



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Τεχνολογίας, Πληροφορικής & Υπολογιστών

Ανίχνευση Ψευδών Ειδήσεων με χρήση Νευρωνικών Δικτύων Γρά- φων και Τεχνικών Επεξεργασίας Φυσικής Γλώσσας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μιχάλης Κωνσταντίνου

Επιβλέπων: Γεώργιος Στάμου

Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2023



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Τεχνολογίας, Πληροφορικής & Υπολογιστών

Ανίχνευση Ψευδών Ειδήσεων με χρήση Νευρωνικών Δικτύων Γρά- φων και Τεχνικών Επεξεργασίας Φυσικής Γλώσσας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μιχάλης Κωνσταντίνου

Επιβλέπων: Γεώργιος Στάμου

Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 28^η Μαρτίου 2023.

.....
Γεώργιος Στάμου

Καθηγητής Ε.Μ.Π.

.....
Αθανάσιος Βουλόδημος

Επ. Καθηγητής Ε.Μ.Π.

.....
Ανδρέας-Γεώργιος Στα-
φυλοπάτης

Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2023

.....
ΜΙΧΑΛΗΣ ΚΩΝΣΤΑΝΤΙΝΟΥ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός
και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © – All rights reserved by Michalis Constantinou, February 2023.

Με επιφύλαξη παντός δικαιώματος.

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

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Η διάδοση ψευδών ειδήσεων αποτελεί ένα πρόβλημα, το οποίο ταλανίζει τον κόσμο εδώ και πολλά χρόνια. Με την αύξηση των μέσων διάδοσης και την πλέον εκτεταμένη χρήση των κοινωνικών δικτύων όπως του Facebook και του Twitter, ο ρυθμός διάδοσης των ψευδών ειδήσεων αυξήθηκε εκθετικά. Στην παρακάτω διπλωματική εργασία, αναλύεται η προϋπάρχουσα βιβλιογραφία πάνω στον εντοπισμό ψευδών ειδήσεων, η οποία περιλαμβάνει την μελέτη με βάση αποκλειστικά το περιεχόμενο των ειδήσεων, το περιεχόμενο των ειδήσεων με βάση τα συμφραζόμενα, τη διάδοση ειδήσεων και τη γνώση που περιέχουν οι ειδήσεις και γίνεται προσπάθεια μέσω ενός συνόλου δεδομένων, το οποίο είναι εξ ολοκλήρου στα ελληνικά, να εντοπιστεί η επιρροή της κάθε μεθόδου και να εξακριβωθεί ποια είναι η πιο αποδοτική. Αρχικά αναλύεται το θεωρητικό υπόβαθρο και παρατίθενται έννοιες, των οποίων η γνώση θεωρείται απαραίτητη όπως Μηχανική Μάθηση, ο ορισμός των Γράφων, των Νευρωνικών Δικτύων Γράφων και των Γράφων Γνώσης. Έπειτα αναλύεται το σύνολο δεδομένων το οποίο είναι καινοτόμο, ως προς το γεγονός ότι δεν υπάρχει άλλο ολοκληρωμένο σύνολο δεδομένων στην ελληνική γλώσσα που περιέχει επιβεβαιωμένες ψευδείς και αληθείς ειδήσεις και παρουσιάζεται η δομή του και το περιεχόμενό του. Ο εντοπισμός των ψευδών ειδήσεων, βασίζεται στο κείμενο που περιέχουν, με εξαίρεση την περίπτωση της διάδοσης, όπου χρησιμοποιείται και η πηγή του κάθε κειμένου. Στη συνέχεια, παρατίθενται αποτελέσματα από κάποιες baseline τεχνικές, οι οποίες μελετήθηκαν για να συγκριθούν με το βέλτιστο μοντέλο, οι οποίες περιλαμβάνουν τεχνικές διανυσμάτων λέξεων όπως FastText, Word2Vec, TF-IDF, GloVe, SpaCy, BERT. Εν συνεχεία, υλοποιήθηκαν πειράματα με BiGRU, τεχνικές δικτύων γράφων μέσω του Δικτύου Γράφων Προσοχής (Graph Attention Network - GAT) – το οποίο συγκρίθηκε με το Συνελκτικό Δίκτυο Γράφων (Graph Convolutional Network – GCN) - και συνδυάζονται οι πιο αποδοτικές τεχνικές, οι οποίες χρησιμοποιούνται, για την εξακρίβωση της απόδοσης με βάση τη γνώση. Εν συνεχεία παρουσιάζονται οι επιδόσεις του κάθε μοντέλου, γίνεται σύγκριση με σύνολα δεδομένων που χρησιμοποιήθηκαν εκτενώς στη βιβλιογραφία και συμπεραίνουμε ότι με βάση την ακρίβεια (accuracy) και το f1-score, το καλύτερο μοντέλο ανίχνευσης ειδήσεων χρησιμοποιεί τη BiGRU με ένα απλό νευρωνικό δίκτυο, το οποίο υπάγεται στη μέθοδο εντοπισμού ψευδών ειδήσεων σύμφωνα με το περιεχόμενο βασισμένο στα συμφραζόμενα και αποτελείται από επίπεδο διανυσμάτων, επίπεδο dropout για αποτροπή της υπερπροσαρμογής, επίπεδο BiGRU και επίπεδο dense για την έπειτα ταξινόμηση μέσω σιγμοειδούς συνάρτησης, με επίδοση που ανέρχεται στο 95%.

Λέξεις – κλειδιά: Μηχανική Μάθηση, Νευρωνικά Δίκτυα Γράφων, Γράφοι Γνώσης, Ανίχνευση Ψευδών Ειδήσεων, Εκπαίδευση Μοντέλου, Αξιολόγηση Μοντέλου, Ακρίβεια

Abstract

The spread of fake news is a problem which has plagued the world for a very long time. Nowadays, the extensive use of social networks such as Facebook and Twitter, has increased exponentially the fake news rate of spread. In the following thesis, the existing literature on fake news is analysed, which includes studies based exclusively on news content, based on news content according to the context, based on news propagation and knowledge based fake news detection and the impact of each method is identified on a custom dataset written exclusively in Greek, in order to get the most efficient method. Firstly, we analyse the theoretical background and other concepts, which are vital such as Machine Learning, the definition of Graphs, Graph Neural Networks and Knowledge Graphs. Then, the Greek dataset is presented, which is innovative in terms of the fact that there is no other dataset exclusively in Greek, which contains confirmed true and fake news. The detection of fake news is based on the text of each news article, except the case of news propagation where the source of each news article is also used. In addition, results from baseline techniques are listed, which were studied to be compared to the most optimal model, which include embedding techniques such as FastText, Word2Vec, TF-IDF, GloVe, SpaCy, BERT. Subsequently, we show the results of our own experiments which utilize BiGRU in terms of deep learning techniques, GAT (Graph Attention Network) in terms of propagation techniques – which is then compared to GCN (Convolutional Neural Network) – and the best performing models are combined, to be used in the examination of the knowledge-based performance. Afterwards, the performance of each model is compared with the performances on datasets which have been extensively used throughout literature and we conclude that based on accuracy and f1-score, the most effective fake news detection model utilizes BiGRU with a basic neural network, which is under context-based methods based on deep learning and consists of an embedding layer, a dropout layer to prevent overfitting, a BiGRU layer to form embeddings which include semantic information included in the text and a dense layer which is used to classify the news to fake or real, using a sigmoid function, with an f1-score upwards to 95%.

Keywords: Machine Learning, Graph Neural Networks, Knowledge Graphs, Fake News Detection, Model training, Model evaluation, Accuracy

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά την οικογένειά μου και όλους μου τους φίλους που με στήριξαν και πίστεψαν σε εμένα καθ' όλη την πάροδο των τελευταίων πέντε χρόνων. Επίσης θα ήθελα να ευχαριστήσω τον Γιώργο Μανωλιάδη, ο οποίος μέσω της βοήθειάς που προσέφερε, της συνεργασίας και τις καθημερινές συζητήσεις που είχαμε, συνέδραμε στην παρούσα διπλωματική εργασία.

Μιχάλης Κωνσταντίνου, Φεβρουάριος 2023

Περιεχόμενα

| | |
|-------------------------------------------------------------|--------|
| Περίληψη | - 7 - |
| Abstract | - 9 - |
| Ευχαριστίες | - 11 - |
| Πίνακας Περιεχομένων Εικόνων | - 16 - |
| 1. Εισαγωγή | - 20 - |
| 1.1. Ψευδείς Ειδήσεις | - 20 - |
| 1.1.1 Ορισμός ψευδών ειδήσεων | - 20 - |
| 1.1.2 Ελληνικά Μέσα Μαζικής Ενημέρωσης | - 21 - |
| 1.1.3 Παραδείγματα ψευδών ειδήσεων | - 23 - |
| 1.1.4 Κλασσική Μεθοδολογία εντοπισμού ψευδών ειδήσεων | - 26 - |
| 1.2. Κίνητρο Υλοποίησης | - 27 - |
| 2. Μηχανική Μάθηση | - 29 - |
| 2.1 Τύποι Εργασιών Μηχανικής Μάθησης | - 30 - |
| 2.2. Επεξήγηση Χρήσιμων Εννοιών | - 31 - |
| 2.2.1. Εξόρυξη Γνώσης Δεδομένων | - 31 - |
| 2.2.2. Εξόρυξη Γνώσης Κειμένου | - 32 - |
| 2.2.2.1. Συλλογή Δεδομένων | - 32 - |
| 2.2.2.2. Μετατροπή Κειμένου | - 33 - |
| 2.2.2.3. Κατακερματισμός | - 33 - |
| 2.2.2.4. Αφαίρεση Σημείων Στίξης | - 33 - |
| 2.2.2.5. Αφαίρεση Stopwords | - 33 - |
| 2.2.2.6. Λημματοποίηση - Ανακοπή | - 33 - |
| 2.2.2.7. Λεξικογραφική Ανάλυση | - 34 - |
| 2.2.3. Επεξεργασία Φυσικής Γλώσσας | - 34 - |
| 2.2.3.1. Αναπαράσταση Φυσικής Γλώσσας σε διανύσματα | - 34 - |
| 2.2.3.2. Προεπεξεργασία Δεδομένων | - 35 - |
| 2.2.4. Υπερπροσαρμογή και Υποπροσαρμογή Δεδομένων | - 36 - |
| 3. Γράφοι | - 38 - |
| 3.1. Ορισμοί | - 38 - |
| 3.2. Αναπαράσταση Δεδομένων Γράφου | - 38 - |
| 3.2.1 Πίνακας Γειτνίασης | - 38 - |

| | |
|-----------------------------------------------------------------------------------|--------|
| 3.2.2. Λίστα Γειτνίασης..... | - 39 - |
| 3.3 Χρήση Γράφων για αναπαράσταση δεδομένων..... | - 40 - |
| 4. Νευρωνικά Δίκτυα Γράφων..... | - 42 - |
| 4.1 Βασικές Έννοιες..... | - 42 - |
| 4.1.1. Βασικά Χαρακτηριστικά..... | - 42 - |
| 4.1.2. Είδη προβλέψεων μέσω Νευρωνικών Δικτύων Γράφων..... | - 42 - |
| 4.1.3. Χρήση Νευρωνικών Δικτύων Γράφων στην Μηχανική Μάθηση..... | - 42 - |
| 4.2. Τρόπος λειτουργίας των Νευρωνικών Δικτύων Γράφου..... | - 43 - |
| 4.2.1. Προβλέψεις μέσω Pooling..... | - 43 - |
| 4.2.2. Διάδοση μηνυμάτων..... | - 44 - |
| 4.2.3. Καθολικές Αναπαραστάσεις..... | - 44 - |
| 4.3. Διαφορετικοί τύποι γράφων..... | - 44 - |
| 4.4. Υποκατηγορίες Νευρωνικών Δικτύων Γράφου..... | - 45 - |
| 4.4.1. Νευρωνικό δίκτυο γράφων με ξεχωριστό MLP(Multilayer perceptron)..... | - 45 - |
| 4.4.2. Συνελκτικό Δίκτυο Γράφου..... | - 45 - |
| 4.4.3. Δίκτυο Γράφου Προσοχής..... | - 46 - |
| 4.5. Μεταγωγική – Επαγωγική Μάθηση..... | - 48 - |
| 5. Γράφοι Γνώσης..... | - 51 - |
| 5.1. Εισαγωγή..... | - 51 - |
| 5.2. Αρχιτεκτονική και κατασκευή ενός γράφου γνώσης..... | - 51 - |
| 5.2.1. Εξαγωγή γνώσης..... | - 52 - |
| 5.2.2. Συγχώνευση γνώσεων..... | - 53 - |
| 5.2.3. Αποθήκευση του γράφου..... | - 54 - |
| 5.2.4. Ανάκτηση και οπτικοποίηση γνώσης..... | - 54 - |
| 6. Ανασκόπηση Προϋπάρχουσας Βιβλιογραφίας..... | - 56 - |
| 6.1. Εντοπισμός Ψευδών Ειδήσεων μέσω περιεχομένου..... | - 56 - |
| 6.1.1. Μεθόδοι βασιζόμενοι αποκλειστικά στο περιεχόμενο..... | - 56 - |
| 6.1.2. Μεθόδοι βασιζόμενοι στο περιεχόμενο, λαμβάνοντας υπόψη τα συμφραζόμενα ... | - 59 - |
| 6.3. Εντοπισμός Ψευδών Ειδήσεων μέσω Δικτύου Διάδοσης..... | - 63 - |
| 6.4. Εντοπισμός Ψευδών Ειδήσεων με βάση τη Γνώση..... | - 66 - |
| 7. Πειράματα..... | - 75 - |
| 7.1. Συγκέντρωση Συνόλου Δεδομένων..... | - 75 - |

| | |
|----------------------------------------------------------------------------------|---------|
| 7.2. Μελέτη Συνόλου Δεδομένων | - 76 - |
| 7.3. Μετρικές Αξιολόγησης | - 78 - |
| 7.4. Διεκπεραίωση Πειραμάτων..... | - 79 - |
| 7.4.1. Πειράματα βασισμένα αποκλειστικά στο περιεχόμενο | - 80 - |
| 7.4.1.1. FastText..... | - 80 - |
| 7.4.1.2. Word2Vec | - 83 - |
| 7.4.1.3. TF-IDF | - 85 - |
| 7.4.1.4. GloVe..... | - 86 - |
| 7.4.1.5. Διανύσματα μέσω SpaCy..... | - 88 - |
| 7.4.2. Πειράματα βασισμένα στο περιεχόμενο, λαμβάνοντας υπόψη τα συμφραζόμενα .. | - 88 - |
| 7.4.2.1. BERT | - 88 - |
| 7.4.2.2. GRU-BiGRU..... | - 90 - |
| 7.4.3. Πειράματα βάσει Δικτύου Διάδοσης Ειδήσεων | - 96 - |
| 7.4.3.1. Δίκτυο Γράφων Προσοχής..... | - 96 - |
| 7.4.3.2. Συνελκτικό Δίκτυο Γράφων..... | - 99 - |
| 7.3.4. Πειράματα με βάση τη Γνώση..... | - 100 - |
| 7.3.4.1 Πείραμα βασισμένο στο DETERRENT | - 100 - |
| 8. Συμπεράσματα και Μελλοντικές Βελτιώσεις..... | - 111 - |
| 9. Βιβλιογραφία | - 116 - |

Πίνακας Περιεχομένων Εικόνων

| | |
|----------------------------------------------------------------------------------|--------|
| Εικόνα 1: Ποσοστό που εμπιστεύεται τις ειδήσεις σε κάθε χώρα..... | - 22 - |
| Εικόνα 2: Αίσθηση ανεξαρτησίας ΜΜΕ από αθέμιτες πολιτικές επιρροές ανά χώρα..... | - 22 - |
| Εικόνα 3: Αίσθηση ανεξαρτησίας ΜΜΕ από αθέμιτες επιχειρηματικές επιρροές | - 23 - |
| Εικόνα 4: Διάφοροι τρόποι μάθησης | - 30 - |
| Εικόνα 5: Συνθήκη συμπλήρωσης πίνακα γειτνίασης..... | - 38 - |
| Εικόνα 6: Γράφος Αναπαράστασης Δεδομένων..... | - 39 - |
| Εικόνα 7: Πίνακας Γειτνίασης..... | - 39 - |
| Εικόνα 8: Λίστα Γειτνίασης..... | - 40 - |
| Εικόνα 9: Μετατροπή κειμένου σε Πίνακα Γειτνίασης..... | - 40 - |
| Εικόνα 10: Χαρακτηριστικά εξόδου του κόμβου i..... | - 46 - |
| Εικόνα 11: Μη κανονικοποιημένοι συντελεστές..... | - 46 - |
| Εικόνα 12: Κανονικοποιημένοι συντελεστές | - 47 - |
| Εικόνα 13: Χαρακτηριστικά εξόδου..... | - 47 - |
| Εικόνα 14: Single-head attention(δεξιά) και Multi-head attention(αριστερά) | - 47 - |
| Εικόνα 15: Αρχιτεκτονική ενός γράφου γνώσης | - 52 - |
| Εικόνα 16: Πίνακας Βιβλιογραφίας..... | - 71 - |
| Εικόνα 17: Απεικόνιση συνόλου δεδομένων..... | - 76 - |
| Εικόνα 18: Top10 πηγών ψευδών ειδήσεων..... | - 76 - |
| Εικόνα 19: Ποσοστό ψευδών – αληθινών ειδήσεων | - 77 - |
| Εικόνα 20: Σύννεφο συχνότερα εμφανιζόμενων λέξεων ψευδών και αληθινών ειδήσεων - | 77 - |
| Εικόνα 21: Μέσος όρος λέξεων σε αληθινές και ψευδείς ειδήσεις..... | - 78 - |
| Εικόνα 22: Ορισμός Παραμέτρων Μετρικών Αξιολόγησης | - 78 - |
| Εικόνα 23: Ορισμός Μετρικών Αξιολόγησης | - 79 - |
| Εικόνα 24: Σχηματική Αναπαράσταση των μοντέλων CBOW και SKIPGRAM..... | - 80 - |
| Εικόνα 25: Κώδικας μορφοποίησης του συνόλου δεδομένων | - 81 - |
| Εικόνα 26: Απεικόνιση του μοντέλου εκπαίδευσης | - 82 - |
| Εικόνα 27: Συνάρτηση Πειράματος..... | - 82 - |
| Εικόνα 28: Έκθεση ταξινόμησης μέσω FastText | - 83 - |
| Εικόνα 29: Εξίσωση Πιθανότητας Εντοπισμού λέξης..... | - 83 - |
| Εικόνα 30: Μοντέλο Word2Vec..... | - 84 - |
| Εικόνα 31: Παρόμοιες λέξεις, με τη λέξη εισόδου | - 84 - |

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| Εικόνα 32: Έκθεση ταξινόμησης μέσω word2vec..... | - 85 - |
| Εικόνα 33: Εξίσωση Υπολογισμού Inverse Document Frequency | - 85 - |
| Εικόνα 34: Έκθεση ταξινόμησης μέσω TF-IDF..... | - 86 - |
| Εικόνα 35: Εξίσωση Συχνότητας Εμφάνισης κάθε ζεύγους λέξεων στο σώμα κειμένου .. | - 87 - |
| Εικόνα 36: Έκθεση ταξινόμησης μέσω GloVe..... | - 87 - |
| Εικόνα 37: Έκθεση ταξινόμησης μέσω διανυσμάτων Spacy | - 88 - |
| Εικόνα 38: Προ-εκπαίδευση και fine-tuning του BERT | - 89 - |
| Εικόνα 39: Έκθεση ταξινόμησης μέσω BERT multi-cased L-12 H-768 | - 90 - |
| Εικόνα 40: Αρχιτεκτονική RNN..... | - 92 - |
| Εικόνα 41: Αρχιτεκτονική LSTM..... | - 94 - |
| Εικόνα 42: Αρχιτεκτονική GRU..... | - 95 - |
| Εικόνα 43: Αποτελέσματα ταξινόμησης μέσω BiGRU | - 95 - |
| Εικόνα 44: Ιστόγραμμα συχνότητας εμφάνισης λέξεων | - 97 - |
| Εικόνα 45: Χαρακτηριστικά του Pytorch Geometric συνόλου δεδομένων..... | - 97 - |
| Εικόνα 46: Συνάρτηση Δικτύου Γράφων Προσοχής..... | - 98 - |
| Εικόνα 47: Υλοποίηση συνάρτησης Συνελκτικού Δικτύου Γράφων | - 99 - |
| Εικόνα 48 : Πιθανές οντότητες, με το μέρος που είναι στην πρόταση, την λέξη από την οποία εξαρτώνται και το μέρος του λόγου που αποτελεί αυτή η λέξη | - 101 - |
| Εικόνα 49: Κείμενο με τα αντίστοιχα του triplets και ετικέτα | - 101 - |
| Εικόνα 50: Πίνακας triplets | - 102 - |
| Εικόνα 51: Ενημέρωση κόμβων σε Σχεσιακό – Συνελκτικό Δίκτυο Γράφων | - 102 - |
| Εικόνα 52: Ενημέρωση κόμβων σε Σχεσιακό –Δίκτυο Γράφων Προσοχής | - 102 - |
| Εικόνα 53: Μηχανισμός Προσοχής | - 103 - |
| Εικόνα 54: Υπολογισμός διανύσματος του άρθρου | - 103 - |
| Εικόνα 55: Σημασία της εκάστοτε λέξης στο άρθρο | - 103 - |
| Εικόνα 56: Υπολογισμός διανύσματος κάθε άρθρου | - 104 - |
| Εικόνα 57: Χρήση γνώσης σε συνδυασμό με τις πληροφορίες κειμένου | - 104 - |
| Εικόνα 58: Μείωση του Binary Cross Entropy Loss..... | - 104 - |
| Εικόνα 59: Στατιστικά Συνόλων Δεδομένων..... | - 106 - |
| Εικόνα 60: Baseline υλοποιήσεις ως προς το περιεχόμενο σε άλλα σύνολα δεδομένων . | - 106 - |
| Εικόνα 61: : Baseline υλοποιήσεις με τα συμφραζόμενα σε άλλα σύνολα δεδομένων ... | - 107 - |
| Εικόνα 62: Σύγκριση απόδοσης BiGRU, με το βέλτιστο μοντέλο OPCNN..... | - 107 - |
| Εικόνα 63: Απόδοση συνόλων δεδομένων με την υλοποίηση με χρήση GAT | - 108 - |

Εικόνα 64: Απόδοση συνόλων δεδομένων υλοποίησης με βάση το DETERRENT - 108 -
Εικόνα 65: Απόδοση baseline μοντέλων και δικών μας υλοποιήσεων στο ελληνικό σύνολο
δεδομένων - 111 -

1. Εισαγωγή

1.1. Ψευδείς Ειδήσεις

1.1.1 Ορισμός ψευδών ειδήσεων

Ο άνθρωπος σαν σκεπτόμενο και έλλογο ον, σκέπτεται, μαθαίνει και διαμορφώνει τις απόψεις του, με βάση το κοινωνικό πλαίσιο στο οποίο υπάγεται. Ο κατακλυσμός των δεδομένων και των πληροφοριών, με τον οποίο βομβαρδίζεται καθημερινά, τον επηρεάζει άμεσα ως προς τη διαμόρφωση των απόψεών του, είτε αυτές οι πληροφορίες προέρχονται από τον περίγυρό του, είτε από ανθρώπους που εμπιστεύεται ή ακόμα και από πηγές πληροφόρησης τις οποίες θεωρεί αξιόπιστες. Η εκτεταμένη μετάδοση της γνώσης, ιδιαίτερα μετά την εφεύρεση του τύπου και των κοινωνικών μέσων μαζικής δικτύωσης, έχει μετατρέψει την διαδικασία ενημέρωσης, σε ένα όπλο χειραγώγησης του μυαλού του κάθε ανθρώπου. Πίσω από την λανθασμένη ενημέρωση – είτε θελημένα είτε ακούσια - ελλοχεύουν αρκετοί κίνδυνοι που απειλούν την ακεραιότητα του τρόπου σκέψης του κάθε ανθρώπου, έχοντας ως αποτέλεσμα τη διαμόρφωση στοχευμένων απόψεων, που πρόσκεινται στο εκάστοτε μέσο, το οποίο τις διαδίδει.

