αυτής, για την οξυδέρκεια και το ομαδικό του πνεύμα. Η συνεργασία μας κατέστησε τις ώρες έρευνας και δουλείας ακόμα πιο ευχάριστες.

- Παναγιώτης

Table of Contents

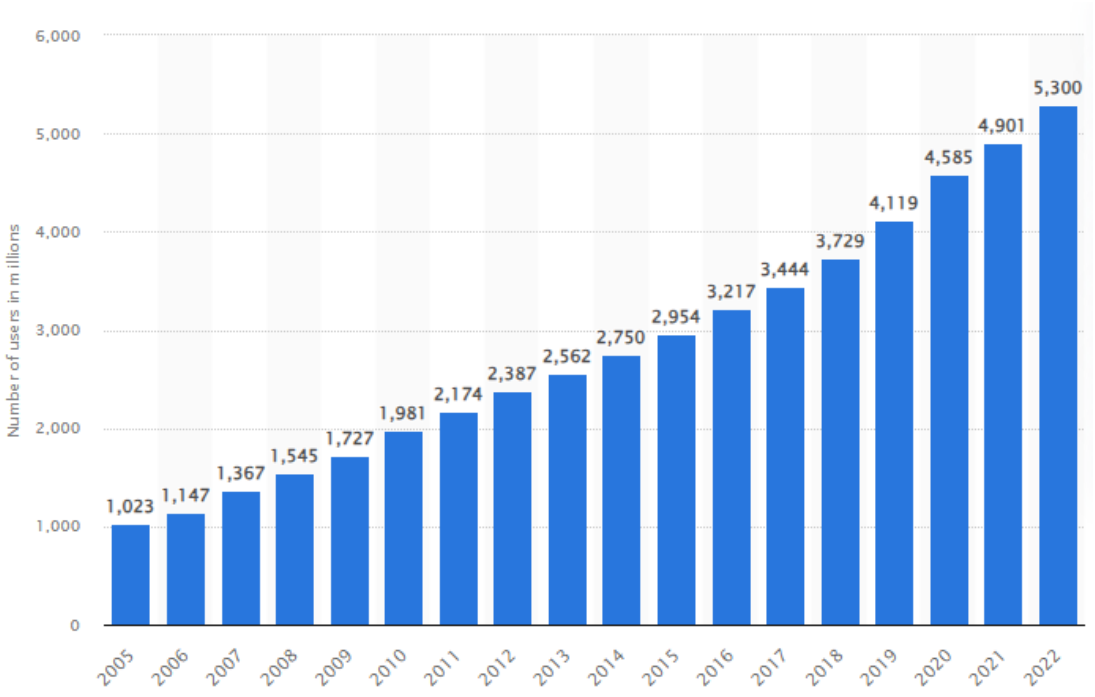
List of Figures

List of Tables

List of Procedures

ΚΕΦΑΛΑΙΟ 1 Εισαγωγή

Η έλευση της ψηφιακής εποχής έχει εγκαινιάσει μια εποχή άνευ προηγουμένου πρόσβασης στην πληροφορία. Στον πλέον υπερσυνδεδεμένο κόσμο, εικόνα 1.1 [1], οι πληροφορίες ρέουν με ιλιγγιώδη ταχύτητα προκαλώντας το φαινόμενο της υπερφόρτωσης πληροφοριών. Το φαινόμενο αυτό εκτός ότι μπορεί να επηρεάσει την διαδικασία λήψης αποφάσεων, έχει αποδειχθεί ότι στο καταναλωτικό κοινό μπορεί να προκαλέσει άγχος, σύγχυση και αναστάτωση [2]. Σε φοιτητές μεταπτυχιακών εκτός του άγχους, της αποθάρρυνσης, της απογοήτευσης και της σύγχυσης μπορεί να έχει αρνητικές επιδράσεις και στις έρευνες τους όπου η κακή ποιότητα έρευνας και η χαμηλή παραγωγικότητα [3] μεταξύ αυτών προκαλεί ένα μείζων πρόβλημα. Εκτός αυτών, δημιουργούνται προκλήσεις στο σύνολο των πληροφοριών όπως το διάχυτο και ύπουλο φαινόμενο των Fake News. Τα fake news, είναι ένας όρος που έχει αποκτήσει ευρεία φήμη και υποδηλώνει τη σκόπιμη κατασκευή ή διάδοση παραπλανητικών πληροφοριών που μεταμφιέζονται ως πραγματικές ειδήσεις. Γι' αυτό το λόγο αντιμετωπίζονται ως ένας τρομερός αντίπαλος στη μάχη για την ακεραιότητα των πληροφοριών στη σύγχρονη κοινωνία.



Εικόνα 1.1: Αριθμός χρηστών του Διαδικτύου παγκοσμίως από το 2005 έως το 2022

(Υποκεφάλαιο 1.1) **Ιστορία και αντίκτυπος των Fake News**

Ας εμβαθύνουμε σε μια διερεύνηση των ψευδών ειδήσεων και της πολύπλευρης εξέλιξής τους, καθώς και των επιπτώσεων που δημιούργησε σε διάφορους τομείς με την πάροδο του χρόνου. Οι ψευδείς ειδήσεις θα μπορούσαν να θεωρηθούν ως μια Οδύσσεια καθώς το φαινόμενο αυτό απέχει πολύ από κάτι καινούργιο στην ανθρώπινη ιστορία. Πριν ακόμα την ανακάλυψη του ραδιοφώνου, της τηλεόρασης και την έναρξη του World Wide Web (WWW) υπάρχουν άρθρα [4] που εμβαθύνουν στις ιστορικές ρίζες των ψευδών ειδήσεων που χρονολογούνται από τον 18ο αιώνα με την μορφή του Τύπου. Συγκεκριμένα, όταν ο Γάλλος φιλόσοφος Condorcet και ο πρόεδρος της Αμερικής John Adams συμμετείχαν σε συζητήσεις σχετικά με τον ρόλο του ελεύθερου τύπου

ειπώθηκε πως «τα τελευταία δέκα χρόνια έχουν διαδοθεί περισσότερα νέα σφάλματα από τον Τύπο από ό,τι σε εκατό χρόνια πριν από το 1798».

Αυτό σημαίνει ότι η ιστορία είναι γεμάτη με περιπτώσεις παραπληροφόρησης, προπαγάνδας και κατάλληλα σχεδιασμένων άρθρων ή ομιλιών για να ασκήσουν επιρροή, να χειραγωγήσουν την κοινή γνώμη και κατ' επέκταση να επωφεληθούν πολιτικά ή κοινωνικά διάφοροι άνθρωποι. Ακολουθούν μερικά παραδείγματα αυτών των φαινομένων κατά το πέρασμα των χρόνων:

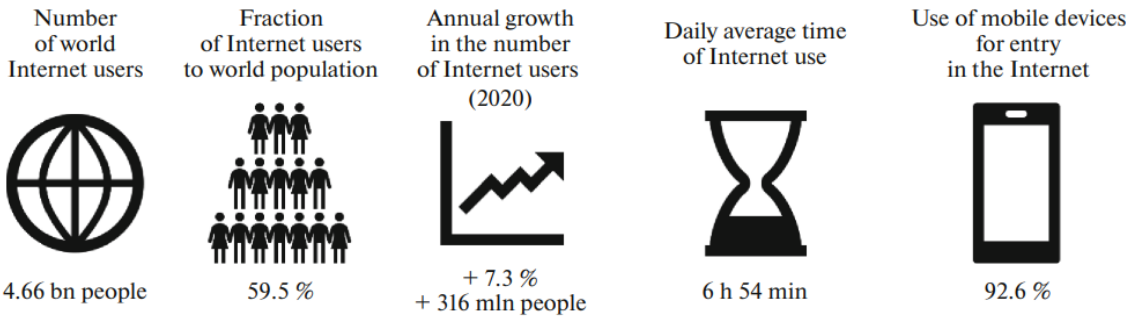
- Το "**Great Moon Hoax**" ή η "**Μεγάλη Φεγγαρική Απάτη**" αν και είναι μια φάρσα των εφημερίδων του δέκατου ένατου αιώνα, συγκεκριμένα της εφημερίδας «The New York Sun», που αποτελεί ένα από τα πρώτα τεκμηριωμένα περιστατικά ψευδών ειδήσεων. Η δημοσίευση περιέχει μια σειρά άρθρων που υποτίθεται ότι αποκάλυψε την ανακάλυψη της ζωής στο φεγγάρι. Ήταν ένα ελκυστικό έργο φαντασίας που αιχμαλώτισε τις καρδιές και το μυαλό των αναγνωστών, συμβολίζοντας τελικά τη διαρκή γοητεία των ψευδών ειδήσεων.

- Η εποχή της Κίτρινης Δημοσιογραφίας υπήρξε παράγοντας του πολλαπλασιασμού των συγκλονιστικών ρεπορτάζ που χαρακτηρίζονται από υπερβολές, στρεβλώσεις και υπερβολικές αφηγήσεις για την τόνωση της αγοραστικής κυκλοφορίας και της διαμόρφωσης της αντίληψης του κοινού.

- Κατά τη διάρκεια των κατακλυσμικών γεγονότων του Α' και του Β' Παγκοσμίου Πολέμου οι κυβερνήσεις και στις δύο πλευρές αυτών των συγκρούσεων χρησιμοποίησαν την προπαγάνδα ως ένα ισχυρό εργαλείο για να διαμορφώσουν το δημόσιο αίσθημα και να συγκεντρώσουν υποστήριξη για τις αντίστοιχες πολεμικές τους προσπάθειες. Αυτές οι προπαγανδιστικές εκστρατείες συχνά διέδιδαν ψευδείς πληροφορίες και εξωραΐζονταν γεγονότα για να επηρεάσουν το αίσθημα και το μυαλό ολόκληρων εθνών.

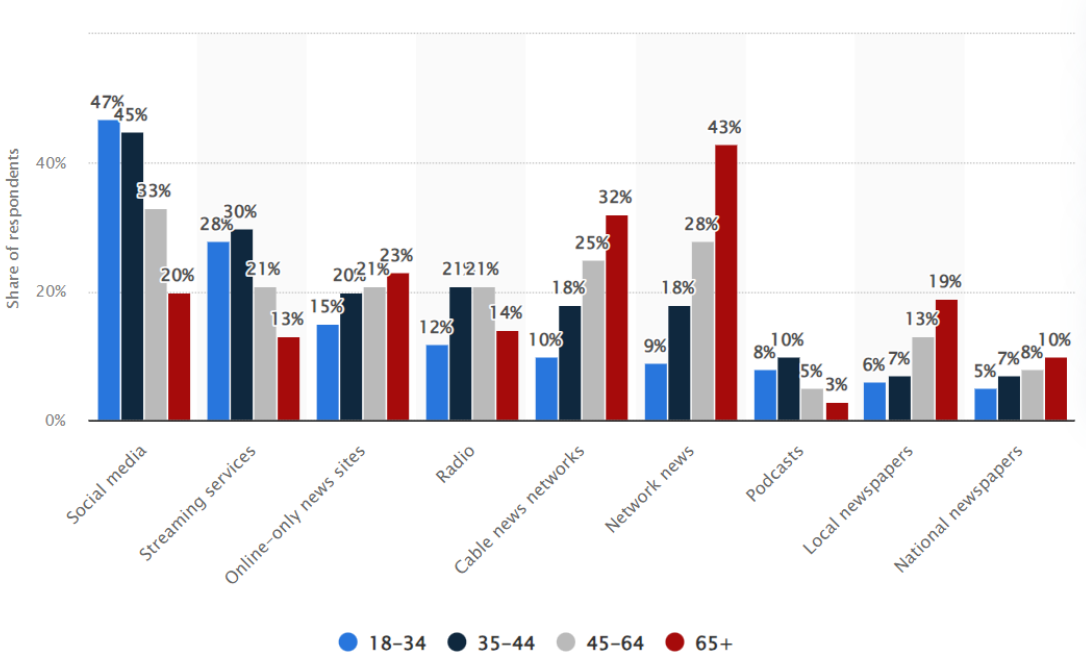
- Η επανάσταση του διαδικτύου απελευθέρωσε μια νέα εποχή διάδοσης ψευδών ειδήσεων. Με την άνοδο των φόρουμ (forum), των προσωπικών ιστολογίων (personal blogs) και την εκρηκτική ανάπτυξη των μέσων κοινωνικής δικτύωσης, η διάδοση ψευδών πληροφοριών έγινε πιο προσιτή και διαδεδομένη από ποτέ. Ο εκδημοκρατισμός της πληροφορίας κατέληξε να έχει ένα τίμημα αυτό της διάδοσης των ψευδών ειδήσεων.

Η ψηφιακή εποχή έχει προκαλέσει μια άνευ προηγουμένου ταχύτητα διάδοσης πληροφοριών. Σε έναν κόσμο όπου οι πληροφορίες είναι μόνο ένα κλικ μακριά, η ιδέα των ψεύτικων ειδήσεων είναι συγκλονιστική. «Το 2020 ο μέσος χρήστης ξοδεύει σχεδόν επτά ώρες την ημέρα στο διαδίκτυο το οποίο αποτελεί αύξηση 9% από την προηγούμενη χρονιά» [5] εικόνα 1.2. Καταλαβαίνουμε ότι η έκθεση του μέσου χρήστη στις πληροφορίες και ιδίως στις ψευδές είναι μεγάλη αποτελώντας σημαντική απειλή για τη λήψη αποφάσεων με ενημέρωση και την δημοκρατική και κοινωνική αρμονία.



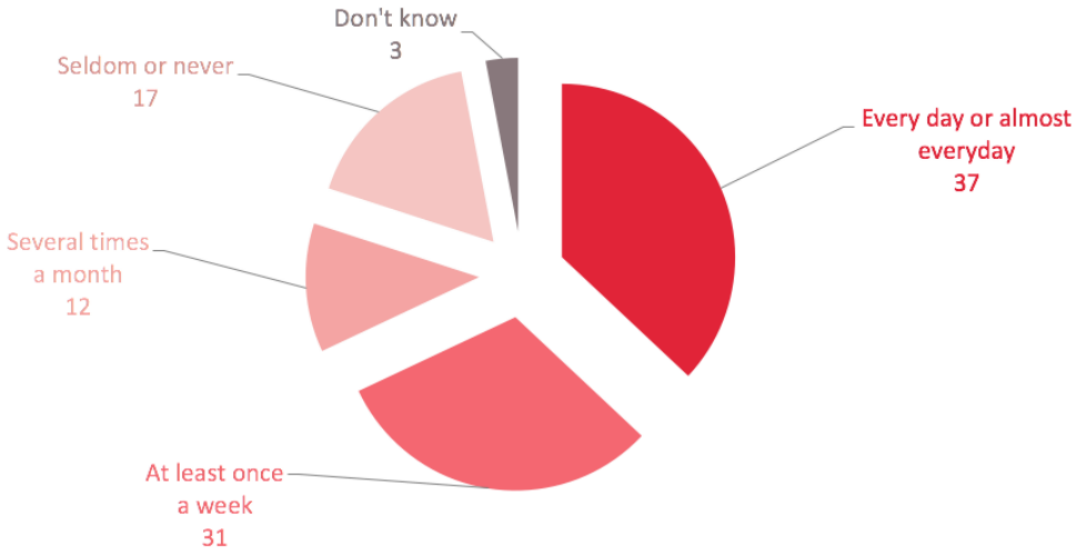
Εικόνα 1.2: Μια επισκόπηση της παγκόσμιας χρήσης του Διαδικτύου

Η πλειοψηφία των ανθρώπων έχουν επιλέξει να καταναλώνουν νέα μέσω των κοινωνικών δικτύων έναντι οποιασδήποτε άλλης πλατφόρμας, εικόνα 1.3 [6]. Υπάρχουν βέβαια πολύπλευροι κίνδυνοι στα μέσα κοινωνικής δικτύωσης θέτοντας μοναδικές προκλήσεις σε σύγκριση με τις παραδοσιακές πλατφόρμες ειδήσεων. Η ταχεία διάδοση είναι από τις πιο σημαντικές διότι οι αλγόριθμοι που προτείνουν περιεχόμενο στον χρήστη έχουν την τάση να προωθούν παραπλανητικές ή ψευδές πληροφορίες που έγιναν διάσημες μέσα σε σύντομο χρονικό διάστημα φτάνοντας σε εκατομμύρια χρήστες προτού μπορέσουν να απαντήσουν οι ελεγκτές στοιχείων. Υπάρχουν φορές που η κοινοποίηση ψευδών ειδήσεων οφείλεται και στην άγνοια του αναγνώστη καθώς έρευνα έχει δείξει πως Αμερικανοί μοιράζονται ψεύτικες ειδήσεις στα μέσα κοινωνικής δικτύωσης επειδή απλώς δεν δίνουν προσοχή στο αν το περιεχόμενο είναι ακριβές και όχι απαραίτητα επειδή δεν μπορούν να ξεχωρίσουν εάν μια είδηση είναι ψευδής ή όχι [7]. Στα κοινωνικά δίκτυα βέβαια μπορεί εύκολα να χειριστεί η κοινή γνώμη από κακόβουλους παράγοντες. Για παράδειγμα οι ψεύτικες ειδήσεις μπορούν να χρησιμοποιηθούν στρατηγικά για να χειραγωγήσουν την κοινή γνώμη και να επηρεάσουν τις εκλογές, τα δημοψηφίσματα και τον κοινωνικό λόγο. Έρευνα μεγάλης κλίμακας που διεξήχθη στη Γερμανία και το Ηνωμένο Βασίλειο έχει δείξει ότι ενώ η πλειονότητα της κοινοποίησης ψευδών ειδήσεων συμβαίνει ακούσια, με μόνο ένα μικρότερο μέρος να είναι σκόπιμη, οι νεότεροι και οι δεξιοί τείνουν να μοιράζονται πιο συχνά ψεύτικες ειδήσεις [8]. Ένας εξίσου σημαντικός κίνδυνος είναι αυτός της προκατάληψης επιβεβαίωσης (confirmation bias) όπου οι αλγόριθμοι των μέσων κοινωνικής δικτύωσης δίνουν συχνά προτεραιότητα σε περιεχόμενο που ευθυγραμμίζεται με τις υπάρχουσες πεποιθήσεις και προτιμήσεις των χρηστών. Έτσι, ενισχύεται η προκατάληψη επιβεβαίωσης, με τους χρήστες να εκτίθενται και να μοιράζονται πληροφορίες που ενισχύουν τις προκαταλήψεις τους, εμβαθύνοντας τις ιδεολογικές τους διαφορές και καμιά φορά έχοντας αντίκτυπο στον πραγματικό κόσμο. Τέτοιο παράδειγμα παραπληροφόρησης αφορά τον τομέα της υγείας όπου «ένα κατασκευασμένο επιστημονικό άρθρο που ισχυριζόταν ότι το εμβόλιο ιλαράς, παρωτίτιδας και ερυθράς προκαλεί αυτισμό οδήγησε σε ευρεία διάδοση αυτής της παραπληροφόρησης, ειδικά μέσω των μέσων κοινωνικής δικτύωσης. Αυτό, με τη σειρά του, οδήγησε όχι μόνο σε επίπεδα ρεκόρ της επίπτωσης της ιλαράς στην Ευρώπη το 2018, αλλά διεύρυνε επίσης το φάσμα των λεγόμενων αντιεμβολιαστών» [5].



Εικόνα 1.3: Οι πιο δημοφιλείς πλατφόρμες για καθημερινή κατανάλωση ειδήσεων στις Ηνωμένες Πολιτείες από τον Αύγουστο του 2022, ανά ηλικιακή ομάδα

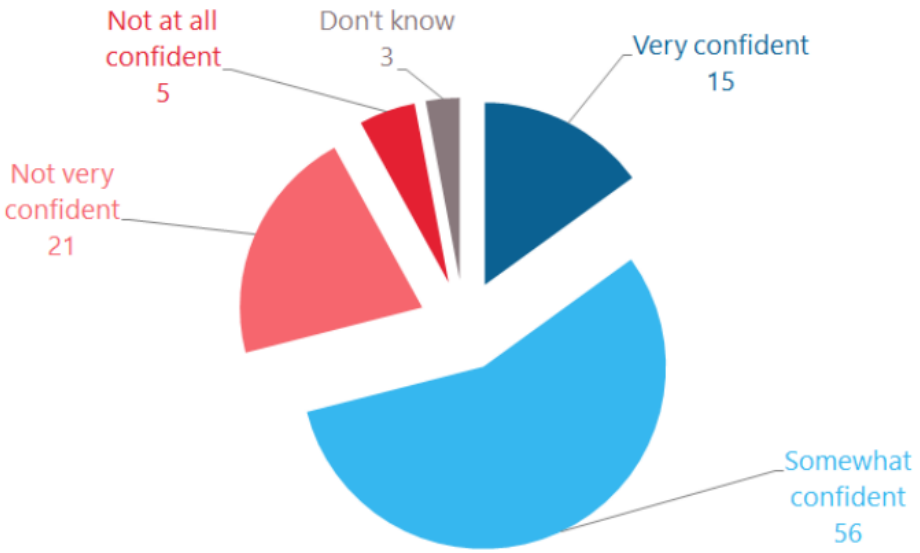
Υπάρχει ακόμη μια έρευνα που έχει διεξαχθεί για την Ευρωπαϊκή Επιτροπή το 2018 η οποία περιελάμβανε συνεντεύξεις με περισσότερους από 26.000 ερωτηθέντες στα 28 κράτη μέλη της ΕΕ. Βρέθηκε ότι μια σημαντική μερίδα των ερωτηθέντων (31%) αντιμετωπίζει ψεύτικες ειδήσεις τουλάχιστον μία φορά την εβδομάδα και το 37% να τις συναντά καθημερινά ή σχεδόν καθημερινά, εικόνα 1.4 [9]. Λαμβάνοντας υπόψιν ότι η έρευνα είναι πέντε χρόνων μπορούμε να υποθέσουμε με σιγουριά ότι περισσότερα άτομα θα απαντούσαν ότι συναντάνε ψευδή νέα συχνότερα.



Εικόνα 1.4: Γράφημα πίτα με τα ποσοστά των απαντήσεων των ερωτηθέντων στην ερώτηση «πόσο συχνά συναντάτε ειδήσεις ή πληροφορίες που πιστεύετε ότι παραποιούν την πραγματικότητα ή είναι ακόμη και ψευδείς;»

(Υποκεφάλαιο 1.2) **Προβλήματα ανίχνευσης και αντιμετώπισης**

Η έρευνα για την Ευρωπαϊκή Επιτροπή [9] έχει δείξει ακόμη ότι οι ερωτηθέντες εμπιστεύονται τις παραδοσιακές πηγές μέσων όπως το ραδιόφωνο (70%), την τηλεόραση (66%) και τα έντυπα μέσα (63%) περισσότερο από τις διαδικτυακές πηγές όπως διαδικτυακές εφημερίδες και περιοδικά (47%), ιστότοπους με βίντεο ή podcast (27%) και διαδικτυακά κοινωνικά δίκτυα (26%). Από την άλλη, το 15% είναι απόλυτα σίγουροι και το 56% κάπως σίγουροι ότι είναι σε θέση να αναγνωρίσουν ειδήσεις ή πληροφορίες που παραποιούν την πραγματικότητα ή είναι ψευδείς ενώ το 26% δεν έχει αρκετή αυτοπεποίθηση, εικόνα 1.5. Από τα παραπάνω θα μπορούσαμε να συσχετίσουμε την έλλειψη εμπιστοσύνης των ειδήσεων που προέρχονται από τα κοινωνικά μέσα με την βεβαιότητα αναγνώρισης αλλά αυτό δεν σημαίνει ότι η αναγνώριση των πιθανών ψευδών ειδήσεων που εκλαμβάνουν είναι σωστή. Τα κατασκευασμένα νέα είναι πιο σύνθετα επειδή είναι πολύ εύκολο να παραποιηθεί ή αποκρυφθεί η αλήθεια πράγμα που δημιουργεί και το πρόβλημα της ανίχνευσης.



Εικόνα 1.5: Γράφημα πίτας που αναδεικνύει την σιγουριά των ερωτηθέντων όταν τους ρωτήθηκε «πόσο σίγουροι ή όχι είστε σε θέση να αναγνωρίσετε ειδήσεις ή πληροφορίες που παραποιούν την πραγματικότητα ή είναι ακόμη και ψευδείς;»

Στον ψηφιακό κόσμο, οι ψεύτικες ειδήσεις ευδοκιμούν λόγω ποικίλων παραγόντων. Οι τεχνολογικές εξελίξεις έχουν ενδυναμώσει τόσο τους δημιουργούς περιεχομένου όσο και τους καταναλωτές, επιτρέποντας τη δημιουργία περίπλοκων ψευδών ειδήσεων και την ταχεία ανταλλαγή πληροφοριών σε διάφορες πλατφόρμες κοινωνικής δικτύωσης. Η διαδικασία της επικύρωσης της αλήθεια καθίσταται αδύνατη από ανθρώπινο παράγοντα λόγω του όγκου των δεδομένων που πρέπει να ελέγχονται συνεχώς κάνοντας το πρόβλημα της ανίχνευσης ακόμη πιο προκλητικό. Υπάρχουν τεχνικές πειθούς όπου χρησιμοποιούν επικλήσεις στο συναίσθημα και άλλες μορφές αισθησιασμού ώστε να κεφαλαιοποιούν τον εντυπωσιασμό, αξιοποιώντας συναισθηματικά φορτισμένη γλώσσα και κατάλληλες εικόνες που χειραγωγούν τα συναισθήματα και τις ευαισθησίες των αναγνωστών. Επιπρόσθετα, η σάτιρα που πολλές φορές είναι ασαφής μπορεί εύκολα να εκληφθεί ως πραγματική είδηση. Για παράδειγμα οι χειραγωγικές σατιρικές εικόνες και το σατιρικό περιεχόμενο είναι κοινά χαρακτηριστικά των πολυσχιδών ψευδών νέων που μπορούν να κάνουν τον κύκλο του διαδικτύου αφήνοντας υποσυνείδητα το μήνυμα τους στον αναγνώστη. Υπάρχει και ένας σημαντικός παράγοντας ηθικής όπου οι εταιρίες που κατέχουν κοινωνικές πλατφόρμες πρέπει να ισορροπήσουν την καταπάτηση της ελευθερίας του λόγου και την αντιμετώπιση των ψευδών ειδήσεων. Τελικά, η ανωνυμία επιδεινώνει περαιτέρω το πρόβλημα επειδή προστατεύονται τα άτομα που παραπληροφορούν ή διαδίδουν ψευδή νέα και δεν μπορούν να τους επωμιστούν ευθύνες.

Μέσα στο περίπλοκο και εξελισσόμενο τοπίο των ψεύτικων ειδήσεων, υπάρχει ανάγκη για ισχυρές λύσεις. Πρέπει να αναπτυχθούν στρατηγικές ανίχνευσης και μετριασμού για την αντιμετώπιση των επιπτώσεων των ψευδών ειδήσεων. Σε αυτήν την επιδίωξη, η μηχανική μάθηση που είναι πεδίο της τεχνητής νοημοσύνης, μπορεί να δώσει λύση στο πρόβλημα. Αξιοποιώντας τη δύναμη των αλγορίθμων που βασίζονται σε δεδομένα και της επεξεργασίας φυσικής γλώσσας (NLP), οι ερευνητές έχουν την δυνατότητα να αναπτύξουν μοντέλα ικανά να διακρίνουν τα τεχνητά κείμενα που διαστρεβλώνουν την πραγματικότητα. Φυσικά, δεν είναι μόνο στο χέρι των ερευνητών να αντιμετωπίσουν αυτό το πρόβλημα καθώς οι άνθρωποι που ερωτήθηκαν «ποιοι είναι οι πιο καθορισμένοι παράγοντες για να σταματήσουν τη διάδοση ψευδών ειδήσεων;» στην έρευνα για την Ευρωπαϊκή Επιτροπή[9] απάντησαν πως οι δημοσιογράφοι πρέπει να δράσουν για να σταματήσουν τη διάδοση ψεύτικων ειδήσεων (45%), ακολουθούμενες από τις εθνικές αρχές (39%), τη διαχείριση του Τύπου και των ραδιοτηλεοπτικών εκπομπών (36%), τους ίδιους τους πολίτες (32%), τα διαδικτυακά κοινωνικά δίκτυα (26%), τα θεσμικά όργανα της ΕΕ (21%) και τις μη κυβερνητικές οργανώσεις (15%), εικόνα 1.6. Υπάρχει και έρευνα που θεωρεί την έννοια της «κοινωνικής αποδοχής» ως

κρίσιμο παράγοντα για την κατανόηση του τρόπου με τον οποίο τα ψευδή νέα εξαπλώνονται και επηρεάζουν τους ανθρώπους [10]. Υποστηρίζει ακόμη ότι υπάρχει ανάγκη για τον έλεγχο γεγονότων (fact checking) και την διαχείριση περιεχομένου των ψευδών ειδήσεων σε πλατφόρμες κοινωνική δικτύωσης και προτείνει οι μεγάλοι οργανισμοί που τους ανήκουν αυτές οι πλατφόρμες να είναι καλύτερα εξοπλισμένοι για να ξεκινήσουν αυτές τις προσπάθειες. Ενώ άλλη έρευνα που αναφέραμε [7] υποστηρίζει μια πιο καινοτόμα λύση όπου εμπλέκεται ο ίδιο ο χρήστης. Προτείνουν οι πλατφόρμες να προτρέπουν περιοδικά τους χρήστες να αξιολογούν την ακρίβεια των τυχαία επιλεγμένων επικεφαλίδων. Αυτό, αναφέρουν, πως θα χρησίμευε ως μια λεπτή υπενθύμιση σχετικά με τη σημασία της ακρίβειας και θα δημιουργούσε επίσης πολύτιμες αξιολογήσεις χρηστών για τον εντοπισμό παραπληροφόρησης ή ψευδών ειδήσεων.

Εικόνα 1.6: Ραβδόγραμμα που δείχνει τους παράγοντες σε ποσοστά που πιστεύουν οι ερωτηθέντες ότι πρέπει να δράσουν για να σταματήσουν τη διάδοση ψεύτικων ειδήσεων

ΚΕΦΑΛΑΙΟ 2 Βιβλιογραφική Επισκόπηση

Η εξερεύνηση των σχετικών εργασιών στον τομέα της ευφυής αναγνώρισης ψευδών ειδήσεων δεν είναι απλώς μια επιπόλαια άσκηση, αλλά μια ηράκλεια προσπάθεια από μια μεγάλη επιστημονική κοινότητα. Σε αυτό το κεφάλαιο, θα παρουσιάσουμε την βιβλιογραφία και τις έρευνες που έχουν διεξαχθεί και σχετίζονται με την μελέτη μας. Το συγκεκριμένο κεφάλαιο αποτελεί τα θεμέλια της έρευνας μας καθώς συνδέεται η προηγούμενη γνώση με τη δική μας προσπάθεια για ανάπτυξη μοντέλων για τον σκοπό της αναγνώρισης ψευδών ειδήσεων. Η σημασία αυτών των επιστημονικών προσπαθειών διατυπώνεται σχολαστικά, παρέχοντας την δυνατότητα της σύνθεσης των συνεχιζόμενων ερευνών και της υπάρχουσας έρευνας.

Εξετάζοντας προηγούμενες μελέτες, στοχεύουμε να αποκτήσουμε πολύτιμες γνώσεις, και να δημιουργήσουμε το πλαίσιο εντός του οποίου οι δικές μας συνεισφορές βρίσκουν τη θέση τους. Η σύνθεση της υπάρχουσας έρευνας δεν είναι μια απλή επισήμανση των ευρημάτων, αλλά μια διαδικασία εναρμόνισης διαφορετικών πτυχών της γνώσης. Αναγνωρίζουμε την ανεκτίμητη συνεισφορά όλων όσων ακολούθησαν τον ίδιο δρόμο πριν από εμάς και αυτό μας δίνει παραπάνω κίνητρο να γεφυρώσουμε το παρελθόν με το παρόν. Αυτό το κεφάλαιο, λοιπόν, χρησιμεύει ως πυξίδα που μας καθοδηγεί στα ευρήματα του πεδίου μας. Ο παρακάτω πίνακας υποδεικνύει τις έρευνες που διεξήχθησαν στα πλαίσια των ψευδών ειδήσεων.

1	Tavishee Chauhan, Hemant Palivela, “Optimization and improvement of fake news detection using deep learning approaches for societal benefit”	[11]
2	Akhtar, P., Ghouri, A.M., Khan, H.U.R. et al. “Detecting fake news and disinformation using artificial intelligence and machine learning to avoid supply chain disruptions”	[12]
3	S. Rastogi, D. Bansal, “A review on fake news detection 3T’s: typology, time of detection, taxonomies.”	[13]
4	Ihsan Ali, Mohamad Nizam Bin Ayub, Palaiahnakote Shivakumara, Nurul Fazmidar Binti Mohd Noor, “Fake News Detection Techniques on Social Media: A Survey”	[14]
5	Nicole O'Brien, “Machine learning for detection of fake news”	[15]
6	S.A. Khan, K. Shahzad, O. Shabbir, A. Iqbal “Developing a Framework for Fake News Diffusion Control (FNDC) on Digital Media (DM): A Systematic Review 2010–2022”	[16]

Πίνακας 2.1: Έρευνες που μελετήσαμε για τους στόχους την πτυχιακής εργασίας

1. Tavishee Chauhan, Hemant Palivela, “Optimization and improvement of fake news detection using deep learning approaches for societal benefit”

   Στο παρόν άρθρο γίνεται λόγος για την χρήση βαθιάς μάθησης ώστε να καταπολεμηθεί η προπαγάνδα και παραπληροφόρηση μέσω της πρόληψης κατά των ψευδών ειδήσεων, ειδικότερα σε μια περίοδο ψηφισμού νέας κυβέρνησης σε ένα Δημοκρατικό Κράτος. Πιο συγκεκριμένα, εξηγείται αναλυτικά η προεπεξεργασία των δεδομένων και η υλοποίηση του LSTM αλγορίθμου.

2. *Akhtar, P., Ghouri, A.M., Khan, H.U.R. et al. “Detecting fake news and*

disinformation using artificial intelligence and machine learning to avoid supply chain disruptions"

Τα εμπόδια που παρουσιάζει η παραπληροφόρηση στην Εφοδιαστική Αλυσίδα είναι ένα από τα θέματα που αναλύεται σε αυτή την έρευνα. Χρησιμοποιείται ο αλγόριθμος Μηχανικής Μάθησης Support Vector Machine (SVM), σε συνδυασμό με την Τεχνητή Νοημοσύνη.

3. S. Rastogi, D. Bansal, "A review on fake news detection 3T's: typology, time of detection, taxonomies."

   Στο παρόν γίνεται ο διαχωρισμός μεταξύ της παραπληροφόρησης, την ελλιπούς πληροφόρησης, καθώς και της σάτιρας που παρατηρείται στα μέσα μαζικής ενημέρωσης. Μέσω παραδειγμάτων εξηγείται ο αντίκτυπος της φήμης, της παραπληροφόρησης αλλα και της απάτης στο διαδίκτυο. Αναλύεται ο τρόπος αναγνώρισης τους και ανάκτησης συνόλων δεδομένων κατάλληλων για έρευνα.

4. Ihsan Ali, Mohamad Nizam Bin Ayub, Palaiahnakote Shivakumara, Nurul Fazmidar Binti Mohd Noor, "Fake News Detection Techniques on Social Media: A Survey"

   Στην έρευνα αυτή εξηγείται η ανάγκη καθώς και η διαδικασία της αναγνώρισης ψευδών ειδήσεων. Πέρα από την αναφορά σε αλγορίθμους όπως ο SVM και ο Naive Bayes Classifier, γίνεται λόγος και για τις προκλήσεις που παρουσιάζονται κατά την υλοποίηση τους.

5. Nicole O'Brien, "Machine learning for detection of fake news"

   Στο συγκεκριμένο άρθρο γίνεται συνοπτικός λόγος για την αναγνώριση spam αλλά και ψευδών ειδήσεων. Μελετήσαμε τις μεθόδους Μηχανικής Μάθησης που υλοποιούνται, πιο συγκεκριμένα τον SVM και Logistic Regression.