Οι ψευδείς ειδήσεις και η μετάδοσή τους είναι ένα πάγιο πρόβλημα, το οποίο οξύνθηκε, παράλληλα με την πρόοδο του ανθρώπου ως προς τον τομέα της ενημέρωσης. Η ευρύτερη διασπορά των ψευδών ειδήσεων, οφείλεται στις πλατφόρμες κοινωνικής δικτύωσης, καθώς και στην εκτενέστερη χρήση του διαδικτύου και μπορούν να έχουν αρνητικές επιρροές σε διάφορους τομείς της καθημερινότητας όπως στην πολιτική, σε θέματα υγείας και στην οικονομία. Από τη στιγμή που η προσέλκυση του ενδιαφέροντος του κάθε πολίτη έχει γίνει αυτοσκοπός, η ανάγκη σωστής ενημέρωσης έχει περάσει σε δεύτερη μοίρα και τα μέσα μαζικής ενημέρωσης έχουν καταφύγει σε θεμιτούς και αθέμιτους τρόπους, με σκοπό την αύξηση της επισκεψιμότητας στους ιστοτόπους τους. Όσο πιο ακραίος είναι ο τίτλος μιας είδησης, τόσο περισσότερους ανθρώπους θα έχουν την δυνατότητα να προσελκύσουν και κατ' επέκταση όσο περισσότερες επισκέψεις έχει μια ιστοσελίδα, εξυπηρετείται ο σκοπός της αποκόμισης εσόδων μέσω των διαφημίσεων τους. Έτσι, τίθεται σαν πρωταρχική η ανάγκη εντυπωσιασμού και παραγκωνίζεται το καθήκον της ενημέρωσης.

Το 2017, η Claire Wardle υποστήριξε ότι όρος «ψευδείς ειδήσεις» είναι λανθασμένος και χρησιμοποιείται για να περιγράψει θέματα, στα οποία δεν αρμόζει. Το κύριο θέμα που έπρεπε να απασχολεί τους πάντες είναι η ρίζα προέλευσης των ειδήσεων, η οποία αποτελείται από την ενημέρωση. Τα ψευδή γεγονότα είναι προϊόντα κακής ενημέρωσης, πράγμα το οποίο καθιστά προφανή την ανάγκη βελτίωσης της ποιότητας ενημέρωσης. Υπάρχουν δύο τύποι λανθασμένης πληροφόρησης: Η επιτηδευμένη παραπληροφόρηση και η κακή πληροφόρηση, οι οποίες διαφέρουν ως προς την πρόθεση του αρθρογράφου ή του εκάστοτε μέσου ενημέρωσης. Η επιτηδευμένη παραπληροφόρηση, χωρίζεται σε δύο τύπους: στη διάδοση κατασκευασμένων πληροφοριών με σκοπό την στοχοποίηση προσώπων και πρόκληση βλαβών και στη διάδοση φαινομενικά αληθών ειδήσεων, οι οποίες όμως περνάνε από μια διαδικασία επεξεργασίας με σκοπό και πάλι την στοχοποίηση. Από την άλλη, στην κακή πληροφόρηση, η ψευδής πληροφορία προέκυψε ακούσια, χωρίς να υπάρχει πρόθεση εξαπάτησης ή παραπληροφόρησης του

κοινού, έχοντας όμως διαδοθεί σαν αληθής [1]. Γενικότερα, η οποιαδήποτε μορφή παραπληροφόρησης, κρίνεται ζημιογόνα και θέτει εμπόδια στην σωστή ενημέρωση και στο έργο των δημοσιογράφων που προσπαθούν να διαδώσουν την αλήθεια.

Στην πιο πρόσφατη βιβλιογραφία, τα ψευδή γεγονότα διαχωρίζονται σε επτά τύπους [2]: Τις ειδήσεις σάτιρας, οι οποίες έχουν σκοπό να διακωμωδήσουν καταστάσεις με χιουμοριστικό ύφος κείμενο, αλλά λόγω δυσκολίας εντοπισμού του σαρκασμού στον γραπτό λόγο, άθελα τους παραπλανούν τον αναγνώστη. Έπειτα, υπάρχουν οι ειδήσεις γνωστές και σαν «clickbaits», όπου σκοπό έχουν τον εντυπωσιασμό του αναγνώστη με ακραίους τίτλους, που θέλουν να τραβήξουν την προσοχή, χωρίς να τους απασχολεί το περιεχόμενο της εκάστοτε είδησης, ακόμα και αν ο τίτλος δεν συνάδει με τα γραφόμενα. Στη συνέχεια, υπάρχουν οι ειδήσεις παραπληροφόρησης, οι οποίες στοχοποιούν άτομα και καταστάσεις και χρησιμοποιούν επιλεκτικά πληροφορίες, με σκοπό την επίθεση ή και να βλάψουν. Μετά, υπάρχουν οι ειδήσεις λανθασμένου περιεχομένου, όπου έχουν σαν στόχο την ανάμειξη γεγονότων με ψευδείς πληροφορίες, με σκοπό της ενίσχυσης της θέσης που θέλουν να προβάλλουν και να παραποιήσουν την αλήθεια κατά το συμφέρον τους. Ακόμα, υπάρχουν οι απάτες, στις οποίες γίνεται κλήση σε ανυπόστατες πληροφορίες και ανύπαρκτες πηγές, έχοντας σαν σκοπό να δοθεί μια λανθασμένη εντύπωση εγκυρότητας και ο αναγνώστης να παραπλανηθεί ευκολότερα. Υπάρχουν και οι ειδήσεις με σκοπό να χειραγωγήσουν το κοινό, χρησιμοποιώντας παραποιημένες ειδήσεις και δεδομένα, με σκοπό να την διάδοση ιδεών, οι οποίες βασίζονται σε γεγονότα, τα οποία όμως είναι στοχευμένα ψευδή. Τέλος, υπάρχουν και οι κατασκευασμένες ειδήσεις, οι οποίες είναι απόλυτα ψευδής και σαν σκοπό έχουν να εξαπατήσουν και να κάνουν κακό.

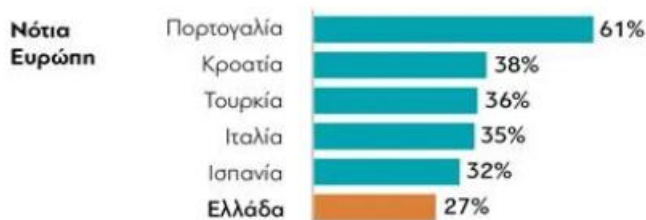
Οι λόγοι δημιουργίας ψευδών ειδήσεων ποικίλουν και μπορεί να είναι τόσο οικονομικοί, όσο και πολιτικοί και οφείλονται στην ύπαρξη αντικρουόμενων ειδήσεων ή ακόμα και ερευνών, οι οποίες ενδεχομένως να πείσουν τον χρήστη να διαδώσει ακόμα και ο ίδιος ψευδείς ειδήσεις [3]. Κάποιες έρευνες δείχνουν ότι οι υπεύθυνοι για διάδοση ψευδών ειδήσεων είναι τα bots, τα οποία είναι κακόβουλοι εισβολείς, που κατακλύζουν το διαδίκτυο και έχουν σαν σκοπό την αυτοματοποίηση εκστρατιών ανεπιθύμητης αλληλογραφίας, την κατασκοπεία σε ανταγωνιστές ή ακόμα και την επίθεση. Η διάδοση βεβαίως μπορεί να γίνει και στοχευμένα από πρόσωπα, τα οποία έχουν ως σκοπό την εξυπηρέτηση συμφερόντων.

1.1.2 Ελληνικά Μέσα Μαζικής Ενημέρωσης

Στην Ελλάδα, το συντριπτικό ποσοστό των Ελλήνων πολιτών, έχει γυρίσει την πλάτη στα Ελληνικά μέσα μαζικής ενημέρωσης, καθώς τα θεωρεί αναξιόπιστα και φερέφωνα τρίτων. Σύμφωνα με την ετήσια έκθεση για την ενημέρωση στο διαδίκτυο, του ινστιτούτου Reuters για την μελέτη της δημοσιογραφίας [4], η Ελλάδα καταλαμβάνει την τελευταία θέση ως προς την ανεξαρτησία των μέσων, τόσο από πολιτικές, όσο και από επιχειρηματικές επιρροές. Συγκεκριμένα, μόλις το 27% του δείγματος της έρευνας, θεωρεί αξιόπιστες τις ειδήσεις τις περισσότερες φορές. Μάλιστα χαρακτηριστική είναι η πτώση κατά 5 ποσοστιαίων μονάδων, σε σχέση με

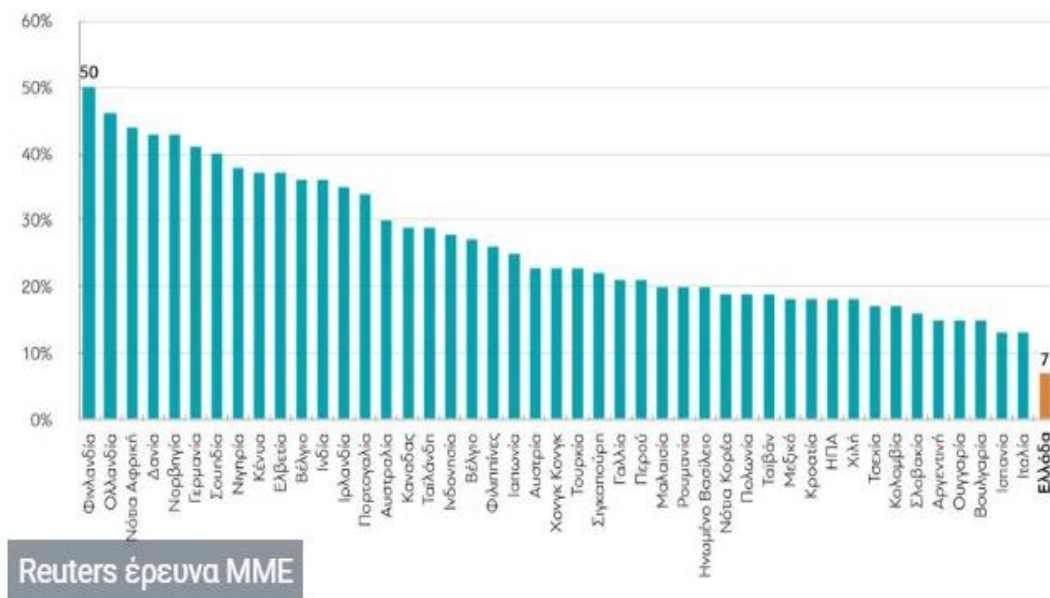
την προηγούμενη χρονιά, παρά το γεγονός ότι κατά την πρώτη φάση της πανδημίας, η εμπιστοσύνη των πολιτών στην ενημέρωση είχε αυξηθεί.

Εικόνα 1: Ποσοστό που εμπιστεύεται τις ειδήσεις σε κάθε χώρα

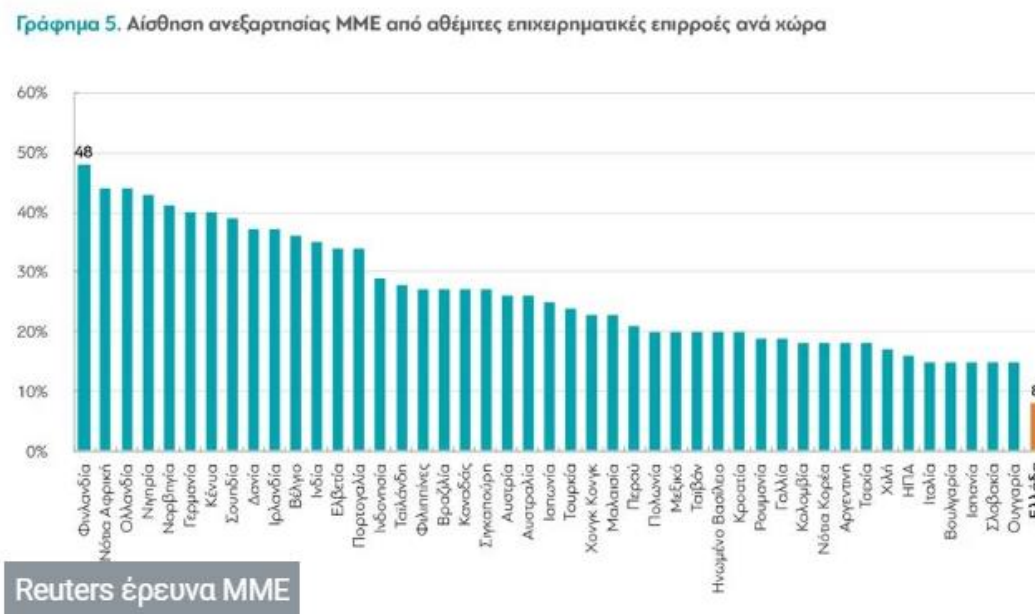


Εύλογη κρίνεται η απορία, ως προς το ρόλο που διαδραματίζει ο πολιτικός προσανατολισμός του κάθε ανθρώπου. Αν το δείγμα διαχωριστεί σε «αριστερούς», «κεντρώους» και «δεξιούς», οι διαφορές είναι μεγάλες. Οι «αριστεροί» είναι η μοναδική πολιτική ομάδα, στην οποία έχει παρατηρηθεί πτώση ως προς την εμπιστοσύνη τους στα μέσα (γύρω στο 2%), ενώ στις άλλες δύο ομάδες έχει παρατηρηθεί άνοδος ως προς την εμπιστοσύνη, αντίθεση η οποία αποτυπώνεται στις κόντρες που δημιουργούνται μεταξύ κυβέρνησης και αντιπολίτευσης. Επιπροσθέτως, μόλις το 7% των ερωτηθέντων θεωρεί ανεξάρτητα τα ΜΜΕ από αθέμιτες πολιτικές επιρροές, ενώ 8% θεωρεί ανεξάρτητα τα ΜΜΕ από επιχειρηματικές επιρροές.

Εικόνα 2: Αίσθηση ανεξαρτησίας ΜΜΕ από αθέμιτες πολιτικές επιρροές ανά χώρα



Εικόνα 3: Αίσθηση ανεξαρτησίας MME από αθέμιτες επιχειρηματικές επιρροές



1.1.3 Παραδείγματα ψευδών ειδήσεων

Η επιρροή, καθώς και οι επιπτώσεις των ψευδών ειδήσεων στην καθημερινότητά μας, γίνονται εύλογα κατανοητές. Παρακάτω, παρατίθενται κάποια παραδείγματα παραπληροφόρησης, καθώς και η επιρροή τους που είχαν στην κοινωνία:

- Ειδηση Παραπληροφόρησης:

«Pizzagate», μια θεωρία συνωμοσίας, η οποία διαδόθηκε κατά τις προεδρικές εκλογές του 2016 στις ΗΠΑ: Λίγες ημέρες πριν τις Αμερικανικές εκλογές, ένας οπαδός της λεγόμενης «λευκής ανωτερότητας», διέδωσε στο twitter πως βρέθηκαν στον προσωπικό υπολογιστή του πολιτικού Άντονι Γουίνερ, μηνύματα που αποκάλυπταν την ύπαρξη ενός μεγάλου κυκλώματος παιδόφιλων, με διασυνδέσεις στους Δημοκράτες. Λίγες μέρες αργότερα, ο λογαριασμός με το όνομα Ντέιβιντ Γκόλτμπεργκ, γνωστός για τις υπερεθνικιστικές αναρτήσεις του και οπαδός του Trump, διέδωσε στο twitter πως αυτή τη φορά, σύμφωνα με πληροφορίες από την αστυνομία της Νέας Υόρκης, υπάρχει ισχυρός πυρήνας παιδόφιλων στους Δημοκράτες. Η πληροφορία προερχόταν από ανώνυμο λογαριασμό στο facebook, ο οποίος επικαλέστηκε ανώνυμη πηγή. Στα emails, σύμφωνα με τη θεωρία, ο επικεφαλής της προεκλογικής καμπάνιας της Χίλαρι Κλίντον, Τζον Ποτέστα, χρησιμοποιούσε τακτικά τη λέξη "pizza" κάτι που σύμφωνα με τους υποστηρικτές της συνωμοσίας, ήταν η κωδική λέξη για τη παιδοφιλία. Οι ισχυρισμοί υποστήριζαν ότι τα όργανα των παιδόφιλων λάμβαναν χώρα σε μυστικά δωμάτια στην πιτσαρία Comet Ping Pong, ιδιοκτησίας του ελληνικής καταγωγής Τζέιμς Αλεφάντης. Στην ιστορία, προστέθηκαν λίγο αργότερα βασανισμοί και θυσίες παιδιών στο σατανά. Τίποτα απ' όλα αυτά δε στηριζόταν σε στοιχεία ή αποδείξεις, πέρα των εξωφρενικών ισχυρισμών που διαδίδονταν

από αμφιβόλου ποιότητας ιστοσελίδες και χρήστες των κοινωνικών δικτύων. Ο Τζέιμς Αλεφάντης που ακόμη δεν είχε πληροφορηθεί για την εμπλοκή του στη συνωμοσία, άρχισε να λαμβάνει μηνύματα μίσους και απειλές μέσω κοινωνικών δικτύων. Αμέσως, απευθύνθηκε στην αστυνομία, ενώ το γεγονός πως ήταν ομοφυλόφιλος, φαίνεται πως έπαιξε ρόλο στην εμπλοκή του ονόματός του. Ο Τζέιμς Αλεφάντης, αν και υποστηρικτής της Κλίντον, ουδέποτε την είχε γνωρίσει, δεν πουλούσε, ούτε κακοποιούσε παιδιά και δεν τον παρακολουθούσαν οι αρχές για τίποτα από τα ανωτέρω, σύμφωνα και με τους New York Times. Μετέπειτα δημοσιεύματα από "αφυπνισμένα" σάιτ του εξωτερικού, απλώς αναπαρήγαγαν την είδηση, διαδίδοντας ακόμα περισσότερο μια αβάσιμη θεωρία. Η κορύφωση θα ερχόταν, όταν ένας άνδρας από τη Βόρεια Καρολίνα, συνελήφθη όταν εισέβαλε στη πιτσαρία Comet Ping Pong, οπλισμένος, για να ερευνήσει τις καταγγελίες στο συγκεκριμένο κατάστημα. Ο δράστης άρχισε να πυροβολεί προς κάθε κατεύθυνση, χωρίς ευτυχώς να υπάρξουν νεκροί ή τραυματίες. Η αστυνομία που έφτασε γρήγορα στο σημείο, συνέλαβε τον 28χρονο δράστη. Όπως καταλαβαίνει κανείς, η διάδοση μιας κατασκευασμένης πληροφορίας, ειδικά στα μέσα κοινωνικής δικτύωσης, μπορεί να έχει καταστροφικές επιπτώσεις στις ζωές αθώων πολιτών.

- Είδηση λανθασμένου περιεχομένου:

«Η σύζυγος του Άλμπερτ Μπουρλά πέθανε ύστερα από επιπλοκές του εμβολιασμού της κατά της COVID-19». Στα μέσα Νοεμβρίου του 2021, κυκλοφόρησε ο ισχυρισμός ότι η σύζυγος του Άλμπερτ Μπουρλά πέθανε ύστερα από επιπλοκές του εμβολιασμού της κατά της COVID-19. Σε αρκετά Ελληνικά σάιτ, αναφέρεται χαρακτηριστικά: Πέθανε η σύζυγος του Άλμπερτ Μπουρλά, η σύζυγος του διευθύνοντος συμβούλου της Pfizer, μετά από επιπλοκές με το αντίστοιχο εμβόλιο. Είναι μια είδηση που αναρτήθηκε από την Καναδέζικη ιστοσελίδα <https://www.conservativebeaver.com/>, η οποία εδρεύει στον Καναδά και δεν υπόκειται στη δικαστική απόφαση για αποσιώπηση του γεγονότος στις ΗΠΑ. Ωστόσο, αυτός ο ισχυρισμός όχι μόνο δεν ανταποκρίνεται στην πραγματικότητα, αλλά εκπρόσωποι επικοινωνίας της Pfizer διέψευσαν επίσημα τον υπό εξέταση ισχυρισμό, δείχνοντας και την επιρροή μιας ψευδούς είδησης στην κοινωνική ζωή: "Παρά τα όσα κυκλοφόρησαν σήμερα στο διαδίκτυο, η σύζυγος του διευθύνοντα σύμβουλου της εταιρείας μας είναι ζωντανή και καλά στην υγεία της. Είναι ανήθικο κάποιος που παρουσιάζεται ως δημοσιογράφος να αναπαράγει εξωφρενικά ψεύδη για τον διευθύνοντα σύμβολο και την οικογένειά του με σκοπό να υπονομεύσει την εμπιστοσύνη σε ένα εμβόλιο, το οποίο έχει χορηγηθεί σε εκατοντάδες εκατομμύρια ανθρώπους παγκοσμίως. Προσπαθούν εσκεμμένα και κακόβουλα να προκαλέσουν συναισθηματική πίεση στην οικογένεια Μπουρλά. Ειδικότερα, κάνουν κακό σε αθώους ανθρώπους σε όλο τον κόσμο, οι οποίοι βάζουν σε κίνδυνο τις ζωές τους με το να πιστεύουν σε τέτοιες ανοησίες".

- Είδηση χειραγώγησης:

"Rosswell Incident": Την 8η Ιουλίου 1947, η RAAF εξέδωσε δελτίο τύπου δηλώνοντας ότι προσωπικό του, ανέκτησε τα συντρίμια ενός "ιπτάμενου δίσκου" σε ράντζο κοντά στο Ρόσγουελ, πυροδοτώντας το έντονο ενδιαφέρον των ΜΜΕ. Αργότερα την ίδια μέρα, ο Στρα-

τηγός διοικητής, δήλωσε ότι στην πραγματικότητα, ένα μετεωρολογικό μπαλόνι είχε ανακτηθεί από προσωπικό της RAAF και όχι "ιπτάμενος δίσκος". Ακολούθησε συνέντευξη τύπου, αναδεικνύοντας χαλάσματα φερόμενα ότι ανήκαν στο κατεστραμμένο αντικείμενο, τα οποία έδειχναν να επαληθεύουν την περιγραφή του μετεωρολογικού μπαλονιού. Η υπόθεση ξεχάστηκε γρήγορα και αγνοήθηκε για περισσότερο από 30 χρόνια, όμως τότε, το 1978 ο ειδικός των ΑΤΙΑ, Στάντον Φρίντμαν, κάλεσε σε συνέντευξη τον Ταγματάρχη Τζέσιε Μαρσέλ, ο οποίος είχε αναμειχθεί στην αρχική ανάκτηση των χαλασμάτων το 1947. Ο Μαρσέλ εξέφρασε την πεποίθηση ότι ο στρατός είχε συγκαλύψει την ανάκτηση ενός εξωγήινου σκάφους, ιστορία η οποία διαδόθηκε στους κύκλους εκείνης της εποχής. Το 1989 ο πρώην εργολάβος κηδειών Γκλεν Ντένις, έφερε στην επιφάνεια λεπτομερή λογαριασμό, ισχυριζόμενος ότι αυτοψίες εξωγήινων όντων έλαβαν χώρα στη βάση του Ρόσγουελ. Σε απάντηση στις παραπάνω αναφορές και μετά από ερωτήσεις του Κογκρέσου, το Κυβερνητικό Γραφείο Λογοδοσίας ξεκίνησε διερεύνηση και οδήγησε το Γραφείο της Γραμματείας Πολεμικής Αεροπορίας (SECAF) να διεξάγει εσωτερική έρευνα. Το αποτέλεσμα περιλήφθηκε σε δύο αναφορές. Η πρώτη, με χρονολογία έκδοσης 1995, κατέληγε ότι το ανακτηθέν το 1947 υλικό ήταν πιθανώς συντρίμια ενός μυστικού κυβερνητικού προγράμματος με την ονομασία "Μόγκουλ". Η δεύτερη αναφορά, με χρονολογία έκδοσης 1997, κατέληγε ότι οι αναφορές για περισυλλογή πτωμάτων εξωγήινων όντων πιθανώς προέρχονταν από ακουσίως ταραγμένες αναμνήσεις στρατιωτικών ατυχημάτων κατά τα οποία προσωπικό είχε τραυματιστεί ή σκοτωθεί και στην ανάκτηση ανθρωπόμορφων ομοιωμάτων κατά την διεξαγωγή στρατιωτικών προγραμμάτων όπως ένα πρόγραμμα το οποίο έλαβε χώρα την δεκαετία του '50 καθώς επίσης και σε φάρσες από διάφορους μάρτυρες και υποστηρικτές των ΑΤΙΑ.

- Είδηση σάτιρας:

Παρά το εμφανές χιουμοριστικό ύφος γραφής, το παρακάτω άρθρο, αναπαράχθηκε ως είδηση σε αρκετούς ιστοτόπους: «Πέθανε ο τελευταίος κάτοχος μηχανακιού της AGB στην Ελλάδα», μια είδηση χιουμοριστικής φύσης, η οποία αναφέρει χαρακτηριστικά: Έσβησε αθόρυβα χθες αργά το βράδυ ο Χαράλαμπος Βουρεξάκης, κάτοικος Στούπας Μεσσηνίας, ο οποίος ήταν και ο τελευταίος εν ζωή κάτοχος μηχανακιού της AGB στη χώρα μας. Ο 89χρονος, είχε παραλάβει τη συσκευή που μετράει τις τηλεθέσεις πίσω στο 2005, και την είχε χρησιμοποιήσει για να καταγράψει κυρίως την παρακολούθηση ντοκιμαντέρ στην ΕΡΤ, αθλητικών μεταδόσεων και δελτίων ειδήσεων, ενώ σύμφωνα με πληροφορίες, αυτή είχε σταματήσει να λειτουργεί από το 2012. Τη θλιβερή είδηση επιβεβαίωσε η σύζυγος του κυρίου Χαράλαμπου, Αγλαΐα, η οποία μάλιστα κάλεσε την εταιρεία Nielsen στο τηλέφωνο, ενημερώνοντάς την πως θα αναλάβει εκείνη το χρέος του άνδρα της. Οι υπάλληλοι της εταιρείας, νέοι στην πλειονότητά τους, δεν γνωρίζουν καν τι είναι τα μηχανάκια αυτά, ωστόσο έσπευσαν να της εκφράσουν τα συλλυπητήριά τους, ενώ προς τιμήν του κυρίου Χαράλαμπου, ανέβασαν χθες την τηλεθέαση της ΕΡΤ3 κατά 1%. Τέλος, διαψεύδονται οι φήμες πως στην κηδεία του θα χρησιμοποιηθεί ειδικό μηχανάκι καταγραφής κλάματος, με το «Κουλούρι» να επανέρχεται ωστόσο για οτιδήποτε νεότερο.

1.1.4 Κλασσική Μεθοδολογία εντοπισμού ψευδών ειδήσεων

Ο εντοπισμός ψευδών ειδήσεων, μπορεί να γίνει και χωρίς κάποιο αλγόριθμο, αλλά μέσω έρευνας, στην οποία συμμετέχει ο ανθρώπινος παράγοντας, με ό,τι συνέπειες μπορεί να συνεπάγεται αυτό. Δεδομένου ότι το υλικό που καλούμαστε να διαχειριστούμε μπορεί να κυμαίνεται από τον έλεγχο μιας εικόνας με σκοπό την εξακρίβωση επεξεργασίας, μέχρι τον έλεγχο ψευδοϊατρικών ισχυρισμών, ή ακόμη και τον έλεγχο και την επεξήγηση ενός Σχεδίου Νόμου, δεν μπορούμε να ισχυριστούμε πως υπάρχει μια και μοναδική μεθοδολογία που αρμόζει σε όλες τις περιπτώσεις. Τα δεδομένα που χρησιμοποιούμε προσπαθούμε πάντα να είναι τα πιο πολύπλευρα, αμερόληπτα, έγκριτα και πρόσφατα πάνω στο εκάστοτε θέμα που ερευνούμε και πάντοτε οφείλουμε να προσπαθούμε να τα διασταυρώσουμε από πολλαπλές πηγές. Όπως γνωρίζουν οι περισσότεροι ελεγκτές ειδήσεων, κάθε περίπτωση είναι διαφορετική, ωστόσο η αξιοποίηση μια σειράς γενικών βημάτων [5], η οποία θα δίνει και τη δυνατότητα ευελιξίας στον εκάστοτε ελεγκτή ειδήσεων μπορεί να βοηθήσει στο να περιορίζεται ο χρόνος που χρησιμοποιείται σε κάθε ανάλυση.

- Βήμα 1^ο: Εντοπισμός δυνητικά ύποπτου υλικού

Σε κάθε μορφή πληροφορίας που εξετάζουμε, είτε πρόκειται για άρθρο, φωτογραφία ή βίντεο, μπορούμε να αναζητήσουμε συγκεκριμένες λεπτομέρειες οι οποίες μας οδηγούν στο συμπέρασμα πως το υλικό που εξετάζουμε μπορεί να μην είναι έγκυρο. Χαρακτηριστικά παραδείγματα αποτελούν οι τίτλοι με κεφαλαία γράμματα και έντονα συναισθηματική γλώσσα, η απουσία αρθρογράφου, η πρόταση για άμεση κοινοποίηση και ούτω καθεξής. Αυτές οι λεπτομέρειες δε καθιστούν αυτόματα το υπό εξέταση υλικό ψευδές απλά δεδομένου του τεράστιου όγκου πληροφοριών με τον οποίο ερχόμαστε καθημερινά σε επαφή, μπορούν να λειτουργήσουν ως σημάδια για να εντοπίσουμε δυνητικά ύποπτο υλικό το οποίο χρίζει περαιτέρω ανάλυσης.

- Βήμα 2^ο: Ανάλυση του περιεχομένου

Αφού επιλέξουμε το υλικό που θέλουμε να αξιολογήσουμε, εξετάζουμε προσεκτικά το περιεχόμενο κρατώντας μια μικρή λίστα με όλους τους ισχυρισμούς που γίνονται εντός ώστε να τους ελέγξουμε ξεχωριστά.

- Βήμα 3^ο: Έρευνα στο οπτικοακουστικό υλικό

Στη περίπτωση που το υλικό στο οποίο δουλεύουμε είναι ένα άρθρο το οποίο περιέχει εικόνες ή βίντεο σχετικά με τους ισχυρισμούς που γίνονται εντός του κειμένου πρέπει να εντοπίσουμε αν αυτό το οπτικοακουστικό υλικό σχετίζεται όντως με τους εν λόγω ισχυρισμούς. Αφού εντοπίσουμε τη πηγή της εικόνας ή του βίντεο, παρατηρούμε αν έχουν όντως παραποιηθεί ή αν το νόημα τους έχει αλλοιωθεί στο υπό εξέταση άρθρο.

- Βήμα 4^ο: Εξέταση επιστημονικών μελετών σε περιπτώσεις ψευδοεπιστημονικών ισχυρισμών

Όταν εξετάζουμε ένα άρθρο το οποίο κάνει ισχυρισμούς επιστημονικού περιεχομένου το πρώτο βήμα είναι να κάνουμε μια λίστα με κάθε ισχυρισμό και να ψάξουμε τον κάθε ένα με χρήση της δημοσιευμένης επιστημονικής βιβλιογραφίας.

1.2. Κίνητρο Υλοποίησης

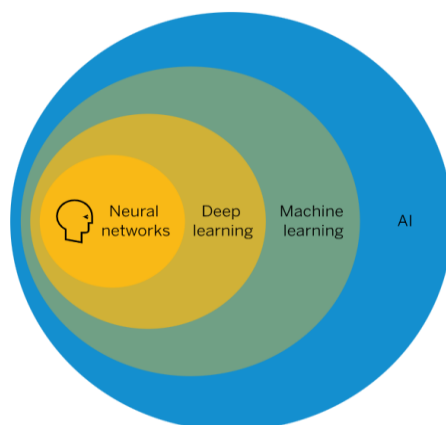
Το πρόβλημα εντοπισμού των ψευδών ειδήσεων και το ευρύτερο ζήτημα εξάλειψης της διάδοσής τους, εξακολουθούν να προσελκύουν την προσοχή τόσο των «καταναλωτών» τους, όσο και των δημιουργών τους. Στην πρόσφατη έξαρση της διάδοσης των ψευδών ειδήσεων, έχουν συμβάλει τόσο το διαδίκτυο όσο και διάφορα γεγονότα, όπως ο κορονοϊός, τα οποία έχουν σπείρει την διχόνοια, στρέφοντας τους ανθρώπους τον έναν, ενάντια στον άλλο, δίνοντας ανεξέλεγκτα βήμα σε διάφορες αντικρουόμενες και πολλές φορές, ατεκμηρίωτες απόψεις. Το γεγονός αυτό έχει αναγνωριστεί από διάφορους φορείς και γίνεται μια αξιόπαινη προσπάθεια καταπολέμησης του φαινομένου αυτού, γεγονός που ενισχύεται από το ότι ακόμα και σε ιστοτόπους κοινωνικής δικτύωσης, επιστρατεύονται μέσα για να περιοριστεί η διάδοση των ψευδών ειδήσεων, όπως γίνεται π.χ. στο Facebook. Δυστυχώς, ο χειροκίνητος έλεγχος δεδομένων μέσω των κλασικών μεθοδολογιών είναι πολύ επίπονος και χρονοβόρος για να εφαρμοστεί σε κάθε ανάρτηση που δημοσιεύεται σε αυτές τις πλατφόρμες και αλλού. Έτσι εξετάζουμε τις σύγχρονες τεχνικές που θα μπορούσαν να αξιολογήσουν αυτόματα την αξιοπιστία των ειδήσεων που δημοσιεύονται στο διαδίκτυο, οι οποίες θα μπορούσαν επίσης να είναι χρήσιμες για άλλα ενδιαφερόμενα για να μειώσουν τον αντίκτυπο της παραπληροφόρησης, συμπεριλαμβανομένων των δημοσιογράφων και των χρηστών του διαδικτύου. Άρα, μείζονος σημασίας κρίνεται η περαιτέρω ανάπτυξη τεχνικών και μεθόδων αυτόματης αναγνώρισης των ψευδών ειδήσεων, διαδικασία κατά την οποία το περιεχόμενο μιας είδησης επιβεβαιώνεται ή αμφισβητείται.

2. Μηχανική Μάθηση

Η μηχανική μάθηση ορίζεται ως ένα υποσύνολο της επιστήμης των υπολογιστών, που αναπτύχθηκε από την μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη [6]. Η μηχανική μάθηση εμβαθύνει ως προς την μελέτη και την κατασκευή αλγορίθμων, οι οποίοι μέσω εκμάθησης δεδομένων, θα μπορούν να κάνουν προβλέψεις ως προς αυτά. Ο αυτοσκοπός της είναι η διδασκαλία των υπολογιστών για να τους εκπαιδεύσει ως προς την εκμάθηση δεδομένων και να βελτιώνονται μέσω εκπαίδευση. Στη μηχανική μάθηση, οι αλγόριθμοι εκπαιδεύονται να βρίσκουν μοτίβα και συσχετίσεις σε σύνολα δεδομένων και μέσω διαφόρων μοντέλων κάνουν προβλέψεις και λαμβάνουν αποφάσεις. Οι εφαρμογές της μηχανικής μάθησης στην καθημερινή μας ζωή ποικίλουν από την υγεία, στο επιχειρείν ή ακόμα και στα σπίτια μας.

Αρχικά, ξεκινώντας από τα νευρωνικά δίκτυα, ένα τεχνητό νευρωνικό δίκτυο μοντελοποιεί τους νευρώνες σε έναν βιολογικό εγκέφαλο. Οι τεχνητοί νευρώνες ονομάζονται κόμβοι και ομαδοποιούνται σε πολλαπλά επίπεδα, τα οποία λειτουργούν παράλληλα και όταν λαμβάνουν ένα αριθμητικό σήμα, το επεξεργάζονται και σηματοδοτούν στους άλλους νευρώνες που συνδέονται με αυτό. Όπως και στον ανθρώπινο εγκέφαλο, η νευρωνική ενίσχυση έχει ως αποτέλεσμα τη βελτιωμένη αναγνώριση προτύπων, την εμπειρογνωμοσύνη και τη συνολική μάθηση. Στη συνέχεια, έχουμε τη βαθιά μάθηση, η οποία περιλαμβάνει πολλά επίπεδα του νευρωνικού δικτύου και τεράστιους όγκους πολύπλοκων και ανόμοιων δεδομένων. Για την επίτευξη βαθιάς μάθησης, το σύστημα εμπλέκεται με πολλαπλά επίπεδα στο δίκτυο, εξάγοντας όλο και πιο υψηλά επίπεδα εξόδου. Παραδείγματα εφαρμογών βαθιάς μάθησης περιλαμβάνουν την αναγνώριση ομιλίας, την ταξινόμηση εικόνας και τη φαρμακευτική ανάλυση. Έπειτα έχουμε την μηχανική μάθηση, η οποία αποτελείται από διαφορετικούς τύπους μοντέλων μηχανικής μάθησης, χρησιμοποιώντας διάφορες αλγοριθμικές τεχνικές. Ανάλογα με τη φύση των δεδομένων και το επιθυμητό αποτέλεσμα, ένα εκ των τεσσάρων μοντέλων μάθησης: επιβλεπόμενη, μη επιβλεπόμενη, ημι-επιβλεπόμενη και ενισχυμένη, μπορούν να χρησιμοποιηθούν. Σε καθένα από τα μοντέλα αυτά, μπορούν να εφαρμόζονται μία ή περισσότερες αλγοριθμικές τεχνικές – σε σχέση με τα σύνολα δεδομένων που χρησιμοποιούνται και τα επιδιωκόμενα αποτελέσματα. Οι αλγόριθμοι μηχανικής μάθησης είναι βασικά σχεδιασμένοι για να ταξινομούν τα δεδομένα, να βρίσκουν πρότυπα, να προβλέπουν αποτελέσματα και να λαμβάνουν ενημερωμένες αποφάσεις και μπορούν να χρησιμοποιηθούν ή να συνδυαστούν για να επιτύχουν την καλύτερη δυνατή ακρίβεια όταν εμπλέκονται πολύπλοκα και πιο απρόβλεπτα δεδομένα.

Εικόνα 4: Διάφοροι τρόποι μάθησης [7]



2.1 Τύποι Εργασιών Μηχανικής Μάθησης

Όπως προαναφέρθηκε, οι εργασίες μηχανικής μάθησης, ταξινομούνται σε τέσσερις μεγάλες κατηγορίες, αναλόγως της φύσεως του εκπαιδευτικού σήματος που παρέχεται. Οι κατηγορίες εξηγούνται και αναλύονται παρακάτω:

- Επιτηρούμενη μάθηση ή Προβλεπόμενη μάθηση

Ορίζεται ως το υπολογιστικό πρόγραμμα που διδάσκεται μέσω παραδειγμάτων – εισόδων. Τα μοντέλα επιτηρούμενης μάθησης, αποτελούνται από ζεύγη εισόδου – εξόδου, όπου η έξοδος έχει μια συγκεκριμένη τιμή, η οποία ορίζεται ως επιθυμητή προβλεπόμενη τιμή. Για παράδειγμα, στο θέμα των ψευδών ειδήσεων, έχουμε τις εισόδους και μια τιμή που καταδεικνύει το κατά πόσο μια είδηση είναι αληθής ή ψευδής και ο σκοπός του αλγορίθμου είναι να μας επιστρέψει την σωστή τιμή και κατ' επέκταση να ταξινομήσει σωστά τις ειδήσεις. Μέσω ενός αλγορίθμου, το σύστημα συλλέγει όλα αυτά τα δεδομένα εκπαίδευσης με την πάροδο του χρόνου και αρχίζει να καθορίζει τις συσχετίσεις, τις διαφορές και άλλα σημεία της λογικής, μέχρι να μπορεί να προβλέψει σωστά την ετικέτα κάθε είδησης (αληθής ή ψευδής). Κάποια άλλα παραδείγματα εφαρμογών επιτηρουμένων μοντέλων μάθησης είναι οι μηχανές συστάσεων για προϊόντα και εφαρμογές ανάλυσης κυκλοφορίας, που προβλέπουν την ταχύτερη διαδρομή σε διαφορετικές ώρες της ημέρας.