6. S.A. Khan, K. Shahzad, O. Shabbir, A. Iqbal "Developing a Framework for Fake News Diffusion Control (FNDC) on Digital Media (DM): A Systematic Review 2010–2022"

   Τα κύρια θέματα της παρούσας έρευνας αυτής αποτελούν: τις ψευδές ειδήσεις και διάφορες πηγές τους, ιστοσελίδες διασταύρωσης γεγονότων, την παραπληροφόρηση σαν φαινόμενο, τεχνικές Βαθιάς Μάθησης και Τεχνητής Νοημοσύνης για την ανίχνευσή τους κ.α.

ΚΕΦΑΛΑΙΟ 3 Σύνολα δεδομένων και προεπεξεργασία

Σε αυτό το κεφάλαιο θα ασχοληθούμε με τα Datasets που επιλέξαμε καθώς και τους τρόπους που τα προεπεξεργαστήκαμε προκειμένου να εκπαιδεύσουμε τα μοντέλα για τον εντοπισμό των ψευδών ή αληθών ειδήσεων. Μελετώντας τα Datasets παρατηρήσαμε ότι στην αρχική τους μορφή δεν ικανοποιούσαν τις απαιτήσεις μας προκειμένου να εκπαιδεύσουμε τα μοντέλα. Γι' αυτό χρησιμοποιήσαμε τεχνικές προεπεξεργασίας στην NLP (Natural Language Processing) όπου περιλαμβάνουν τον καθαρισμό και τη μετατροπή δεδομένων ακατέργαστου κειμένου σε μορφή που είναι κατάλληλη για τη μηχανική μάθηση. Ακόμη, χρησιμοποιήσαμε τα Word Embeddings στα μοντέλα με τα νευρωνικά δίκτυα για να αντιστοιχήσουμε τις λέξεις από το λεξιλόγιο μας με τις αντίστοιχες διανυσματικές αναπαραστάσεις λέξεων ώστε να υπάρχουν τα αρχικά βάρη στα embedding layers κατά την διάρκεια της εκπαίδευσης του μοντέλου.

Για τους σκοπούς αυτής της εργασίας, επιλέγουμε να χρησιμοποιήσουμε τρία Datasets, το Fake News Corpus Dataset [17], το WELFake Dataset [18] και το LIAR Dataset [19]. Και τα τρία Datasets επικεντρώνονται γύρω από άρθρα ειδήσεων, αντανακλώντας τη σημασία των ειδήσεων σε διάφορους τομείς. Υπάρχει ποικιλομορφία στα Datasets και γενικεύονται σε δεδομένα του πραγματικού κόσμου. Αποτελούνται από διάφορες στήλες με κυρίαρχες για την δική μας εργασία να είναι οι τίτλοι και τα κείμενα που αντιπροσωπεύουν την είδηση καθώς και οι ταμπέλες με το εάν είναι αληθής η ψευδής είδηση.

(Υποκεφάλαιο 3.1) **Ανάλυση προεπεξεργασίας**

Σε αυτό το τμήμα θα αναλύσουμε την διαδικασία προεπεξεργασίας των Datasets. Καταλήξαμε πως για όλα τα Datasets θα χρησιμοποιήσουμε την ίδια διαδικασία που αναλύουμε παρακάτω εκτός από το WELFake Dataset όπου υλοποιούμε ένα επιπλέον βήμα. Μας απασχολούν μόνο συγκεκριμένες στήλες σε κάθε Dataset που περιέχουν τα δεδομένα που χρειαζόμαστε για την εκπαίδευση των μοντέλων. Χρειαζόμαστε τις στήλες που απέμειναν στα Datasets που περιέχουν κείμενο, εκτός από τη στήλη με τις ταμπέλες για κάθε είδηση (αληθής ή ψευδής είδηση), να έχουν την μορφή μόνο κειμένου αφαιρώντας όλα τα υπόλοιπα στοιχεία ή και θόρυβο που μπορεί να προκύπτουν στα κείμενα. Οποιαδήποτε μορφή ελλιπών δεδομένων ή κενών κειμένων σε κάθε σειρά στα Datasets θα αφαιρείται ολόκληρη η σειρά. Όταν θα χρησιμοποιείται μία τεχνική για Feature Extraction θα αναφέρεται ξεχωριστά στο μοντέλο που χρησιμοποιείται, ενώ αντίστοιχα και για οποιοδήποτε Word Embedding μοντέλο.

Στο WELFake Dataset παρατηρήθηκε συχνή εμφάνιση της ταμπέλας '(Reuters)', ακολουθούμενη από διάφορες τοποθεσίες και μια παύλα ("-"), η οποία χώριζε την πηγή της είδησης από το κυρίως κείμενο. Ειδήσεις που περιείχαν αυτό το μοτίβο στο Dataset ήταν συνήθως αληθείς, με αποτέλεσμα να δημιουργείται τάση-προκατάληψη (bias) στα εκπαιδευμένα μοντέλα να αναγνωρίζουν παρόμοιου τύπου κείμενα ως αληθή στις περισσότερες περιπτώσεις, χωρίς αυτό να είναι ορθό. Για αυτό το λόγο, κατά την επεξεργασία του WELFake Dataset, χωρίσαμε τα κείμενα στις ταμπέλες των πηγών και στα κύρια μέρη τους, αμέσως μετά τη ανίχνευση παύλας στην αρχή του κειμένου. Έπειτα, αφαιρέσαμε κάθε ταμπέλα ώστε να εξαλείψουμε αυτού του είδους τάση από το σύνολο δεδομένων μας και να φέρουμε εις πέρας καλύτερα ποσοστιαία αποτελέσματα. Αυτό λοιπόν είναι το πρώτο βήμα που υλοποιούμε μόνο στο WELFake Dataset.

Για τα υπόλοιπα Datasets το αρχικό βήμα είναι να αφαιρέσουμε οποιαδήποτε γραμμή η οποία περιέχει ελλιπή ή απροσδιόριστα δεδομένα που μπορεί να προέκυψαν καταλάθος από τους δημιουργούς ή κατά το φόρτωμα. Συνεχίζουμε με συνενώνοντας την στήλη των τίτλων, αν υπάρχει, με την στήλη των κειμένων με τη μορφή τίτλος, κενό, υπόλοιπο κείμενο και αφαιρούμε την στήλη των τίτλων καθώς και όλες τις υπόλοιπες στήλες κρατώντας μόνο την στήλη που περιέχει τα κείμενα από τα νέα και την στήλη με τις ταμπέλες για το εάν η είδηση είναι αληθής η ψευδής. Έπειτα, αφαιρούμε τις γραμμές κειμένων που δεν έχουν κείμενο (δηλαδή είναι εντελώς κενές), μετατρέπουμε τους Unicode χαρακτήρες σε ASCII καθώς μπορεί να υπάρχουν 'περίεργοι' ή άγνωστοι μη αγγλικοί χαρακτήρες που μπορεί να προήλθαν κατά λάθος από τους δημιουργούς. Λέξεις όπως το he'll ή το you're ή το can't μετατρέπονται σε he will, you are και cannot αντίστοιχα, το βήμα αυτό το

θεωρούμε σημαντικό ώστε να υπάρχει συνοχή στα κείμενα και να εκπαιδεύονται τα μοντέλα πάνω σε λέξεις που δεν συνενώνονται τα φωνήεντα ή οι δίφθογγοι σε ένα φωνήεν ή ένα δίφθογγο, ενώ ακόμη μειώνεται και ο θόρυβος στο Dataset. Συνεχίζουμε μετατρέποντας όλα τα γράμματα σε πεζά, αφαιρούμε όλους τους συνδέσμους, τους ειδικούς χαρακτήρες, τις ετικέτες html, όλους τους αριθμούς και τελικά τα σημεία στίξεως που μπορεί να απέμειναν από την αφαίρεση των ειδικών χαρακτήρων. Τελικά αποθηκεύουμε το προεπεξεργασμένο Dataset για μελλοντική χρήση.

(Υποκεφάλαιο 3.2) **Ανάλυση συνόλων δεδομένων**

Σε αυτό το τμήμα θα μιλήσουμε για κάθε ένα Dataset ξεχωριστά. Αξίζει να αναφέρουμε ότι τα Datasets έχουν τρία διαφορετικά μεγέθη, πολύ μεγάλο, μεσαίο και μικρό. Τα Datasets Για το καθένα θα τα παρουσιάσουμε με αρχική τους μορφή και έπειτα με την τελική τους μορφή, δηλαδή αφού υλοποιήσουμε τα βήματα που αναφέρθηκαν παραπάνω για την προεπεξεργασία.

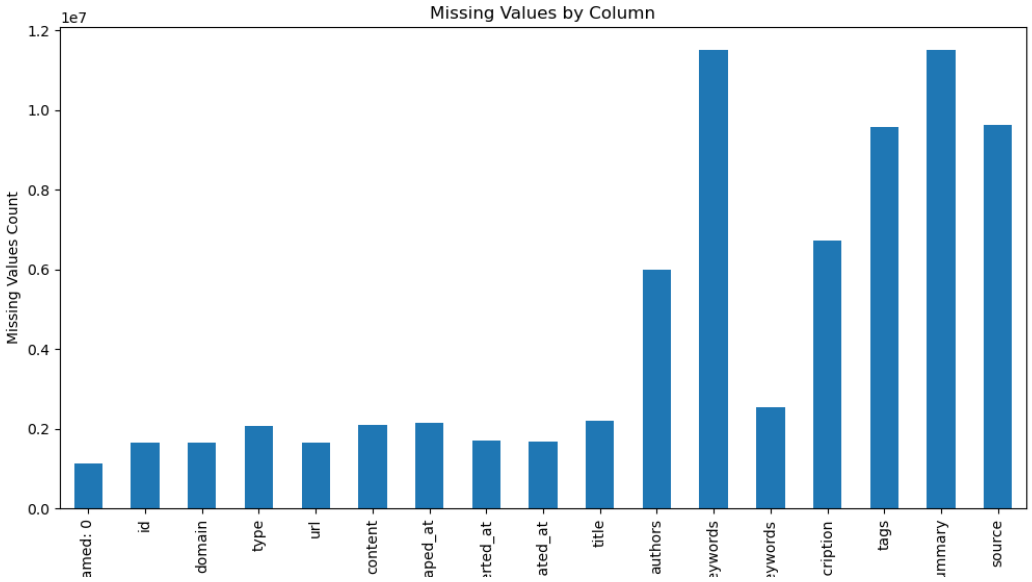
(Ενότητα 3.2.α) **Fake News Corpus σύνολο δεδομένων**

Το συγκεκριμένο Dataset σύμφωνα με τον δημιουργό “είναι μια τεράστια συλλογή από άρθρα από διάφορες πηγές κυρίως από τα domains του opensources.co όπου έχουν αποδοθεί σε αυτά μια ετικέτα που σχετίζεται με τον τομέα του άρθρου” [20]. Το Dataset περιλαμβάνει κείμενα των άρθρων ειδήσεων, τα οποία μπορούν να χρησιμοποιηθούν ως είσοδοι για την εκπαίδευση μοντέλων μηχανικής εκμάθησης ή τη εξαγωγή διαφόρων εργασιών επεξεργασίας φυσικής γλώσσας (NLP) όπως οι ενσωματώσεις λέξεων (Word Embeddings) που υλοποιούμε αργότερα. Λόγω του μεγέθους του και της ποικιλομορφίας σε κείμενα άρθρων το θεωρούμε μια πολύτιμη πηγή για τη μελέτη των χαρακτηριστικών και των προτύπων των άρθρων ψευδών ειδήσεων και την εξερεύνηση τεχνικών και μοντέλων βαθιάς μάθησης για τον για αυτοματοποιημένο εντοπισμό.

Φορτώνουμε και επισκοπούμε το DataFrame (εικόνα 3.1) παρατηρώντας πως υπάρχουν εκατομμύρια γραμμές καθώς και τυπώνοντας τον αριθμό των γραμμών που περιέχουν null (εικόνα 3.2) διακρίνουμε πως υπάρχουν αρκετές ‘προβληματικές’, για το DataFrame, γραμμές.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11558723 entries, 0 to 11558722
Data columns (total 17 columns):
 #   Column            Dtype
---  ------            -----
 0   Unnamed: 0        object
 1   id                object
 2   domain            object
 3   type              object
 4   url               object
 5   content           object
 6   scraped_at        object
 7   inserted_at       object
 8   updated_at        object
 9   title             object
 10  authors           object
 11  keywords          object
 12  meta_keywords     object
 13  meta_description  object
 14  tags              object
 15  summary           object
 16  source            object
dtypes: object(17)
memory usage: 1.5+ GB
```

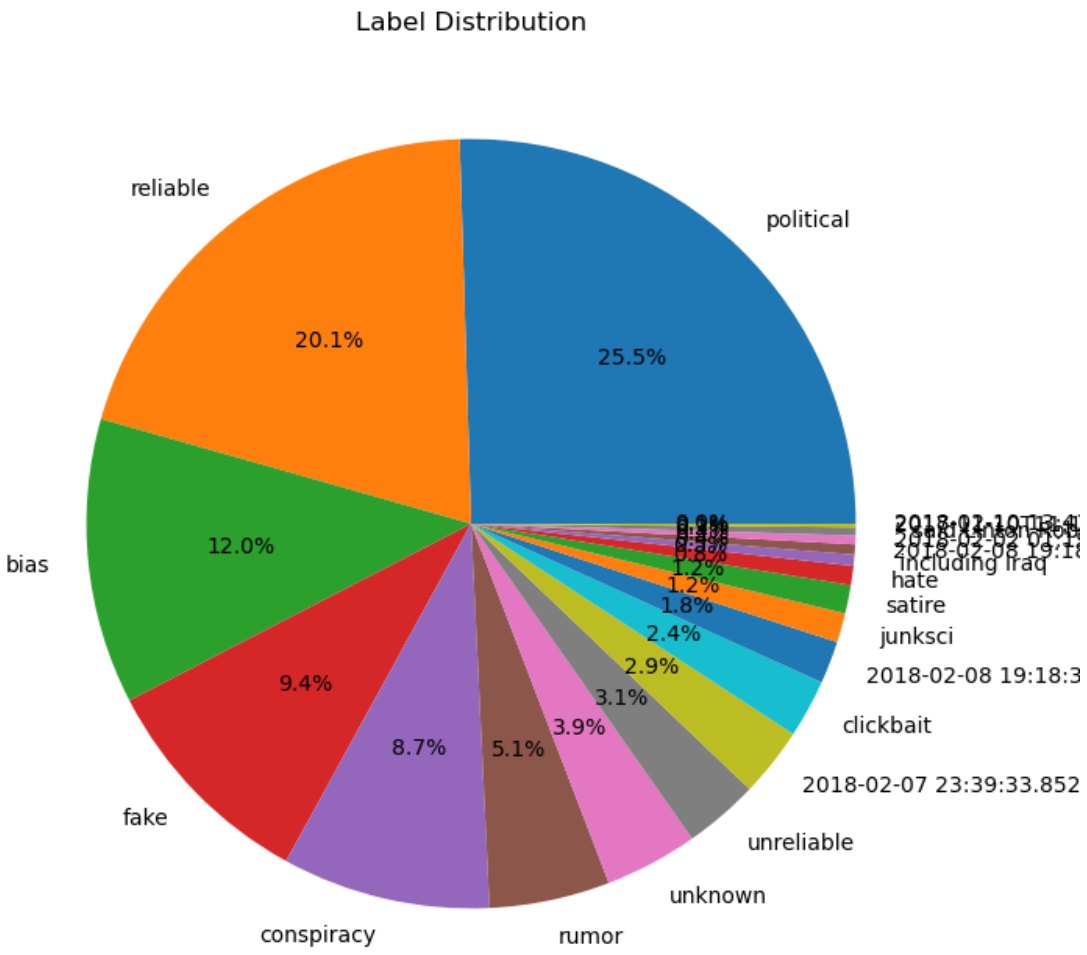
Εικόνα 3.1: Συνοπτικής περίληψης των βασικών πληροφοριών του συνόλου δεδομένων Fake News Corpus. Παρέχονται πληροφορίες όπως ο αριθμός καταχωρίσεων, τα ονόματα στηλών και η μνήμη που χρησιμοποιείται



Εικόνα 3.2: Η γραφική παράσταση ράβδων αναπαριστά πόσες τιμές που λείπουν υπάρχουν σε κάθε στήλη του συνόλου δεδομένων

Πρέπει να αφαιρέσουμε και τις στήλες που δεν μας απασχολούν όπως η στήλη “authors” ή “meta_keywords”, και γι’ αυτό κρατάμε μόνο τις στήλες “type”, “title”, “content” όπου περιέχουν τις ταμπέλες, τους τίτλους και τα κείμενα των άρθρων ειδήσεων αντίστοιχα. Έπειτα, συνενώνουμε τον τίτλο με τα κείμενα των άρθρων και αφαιρούμε και την στήλη “title”.

Αναπαριστώντας τις αναλογίες που υπάρχουν οι ταμπέλες (εικόνα 3.3) παρατηρούμε πως υπάρχουν ταμπέλες όπου δεν μας είναι ξεκάθαρο για το εάν μια είδηση είναι αληθής ή ψευδής ενώ ακόμη μας προτρέπει σε ταξινόμησης πολλαπλών τάξεων (Multiclass Classification). Παρόλα αυτά ο δημιουργός μας παρέχει έναν πίνακα (πίνακας 3.1) που εξηγεί τι τύπος είναι κάθε ταμπέλα, πόσες φορές εμφανίζεται και μια μικρή περιγραφή που μας υποδεικνύει για το εάν πρόκειται για μια πλήρως αληθής ή ψευδής είδηση.

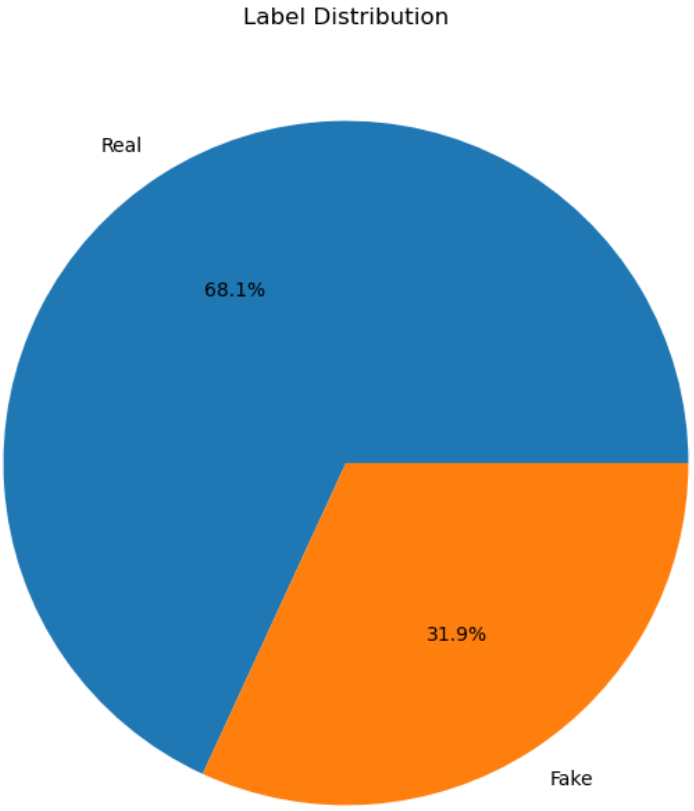


Εικόνα 3.3: Το γράφημα πίτας αντιπροσωπεύει την κατανομή των ετικετών στο σύνολο δεδομένων, δείχνοντας την αναλογία κάθε κατηγορίας ετικετών σε σχέση με τον συνολικό αριθμό παρουσιών

Type	Tag	Count (so far)	Description
Fake News	fake	928,083	Sources that entirely fabricate information, disseminate deceptive content, or grossly distort actual news reports
Satire	satire	146,080	Sources that use humor, irony, exaggeration, ridicule, and false information to comment on current events.
Extreme Bias	bias	1,300,444	Sources that come from a particular point of view and may rely on propaganda, decontextualized information, and opinions distorted as facts.
Conspiracy Theory	conspiracy	905,981	Sources that are well-known promoters of kooky conspiracy theories.
State News	state	0	Sources in repressive states operating under government sanction.
Junk Science	junksci	144,939	Sources that promote pseudoscience, metaphysics, naturalistic fallacies, and other scientifically dubious claims.
Hate News	hate	117,374	Sources that actively promote racism, misogyny, homophobia, and other forms of discrimination.
Clickbait	clickbait	292,201	Sources that provide generally credible content, but use exaggerated, misleading, or questionable headlines, social media descriptions, and/or images.
Proceed With Caution	unreliable	319,830	Sources that may be reliable but whose contents require further verification.
Political	political	2,435,471	Sources that provide generally verifiable information in support of certain points of view or political orientations.
Credible	reliable	1,920,139	Sources that circulate news and information in a manner consistent with traditional and ethical practices in journalism (Remember: even credible sources sometimes rely on clickbait-style headlines or occasionally make mistakes. No news organization is perfect, which is why a healthy news diet consists of multiple sources of information).

Πίνακας 3.1: Αναλυτική περιγραφή των ταμπελών του συνόλου δεδομένων Fake News Corpus

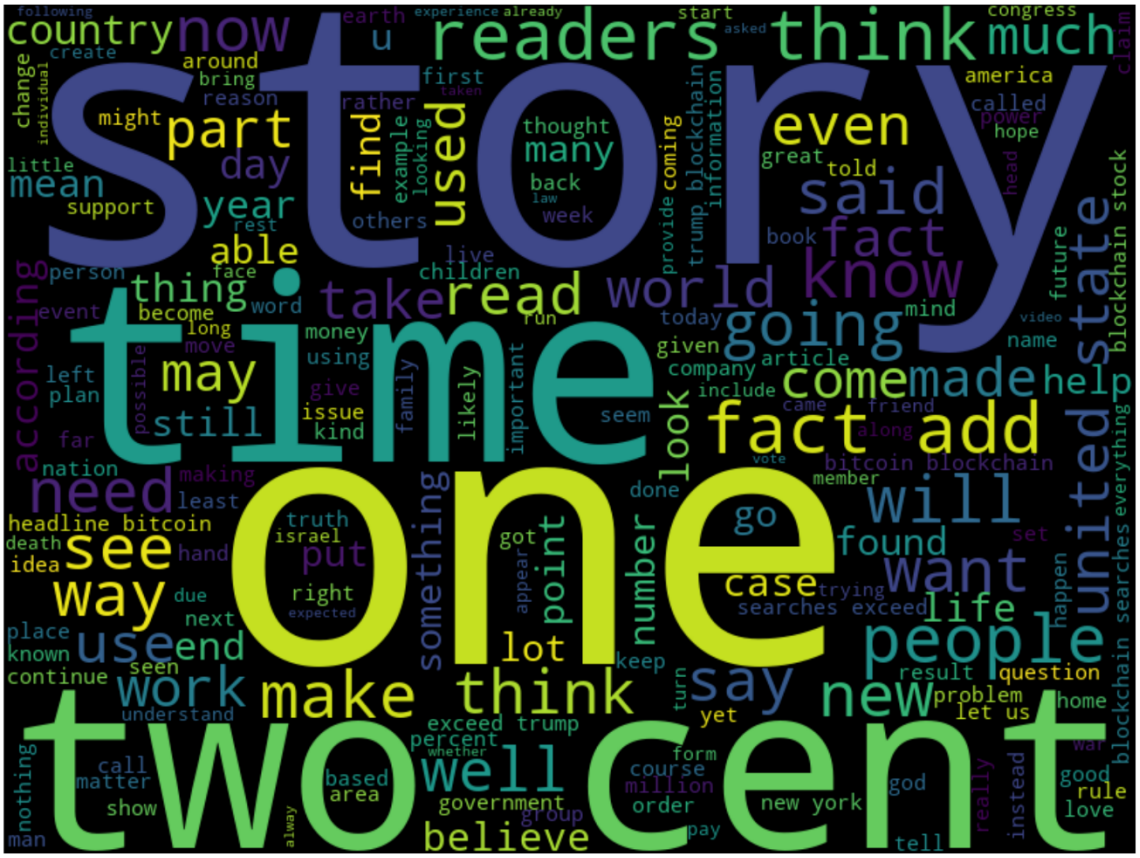
Μελετώντας τον πίνακα 3.1 καταλήγουμε να κρατήσουμε μόνο τις ταμπέλες “reliable” ως την ταμπέλα με τις “αληθής” ειδήσεις και “fake” ως την ταμπέλα με τις “ψευδής” ειδήσεις. Συνεχίζουμε, εφαρμόζοντας τα βήματα της προεπεξεργασίας που αναφέρθηκαν και αποθηκεύουμε για μελλοντική χρήση. Έτσι, έχουμε δυαδικές ταμπέλες αλλά πλέον στο Dataset υπάρχει μια ανομοιομορφία όπως φαίνεται στην εικόνα 3.4, τα αληθή νέα είναι αρκετά περισσότερα από τα ψευδή.



Εικόνα 3.4: Η αναλογία των δυαδικών ταμπελών του συνόλου δεδομένων Fake News Corpus

Τελικά, στις εικόνες 3.5 και 3.6 αναπαριστούμε τις συχνότερες λέξεις που εμφανίζονται στην στήλη με τα κείμενα για τις δύο ταμπέλες. Για τις συχνότερες λέξεις στις αληθής ταμπέλες χρειαστήκαμε να χρησιμοποιήσουμε το 50% των δεδομένων λόγω περιορισμένης μνήμης.

Fake



Εικόνα 3.5: Οι συχνότερες λέξεις των κειμένων που έχουν ταξινομηθεί ως ψευδή νέα μετά την προεπεξεργασία του συνόλου δεδομένων Fake News Corpus

Εικόνα 3.6: Οι συχνότερες λέξεις των κειμένων που έχουν ταξινομηθεί ως αληθή νέα μετά την προεπεξεργασία του συνόλου δεδομένων Fake News Corpus

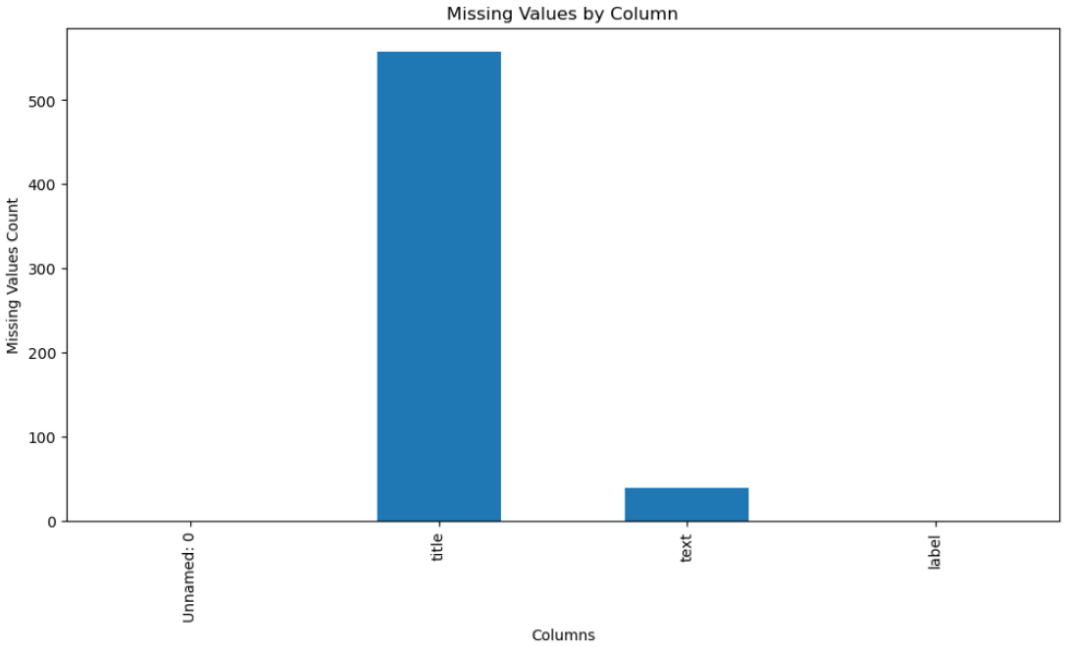
(Ενότητα 3.2.β) **WELFake σύνολο δεδομένων**

Το WELFake Dataset είναι ένα Dataset που χρησιμοποιείται συνήθως στην έρευνα που σχετίζεται με τον εντοπισμό και την ταξινόμηση ψευδών ειδήσεων. Είναι μια συλλογή ειδησεογραφικών άρθρων που επισημαίνονται είτε ως “ψευδής” ή “αληθής”. Το μέγεθος του WELFake Dataset το θεωρούμε ως μεσαίο σε σχέση με τα υπόλοιπα δύο. Το Dataset στοχεύει να παρέχει μια ισορροπημένη αναπαράσταση δειγμάτων ψευδών και πραγματικών ειδήσεων και στοχεύει στην δυαδική ταξινόμηση, με δύο ετικέτες 0 και 1. Το σύνολο δεδομένων WELFake μπορεί να χρησιμοποιηθεί για διάφορες εφαρμογές που σχετίζονται με τον εντοπισμό, την ταξινόμηση και την ανάλυση ψευδών ειδήσεων και το θεωρούμε ιδανικό για την δυαδική ταξινόμηση ειδήσεων. Ερευνητές και οι επαγγελματίες έχουν χρησιμοποιήσει αυτό το Dataset για να αναπτύξουν και να αξιολογήσουν μοντέλα μηχανικής [18], ή υβριδικά μοντέλα που περιέχουν μοντέλα επεξεργασίας φυσικής γλώσσας [21] για τον εντοπισμό ψευδών ειδήσεων.

Φορτώνουμε και επισκοπούμε το DataFrame (εικόνα 3.7) για να δούμε από πόσες σειρές αποτελείται, πόσες σειρές είναι μη ελλιπής για κάθε στήλη, καθώς και τον τύπο των δεδομένων. Έτσι, έχουμε μια συνοπτική εικόνα του DataFrame που φορτώσαμε ώστε να προχωρήσουμε στην ανάλυση του για να το ετοιμάσουμε για τα μοντέλα που χρησιμοποιήσαμε για το classification. Τυπώνουμε και τον αριθμό των γραμμών που περιέχουν null σε κάθε στήλη.

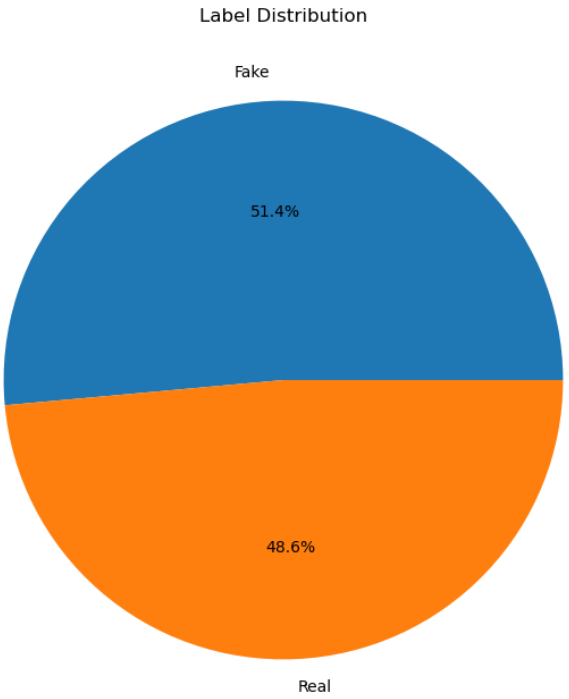
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 72134 entries, 0 to 72133
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Unnamed: 0  72134 non-null  int64
 1   title       71576 non-null  object
 2   text        72095 non-null  object
 3   label       72134 non-null  int64
dtypes: int64(2), object(2)
memory usage: 2.2+ MB
```

Εικόνα 3.7: Συνοπτικής περίληψης των βασικών πληροφοριών του συνόλου δεδομένων WELFake

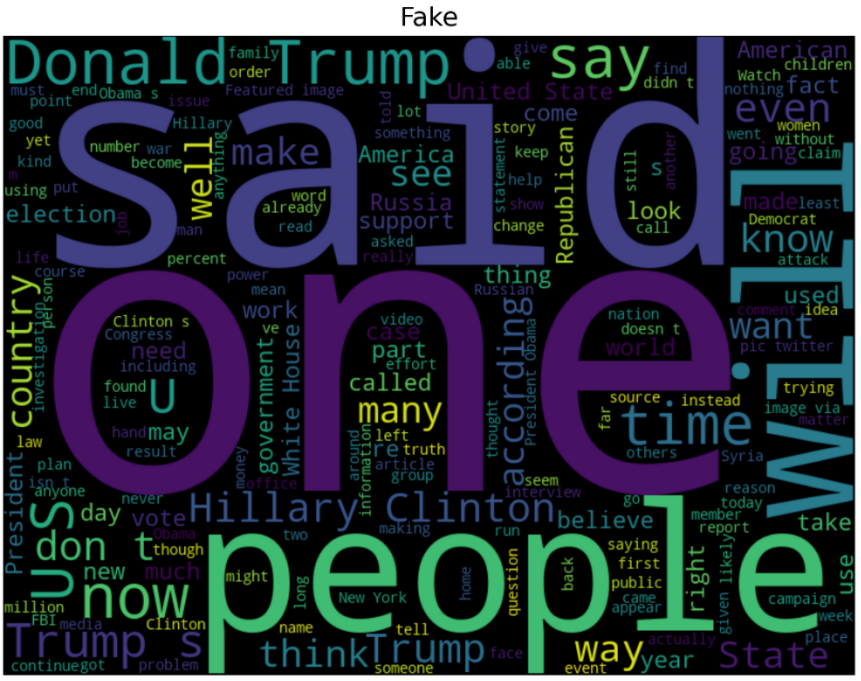


Εικόνα 3.8: Γραφική παράσταση ράβδων για τις ελλιπής τιμές για κάθε στήλη του WELFake συνόλου δεδομένων

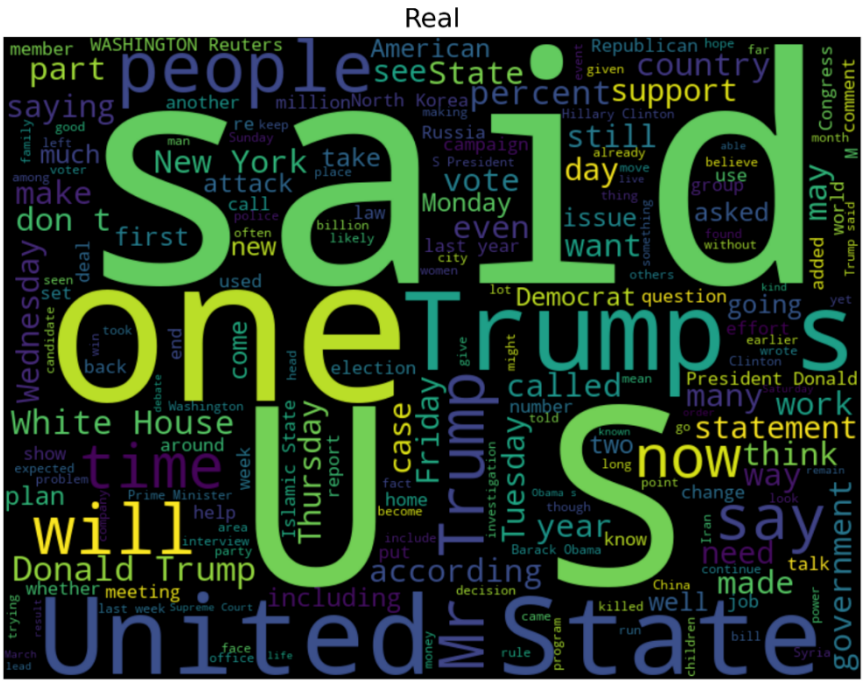
Έπειτα, αναπαριστούμε σε τι αναλογίες υπάρχουν οι ταμπέλες, στο συγκεκριμένο Dataset λόγω της δυαδικότητας των ταμπελών δεν χρειάζεται να αφαιρέσουμε κάποια ταμπέλα η οποία μπορεί να μην μας ξεκαθαρίζει για το εάν μια είδηση είναι αληθής ή ψευδής. Επιπλέον, παρατηρήσαμε ότι οι δημιουργοί αναφέρουν πως στις ταμπέλες το 0 ισούται με ψεύτικη είδηση και 1 ισούται με αληθής είδηση, αλλά διαβάζοντας το Dataset υπήρχαν πολλά άρθρα όπου ήταν προφανές πως ίσχυε το αντίθετο. Για παράδειγμα η πλειονότητα των άρθρων των Reuters και των New York Times επισημαίνεται ως 0 δηλαδή ψευδής είδηση. Για αυτόν τον λόγο, εμείς χρησιμοποιούμε τις ταμπέλες ανάποδα δηλαδή 1 για ψευδής είδηση και 0 για αληθής είδηση. Συνεχίζουμε, ακολουθώντας τα βήματα που περιγράψαμε στο τμήμα 3.1 με την προεπεξεργασία και έχουμε το ολοκληρωμένο προεπεξεργασμένο Dataset το οποίο από εδώ και πέρα θα χρησιμοποιούμε στα μοντέλα που υιοθετήσαμε. Από την εικόνα 3.9 βλέπουμε πως η κατανομή των ετικετών είναι γενικά ισορροπημένη, πράγμα που σημαίνει ότι περίπου τα μισά από τα άρθρα φέρουν την ετικέτα “Ψευδής” και τα άλλα μισά ως "Αληθής". Αυτή η ισορροπημένη κατανομή επιτρέπει την αμερόληπτη αξιολόγηση των μοντέλων και διασφαλίζει την ίση εκπροσώπηση και των δύο κατηγοριών. Τελικά, στις εικόνες 3.10 και 3.11 αναπαριστούμε τις συχνότερες λέξεις που εμφανίζονται στην στήλη με τα κείμενα για τις δύο ταμπέλες.