- Μη επιτηρούμενη μάθηση ή Μη επιβλεπόμενη μάθηση

Ορίζεται ως η εύρεση της δομής δεδομένων εισόδου, χωρίς να παρέχεται κάποια «εμπειρία» στον αλγόριθμο μάθησης. Η μηχανή μελετά τα δεδομένα εισόδου, μεγάλο μέρος των οποίων ενδέχεται να είναι χωρίς ετικέτες και μη δομημένα και αρχίζει να αναγνωρίζει μοτίβα και συσχετίσεις, χρησιμοποιώντας όλα τα σχετικά, προσβάσιμα δεδομένα. Γενικότερα στην μη επιτηρούμενη μάθηση, χρησιμοποιείται η «διαίσθηση» και η «εμπειρία» για να ομαδοποιηθούν τα δεδομένα. Για τις μηχανές, η «εμπειρία» ορίζεται από την ποσότητα των δεδομένων που

εισάγονται και διατίθενται. Κοινά παραδείγματα μη εποπτευόμενων εφαρμογών μάθησης περιλαμβάνουν την αναγνώριση προσώπου, την ανάλυση γονιδιακής αλληλουχίας, την έρευνα αγοράς και την ασφάλεια στον κυβερνοχώρο.

- Ημι-επιτηρούμενη μάθηση

Ορίζεται ως το υπολογιστικό πρόγραμμα που διδάσκεται μέσω συνδυασμού των δύο παραπάνω μεθόδων. Η ημι-επιτηρούμενη μάθηση γίνεται μια λειτουργική λύση όταν υπάρχουν τεράστιες ποσότητες ανεπεξέργαστων, μη δομημένων δεδομένων. Αυτό το μοντέλο αποτελείται από την εισαγωγή μικρών ποσοτήτων δεδομένων με ετικέτες για την αύξηση συνόλων δεδομένων χωρίς ετικέτες. Ουσιαστικά, τα επισημασμένα δεδομένα ενεργούν για να δώσουν μια αρχή στο σύστημα και μπορούν να βελτιώσουν σημαντικά την ταχύτητα και την ακρίβεια της μάθησης. Ένας ημι-εποπτευόμενος αλγόριθμος εκμάθησης δίνει οδηγίες στη μηχανή να αναλύσει τα δεδομένα με ετικέτα για τις σχετικές ιδιότητες που θα μπορούσαν να εφαρμοστούν στα μη επιτηρούμενα δεδομένα. Η ημι-επιβλεπόμενη μάθηση χρησιμοποιείται στην ανάλυση της ομιλίας και της γλώσσας, στην πολύπλοκη ιατρική έρευνα όπως η κατηγοριοποίηση πρωτεϊνών και η ανίχνευση απάτης υψηλού επιπέδου.

- Ενισχυτική μάθηση

Ορίζεται ως ένα πρόγραμμα υπολογιστή αλληλεπιδρά με ένα δυναμικό περιβάλλον, στο οποίο πρέπει να επιτευχθεί ένας συγκεκριμένος στόχος, χωρίς να υπάρχει κάποιος «δάσκαλος», να του λέει ρητά αν έχει φτάσει κοντά στο στόχο του. Το μοντέλο εκμάθησης ενίσχυσης δεν περιλαμβάνει ένα κλειδί απάντησης αλλά, εισάγει ένα σύνολο επιτρεπόμενων δράσεων, κανόνων και πιθανών τελικών καταστάσεων. Όταν ο επιθυμητός στόχος του αλγορίθμου είναι σταθερός ή δυαδικός, οι μηχανές μπορούν να μάθουν με παράδειγμα. Αλλά σε περιπτώσεις όπου το επιθυμητό αποτέλεσμα είναι αμετάβλητο, το σύστημα πρέπει να μαθαίνει από την εμπειρία και την ανταμοιβή. Στα ενισχυτικά μαθησιακά μοντέλα, η “ανταμοιβή” είναι αριθμητική και προγραμματίζεται στον αλγόριθμο ως κάτι που το σύστημα επιδιώκει να συλλέξει. Σαν παράδειγμα μπορεί να δοθεί, η εκμάθηση οδήγησης, χωρίς καθοδήγηση ή η διδασκαλία σε κάποιον πώς να παίζει σκάκι.

2.2. Επεξήγηση Χρήσιμων Εννοιών

2.2.1. Εξόρυξη Γνώσης Δεδομένων

Η εξόρυξη γνώσης δεδομένων είναι μια μεθοδολογία, μέσω της οποίας επιτυγχάνεται ανακάλυψη χρήσιμης γνώσης από δεδομένα η οποία υπό φυσιολογικές συνθήκες δεν μπορεί να εξαχθεί από τον άνθρωπο με ευκολία και η οποία θα ήταν πρακτικά αδύνατη χωρίς τη χρήση ηλεκτρονικού υπολογιστή. Η διαφορά της εξόρυξης γνώσης από δεδομένα, σε σχέση με την εξόρυξη γνώσης από κείμενα είναι στο γεγονός ότι στην εξόρυξη γνώσης από κείμενο, οι συσχετίσεις και η γνώση, εξάγεται από αδόμητα κείμενα και όχι από δομημένες βάσεις δεδομένων [8]. Οι τεχνικές εξόρυξης κειμένου είναι πιο περίπλοκες σε σχέση με τεχνικές εξόρυξης

από δεδομένα, γεγονός που οφείλεται και στην ασαφή και αδόμητη φύση των κειμένων φυσικής γλώσσας.

2.2.2. Εξόρυξη Γνώσης Κειμένου

Η εξόρυξη κειμένου – text mining είναι η διαδικασία κατά την οποία εξάγονται υψηλής ποιότητας πληροφορίες από κείμενα, με τη χρήση πληροφοριακών συστημάτων και εντοπίζει με αυτόματο τρόπο πληροφορίες από διάφορους γραπτούς πόρους όπως ιστοτόπους, βιβλία, μηνύματα ηλεκτρονικού ταχυδρομείου, κριτικές και άρθρα [9]. Η εξόρυξη κειμένων χρησιμοποιεί τεχνικές από διάφορα επιστημονικά πεδία, όπως την εξόρυξη δεδομένων, την μηχανική μάθηση, την ανάκτηση πληροφοριών, την στατιστική, την επεξεργασία φυσικής γλώσσας και την διαχείριση γνώσεων, ώστε να μπορούμε να κατανοήσουμε τα δεδομένα και να μας δίνεται η δυνατότητα να εξάγουμε συμπεράσματα ως προς αυτά, αφού τύχουν της κατάλληλης μεταχείρισης. Οι στόχοι της εξόρυξης κειμένου [10] αναφέρονται ως εξής:

- Η εξόρυξη πληροφορίας από αδόμητα ή ημι-δομημένα δεδομένα.
- Εξαγωγή στοιχείων που προέρχονται από μεγάλες ποσότητες, πηγών δεδομένων.
- Αναπαράσταση σημασίας σχετιζόμενων στοιχείων
- Δημιουργία τεχνικών που θα εξάγουν οντότητες και τα θα δημιουργήσουν συνδέσεις μεταξύ τους.
- Αποτελεσματική διαχείριση γνώσης.

Τα ζητήματα τα οποία είναι χρήσιμα κατά τη διάρκεια της εξόρυξης γνώσης ποικίλουν από την εξαγωγή πληροφοριών, την κατηγοριοποίηση, την ομαδοποίηση, την συνόψιση, την απεικόνιση πληροφορίας, την διασύνδεση εννοιών και την εξαγωγή οντολογιών. Παρότι δεν υπάρχουν προγράμματα που να μπορούν να διαβάσουν και να κατανοήσουν πλήρως κείμενο, το οποίο είναι γραμμένο σε φυσική γλώσσα, έχει αναπτυχθεί ο τομέας επεξεργασίας φυσικής γλώσσας, ο οποίος σημείωσε σημαντική πρόοδο στην ανάλυση κειμένου.

2.2.2.1. Συλλογή Δεδομένων

Το πρώτο βήμα στην εξόρυξη κειμένου είναι η συλλογή των δεδομένων. Σε αρκετές περιπτώσεις, η διαδικασία εξόρυξης κειμένων ενδέχεται να έχει ήδη πραγματοποιηθεί. Σε αντίθετη περίπτωση, ενδέχεται να χρειαστεί να αναπτυχθεί μια διαδικασία συλλογής δεδομένων, η οποία θα συλλέγει έγγραφα από τον ιστό, μέσω χρήσης διαφόρων λογισμικών ή βιβλιοθηκών, όπως για παράδειγμα το BeautifulSoup και το Trafilatura. Αφού συλλεχθούν τα δεδομένα, τότε το κύριο ζήτημα που εκκρεμεί είναι ο καθαρισμός των δεδομένων και η διασφάλιση της ποιότητάς τους. Στις περιπτώσεις όπου τα έγγραφα έχουν ληφθεί από αποθήκες εγγράφων ή βάσεις δεδομένων, αναμένεται ότι ο καθαρισμός δεδομένων έγινε πριν από την κατάθεση και μπορούμε να είμαστε σίγουροι για την ποιότητα των εγγράφων.

2.2.2.2. Μετατροπή Κειμένου

Αφού συλλεχθούν τα έγγραφα, ενδέχεται να βρίσκονται σε αρκετές μορφές, είτε λόγω του ότι έχουν δημιουργηθεί από ένα πρόγραμμα επεξεργασίας κειμένου, είτε από ένα επεξεργαστή. Η πιο συχνή μορφή αποθήκευσης κειμένου είναι γνωστή ως ASCII και για να μπορέσει να γίνει η επεξεργασία, πρέπει να μετατραπούν σε μια τυπική μορφή, ώστε να αφαιρεθούν τα δομικά τους χαρακτηριστικά. Στην περίπτωση που τα κείμενα είναι της μορφής, HTML ή και άλλης μορφής, επεξεργάζονται με τέτοιο τρόπο, ώστε να αφαιρεθούν τα δομικά τους συστατικά, όπως π.χ. στην περίπτωση του HTML, τα HTML tags και να εξαχθεί το απλό κείμενο.

2.2.2.3. Κατακερματισμός

Το πρώτο βήμα στην επεξεργασία κειμένου είναι ο διαχωρισμός της ροής των χαρακτήρων σε λέξεις ή όπως ονομάζονται σε κέρματα (tokens). Τα κέρματα αυτά μπορεί να είναι λέξεις, σύμβολα ή ακόμα και σημεία στίξης. Όμως, χαρακτήρες όπως κενά, tabs και χαρακτήρες νέας γραμμής, δεν υπολογίζονται ως κέρματα και ορίζονται ως λευκά κενά. Για το καλύτερο δυνατό αποτέλεσμα, ο κατακερματιστής, θα πρέπει να προσαρμόζεται στο διαθέσιμο κείμενο και να λαμβάνεται υπόψη ότι η διαδικασία διαμόρφωσης, εξαρτάται από τη γλώσσα.

2.2.2.4 Αφαίρεση Σημείων Στίξης

Λόγω του θορύβου που προκαλούν τα σημεία στίξης (punctuations), εκτός των περιπτώσεων που θεωρούνται μέρη ενός αριθμού (π.χ. 3,14), αφαιρούνται πλήρως από το σώμα κειμένου, αφού δεν προσφέρουν καμιά πληροφορία και δεν συμβάλουν καθόλου στο εννοιολογικό περιεχόμενο του κειμένου. Έτσι, απομένουν οι λέξεις που θεωρούνται χρήσιμες ως προς το κείμενο.

2.2.2.5. Αφαίρεση Stopwords

Stopwords θεωρούνται οι όροι, οι οποίοι έχουν μεγάλη συχνότητα εμφάνισης μέσα στο κείμενο και δεν προσφέρουν κάτι ως προς το περιεχόμενο του κειμένου. Συνήθως αποτελούν, προθέσεις, άρθρα ή συνδέσμους. Ενδεχόμενη ύπαρξή τους στο σώμα κειμένου, θα προκαλέσει θόρυβο. Στις περιπτώσεις όπου η εργασία που επιτελείται, αφορά σχηματισμό γράφου γνώσης, τότε η διαδικασία αυτή κρίνεται ζημιογόνα, καθώς σε αρκετά παραδείγματα επηρεάζει το νόημα της πρότασης και τη λεξικογραφική της ανάλυση.

2.2.2.6 Λημματοποίηση - Ανακοπή

Έχοντας πλέον σχηματίσει μια ροή χαρακτήρων και έχοντας την τμηματοποιήσει σε μια ακολουθία κερμάτων, το επόμενο βήμα είναι η μετατροπή του κέρματος σε μια τυπική φόρμα, μια

διαδικασία που ονομάζεται ανακοπή (stemming) ή λημματοποίηση (lemmatization). Οι διαδικασίες αυτές αποκόπτουν κάθε λέξη, παίρνοντας μόνο την ρίζα της ή κανονικοποιούν τις παραλλαγές ενικού/πληθυντικού, παρόν/παρελθόντος και άλλως καταλήξεων, αντιστοίχως. Η εφαρμογή του βήματος αυτού εξαρτάται από το θέμα το ζήτημα το οποίο επεξεργαζόμαστε.

2.2.2.7. Λεξικογραφική Ανάλυση

Η λεξικογραφική ανάλυση (part of speech tagging), αναγνωρίζει το μέρος του λόγου που αποτελεί η κάθε λέξη, δηλαδή ρήμα, ουσιαστικό, κτλ.

2.2.3. Επεξεργασία Φυσικής Γλώσσας

2.2.3.1. Αναπαράσταση Φυσικής Γλώσσας σε διανύσματα

Η ύπαρξη πολλών δεδομένων, σε μορφή κειμένων είναι συνεχώς αυξανόμενη και αποτέλεσε αφορμή για την άνθιση του κλάδου της φυσικής επεξεργασίας γλώσσας – NLP (Natural Language Processing). Ο στόχος της είναι να χρησιμοποιήσει την υπολογιστική ισχύ των μηχανών για να μελετά και να αναλύει κείμενα με τρόπο παρόμοιο με τον ανθρώπινο εγκέφαλο. Οι εφαρμογές του NLP ποικίλουν και χρησιμοποιείται μεταξύ άλλων για τη μετάφραση κειμένων, την αναγνώριση φωνής, την ανίχνευση συναισθημάτων αλλά και τη τεχνητή δημιουργία κειμένων σε φυσική γλώσσα.

Αξίζει να αναφερθεί ότι για να αντιληφθούν οι υπολογιστές τη φυσική γλώσσα, την αναπαριστούν ως διάνυσμα, η ύπαρξη του οποίου αποσκοπεί στη συγκέντρωση των χαρακτηριστικών της κάθε λέξης, σε σχέση με το νόημά της και τη θέση της στην εκάστοτε πρόταση. Η αναπαράσταση αυτή αποτελεί ένα διάνυσμα, το οποίο ουσιαστικά κωδικοποιεί τη σημασία της λέξης και προκύπτει χρησιμοποιώντας τεχνικές μοντελοποίησης γλώσσας και τεχνικές εκμάθησης χαρακτηριστικών. Οι μέθοδοι, οι οποίες χρησιμοποιούνται για την δημιουργία των διανυσμάτων, περιλαμβάνουν νευρωνικά δίκτυα, πίνακα εμφανίσεων λέξεων που περιέχει τη μοναδική ύπαρξη κάθε λέξης, πιθανοτικά μοντέλα και γνώση βάσης. Αυτά τα διανύσματα χρησιμοποιούνται ως εισόδοι σε διάφορους τομείς της μηχανικής μάθησης και βοηθούν στην βελτίωση της απόδοσης της επεξεργασίας φυσικής γλώσσας.

Η σημασιολογική ενσωμάτωση κατά την οποία μετατρέπεται κείμενο σε ένα σύνολο διανυσμάτων, αναφέρεται σε μια σειρά τεχνικών εκμάθησης αναπαράστασης (ή εκμάθησης χαρακτηριστικών) που κωδικοποιούν τη σημασιολογία των δεδομένων, όπως οι ακολουθίες και τα γραφήματα σε διανύσματα, έτσι ώστε να μπορούν να χρησιμοποιηθούν από εργασίες πρόβλεψης μηχανικής μάθησης και στατιστικής ανάλυσης. Για την δημιουργία διανυσμάτων λέξεων, θεμελιώδης είναι η σημασία κατασκευής σημασιολογικών χώρων, οι οποίοι κατασκευάζονται αυτόματα αναλύοντας την συνύπαρξη των λέξεων στα κείμενα, στα οποία περιέχονται. Οι λέξεις, οι οποίες εμφανίζονται σε κοντινές θέσεις στο σημασιολογικό χώρο τείνουν να έχουν την ίδια ή παρόμοια σημασία. Οι αρχικές μέθοδοι απεικόνισης λέξεων σε διανύσματα, είχαν τον

περιορισμό ότι μία λέξη απεικονιζόταν πάντα με το ίδιο διάνυσμα, δηλαδή η συνάρτηση απεικόνισης της λέξης, χρησιμοποιούσε ως είσοδο μόνο την ίδια τη λέξη, αγνοώντας την ενδεχόμενη αμφισημία των λέξεων. Σε κατοπινό στάδιο, δημιουργήθηκαν μοντέλα απεικόνισης που να λάμβαναν υπόψη και τα συμφραζόμενα της λέξης.

2.2.3.2. Προεπεξεργασία Δεδομένων

Η επεξεργασία δεδομένων φυσικής γλώσσας είναι μια διαδικασία, η οποία συμπεριλαμβάνει αρκετές προκλήσεις, καθώς τα δεδομένα πάσχουν από αρκετά προβλήματα [11]. Όμως, η ύπαρξη προβλημάτων σε δεδομένα πραγματικού κόσμου είναι σχεδόν καθιερωμένη. Τα προβλήματα των δεδομένων προκύπτουν από την ανάγκη για συγχώνευση διάσπαρτων δεδομένων από διάφορες πηγές, χωρίς αυτό να αποτελεί το μοναδικό λόγο. Κάποια από τα προβλήματα που ενδέχεται να παρατηρηθούν σε ένα σύνολο δεδομένων είναι διπλοκαταχωρημένες εγγραφές, εγγραφές με αντικρουόμενο περιεχόμενο, τιμές που παραβιάζουν λογικούς κανόνες, εσφαλμένες τιμές ή ακόμα και χρήση συνώνυμων τιμών, καθώς και αρκετά άλλα. Μερικά από τα πιο συχνά θέματα ενός συνόλου δεδομένων είναι: Η ύπαρξη χαμένων τιμών, οι οποίες μπορεί να οφείλονται ενδεχομένως στο γεγονός ότι δεν ήταν διαθέσιμες κατά τον σχηματισμό του συνόλου δεδομένων, στο ότι έχουν διαγραφεί λόγω κάποιου σφάλματος ή επειδή δεν θεωρήθηκαν σημαντικές για να καταγραφούν. Ένα άλλο πρόβλημα που προκύπτει σχεδόν πάντα είναι η ύπαρξη θορύβου στα δεδομένα, όπου τα δεδομένα μπορεί να περιέχουν λανθασμένες τιμές, είτε λόγω λάθους από τον ανθρώπινο παράγοντα, είτε λόγω τεχνικών σφαλμάτων. Κάποια σφάλματα που ενδέχεται να συναντηθούν είναι η ύπαρξη ακραίων τιμών στα δεδομένα, όπου δύνανται να αποπροσανατολίσουν τους αλγορίθμους εξόρυξης και να τους οδηγήσουν σε μεροληπτικά ή ακόμα και εσφαλμένα αποτελέσματα (αντιθέτως, σε ορισμένες εργασίες, η εξόρυξη δεδομένων με ακραίες τιμές, αποτελεί αυτοσκοπό). Ένας άλλος όρος που υπάρχει στη βιβλιογραφία και χαρακτηρίζει τα δεδομένα με χαμένες τιμές, θόρυβο και άλλα προβλήματα είναι τα «ακάθαρτα δεδομένα». Η εκάστοτε εταιρεία που θέλει να εξασφαλίσει την ποιότητα των δεδομένων που έχουν εξορυχθεί, ορίζει κάποιους κανόνες, με σκοπό τον περιορισμό των προβλημάτων. Σύμφωνα με μελέτες, το ποσοστό που καταδεικνύει την ύπαρξη σφαλμάτων σε ένα σύνολο δεδομένων ανέρχεται στο 5%, ενώ το συνολικό ποσοστό των δεδομένων που μπορούν να χαρακτηριστούν ως ακάθαρτα, ανέρχεται στο 40%.

Η διαδικασία αντιμετώπισης των ακάθαρτων τιμών ενός συνόλου δεδομένων, ονομάζεται καθαρισμός δεδομένων και αποτελεί μέρος της προεπεξεργασίας δεδομένων. Ακόμη μια διαδικασία που αποτελεί μέρος της προεπεξεργασίας είναι η διακριτοποίηση, κατά την οποία τα δεδομένα απαιτείται να προσαρμοστούν στις απαιτήσεις των μεθόδων επεξεργασίας, όπως π.χ. σε περίπτωση που κάποια μέθοδος δεν μπορεί να διαχειριστεί αριθμούς και χρειάζεται ονομαστικές τιμές, τότε πρέπει να δεδομένα να μετατραπούν σε ονομαστικές τιμές. Επίσης υπάρχει και η κανονικοποίηση, η οποία συνήθως εμφανίζεται σε Νευρωνικά Δίκτυα στα οποία σε αρκετές περιπτώσεις, οι τιμές πρέπει να κυμαίνονται από το 0 μέχρι το 1, έτσι τροποποιούνται τα

δεδομένα για να μπορούν να χρησιμοποιηθούν. Η γενικότερη έννοια μετατροπής των δεδομένων, ορίζεται ως μετασχηματισμός. Μια άλλη εργασία, η οποία συντελείται κατά την προεπεξεργασία δεδομένων είναι η μείωση των δεδομένων, καθώς δεδομένα μεγάλου όγκου, μπορούν να προκαλέσουν προβλήματα στις μεθόδους επεξεργασίας και μεγάλες καθυστερήσεις στη εξαγωγή αποτελεσμάτων. Μια ειδική περίπτωση μείωσης του όγκου είναι η επιλογή των σημαντικών χαρακτηριστικών, κατά την οποία επιλέγονται μόνο τα χαρακτηριστικά, τα οποία περιέχουν χρήσιμες πληροφορίες, που αφορούν το αντικείμενο ανάλυσης.

Η προεπεξεργασία δεδομένων αποτελεί ένα αναπόσπαστο στάδιο της διαδικασίας της εξόρυξης γνώσης και στα πλαίσια της, μπορούν να χρησιμοποιηθούν όλα τα παραπάνω ή μόνο αυτά που είναι χρήσιμα για το εκάστοτε σύνολο δεδομένων και το σενάριο που μελετούν.

2.2.4. Υπερπροσαρμογή και Υποπροσαρμογή Δεδομένων

Ένας κίνδυνος που ελλοχεύει κατά την μηχανική μάθηση είναι η μηχανή να μάθει υπερβολικά καλά μια ομάδα δεδομένων και εκ του αποτελέσματος να είναι λιγότερο ακριβής στον έλεγχο. Η παραπάνω κατάσταση, ονομάζεται υπερπροσαρμογή δεδομένων (overfitting) και για να μπορέσει να αντιμετωπιστεί, χωρίζει τα δεδομένα σε δύο κομμάτια: Το κομμάτι της εκπαίδευσης δίνεται στη μηχανή και το κομμάτι ελέγχου, φυλάσσεται ώστε να χρησιμοποιηθεί σε κατοπινό στάδιο, για να γίνει έλεγχος κατά πόσο μπορεί να λειτουργήσει ένα μοντέλο, βασισμένο στα δεδομένα. Για την εξακρίβωση της ακρίβειας ενός μοντέλου, πρέπει να χρησιμοποιηθούν διαφορετικά δεδομένα από αυτά που χρησιμοποιήθηκαν κατά την εκπαίδευση. Μέσω της υπερπροσαρμογής, το μοντέλο εξετάζει περιπτώσεις, οι οποίες υπάρχουν και στα δεδομένα εκπαίδευσης, έχοντας έτσι παραπλανητικά αποτελέσματα, καθώς ένα μοντέλο το οποίο έχει υπερβολικά προσαρμοσμένα δεδομένα, ενδεχομένως να ενσωματώσει και τον θόρυβό τους ή ακόμα και να προβλέψει λανθασμένα την έξοδο νέων δειγμάτων. Όπως γίνεται κατανοητό, ένα υπερπροσαρμοσμένο μοντέλο πετυχαίνει πολύ υψηλές επιδόσεις έναντι των δεδομένων εκπαίδευσης, όμως δεν ισχύει και το ίδιο ως προς άγνωστες παρατηρήσεις. Έτσι, η εξακρίβωση της ακρίβειας ως προς το μοντέλο εκπαίδευσης, δεν αποτελεί ασφαλές μέτρο επιτυχίας ενός μοντέλου, αλλά αντιθέτως μπορεί να αποτελέσει ένδειξη πιθανής υπερπροσαρμογής του. Το ακριβώς αντίθετο φαινόμενο της υπερπροσαρμογής είναι η υποπροσαρμογή, στην περίπτωση της οποίας το μοντέλο είναι υπερβολικά απλό για να ενσωματώσει ουσιαστικές σχέσεις, με αποτέλεσμα στις πλείστες των περιπτώσεων να έχει χαμηλή ακρίβεια ως προς τα δεδομένα εκπαίδευσης, αλλά και ως προς τις άγνωστες παρατηρήσεις.

3. Γράφοι

3.1. Ορισμοί

Γράφος ορίζεται ως μια αφηρημένη αναπαράσταση ενός συνόλου δεδομένων, η οποία αποτελείται από μια συλλογή από κόμβους και ακμές [12]. Οι ακμές αποτελούν το σύνολο των ζευγών των κόμβων, που συνδέονται μεταξύ τους. Τα ζεύγη (V,E) συνθέτουν τον γράφο G , όπου V οι κόμβοι και E οι ακμές, αποτελούν πεπερασμένα σύνολα, με το V να είναι μη-κενό σύνολο n στοιχείων και το E σύνολο μη-διατεταγμένων ζευγών με στοιχεία του συνόλου V .

Οι γράφοι μπορούν να χαρακτηριστούν ως κατευθυνόμενοι, αν κάθε μια από τις ακμές τους είναι προσανατολισμένη προς μια κατεύθυνση. Αντιστοίχως, χαρακτηρίζονται ως μη-κατευθυνόμενοι αν οι ακμές τους δεν είναι προσανατολισμένες. Το πλήθος των ακμών που έχουν σαν αφετηρία ένα κόμβο, ορίζεται ως ο βαθμός εξόδου του κόμβου. Το πλήθος των βελών που καταλήγουν σε ένα κόμβο, ορίζεται ως ο βαθμός εισόδου του κόμβου. Οι κόμβοι που αποτελούν μια ακμή, ονομάζονται γειτονικοί. Κάθε γράφος αποτελείται από διάφορα μονοπάτια, τα οποία ορίζονται ως μια ακολουθία κόμβων οι οποίοι συνδέονται μεταξύ τους με ακμές. Ως μήκος του μονοπατιού, ορίζεται ο αριθμός των ακμών που περιέχει. Επιπροσθέτως, κύκλος σε γράφο, ονομάζεται μια διαδρομή με μήκος > 1 , όπου ο κόμβος από τον οποίο ξεκινά ένα μονοπάτι, αποτελεί ταυτοχρόνως τόσο την αφετηρία, όσο και την λήξη. Ένας γράφος, μπορεί να αποτελείται από διάφορους υπογράφους, οι οποίοι αποτελούν υποσύνολο του αρχικού γράφου και κατ' επέκταση έχουν υποσύνολα ακμών και κόμβων. Ένας μη κατευθυνόμενος γράφος, λέγεται συνεκτικός αν για κάθε ζευγάρι κορυφών, υπάρχει διαδρομή που τις συνδέει. Από την άλλη, κάθε κατευθυνόμενος γράφος που ικανοποιεί την ίδια προϋπόθεση, ονομάζεται ισχυρά συνεκτικός. Αν ο μη κατευθυνόμενος γράφος στον οποίο αντιστοιχεί είναι συνεκτικός, τότε ο γράφος είναι ελαφρά συνεκτικός.

3.2. Αναπαράσταση Δεδομένων Γράφου

3.2.1 Πίνακας Γειτνίασης

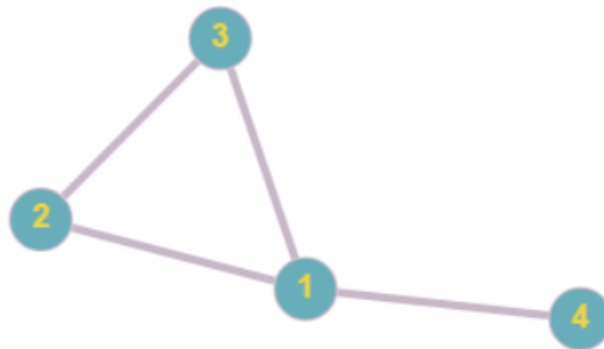
Η δομή του γράφου, αναπαρίσταται μέσω ενός πίνακα γειτνίασης, μεγέθους $[i \times j]$, όπου i και j ο αριθμός κόμβων. Η κάθε γραμμή του πίνακα, αποτελείται από δυαδικές τιμές, οι οποίες επιλέγονται ως εξής: Στην περίπτωση που οι δύο κόμβοι συνδέονται, συνεπάγεται ότι υπάρχει ακμή μεταξύ τους, άρα στην αντίστοιχη θέση του πίνακα, η τιμή είναι ίση με 1. Στην αντίθετη περίπτωση, όπου δύο κόμβοι δεν συνδέονται, σημαίνει ότι δεν υπάρχει ακμή μεταξύ τους, άρα στην αντίστοιχη θέση του πίνακα, η τιμή είναι ίση με 0. Ως προς τον πίνακα γειτνίασης, ισχύει ότι εάν ένα ζεύγος κόμβων συνδέεται, τότε η τιμή τίθεται ίση με 1, τόσο στη θέση όπου ισχύει $i-j$, όσο και στη θέση όπου έχουμε $j-i$.

Εικόνα 5: Συνθήκη συμπλήρωσης πίνακα γειτνίασης

$$A_{IJ} = \begin{cases} 1, & \text{αν } (i,j) \text{ ανήκουν στα ζεύγη των ακμών} \\ 0, & (i,j) \text{ δεν ανήκουν στα ζεύγη των ακμών} \end{cases}$$

Αν ο γράφος έχει βάρη, τότε οι δυαδικές τιμές 0,1, αντικαθίστανται από τα αντίστοιχα βάρη. Η απαιτούμενη μνήμη για την αποθήκευση του πίνακα A είναι $\Theta(n^2)$. Έτσι, όταν ο αριθμός των κόμβων σε ένα γράφο είναι μεγάλος και ο αριθμός των ακμών ανά κόμβο είναι μεταβλητός, ο αντίστοιχος πίνακας γειτνίασης είναι αραιός και μη αποδοτικός ως προς το χώρο.

Εικόνα 6: Γράφος Αναπαράστασης Δεδομένων



Εικόνα 7: Πίνακας Γειτνίασης

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 1 |
| 2 | 1 | 0 | 1 | 0 |
| 3 | 1 | 1 | 0 | 0 |
| 4 | 1 | 0 | 0 | 0 |

3.2.2. Λίστα Γειτνίασης

Ένας γράφος, μπορεί να αναπαρασταθεί και ως μια λίστα γειτνίασης ως εξής: Ο γράφος μπορεί να αναπαρίσταται ως ένας μονοδιάστατος πίνακας A. Για κάθε κορυφή ω , $A[\omega]$ αποτελεί ένα δείκτη σε μια συνδεδεμένη λίστα, στην οποία αποθηκεύονται οι κορυφές που γειτνιάζουν με την ω . Η λίστα γειτνίασης, αποτελεί ένα πιο κομψό και αποδοτικό τρόπο αναπαράστασης γράφων, οι οποίοι είναι αραιοί, δηλαδή έχουν μικρό μήκος ακμών. Σε σχέση με τον πίνακα γειτνίασης, ο τομέας στον οποίο υστερούν οι λίστες γειτνίασης είναι ότι μπορεί να απαιτήσουν μέχρι και $\Theta(|V|)$ συγκρίσεις, στην περίπτωση που πρέπει να αναζητήσουμε τη λίστα των κόμβων που είναι γειτονικές, ενώ στον πίνακα γειτνίασης η χρονική πολυπλοκότητα είναι $O(1)$,

καθώς χρειάζεται απλώς να εξεταστεί μια καταχώρηση του πίνακα. Παρακάτω στην Εικόνα 8, παρατίθεται μια λίστα γειτνίασης, βασισμένη στο παραπάνω γράφημα.

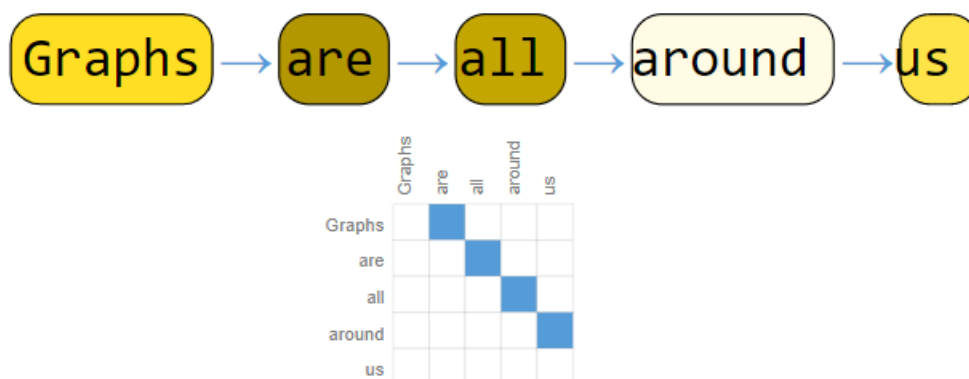
Εικόνα 8: Λίστα Γειτνίασης

| Κόμβος | Γειτονικοί Κόμβοι |
|--------|-------------------|
| 1 | 2,3,4 |
| 2 | 1,3 |
| 3 | 1,2 |
| 4 | 1 |

3.3 Χρήση Γράφων για αναπαράσταση δεδομένων

Τα γραφήματα είναι δομές που έχουν μελετηθεί ευρέως και που χρησιμοποιούνται για να μοντελοποιήσουν διάφορους τύπους οντοτήτων και τις μεταξύ τους σχέσεις. Τα τελευταία χρόνια, οι αναπαραστάσεις δεδομένων ως γραφήματα έχουν διαδοθεί σε μεγάλο βαθμό και χρησιμοποιούνται σε πολλές εφαρμογές. Για παράδειγμα, στην υπολογιστική χημεία, τα γραφήματα χρησιμοποιούνται για την αναπαράσταση μορίων, ενώ στην επεξεργασία φυσικής γλώσσας χρησιμοποιούνται για την αναπαράσταση κειμένων. Σε πολλές εφαρμογές, είναι απαραίτητο να εκτελέσουμε εργασίες μηχανικής μάθησης πάνω σε γραφήματα όπως για παράδειγμα να ταξινομήσουμε γραφήματα ή να ομαδοποιήσουμε γραφήματα. Στην Εικόνα 9 παρουσιάζεται ένα παράδειγμα αναπαράστασης κειμένου σε γράφο:

Εικόνα 9: Μετατροπή κειμένου σε Πίνακα Γειτνίασης



4. Νευρωνικά Δίκτυα Γράφων

4.1 Βασικές Έννοιες

4.1.1. Βασικά Χαρακτηριστικά

Τα νευρωνικά δίκτυα έχουν προσαρμοστεί, ώστε να έχουν την δυνατότητα να αξιοποιούν την δομή και τις ιδιότητες των γράφων. Κατ' επέκταση, τα νευρωνικά δίκτυα γράφων (Graph Neural Networks – GNNs) αναπαριστούν τα δεδομένα σε δομές γνωστές ως ενσωματώσεις – διανύσματα (embeddings) για κάθε κόμβο, τα οποία θα αντιπροσωπεύουν τα περιεχόμενα – δεδομένα, κάθε κόμβου [13]. Οι ενσωματώσεις χρησιμοποιούνται σε λειτουργίες προβλέψεων, όπως και στην περίπτωση μας που αφορά την πρόβλεψη της ετικέτας ενός κόμβου. Η δομή του γράφου, καθώς και τα χαρακτηριστικά κάθε κόμβου, θα αξιοποιηθούν για να αναπαράχθουν στο μέγεθος της ενσωμάτωσης που έχει προκαθοριστεί από μια συγκεκριμένη υπερπαράμετρο. Οι κόμβοι συνδέονται λόγω της συσχέτισεως που έχουν μεταξύ τους ή λόγω του ότι είναι παρόμοιοι, φαινόμενο γνωστό και σαν homophily. Τα κύρια δομικά στοιχεία ενός νευρωνικού δικτύου γράφου είναι τα επίπεδα μεταφοράς μηνυμάτων (message passing layer), τα οποία συλλέγουν πληροφορίες από τους γειτονικούς κόμβους και την συνδυάζουν σε ενσωματώσεις και ενημερώνουν με τα χαρακτηριστικά του κόμβου, διαδικασία η οποία αφορά τη συνέλιξη του δικτύου.

4.1.2. Είδη προβλέψεων μέσω Νευρωνικών Δικτύων Γράφων

Τα είδη των προβλέψεων που επιτελούν τα Νευρωνικά Δίκτυα Γράφων είναι τα εξής: Πρόβλεψη αναφορικά με τους κόμβους (Node-level prediction), η οποία μπορεί να απαιτεί την εύρεση ιδιοτήτων συγκεκριμένων κόμβων ή ακόμα την κατηγοριοποίηση του κάθε κόμβου. (Classification) Έπειτα υπάρχει και η πρόβλεψη αναφορικά με τις ακμές (Link-level prediction), η οποία μπορεί να εξετάζει τη συσχέτιση μεταξύ κόμβων του γράφου ή ακόμα και την πρόβλεψη των ιδιοτήτων των συσχέτισεων. Επιπροσθέτως, υπάρχει και η πρόβλεψη αναφορικά με ολόκληρο τον γράφο (Graph-level prediction), η οποία προβλέπει την ιδιότητα ολόκληρου του γράφου ή ακόμα και την κατηγοριοποίησή του.

4.1.3. Χρήση Νευρωνικών Δικτύων Γράφων στην Μηχανική Μάθηση

Αρχικά, οι γράφοι πρέπει να αναπαρασταθούν με τέτοιο τρόπο, ώστε να είναι συμβατοί με τα νευρωνικά δίκτυα. Τα νευρωνικά δίκτυα, συνήθως λαμβάνουν ως είσοδο ορθογώνιους ή πλεγματοειδείς πίνακες, πράγμα το οποίο καθιστά αρκετά δύσκολη την διαδικασία αναπαράστασής τους σε μια μορφή συμβατή με την βαθιά μάθηση. Οι γράφοι, αποτελούνται από έως και τέσσερις τύπους πληροφοριών, οι οποίες ενδέχονται να αξιοποιηθούν για να εξαχθούν οι προβλέψεις: τους κόμβους, τις ακμές, το περιεχόμενο και την συνδεσιμότητα. Οι τρεις πρώτες δομές είναι πιο εύκολο να μορφοποιηθούν, χωρίς να χρειάζονται ιδιαίτερες τεχνικές: Ξεκινώντας από τους κόμβους, σχηματίζεται ένας πίνακας χαρακτηριστικών (node features) τύπου λεξικού, ο

οποίος αντιστοιχίζει ένα δείκτη για κάθε κόμβο σαν το κλειδί του λεξικού και τους δείκτες των κόμβων οι οποίοι συσχετίζονται μαζί του σαν οι τιμές του λεξικού. Οι σχέσεις μεταξύ των κόμβων όπως έχει προαναφερθεί, αναπαρίστανται μέσω των πινάκων γειτνίασης. Από την άλλη, η αναπαράσταση της συνδεσιμότητας ενός γράφου είναι ένα πολύ πιο περίπλοκο θέμα, καθώς η ιδέα χρήσης πίνακα γειτνίασης δεν είναι η βέλτιστη. Ο πίνακας γειτνίασης εξαρτάται άμεσα από τον αριθμό των κόμβων του γράφου, κάτι το οποίο είναι αρκετά προβληματικό, καθώς σε περίπτωση που ο αριθμός των κόμβων είναι της τάξης του εκατομμυρίου, απαιτεί αρκετό χρόνο. Επίσης, ο αριθμός των ακμών ανά κόμβο είναι μεταβλητός, κάτι το οποίο σε συνδυασμό με τα προαναφερθέντα, μπορεί να οδηγήσει σε αραιούς γράφους, οι οποίοι είναι αναποτελεσματικοί ως προς τον χώρο. Επιπροσθέτως, υπάρχουν πολλαπλοί πίνακες γειτνίασης που μπορούν να αναπαραστήσουν την ίδια συνδεσιμότητα, αλλά δεν σημαίνει απαραίτητα ότι θα παράγουν το ίδιο αποτέλεσμα σε ένα βαθύ νευρωνικό δίκτυο. Ένας κομψός και αποδοτικός τρόπος αναπαράστασης αραιών πινάκων είναι οι λίστες γειτνίασης, που αναπαριστούν τη συνδεσιμότητα μιας ακμής ως προς το ζεύγος κόμβων που συνδέει. Ο αριθμός των ακμών θα είναι πολύ μικρότερος από τον αριθμό συνδέσεων κόμβο σε κόμβο και έτσι αναπαρίστανται μόνο τα κομμάτια που συνδέονται μεταξύ τους.