Εικόνα 3.9: Ισορροπημένη κατανομή των ταμπελών του WELFake συνόλου



Εικόνα 3.10: Οι συχνότερες λέξεις των κειμένων που έχουν ταξινομηθεί ως ψευδή νέα μετά την προεπεξεργασία του συνόλου δεδομένων WELFake



Εικόνα 3.11: Οι συχνότερες λέξεις των κειμένων που έχουν ταξινομηθεί ως αληθή νέα μετά την προεπεξεργασία του συνόλου δεδομένων WELFake

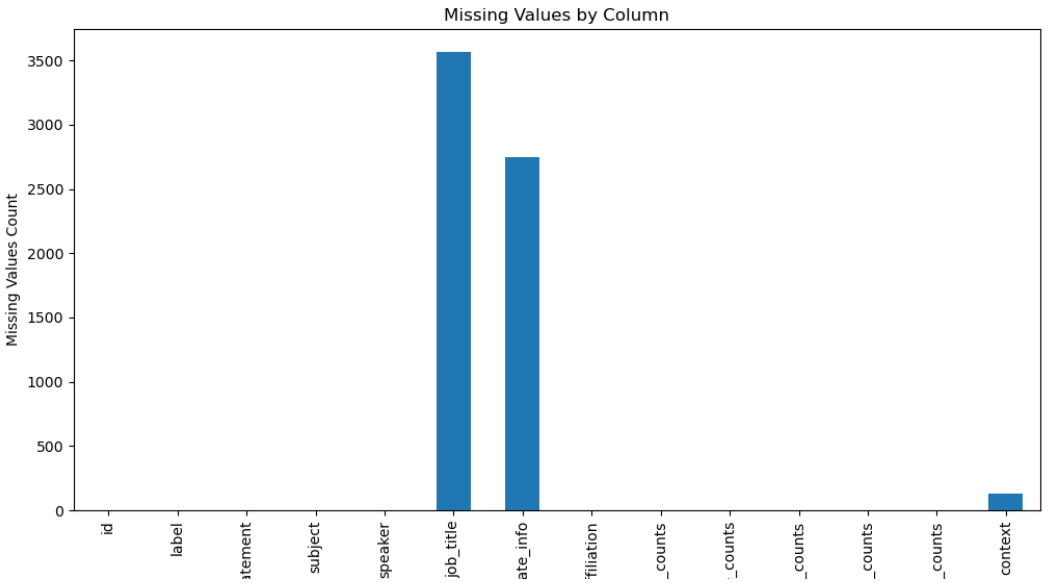
(Ενότητα 3.2.γ) **LIAR σύνολο δεδομένων**

Το LIAR Dataset είναι ένα Dataset με ευρεία αναφορά στον τομέα του εντοπισμού και της ανάλυσης ψευδών ειδήσεων. Έχει σχεδιαστεί για να διευκολύνει την έρευνα στον εντοπισμό και την ταξινόμηση ψευδών ειδήσεων καθώς μπορεί επίσης να χρησιμοποιηθεί για έρευνα ελέγχου γεγονότων[3]. Το συγκεκριμένο Dataset δημιουργήθηκε με τη συλλογή δηλώσεων από διάφορες πηγές, συμπεριλαμβανομένων ομιλιών πολιτικών, συνεντεύξεων, ειδησεογραφικών άρθρων και ιστοτόπων ελέγχου στοιχείων. Τα άρθρα ελέγχθηκαν από πολλές πηγές και αντιστοιχίστηκαν με ετικέτες από το PolitiFact, όπου είναι ένας μη κερδοσκοπικός οργανισμός που ελέγχει την εγκυρότητα και την ακρίβεια διαφόρων ειδήσεων. Το μέγεθος του LIAR Dataset το θεωρούμε ως μικρό σε σχέση με τα υπόλοιπα δύο. Αποτελείται φυσικά από ειδησεογραφικά άρθρα που έχουν επισημανθεί με διαφορετικά επίπεδα ακρίβειας, όπως αναφέρθηκε πιο πάνω, που αυτά είναι: "True", "Mostly True", "Half True", "Barely True", "False," και "Pants on Fire". Το Dataset περιλαμβάνει τόσο δηλώσεις-κείμενα όσο και τις αντίστοιχες ετικέτες τους που υποδεικνύουν το επίπεδο αληθείας. Ακόμη και εάν το LIAR Dataset παρέχει πολλαπλές ετικέτες και υποδεικνύεται για ταξινόμησης πολλαπλών τάξεων (Multiclass Classification), εμείς επιλέγουμε να κρατήσουμε αυτές που δηλώνουν ρητά εάν η είδηση/δήλωση είναι ψευδής ή αληθής. Η κατανομή, παρόλα αυτά, των ετικετών εντός του Dataset ποικίλλει, αλλά είναι γενικά ισορροπημένη ώστε να περιλαμβάνει έναν συνδυασμό αληθών και ψευδών δηλώσεων στις διάφορες κλάσεις.

Το συγκεκριμένο Dataset έρχεται ήδη χωρισμένο σε τρία αρχεία, ένα για την εκπαίδευση του μοντέλου, ένα για την δοκιμή και ένα που μπορεί να χρησιμοποιηθεί για την επικύρωση κατά την διάρκεια της εκπαίδευσης. Στο συγκεκριμένο Dataset φορτώνουμε και συνενώνουμε τα DataFrames πριν τα επεξεργαστούμε. Επισκοπούμε αντίστοιχα όπως φαίνεται στην εικόνα 3.12 για να πληροφορηθούμε για το Dataset μήπως και χρειαστεί να προσέξουμε σε κάποιο βήμα (να αφαιρέσουμε ή να προσθέσουμε κάποιο βήμα) ώστε να έχουμε το επιθυμητό DataFrame. Επιπλέον, αναπαριστούμε τις πιθανών ελλιπής τιμές που μπορεί να εμφανίζονται σε κάποια στήλη (εικόνα 3.13).