4.2. Τρόπος λειτουργίας των Νευρωνικών Δικτύων Γράφου

Έχοντας διευκρινίσει την μορφή που πρέπει να πάρουν οι δομές του γράφου, για να μπορούν να χρησιμοποιηθούν στα νευρωνικά δίκτυα, θα περιγράψουμε την χρήση νευρωνικών δικτύων γράφων για την επίλυση προβλημάτων πρόβλεψης. Το νευρωνικό δίκτυο γράφων είναι ένας βελτιστοποιημένος μετασχηματισμός για όλα τα χαρακτηριστικά του γραφήματος, ο οποίος διατηρεί και τις συμμετρίες του γράφου. Το νευρωνικό δίκτυο γράφων χρησιμοποιεί δίκτυο μεταφοράς μηνυμάτων, το οποίο είχε προτείνει ο Gilmer, χρησιμοποιώντας την αρχιτεκτονική που προτάθηκε από τον Battaglia. Χρησιμοποιούν μια αρχιτεκτονική «graph-in, graph-out», που σημαίνει ότι δέχονται γράφο σαν είσοδο, με τις απαραίτητες πληροφορίες στους κόμβους του, ακμές και περιεχόμενο και σταδιακά τα μετατρέπουν σε ενσωματώσεις, χωρίς να επηρεάζουν την συνδεσιμότητα του γράφου.

4.2.1. Προβλέψεις μέσω Pooling

Στην περίπτωση της ταξινόμησης, για κάθε διάνυσμα κόμβου, εφαρμόζεται ένας γραμμικός ταξινομητής. Εάν όμως, υπάρχουν πληροφορίες στις ακμές και όχι στους κόμβους, τότε χρειάζεται κάποιος τρόπος για να συλλεχθούν οι πληροφορίες από τις ακμές και να χρησιμοποιηθούν στους κόμβους για ταξινόμηση. Η διαδικασία αυτή ονομάζεται pooling και μπορεί να εφαρμοστεί τόσο σε κόμβους, όσο και σε ακμές και έχει ως εξής:

- Μαζεύονται τα διανύσματα και ενώνονται σε ένα πίνακα.
- Τα ενωμένα διανύσματα αθροίζονται, μέσω συνάρτησης.

4.2.2 Διάδοση μηνυμάτων

Υπάρχει η δυνατότητα για προβλέψεις με συνδυασμό pooling με GNN επίπεδο, μέσω της διάδοσης μηνυμάτων, όπου κάθε κόμβος αλληλεπιδρά με τους γειτονικούς του κόμβους, δηλαδή αυτούς που συνδέεται και ανταλλάζουν «πληροφορίες» μεταξύ τους, επηρεάζοντας άμεσα τις ενημερωμένες ενσωματώσεις. Η διαδικασία διάδοσης μηνυμάτων, έχει ως εξής:

- Για κάθε κόμβο στο γράφο, συγκεντρώνονται όλες οι πληροφορίες των γειτονικών του κόμβων, μέσω των διανυσμάτων τους.
- Αθροίζονται όλα τα διανύσματα, μέσω συνάρτησης.
- Όλα τα pooled μηνύματα – ενσωματώσεις, διαδίδονται μέσω μιας συνάρτησης ενημέρωσης, που ονομάζεται νευρωνικό δίκτυο στο οποίο έγινε η εκμάθηση.

Η διάδοση μηνυμάτων, μπορεί επίσης να εφαρμοστεί τόσο σε κόμβους, όσο και σε ακμές και μέσω των παραπάνω διαδικασιών, αλιεύεται η συνδεσιμότητα του γράφου. Η διάδοση μηνυμάτων και η συνέλιξη (convolution) είναι λειτουργίες για το άθροισμα επεξεργασία των πληροφοριών των γειτόνων, για να μπορούν να παραχθούν οι ενημερωμένες ενσωματώσεις. Στοιβάζοντας επίπεδα διάδοσης μηνυμάτων, κάθε κόμβος μπορεί να ενσωματώσει πληροφορίες από ολόκληρο τον γράφο. Κάθε επίπεδο, σημαίνει ότι ο κόμβος μπορεί να πάρει πληροφορίες - ευθέως ανάλογες με τον αριθμό των βημάτων – μακριά από αυτόν.

4.2.3. Καθολικές Αναπαραστάσεις

Τα δίκτυα που περιγράψαμε πιο πάνω, έχουν εμφανή ελαττώματα. Αρχικά, οι κόμβοι, ενδέχονται να βρίσκονται μακριά ο ένας προς τον άλλο στο γράφημα, πράγμα το οποίο στην πράξη υπάρχει περίπτωση να σημαίνει ότι δεν θα είναι σε θέση να μεταφέρουν αποτελεσματικά τις πληροφορίες, ακόμα και αν υπάρχουν πολλαπλά επίπεδα διάδοσης μηνυμάτων. Όπως αναφέρθηκε πιο πάνω, εάν έχουμε x επίπεδα, οι πληροφορίες θα μεταφερθούν x -βήματα μακριά, πράγμα το οποίο μπορεί να είναι ελαττωματικό, εάν η πρόβλεψη εξαρτάται από κόμβους ή ομάδες κόμβων που απέχουν πολύ μεταξύ τους. Μια λύση είναι να μπορούν οι κόμβοι να μεταφέρουν μηνύματα ο ένας στον άλλο, πράγμα το οποίο δυστυχώς είναι υπολογιστικά ακριβό. Η λύση του προβλήματος αυτού είναι η χρήση καθολικής αναπαράστασης, που ονομάζεται κύριος κόμβος. Αυτό το καθολικό διάνυσμα περιβάλλοντος συνδέεται με όλους τους άλλους κόμβους και ακμές του δικτύου και μπορεί να λειτουργήσει ως γέφυρα μεταξύ τους για τη μετάδοση πληροφοριών, δημιουργώντας μια αναπαράσταση για το γράφημα πιο πλούσια και πιο σύνθετη.

4.3. Διαφορετικοί τύποι γράφων

Οι δομές γράφων είναι πιο ευέλικτες και μπορούν να αναπαραστήσουν διάφορους τύπους πληροφοριών. Η διάδοση μηνυμάτων είναι επίσης αρκετά ευέλικτη διαδικασία, ώστε η προσαρμογή των νευρωνικών δικτύων γράφων σε πιο σύνθετες δομές γραφημάτων, να αφορά τον

τρόπο μετάδοσης των πληροφοριών και πως ενημερώνονται από νέα χαρακτηριστικά του γράφου. Όταν θελήσουμε να μοντελοποιήσουμε τις αλληλεπιδράσεις μεταξύ κόμβων διαφορετικού τύπου, χρησιμοποιούνται γραφήματα πολλαπλών τύπων ακμών (multi-edge graphs), όπου ένα ζεύγος κόμβων μπορεί να μοιράζεται διαφορετικούς τύπους ακμών, όπως συμβαίνει για παράδειγμα στους τύπους σχέσεων σε ένα γράφο που μοντελοποιεί ένα γενεαλογικό δέντρο (αδέρφια, γονείς, θείοι, κτλ.). Επίσης υπάρχουν και οι ένθετοι γράφοι (nested graphs), οι οποίοι είναι χρήσιμοι για την αναπαράσταση ιεραρχικών πληροφοριών και ονομάζονται αλλιώς σαν γράφοι υπερκόμβων, όπως για παράδειγμα σε ένα δίκτυο χημικών στοιχείων, όπου κάθε κόμβος αντιπροσωπεύει ένα στοιχείο και συνδέεται με άλλο κόμβο εάν μπορούν να συνθέσουν ένα φάρμακο. Ένας άλλος τύπος γράφου είναι ένας υπεργράφος, όπου μια ακμή μπορεί να συνδεθεί με πολλούς κόμβους αντί για μόνο δύο. Παραδείγματος χάριν, για ένα δεδομένο γράφημα, μπορούμε να δημιουργήσουμε έναν υπεργράφο προσδιορίζοντας κοινότητες κόμβων και αναθέτοντας μια υπερ-ακμή που συνδέεται με όλους τους κόμβους μιας κοινότητας.

4.4. Υποκατηγορίες Νευρωνικών Δικτύων Γράφου

4.4.1. Νευρωνικό δίκτυο γράφων με ξεχωριστό MLP (Multilayer perceptron)

Έχοντας μετατρέψει τις πληροφορίες του γράφου σε διανύσματα, πλέον υπάρχει η δυνατότητα δημιουργίας νευρωνικού δικτύου γράφου. Η πιο απλή αρχιτεκτονική νευρωνικού δικτύου γράφου είναι αυτή κατά την οποία μαθαίνονται νέες ενσωματώσεις για όλα τα χαρακτηριστικά του γραφήματος, αλλά δεν χρησιμοποιείται η συνδεσιμότητα. Χρησιμοποιεί ένα MLP σε κάθε στοιχείο του γράφου, αποτελώντας έτσι το GNN επίπεδο (layer) [14]. Για κάθε διάνυσμα κόμβου, εφαρμόζεται το MLP και παίρνουμε πίσω ένα εκπαιδευμένο πλέον διάνυσμα κόμβου. Το ίδιο συμβαίνει και για κάθε ακμή, καθώς και το περιεχόμενο, μαθαίνοντας έτσι μια ενσωμάτωση για ολόκληρο το γράφημα, στοιβάζοντας αυτά τα τρία επίπεδα. Το γράφημα εξόδου που προκύπτει έχει ενημερωμένες ενσωματώσεις, για τις δομές που εφαρμόστηκε το MLP, διατηρώντας ίδιο την λίστα γειτνίασης (αφού δεν έχει χρησιμοποιηθεί η συνδεσιμότητα).

4.4.2. Συνελικτικό Δίκτυο Γράφου

Μπορούμε να σκεφτούμε τους γράφους σαν να κωδικοποιούν μια μορφή ακανόνιστης χωρικής δομής και επομένως, κρίνεται αναγκαίο αν μπορούσαμε με κάποιο τρόπο να γενικεύσουμε τον συνελκτικό τελεστή (όπως χρησιμοποιείται στα CNN) για να λειτουργεί σε γράφους. Τα CNN αποτελούν σημαντικό εργαλείο όσον αφορά την εργασία με δεδομένα εικόνας. Εκμεταλλεύονται το γεγονός ότι οι εικόνες έχουν ένα εξαιρετικά άκαμπτο και κανονικό μοτίβο συνδεσιμότητας, όπου κάθε pixel «συνδέεται» με τα οκτώ γειτονικά του pixels. Τα αυθαίρετα γραφήματα είναι μια πολύ πιο δύσκολη πρόκληση, καθώς ιδανικά, θα θέλαμε να συγκεντρώνουμε πληροφορίες σε κάθε γειτονιά των κόμβων, αλλά δεν εξασφαλίζεται η ακαμψία δομής. Ένα ιδανικό συνελκτικό επίπεδο γραφήματος έχει τα εξής χαρακτηριστικά: Έχει υπολογιστική αποτελεσματικότητα αποθήκευσης που δεν απαιτεί περισσότερο από $O(V+E)$ χρόνο και μνήμη. Επιπρόσθετα έχει σταθερό αριθμό παραμέτρων, ανεξάρτητο από το μέγεθος εισόδου του γράφου.

Δύναται να ενεργεί σε μια τοπική γειτονιά ενός κόμβου και να καθορίζει τη σημασία σε διαφορετικούς γείτονες. Επίσης, σημαντική είναι και η δυνατότητα εφαρμογής σε επαγωγικά προβλήματα. Ωστόσο, η ταυτόχρονη ικανοποίηση όλων των παραπάνω έχει αποδειχθεί αρκετά προκλητική και, πράγματι, καμία από τις προηγούμενες τεχνικές δεν ήταν επιτυχής στο να τα επιτύχει ταυτόχρονα.

Για να κατανοήσουμε τη διαδικασία συνέλιξης του γράφου: Αρχικά, θεωρούμε ένα γράφο n κόμβων, ο οποίος αποτελείται από σεν χαρακτηριστικών κόμβων και από τη συνήθη δομή του πίνακα γειτνίασης. Ένα επίπεδο συνέλιξης γράφου, υπολογίζει τα νέα χαρακτηριστικά κάθε κόμβου, βασισμένο στα χαρακτηριστικά εισόδου, καθώς και στη δομή του γράφου. Κάθε συνελκτικό επίπεδο του γράφου, ξεκινά με ένα κοινό μετασχηματισμό χαρακτηριστικών ανά κόμβο, προκειμένου να επιτευχθεί η αναπαράσταση σε υψηλότερο επίπεδο, που καθορίζεται από πίνακα βάρους W . Έτσι μετατρέπονται τα διανύσματα χαρακτηριστικών ως εξής:

Εικόνα 10: Χαρακτηριστικά εξόδου του κόμβου i

$$H^{(l+1)} = \sigma \left(\frac{\hat{A} H^l W^l}{d_i d_j} \right)$$

όπου \hat{A} ο πίνακας γειτνίασης του γράφου που μελετάται, ο οποίος περιέχει και συνδέσεις - ακμές κάθε κόμβου με τον εαυτό του. Το W^l αποτελεί τον πίνακα μετασχηματισμού του εκάστοτε επιπέδου και το H^l τα χαρακτηριστικά του εκάστοτε κόμβου στο τρέχον επίπεδο. Αξίζει να σημειωθεί ότι το H^0 αρχικοποιείται με τον πίνακα X , ο οποίος περιέχει διανύσματα για κάθε κόμβο. Τα d_i και d_j , αποτελούν τον αριθμό των συνδέσεων του εκάστοτε κόμβου, με τους άλλους που περιέχονται στο γράφο. Τέλος, σ αποτελεί την συνάρτηση ενεργοποίησης, η οποία είναι σιγμοειδής.

4.4.3 Δίκτυο Γράφου Προσοχής

Στο δίκτυο γράφου προσοχής, χρησιμοποιείται το self-attention για να προσδιοριστεί η σημασία του κόμβου j στον κόμβο i . Το self-attention έχει αποδειχθεί ότι είναι αποτελεσματικό για state-of-the-art αποτελέσματα, όπως αποδεικνύεται και από την αρχιτεκτονική των Transformers, σύμφωνα με τον Vaswani. Το a_{ij} υπολογίζεται μέσω του μηχανισμού προσοχής, που υπολογίζει μη κανονικοποιημένους συντελεστές e_{ij} για κάθε ζεύγος κόμβων i, j , βασισμένο στα χαρακτηριστικά τους.

Εικόνα 11: Μη κανονικοποιημένοι συντελεστές

$$e_{ij} = a(\bar{h}_i \bar{h}_j)$$

Χρησιμοποιούμε την δομή του γράφου, χρησιμοποιώντας μόνο τους κόμβους που συνδέονται μεταξύ τους. Οι συντελεστές που προκύπτουν, κανονικοποιούνται με τη χρήση της συνάρτησης softmax, για να είναι συγκρίσιμοι σε όλες τις γειτονιές.

Εικόνα 12: Κανονικοποιημένοι συντελεστές

$$a_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})}$$

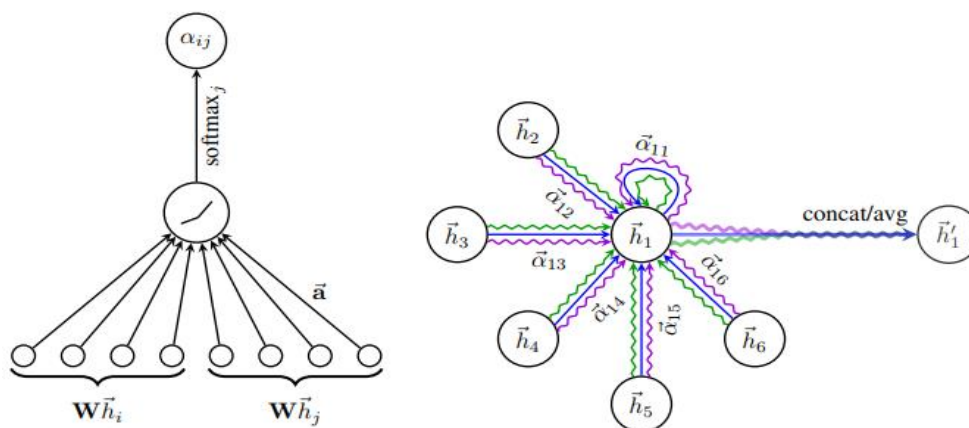
Για τη σταθεροποίηση της μάθησης κατά το self-attention, χρησιμοποιείται multi-head attention, όπου οι πράξεις εκτελούνται K διαφορετικές και ανεξάρτητες φορές, με διαφορετικές παραμέτρους και τα χαρακτηριστικά αθροίζονται ως προς το κάθε κεφάλι.

Εικόνα 13: Χαρακτηριστικά εξόδου

$$\vec{h}_i' = \parallel_{k=1}^K \sigma \left(\sum_{j \in N_i} a_{ij}^k W^k \vec{h}_j \right)$$

όπου a_{ij}^k οι συντελεστές προσοχής που προκύπτουν από το k κεφάλι και από το W^k πίνακα βαρών, ο οποίος προσδιορίζεται από τη γραμμική μεταμόρφωση στο k κεφάλι.

Εικόνα 14: Single-head attention(δεξιά) και Multi-head attention(αριστερά)



4.5. Μεταγωγική – Επαγωγική Μάθηση

Στην προσέγγιση της επαγωγικής μάθησης, ένα μοντέλο I_w μαθαίνεται, προσαρμόζοντας τα βάρη του w με βάση ένα σύνολο δεδομένων με ετικέτες, δηλαδή στις πλείστες φορές το σύνολο εκπαίδευσης. Κάθε παράδειγμα επεξεργάζεται ανεξάρτητα από τα άλλα, όμως τα γενικά στατιστικά επιτρέπουν στον αλγόριθμο μάθησης να εξάγει ένα γενικό μοντέλο για την επίλυση της απαιτούμενης εργασίας. Η πρόβλεψη μοντέλου βασίζεται μόνο στα χαρακτηριστικά που περιγράφουν κάθε διαφορετική είσοδο. Μόλις μάθει το μοντέλο, νέες μη ορατές εισόδους μπορούν να υποβληθούν σε επεξεργασία, μία τη φορά, για να υπολογιστεί η έξοδος του μοντέλου, δηλαδή η προβλεπόμενη κλάση. Αντιθέτως, στη μεταγωγική μάθηση, ο αλγόριθμος έχει σχεδιαστεί για να μπορεί να μελετήσει, τόσο τα παραδείγματα με ετικέτες, όσο και χωρίς, εκμεταλλευόμενη τις σχέσεις μεταξύ διαφορετικών δειγμάτων. Οι σχέσεις μεταξύ των δεδομένων μπορούν να αξιοποιηθούν είτε στη φάση της μάθησης, είτε στη φάση της πρόβλεψης ή ακόμα και στα δύο. Η πρόβλεψη για τα δεδομένα χωρίς ετικέτα, γίνεται μέσω της διάδοσης των πληροφοριών που είναι διαθέσιμες για τα κοντινά παραδείγματα, μέσω των δοθέντων μεταξύ τους σχέσεων. Για παράδειγμα εάν το n είναι παράδειγμα κατά το test time, τότε οι στόχοι που είναι διαθέσιμοι στη γειτονιά του, μπορούν να αξιοποιηθούν μαζί με τα τοπικά χαρακτηριστικά του n , ως εισόδοι για να υπολογιστεί η έξοδος. Η μεταγωγική μάθηση είναι ιδιαίτερα χρήσιμη, όταν έχουμε μικρό σύνολο δεδομένων με ετικέτες, που προέρχονται από μια άγνωστη στοχαστική διαδικασία και όμως ένα μικρό δείγμα δεν μπορεί να είναι στατιστικά σημαντικό για την επαγωγή ενός γενικού κανόνα πρόβλεψης, που βασίζεται μόνο σε τοπικά χαρακτηριστικά. Η μεταγωγική μάθηση σύμφωνα με τη βιβλιογραφία [15] είναι βασισμένη στους γράφους. Τα τελευταία χρόνια αυτές οι μέθοδοι εφαρμόζονται σε πολλούς τομείς, χάρη στην ικανότητά τους να προσαρμόζονται σε διαφορετικές εφαρμογές του πραγματικού κόσμου, όπως στην επεξεργασία φυσικής γλώσσας, επιτήρηση και ανακατασκευή πειραμάτων.

Τα νευρωνικά δίκτυα γράφων μπορούν να χρησιμοποιηθούν σε συνδυασμό με επαγωγική προσέγγιση ή ακόμα και με ένα συνδυασμό επαγωγικής – μεταγωγικής. Παίρνοντας ως είσοδο ένα γράφο $G = (V, E)$, το σύνολο των κόμβων V μπορεί να διαχωριστεί σε ένα σύνολο κόμβων με εποπτεία (S) και σύνολο μη εποπτευόμενων κόμβων (U). Όταν χρησιμοποιείται μια καθαρά επαγωγική προσέγγιση, δίνεται στο νευρωνικό δίκτυο γράφων ως είσοδος μια ή ακόμα και περισσότερες περιπτώσεις του γράφου που πρόκειται να μάθει και οι στόχοι του για τους εποπτευόμενους κόμβους στο S , που χρησιμοποιούνται μόνο για την εκμάθηση των παραμέτρων του νευρωνικού δικτύου. Το εκπαιδευμένο νευρωνικό δίκτυο γράφων μπορεί να αξιοποιηθεί για την επεξεργασία τόσο των αρχικών γράφων στο σύνολο εκμάθησης για τον υπολογισμό των προβλέψεων εξόδου για τους κόμβους του U που δεν έχουν ετικέτες και για την επεξεργασία μη ορατών γραφημάτων χωρίς εποπτευόμενους κόμβους. Θα πρέπει να σημειωθεί ότι και σε αυτή την περίπτωση το νευρωνικό δίκτυο γράφων εκμεταλλεύεται την τοπολογία των σχέσεων μεταξύ των κόμβων, μέσω της διαδικασίας διάχυσης που χρησιμοποιείται για τον υπολογισμό των καταστάσεων. Οι κόμβοι στα S και U , εμπλέκονται σε αυτό τον υπολογισμό, αλλά δεν αξιοποιούνται οι πληροφορίες σχετικά με τους στόχους των κόμβων στο S . Σε αυτή

την περίπτωση, δεν θεωρείται σωστή διαδικασία επαγωγικής μάθησης, καθώς κατά τον χρόνο δοκιμής μόνο τα χαρακτηριστικά του κόμβου και το περιβάλλον του στην πολλαπλότητα δεδομένων επηρεάζουν το αποτέλεσμα του υπολογισμού.

Στα επαγωγικά – μεταγωγικά GNN, υποθέτουμε ότι εμπλουτίζουμε τα χαρακτηριστικά του κόμβου με την ετικέτα στόχο, έτσι ώστε να αξιοποιείται στη διαδικασία της διάχυσης, προσφέροντας άμεση μεταγωγική συνεισφορά. Ο τρόπος με τον οποίο οι στόχοι διαχέονται και συμβάλουν στις τελικές εξόδους – αποτελέσματα, μαθαίνεται μέσω παραδειγμάτων. Υποθέτουμε ότι το σύνολο εκμάθησης, περιέχει μερικώς εποπτευόμενους γράφους και για κάθε γράφο. Χωρίζουμε το σύνολο των εποπτευόμενων κόμβων σε δύο υποσύνολα: το σύνολο των κόμβων που χρησιμοποιούνται για τον υπολογισμό της απώλειας και το σύνολο των μεταγωγικών κόμβων. Για το τελευταίο υποσύνολο, ο διαθέσιμος στόχος συνδέεται με τα διανύσματα χαρακτηριστικών εισόδου, ενώ για τους κόμβους που υπολογίζουν την απώλεια, χρησιμοποιείται ένας μηδενικός στόχος, όπως π.χ. ένα διάνυσμα μηδενικών. Η ρύθμιση αυτή, αντιστοιχεί σε μια περίπτωση μεταγωγική, στην οποία είναι διαθέσιμοι οι στόχοι στους κόμβους στο σύνολο των μεταγωγικών.

5. Γράφοι Γνώσης

5.1. Εισαγωγή

Ο γράφος γνώσης είναι ένα μοντέλο γνώσης, το οποίο αναπαριστά αλληλένδετες έννοιες, οντότητες, σχέσεις και γεγονότα [16]. Οι γράφοι γνώσης, οργανώνουν δεδομένα από πολλαπλές πηγές, για να εκμαιεύσουν πληροφορίες για τις οντότητες που περιέχουν και να δημιουργηθούν συσχετίσεις μεταξύ τους. Αποτελούνται από κόμβους και ακμές, οι οποίοι αναπαριστούν οντότητες και σχέσεις, αντίστοιχα. Ο γράφος γνώσης, θεωρείται χρήσιμος για αναπαράσταση γνώσης, η οποία θα χρησιμοποιηθεί τόσο στον εμπορικό και επιστημονικό τομέα, όσο και σε πειράματα όπως: γραφήματα ροής δεδομένων, δυαδικά γραφήματα απόφασης, απάντηση ερωτήσεων και οργάνωση γνώσης στο διαδίκτυο. Η εκτεταμένη χρήση των γραφημάτων γνώσης, οφείλεται στη συμβολή τους στη μηχανική μάθηση, όπου οι ισχυρές σχέσεις μεταξύ των οντοτήτων, βοηθούν στην εξακρίβωση της σημασίας των δεδομένων, λόγω της σημασιολογικής τους φύσης.

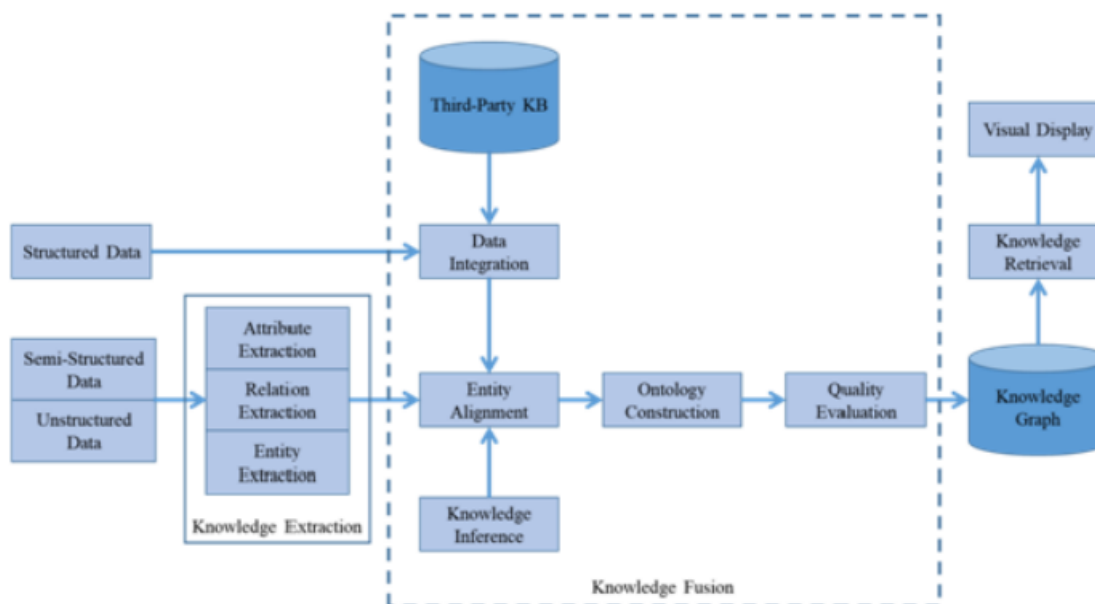
Κάποιοι από τους πιο γνωστούς γράφους γνώσης, ανήκουν στις Microsoft, Google, Facebook, IBM και eBay. Ο γράφος γνώσης μας δίνει τη δυνατότητα να αναζητήσουμε μέρη, πρόσωπα ή ακόμα και πράγματα, τα οποία περιέχει, με σκοπό να λάβουμε πληροφορίες που συσχετίζονται άμεσα με το ερώτημα που αναζητείται. Η μεθοδολογία αποτελεί ένα κρίσιμο βήμα προς την οικοδόμηση της επόμενης γενιάς αναζήτησης, η οποία αξιοποιεί τη συλλογική νοημοσύνη του ιστού και προσπαθεί να κατανοήσει τον κόσμο όπως ο άνθρωπος. Για παράδειγμα ο γράφος γνώσης της Google δεν βασίζεται μόνο σε δημόσιες πηγές πληροφοριών όπως το Freebase, η Wikipedia και το CIA World Factbook, αλλά είναι επίσης επαυξημένο σε πολύ μεγαλύτερη κλίμακα. Αυτή τη στιγμή περιέχει περισσότερα από 500 εκατομμύρια αντικείμενα, καθώς και περισσότερα από 3,5 δισεκατομμύρια στοιχεία για και σχέσεις μεταξύ αυτών των διαφορετικών αντικειμένων. Και ρυθμίζεται με βάση το τι αναζητούν οι άνθρωποι και τι ανακαλύπτουμε στον ιστό.

5.2. Αρχιτεκτονική και κατασκευή ενός γράφου γνώσης

Παρά την ύπαρξη πολλαπλών γράφων γνώσης, καθώς και την εκτεταμένη έρευνα στον τομέα τους, λόγω του ότι υπάρχουν αντικρουόμενα συμφέροντα μεταξύ των προαναφερθέντων εταιρειών – ως προς την ποιότητα των γράφων γνώσης τους - δεν υπάρχει κάποια προκαθορισμένη μεθοδολογία, τόσο για την κατασκευή τους, όσο και για την εξαγωγή γνώσης. Σύμφωνα με τους Zhao, Z., Han, S., & So, I. (2018), για τη δημιουργία ενός γράφου γνώσης υπάρχουν δύο προσεγγίσεις: από πάνω προς τα κάτω και από κάτω προς τα πάνω. Κατά την πρώτη προσέγγιση, αρχικά ορίζεται η οντολογία και το σχήμα του γράφου και στη συνέχεια τα δεδομένα εισάγονται στη βάση γνώσης. Στη δεύτερη, η γνώση εξάγεται από διάφορες πηγές όπως επίσης και από διάφορα ανοιχτά συνδεδεμένα δεδομένα και αφού συγχωνευτεί, κατασκευάζεται η οντολογία και το σχήμα του γράφου γνώσης. Αυτή είναι και η προσέγγιση που θα αναλυθεί περισσότερο στη συνέχεια.

Κατά τη δημιουργία ενός γράφου από κάτω προς τα πάνω, η γνώση μπορεί να προέρχεται από δομημένα δεδομένα, από ημιδομημένα δεδομένα ή ακόμη και από αδόμητα δεδομένα. Στη συνέχεια, η συγχώνευση όλης αυτής της γνώσης αποτελεί μια επαναλαμβανόμενη διαδικασία, κατά την οποία με βάση τα δεδομένα που εισάγονται, κατασκευάζεται η οντολογία του γράφου και αξιολογείται συνεχώς η ποιότητά της. Αφού ολοκληρωθεί η διαδικασία αυτή, οι γνώσεις αποθηκεύονται σε βάσεις δεδομένων και αφού ολοκληρωθεί και η αποθήκευση, οι χρήστες μπορούν να θέσουν ερωτήματα στο γράφο γνώσης και να πάρουν τις απαντήσεις τους είτε σε μορφή κειμένου σε κάποια γλώσσα, είτε ακόμη και σε μια πιο οπτικοποιημένη μορφή με τη βοήθεια κατάλληλων προγραμμάτων. Η παραπάνω διαδικασία αναπαρίσταται στην Εικόνα 15 και συνοψίζεται σε 4 βήματα: α) Εξαγωγή γνώσης, β) Συγχώνευση γνώσεων, γ) Αποθήκευση γράφου και δ) Ανάκτηση και οπτικοποίηση γνώσεων.

Εικόνα 15: Αρχιτεκτονική ενός γράφου γνώσης



5.2.1. Εξαγωγή γνώσης

Οι γνώσεις του γράφου είναι πληροφορίες που εμπεριέχονται σε 3 είδη δεδομένων, από τα οποία παίρνουμε την επιθυμητή πληροφορία, δηλαδή: Τα δομημένα δεδομένα που προκύπτουν από οργανωμένες βάσεις δεδομένων, τα ημιδομημένα δεδομένα, δηλαδή αρχεία που βρίσκονται σε μορφές όπως HTML, XML και JSON και τα αδόμητα δεδομένα όπως εικόνες, έγγραφα και αρχεία κειμένου. Αναλόγως με τα δεδομένα που θα χρησιμοποιηθούν, επιλέγεται και η κατάλληλη στρατηγική για την κατασκευή της οντολογίας και κατ' επέκταση ολόκληρου του γράφου. Υπάρχουν 3 διαφορετικοί τύποι εργαλείων εξαγωγής γνώσης από τα δεδομένα: η

εξαγωγή οντοτήτων, η εξαγωγή σχέσεων μεταξύ οντοτήτων και η εξαγωγή των χαρακτηριστικών των οντοτήτων. Στον πρώτο τύπο εξαγωγής, γίνεται η προσπάθεια του προσδιορισμού διαφορετικών οντοτήτων μέσα από την «ανάγνωση» μιας μεγάλης πηγής γνώσης. Όσο πιο σωστός είναι αυτός ο διαχωρισμός των οντοτήτων τόσο καλύτερη θα είναι και η ποιότητα της γνώσης που θα προκύψει. Το εργαλείο που χρησιμοποιείται πιο συχνά είναι η αναγνώριση επώνυμης οντότητας (Named Entity Recognition – NER), η οποία εξειδικεύεται στην εύρεση ατόμων, οργανισμών, τοποθεσιών και γεγονότων, με παράδειγμα το spaCy8. Για να βεβαιωθεί ότι εξάγονται όλες οι χρήσιμες πληροφορίες που αφορούν την εκάστοτε οντότητα, χρησιμοποιείται εργαλείο αποσαφήνισης οντότητας (Named Entity Disambiguation - NED), το οποίο ουσιαστικά εξάγει το νόημα της κάθε πρότασης αποσαφηνίζοντας την οντότητα την οποία αφορά η πληροφορία, το οποίο χρησιμοποιείται σε προτάσεις όπου η πληροφορία αφορά οντότητες οι οποίες είναι ασαφείς (π.χ. αυτός, αυτό, κτλ.) αλλά αναφέρονται σε οντότητες που προσδιορίζονται σε προηγούμενη ή επόμενη πρόταση, με παράδειγμα το NeuralCoref. Τέλος, ανατίθεται στην κάθε πρόταση μια μοναδική ταυτότητα, διαδικασία γνωστή ως σύνδεση οντοτήτων (Named Entity Linking - NEL), ενώ εξάγονται όλες οι πληροφορίες που αφορούν την κάθε οντότητα, μέσω σύνδεσης με μια βάση γνώσης. Στον δεύτερο τύπο εξαγωγής, προσδιορίζονται οι σχέσεις που υπάρχουν ανάμεσα στις οντότητες που έχουν προκύψει, με σκοπό την ανάδειξη των πληροφοριών σημασιολογικού περιεχομένου για την κατασκευή του γράφου. Στον τρίτο και τελευταίο τύπο, προσδιορίζονται τα μοναδικά χαρακτηριστικά της κάθε οντότητας, καθορίζοντας τη σημασιολογία της.

Για να μπορεί να επιτελεστεί η παραπάνω διαδικασία, απαραίτητη είναι η επεξεργασία της πηγής, την οποία επεξεργαζόμαστε. Στην περίπτωση μας, έχοντας να κάνουμε με κείμενα, η διαδικασία αυτή, γίνεται μέσω της επεξεργασίας φυσικής γλώσσας (NLP), κατά την οποία ο υπολογιστής, με τη βοήθεια αλγορίθμων ουσιαστικά αποκτά τη δυνατότητα να καταλαβαίνει την ανθρώπινη γλώσσα εισαγωγής. Κάποιοι άλλοι τρόποι επεξεργασίας είναι η εξόρυξη κειμένου (text mining) κατά την οποία οι λέξεις, με τη βοήθεια αλγορίθμων, μετατρέπονται σε αριθμητικές παραστάσεις με σκοπό την ανίχνευση διαφόρων μοτίβων και σχέσεων μέσα σε ένα μεγάλο μέρος κειμένου και η μηχανική μάθηση (machine learning), κατά την οποία το σύστημα αποκτά τη δυνατότητα να μαθαίνει και να βελτιώνεται αυτόματα, χρησιμοποιώντας και πάλι αλγορίθμους.

5.2.2. Συγχώνευση γνώσεων

Ο στόχος σε αυτό το στάδιο είναι να ταξινομηθούν οι οντότητες και να πραγματοποιηθεί μια ευθυγράμμιση τους ώστε να καθοριστεί αν διαφορετικές οντότητες αναφέρονται σε ίδια αντικείμενα του πραγματικού κόσμου. Για να πραγματοποιηθεί η ευθυγράμμιση οντοτήτων χρησιμοποιούνται διάφορες τεχνικές οι οποίες αντιστοιχίζουν μεταξύ τους, τις οντότητες που αναφέρονται σε ίδια αντικείμενα.

Κατά την ευθυγράμμιση αυτή λοιπόν, γίνεται μια πρώιμη επεξεργασία των δεδομένων, κατά την οποία ελέγχεται, αρχικά κατά ζεύγη και έπειτα συλλογικά, η ακεραιότητα και η συνέπεια

αυτών ανάμεσα στις διαφορετικές πηγές που μπορεί να εμφανίζονται. Στη συνέχεια, πραγματοποιείται η αντιστοίχιση χαρακτηριστικών. Οι συναρτήσεις που χρησιμοποιούνται γι' αυτό είναι δύο. Η μία είναι η συνάρτηση ομοιότητας κειμένου κατά την οποία αντιστοιχίζονται και συγκρίνονται τα χαρακτηριστικά και η δεύτερη είναι η συνάρτηση δομικής ομοιότητας που χρησιμοποιείται για την αντιστοίχιση και τη σύγκριση σχέσεων ανάμεσα στις οντότητες. Έχοντας ολοκληρωθεί η ευθυγράμμιση των οντοτήτων, δημιουργείται η οντολογία και ολοκληρώνεται η κατασκευή του γράφου γνώσης. Έπειτα από αξιολογήσεις που γίνονται στην ποιότητα της οντολογίας και του γράφου που προκύπτει, στην περίπτωση που δεν πληρούν τις απαιτήσεις που έχουν τεθεί, η διαδικασία συγχώνευσης των γνώσεων και κατ' επέκταση η δημιουργία του γράφου επαναλαμβάνονται από την αρχή.

5.2.3. Αποθήκευση του γράφου

Αφού ολοκληρωθεί η κατασκευή ενός γράφου γνώσης χρειάζεται κάπου να αποθηκευτεί. Όπως έχει αναφερθεί ξανά, συνήθως η αποθήκευση γίνεται σε NoSQL βάσεις δεδομένων. Οι δύο κατηγορίες που συναντάμε συχνότερα είναι η αποθήκευση βασισμένη σε RDF (Resource Description Framework) τριάδες και η αποθήκευση σε βάση δεδομένων γράφων (Graph DB).

5.2.4. Ανάκτηση και οπτικοποίηση γνώσης

Αφού έχει ολοκληρωθεί και η αποθήκευση ενός γράφου γνώσης, πλέον είναι έτοιμος να αξιοποιηθεί για οποιοδήποτε σκοπό έχουμε. Ένας από αυτούς τους σκοπούς μπορεί να είναι να απαντήσει στα ερωτήματα που θέτουν οι χρήστες του, να εντοπίσει απάτες και ψεύδη, να αναζητήσει κάποια δεδομένα ή και να προτείνει κάτι στον εκάστοτε χρήστη του (σαν recommender) . Μέσω των αποτελεσμάτων μπορούμε να διακρίνουμε ότι η απεικόνιση σε μορφή κειμένου μπορεί να δώσει πολλές λεπτομέρειες που βοηθούν στην καλύτερη και ευκολότερη ανάλυση του γράφου. Από την άλλη, η γραφική απεικόνιση επιτρέπει στο χρήστη την πλοήγηση δίνοντάς του έτσι τη δυνατότητα να ανακαλύψει τις γνώσεις που επιθυμεί, βλέποντας μία μεγαλύτερη, συνολικότερη εικόνα του γράφου.

6. Ανασκόπηση Προϋπάρχουσας Βιβλιογραφίας

Προτού αναλύσουμε τις δικές μας προσεγγίσεις και τα πειράματα που εκτελέστηκαν είναι αρκετά σημαντικό να μελετήσουμε τις υλοποιήσεις που ήδη υπάρχουν, λαμβάνοντας υπόψη τόσο τα πλεονεκτήματά τους, όσο και τα μειονεκτήματά τους. Στόχος είναι η εξαγωγή συμπερασμάτων, ως προς τη φιλοσοφία τους και η αξιοποίηση τους πάνω στο δικό μας σύνολο δεδομένων σε κατοπινό στάδιο, με απώτερο σκοπό την εύρεση της βέλτιστης δυνατής μεθοδολογίας εντοπισμού ψευδών ειδήσεων.