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12791 entries, 0 to 12790
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   id                    12791 non-null  object
 1   label                 12791 non-null  object
 2   statement             12791 non-null  object
 3   subject               12789 non-null  object
 4   speaker               12789 non-null  object
 5   job_title             9224 non-null   object
 6   state_info            10042 non-null  object
 7   party_affiliation     12789 non-null  object
 8   barely_true_counts    12789 non-null  float64
 9   false_counts          12789 non-null  float64
 10  half_true_counts      12789 non-null  float64
 11  mostly_true_counts    12789 non-null  float64
 12  pants_on_fire_counts  12789 non-null  float64
 13  context               12660 non-null  object
dtypes: float64(5), object(9)
memory usage: 1.4+ MB
```

Εικόνα 3.12: Συνοπτικής περίληψης των βασικών πληροφοριών του συνόλου δεδομένων LIAR



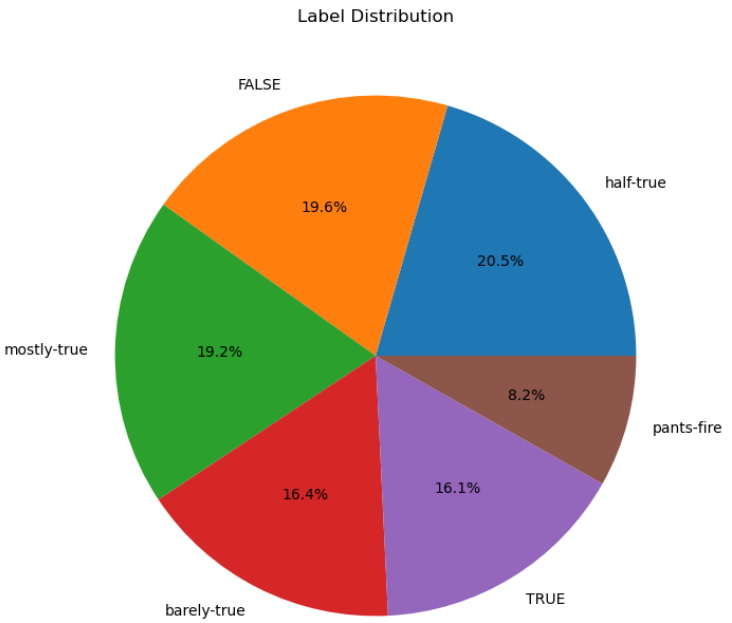
Εικόνα 3.13: Γραφική παράσταση ράβδων για τις ελλιπής τιμές για κάθε στήλη του LIAR συνόλου δεδομένων

Υπάρχουν στήλες στο DataFrame όπου δίνουν πληροφορίες όπως, τι είδους είδηση είναι, από ποιον ειπώθηκε ή από ποιον πήραν την συνέντευξη, σε ποιο κώμα ανήκει κ.λπ. όπου είναι πληροφορίες που δεν μας εξυπηρετούν για τον σκοπό αυτής της εργασίας γι' αυτό τις αφαιρούμε από το Dataset και κρατάμε μόνο τις στήλες των κειμένων και τις ετικέτες.

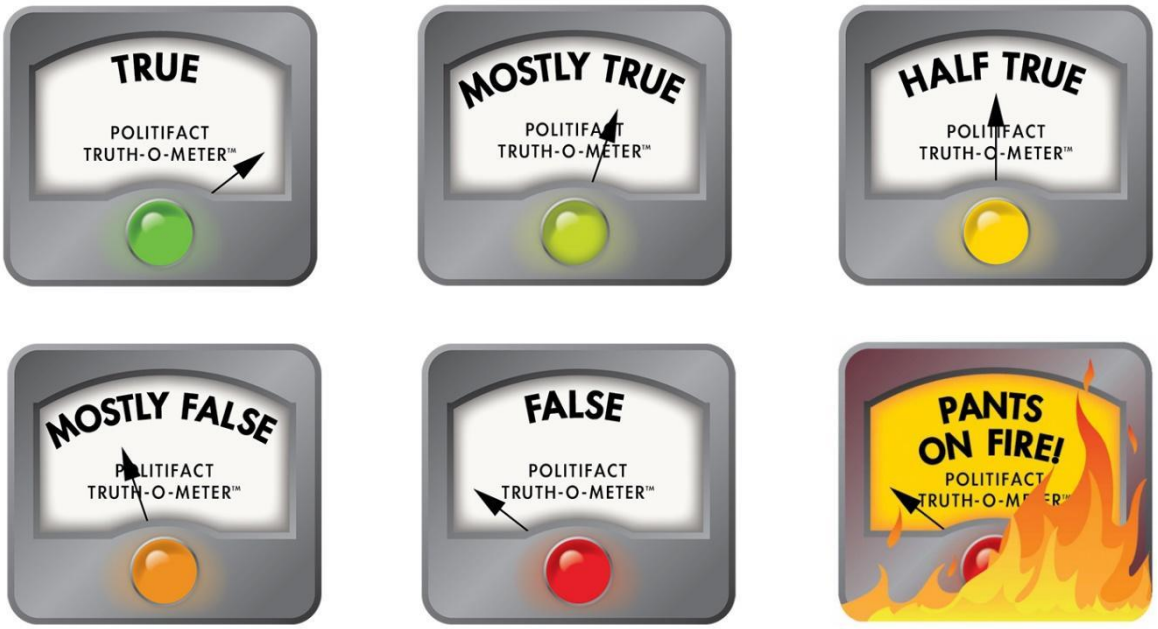
Αναπαριστώντας στην εικόνα 3.14 τις αναλογίες που υπάρχουν στις ταμπέλες του DataFrame βλέπουμε ότι υπάρχουν ταμπέλες που δεν μας διευκρινίζουν ρητά την ακεραιότητα της αλήθειας ή του ψέματος. Συγκεκριμένα στην εικόνα 3.15 βλέπουμε τον δείκτη αλήθειας (truth-o-meter) που υπάρχει στην σελίδα του politifact[1] και σύμφωνα με τον πίνακα βαθμολογίας αξιολογούμε ποιες ταμπέλες είναι ιδανικές για εμάς και ποιες μπορούμε να αφαιρέσουμε:

[1] https://www.politifact.com/truth-o-meter/

- **True**: Η δήλωση είναι ακριβής και δεν λείπει τίποτα σημαντικό
- **Mostly-True**: Η δήλωση είναι ακριβής αλλά χρειάζεται διευκρίνιση ή πρόσθετες πληροφορίες
- **Half True**: Η δήλωση είναι εν μέρει ακριβής, αλλά παραλείπει σημαντικές λεπτομέρειες ή δεν υπάρχει συνδετική δομή/συμφραζόμενα.
- **Mostly False**: Η δήλωση περιέχει ένα στοιχείο αλήθειας αλλά αγνοεί κρίσιμα γεγονότα που θα έδιναν διαφορετική άποψη.
- **False**: Η δήλωση δεν είναι ακριβής.
- **Pants on Fire**: Η δήλωση δεν είναι ακριβής και κάνει έναν γελοίο ισχυρισμό.

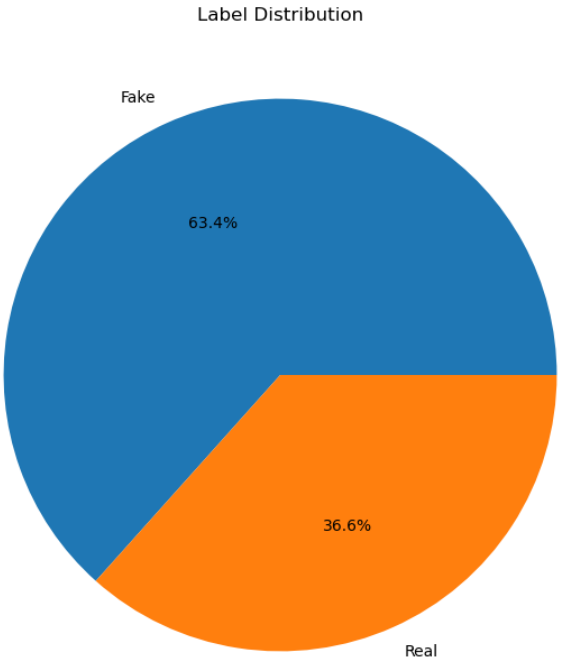


Εικόνα 3.14: Γράφημα πίτας που αντιπροσωπεύει την κατανομή των ετικετών στο σύνολο δεδομένων LIAR

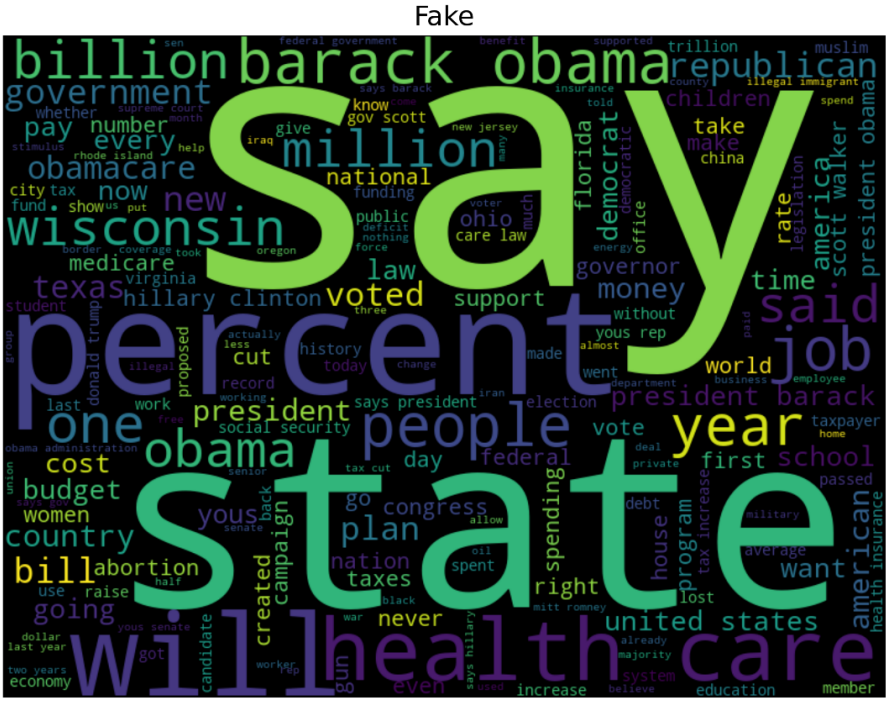


Εικόνα 3.15: Δείκτης αλήθειας (truth-o-meter) που υποδεικνύει τις ταμπέλες που μπορεί να αντιστοιχιστεί μια είδηση/δήλωση/ισχυρισμό

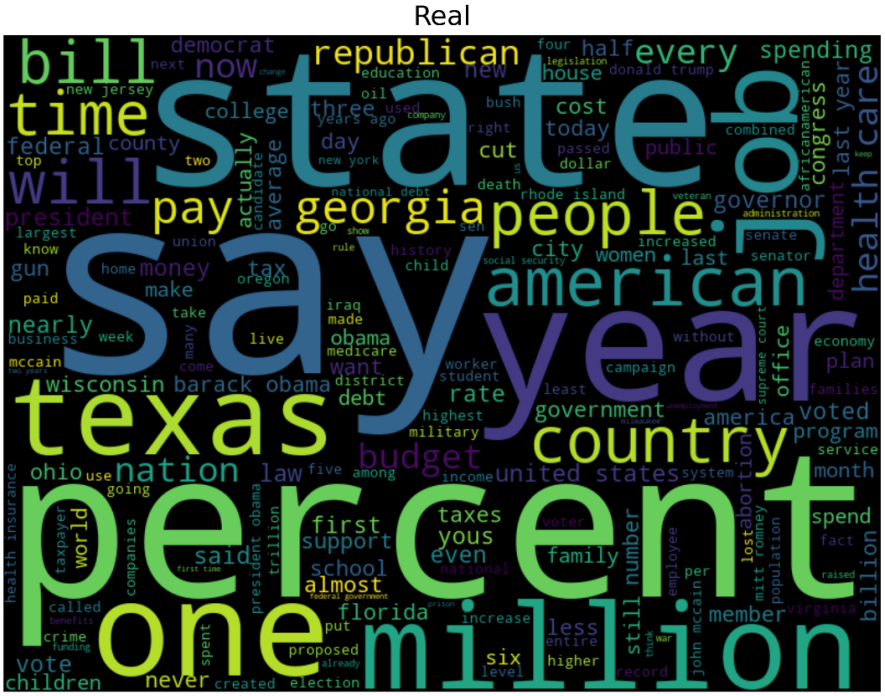
Σύμφωνα με τα παραπάνω επιλέγουμε λόγω δυαδικής ταξινόμησης τις ταμπέλες που δείχνουν ότι μια είδηση είναι απολύτως αληθής ή απολύτως ψευδής. Αυτές είναι: True και την αντιστοιχούμε στην ταμπέλα "αληθής" και False, Pants on Fire όπου τις αντιστοιχούμε στην ταμπέλα "ψευδής". Συνεχίζουμε με τα βήματα της προεπεξεργασία και αποθηκεύουμε για να χρησιμοποιήσουμε αργότερα στην εκπαίδευση των μοντέλων. Με τις τροποποιημένες ταμπέλες του DataFrame σε δυαδικές παίρνουμε ξανά τις αναλογίες και παρατηρούμε στην εικόνα 3.16 πως οι υπάρχουν περισσότερες ταμπέλες με ψευδή νέα παρά με αληθή. Τέλος, αναπαριστούμε στις εικόνες 3.17 και 3.18 τις συχνότερες λέξεις που εμφανίζονται στην στήλη με τα κείμενα για τις δύο ταμπέλες.



Εικόνα 3.16: Ανομοιόμορφη αναλογία των δυαδικών ταμπελών του συνόλου δεδομένων LIAR



Εικόνα 3.17: Οι συχνότερες λέξεις των κειμένων που έχουν ταξινομηθεί ως ψευδή νέα μετά την προεπεξεργασία του συνόλου δεδομένων LIAR



Εικόνα 3.18: Οι συχνότερες λέξεις των κειμένων που έχουν ταξινομηθεί ως αληθή νέα μετά την προεπεξεργασία του συνόλου δεδομένων LIAR

ΚΕΦΑΛΑΙΟ 4 Μοντέλα που χρησιμοποιήθηκαν

Σκοπός του κεφαλαίου είναι να παρουσιάσουμε τα μοντέλα που επιλέξαμε καθώς και οποιαδήποτε τεχνική χρησιμοποιήθηκε πριν αρχίσουμε να τα εκπαιδεύουμε με απώτερο σκοπό την βελτίωση του accuracy.

Αρχικά, πρέπει να αναφέρουμε πως όλες οι μετρήσεις για τα Datasets "Fake News Corpus" και "WELFake" έχουν γίνει εκπαιδεύοντας τα μοντέλα σε δέκα χιλιάδες (10.000) δείγματα από το αντίστοιχο Dataset τα οποία έχουν χωριστεί κατάλληλα για να αντιπροσωπεύουν το ποσοστό 45% δείγματα με ταμπέλα ψευδής νέα και 55% δείγματα με ταμπέλα αληθής νέα. Όσον αφορά το LIAR Dataset το χρησιμοποιούμε ολόκληρο για την εκπαίδευση των μοντέλων. Κάθε μοντέλο το εκπαιδεύουμε για κάθε Dataset που έχουμε προεπεξεργαστεί και έπειτα προβλέπουμε τις ταμπέλες που θα επέλεγε το μοντέλο και για τα τρία Dataset. Το Fake News Corpus Dataset έχει περιορισμένο αριθμό μεγέθους που φορτώνουμε για την πρόβλεψη ενώ τα δύο υπόλοιπα Datasets τα φορτώνουμε ολόκληρα. Επιπλέον, χωρίζουμε τα δεδομένα σε εκπαίδευσης και εξέτασης χρησιμοποιώντας την βιβλιοθήκη sklearn[2] σε ποσοστά 70% και 30% αντίστοιχα.

(Υποκεφάλαιο 4.1) **Machine Learning μοντέλα**

Σε αυτό το τμήμα θα αναπτύξουμε τα Machine Learning μοντέλα που επιλέχθηκαν, τις αποδόσεις τους σε κάθε σύνολο δεδομένων καθώς και τις επιπλέον τεχνικές που επιλέξαμε θεωρώντας ότι θα πετύχουμε ένα υψηλότερο accuracy. Ένα σύστημα μηχανικής μάθησης μπορούμε να πούμε ότι εκπαιδεύεται αντί να προγραμματίζεται, καθώς του παρουσιάζονται συγκεκριμένοι είσοδοι δεδομένων σχετικοί με την εκάστοτε εργασία. Έτσι, σχηματίζει μια στατιστική δομή πάνω σε αυτό το σύνολο και επινοεί κανόνες για να αυτοματοποιήσει την εργασία**.** Σε αντίθεση όμως με τις κλασσικές τεχνικές ανάλυσης στατιστικής, οι αλγόριθμοι μηχανικής μάθησης έχουν την δυνατότητα να επεξεργαστούν μεγάλα και πολύπλοκα σύνολα δεδομένων, όπως εκατομμύριες εικόνες ή κείμενα. Η εικόνα 4.1 [22] παρουσιάζει το πως λειτουργεί ο κλασικός προγραμματισμός και η μηχανική μάθηση.

Εικόνα 4.1: Λειτουργία κλασικού προγραμματισμού έναντι της μηχανικής μάθησης

Οι περισσότεροι από τους αλγόριθμους που υλοποιούμε σε αυτήν εργασία ειδικεύονται στην ανίχνευση ψευδών ειδήσεων πάνω σε μικρότερα κείμενα, όπως άρθρα ή δημοσιεύσεις σε μέσα κοινωνικής δικτύωσης (social media), και μερικοί από αυτούς είναι ιδιαίτερα αποδοτικοί κατά την επεξεργασία φυσικής γλώσσας (NLP). Επιλέγουμε λοιπόν, να χρησιμοποιήσουμε τους εξής Machine Learning αλγορίθμους και για τους εξής λόγους:

- **Logistic Regression:** Είναι ένα μοντέλο στατιστικής και μηχανικής μάθησης και θεωρείται μια επέκταση της γραμμικής παλινδρόμησης. Χρησιμοποιείται για προβλήματα δυαδικής ταξινόμησης, όπου η μεταβλητή αποτελέσματος είναι κατηγορική και έχει μόνο δύο κατηγορίες, στην δικιά μας περίπτωση αληθής ή ψευδής. Χρησιμοποιείται ευρέως στην ταξινόμηση κειμένων λόγω της απλότητας και της ερμηνευτικότητάς του. Ενώ ακόμη, λειτουργεί καλά όταν η σχέση μεταξύ των χαρακτηριστικών εισαγωγής (δεδομένα κειμένου) και της δυαδικής ταμπέλας είναι περίπου γραμμική. Ο κύριος τρόπος λειτουργίας της είναι η

[2] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html

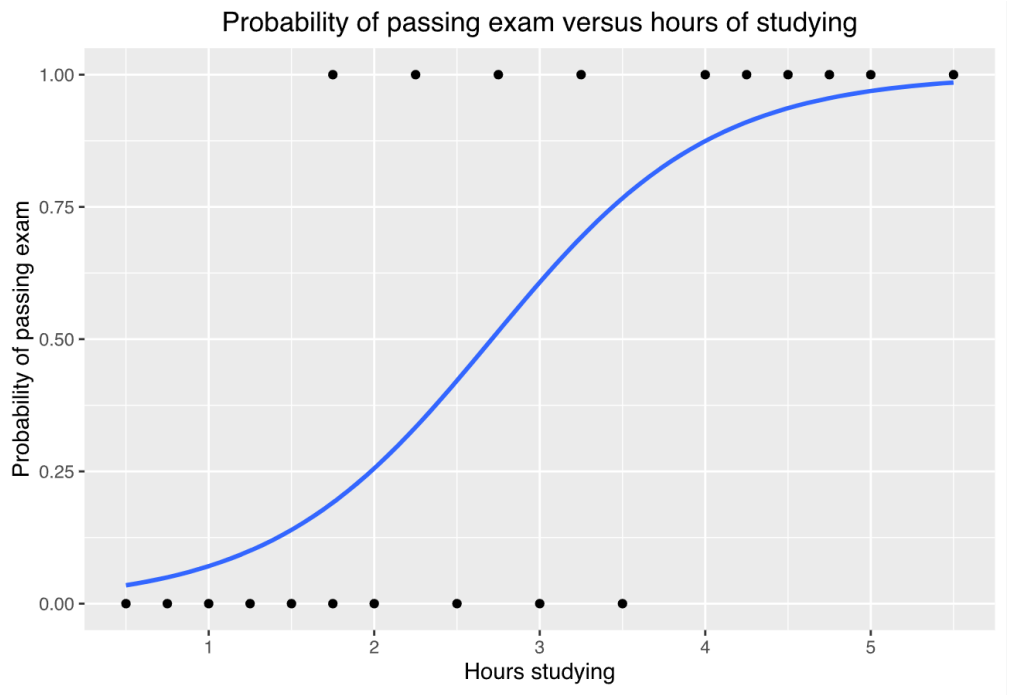
σιγμοειδής συνάρτηση (σ) όπου αντιστοιχίζει οποιονδήποτε αριθμό πραγματικής αξίας σε μια τιμή μεταξύ 0 και 1. Ο τύπος της σιγμοειδούς συνάρτησης είναι:

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Όπου z είναι ένας γραμμικός συνδυασμός των χαρακτηριστικών εισόδου και των παραμέτρων του μοντέλου:

$$z = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$$

Όπου τα $b_0, b_1, b_{2,\ldots}, b_n$ είναι οι συντελεστές (βάρη) που σχετίζονται με τα χαρακτηριστικά εισόδου $x_1, x_{2,\ldots}, x_n$. Η εικόνα 4.2[3] αναπαριστά ένα παράδειγμα γραφήματος μιας καμπύλης λογιστικής παλινδρόμησης προσαρμοσμένης σε δεδομένα.



Εικόνα 4.2: Παράδειγμα της μορφής της καμπύλης μιας λογιστικής παλινδρόμησης

- **Passive Aggressive Classifier:** είναι ένας αλγόριθμος ταξινόμησης μηχανικής μάθησης που χρησιμοποιείται κυρίως για διαδικτυακή ή σταδιακή μάθηση. Μπορεί να είναι χρήσιμος όσων αφορά την ταξινόμηση κειμένου όταν υπάρχει μια συνεχή ροή δεδομένων κειμένου, όπως άρθρα ειδήσεων. Μπορεί να προσαρμοστεί σε μεταβαλλόμενα μοτίβα, να κάνει προβλέψεις σε πραγματικό χρόνο καθώς φτάνουν νέα δεδομένα, «ενώ το μοντέλο ενημερώνεται μόνο όταν αποτυγχάνει να ταξινομήσει σωστά ένα παράδειγμα με υψηλή εμπιστοσύνη» [23]. Κάθε δείγμα δεδομένων αναπαρίσταται ως ένα διάνυσμα χαρακτηριστικών x. Ο αλγόριθμος ξεκινά με ένα αρχικό μοντέλο, συνήθως αρχικοποιημένο με μηδενικά βάρη ή μικρές τυχαίες τιμές. Το μοντέλο δίνοντας το x υπολογίζει μια πρόβλεψη y χρησιμοποιώντας την ακόλουθη εξίσωση:

$$y = sign\left(w^T \cdot x + b\right)$$

Όπου το w είναι ένα διάνυσμα βάρους, το b είναι ο όρος μεροληψίας (bias term), το w^T αντιπροσωπεύει τη μετάθεση του w και η συνάρτηση $sign$ επιστρέφει +1 εάν η έκφραση $w^T \cdot x + b$ είναι μεγαλύτερη ή ίση με το μηδέν και -1 εάν είναι αρνητική. Αφού κάνει μια

[3] https://en.wikipedia.org/wiki/Logistic_regression

πρόβλεψη, ο αλγόριθμος υπολογίζει μια απώλεια (Δ) με βάση την αληθινή ετικέτα y_t και την προβλεπόμενη ετικέτα y χρησιμοποιώντας τον παρακάτω τύπο:

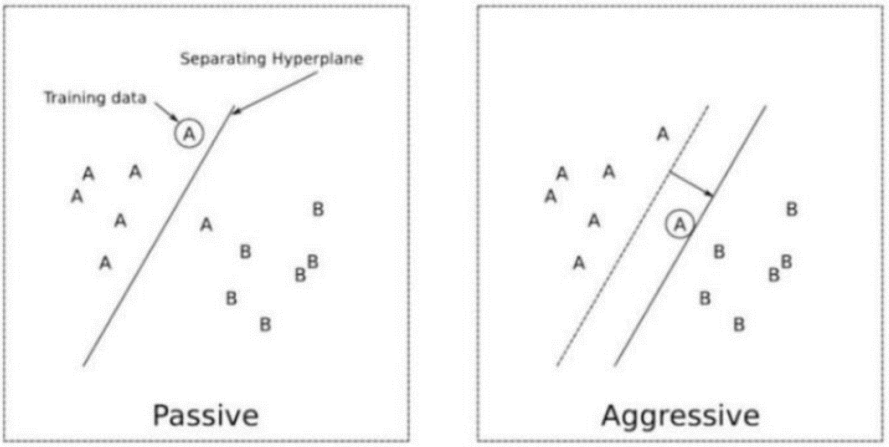
$$\Delta = \left(0, 1 - y_t \cdot y\right)$$

Εάν $\Delta = 0$ τότε σημαίνει ότι η πρόβλεψη ήταν σωστή και δεν χρειάζονται ενημερώσεις. Εάν $\Delta > 0$, δηλαδή η πρόβλεψη είναι λάθος, τότε ο αλγόριθμος ενημερώνει τις παραμέτρους του μοντέλου (w και b) για μείωση της απώλειας. Η ενημέρωση γίνεται με τέτοιο τρόπο ώστε να προσπαθεί να κάνει σωστή την πρόβλεψη για το λανθασμένα ταξινομημένο δείγμα, ενώ ελαχιστοποιεί τις αλλαγές στο μοντέλο για σωστά ταξινομημένα δείγματα. Οι κανόνες ενημέρωσης έχουν ως εξής:

Εάν $y_t \cdot y \leq 1$ τότε ενημερώνονται τα βάρη και η προκατάληψη ως εξής:

$$w = w + a \cdot \Delta \cdot x$$
$$b = b + a \cdot \Delta$$

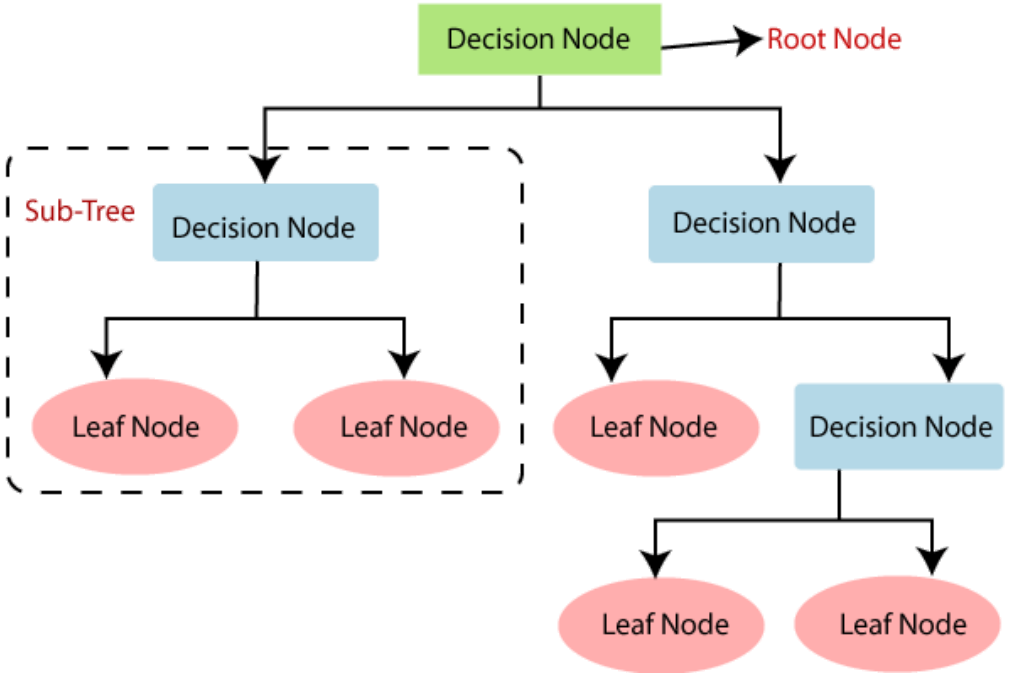
Όπου το a είναι μια παράμετρος συστηματοποίησης (regularization parameter) όπου ονομάζεται παράμετρος επιθετικότητας. Αυτή ελέγχει το μέγεθος βήματος της ενημέρωσης και καθορίζει πόσο επιθετικά προσαρμόζεται το μοντέλο στα νέα δεδομένα. Στην εικόνα 4.3 [24] βλέπουμε μια οπτικοποίηση του μοντέλου στα στάδια όπου το μοντέλο είναι παθητικό και επιθετικό.



Εικόνα 4.3: Καταστάσεις παθητικότητας όπου δεν γίνεται αλλαγή εάν η πρόβλεψη είναι σωστή και κατάσταση επιθετικότητας όταν η πρόβλεψη είναι λάθος το μοντέλο κάνει αλλαγές

- **Decision Tree Classifier:** είναι ένας δημοφιλής αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται για προβλήματα ταξινόμησης. Ο αλγόριθμος αυτός βασίζεται σε διακλαδώσεις αποφάσεων (decision trees) και επιτρέπει την κατηγοριοποίηση εισόδου σε διάφορες κλάσεις ή κατηγορίες με βάση τις χαρακτηριστικές παραμέτρους της εισόδου. Οι αποφάσεις αναπαρίστανται σε μορφή δέντρων, όπου κάθε κόμβος αποτελεί ένα χαρακτηριστικό (feature) της εισόδου. Σε κάθε επίπεδο του δέντρου επιλέγονται τα χαρακτηριστικά τα οποία θα χρησιμοποιηθούν για τον διαχωρισμό των δεδομένων. Η διακλάδωση συνεχίζεται σε υπό-δέντρα μέχρι να φτάσουν στα φύλλα, δηλαδή το τέλος του δέντρου. Η εικόνα 4.4[4] απεικονίζει ένα δένδρο αποφάσεων. Παρόλη την ευκολία στην χρήση του αλγορίθμου σε διάφορους τύπους δεδομένων εισόδου, ο αλγόριθμος Decision Tree Classifier έχει την τάση υπερπροσαρμογής (overfitting), και δεν ήταν ιδανικός για το πρόβλημα στο οποίο ερευνάται σε αυτήν την εργασία.

[4] https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm



Εικόνα 4.4: Μια απλή οπτικοποίηση ενός δέντρου αποφάσεων

- **Multinomial Naive Bayes:** Ο Multinomial Naive Bayes (που βασίζεται στον αλγόριθμο Naive Bayes), είναι ιδιαίτερα δημοφιλής στην επεξεργασία φυσικής γλώσσας (NLP). Ονομάζεται Multinomial καθώς χρησιμοποιείται σε προβλήματα με πολλαπλές κατηγορίες και κατανέμει τα χαρακτηριστικά της εισόδου πολυωνυμικά. Ο αλγόριθμος χρησιμοποιεί το Θεώρημα Bayes με 'αφελείς' υποθέσεις, δηλαδή θεωρεί πως τα χαρακτηριστικά είναι ανεξάρτητα μεταξύ τους, παρόλο που συχνά αυτό δεν ισχύει για τα δεδομένα. Το θεώρημα υπολογίζει πιθανότητες που σχετίζονται με συμβάντα. Με κάθε δείγμα εκπαίδευσης μπορεί να αυξηθεί/μειωθεί η πιθανότητα του να είναι εύστοχη μια υπόθεση, καθώς η προηγούμενη γνώση μπορεί να συνδυαστεί με τα δεδομένα που έχουν παρατηρηθεί. Η εικόνα 4.5 δείχνει ένα παράδειγμα της εφαρμογής του Multinomial Naive Bayes σε δεδομένα. Παρατηρείται πως ο MNB είναι ιδιαίτερα γρήγορος και αποτελεσματικός στην κατηγοριοποίηση κειμένων. Χρησιμοποιείται συχνά στην ανίχνευση μαζικής αποστολής ηλεκτρονικών μηνυμάτων (spam) και την αναγνώριση συναισθημάτων μέσα σε ένα κείμενο. Παρόλα αυτά, ο ιδιαίτερα 'υποθετικός' του χαρακτήρας, πολλές φορές επηρεάζει την ευστοχία προβλέψεων.

$$P(C_i|\mathbf{X}) = P(\mathbf{X}|C_i)P(C_i)$$

$$P(\mathbf{X}|C_i) = \prod_{k=1}^{n} P(x_k|C_i) = P(x_1|C_i) \times P(x_2|C_i) \times \ldots \times P(x_n|C_i)$$

Example

age	income	student	credit_rating	com
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

$P(C_i)$: P(buys_computer = "yes") = 9/14 = 0.643

P(buys_computer = "no") = 5/14= 0.357

Compute $P(X|C_i)$ for each class

P(age = "<=30" | buys_computer = "yes") = 2/9 = 0.222

P(age = "<= 30" | buys_computer = "no") = 3/5 = 0.6

P(income = "medium" | buys_computer = "yes") = 4/9 = 0.444

P(income = "medium" | buys_computer = "no") = 2/5 = 0.4

P(student = "yes" | buys_computer = "yes) = 6/9 = 0.667

P(student = "yes" | buys_computer = "no") = 1/5 = 0.2

P(credit_rating = "fair" | buys_computer = "yes") = 6/9 = 0.667

P(credit_rating = "fair" | buys_computer = "no") = 2/5 = 0.4

X = (age <= 30 , income = medium, student = yes, credit_rating = fair)

$P(X|C_i)$: P(X|buys_computer = "yes") = 0.222 x 0.444 x 0.667 x 0.667 = 0.044

P(X|buys_computer = "no") = 0.6 x 0.4 x 0.2 x 0.4 = 0.019

$P(X|C_i)*P(C_i)$: P(X|buys_computer = "yes") * P(buys_computer = "yes") = 0.028

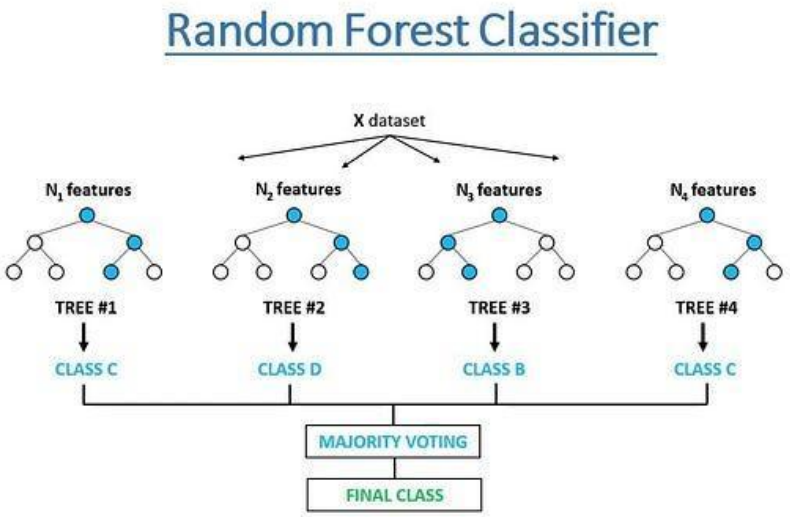
P(X|buys_computer = "no") * P(buys_computer = "no") = 0.007

Therefore, X belongs to class ("buys_computer = yes")

30

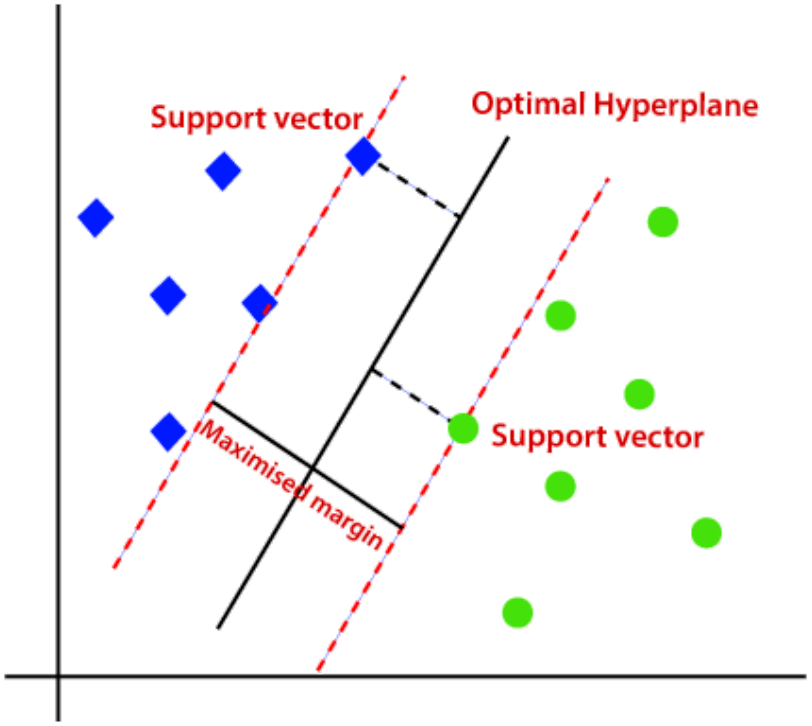
Εικόνα 4.5: Εκτέλεση του Multinomial Naive Bayes σε πίνακα με δεδομένα

- **Random Forest Classifier:** Ο Random Forest Classifier βασίζεται στην συνδυαστική χρήση πολλών απλών μοντέλων μηχανικής μάθησης (ensemble learning). Συγκεκριμένα, ο αλγόριθμος υλοποιεί πολλά Decision Trees ώστε να βρεθεί μια πιο ακριβής και δοκιμασμένη πρόβλεψη. Κάθε δέντρο παίρνει τυχαία δεδομένα από το dataset και εργάζεται πάνω σε αυτά. Κατά την ταξινόμηση, αφήνει κάθε δέντρο να ψηφίσει για την κλάση. Συγκρίνοντας τα αποτελέσματά τους και συνδυάζοντας τις αποφάσεις κάθε δέντρου, δημιουργείται ένα μοντέλο που θα έχει πιο ακριβείς προβλέψεις από ένα απλό Decision Tree. Η εικόνα 4.6 [25] απεικονίζει την λειτουργία ενός random forest classifier. Ο Random Forest Classifier είναι αρκετά δημοφιλής διότι αντιμετωπίζει το overfitting και έχει την δυνατότητα να χρησιμοποιηθεί χωρίς ιδιαίτερη προεπεξεργασία των δεδομένων. Χρησιμοποιείται ευρέως σε πολλές περιπτώσεις, συμπεριλαμβανομένων της ταξινόμησης εικόνων, ανίχνευσης απάτης, και πρόβλεψης τιμών, μεταξύ άλλων.



Εικόνα 4.6: Οπτική αναπαράσταση ενός Random Forest Classifier

- **Support Vector Machines:** Ο αλγόριθμος αυτός (SVM), ανήκει στην κατηγορία μοντέλων επιβλεπόμενης μάθησης. Είναι αποτελεσματικός όταν οι κατηγορίες που πρέπει να ταξινομηθούν είναι γραμμικά ασυμβίβαστες, δηλαδή δεν μπορούν να διαχωριστούν από ένα απλό γραμμικό όριο. Το SVC αναζητά ένα υπερεπίπεδο (Hyperplane) που διαχωρίζει τις κλάσεις στον χώρο των χαρακτηριστικών. Δημιουργεί υποστηρικτικά διανύσματα, τα οποία αποτελούν δείγματα της εισόδου, και επιλέγει αυτά που είναι σε μικρότερη απόσταση από το υπερεπίπεδο ως πιο σημαντικά. Στην εικόνα 4.7 [26] απεικονίζεται «το καλύτερο υπερεπίπεδο που έχει τη μέγιστη απόσταση και από τις δύο κατηγορίες όπου είναι και κύριος στόχος του SVM». Ο συγκεκριμένος αλγόριθμος είναι ανθεκτικός στο overfitting και αποτελεί ισχυρό μοντέλο μηχανικής μάθησης που χρησιμοποιείται σε αναγνώριση εικόνων, ανίχνευση μαζικής αποστολής ηλεκτρονικών μηνυμάτων (spam) και κατηγοριοποίηση κειμένων.



Εικόνα 4.7: «Το καλύτερο υπερεπίπεδο ενός Support Vector Machines είναι εκείνο το επίπεδο που έχει τη μέγιστη απόσταση και από τις δύο κατηγορίες. Αυτό γίνεται με την εύρεση διαφορετικών υπερεπιπέδων που ταξινομούν τις ετικέτες με τον καλύτερο τρόπο και, στη συνέχεια, θα επιλέξει αυτό που είναι πιο μακριά από τα σημεία δεδομένων ή αυτό που έχει μέγιστο περιθώριο.»

Αρχικά, αναφέρουμε σε μερικά μοντέλα χρησιμοποιήσαμε την σακούλα με λέξεις (Bag of Words) και σε άλλα την αντίστροφη συχνότητα εγγράφων (IDF) για την διανυσματοποίηση[5] (Vectorization) των λέξεων. Θα αναφερόμαστε ως Count Vectorizer την συνάρτηση διανυσματοποίησης για το bag of words. Αφού χωρίσαμε τα δεδομένα σε εκπαίδευσης και εξέτασης χρησιμοποιούμε την συνάρτηση GridSearchCV[6] και σε συνδυασμό με το Pipelining[7] συνδυάζουμε το αντίστοιχο μοντέλο και τον vectorizer ώστε να κάνουμε εξαντλητική αναζήτηση

[5] https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

[6] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

[7] https://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html

τον καλύτερο συνδυασμό υπερπαραμέτρων από ένα σύνολο που δώσαμε ώστε να επιτύχουμε μεγαλύτερη ακρίβεια.

Το Count Vectorizer είναι ένα εργαλείο επεξεργασίας της φυσικής γλώσσας, που παρέχεται από την βιβλιοθήκη scikit-learn της Python[8], το οποίο χρησιμοποιείται συχνά σε αλγορίθμους μηχανικής μάθησης οι οποίοι επεξεργάζονται κείμενα. Σκοπός του είναι η μετατροπή ενός συνόλου κειμένων σε αναπαραστάσεις διανυσμάτων με βάση την συχνότητα εμφάνισης των λέξεων στα κείμενα αυτά. Δημιουργεί έναν πίνακα στον οποίο κάθε στήλη αναπαριστά μια ξεχωριστή λέξη και κάθε γραμμή ένα δείγμα κειμένου. Η τιμή σε κάθε κελί είναι ο αριθμός εμφάνισης κάθε λέξης στο κείμενο που αντιστοιχεί. Για παράδειγμα, οι προτάσεις *«Το μήλο είναι φρούτο»* και *«Το μήλο και η φράουλα είναι νόστιμα»* αντιστοιχούν στον πίνακα 4.1. Η αναπαράσταση αυτή μπορεί να χρησιμοποιηθεί ως είσοδος αλγορίθμων μηχανικής μάθησης σε προβλέψεις κειμένων, κατηγοριοποίηση και άλλες επεξεργασίες φυσικής γλώσσας (NLP).

	Το	μηλο	είναι	φρούτο	η	φραουλα	και	νόστιμα
Κείμενο[1]	1	1	1	1	0	0	0	0
Κείμενο[2]	1	1	1	0	1	1	1	1

Πίνακας 4.1: Εφαρμογή του Bag of Words σε δύο κείμενα

Τόσο για τα μοντέλα που χρησιμοποιούν τον Count Vectorizer όσο και αυτά που χρησιμοποιούν το IDF χρησιμοποιούνται οι υπερπαράμετροι stop_words, max_df, min_df και n-gram range. Τα stopwords είναι οι λέξεις σε κάθε γλώσσα που δεν προσθέτουν ουσιαστικό νόημα σε μια πρόταση και μπορούν να αγνοηθούν. Αυτές οι λέξεις αντιμετωπίζονται με ενσωματωμένη λίστα Stopwords της βιβλιοθήκης sklearn (για την αγγλική γλώσσα: *CountVectorizer(stop_words ='english')*. Το max_df είναι η μέγιστη συχνότητα στο έγγραφο (Max document frequency). Αγνοούνται οι λέξεις που εμφανίζονται παραπάνω φορές από ότι αναγράφει ο αριθμός της παραμέτρου. Ενώ το min_df είναι η ελάχιστη συχνότητα στο έγγραφο (Minimum document frequency). Δρα αντίστοιχα με το max_df, και μαζί αγνοούν λέξεις που δεν επηρεάζουν το νόημα του κειμένου. Το n-gram range είναι το εύρος μίας σειράς από συνεχόμενες λέξεις στο εκάστοτε κείμενο. Στα υλοποιημένα μοντέλα μηχανικής μάθησης, επεξεργαζόμαστε τα κείμενα με εύρη n-gram (1,1), (1,2), (1,3). Παρακάτω, παρουσιάζεται παράδειγμα για την κατανόηση του εύρους n-gram.

Κείμενο: *«Το μήλο και η φράουλα είναι νόστιμα»*

n-gram range = (1,3)

Αφού υλοποιηθεί το n-gram εύρος, το αποτέλεσμα θα είναι:

[‘Το’, ‘μήλο’, ‘και’, ‘η’, ‘φράουλα’, ‘είναι’, ‘νόστιμα’, ‘Το μήλο’, ‘μήλο και’, ‘και η’, ‘η φράουλα’, ‘φράουλα είναι’, ‘είναι νόστιμα’, ‘το μήλο και’, ‘μήλο και η’, ‘και η φράουλα’, ‘η φράουλα είναι’, ‘φράουλα είναι νόστιμα’]

Το IDF είναι ένα αριθμητικό στατιστικό στοιχείο που αντικατοπτρίζει τη σημασία μιας λέξης σε ένα έγγραφο σε σχέση με μια συλλογή εγγράφων, συνήθως ένα σώμα (corpus). Μπορούμε να πούμε ότι ποσοτικοποιεί την σημασία ενός όρου σε πολλά έγγραφα. Χρησιμοποιείται συνήθως στην επεξεργασία φυσικής γλώσσας και στην ανάκτηση πληροφοριών για διάφορες εργασίες που σχετίζονται με κείμενο, συμπεριλαμβανομένης της ταξινόμησης κειμένου. Το IDF μετρά τη σπανιότητα ή τη σημασία ενός όρου σε μια συλλογή εγγράφων και υπολογίζεται ως ο λογάριθμος του συνολικού αριθμού εγγράφων στο σώμα διαιρεμένος με τον αριθμό των εγγράφων που

8 https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html

περιέχουν τον όρο, συν ένα για να αποφευχθεί η διαίρεση με το μηδέν[9]. Η εικόνα 4.8 δείχνει τον τύπο που μόλις περιγράψαμε.

$$IDF(t,D) = \frac{\text{Συνολικός αριθμός εγγράφων στο corpus D}}{\text{(Αριθμός εγγράφων που περιέχουν τον όρο t στο corpus D) + 1}}$$

Εικόνα 4.8: Ο τύπος της αντίστροφη συχνότητα εγγράφων (IDF)

Επομένως, για το IDF που χρησιμοποιούμε δηλώνουμε το εύρος n-gram [(1,1), (1,2), (1,3)] για όλα τα σύνολα δεδομένων. Χρησιμοποιούμε εύρος (0.0, 0.25, 0.5, 1.0) για το min_df και max_df γιατί θέλουμε με το min_df να δούμε άμα αποδίδει καλύτερα το μοντέλο μας με το να αγνοεί λέξεις που εμφανίζονται κάτω από το ορισμένο ποσοστό σε κάθε γραμμή του Dataset ή συγκεκριμένα σε κάθε κείμενο, και με το max_df να αγνοήσουμε λέξεις που ίσως εμφανίζονται περισσότερο από το ορισμένο ποσοστό σε κάθε κείμενο. Επίσης, χρησιμοποιούμε και την παράμετρο C και για τα δύο μοντέλα με εύρος (10^{-5}, 10^{-3}, 10^{-1}, 10^{0}, 10^{1}, 10^{3}, 10^{5}) ώστε να βρει την καλύτερη αντίστροφη της ισχύος κανονικοποίησης και να έχουμε αποφυγή της υπερπροσαρμογής (Overfitting) των μοντέλων.

Για τα μοντέλα που χρησιμοποιούν το bag of words (count vectorizer) δηλώνουμε εύρος max_features (1000, 5000, 10000, None). Αυτή η παράμετρος επιτρέπει να περιορίσουμε το μέγεθος του λεξιλογίου στις πιο συχνές λέξεις οι οποίες μπορεί να είναι χρήσιμες για τον έλεγχο της διάστασης των διανυσμάτων χαρακτηριστικών (feature vectors) και κατ' επέκταση τη διαχείριση της χρήσης μνήμης. Για το εύρος n-gram επιλέγεται η λίστα [(1, 1), (1, 2), (1, 3), (2, 3), (3, 3)], ενώ το max_df και min_df παραμένει ως εύρος (0.0, 0.25, 0.5, 1.0). Επιλέγουμε για την μεταβλητή stop_words το εύρος ['english', None]. Για τα τέσσερα μοντέλα παρουσιάζουμε τον παρακάτω πίνακα 4.2 που συμπεριλαμβάνει τα διάφορα εύρη για τις διαφορετικές τιμές του κάθε μοντέλου.

Models / Variables	Decision Tree Classifier
criterion	['gini', 'entropy']
min_samples_leaf	[1, 2, 3, 4, 5, 6, 7, 8]
max_features	[1, 2, 3, 4, 5, 6, 7, 8]

Models / Variables	Multinomial Naive Bayes
alpha	[0.5, 0.6, 0.7, 0.8, 0.9, 1]
fit_prior	[True, False]

Models / Variables	Random Forest Classifier
criterion	['gini', 'entropy']
n_estimators	[50,100,200,300]
max_depth	[4,6,8,10]

Models / Variables	Support Vector Machines
C	[2,4,6,8,10]

Πίνακας 4.2: Τα εύρη τιμών για τα μοντέλα που χρησιμοποίησαν Decision Tree Classifier, Multinomial Naïve Bayes, Random Forest Classifier και Support Vector Machines

Καταλήγουμε λοιπόν σε έξι machine learning μοντέλα δύο εκ των οποίων (Logistic Regression και Passive Aggressive Classifier) χρησιμοποιούν το IDF ως τεχνική για την NLP και την αναπαράσταση και την κωδικοποίηση δεδομένων κειμένου, και τα υπόλοιπα τέσσερα μοντέλα που χρησιμοποιούν το BoW (Bag of Words). Στον πίνακα 4.3 παρουσιάζονται οι τελικές ακρίβειες για όλα τα μοντέλα. Οι δύο πρώτοι πίνακες αφορούν τα μοντέλα που χρησιμοποίησαν το IDF ως τεχνική και οι υπόλοιποι τέσσερις αφορούν τα μοντέλα που χρησιμοποίησαν το Bag of Words ως τεχνική.

[9] https://en.wikipedia.org/wiki/Tf%E2%80%93idf#Inverse_document_frequency

Model	Datasets Accuracy				Datasets f1-Score			Datasets Precision			Datasets Recall		
Logistic Regression	Fake News Cropus	WELFake	LIAR	Overall Accuracy	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR
Train with Fake News Corpus	94%	67%	47%	**69.33%**	94%	64%	47%	94%	74%	55%	94%	67%	47%
Train with WELFake	68%	95%	60%	**74.33%**	67%	95%	55%	71%	95%	55%	68%	95%	60%
Train with LIAR	52%	56%	76%	**61.33%**	47%	51%	73%	58%	58%	77%	52%	56%	76%
Model	**Datasets Accuracy**				**Datasets f1-Score**			**Datasets Precision**			**Datasets Recall**		
Passive Aggressive Classifier	Fake News Cropus	WELFake	LIAR	Overall Accuracy	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR
Train with Fake News Corpus	94%	72%	50%	**72.00%**	94%	71%	51%	94%	77%	55%	94%	72%	50%
Train with WELFake	69%	95%	61%	**75.00%**	69%	95%	56%	73%	95%	56%	69%	95%	61%
Train with LIAR	51%	55%	91%	**65.67%**	48%	52%	90%	55%	56%	90%	51%	55%	91%
Model	**Datasets Accuracy**				**Datasets f1-Score**			**Datasets Precision**			**Datasets Recall**		
Decision Tree Classifier	Fake News Cropus	WELFake	LIAR	Overall Accuracy	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR
Train with Fake News Corpus	93%	65%	49%	**69.00%**	93%	65%	53%	93%	68%	67%	93%	65%	49%
Train with WELFake	64%	90%	48%	**67.33%**	63%	90%	54%	69%	90%	73%	64%	90%	48%
Train with LIAR	53%	44%	61%	**52.67%**	53%	45%	61%	53%	47%	61%	53%	44%	61%
Model	**Datasets Accuracy**				**Datasets f1-Score**			**Datasets Precision**			**Datasets Recall**		
Multinomial Naive Bayes	Fake News Cropus	WELFake	LIAR	Overall Accuracy	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR
Train with Fake News Corpus	86%	68%	57%	**70.33%**	86%	70%	60%	88%	82%	66%	86%	68%	57%
Train with WELFake	66%	88%	56%	**70.00%**	64%	88%	56%	67%	88%	57%	66%	88%	56%
Train with LIAR	68%	54%	65%	**62.33%**	73%	59%	61%	82%	75%	67%	68%	54%	65%
Model	**Datasets Accuracy**				**Datasets f1-Score**			**Datasets Precision**			**Datasets Recall**		
Random Forest Classifier	Fake News Cropus	WELFake	LIAR	Overall Accuracy	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR
Train with Fake News Corpus	87%	57%	56%	**66.67%**	87%	65%	70%	89%	90%	95%	87%	57%	56%
Train with WELFake	49%	85%	44%	**59.33%**	50%	85%	61%	76%	85%	100%	49%	85%	44%
Train with LIAR	69%	49%	59%	**59.00%**	81%	66%	46%	100%	100%	69%	69%	49%	59%
Model	**Datasets Accuracy**				**Datasets f1-Score**			**Datasets Precision**			**Datasets Recall**		
Support Vector Machines	Fake News Cropus	WELFake	LIAR	Overall Accuracy	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR
Train with Fake News Corpus	96%	78%	52%	**75.33%**	96%	78%	52%	96%	79%	54%	96%	78%	52%
Train with WELFake	70%	96%	47%	**71.00%**	69%	96%	53%	76%	96%	74%	70%	96%	47%
Train with LIAR	67%	53%	67%	**62.33%**	71%	56%	66%	77%	66%	66%	67%	53%	67%

Πίνακας 4.3: Αναλυτικές ακρίβειες για κάθε μοντέλο μηχανικής μάθησης που υλοποιήθηκε, εκπαιδεύτηκε και χρησιμοποιήθηκε για τις προβλέψεις όλων των συνόλων δεδομένων

Μετά από τα παραπάνω αποτελέσματα μπορούμε να πούμε ότι τα καλύτερα μέσα αποτελέσματα τα είχε το μοντέλο που χρησιμοποίησε τον Passive Aggressive Classifier με μέση ακρίβεια 70.89%. Από την άλλη το μοντέλο με το Random Forest έδωσε τα χειρότερα αποτελέσματα με μέση ακρίβεια 61.67% και το δεύτερο χειρότερο ήταν το Decision Tree με 63%. Παρόλα αυτά και τα έξι μοντέλα δεν είχαν κακές αποδώσεις για κάθε σύνολο δεδομένων ενώ κάνοντας εκπαίδευση οποιοδήποτε μοντέλο χρησιμοποιώντας το Fake News Corpus βλέπουμε πως συχνά δίνει τις μεγαλύτερες ακρίβειες σε σχέση με το εάν κάναμε εκπαίδευση με το WELFake ή το LIAR σύνολο δεδομένων. Η ικανότητα των μοντέλων να γενικεύουν από το ένα σύνολο δεδομένων στο άλλο είναι εξίσου σημαντική διότι παρατηρούμε πως τα σύνολα δεδομένων δίνουν την μεγαλύτερη απόδοση όταν προβλέπουν στα δικά τους αλλά μεγαλύτερα σύνολα, ενώ η απόδοση τους μειώνεται όταν προκύπτει η πρόβλεψη στα υπόλοιπα δύο σύνολα. Συγκεκριμένα, στην εκπαίδευση με το Fake News Corpus και το WELFake σε όλα τα μοντέλα η ακρίβεια όταν γίνεται πρόβλεψη, παραλείποντας το σύνολο δεδομένων του εαυτού του, μειώνεται και έπειτα έχει την μικρότερη τιμή στο LIAR σύνολο δεδομένων. Βγάζουμε το συμπέρασμα ότι τα μοντέλα δυσκολεύονται στην προσαρμογή τομέα του LIAR συνόλου όπου είναι κάτι αναμενόμενο διότι είναι ένα σύνολο δεδομένων που έχει λίγα και μικρά δεδομένα κειμένου. Βέβαια, δεν μπορούμε να ισχυριστούμε πως κάποιο dataset είναι κατώτερο από κάποιο άλλο, αντίθετα ανάλογα με το μοντέλο που χρησιμοποιείται συγκεκριμένα χαρακτηριστικά των συνόλων δεδομένων όπως η πολυπλοκότητα κειμένου μπορεί να επηρεάσει την απόδοση. Επιπλέον, και στα ίδια τα μοντέλα μηχανικής μάθησης υπάρχουν δυνατά και αδύνατα σημεία όπως για παράδειγμα ότι τα Decision Trees μπορεί να είναι ερμηνεύσιμα αλλά επιρρεπή σε υπερπροσαρμογή, ενώ ο Passive Aggressive Classifier μπορεί να είναι αποδοτικός και να χειριστεί ροές δεδομένων κειμένου αλλά μπορεί να δυσκολευτεί να καταγράψει πολύπλοκες σχέσεις σε δεδομένα κειμένου ενώ ο συντονισμός υπερπαραμέτρων μπορεί να απαιτεί παραπάνω πειραματισμό.

(Υποκεφάλαιο 4.2) **FastText μοντέλο**

Η FastText βιβλιοθήκη είναι ένα ευέλικτο εργαλείο για την ταξινόμηση κειμένων και η ταχύτητα και η ικανότητά του να χειρίζεται πληροφορίες λέξεων την καθιστούν πολύτιμη για διάφορες εφαρμογές στην NLP. Στην δική μας περίπτωση χρησιμοποιούμε την FastText για να δημιουργήσουμε ενσωματώσεις λέξεων (word embeddings) από το corpus του κάθε Dataset,

όπου θα χρησιμοποιηθούν αργότερα σε επόμενο κεφάλαιο που μελετάμε τα νευρωνικά δίκτυα ως μοντέλα. Επιπλέον, την χρησιμοποιούμε για την ταξινόμηση κειμένων όπου φέρνοντας στην κατάλληλη μορφή τις ετικέτες του Dataset, χρησιμοποιεί ρηχό νευρωνικό δίκτυο (shallow neural network) κατά την εκπαίδευση του μοντέλου. Για την χρήση του μοντέλου χωρίς επίβλεψη (unsupervised), με σκοπό την δημιουργία των ενσωματώσεων λέξεων οι τιμές των υπερπαραμέτρων παρουσιάζονται στον πίνακα 4.4. Η διαδικασία που χρησιμοποιούμε για τη δημιουργία ενσωματώσεων λέξεων απεικονίζεται στη διαδικασία 1.

Variables \ Models	FastText Unsupervised Fake News Corpus
dim	300
lr	0.1
epoch	7

Variables \ Models	FastText Unsupervised WELFake / LIAR
dim	100
lr	0.1
epoch	7

Πίνακας 4.4: Οι τιμές των υπερπαραμέτρων για την δημιουργία ενσωματώσεων λέξεων για το Fake News Corpus και τα WELFake και LIAR σύνολα δεδομένων

Διαδικασία 1: Διαδικασία δημιουργίας των ενσωματώσεων λέξεων

```
Φόρτωση στην μνήμη του συγκεκριμένου συνόλου δεδομένων
Διαχωρισμός του συνόλου δεδομένων σε μικρότερες παρτίδες
Για κάθε παρτίδα
    Μετατροπή του κειμένου σε μεμονωμένες λέξεις και αποθήκευση σε συγκεκριμένη λίστα
    Αποθήκευση της λίστας σε αντίστοιχο αρχείο κειμένου με αριθμό τον συγκεκριμένο αριθμό παρτίδας
Τέλος επανάληψης
Για κάθε αρχείο κειμένου
    Φόρτωση του αντίστοιχου αρχείου κειμένου
    Εκπαίδευση χωρίς επίβλεψη με συγκεκριμένες μεταβλητές του μοντέλου FastText με το αντίστοιχο αρχείο κειμένου
Τέλος επανάληψης
Αποθήκευση του μοντέλου που μόλις εκπαιδεύτηκε για μελλοντική χρήση
```

Όσων αφορά την ταξινόμηση κειμένου χρησιμοποιώντας την βιβλιοθήκη FastText, επιλέγουμε να κάνουμε πιο απαιτητική σε υπολογιστική δύναμη τη διαδικασία της εκπαίδευσης του μοντέλου μόνο για όλα τα σύνολα δεδομένων, με απώτερο σκοπό την αύξηση της ακρίβειας. Ορίζουμε λοιπόν τις τιμές δύο μεταβλητών "dim" (Dimensionality of Word Vectors) ίσες με τριακόσια (300) και "wordNgrams" (μέγιστο μήκος λέξης n-grams) ίσον με τρία (3). Με την αύξηση της διάστασης των διανυσμάτων λέξεων μπορούμε να καταγράψουμε πιο σύνθετες σχέσεις μεταξύ των λέξεων και ως αποτέλεσμα αυτό να οδηγήσει σε καλύτερη απόδοση του μοντέλου. Με

την επιλογή του μέγιστου μήκους n-grams μιας λέξης επηρεάζεται το επίπεδο γλωσσικών πληροφοριών που το μοντέλο μπορεί να εκλάβει. Ρυθμίζοντας σε υψηλότερη τιμή, δηλαδή τρία (3), επιτρέπουμε στο μοντέλο να λάβει υπόψη όχι μόνο μεμονωμένες λέξεις (μονόγραμμα ή unigrams), αλλά και ζεύγη διαδοχικών λέξεων (διγράμματα ή bigrams) και τριπλέτες διαδοχικών λέξεων (τριγράμματα ή trigrams). Αυτό χρησιμεύει στο να μπορεί το μοντέλο να εκλάβει ορισμένους τύπους πληροφοριών με βάση τα συμφραζόμενα. Ωστόσο, γνωρίζουμε ότι με αυτόν τον τρόπο αυξάνεται επίσης ο χώρος χαρακτηριστικών, ο οποίος μπορεί να απαιτεί περισσότερα δεδομένα εκπαίδευσης για να γενικευθεί καλά. Συνοπτικά παρουσιάζονται οι τιμές των υπερπαραμέτρων στον πίνακα 4.5. Έτσι λοιπόν, λαμβάνοντας υπόψη τα παραπάνω εκπαιδεύουμε το μοντέλο και με τα τρία Datasets και κάνοντας predict σε κάθε ένα από αυτά στον πίνακα 4.6 παρουσιάζονται τα τελικά αποτελέσματα.

Models / Variables	FastText Supervised Text Classification
dim	300
lr	0.3
epoch	100
wordNgrams	3

Πίνακας 4.5: Οι τιμές των υπερπαραμέτρων για την εκπαίδευση του FastText μοντέλου με σκοπό την ταξινόμηση κειμένου για όλα τα σύνολα δεδομένων

Model	Datasets Accuracy				Datasets f1-Score			Datasets Precision			Datasets Recall		
Fasttext	Fake News Cropus	WELFake	LIAR	Overall Accuracy	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR
Train with Fake News Corpus	94%	63%	50%	**69.00%**	94%	59%	51%	94%	73%	54%	94%	63%	50%
Train with WELFake	67%	95%	59%	**73.67%**	67%	95%	55%	70%	95%	55%	67%	95%	59%
Train with LIAR	49%	52%	90%	**63.67%**	44%	45%	90%	55%	53%	90%	49%	52%	90%

Πίνακας 4.6: Αναλυτικές ακρίβειες για το μοντέλο FastText που υλοποιήθηκε, προπονήθηκε και χρησιμοποιήθηκε για τις προβλέψεις όλων των συνόλων δεδομένων

Παρατηρώντας τα αποτελέσματα για το FastText μοντέλο βλέπουμε πως η μέση απόδοση είναι 68.78%. Αυτή την φορά η εκπαίδευση με το WELFake σύνολο δεδομένων μας δίνει την μεγαλύτερη ακρίβεια 73.67% ενώ ακόμη τα πηγαίνει καλύτερα στη γενίκευση, κρίνοντας από τις ακρίβειες, στα υπόλοιπα σύνολα δεδομένων. Τα αποτελέσματα, ειδικά για το LIAR σύνολο δεδομένων, είναι ικανοποιητικά δεδομένου του μεγέθους του και των λίγων πληροφοριών που έχει στα κείμενα του. Φυσικά, παρατηρούμε ότι οι προβλέψεις στα ίδια τα σύνολα δεδομένων που εκπαιδεύτηκαν δίνουν πάρα πολύ καλές ακρίβειες, Fake News Corpus με 94%, WELFake με 95% και LIAR με 90%. Αυτό σημαίνει ότι ενώ το μέγεθος των συνόλων δεδομένων ήταν περιορισμένο το μοντέλο FastText κατάφερε να εξάγει χαρακτηριστικά και να μάθει μοτίβα και ως αποτέλεσμα να γενικεύεται πολύ αποτελεσματικά στο συγκεκριμένο σύνολο δεδομένων.

(Υποκεφάλαιο 4.3) **BERT μοντέλο**

Προτού αναλύσουμε τα μοντέλα που χρησιμοποιούν νευρωνικά δίκτυα, θα αναφερθούμε στο BERT μοντέλο το οποίο ανήκει σε αυτή την κατηγορία και ειδικότερα στην κατηγορία των Transformers. Όντας «μια αρχιτεκτονική μοντέλου που αποφεύγει την επανάληψη και αντ' αυτού βασίζεται εξ ολοκλήρου σε έναν μηχανισμό προσοχής για να αντλήσει παγκόσμιες εξαρτήσεις μεταξύ εισόδου και εξόδου» [27], έχει αποδειχθεί κατάλληλη για την αναγνώριση αληθών ή ψευδών ειδήσεων [28]. Χρησιμοποιούμε λοιπόν, τα προ-εκπαιδευμένα βάρη του μοντέλου BERT ως σημείο εκκίνησης (ώστε να εξοικονομήσουμε χρόνο και υπολογιστικούς πόρους σε σύγκριση αν επιλέγαμε

να εκπαιδεύσουμε από την αρχή), και τα προσαρμόζουμε με ακρίβεια στα δεδομένα των Datasets. Πριν περιγράψουμε τη διαδικασία της εκπαίδευσης του μοντέλου θα παρουσιάσουμε τον πίνακα 4.7 που αναφέρει συνοπτικά επιπλέον στρώματα καθώς και τις τιμές των υπερπαραμέτρων τους.

Models / Variables	pre-trained BERT for Text Classification
Dropout Layer 1	0.6
Dropout Layer 2	0.55
Fully Connected Dense Layers	2
Dense Layers Units	64, 32
Dense Layers Actication Function	ReLU
Optimizer	Adam
Learning Rate	1.00E-05

Πίνακας 4.7: Επιπλέον στρώματα και οι τιμές των υπερπαραμέτρων για την εκπαίδευση του BERT μοντέλου με σκοπό την ταξινόμηση κειμένου για όλα τα σύνολα δεδομένων

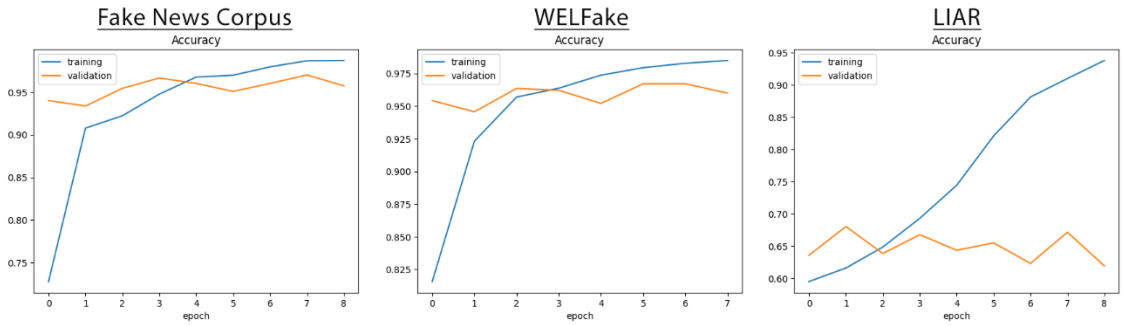
Η διαδικασία που ακολουθούμε για την εκπαίδευση του μοντέλου αναλυτικά παρέχεται στο διάγραμμα διαδικασία 2.

ΔΙΑΔΙΚΑΣΙΑ 2: ΔΙΑΔΙΚΑΣΙΑ ΓΙΑ ΤΗΝ ΕΚΠΑΊΔΕΥΣΗ ΤΟΥ ΜΟΝΤΕΛΟΥ BERT

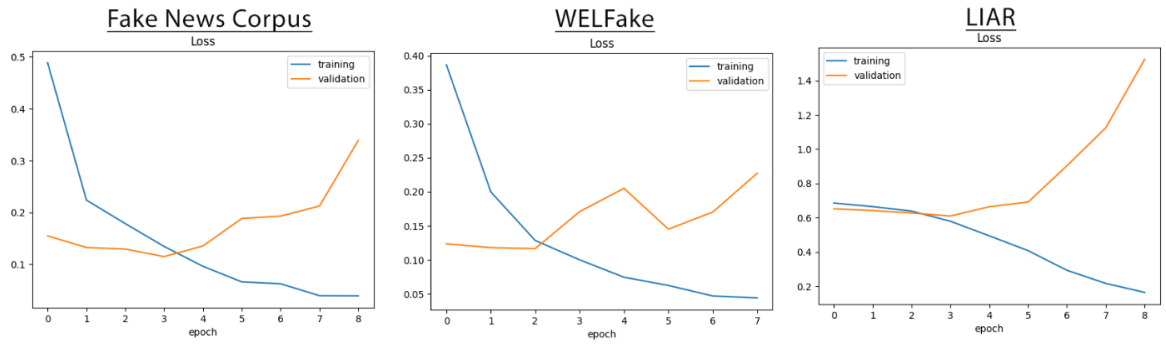
1	Φόρτωση στην μνήμη το Dataset
2	Χωρισμός Dataset σε 10.000 samples. 55% Real – 45% Fake ταμπέλες
3	Διαχωρισμός σε σετ εκπαίδευσης και δοκιμών
4	Εισαγωγή του προεκπαιδευμένο BERT: bert_en_uncased_L-12_H-768_A-12
5	Φόρτωση του λεξιλογίου και του tokenizer
6	Είσοδος των κειμένων για το tokenization
7	Για κάθε κείμενο:
8	Προσαρμογή κειμένου χρησιμοποιώντας τον tokenizer και περιορισμός της tokenized ακολουθίας σε max_len-2 tokens, κρατώντας χώρο για τα ειδικά tokens CLS και SEP
9	Μετατροπή τα tokens στα αντίστοιχα αναγνωριστικά τους
10	Δημιουργία μιας δυαδική μάσκα για να υποδείξει ποια tokens αποτελούν μέρος της εισόδου και ποια συμπληρώνονται
11	Αρχικοποίηση των αναγνωριστικών τμημάτων
12	Επιστροφή τριών πινάκων που περιέχουν αναγνωριστικά tokens, τις τιμές της μάσκας και τα αναγνωριστικά τμημάτων για όλα τα κείμενα εισόδου
13	Τέλος επανάληψης
14	Εισαγωγή τριών εισόδων (αναγνωριστικά tokens, τιμές μάσκας και αναγνωριστικά τμημάτων)

```
15      Επεξεργασία των εισόδων και επιστροφή δύο εξόδων sequence_output και pooled_output
16
17      Εξαγωγή του CLS token από το sequence_output
18
19      Προσθήκη πυκνών (dense) στρωμάτων και στρωμάτων εγκατάλειψης (dropout)
20
21      Μεταγλώττιση μοντέλου με βελτιστοποιητή Adam
22
23      Χρήση επανάκλησης όπως checkpointing και early stopping παρακολουθώντας την απώλεια
        επικύρωσης (validation loss)
24
25      Εκπαίδευση μοντέλου στα δεδομένα εκπαίδευσης
26
27      Αποθήκευση μοντέλου
28
29      Χρήση συναρτήσεων για την παρακολούθηση της ακρίβειας και του απώλειας (loss) του
        μοντέλου για κάθε epoch
30
31      Φόρτωση στην μνήμη των συνόλων δεδομένων και πρόβλεψη των ταμπελών.
32
```

Οι εικόνες 4.9 και 4.10 δείχνουν την ακρίβεια/απώλεια και ακρίβεια/απώλεια επικύρωσης αντίστοιχα κατά την διάρκεια της εκπαίδευσης ανά epoch. Στον πίνακα 4.8 βλέπουμε την ακρίβεια που είχε τελικά το Bert μοντέλο στις προβλέψεις για κάθε Dataset.



Εικόνα 4.9: Ακρίβεια και ακρίβεια επικύρωσης κατά την διάρκεια της εκπαίδευσης του μοντέλου BERT για όλα τα epoch και με τα τρία σύνολα δεδομένων



Εικόνα 4.10: Απώλεια και απώλεια επικύρωσης κατά την διάρκεια της εκπαίδευσης του μοντέλου BERT για όλα τα epoch και με τα τρία σύνολα δεδομένων

Model	Datasets Accuracy				Datasets f1-Score			Datasets Precision			Datasets Recall		
BERT	Fake News Cropus	WELFake	LIAR	Overall Accuracy	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR
Train with Fake News Corpus	96%	75%	64%	**78.33%**	96%	75%	52%	96%	76%	61%	96%	75%	64%
Train with WELFake	52%	97%	48%	**65.67%**	46%	97%	46%	63%	97%	58%	52%	97%	48%
Train with LIAR	46%	52%	75%	**57.67%**	35%	41%	75%	50%	57%	76%	46%	52%	75%

Πίνακας 4.8: Αναλυτικές ακρίβειες για το μοντέλο BERT που υλοποιήθηκε, προπονήθηκε και χρησιμοποιήθηκε για τις προβλέψεις όλων των συνόλων δεδομένων

Παρακολουθώντας τα αποτελέσματα η μέση ακρίβεια που προκύπτει και από τις τρεις ακρίβειες είναι 67.22%. Όπως έχουμε αναφέρει και στα προηγούμενα αποτελέσματα μοντέλων, το LIAR σύνολο δεδομένων, λόγω της φύσης του, δυσκολεύεται να γενικεύσει και να προσαρμοστεί στους τομείς των υπόλοιπων συνόλων δίνοντας την μικρότερη ακρίβεια μεταξύ των υπόλοιπων. Για άλλη μία φορά κυριαρχεί στην ακρίβεια η εκπαίδευση με το Fake News Corpus καθώς γενικεύεται καλύτερα στα υπόλοιπα σύνολα δίνοντας την μεγαλύτερη ακρίβεια ίση με 78.33%. Το WELFake σύνολο δεδομένων όπως και παρατηρήθηκε και τις προηγούμενες φορές δίνει ενδιάμεσα αποτελέσματα από το Fake News Corpus και το LIAR. Η μεταφορά μάθησης του προεκπαιδευμένου BERT που αφορά μια γενική εργασία κατανόησης γλώσσας στην συγκεκριμένη εργασία, ταξινόμηση ψευδών ειδήσεων, επηρέασε πιθανώς την απόδοση των μοντέλων ανάλογα με το σύνολο δεδομένων που χρησιμοποιήθηκε. Για παράδειγμα το Fake News Corpus λόγω της περιπλοκότητας του κειμένου του κατά τη διαδικασία της εκπαίδευσης επιτρέπει στο BERT να μάθει πλούσιες και συναφείς ενσωματώσεις λέξεων και κατά συνέπεια, αποκτά μια βαθιά κατανόηση της γλωσσικής δομής και της σημασιολογίας με αποτέλεσμα μιας μεγαλύτερης ακρίβειας. Από την άλλη η γνώση που αποκτήθηκε κατά τη διάρκεια της προεκπαίδευσης του μοντέλου BERT μεταφέρεται στην εργασία στόχο. Αυτή η προσαρμογή απαιτεί συχνά λιγότερα δεδομένα εκπαίδευσης και μικρότερους χρόνους εκπαίδευσης από την εκπαίδευση ενός μοντέλου από το μηδέν, κάνοντας το ιδανικό για μικρά σύνολα δεδομένων όπως το LIAR που παρατηρήσαμε ότι τα αποτελέσματα του ήταν ικανοποιητικά.

ΚΕΦΑΛΑΙΟ 5 Μοντέλα με νευρωνικά δίκτυα

Τα νευρωνικά δίκτυα ανήκουν στην κατηγορία μοντέλων μηχανικής μάθησης και είναι εμπνευσμένα από τη δομή του ανθρώπινου εγκεφάλου. Αποτελούνται από τεχνητούς νευρώνες που είναι συνδεδεμένοι μεταξύ τους και δίνεται η δυνατότητα να μπορούν να μάθουν απλά αλλά και περίπλοκα μοτίβα και αναπαραστάσεις από δεδομένα. Έχουν κερδίσει μια σημαντική θέση στην NLP λόγω της ικανότητάς τους να μαθαίνουν αυτόματα πολύπλοκα μοτίβα και αναπαραστάσεις από δεδομένα κειμένου που τα καθιστά κατάλληλα για εργασίες όπως η ταξινόμηση κειμένου. Στα περίπλοκα κείμενα μπορεί να συναντήσουμε σύνθετα μοτίβα και λεπτές γλωσσικές ενδείξεις όπου τα νευρωνικά δίκτυα μπορούν αυτόματα να τα μάθουν. Ακόμη, τα συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks, CNN) και τα επαναλαμβανόμενα νευρωνικά δίκτυα (Recurrent Neural Networks, RNN), που θα αναλύσουμε παρακάτω, είναι αποτελεσματικά στην εξαγωγή σχετικών χαρακτηριστικών από το κείμενο. Για οποιονδήποτε όγκο συνόλων δεδομένων έχουν την δυνατότητα να κλιμακωθούν για να τα διαχειριστούν ενώ ακόμη προσαρμόζονται σε διαφορετικές γλώσσες και στυλ γραφής καθιστώντας τα ευέλικτα για ένα ευρύ φάσμα πηγών ειδήσεων. Τελικά, μπορούν να γενικεύουν από τα μοτίβα που μαθαίνουν, καθιστώντας τα ικανά να κάνουν προβλέψεις σε άρθρα ειδήσεων που δεν είχαν δει στο παρελθόν. Για τη συγκεκριμένη εργασία έχουμε επιλέξει να δημιουργήσουμε τα μοντέλα χρησιμοποιώντας την κλάση tf.keras.Model[10] έναντι της tf.keras.Sequential για να ομαδοποιήσουμε τα στρώματα σε ένα αντικείμενο. Η κλάση αυτή αποτελεί μέρος του Keras API του TensorFlow, το οποίο είναι ένα API βαθιάς εκμάθησης υψηλού επιπέδου που παρέχεται από το TensorFlow[11]. Το Keras είναι ενσωματωμένο στο TensorFlow, καθιστώντας το ως υψηλής προτίμησης API υψηλού επιπέδου για τη δημιουργία και την εκπαίδευση νευρωνικών δικτύων εντός του TensorFlow.

Σε αυτό το κεφάλαιο λοιπόν, θα αναλύσουμε τα νευρωνικά δίκτυα που επιλέξαμε και προσαρμόσαμε με διαφορετικούς τρόπους ώστε αυξήσουμε την ακρίβεια και να μειώσουμε την απώλεια ώστε να μην είναι υπερπροσαρμοσμένο το τελικό μοντέλο. Όπως είχαμε αναφέρει και παραπάνω, χρησιμοποιήσαμε την βιβλιοθήκη FastText για την δημιουργία ενσωματώσεων λέξεων (word embeddings) που εισάγαμε σε κάθε μοντέλο ανάλογα με το Dataset που χρησιμοποιούσαμε για την εκπαίδευση. Κύριοι σκοποί για τη χρήση τους ήταν:

- **Η σημασιολογική κατανόηση**. Χρησιμοποιώντας αυτές ως αρχικά βάρη, το μοντέλο έχει την δυνατότητα να ξεκινά με κάποια κατανόηση των σημασιών και των σχέσεων των λέξεων που παρείχαμε.
- **Καλύτερος χειρισμός λέξεων εκτός του λεξιλογίου**. Λέξεις οι οποίες δεν εμφανίστηκαν πριν την εκπαίδευση μπορούν να αναπαρασταθούνε με έναν πιο ουσιώδη τρόπο.
- **Προσαρμογή σε κάποιον τομέα**. Παρέχοντας τις ενσωματώσεις στην προεκπαίδευση του μοντέλου παρέχουμε και γνώσεις για κάποιον συγκεκριμένο τομέα (π.χ. πολιτικά) όπου μερικές ορολογίες, ονόματα ή ακόμη και το πλαίσιο μερικών κειμένων έχουν ζωτική σημασία.

[10] https://www.tensorflow.org/api_docs/python/tf/keras/Model
[11] https://www.tensorflow.org/guide/keras

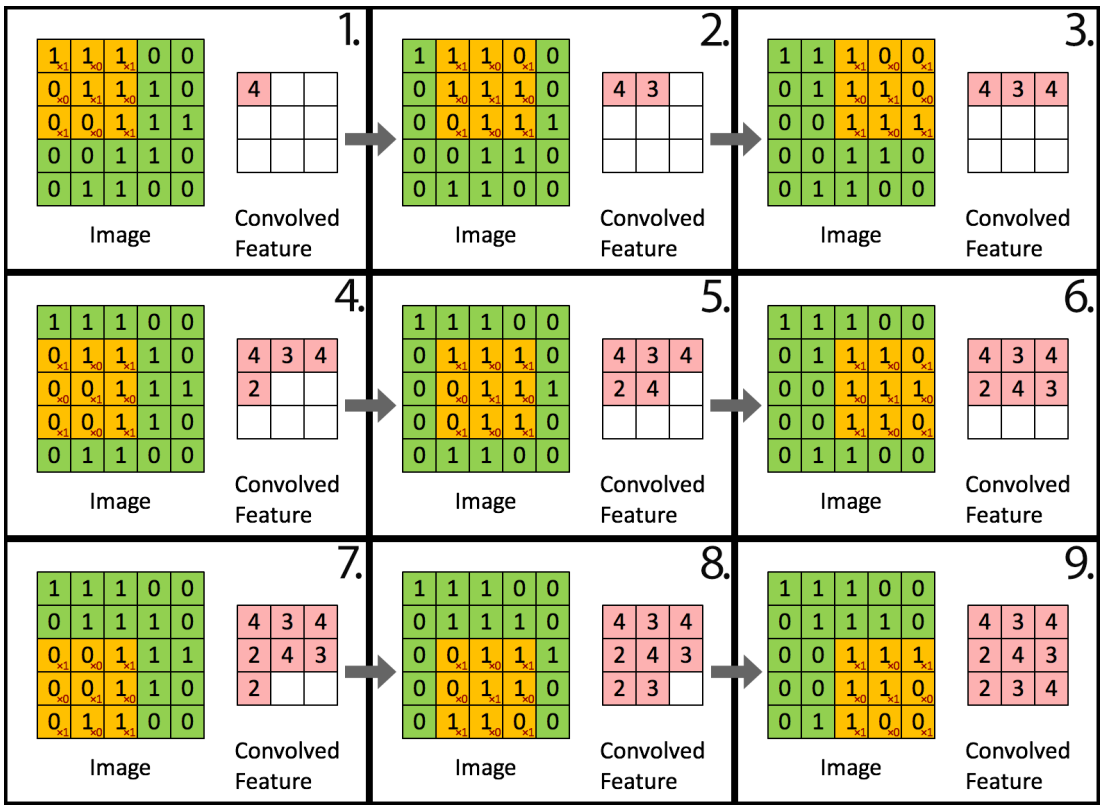
- **Βελτιωμένη σύγκλιση**. Επειδή οι ενσωματώσεις λέξεων παρέχουν ένα καλό σημείο εκκίνησης για το μοντέλο, μειώνεται ο όγκος της εκπαίδευσης που απαιτείται από το νευρωνικό δίκτυο ώστε να μάθει ουσιαστικές αναπαραστάσεις.
- **Μείωση της υπολογιστικής δύναμης**. Εάν εκπαιδεύαμε το μοντέλο από το μηδέν, θα χρειαζόμασταν πολύ μεγαλύτερο όγκο δεδομένων για την επίτευξη ουσιαστικών αποτελεσμάτων πράγμα που θα μας περιόριζε περισσότερο στην υπολογιστική δύναμη.
- **Μικρότεροι χρόνοι εκπαίδευσης**. Το μοντέλο θα έθετε τυχαία βάρη και θα χρειαζόταν περισσότερες επαναλήψεις για να καταλάβει ουσιαστικές αναπαραστάσεις.
- **Υψηλότερος κίνδυνος υπερπροσαρμογής**. Το μοντέλα που δεν χρησιμοποιούν ενσωματώσεις λέξεων είναι πιο επιρρεπή στην υπερπροσαρμογή διότι έχουν περισσότερες παραμέτρους που πρέπει να εκπαιδευτούν και μπορούν εύκολα να απομνημονεύσουν τα δεδομένα της εκπαίδευσης.

Επιλέξαμε να χρησιμοποιήσουμε δύο από τα πιο γνωστά νευρωνικά δίκτυα, τα συνελικτικά νευρωνικά δίκτυα (CNNs) και μια παραλλαγή των επαναλαμβανομένων νευρωνικών δικτύων (RNNs), τα επαναλαμβανόμενα νευρωνικά δίκτυα μακροπρόθεσμης μνήμης (LSTMs) καθώς και ένα υβριδικό μοντέλο που χρησιμοποιεί το CNN και μια παραλλαγή του RNN (πέραν του LSTM) τις περιφραγμένες επαναλαμβανόμενες μονάδες (Gated Recurrent Units, GRU). Κυρίως στα CNN και τα υβριδικά μοντέλα τροποποιήσαμε αλλάζοντας μερικές μεταβλητές ή/και προσθέτοντας περισσότερα στρώματα προσπαθώντας να πετύχουμε μια μεγαλύτερη ακρίβεια όταν κάναμε προβλέψεις στα Datasets.

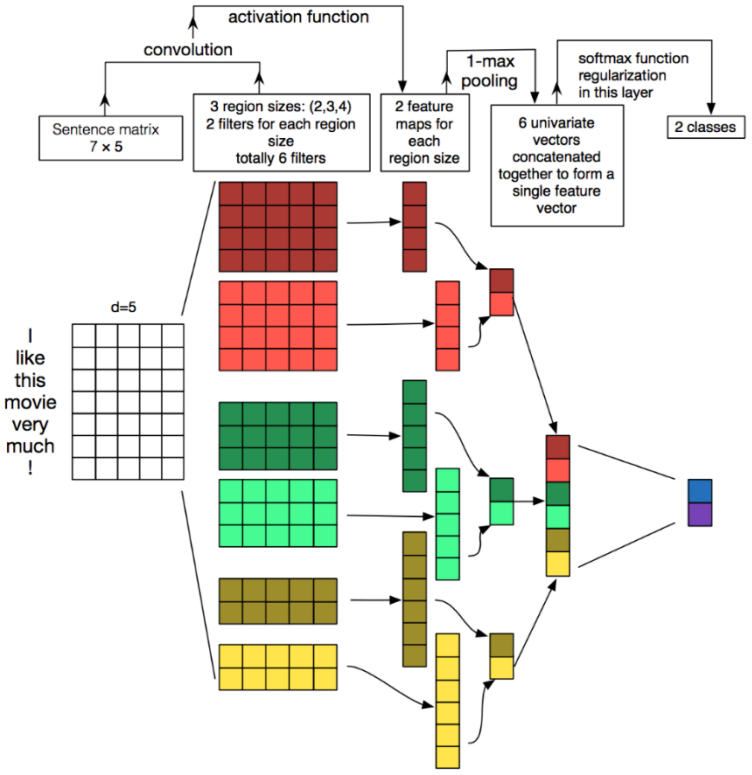
(Υποκεφάλαιο 5.1) **Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks, CNN)**

Τα συνελικτικά νευρωνικά δίκτυα (CNN) ανήκουν στην κατηγορία μοντέλων deep learning που έχουν σχεδιαστεί για την επεξεργασία και την εξαγωγή μοτίβων από δεδομένα τύπου πλέγματος (grid), όπως εικόνες αλλά έχει αποδειχθεί πως είναι σαφώς αποτελεσματικά σε εργασίες ταξινόμησης κειμένου [29]. Η αρχιτεκτονική τους δίνει την δυνατότητα να προσαρμοστούν για εργασίες ταξινόμησης κειμένου, αντιμετωπίζοντας τα δεδομένα κειμένου ως μονοδιάστατες ακολουθίες και εφαρμόζοντας συνελικτικά και ομαδικά επίπεδα για την εξαγωγή χαρακτηριστικών. Θα αναφέρουμε περιληπτικά την διαδικασία όπου λειτουργούν τα συνελικτικά νευρωνικά δίκτυα όταν αντιμετωπίζουν δεδομένα κειμένου. Αρχικά, τα δεδομένα κειμένου πρέπει να υποβληθούν σε προεπεξεργασία και να μετατραπούν σε αριθμητική μορφή όπου θα χρησιμοποιηθούν ως είσοδοι για το CNN. Τα συνελικτικά στρώματα (convolutional layers) είναι τα βασικά δομικά στοιχεία των CNN. Αυτά εφαρμόζονται 'σύροντας' μικρά παράθυρα (ή πυρήνες) σταθερού μεγέθους πάνω από το κείμενο εισαγωγής για να ανιχνεύσουν τοπικά μοτίβα και χαρακτηριστικά. Περιγραφική εικόνα 5.1[12] του παραπάνω βήματος. Οι πυρήνες (kernels) είναι φίλτρα με δυνατότητα εκμάθησης που καταγράφουν συγκεκριμένα μοτίβα ή n-grams στο κείμενο εισαγωγής. Η λειτουργία συνέλιξης περιλαμβάνει τη λήψη του γινόμενου κουκίδων του πυρήνα και των διανυσμάτων χαρακτηριστικών εισόδου μέσα στο συρόμενο παράθυρο δημιουργώντας χάρτες χαρακτηριστικών που τονίζουν την παρουσία συγκεκριμένων μοτίβων. Μετά τη λειτουργία συνέλιξης, μια συνάρτηση ενεργοποίησης (activation function) εφαρμόζεται στα στοιχεία για την εισαγωγή μη γραμμικότητας στο μοντέλο. Τα επίπεδα συγκέντρωσης (pooling layers) εφαρμόζονται για τη μείωση της διάστασης των χαρτών χαρακτηριστικών, διατηρώντας παράλληλα τις πιο σημαντικές πληροφορίες. Η εικόνα 5.2 απεικονίζει αναλυτικά την αρχιτεκτονική CNN για ταξινόμηση προτάσεων [30].

[12] http://deeplearning.stanford.edu/tutorial/supervised/FeatureExtractionUsingConvolution/



Εικόνα 5.1: Διαδικασία του πως σέρνεται ένας πυρήνας ενός συνελικτικού στρώματος πάνω σε δεδομένα για την ανίχνευση χαρακτηριστικών



Εικόνα 5.2: «Απεικόνιση αρχιτεκτονικής CNN για ταξινόμηση προτάσεων. Απεικονίζονται τρία μεγέθη περιοχής φίλτρου (filter region sizes): 2, 3 και 4, καθένα από τα οποία έχει 2 φίλτρα. Τα φίλτρα εκτελούν συνελίξεις στον πίνακα προτάσεων και δημιουργούν χάρτες χαρακτηριστικών (μεταβλητού μήκους). Το 1-max

pooling εκτελείται σε κάθε χάρτη, δηλαδή καταγράφεται ο μεγαλύτερος αριθμός από κάθε χάρτη χαρακτηριστικών. Έτσι, δημιουργείται ένα μονομεταβλητό διάνυσμα χαρακτηριστικών και από τους έξι χάρτες, και αυτά τα 6 χαρακτηριστικά συνδέονται για να σχηματίσουν ένα διάνυσμα χαρακτηριστικών για το προτελευταίο στρώμα. Το τελικό επίπεδο softmax λαμβάνει στη συνέχεια αυτό το διάνυσμα χαρακτηριστικών ως είσοδο και το χρησιμοποιεί για να ταξινομήσει την πρόταση. Στην εικόνα υποθέτετε δυαδική ταξινόμηση και επομένως απεικονίζονται δύο πιθανές καταστάσεις εξόδου»

(Ενότητα 5.1.α) **Αρχιτεκτονική**

Η αρχιτεκτονική του αποτελείται από:

- **Συνελικτικά επίπεδα (Convolutional Layers)**: Για την ταξινόμηση κειμένου, η σάρωση του κειμένου, που δέχεται το μοντέλο ως είσοδο, γίνεται με τα συνελικτικά επίπεδα χρησιμοποιώντας μικρά φίλτρα (ή πυρήνες) ώστε να καταγραφούν τοπικά μοτίβα και χαρακτηριστικά. Αυτά τα φίλτρα λέμε ότι 'γλιστρούνε' πάνω από το κείμενο εισόδου χρησιμοποιώντας τον τελεστή του πολλαπλασιασμού στα στοιχεία και συγκεντρώνοντας τα αποτελέσματα. Με αυτόν τον τρόπο δημιουργούνται χάρτες χαρακτηριστικών που δίνουν έμφαση στην παρουσία συγκεκριμένων μοτίβων στο κείμενο. Υπάρχει πάντοτε και η δυνατότητα προσθήκης περισσοτέρων φίλτρων και αλλαγής των μεγεθών τους για διαφορετική καταγραφή μοτίβων και ως επί το πλείστον, αποτελεσμάτων.

- **Επίπεδα μέγιστης ομαδοποίησης (Max-Pooling Layers)**: Έπειτα από την προσθήκη και χρήση των συνελικτικών στρωμάτων, ακολουθούν πολύ συχνά στρώματα ομαδοποίησης όπου μειώνουν τις διαστάσεις του χώρου που αντιστοιχούν στους χάρτες των χαρακτηριστικών επιλέγοντας μια μέγιστη τιμή σε μια τοπική περιοχή. Αυτό έχει ως αποτέλεσμα την μείωση της υπολογιστικής ισχύος και της διατήρησης των χαρακτηριστικών που θεωρεί το μοντέλο πιο σημαντικά.

- **Πρόσθετα επίπεδα**: Λαμβάνοντας υπόψιν την πολυπλοκότητα της εργασίας που έχουμε να αντιμετωπίσουμε μπορούμε να συμπεριλάβουμε πρόσθετα επίπεδα όπως τα συνδετικά στρώματα (Concatenating Layers), τα στρώματα εγκατάλειψης (Dropout Layers), ισοπεδωτικό στρώμα (Flatten Layer), πλήρως συνδεδεμένα πυκνά στρώματα (Fully Connected Dense Layers), και ένα μοναδικό πυκνό στρώμα (Dense Layer). Τα συνδετικά στρώματα συγχωνεύουν χαρακτηριστικά που εξάγονται από διαφορετικά μεγέθη φίλτρων στα συνελικτικά στρώματα επιτρέποντας το μοντέλο να εξετάσει μοτίβα διαφορετικών μεγεθών στο εισαγόμενο κείμενο. Τα στρώματα εγκατάλειψης εμποδίζουν την υπερπροσαρμογή ρίχνοντας ένα κλάσμα νευρώνων κατά τη διάρκεια της προπόνησης. Το ισοπεδωτικό είναι η διαδικασία της προετοιμασίας των δεδομένων για το μοναδικό πυκνό στρώμα, το οποίο αναμένει μονοδιάστατη είσοδο. Τα πλήρως συνδεδεμένα πυκνά στρώματα μπορούν να βοηθήσουν στην εκμάθηση αφαιρέσεων υψηλού επιπέδου (high-level abstractions) από τα εξαγόμενα χαρακτηριστικά. Το μοναδικό πυκνό στρώμα είναι υπεύθυνο για την πραγματοποίηση προβλέψεων και σε συνδυασμό με υπερπαραμέτρους υποδεικνύουμε ότι αυτό το μοντέλο χρησιμοποιείται για δυαδική ταξινόμηση.

(Ενότητα 5.1.β) **Εξαγωγή χαρακτηριστικών**

Η καταγραφή των τοπικών μοτίβων από τα μοντέλα CNN είναι αποτελεσματική επειδή ο τρόπος που δέχονται τα κείμενα ορίζεται από το μέγεθος των συνελικτικών φίλτρων. Τα μοτίβα

αναγνωρίζονται σύροντας τα φίλτρα πάνω στο κείμενο και έτσι τα κείμενα που περιέχουν συνδυασμούς λέξεων μπορούν να παρέχουν συγκεκριμένα χαρακτηριστικά ή συμφραζόμενα στο μοντέλο. Οι ιεραρχικές αναπαραστάσεις είναι επίσης κάτι που μπορούν τα CNN να μάθουν όπως τα χαμηλότερα επίπεδα που μπορεί να αποτελούνται από μεμονωμένες λέξεις ή συνδυασμούς λέξεων και τα υψηλότερα επίπεδα με πιο αφηρημένες έννοιες ή χαρακτηριστικά με τελικό αποτέλεσμα να καταγράφουν ένα ευρύ φάσμα μοτίβων. Όταν η θέση ενός στοιχείου μπορεί να ποικίλλει μέσα στο κείμενο αυτό μπορεί να δώσει μια μεταβαλλόμενη μετάφραση, τα CNN μπορούν να ανιχνεύσουν μοτίβα ανεξάρτητα από τη θέση τους στο εισαγόμενο κείμενο καθιστώντας τα μεταφραστικά αμετάβλητα.

(Υποκεφάλαιο 5.2) **Επαναλαμβανόμενα νευρωνικά δίκτυα (Recurrent Neural Networks, RNN)**

Τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) είναι μια κατηγορία τεχνητών νευρωνικών δικτύων που έχουν σχεδιαστεί για την επεξεργασία ακολουθιών δεδομένων. Έχουν εφαρμογή σε οποιαδήποτε εργασία περιλαμβάνει διαδοχικά δεδομένα όπως την NLP στην δικιά μας περίπτωση, αφού επεξεργάζονται δεδομένα εισόδου καταγράφονται εξαρτήσεις με την πάροδο του χρόνου. Τα RNN έχουν επαναλαμβανόμενες συνδέσεις, πράγμα που σημαίνει ότι οι πληροφορίες κάνουν κύκλους μέσα στο δίκτυο. Υπάρχει όμως, και η κρυφή κατάσταση που χρησιμεύει ως μνήμη για τα προηγούμενα βήματα που έκανε το δίκτυο επιτρέποντας το να διατηρεί τα συμφραζόμενα και να λαμβάνει διαδοχικές πληροφορίες. Έτσι, για κάθε χρονικό βήμα, όταν το δίκτυο λαμβάνει μια είσοδο ενημερώνει την κρυφή του μνήμη. Παρόλα αυτά τα παραδοσιακά RNN παρουσιάζουν περιορισμούς όπου έφεραν το κίνητρο να αναπτυχθούν εξειδικευμένες παραλλαγές όπως το LSTM και GRU. Θα αναλύσουμε αυτούς τους περιορισμούς για να τονίσουμε την σημαντικότητα της χρήσης αυτών των μοντέλων στην NLP:

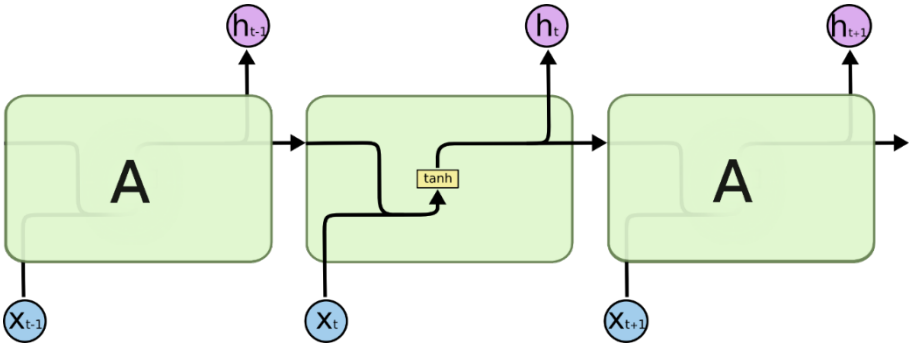
- Ένας από τους κύριους περιορισμούς των παραδοσιακών RNN είναι το πρόβλημα της **εξαφάνισης της κλίσης (Vanishing Gradient Problem)**. Κατά τη διάρκεια της εκπαίδευσης, όταν οι κλίσεις οπισθοπολλαπλασιάζονται μέσω του δικτύου, τείνουν να γίνονται εξαιρετικά μικρές καθώς διασχίζουν το δίκτυο από το επίπεδο εξόδου στο επίπεδο εισόδου. Αυτό δημιουργεί πρόβλημα στην καταγραφή εξαρτήσεων που επεκτείνονται σε πολλά χρονικά βήματα σε μια ακολουθία, με αποτέλεσμα το μοντέλο να αδυνατεί να μάθει από πληροφορίες που αφορούν το μακρινό παρελθόν, πράγμα που είναι ζωτικής σημασίας στην κατανόηση των συμφραζομένων ενός κειμένου.

- **Αδυναμία διατήρησης πληροφοριών**. Τα παραδοσιακά RNN έχουν περιορισμένη χωρητικότητα μνήμης που δεν τα καθιστά κατάλληλα για τη διατήρηση πληροφοριών σε εκτεταμένες ακολουθίες. Αυτό συμβαίνει όταν το δίκτυο επεξεργάζεται δεδομένα εισόδου και πρέπει συνεχώς να αντικαθιστά παλιές πληροφορίες στην κρυφή του μνήμη με αποτέλεσμα την αδυναμία καταγραφής περίπλοκων σχέσεων μεταξύ των κειμένων και πτώση απόδοσης όταν η κατανόηση όλων των συμφραζομένων και του ιστορικού είναι σημαντική.

- **Προκλήσεις στην εκπαίδευση**. Η αργή σύγκλιση μπορεί να οδηγήσει σε μεγάλους χρόνους εκπαίδευσης, καθιστώντας τη λιγότερο πρακτική για μεγάλα σύνολα δεδομένων και πολύπλοκα μοντέλα. Ενώ ακόμη, τα παραδοσιακά RNN είναι εγγενώς διαδοχικά στη φύση τους, επεξεργάζονται δεδομένα ένα βήμα τη φορά με αποτέλεσμα την έλλειψη παραλληλισμού και κατ' επέκταση την ανικανότητα της πλήρης εκμετάλλευσης του

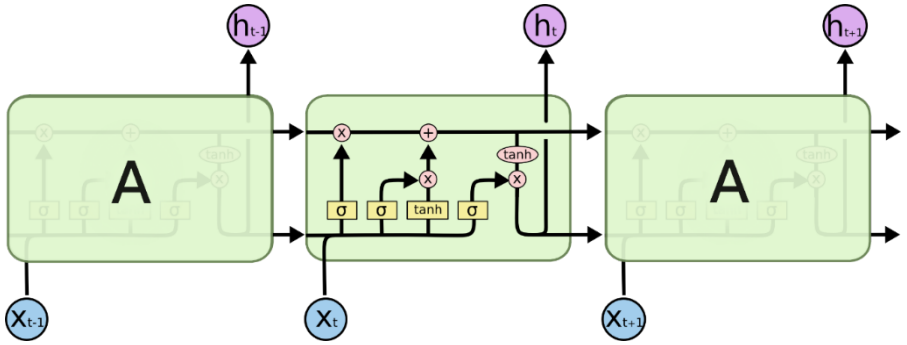
σύγχρονου υλικού όπως των καρτών γραφικών οι οποίες υποστηρίζουν την παράλληλη επεξεργασία.

Για αυτόν τον λόγο τα LSTM έχουν προταθεί. Έρχονται για να αντικαταστήσουν το επίπεδο των Απλών Επαναλαμβανόμενων Νευρωνικών Δικτύων (Simple RNN), καθώς αυτό είναι πολύ απλοϊκό για ρεαλιστική χρήση σε σημερινά προβλήματα βαθιάς μάθησης. Τα LSTM αναπτύχθηκαν από τον Hochreiter και τον Schmidhuber το 1997 [31]. Ο αλγόριθμος αυτός δίνει την δυνατότητα μεταφοράς δεδομένων μέσα από πολλαπλά βήματα χρόνου κατά την εκτέλεση του προγράμματος. Με άλλα λόγια, ο LSTM 'αποθηκεύει' πληροφορίες για αργότερα και έτσι αποτρέπει την εξαφάνιση παλαιότερων πληροφοριών κατά την επεξεργασία. Όλα τα επαναλαμβανόμενα νευρωνικά δίκτυα επαναλαμβάνονται αλυσιδωτά μέσα σε ένα νευρωνικό δίκτυο. Στα απλούστερα δίκτυα αυτά, κάθε επανάληψη περιλαμβάνει μόνο ένα επίπεδο. Η εικόνα 5.3 [32] δείχνει ένα απλό RNN ενός στρώματος.



Εικόνα 5.3: Απλό RNN ενός στρώματος. Κάθε επανάληψη περιέχει μόνο ένα επίπεδο Υπερβολικής Εφαπτομένης (tanh)

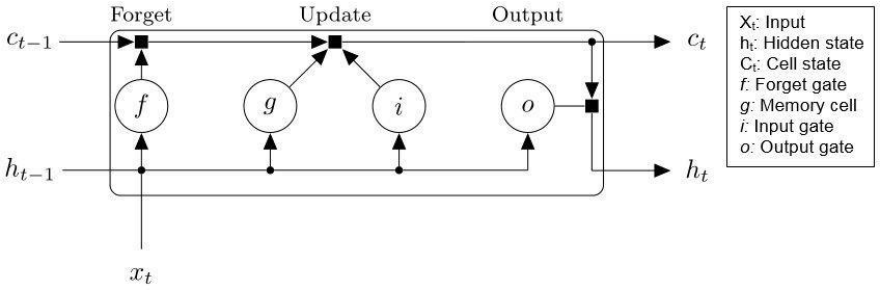
Τα επίπεδα LSTM έχουν παρόμοια αλυσιδωτή δομή, όμως κάθε επανάληψη είναι διαφορετική. Αυτό συμβαίνει διότι αντί να περιέχουν μόνο ένα απλό επίπεδο (όπως κάθε κοινότυπο RNN), χρησιμοποιούν 4 επίπεδα νευρωνικών δικτύων, τα οποία αλληλοεπιδρούν μεταξύ τους. Η εικόνα 5.4 [32] δείχνει τα επίπεδα του LSTM.



Εικόνα 5.4: Η επαναλαμβανόμενη μονάδα και τα τέσσερα επίπεδα του LSTM που αλληλεπιδρούν

Η αρχιτεκτονική τους ενσωματώνουν εξειδικευμένα κελιά μνήμης που μπορούν να αποθηκεύσουν και να ανακτήσουν πληροφορίες σε μεγάλες ακολουθίες. Χρησιμοποιούν τρεις κύριες πύλες, μια πύλη εισόδου (input gate), μια πύλη για να 'ξεχνά' (forget gate) και μια πύλη εξόδου (output gate). Αυτές οι πύλες ελέγχουν τη ροή των πληροφοριών στα κελιά. Η πύλη που ξεχνά επιτρέπει στο μοντέλο να αποφασίσει ποιες πληροφορίες από το προηγούμενο χρονικό βήμα θα ξεχάσει, η πύλη εισόδου καθορίζει ποιες νέες πληροφορίες θα αποθηκεύσει στο κελί μνήμης ενώ η πύλη εξόδου ελέγχει ποιες πληροφορίες εκτίθενται στο επόμενο επίπεδο ή στην έξοδο. Με τις πύλες αυτές επιτυγχάνεται η επεξεργασία και αποθήκευση των δεδομένων κατά την επεξεργασία και επανάληψη

των LSTM, επιτρέποντας να αντιμετωπίζονται ακολουθίες δεδομένων με μακροχρόνιες εξαρτήσεις. Η εικόνα 5.5[13] δείχνει πως μοιάζει μία αρχιτεκτονική ενός LSTM.



Εικόνα 5.5: Το LSTM έχει περισσότερες πύλες σε σχέση με ένα απλό RNN για τον έλεγχο της ροής πληροφοριών

Τα GRU από την άλλη έχουν απλοποιημένη αρχιτεκτονική σε σύγκριση με τα LSTM. Ενώ υπάρχουν οι μηχανισμοί πύλης, αντίθετα με το LSTM ξεχνούν σε μια ενιαία πύλη 'ενημέρωσης' και συγχωνεύουν την κατάσταση κελιού και την κρυφή κατάσταση. Πρόσθετα υπάρχει και μια πύλη επαναφοράς. «Όπως και με τα LSTM, στις πύλες ενημέρωσης και επαναφοράς δίνονται σιγμοειδείς ενεργοποιήσεις»[14]. «Η πύλη επαναφοράς ελέγχει πόσο από την προηγούμενη κατάσταση ίσως θέλει να θυμάται και η πύλη ενημέρωσης επιτρέπει τον έλεγχο για το πόσο από τη νέα κατάσταση είναι απλώς ένα αντίγραφο της παλιάς»[14]. Η υποψήφια κρυφή κατάσταση (candidate hidden state) είναι ένας ενδιάμεσος υπολογισμός εντός του κελιού GRU και αναπαριστά προσωρινά τις πιθανές ενημερώσεις στην κρυφή κατάσταση. Ενσωματώνει πληροφορίες τόσο από την προηγούμενη κρυφή κατάσταση όσο και από την τρέχουσα είσοδο, ενώ οι πύλες επαναφοράς και ενημέρωσης ελέγχουν τη ροή των πληροφοριών και επηρεάζουν την τελική κρυφή κατάσταση. Η κρυφή κατάσταση (hidden state), συνοπτικά, σε ένα GRU είναι μια δυναμική αναπαράσταση που κωδικοποιεί πληροφορίες από προηγούμενα χρονικά βήματα και προσαρμόζεται καθώς επεξεργάζονται νέες πληροφορίες.

(Υποκεφάλαιο 5.3) **Υβριδικά μοντέλα (CNN & GRU)**

Σε αυτό το υποκεφάλαιο θα μιλήσουμε για την επιλογή της χρήσης υβριδικού μοντέλου για τους σκοπούς της ταξινόμησης κειμένου. Θα εμβαθύνουμε στην αρχιτεκτονική και τα πλεονεκτήματα του υβριδικού μας μοντέλου, το οποίο συνδυάζει επίπεδα των συνελικτικών νευρωνικών δικτύων (CNN) και των περιφραγμένων επαναλαμβανόμενων μονάδων (GRU).

(Ενότητα 5.3.α) **Αρχιτεκτονική**

Όταν δημιουργηθούν οι αλληλουχίες κειμένων που έχουν περάσει από την διαδικασία του padding και του tokenization και έχουν εισαχθεί στο στρώμα ενσωμάτωσης (embedding layer), τροφοδοτούνται στη συνέχεια σε μια στοίβα επιπέδων CNN, καθένα από τα οποία εκτελεί λειτουργίες συνέλιξης και συγκέντρωση για να καταγράψει τοπικά μοτίβα και χαρακτηριστικά στο κείμενο. Μετά τα επίπεδα CNN, τα δεδομένα μεταβιβάζονται στα επίπεδα GRU. Τα επίπεδα GRU είναι υπεύθυνα για την καταγραφή των διαδοχικών εξαρτήσεων και του περιβάλλοντος μέσα στο κείμενο. Σε αντίθεση με τα παραδοσιακά RNN, τα GRU μετριάζουν το πρόβλημα της κλίσης που εξαφανίζεται, επιτρέποντάς τους να καταγράφουν εξαρτήσεις μεγάλης εμβέλειας χωρίς απώλεια πληροφοριών. Έπειτα, ακολουθούνε πρόσθετα στρώματα όπως στρώματα εγκατάλειψης και πλήρως

[13] https://www.mathworks.com/discovery/lstm.html
[14] https://d2l.ai/chapter_recurrent-modern/gru.html

συνδεδεμένα πυκνά στρώματα για την μείωση της υπερπροσαρμογής και την δυνατότητα καταγραφής πολύπλοκων μοτίβων και σχέσεων μέσα στα δεδομένα από το μοντέλο. Τελικά, η έξοδος από αυτά τα επίπεδα χρησιμοποιείται στη συνέχεια για εργασίες ταξινόμησης μέσω ενός πυκνού στρώματος εξόδου.

(Ενότητα 5.3.β) **Ενοποίηση CNN και GRU και πλεονεκτήματα**

Η υβριδική αρχιτεκτονική αξιοποιεί τις συμπληρωματικές δυνάμεις των επιπέδων CNN και GRU. Έχει αποδειχθεί ακόμη ως μια αποτελεσματική μέθοδος για την ταξινόμηση κειμένων [33].Τα CNN υπερέχουν στον εντοπισμό τοπικών μοτίβων και χαρακτηριστικών, όπως τα n-grams, ενώ τα GRU είναι ικανά στη σύλληψη διαδοχικών πληροφοριών. Η έξοδος των επιπέδων CNN, η οποία περιέχει πληροφορίες για τοπικά χαρακτηριστικά, ενσωματώνεται απρόσκοπτα στα επίπεδα GRU. Αυτή η ενοποίηση διασφαλίζει ότι τα τοπικά χαρακτηριστικά που ανιχνεύονται από τα CNN λαμβάνονται υπόψη στο πλαίσιο ακολουθίας, επιτρέποντας στο μοντέλο να λάβει πιο ενημερωμένες αποφάσεις ταξινόμησης. Αυτή η ιεραρχική αναπαράσταση επιτρέπει στο μοντέλο μας να συλλαμβάνει τόσο τις λεπτομέρειες μικροεπιπέδου (τοπικά μοτίβα) όσο και το περιβάλλον μακροεπίπεδου (διαδοχικές εξαρτήσεις) του κειμένου εισόδου, καθιστώντας το κατάλληλο για ένα ευρύ φάσμα εργασιών ταξινόμησης κειμένου. Τελικά, η υβριδική αυτή αρχιτεκτονική CNN-GRU προσφέρει πλεονεκτήματα σε σχέση με τα αυτόνομα μοντέλα CNN ή RNN:

- **Αποτελεσματική εξαγωγή χαρακτηριστικών:** Τα CNN είναι ικανά να εξάγουν σχετικές λειτουργίες από το κείμενο, βελτιώνοντας την κατανόηση του περιεχομένου από το μοντέλο.

- **Διαδοχικά συμφραζόμενα:** Οι GRU παρέχουν στο μοντέλο τη δυνατότητα να εξετάζει τη σειρά και το πλαίσιο των λέξεων στο κείμενο, βελτιώνοντας την ικανότητά του να καταγράφει εξαρτήσεις.

- **Βελτιωμένη απόδοση:** Η συγχώνευση των επιπέδων CNN και GRU έχει ως αποτέλεσμα βελτιωμένη απόδοση σε σύγκριση με τη χρήση οποιασδήποτε αρχιτεκτονικής μεμονωμένα.

- **Ευελιξία:** Τα υβριδικά μοντέλα είναι ευέλικτα και μπορούν να προσαρμοστούν σε διάφορες εργασίες ταξινόμησης κειμένου, καθιστώντας τα ένα πολύτιμο εργαλείο για την επεξεργασία φυσικής γλώσσας.

(Υποκεφάλαιο 5.4) **Υλοποίηση μοντέλων**

Σε αυτό το υποκεφάλαιο θα αναλύσουμε τα διαφοροποιημένα μοντέλα που δημιουργήσαμε, τις διαδικασίες που ακολούθησαν πριν την μεταγλώττιση των μοντέλων καθώς και την διαδικασία της εκπαίδευσης. Αρχικά, η διαδικασία που ακολουθήθηκε πριν την μεταγλώττιση των μοντέλων καθώς και η διαδικασία της εκπαίδευσης είναι κοινές σε όλα τα μοντέλα. Για κάθε μοντέλο θα αναφέρουμε τις υπερπαραμέτρους που αλλάξαμε, τα στρώματα που πιθανώς να αφαιρέσαμε ή προσθέσαμε, ή έναν συνδυασμό αυτών των δύο που ενσωματώσαμε σε κάθε παραλλαγή ενός μοντέλου.

(Ενότητα 5.4.α) **Αρχική διαδικασία προεκπαίδευσης**

Τα βήματα που ακολουθούμε για την διαδικασία της προεκπαίδευσης είναι τα εξής:

Βήμα 1: Φορτώνουμε το αντίστοιχο Dataset (Fake News Corpus, WELFake ή LIAR) και χρησιμοποιούμε δέκα χιλιάδες (10.000) δείγματα με αναλογία 45% τα δεδομένα με ταμπέλα ψευδής είδηση και 55% τα δεδομένα με ταμπέλα αληθής είδηση. Το LIAR το φορτώνουμε χωρίς περιορισμό των δειγμάτων.

Βήμα 2: Διαχωρισμός δεδομένων σε εκπαίδευσης και δοκιμής

Βήμα 3: Προετοιμασία και χρήση του tokenizer για την μετατροπή των δεδομένων κειμένου σε αριθμητικά tokens για τη δημιουργία ενός λεξιλογίου και την ανάθεση ενός μοναδικού αριθμητικού ευρετηρίου σε κάθε λέξη.

Βήμα 4: Αποθήκευση των μηκών όλων των ακολουθιών σε μια λίστα, αποθήκευση του μεγίστου μήκους ακολουθίας μεταξύ των ακολουθιών και υπολογισμός του 95^{ου} εκατοστημορίου (εκτός του LIAR Dataset) των μηκών ακολουθίας το οποίο χρησιμοποιείται ως το μέγιστο μήκος ακολουθίας για αλληλουχίες συμπλήρωσης. Το κ-εκατοστημόριο και συγκεκριμένα το 95^{ο} εκατοστημόριο (percentile 95) το χρησιμοποιούμε ώστε να μην έχουμε τεράστια μεγέθη στις ακολουθίες αφού αυτές μπορεί να ανήκουν στο 5% της κατανομής και με αυτόν τον τρόπο έχουμε ένα τυπικό μήκος για τις ακολουθίες στο dataset και γλυτώνουμε σε μερικές περιπτώσεις το να υπάρχουν τεράστια μήκη και να γίνεται πολύ υπολογιστικά ακριβό για εκπαίδευση το μοντέλο μας.

Βήμα 5: Εφαρμογή του padding στις ακολουθίες για να συμπληρωθούν ή να περικοπούν οι ακολουθίες στο καθορισμένο μήκος από το 95^{ο} εκατοστημόριο. Αυτό το βήμα διασφαλίζει ότι όλες οι ακολουθίες έχουν το ίδιο μήκος, το οποίο απαιτείται για είσοδο σε ένα νευρωνικό δίκτυο.

Βήμα 6: Φόρτωση του προεκπαιδευμένου FastText μοντέλου, σύμφωνα με το αντίστοιχο σύνολο δεδομένου, που εκπαιδεύσαμε για την δημιουργία των ενσωματώσεων λέξεων, όπου είναι ουσιαστικά οι συνεχείς διανυσματικές αναπαραστάσεις λέξεων από τα λεξιλόγια κάθε συνόλου δεδομένων.

Βήμα 7: Αρχικοποίηση ενός μηδενικού πίνακα με διαστάσεις του μέγεθος του λεξιλογίου και της διάστασης των ενσωματώσεων λέξεων (η οποία ήταν η τιμή "dim" ίσον με τριακόσια, ή εκατό ανάλογα το Dataset, που θέσαμε κατά την εκπαίδευση του FastText μοντέλου).

Βήμα 8: Στη συνέχεια, ο κώδικας επαναλαμβάνεται για όλες τις λέξεις στο λεξιλόγιο του tokenizer και ελέγχει εάν κάθε λέξη υπάρχει στο λεξιλόγιο του μοντέλου FastText. Εάν βρεθεί μια λέξη στο μοντέλο FastText, το αντίστοιχο διάνυσμα λέξης ανακτάται και αποθηκεύεται στον πίνακα στον δείκτη που αντιστοιχεί στο αριθμητικό δείκτη της λέξης στο λεξιλόγιο του tokenizer. Ο βρόχος διακόπτεται εάν ο αριθμός των λέξεων που υποβάλλονται σε επεξεργασία φτάσει την τιμή του μεγέθους του λεξιλογίου.

Βήμα 9: Το επίπεδο ενσωμάτωσης (embedding layer) ρυθμίζεται κατάλληλα με διαστάσεις που ταιριάζουν με το μεγέθους του λεξιλογίου, της διάστασης των ενσωματώσεων λέξεων και το μήκος εισόδου το οποίο είναι εκατοστημόριο 95. Τα βάρη του επιπέδου ενσωμάτωσης αρχικοποιούνται με τον πίνακα που δημιουργήθηκε από τα προεκπαιδευμένα διανύσματα λέξεων FastText. Αυτό επιτρέπει στο μοντέλο να χρησιμοποιεί αυτές τις ενσωματώσεις ως αρχικά βάρη κατά τη διάρκεια της προπόνησης. Το τελικό αποτέλεσμα είναι ένα επίπεδο ενσωμάτωσης που χρησιμοποιείται ως το πρώτο επίπεδο του νευρωνικού δικτύου.

(Ενότητα 5.4.β) **Διαδικασία εκπαίδευσης & πρόβλεψης**

Σε αυτή την ενότητα θα αναφέρουμε τους τρόπους με τους οποίους εκπαιδεύσαμε τα μοντέλα όπως την επιλογή των υπερπαραμέτρων, καθώς και τις μετρήσεις αξιολόγησης και των μεθόδων που επιλέξαμε να αποθηκεύσουμε το κάθε μοντέλο.

Αρχικά, να αναφέρουμε πως και όλα τα Database που φορτώνουμε στην μνήμη είναι αυτά που έχουν υποστεί προεπεξεργασία όπως αναλύσαμε σε παραπάνω κεφάλαιο. Οι υπερπαράμετροι που κυρίως τροποποιούμε αφορούν τα CNN και τα υβριδικά μοντέλα όπως για παράδειγμα η μεταβλητή των φίλτρων "filters" για το στρώμα της συνέλιξης και η οι μονάδες "units" για τα στρώματα GRU όπου είναι η διάσταση του χώρου εξόδου (δηλαδή ο αριθμός των φίλτρων εξόδου για το στρώμα της συνέλιξης). Μια ακόμη μεταβλητή που τροποποιούμε είναι το μέγεθος πυρήνα "kernel_size" όπου καθορίζει το μήκος του παραθύρου συνέλιξης. Επιπλέον, σε κάποια μοντέλα χρησιμοποιούνται πλήρως συνδεδεμένα πυκνά στρώματα τα οποία έχουν την παρακάτω μορφή:

- 1ο units ίσο με 64 όπου καθορίζουν ότι το στρώμα έχει 64 νευρώνες. Activation ίσον με "relu" όπου ορίζει τη συνάρτηση ενεργοποίησης του επιπέδου σε Διορθωμένη Γραμμική Μονάδα (ReLU) που είναι κοινή επιλογή για κρυφά επίπεδα. kernel_regularizer ίσον με l2(0.08) που εφαρμόζει κανονικοποίηση L2 στα βάρη του στρώματος με ισχύ κανονικοποίησης 0.08. Αυτό σημαίνει ότι ένας όρος ποινής ανάλογος με το άθροισμα των τετραγώνων των βαρών θα προστεθεί στη συνάρτηση απώλειας κατά τη διάρκεια της προπόνησης για να αποφευχθεί η υπερπροσαρμογή.
- 2ο units ίσο με 32, activation ίσον με "relu" και kernel_regularizer ίσο με l2(0.04)
- 3ο units ίσο με 16, activation ίσον με "relu" και kernel_regularizer ίσο με l2(0.02)

Σε όλα τα μοντέλα χρησιμοποιούνται επίσης και δύο στρώματα εγκατάλειψης (εκτός του μοντέλου που υλοποιήθηκε με LSTM που χρησιμοποιήθηκε ένα στρώμα εγκατάλειψης), ένα πριν τα πυκνά στρώματα και ένα μετά την έξοδο των πυκνών στρωμάτων, (εάν δεν εφαρμόζονται τα πυκνά στρώματα τότε τα δύο στρώματα εγκατάλειψης είναι άνω και κάτω του στρώματος ισοπέδωσης), με ποσοστό εγκατάλειψης 50% και 65% αντίστοιχα. Αυτό σημαίνει ότι κατά τη διάρκεια κάθε βήματος εκπαίδευσης, η έξοδος κάθε νευρώνα στο συγκεκριμένο επίπεδο έχει πιθανότητα 50% ή 65% να μηδενιστεί. Αυτή η διαδικασία στοχαστικής εγκατάλειψης βοηθά στην αποτροπή του μοντέλου να βασίζεται πολύ σε οποιονδήποτε μεμονωμένο νευρώνα και ενθαρρύνει πιο ισχυρές και γενικευμένες αναπαραστάσεις στο νευρωνικό δίκτυο.

Στο τελικό μοναδικό πυκνό στρώμα που είναι υπεύθυνο για την παραγωγή της τελικής εξόδου του δικτύου και κατ' επέκταση του αποτελέσματος της πρόβλεψης ή ταξινόμησης, χρησιμοποιούμε την συνάρτηση ενεργοποίησης η οποία είναι σιγμοειδής (sigmoid). Αυτή παράγει μια βαθμολογία πιθανότητας μεταξύ 0 και 1 για μία από τις δύο κατηγορίες. Τελικά, διαμορφώνουμε και μεταγλωττίζουμε το μοντέλο με συγκεκριμένες ρυθμίσεις για το βελτιστοποιητή, τη συνάρτηση απώλειας και τις μετρήσεις. Χρησιμοποιώντας τον Adam σαν αλγόριθμο βελτιστοποίησης προσαρμόζουμε τον ρυθμό μάθησης, ο οποίος έχει τιμή $2{\cdot}10^{-4}$ σε όλα τα μοντέλα (εκτός άμα γράψουμε διαφορετικά σε ένα συγκεκριμένο), κατά τη διάρκεια της εκπαίδευσης για να βελτιώσει τη σύγκλιση. Κατά την μεταγλώττιση ορίζουμε την συνάρτηση απώλειας σε δυαδική σταυροεντροπία (binary crossentropy) όπου είναι μια κοινή επιλογή για εργασίες δυαδικής ταξινόμησης.

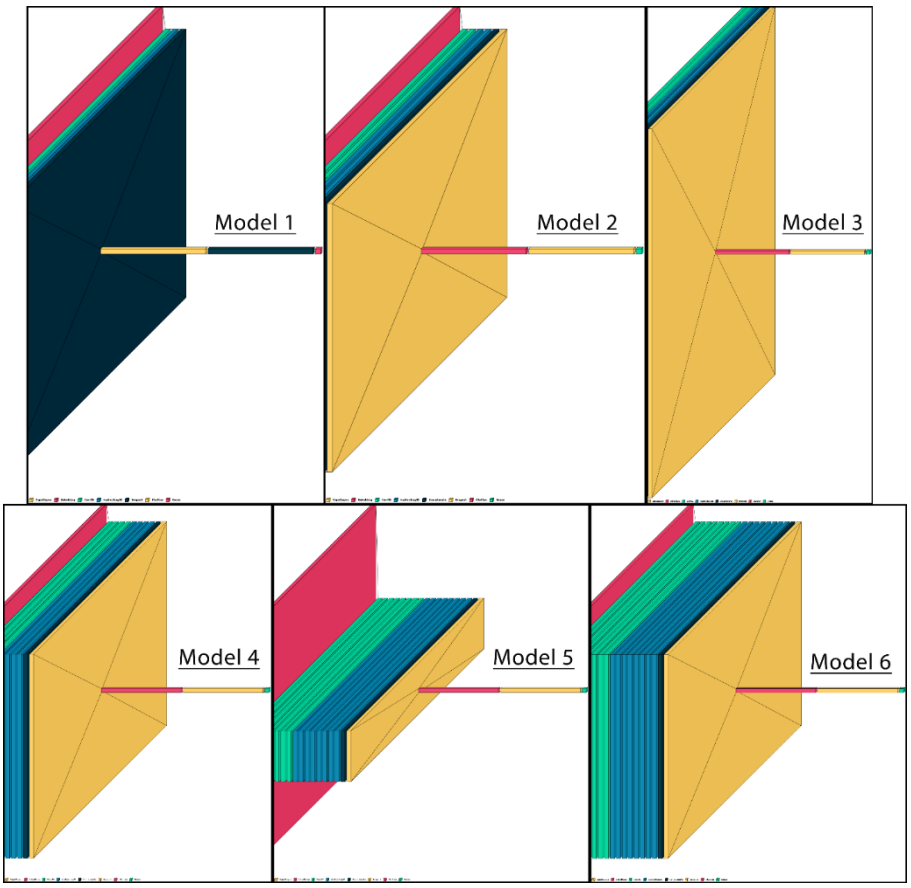
Όσον αφορά την διαδικασία της εκπαίδευσης έχουμε επιλέξει να υιοθετήσουμε την τεχνική της πολλαπλής επικύρωσης (Cross-validation training). Συγκεκριμένα χρησιμοποιούμε αριθμό αναδιπλώσεων ίσον με τρία ενώ ταυτόχρονα η διαδικασία εκπαίδευσης περιλαμβάνει τη ρύθμιση παραμέτρων επανάκλησης όπως το σημείο ελέγχου (checkpoint) και ο διαχωρισμός επικύρωσης (validation splitting). Το σημείο ελέγχου είναι αυτό που χρησιμοποιούμε για να αποθηκεύσουμε το μοντέλο για κάθε αναδίπλωση όπου το έχουμε ρυθμίσει να παρακολουθεί την απώλεια επικύρωσης. Επιπλέον, αποθηκεύουμε τις μετρικές της ακρίβειας, ακρίβειας επικύρωσης, απώλειας και απώλειας επικύρωσης για κάθε εποχή (epoch), όπου για όλα τα μοντέλα είναι ίση με 8 εκτός όταν χρησιμοποιείται το LIAR σύνολο δεδομένων για την εκπαίδευση όπου το epoch είναι ίσο με εκατό, και τα αναπαριστούμε σε γραφήματα για τον έλεγχο της απόδοσης του μοντέλου.

(Υποκεφάλαιο 5.5) **CNN μοντέλα & αποτελέσματα**

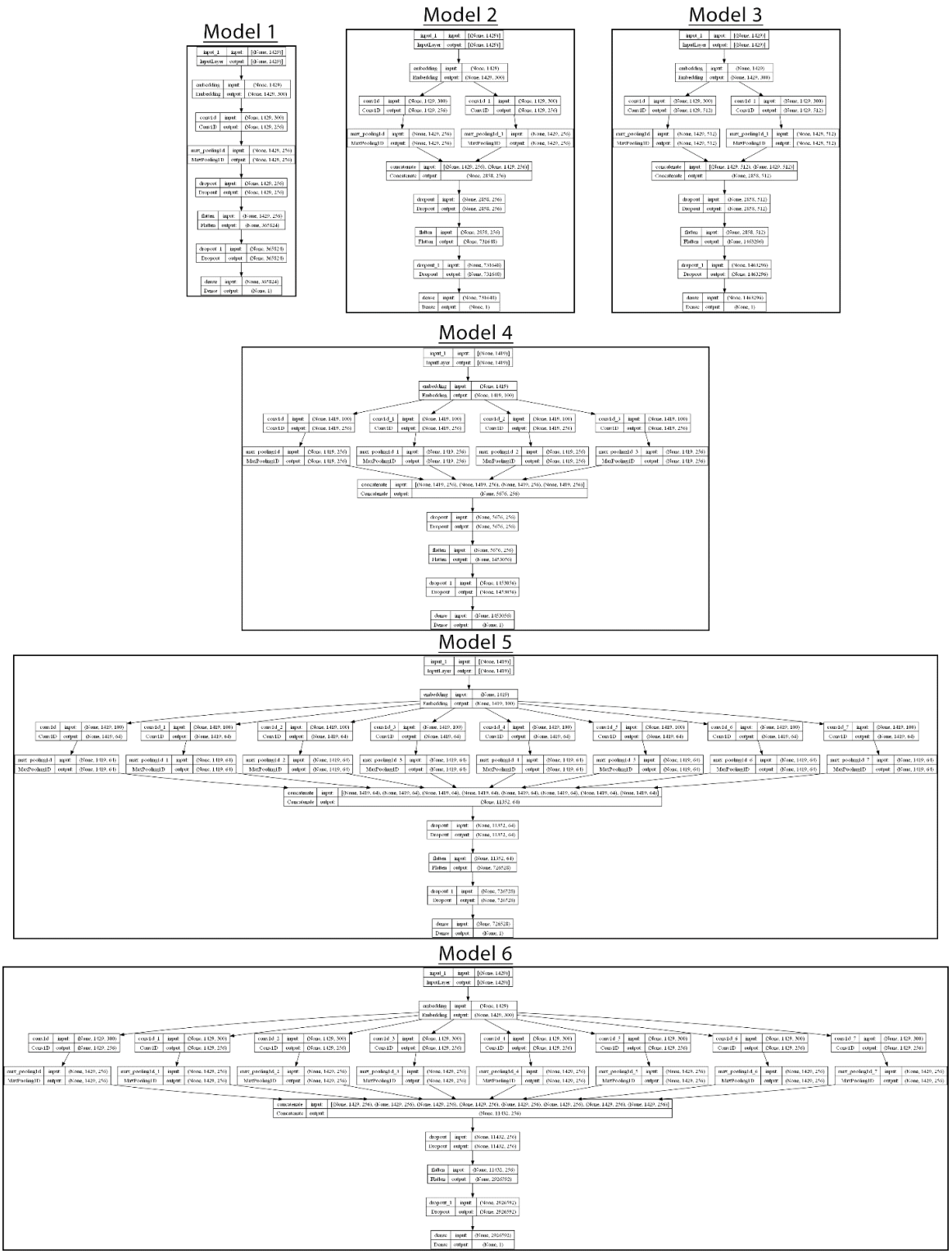
Για τον σκοπό της ταξινόμησης κειμένου δημιουργήσαμε και εξετάσαμε 6 μοντέλα χρησιμοποιώντας τα συνελικτικά νευρωνικά δίκτυα όπου τροποποιήσαμε τις υπερπαραμέτρους και συνδυάσαμε στρώματα και μεταβλητές με διαφορετικές τιμές για να εκπαιδεύσουμε τα μοντέλα. Θα παρουσιάσουμε λοιπόν τα μοντέλα που χρησιμοποιήσαμε, τις παραμέτρους που αλλάξαμε στο καθένα, τις μετρικές τους ανά epoch για κάθε αναδίπλωση και τα τελικά αποτελέσματα που είναι οι αποδόσεις τους κάνοντας προβλέψεις και στα τρία σύνολα δεδομένων. Στον πίνακα 5.1 παρουσιάζουμε τα μοντέλα που χρησιμοποιήσαμε μαζί με τις αλλαγές που έγιναν στο καθένα. Οι εικόνες 5.6 και 5.7 δείχνουν την οπτικοποίηση των μοντέλων ένα έως έξι. Γνωρίζουμε πως λόγω αλλαγών στις τιμές εισόδων στα στρώματα ενσωμάτωσης για κάθε σύνολο δεδομένων οι φωτογραφίες θα είχαν παραλλαγές στα μεγέθη και σε κάποιες τιμές παρόλα αυτά τις χρησιμοποιούμε για να δώσουμε μια γενική ιδέα για το πως μοιάζουν τα μοντέλα. Η εικόνα 5.8 δείχνει την ακρίβεια, ακρίβεια επικύρωσης (πάνω πρώτο μισό μέρος), απώλεια και απώλεια επικύρωσης (κάτω δεύτερο μισό μέρος) ανά epoch για κάθε αναδίπλωση για τα μοντέλα 1 έως 6 για εκπαίδευση με το Fake News Corpus σύνολο δεδομένων. Η εικόνα 5.9 και εικόνα 5.10 δείχνει τα ίδια στοιχεία απλά για τα WELFake και LIAR σύνολα δεδομένων αντίστοιχα. Ο πίνακας 5.2 δείχνει τα τελικά αποτελέσματα για τις προβλέψεις σε κάθε Dataset.

Models / Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Total convolutional layers	1	2	2	4	8	8
Total pooling layers	1	2	2	4	8	8
filters	256	256	512	256	64	256
kernel_size	[3]	[3,4]	[3,4]	[2,3,4,5]	[1,2,3,4,5,6,7,8]	[1,2,3,4,5,6,7,8]

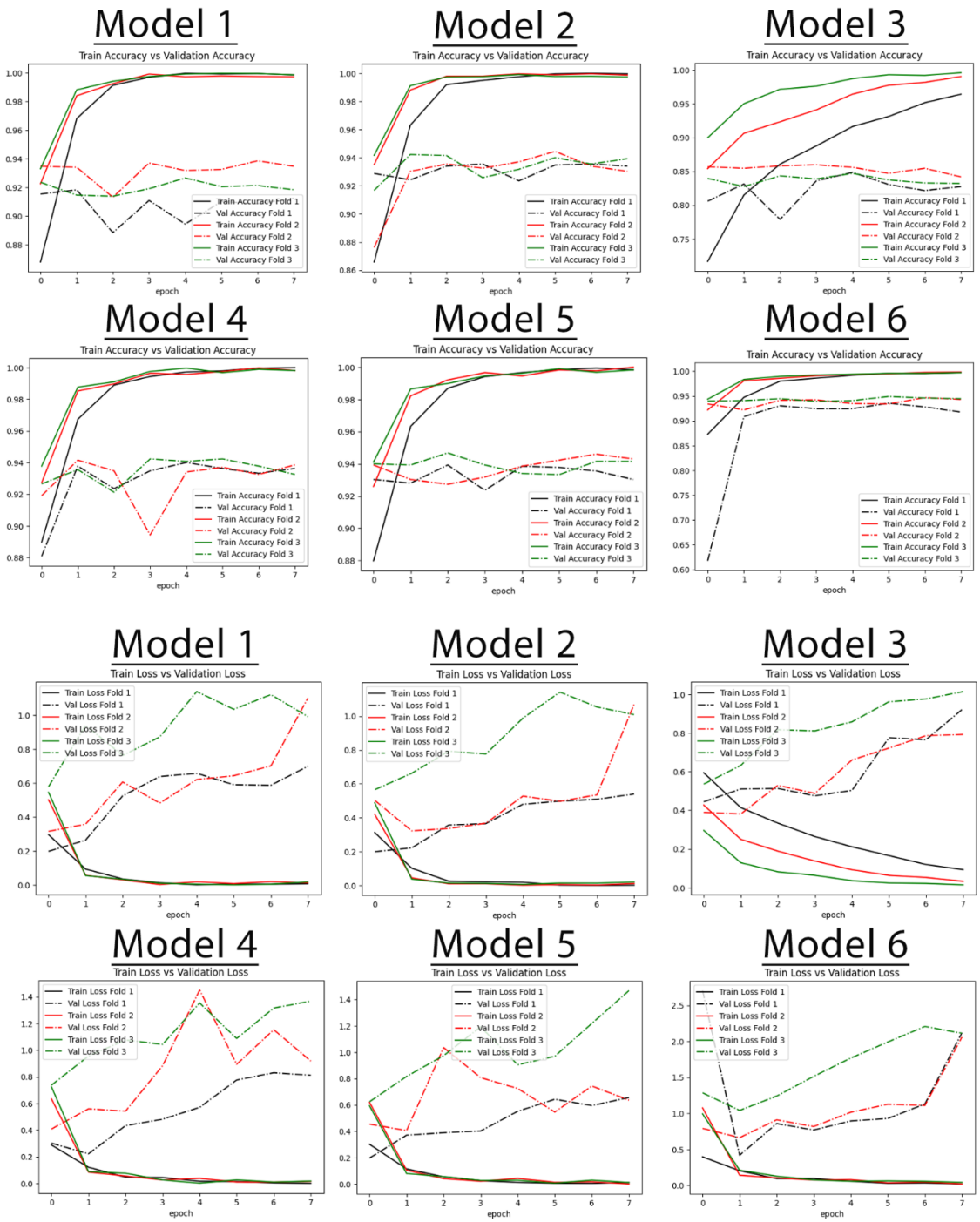
Πίνακας 5.1: Επιπλέον στρώματα και οι τιμές των υπερπαραμέτρων για την εκπαίδευση των CNN μοντέλων ένα έως έξι με σκοπό την ταξινόμηση κειμένου για όλα τα σύνολα δεδομένων



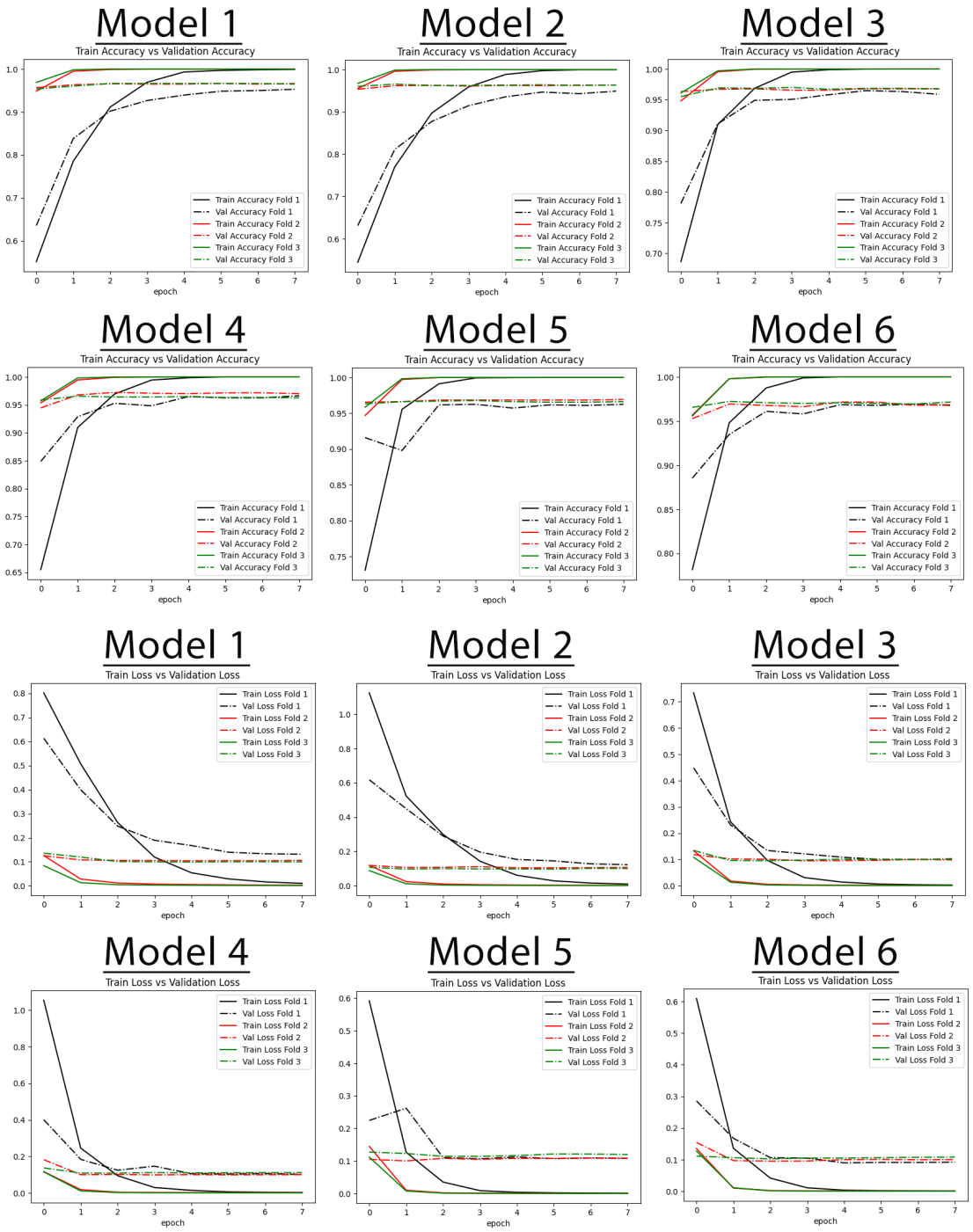
Εικόνα 5.6: Τρισδιάστατη οπτικοποίηση των μοντέλων CNN ένα έως έξι. Μέγιστη ανάλυση φωτογραφίας παρέχεται στον σύνδεσμο



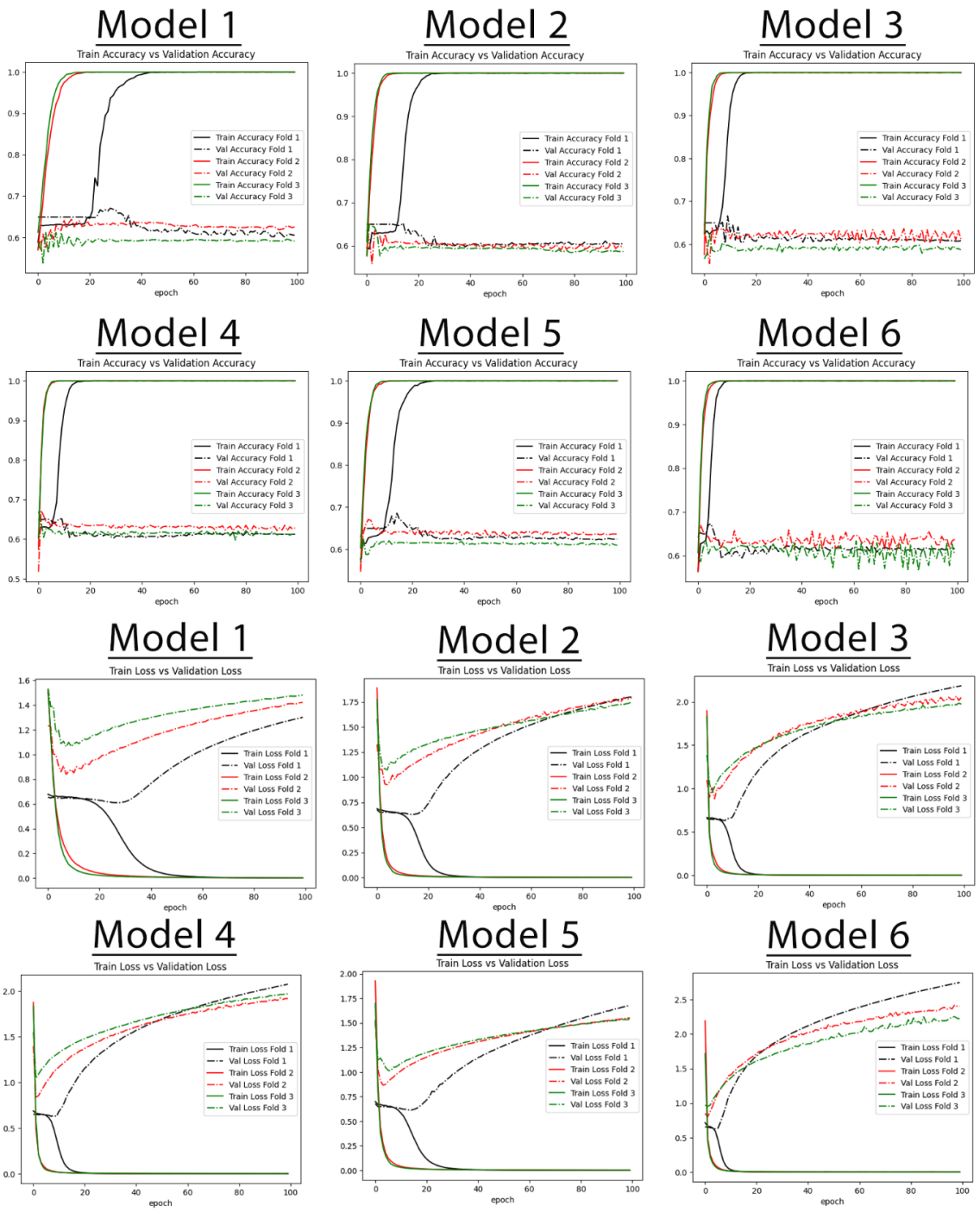
Εικόνα 5.7: Οπτικοποίηση των μοντέλων CNN ένα έως έξι. Μέγιστη ανάλυση φωτογραφίας παρέχεται στον σύνδεσμο



Εικόνα 5.8: Ακρίβεια και ακρίβεια επικύρωσης (πάνω μισό) και απώλεια και απώλεια επικύρωσης (κάτω μισό) κατά την διάρκεια της εκπαίδευσης των μοντέλων CNN ένα έως έξι για όλα τα epoch και για τις τρεις αναδιπλώσεις για το Fake News Corpus σύνολο δεδομένων. Μέγιστη ανάλυση φωτογραφίας παρέχεται στον σύνδεσμο



Εικόνα 5.9: Ακρίβεια και ακρίβεια επικύρωσης (πάνω μισό) και απώλεια και απώλεια επικύρωσης (κάτω μισό) κατά την διάρκεια της εκπαίδευσης των μοντέλων CNN ένα έως έξι για όλα τα epoch και για τις τρεις αναδιπλώσεις για το WELFake σύνολο δεδομένων. Μέγιστη ανάλυση φωτογραφίας παρέχεται στον σύνδεσμο



Εικόνα 5.10: Ακρίβεια και ακρίβεια επικύρωσης (πάνω μισό) και απώλεια και απώλεια επικύρωσης (κάτω μισό) κατά την διάρκεια της εκπαίδευσης των μοντέλων CNN ένα έως έξι για όλα τα epoch και για τις τρεις αναδιπλώσεις για το LIAR σύνολο δεδομένων. Μέγιστη ανάλυση φωτογραφίας παρέχεται στον σύνδεσμο

Model	Datasets Accuracy				Datasets f1-Score			Datasets Precision			Datasets Recall		
Convolutional Neural Network	Fake News Cropus	WELFake	LIAR	Overall Accuracy	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR
model_1 (Fake News Corpus) Fold 3	72%	51%	61%	**61.33%**	72%	50%	54%	74%	52%	55%	72%	51%	61%
model_1 (WELFake) Fold 1	46%	62%	63%	**57.00%**	45%	60%	51%	48%	64%	56%	46%	62%	63%
model_1 (LIAR) Fold 3	54%	53%	67%	**58.00%**	54%	53%	67%	54%	53%	67%	54%	53%	67%
model_2 (Fake News Corpus) Fold 3	75%	51%	59%	**61.67%**	75%	47%	55%	75%	52%	55%	75%	51%	59%
model_2 (WELFake) Fold 2	43%	62%	62%	**55.67%**	43%	61%	54%	43%	62%	56%	43%	62%	62%
model_2 (LIAR) Fold 2	55%	51%	65%	**57.00%**	54%	51%	64%	54%	52%	64%	55%	51%	65%
model_3 (Fake News Corpus) Fold 2	76%	50%	61%	**62.33%**	76%	47%	54%	76%	51%	55%	76%	50%	61%
model_3 (WELFake) Fold 1	45%	62%	62%	**56.33%**	45%	62%	53%	46%	63%	56%	45%	62%	62%
model_3 (LIAR) Fold 3	50%	53%	69%	**57.33%**	50%	52%	68%	52%	53%	68%	50%	53%	69%
model_4 (Fake News Corpus) Fold 3	70%	51%	60%	**60.33%**	70%	50%	56%	71%	52%	56%	70%	51%	60%
model_4 (WELFake) Fold 1	42%	61%	62%	**55.00%**	42%	60%	53%	43%	61%	55%	42%	61%	62%
model_4 (LIAR) Fold 2	58%	49%	66%	**57.67%**	54%	41%	65%	58%	50%	65%	58%	49%	66%
model_5 (Fake News Corpus) Fold 2	77%	52%	54%	**61.00%**	77%	48%	54%	77%	54%	54%	77%	52%	54%
model_5 (WELFake) Fold 1	45%	60%	62%	**55.67%**	45%	60%	53%	46%	60%	55%	45%	60%	62%
model_5 (LIAR) Fold 2	54%	53%	66%	**57.67%**	54%	53%	64%	54%	53%	64%	54%	53%	66%
model_6 (Fake News Corpus) Fold 1	68%	50%	61%	**59.67%**	67%	43%	54%	68%	52%	54%	68%	50%	61%
model_6 (WELFake) Fold 2	45%	60%	60%	**55.00%**	44%	59%	54%	46%	60%	54%	45%	60%	60%
model_6 (LIAR) Fold 1	56%	50%	63%	**56.33%**	49%	38%	60%	56%	58%	60%	56%	50%	63%

Πίνακας 5.2: Αναλυτικές ακρίβειες για τα μοντέλα CNN ένα έως έξι που υλοποιήθηκαν, προπονήθηκαν και χρησιμοποιήθηκαν για τις προβλέψεις όλων των συνόλων δεδομένων

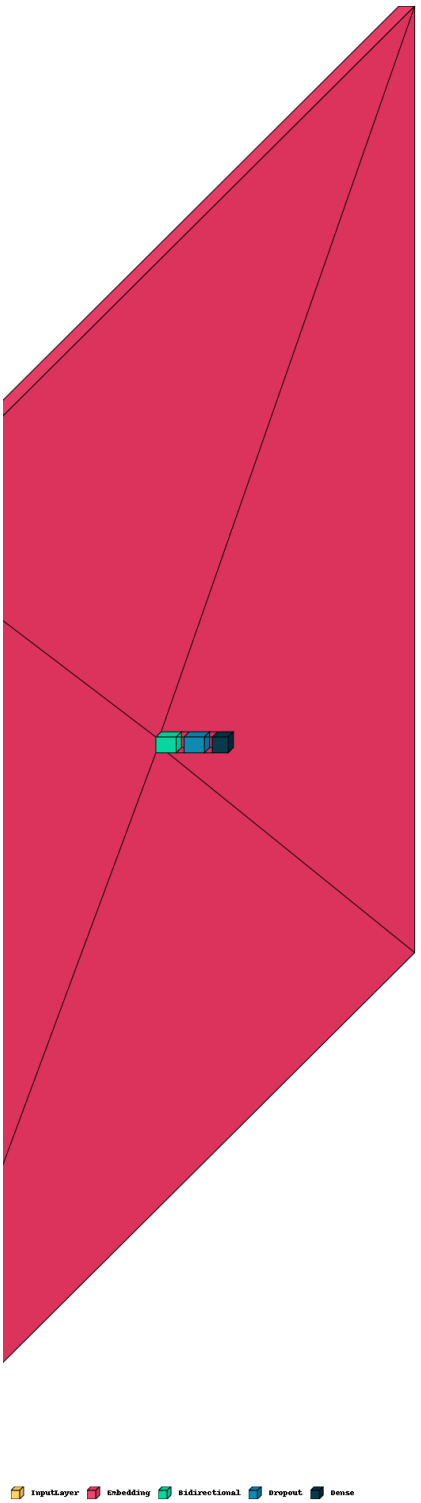
Όσων αφορά τα μοντέλα CNN παρατηρούμε ότι το μοντέλο ένα έχει την μεγαλύτερη μέση ακρίβεια με τιμή 58.78%. Το μοντέλο ένα παρέχει ακόμη και την μεγαλύτερη ξεχωριστή ακρίβεια για εκπαίδευση με το WELFake 57% και το LIAR 58% ενώ για το Fake News Corpus σύνολο δεδομένων το μοντέλο τρία δίνει την μεγαλύτερη ξεχωριστή ακρίβεια με τιμή 62.33%. Αντιληπτό γίνεται ακόμη το γεγονός ότι όταν γίνεται εκπαίδευση με ένα σύνολο δεδομένων και έπειτα πρόβλεψη στο ίδιο (αλλά μεγαλύτερης κλίμακας) σύνολο, η ακρίβεια δεν ξεπερνάει το 77% (μοντέλο πέντε). Αυτό, λαμβάνοντας υπόψιν ότι εκτός του LIAR συνόλου δεδομένων με τα υπόλοιπα δύο γίνεται εκπαίδευση με δέκα χιλιάδες στοιχεία και έπειτα γίνεται πρόβλεψη στα εκατό χιλιάδες (Fake News Corpus) και εβδομήντα δύο χιλιάδες (WELFake), το υποσύνολο είναι σχετικά αντιπροσωπευτικό ολόκληρου του συνόλου δεδομένων γιατί η κατανομή δεδομένων και των μοτίβων στο υποσύνολο ευθυγραμμίζονται καλά με αυτά στο πλήρες σύνολο δεδομένων με αποτέλεσμα το μοντέλο μπορεί να αποδώσει αρκετά καλά. Η πτώση της ακρίβειας στα σύνολα δεδομένων WELFake και LIAR υποδηλώνει ότι το μοντέλο μπορεί να μην γενικεύεται καλά σε σύνολα δεδομένων με διαφορετικά χαρακτηριστικά ή κατανομές δεδομένων και αυτό να οφείλεται στην υπερπροσαρμογή του υποσυνόλου που εκπαιδεύτηκε γιατί βλέπουμε από την εικόνα 5.8 ότι η απώλεια επικύρωσης αυξάνεται ανά epoch. Ένας άλλος λόγος θα μπορούσε να είναι ότι στο μοντέλο καταγράφονται χαρακτηριστικά ειδικά για το συγκεκριμένο σύνολο δεδομένων και δεν μεταφέρονται καλά στα υπόλοιπα.

(Υποκεφάλαιο 5.6) **LSTM μοντέλο & αποτελέσματα**

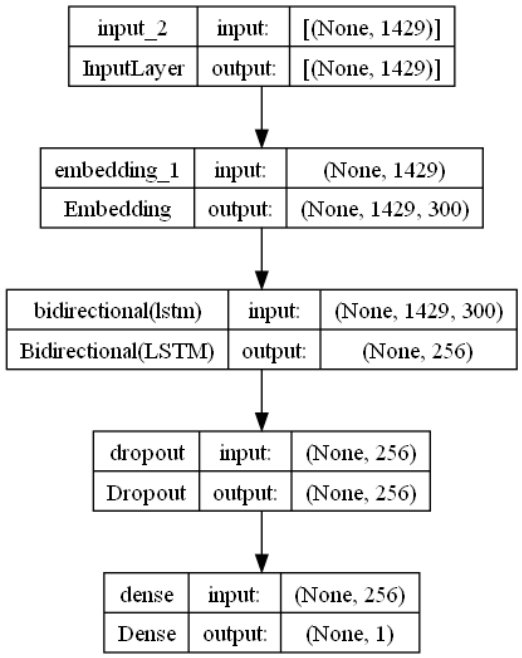
Όσων αφορά την νευρωνικά δίκτυα που χρησιμοποιούσαν LSTM αρχιτεκτονική επιλέξαμε από τις πιο απλούστερες μορφές. Στον πίνακα 5.3 παρουσιάζουμε την δομή του μοντέλου ενώ στις εικόνες 5.11, 5.12 παρατηρούμε την απεικόνιση του μοντέλου σε δύο μορφές. Φυσικά, παρέχουμε τις μετρικές στην εικόνα 5.13 για κάθε σύνολο δεδομένων και τελικά στον πίνακα 5.4 βλέπουμε την τελική ακρίβεια.

Models / Variables	Model 7
Total LSTM layers	1
units	128
Bidirectional	Yes
Fully Connected Dense Layers	No

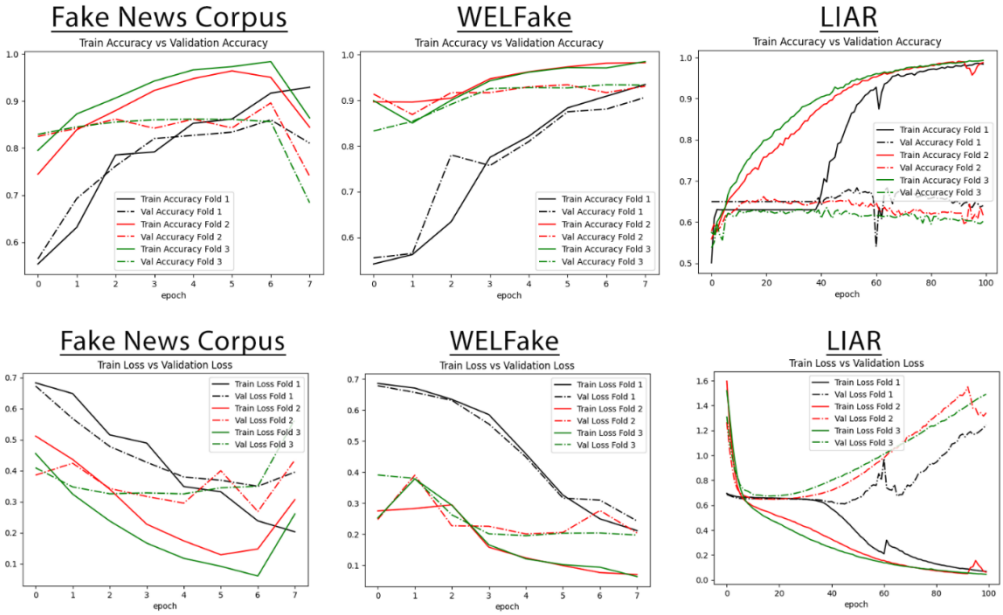
Πίνακας 5.3: Επιπλέον στρώματα και οι τιμές των υπερπαραμέτρων για την εκπαίδευση του LSTM μοντέλου με σκοπό την ταξινόμηση κειμένου για όλα τα σύνολα δεδομένων



Εικόνα 5.11: Τρισδιάστατη οπτικοποίηση του LSTM μοντέλου. Μέγιστη ανάλυση φωτογραφίας παρέχεται στον σύνδεσμο



Εικόνα 5.12: Οπτικοποίηση του LSTM μοντέλου



Εικόνα 5.13: Ακρίβεια και ακρίβεια επικύρωσης (πάνω μισό) και απώλεια και απώλεια επικύρωσης (κάτω μισό) κατά την διάρκεια της εκπαίδευσης του LSTM μοντέλου για όλα τα epoch και για τις τρεις αναδιπλώσεις και για τα τρία σύνολα δεδομένων. Μέγιστη ανάλυση φωτογραφίας παρέχεται στον σύνδεσμο

Model	Datasets Accuracy				Datasets f1-Score			Datasets Precision			Datasets Recall		
Recurrent Neural Network	Fake News Cropus	WELFake	LIAR	Overall Accuracy	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR
model_7 (Fake News Corpus) Fold 3	58%	57%	51%	**55.33%**	56%	56%	52%	57%	59%	53%	58%	57%	51%
model_7 (WELFake) Fold 2	53%	60%	61%	**58.00%**	53%	60%	55%	53%	60%	56%	53%	60%	61%
model_7 (LIAR) Fold 1	54%	53%	65%	**57.33%**	52%	49%	61%	53%	55%	63%	54%	53%	65%

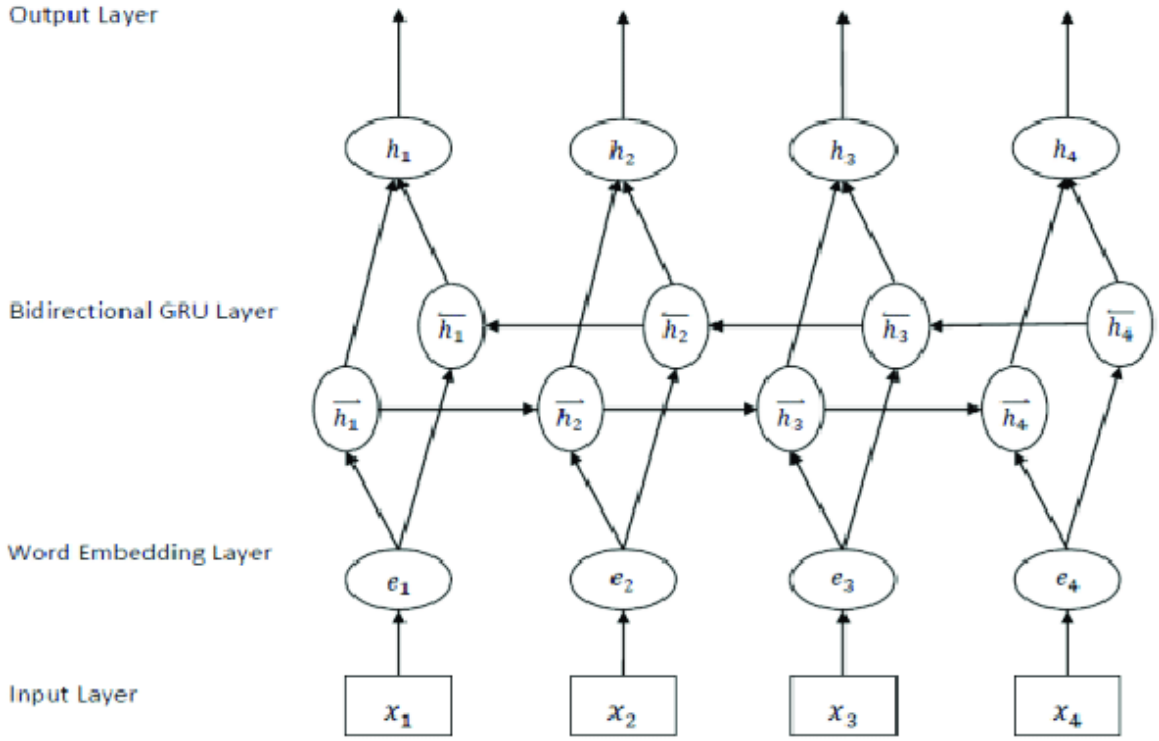
Πίνακας 5.4: Αναλυτικές ακρίβειες για το CNN μοντέλο που υλοποιήθηκε, προπονήθηκε και χρησιμοποιήθηκε για τις προβλέψεις όλων των συνόλων δεδομένων

Στα παραπάνω αποτελέσματα βλέπουμε πως το μοντέλο που χρησιμοποιεί LSTM έχει μέση απόδοση 56.89% πράγμα που μας εκπλήσσει διότι η αρχιτεκτονική του δίνει την δυνατότητα να καταγράφει εξαρτήσεις μεγάλης εμβέλειας επιτρέποντας το να κατανοεί τα συμφραζόμενα. Απροετοίμαστους μας βρήκε το γεγονός ότι η εκπαίδευση με το Fake News Corpus σύνολο

δεδομένων έδωσε τη χειρότερη ακρίβεια με τιμή 55.33% έναντι του LIAR συνόλου το οποίο έρχεται δεύτερο στην ακρίβεια με τιμή 57.33% η οποία είναι πολύ κοντά σε αυτή του WELFake συνόλου 58%. Βέβαια, σε όλες τις προβλέψεις μπορούμε να δούμε πως το συγκεκριμένο μοντέλο γενικεύεται σχεδόν μέτρια στα υπόλοιπα σύνολα δεδομένων. Ίσως η δημιουργία πιο περίπλοκων μοντέλων με διαφορετικό συνδυασμό στρωμάτων ή/και αύξηση των τιμών σε κάποιες υπερπαραμέτρους να μας παρείχε καλύτερα αποτελέσματα από τα συγκεκριμένα. Με σιγουριά μπορούμε να πούμε ότι δεν απορρίπτουμε σε καμία περίπτωση την αρχιτεκτονική του LSTM είτε υπονομεύουμε τις δυνατότητες του, απλώς ίσως να έτυχε ο τρόπος που προεπεξεργαστήκαμε τα σύνολα δεδομένων σε συνδυασμό με την απλή μορφή των στρωμάτων του μοντέλου και τον περιορισμένο αριθμό δεδομένων που τέθηκαν σε εκπαίδευση να μας έδωσε αυτά τα αποτελέσματα.

(Υποκεφάλαιο 5.7) **Υβριδικά μοντέλα & αποτελέσματα**

Για τα υβριδικά μοντέλα έχουμε υιοθετήσει την τεχνική της αμφίδρομης κατεύθυνσης (bidirectional) διότι η κατάσταση του μονοκατευθυντικού (unidirectional) GRU μεταδίδεται μονοκατευθυντικά από εμπρός προς τα πίσω, με άλλα λόγια, δεν μπορεί να λάβει υπόψη την επιρροή των ακόλουθων λέξεων και είναι εύκολο να αγνοηθεί η επιρροή των παρακάτω λέξεων. «Το αμφίδρομο (bidirectional) GRU είναι μια παραλλαγή της μονής κατεύθυνσης GRU, του οποίου η έξοδος εξαρτάται από τις διπλές επιδράσεις των καταστάσεων προς τα εμπρός και προς τα πίσω και έτσι λύνει το πρόβλημα της μονής κατεύθυνσης GRU, καθιστώντας την τελική έξοδο πιο ακριβή» [34]. Η εικόνα 5.14 [34] δείχνει την αρχιτεκτονική του GRU με αμφίδρομη κατεύθυνση.

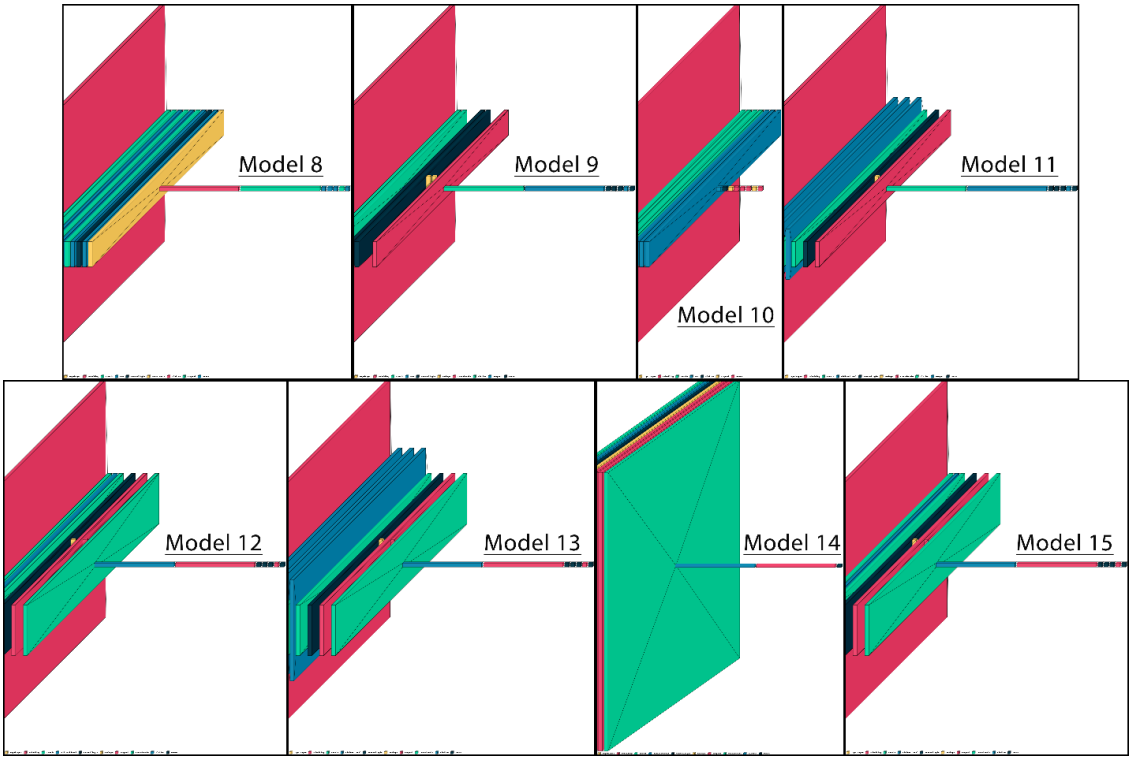


Εικόνα 5.14: Η αρχιτεκτονική ενός αμφίδρομου GRU

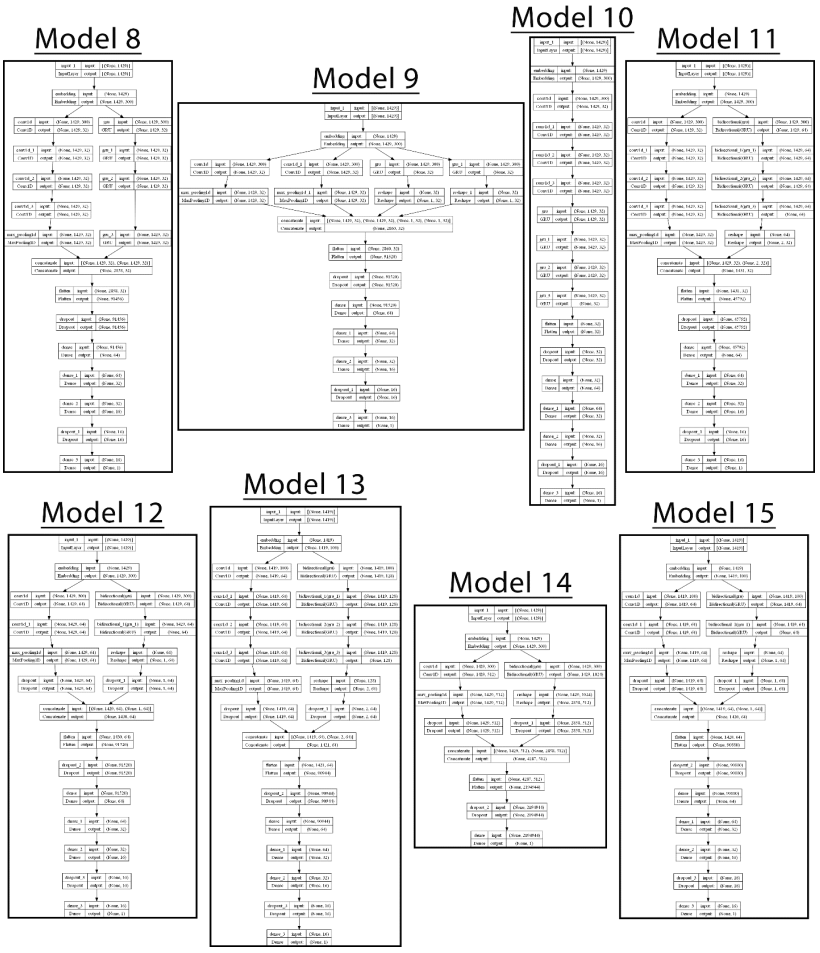
Όπως και με τα παραπάνω μοντέλα ο πίνακας 5.5 δείχνει την απαρίθμηση των μοντέλων καθώς και τις αλλαγές στις μεταβλητές αυτών. Στην εικόνα 5.15 και 5.16 παρουσιάζονται οι οπτικοποιήσεις ενώ στις εικόνες 5.17, 5.18 και 5.19 οι μετρικές για όλα τα σύνολα δεδομένων. Τελικά στον πίνακα 5.6 βλέπουμε τις τελικές ακρίβειες.

Models / Variables	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14	Model 15
Total convolutional layers	4	2	4	4	2	4	1	2
Total pooling layers	1	2	1	1	1	1	1	1
Total GRU layers	4	2	4	4	2	4	1	2
filters	32	32	32	32	64	64	512	64
kernel_size	[3]	[3,4]	[3]	[3]	[3]	[3]	[3]	[3,4]
units	32	32	32	32	32	64	512	32
Bidirectional	No	No	No	Yes	Yes	Yes	Yes	Yes
Fully Connected Dense Layers	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Learning Rate	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.00001

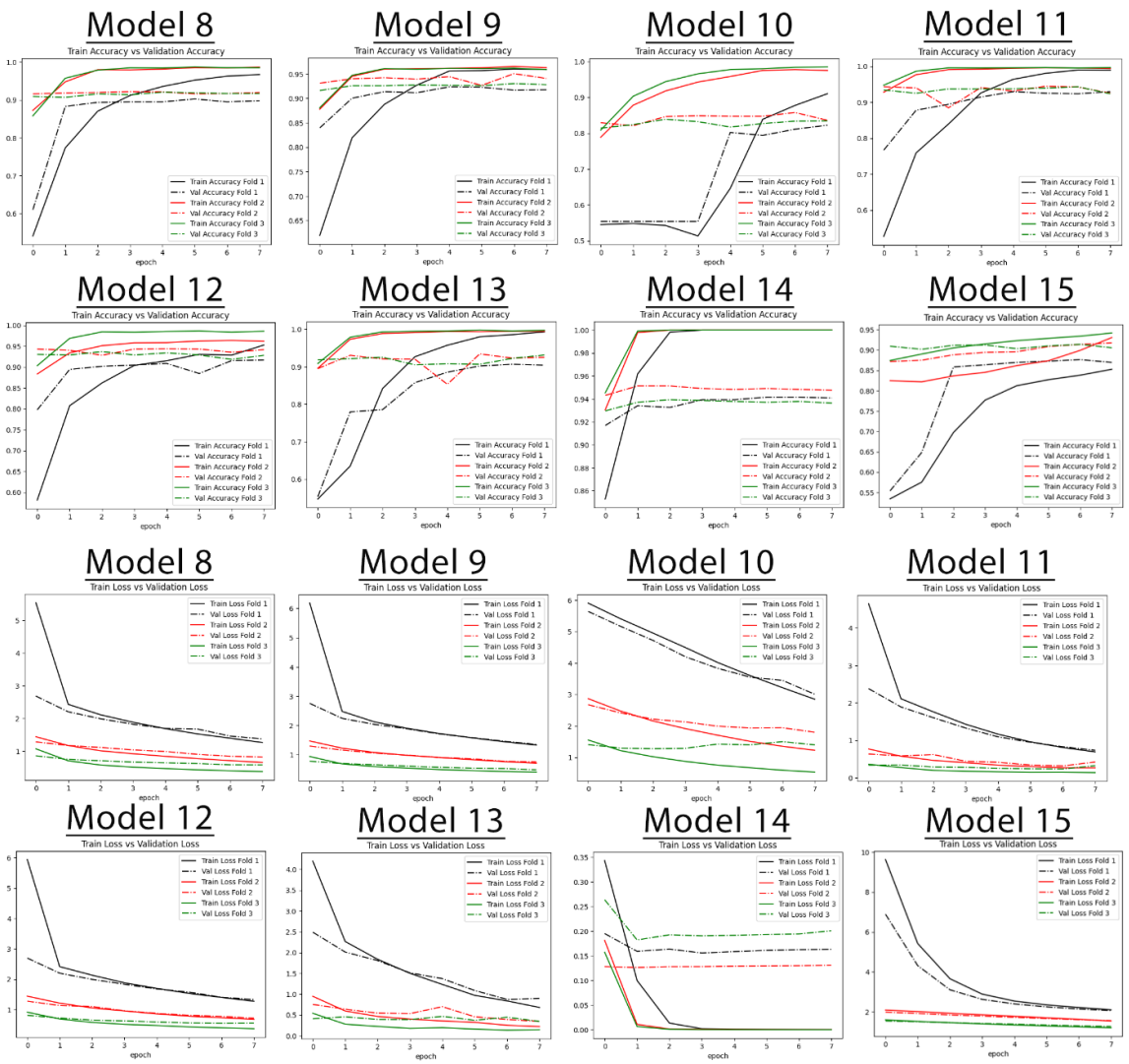
Πίνακας 5.5: Επιπλέον στρώματα και οι τιμές των υπερπαραμέτρων για την εκπαίδευση των υβριδικών μοντέλων οκτώ έως δεκαπέντε με σκοπό την ταξινόμηση κειμένου για όλα τα σύνολα δεδομένων



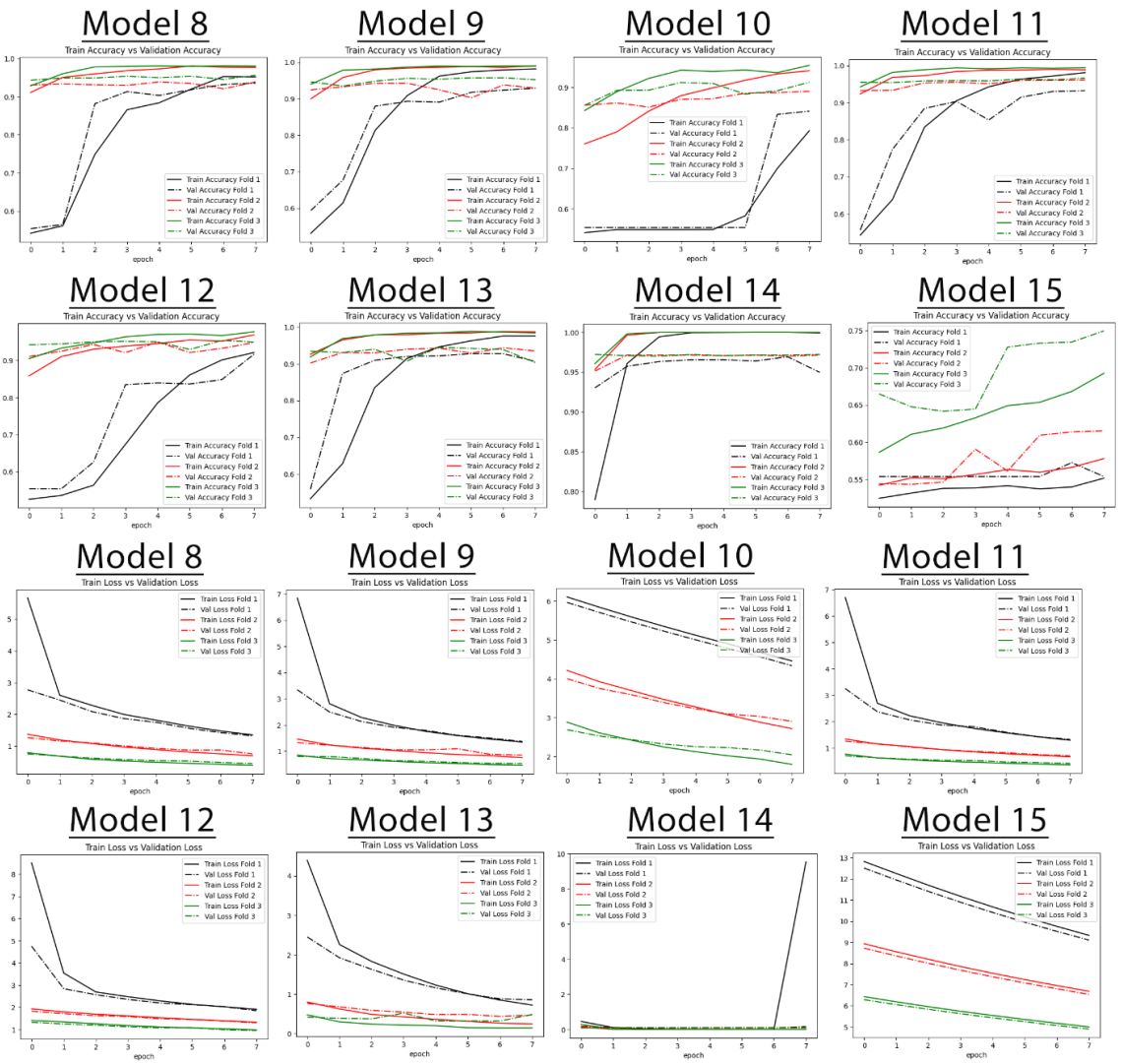
Εικόνα 5.15: Τρισδιάστατη οπτικοποίηση των υβριδικών μοντέλων οκτώ έως δεκαπέντε. Μέγιστη ανάλυση φωτογραφίας παρέχεται στον σύνδεσμο



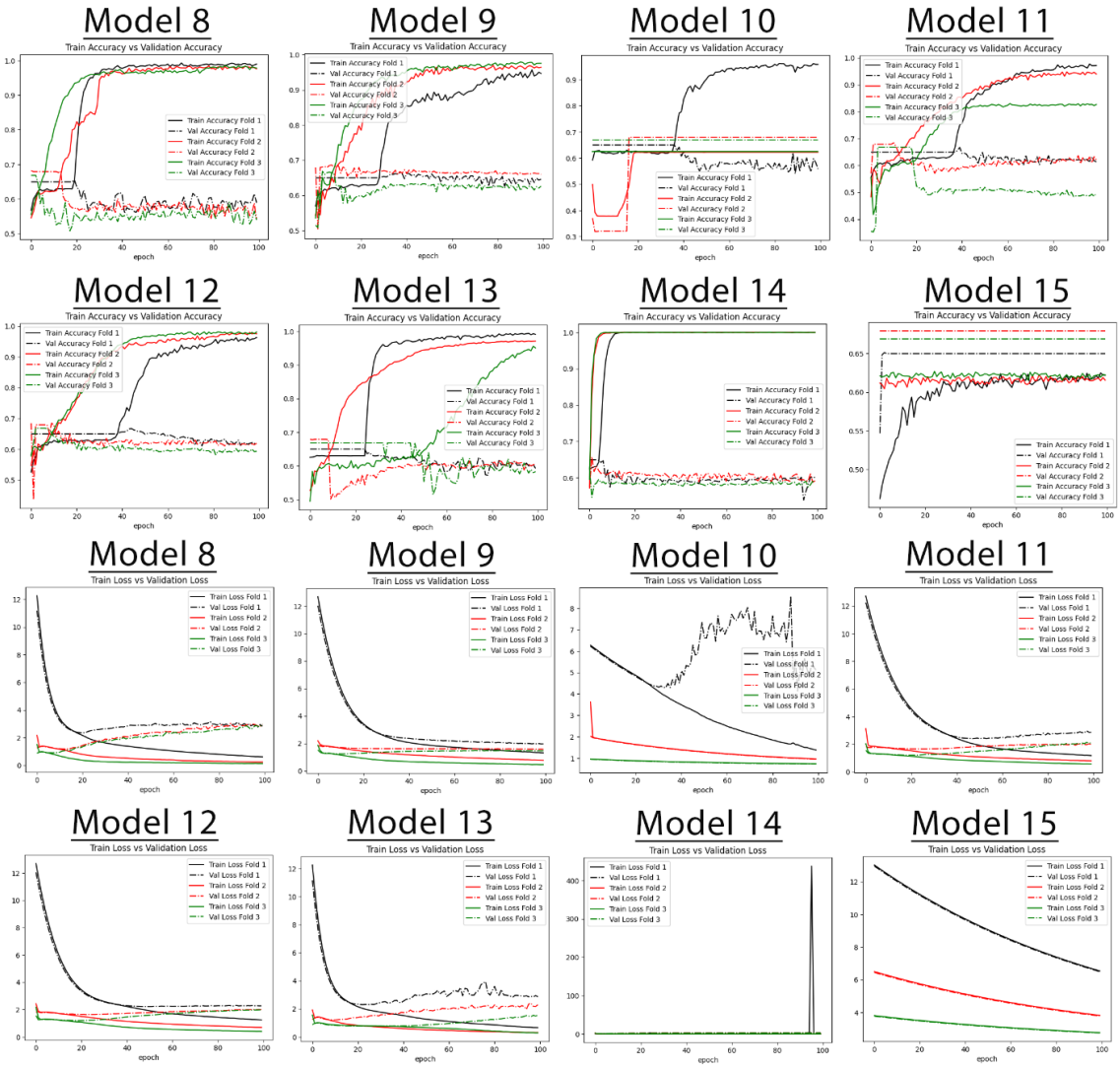
Εικόνα 5.16: Οπτικοποίηση των υβριδικών μοντέλων οκτώ έως δεκαπέντε. Μέγιστη ανάλυση φωτογραφίας παρέχεται στον σύνδεσμο



Εικόνα 5.17: Ακρίβεια και ακρίβεια επικύρωσης (πάνω μισό) και απώλεια και απώλεια επικύρωσης (κάτω μισό) κατά την διάρκεια της εκπαίδευσης των υβριδικών μοντέλων οκτώ έως δεκαπέντε για όλα τα epoch και για τις τρεις αναδιπλώσεις για το Fake News Corpus σύνολο δεδομένων. Μέγιστη ανάλυση φωτογραφίας παρέχεται στον σύνδεσμο



Εικόνα 5.18: Ακρίβεια και ακρίβεια επικύρωσης (πάνω μισό) και απώλεια και απώλεια επικύρωσης (κάτω μισό) κατά την διάρκεια της εκπαίδευσης των υβριδικών μοντέλων οκτώ έως δεκαπέντε για όλα τα epoch και για τις τρεις αναδιπλώσεις για το WELFake σύνολο δεδομένων. Μέγιστη ανάλυση φωτογραφίας παρέχεται στον σύνδεσμο



Εικόνα 5.19: Ακρίβεια και ακρίβεια επικύρωσης (πάνω μισό) και απώλεια και απώλεια επικύρωσης (κάτω μισό) κατά την διάρκεια της εκπαίδευσης των υβριδικών μοντέλων οκτώ έως δεκαπέντε για όλα τα epoch και για τις τρεις αναδιπλώσεις για το LIAR σύνολο δεδομένων. Μέγιστη ανάλυση φωτογραφίας παρέχεται στον σύνδεσμο

Model	Datasets Accuracy				Datasets f1-Score			Datasets Precision			Datasets Recall		
Hybrid Convolutional Neural Network & Gated Recurrent Unit	Fake News Cropus	WELFake	LIAR	Overall Accuracy	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR	Fake News Cropus	WELFake	LIAR
model_8 (Fake News Corpus) Fold 3	81%	52%	56%	**63.00%**	81%	43%	56%	82%	59%	56%	81%	52%	56%
model_8 (WELFake) Fold 1	54%	62%	60%	**58.67%**	54%	62%	55%	56%	62%	55%	54%	62%	60%
model_8 (LIAR) Fold 1	56%	48%	62%	**55.33%**	55%	43%	58%	56%	48%	59%	56%	48%	62%
model_9 (Fake News Corpus) Fold 1	78%	54%	61%	**64.33%**	78%	53%	52%	78%	55%	54%	78%	54%	61%
model_9 (WELFake) Fold 2	47%	58%	61%	**55.33%**	47%	58%	56%	46%	59%	56%	47%	58%	61%
model_9 (LIAR) Fold 3	44%	51%	67%	**54.00%**	30%	36%	65%	40%	49%	65%	44%	51%	67%
model_10 (Fake News Corpus) Fold 3	52%	53%	43%	**49.33%**	46%	50%	40%	48%	54%	54%	52%	53%	43%
model_10 (WELFake) Fold 3	52%	51%	57%	**53.33%**	48%	47%	55%	49%	53%	54%	52%	51%	57%
model_10 (LIAR)	50%	53%	57%	**53.33%**	50%	53%	58%	51%	53%	59%	50%	53%	57%
model_11 (Fake News Corpus) Fold 2	79%	52%	55%	**62.00%**	79%	49%	54%	79%	54%	53%	79%	52%	55%
model_11 (WELFake) Fold 3	47%	57%	58%	**54.00%**	47%	57%	55%	48%	57%	55%	47%	57%	58%
model_11 (LIAR) Fold 1	56%	49%	61%	**55.33%**	50%	38%	62%	56%	50%	62%	56%	49%	61%
model_12 (Fake News Corpus) Fold 1	83%	53%	60%	**65.33%**	83%	47%	54%	84%	57%	54%	83%	53%	60%
model_12 (WELFake) Fold 1	47%	61%	59%	**55.67%**	46%	60%	55%	49%	62%	54%	47%	61%	59%
model_12 (LIAR) Fold 1	45%	50%	66%	**53.67%**	29%	35%	62%	42%	34%	64%	45%	50%	66%
model_13 (Fake News Corpus) Fold 3	77%	51%	59%	**62.33%**	77%	49%	55%	77%	52%	54%	77%	51%	59%
model_13 (WELFake) Fold 1	55%	56%	56%	**55.67%**	54%	55%	55%	54%	56%	54%	55%	56%	56%
model_13 (LIAR) Fold 3	45%	51%	63%	**53.00%**	28%	35%	49%	20%	26%	40%	45%	51%	63%
model_14 (Fake News Corpus) Fold 1	81%	53%	61%	**65.00%**	81%	50%	54%	81%	56%	55%	81%	53%	61%
model_14 (WELFake) Fold 1	45%	61%	62%	**56.00%**	44%	60%	53%	46%	61%	56%	45%	61%	62%
model_14 (LIAR) Fold 2	56%	49%	65%	**56.67%**	51%	38%	64%	56%	49%	64%	56%	49%	65%
model_15 (Fake News Corpus) Fold 3	83%	51%	57%	**63.67%**	83%	46%	54%	83%	52%	53%	83%	51%	57%
model_15 (WELFake) Fold 3	50%	52%	52%	**51.33%**	50%	52%	53%	50%	52%	54%	50%	52%	52%
model_15 (LIAR) Fold 3	45%	51%	63%	**53.00%**	28%	35%	49%	20%	26%	40%	45%	51%	63%

Πίνακας 5.6: Αναλυτικές ακρίβειες για τα υβριδικά μοντέλα οκτώ έως δεκαπέντε που υλοποιήθηκαν, εκπαιδεύτηκαν και χρησιμοποιήθηκαν για τις προβλέψεις όλων των συνόλων δεδομένων

Στον παραπάνω πίνακα παρατηρούμε λοιπόν ότι την μεγαλύτερη ακρίβεια την έδωσε το μοντέλο δεκατέσσερα με μέση ακρίβεια 59.22%. Τα χειρότερα αποτελέσματα ανάμεσα στα οκτώ μοντέλα

που εξετάστηκαν τα έδωσε το μοντέλο δέκα με μέση ακρίβεια 52%. Παρόλα αυτά παρατηρούμε πως με εξαίρεση το μοντέλο δέκα η εκπαίδευση όλων των μοντέλων με το Fake News Corpus σύνολο δεδομένων μας δίνει τις μεγαλύτερες ακρίβειες με 65.33% ως την μεγαλύτερη προερχόμενη από το μοντέλο δώδεκα. Η εκπαίδευση με το ίδιο σύνολο δίνει την μεγαλύτερη ακρίβεια όταν προβλέπει το ίδιο αλλά μεγαλύτερο σύνολο αλλά δυσκολεύεται στην γενίκευση καθώς οι ακρίβειες μειώνονται προβλέποντας στην υπόλοιπα σύνολα. Αυτό βέβαια παρατηρείται εξίσου συχνά και στην εκπαίδευση και των δύο άλλων συνόλων καταλήγοντας στο συμπέρασμα ότι τα μοντέλα ίσως δεν προσαρμόζονται καλά στα χαρακτηριστικά των συνόλων δεδομένων. Γι' αυτό θα μπορούσε να ευθύνεται και ο φορέας της υπερπροσαρμογή, αν και στις μετρικές παρατηρείται ότι η απώλεια επικύρωσης συνήθως μειώνεται ανά epoch και για κάθε αναδίπλωση. Ένας άλλος λόγος θα μπορούσε να είναι οι διαφορές των συνόλων δεδομένων. Δηλαδή, κάθε σύνολο δεδομένων μπορεί να έχει μοναδικά χαρακτηριστικά, όπως λεξιλόγιο και κατανομή θεμάτων. Εάν αυτά τα μοντέλα προσαρμόζονται στα χαρακτηριστικά του συνόλου δεδομένων εκπαίδευσης, μπορεί να δυσκολεύονται να προσαρμοστούν στα διαφορετικά μοτίβα και χαρακτηριστικά των υπόλοιπων συνόλων δεδομένων.

ΚΕΦΑΛΑΙΟ 6 Συμπεράσματα

Στο τελικό κεφάλαιο αυτής της περιεκτικής διατριβής, ξεκινάμε το έργο της ενθυλάκωσης των πολύπλευρων διαστάσεων των ερευνητικών μας προσπαθειών, οι οποίες αφιερώθηκαν στον περίπλοκο τομέα της ανίχνευσης ψευδών ειδήσεων μέσω της χρήσης μιας ποικιλίας μοντέλων μηχανικής μάθησης. Αυτά τα μοντέλα, που κυμαίνονται από τα κλασικά παραδείγματα του Logistic Regression, του Passive Aggressive Classifier, του Random Forest, των Decision Trees, του Multinomial Naive Bayse και των Support Vector Machines έως το πιο σύγχρονο και διαφοροποιημένο BERT, FastText, καθώς και αρχιτεκτονικές νευρωνικών δικτύων όπως CNN, LSTM και οι υβριδικές παραλλαγές CNN+GRU, έχουν υποβληθεί σε εκπαίδευση και σχολαστική αξιολόγηση, καθεμία από τις οποίες εξετάζεται ανεξάρτητα εντός των ορίων των προεπεξεργασμένων συνόλων δεδομένων. Το πλήθος των συνόλων δεδομένων, που περιλαμβάνει τα Fake News Corpus, WELFake και LIAR έπαιξαν καθοριστικό ρόλο στην έρευνά μας. Είναι σημαντικά επειδή καλύπτουν διαφορετικούς τύπους περιεχομένου που σχετίζονται με ψεύτικες ειδήσεις και αποτέλεσαν τη βάση για την ανάλυσή μας.

Κατά τη διάρκεια των μετρήσεων της απόδοσης του κάθε μοντέλου, εμφανίζεται ένα ενδιαφέρον μοτίβο, όπου αυτά τα μοντέλα, όταν υποβάλλονται σε εργασίες πρόβλεψης, εκδηλώνουν (τις περισσότερες φορές) τα υψηλότερα ποσοστά ακρίβειας όταν ασχολούνται με το σύνολο δεδομένων στο οποίο εκπαιδεύτηκαν αρχικά. Ωστόσο, καθώς αντιμετωπίζουν την πρόκληση της γενίκευσης σε σύνολα δεδομένων αντίθετης προέλευσης, γίνεται εμφανής μια σταθερή μείωση της ακρίβειας. Ένας παράγοντας που πιστεύουμε ότι έπαιξε καθοριστικό ρόλο στην δυσκολία της γενίκευσης των μοντέλων είναι ο περιορισμένος αριθμός δειγμάτων που υποβλήθηκαν για την εκπαίδευση. Αυτός ο περιορισμός της χρήσης δέκα χιλιάδων (10.000) δειγμάτων για τα σύνολα δεδομένων Fake News Corpus και WELFake λόγω περιορισμένης υπολογιστικής δύναμης ήταν κάτι που φιλοδοξούμε στο μέλλον να μπορέσουμε να αντιμετωπίσουμε. Η φύση αυτού του προβλήματος της γενίκευσης μας καλεί να διερευνήσουμε βαθύτερα την προέλευσή του, αναφέροντας μερικούς ακόμη παράγοντες. Η προκατάληψη των δεδομένων, οι περίπλοκες αποχρώσεις της προσαρμογής του τομέα και η επιλογή της διαδικασίας με την οποία προεπεξεργαστήκαμε τα σύνολα δεδομένων έρχονται στο προσκήνιο για στοχασμό. Καθώς εμβαθύνουμε στο τι σημαίνουν τα ευρήματά μας, είναι σημαντικό να εξετάσουμε πώς θα μπορούσαν να εφαρμοστούν σε πραγματικές καταστάσεις στον εντοπισμό ψεύτικων ειδήσεων. Πρέπει να αναρωτηθούμε εάν αυτά τα αποτελέσματα έχουν σημασία πέρα από την έρευνά μας. Κοιτάζοντας το μέλλον, έχουμε μερικές πολύτιμες προτάσεις για έρευνα για την αντιμετώπιση του πολύπλοκου ζητήματος της εφαρμογής μοντέλων σε διαφορετικά σύνολα δεδομένων. Αυτές οι ιδέες περιλαμβάνουν τη δοκιμή νέων μεθόδων, τη δοκιμή μοντέλων νευρωνικών δικτύων με πιο σύνθετα ή απλά στρώματα, τη δοκιμή πολλαπλών εκδοχών του LSTM μοντέλου και την εξέταση των πλεονεκτημάτων του συνδυασμού διαφόρων προσεγγίσεων όπως οι μέθοδοι συνόλου (ensemble methods) και η μεταφορά μάθησης (transfer learning).

Αυτή η διατριβή υπογραμμίζει την περίπλοκη φύση του εντοπισμού ψευδών ειδήσεων καθώς δεν πρόκειται για κάτι ακλόνητο. Η παραγωγή ειδήσεων στο απεριόριστο παγκόσμιο ιστό αποτελεί μια αδιάκοπη και πολύπλευρη λειτουργία. Αυτός ο πολλαπλασιασμός των ειδήσεων έχει αυξήσει σημαντικά την πολυπλοκότητα του λεξιλογίου και την ποικιλομορφία των θεμάτων που καλύπτονται. Η τεράστια αυτή έκταση του διαδικτύου παρέχει ένα ανεξάντλητο ρεπερτόριο συνδυασμών λέξεων και γλωσσικών εκφράσεων, κάνοντας αδύνατον για τα σημερινά δεδομένα να αποθηκευτούν σε ένα μεμονωμένο κέντρο δεδομένων, να υποστούν επεξεργασία και να τροφοδοτηθούν σε ένα μοντέλο για εκπαίδευση. Με την συνέχεια της τεχνολογικής εξέλιξης αυτό θα είναι κάποια μέρα ένα τεράστιο επίτευγμα για την ανθρωπότητα αλλά πάντα υπάρχουν ανησυχίες για τυχόν κινδύνους. Παρόλα αυτά οι δημοσιογράφοι και οι δημιουργοί περιεχομένου καλούνται συνεχώς να χρησιμοποιήσουν αποτελεσματικά τη γλώσσα για να τραβήξουν την προσοχή του κοινού τους, να μεταφέρουν ακριβείς πληροφορίες και να προσελκύσουν αναγνώστες ή θεατές. Αυτή η γλωσσική ποικιλομορφία περιλαμβάνει τη χρήση διαφόρων στυλ γραφής, τόνων και τεχνικών για την εξυπηρέτηση διαφορετικών ακροατηρίων και τη μετάδοση του επιδιωκόμενου μηνύματος.

Επιπλέον, η παγκόσμια εμβέλεια του Διαδικτύου δημιουργεί ένα σχεδόν απεριόριστο φάσμα θεμάτων για εξερεύνηση. Ενώ οι παραδοσιακές κατηγορίες ειδήσεων, όπως ο αθλητισμός, η οικονομία και η πολιτική, εξακολουθούν να παραμένουν σταθεροί, συνοδεύονται τώρα από

αναδυόμενους τομείς που καλύπτουν φάσματα της τεχνολογίας και των κοινωνικών τάσεων. Καθώς συνεχίζουμε να ζούμε σε έναν όλο και πιο διασυνδεδεμένο κόσμο, το τοπίο ειδήσεων θα συνεχίσει να επεκτείνεται. Αυτό σημαίνει ότι θα προκύψουν νέα ζητήματα και ιστορίες, συμβάλλοντας στη διαρκώς αυξανόμενη πολυπλοκότητα του λεξιλογίου των ειδήσεων και στην ποικιλία των θεμάτων που καλύπτονται.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Statista, «Number of internet users worldwide from 2005 to 2022,» Νοέμβριος 2022. [Ηλεκτρονικό]. Available: https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/.

[2] E. &. T. M. Özkan, «An Examination on User Generated Content,» *Bogazici Journal,* pp. 27-51, Ιούλιος 2015.

[3] N. H. e. a. Al-Kumaim, «Exploring the Inescapable Suffering Among Postgraduate Researchers: Information Overload Perceptions and Implications for Future Research.,» *International Journal of Information and Communication Technology Education (IJICTE),* τόμ. 17, αρ. 1, pp. 19-41, 2021.

[4] J. Mansky, «Smithsonian Magazine,» 7 Μάϊος 2018. [Ηλεκτρονικό]. Available: https://www.smithsonianmag.com/history/age-old-problem-fake-news-180968945/.

[5] G. TV, «Science Disinformation: On the Problem of Fake News,» *Science and Technology of Information Processing,* τόμ. 48, αρ. 4, p. 290–298, 25 Φεβρουαρίου 2022.

[6] Statista, «Most popular platforms for daily news consumption in the United States as of August 2022, by age group,» Οκτώβριος 2022. [Ηλεκτρονικό]. Available: https://www.statista.com/statistics/717651/most-popular-news-platforms/.

[7] Z. E. M. M. e. a. Gordon Pennycook, «Shifting attention to accuracy can reduce misinformation online,» *Nature,* τόμ. 592, αρ. 7853, p. 590–595, 2021.

[8] D. M. M. T. K.P. Arin, «Ability of detecting and willingness to share fake news,» *Scientific Reports (Sci Rep),* τόμ. 13, αρ. 7298, 2023.

[9] E. Commission, *Fake news and disinformation online,* 2018.

[10] U. J. E. A. e. a. F. Olan, «Fake News on Social Media: the Impact on Society,» *Information Systems Frontiers (Inf Syst Front),* 2022.

[11] H. P. Tavishee Chauhan, «Optimization and improvement of fake news detection using deep learning approaches for societal benefit,» *International Journal of Information Management Data Insights,* τόμ. 1, αρ. 2, 2021.

[12] A. G. H. K. e. a. P. Akhtar, «Detecting fake news and disinformation using artificial intelligence and machine learning to avoid supply chain disruptions,» *Annals of Operations Research,* τόμ. 327, p. 633–657, 2023.

[13] D. B. S. Rastogi, «A review on fake news detection 3T's: typology, time of detection, taxonomies,» *International Journal of Information Security,* τόμ. 22, p. 177–212, 2023.

[14] M. N. B. A. P. S. N. F. B. M. N. Ihsan Ali, «Fake News Detection Techniques on Social Media: A Survey,» *Wireless Communications and Mobile Computing,* τόμ. 2022, 2022.

[15] N. O'Brien, *Machine learning for detection of fake news,* 2018.

[16] K. S. O. S. A. I. S.A. Khan, «Developing a Framework for Fake News Diffusion Control (FNDC) on Digital Media (DM): A Systematic Review 2010–2022,» *Sustainability,* τόμ. 14, p. 15287, 2022.

[17] M. Szpakowski, «Fake News Corpus,» 25 Ιανουαρίου 2020. [Ηλεκτρονικό]. Available: https://github.com/several27/FakeNewsCorpus.

[18] P. A. I. A. R. P. P. K. Verma, «WELFake: Word Embedding Over Linguistic Features for Fake News Detection,» *IEEE Transactions on Computational Social Systems,* τόμ. 8, αρ. 4, pp. 881-893, Αύγουστος 2021.

[19] W. Y. Wang, «"Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection,» *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers),* pp. 422-426, Ιούλιος 2017.

[20] M. Szpakowski, «Fake News Corpus README,» 2020. [Ηλεκτρονικό]. Available: https://github.com/several27/FakeNewsCorpus/blob/master/README.md.

[21] P. A. V. M. a. o. P. K. Verma, «MCred: Multi-modal Message Credibility for Fake News Detection Using BERT and CNN,» *Journal of Ambient Intelligence and Humanized Computing,* τόμ. 14, αρ. Not specified, p. 10617–10629, 2023.

[22] F. Chollet, Deep Learning with Python, Manning Publications, 2017.

[23] L. K. Saul, «An online passive-aggressive algorithm for difference-of-squares classification,» σε *NeurIPS 2021 (Neural Information Processing Systems)*, 2021.

[24] P. N. Kumar, «Detection of Textual Propaganda Using Passive Aggressive Classifiers,» *International Journal of Advanced Trends in Computer Science and Engineering,* τόμ. 12, Μάρτιος - Απρίλιος 2023.

[25] A. Chauhan, «Random Forest Classifier and its Hyperparameters,» 23 Φεβρουαρίου 2021. [Ηλεκτρονικό]. Available: https://medium.com/analytics-vidhya/random-forest-classifier-and-its-hyperparameters-8467bec755f6.

[26] A. Saini, «Guide on Support Vector Machine (SVM) Algorithm,» 12 Οκτωβρίου 2021. [Ηλεκτρονικό]. Available: https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/.

[27] N. S. N. P. J. U. L. J. A. N. G. L. K. I. P. Ashish Vaswani, *Attention Is All You Need,* 7 επιμ., 2023, p. 15.

[28] M. T. I. K. S. A. G. U. A. I. Junaed Younus Khan, «A benchmark study of machine learning models for online fake news detection,» *Machine Learning with Applications,* 2021.

[29] Z. I. M. A. e. a. M. Umer, «Impact of convolutional neural network and FastText embedding on text classification,» *Multimedia Tools and Applications,* τόμ. 82, p. 5569–5585, 2023.

[30] B. W. Ye Zhang, *A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification,* v4 επιμ., 2016.

[31] S. H. a. J. Schmidhuber, «Long Short-Term Memory,» *Neural Computation,* τόμ. 9, αρ. 8, pp. 1735-1780, 1997.

[32] C. Olah, «Understanding LSTM Networks,» 27 Αύγουστος 2015. [Ηλεκτρονικό]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/.

[33] Q. G. J. Z. Z. L. Z. Y. T. Y. H. Z. Lujuan Deng, «News Text Classification Method Based on the GRU_CNN Model,» *International Transactions on Electrical Energy Systems,* τόμ. 2022, p. 11, 2022.

[34] L. Y. G. X. G. Z. T. T. L. Z. Y. Cheng, «Text Sentiment Orientation Analysis Based on Multi-Channel CNN and Bidirectional GRU With Attention Mechanism,» *IEEE Access,* τόμ. 8, pp. 134964-134975, 2020.