6.1. Εντοπισμός Ψευδών Ειδήσεων μέσω περιεχομένου

6.1.1. Μέθοδοι βασιζόμενοι αποκλειστικά στο περιεχόμενο

Η διαδικασία επεξεργασία φυσικής γλώσσας περιλαμβάνει εργασίες όπως προ-επεξεργασία κειμένου, παραγωγή διανυσμάτων λέξεων και εξαγωγή χαρακτηριστικών. Η προ-επεξεργασία δεδομένων βοηθά στην εξοικονόμηση χώρου, υπολογιστικού χρόνου και γενικότερα πόρων, καθώς βοηθά στην εξάλειψη ανεπιθύμητου θορύβου στα δεδομένα. Όπως έχει προαναφερθεί, η αναπαράσταση της φυσικής γλώσσας, γίνεται μέσω αναπαράστασης των λέξεων σε διανύσματα, πράγμα το οποίο υλοποιείται σε πρώτη φάση μέσω αρκετών μεθόδων που υπάγονται στον εντοπισμό μέσω του περιεχομένου. Ξεκινώντας από τα μοντέλα τα οποία δημιουργούν σταθερά διανύσματα για κάθε είσοδο – λέξη, έχουμε το CBOW (Continuous Bag of Words), το οποίο αντιλαμβάνεται κάθε άρθρο σαν ένα κείμενο και υπολογίζει την λέξη – στόχο που θέλει να προβλέψει και χρησιμοποιεί το άθροισμα των διανυσμάτων των λέξεων που υπάρχουν γύρω της, αλλά δεν μπορεί να αντιληφθεί το νόημα της λέξης ως προς τη θέση της στο κείμενο, καθώς και ούτε την σημασία της. Μια άλλη μέθοδος είναι το SkipGram, το οποίο επιτελεί ακριβώς την αντίθετη διαδικασία, όπου χρησιμοποιεί μια λέξη-στόχο, για να γίνει πρόβλεψη των γειτονικών της λέξεων. Στη συνέχεια, έχουμε το TF-IDF (Term Frequency – Inverse Document Frequency) το οποίο χρησιμοποιείται σε διάφορες προσεγγίσεις στα πλαίσια μηχανικής μάθησης για εντοπισμό ψευδών ειδήσεων, υπολογίζοντας το γινόμενο του αριθμού των εμφανίσεων του εκάστοτε όρου στο κείμενο, με το πόσο συνηθισμένη είναι μια λέξη ως προς το ολικό σώμα κειμένων, δηλαδή τον αριθμό των κειμένων που περιέχουν την εκάστοτε λέξη ανά τον αριθμό των κειμένων, υπολογίζοντας το βάρος της κάθε λέξης ως προς το κείμενο που δίνεται, έχοντας όμως σαν βασικό μειονέκτημα το γεγονός ότι παρότι παράγει διανύσματα για τη βαρύτητα της κάθε λέξης, δεν είναι ικανό να αντιληφθεί το νόημα των λέξεων. Έπειτα, έχουμε το CV (Count Vectorizer) το οποίο αποθηκεύει σε λεξικό τον αριθμό των εμφανίσεων κάθε λέξης σε κάθε πρόταση-κείμενο που μελετάται. Το HV (Hashing Vector) διορθώνει ουσιαστικά τις ελλείψεις του CV, χρησιμοποιώντας το hashing trick για να αναπαριστάται η συχνότητα εμφάνισης των λέξεων σε κάθε πρόταση-κείμενο και πλεονεκτεί ως προς την απόδοσή του σε μεγάλα κείμενα. Στη συνέχεια, έρχεται το FastText [17], το οποίο είναι ένα προ-εκπαιδευμένο μοντέλο σε μεγάλα σώματα κειμένων και σε διάφορες γλώσσες και μπορεί μέσω Skipgram, το οποίο προβλέπει μια λέξη στόχο με βάση μια κοντινή και CBOW να δημιουργήσει διανύσματα για κάθε λέξη, ενώ διαθέτει την ικανότητα ταξινόμησης κειμένου. Το

Word2Vec[18] είναι επίσης προ-εκπαιδευμένο σε σύνολα δεδομένων και αναπαριστά κάθε λέξη με ένα ξεχωριστό διάνυσμα σε ένα κοινό διανυσματικό χώρο, χρησιμοποιώντας το Skipgram ή το CBOW, με σκοπό την δημιουργία διανυσμάτων που θα αναπαριστούν παρόμοιου νοήματος λέξεις, κοντά στο διανυσματικό χώρο. Τέλος έχουμε το Glove[19], το οποίο και πάλι είναι προ-εκπαιδευμένο σε τεράστια σώματα κειμένου και δημιουργεί διανύσματα λέξεων στον διανυσματικό χώρο, χρησιμοποιώντας την πιθανότητα συνεμφάνισης λέξεων σε ολόκληρο το σώμα κειμένων, πράγμα το οποίο καθορίζει και τη μεταξύ τους συσχέτιση. Οι παραπάνω μέθοδοι παρά το γεγονός ότι είναι αρκετές αποδοτικές, δεν μπορούν να αντιληφθούν την αμφισημία των λέξεων, αλλά ούτε και τη σειρά με την οποία γράφονται οι λέξεις. Ο λεπτομερής ορισμός των μεθόδων, παρατίθεται στο Κεφάλαιο 7.

Για να γίνει η ταξινόμηση των ειδήσεων σε ψευδείς και αληθείς, παρακάτω θα γίνει αναφορά των αλγορίθμων που χρησιμοποιούνται πιο συχνά στον εντοπισμό ψευδών ειδήσεων, αφού με εξαίρεση το FastText, οι υπόλοιπες μέθοδοι χρειάζονται κάποιο ταξινομητή για την κατανομή των ειδήσεων σε ψευδείς και αληθείς. Αρχικά, έχουμε το SVM (Support Vector Machine), το οποίο εκπαιδεύεται χρησιμοποιώντας δεδομένα που προηγουμένως είχαν χωριστεί σε δύο ομάδες (ψευδείς – αληθείς ειδήσεις) και έχει ως στόχο την εύρεση υπερεπιπέδου, το οποίο θα διαχωρίζει το σύνολο δεδομένων σε δύο ομάδες [20]. Τα σημεία δεδομένων που βρίσκονται πιο κοντά στο υπερεπίπεδο είναι διανύσματα υποστήριξης (support vectors) και η ύπαρξη τους, επηρεάζει τη θέση του υπερεπιπέδου. Το υπερεπίπεδο, περιγράφεται ως μια γραμμή που διαιρεί γραμμικά και ταξινομεί μια ομάδα δεδομένων και όσο πιο μακριά από το υπερεπίπεδο είναι τα σημεία των αντίστοιχων δεδομένων, τόσο πιο μεγάλη είναι η πιθανότητα σωστής κατηγοριοποίησής τους. Το SVM έχει αρκετά μειονεκτήματα, τα κύρια εκ των οποίων είναι: α) Ο αλγόριθμος SVM δεν έχει καλή απόδοση σε μεγάλα σύνολα δεδομένων και β) δεν έχει καλή απόδοση όταν το σύνολο δεδομένων περιέχει αρκετό θόρυβο. Έπειτα έχουμε τον αλγόριθμο Naïve Bayes, ο οποίος βασίζεται στο θεώρημα Bayes και βασίζεται στην αρχή ότι όταν γνωρίζουμε πόσο πιθανό είναι να συμβεί κάτι, τότε μπορούμε να εκτιμήσουμε την πιθανότητα να συμβεί. Για κάθε κλάση που περιέχει το σύνολο δεδομένων, έχει την ικανότητα να προβλέπει την πιθανότητα ένα στοιχείο να ανήκει σε μια κατηγορία, αλλά υστερεί στο γεγονός ότι υπερεκτιμά τις εξαρτήσεις μεταξύ των δεδομένων, παρά την καλή απόδοση που έχει σε μικρά κείμενα. Μεταξύ των δύο, προτιμάται ο Naïve Bayes, διότι έχει καλύτερη απόδοση σε μεγάλα σύνολα δεδομένων. Η παραπάνω θέση αυτή υποστηρίζεται και στο [21], αφού χρησιμοποιήθηκαν και οι δύο αλγόριθμοι για ταξινόμηση κειμένων που περιέχονται στο σύνολο δεδομένων και ο αλγόριθμος SVM έχει την καλύτερη απόδοση, σε μικρότερα σύνολα κειμένων σε σχέση με τον Naïve Bayes, ενώ ο τελευταίος έχει καλύτερη απόδοση σε αντίθετη περίπτωση.

Στο [22] γίνεται ταξινόμηση ειδήσεων, οι οποίες είναι γραμμένες στην Ινδονησιακή γλώσσα και παρατηρείται η επιρροή της τεχνικής TF-IDF, σε μια γλώσσα που δεν διαθέτει τόσα εργαλεία επεξεργασίας. Αφού προηγηθεί η προ-επεξεργασία του κειμένου, χρησιμοποιείται η τεχνική TF-IDF, για να εξακριβωθεί η σημασία της κάθε λέξης ως προς το κάθε κείμενο και γενικά το σώμα κειμένων. Έπειτα συγκρίνονται οι ταξινομητές SVM και SGD για να βρεθεί

ποιος έχει την ικανότητα να ταξινομήσει καλύτερα τις ειδήσεις και παρατηρείται ότι ο SGD σε αυτή την περίπτωση έχει καλύτερη απόδοση.

Στο [23], αφού γίνει η προ-επεξεργασία δεδομένων και εξαγωγή χαρακτηριστικών μέσω TF-IDF και CV (Count Vectorizer), χρησιμοποιούνται εννέα ταξινομητές: DT (Decision Tree), LR (Logistic Regression), XGB (XGBoost), RF (Random Forest), ET (Extra Trees), AB (AdaBoost), SVM, SGD (Stochastic Gradient Descent) και NB, με σκοπό την ταξινόμηση των ειδήσεων. Έπειτα χρησιμοποιείται σύστημα ψήφισης σύμφωνα με την πλειοψηφία, όπου ουσιαστικά με βάση την πρόβλεψη που έδωσαν οι πλείστοι ταξινομητές, καθορίζεται η κλάση της εκάστοτε είδησης. Όπως παρατηρείται, η προσέγγιση αυτή έχει καλύτερη απόδοση σε σχέση με τη χρήση του κάθε ταξινομητή ξεχωριστά.

Στο [24], δημιουργήθηκε ένα μοντέλο εντοπισμού ψευδών ειδήσεων, το οποίο χρησιμοποιεί ένα πολυεπίπεδο σύστημα ψήφισης, το οποίο χρησιμοποιεί δώδεκα διαφορετικούς ταξινομητές (MNB, SVC, NuSVC, LinearSVC, DT, PA, SGD, LR, MLP, AN, GB, Voting) σε συνδυασμό με τρεις μεθόδους εξαγωγής χαρακτηριστικών: το HV, το CV και το TF-IDF. Το πρώτο επίπεδο αποτελείται από τρεις φάσεις, όπου για κάθε μέθοδο επιλέγονται τρεις ταξινομητές ανά φάση - δημιουργώντας για κάθε μέθοδο ένα ταξινομητή ψήφισης - με βάση το χαμηλότερο FP rate (ψευδείς ειδήσεις ταξινομούνται λανθασμένα ως αληθείς) ως προς το σύνολο δεδομένων που μελετάται και επιλέγονται για κάθε φάση δημιουργείται ένας ταξινομητής ψήφισης που περιέχει ένα συνδυασμό των ταξινομητών με κάθε μέθοδο. Ο κάθε ταξινομητής ψήφισης αποτελείται από τρεις άλλους και στο δεύτερο επίπεδο, οι τρεις ταξινομητές ψήφισης που είχαν το χαμηλότερο FP rate ενώνονται. Έπειτα γίνεται σύγκριση με τρεις νέους ταξινομητές για κάθε μέθοδο, οι οποίοι επιλέχθηκαν αυτή τη φορά με βάση το υψηλότερο TP rate (ψευδείς ειδήσεις που ταξινομούνται ορθώς ως ψευδείς), δημιουργώντας ένα νέο ταξινομητή ψήφισης. Στο τρίτο και τελευταίο επίπεδο, οι τρεις ταξινομητές ψήφισης με το χαμηλότερο FP rate, συνενώνονται, δημιουργώντας τον τελικό ταξινομητή ψήφισης, ο οποίος εξάγει τις κλάσεις με βάση την πλειοψηφία. Αυτή η προσέγγιση είχε την κορυφαία απόδοση στη βιβλιογραφία αποκλειστικά με βάση το περιεχόμενο, με απόδοση που ανέρχεται στο 98,5%.

Τέλος, στην ταξινόμηση κειμένου αποκλειστικά με τη χρήση FastText, το οποίο μπορεί να χρησιμοποιηθεί αυτούσιο για την ταξινόμηση κειμένου, παρατηρούνται πολύ υψηλά αποτελέσματα ως προς την απόδοση, τα οποία μάλιστα προκύπτουν σε πολύ μικρότερους χρόνους, σε σχέση με άλλες μεθόδους, γεγονός το οποίο οφείλεται στο ότι το FastText είναι εκπαιδευμένο σε διάφορες γλώσσες, το οποίο βοηθά στην αντίληψη της δομής του λόγου της εκάστοτε γλώσσας, αλλά και λόγω διαφόρων τεχνικών που χρησιμοποιεί για τη δημιουργία διανυσμάτων λέξεων, όπως το n-gram, όπου χρησιμοποιεί συνδυασμό γραμμάτων της λέξης που μελετάται, για την δημιουργία του διανύσμάς της, κάτι το οποίο δημιουργεί ένα πιο καλά δομημένο σημασιολογικό διάνυσμα, το οποίο δηλαδή αναπαριστά πιο αποδοτικά το νόημα της λέξης.

6.1.2. Μεθόδους βασισμένοι στο περιεχόμενο, λαμβάνοντας υπόψη τα συμφοραζόμενα

Τα μοντέλα που αναλύονται παρακάτω, πλεονεκτούν σε σχέση με τα πιο πάνω, καθώς τα διανύσματα που δημιουργούν αντιλαμβάνονται την ενδεχόμενη αμφισημία της εκάστοτε λέξης και το νόημά της σε μια πρόταση σύμφωνα με τα συμφοραζόμενα. Ξεκινώντας από τα βασικά μοντέλα βαθιάς μάθησης, έχουμε τα RNNs (Recurrent Neural Networks), γνωστά και ως αναδρομικά νευρωνικά δίκτυα [25], τα οποία λαμβάνουν πληροφορίες από προηγούμενες εισόδους, οι οποίες θα επηρεάσουν την τρέχουσα είσοδο και έξοδο. Τα αναδρομικά νευρωνικά δίκτυα πλεονεκτούν έναντι άλλων νευρωνικών δικτύων που υποθέτουν ότι οι εισόδοι και οι έξοδοι είναι ανεξάρτητες και επίσης λαμβάνουν υπόψη τη θέση της κάθε λέξης στο κείμενο. Τα αναδρομικά νευρωνικά δίκτυα, δέχονται σαν είσοδο ακολουθιακά δεδομένα – στις παρακάτω περιπτώσεις λέξεις – και αποτελούνται από την τρέχουσα είσοδο που είναι η κάθε λέξη-μέρος της ακολουθίας που μελετάται και την προηγούμενη και τρέχουσα κρυφή κατάσταση, η οποία περιέχει πληροφορίες από όλες τις προηγούμενες λέξεις. Για τον υπολογισμό των βαρών της κάθε λέξης, πραγματοποιείται η διαδικασία της διάδοσης προς τα εμπρός (forward pass) και για την προσαρμογή τους κατά την εκμάθηση, πραγματοποιείται η διαδικασία της διάδοσης προς τα πίσω (back-propagation). Κατά την χρήση του αλγορίθμου της οπισθοδιάδοσης, χρησιμοποιούνται οι κλίσεις – gradients, οι οποίες ανά επίπεδο λέξη μικραίνουν ή μεγαλώνουν εκθετικά. Ενδέχεται εάν υπάρχουν πολλά επίπεδα – λέξεις, οι κλίσεις να γίνουν τόσο μεγάλες που τα βάρη θα δέχονται τεράστιες αλλαγές ή στην αντίθετη περίπτωση αμελητέες, πράγμα που σημαίνει ότι δεν μπορεί να γίνει σωστά η διαδικασία της εκμάθησης. Για να αντιμετωπιστούν αυτά τα προβλήματα, δημιουργήθηκαν τα Δίκτυα Μακράς και Βραχυπρόθεσμης Μνήμης - LSTMs (Long Short Term Memory) [26], τα οποία μέσω των μηχανισμών πύλης μπορούν να ελέγχουν την ροή των πληροφοριών. Αποτελούνται από τρεις πύλες: την πύλη εισόδου που ελέγχει ποιες νέες πληροφορίες θα εισέλθουν στο LSTM, την πύλη λήθης που καθορίζει ποιες πληροφορίες είναι χρήσιμες και ποιες πρέπει να αγνοηθούν και την πύλη εξόδου που καθορίζει ποιες πληροφορίες πρέπει να διατηρηθούν στην κρυφή κατάσταση και ποιες θα χρησιμοποιηθούν στο επόμενο βήμα. Έπειτα έρχονται οι GRUs (Gated Recurrent Unit) [27], οι οποίες επίσης επιλύουν το πρόβλημα μνήμης των RNNs, με τη χρήση αυτή τη φορά δύο πυλών ενημέρωσης, η οποία εξακριβώνει ποιες πληροφορίες είναι χρήσιμες να διατηρηθούν και επαναφοράς, η οποία οι οποίες καθορίζουν ποιες πληροφορίες από την τρέχουσα είσοδο είναι χρήσιμες. Οι GRUs προτιμώνται των LSTMs σε μικρά κείμενα, ενώ τα LSTMs αποδίδουν καλύτερα σε μεγάλα κείμενα. Έπειτα, συναντούμε τα Συνελκτικά Νευρωνικά Δίκτυα – CNN (Convolutional Neural Network) [28], τα οποία αποτελούνται από δύο επίπεδα: το επίπεδο συνέλιξης και το επίπεδο pooling. Κατά τη συνέλιξη εντοπίζονται τα πιο σημαντικά χαρακτηριστικά, χρησιμοποιώντας έναν πίνακα χαρακτηριστικών γνωστό και ως φίλτρο, ο οποίος πολλαπλασιάζεται με τον πίνακα εισόδου ανά στοιχείο. Μέσω pooling, το φίλτρο ολισθαίνει στον πίνακα και εξακριβώνει ποια είναι τα πιο σημαντικά χαρακτηριστικά, με μειωμένες διαστάσεις χωρίς να χάνεται πληροφορία. Αν και η συνήθης εφαρμογή των CNN είναι σε εικόνες, έχουν αρκετά καλά αποτελέσματα, και στην ταξινόμηση κειμένου. Την ανάγκη εξακρίβωσης της σχέσης των λέξεων ήρθαν να λύσουν οι Transformers [29], οι οποίοι χρησιμοποιούν κωδικοποιητές και αποκωδικοποιητές για την αντίστοιχη επεξεργασία του κειμένου και μπορούν να

αντιληφθούν τις σχέσεις των λέξεων σε μεγάλες αποστάσεις σε ένα κείμενο. Οι δύο κύριοι μηχανισμοί των Transformers είναι η κωδικοποίηση ανά θέση, όπου στο διάνυσμα κειμένου λαμβάνεται υπόψη η θέση της κάθε λέξης και ο μηχανισμός προσοχής, ο οποίος εξακριβώνει πόσο σημαντική είναι μια λέξη ως προς τις άλλες που περιέχονται στο κείμενο, κάτι το οποίο καταδεικνύει πόσο σημασιολογικά σημαντική είναι η κάθε λέξη. Τέλος, έχουμε τους BERTs [30], οι οποίοι υποστηρίζονται από τους Transformers και ουσιαστικά χρησιμοποιούν πολλούς κωδικοποιητές το έναν μετά τον άλλο και μέσω προ-εκπαίδευσης αντιλαμβάνονται τα χαρακτηριστικά της εκάστοτε γλώσσας και μέσω fine-tuning, προσαρμόζονται στην κάθε εργασία που χρησιμοποιούνται. Αντί να αποκωδικοποιεί τις κωδικοποιημένες πληροφορίες, ο BERT κωδικοποιεί και δημιουργεί μόνο ένα μοντέλο γλώσσας, επομένως αρκεί ένας κωδικοποιητής. Σε σύγκριση με μοντέλα όπως τα RNN και LSTM, τα οποία έχουν κατεύθυνση και αντιλαμβάνονται κάθε είσοδο διαδοχικά (από αριστερά προς τα δεξιά ή από δεξιά προς τα αριστερά), οι Transformers και οι BERTs δεν έχουν κατεύθυνση, επειδή και τα δύο αυτά μοντέλα διαβάζουν ολόκληρη την πρόταση ως είσοδο αντί για διαδοχική σειρά. Αντιλαμβάνεται το γεγονός ότι μια λέξη αναλόγως της θέσης της στο κείμενο, μπορεί να είναι αμφίσημη και στην αναπαράστασή της ως διάνυσμα, λαμβάνεται υπόψη αυτό το γεγονός. Έχει προ-εκπαιδευτεί σε σώματα κειμένων, πράγμα το οποίο παρέχει πρόγνωση, η οποία βοηθά στην αποδοτική αναπαράσταση των λέξεων ως διανύσματα. Οι αναλυτικοί ορισμοί των παραπάνω μεθόδων, παρουσιάζονται επίσης στο Κεφάλαιο 7.

Στο [31] χρησιμοποιήθηκαν δύο διαφορετικοί τύποι BERT (Bidirectional Encoder Representations from Transformers), οι οποίοι απέδειξαν την αποδοτικότητα τους στην δυαδική ταξινόμηση των ειδήσεων, σε σχέση με του πιο απλούς ταξινομητές, καθώς και το πόσο αποδοτικός είναι ο μηχανισμός προσοχής. Στην πιο πρόσφατη βιβλιογραφία, η χρήση των BERTs είναι εκτενής λόγω της αποτελεσματικότητάς τους όπως καταδεικνύει και στο [32] με τη δημιουργία του FakeBERT. Στο FakeBERT, αρχικά δημιουργούνται διανύσματα λέξεων με τη χρήση του BERT, τα οποία θα χρησιμοποιηθούν σαν εισόδοι σε ένα επίπεδο, το οποίο αποτελείται από τρία άλλα συνελκτικά επίπεδα, που χρησιμοποιούνται για να εξακριβωθούν τα πιο σημαντικά χαρακτηριστικά του κειμένου. Έπειτα χρησιμοποιούνται επίπεδα Max-pooling για την μείωση των υπολογιστικών πράξεων του συστήματος και τη μείωση του αριθμού των παραμέτρων, κρατώντας τις πιο υψηλές τιμές που αποτελούν τις πιο σημαντικές πληροφορίες. Ακολουθεί το επίπεδο Flatten, το οποίο μετατρέπει την είσοδό του σε διάνυσμα, το οποίο μπορεί να δοθεί σαν είσοδος σε οποιοδήποτε πλήρως συνδεδεμένο νευρωνικό δίκτυο, με το αναπαρασταθεί σε ένα μονοδιάστατο πίνακα. Αργότερα, ακολουθεί ένα επίπεδο dense, το οποίο βεβαιώνεται ότι κάθε νευρώνας δέχεται είσοδο από το προηγούμενο επίπεδο, δηλαδή έχουμε πλήρως συνδεδεμένο νευρωνικό δίκτυο. Ακολουθεί ένα επίπεδο dropout, ο σκοπός του οποίου είναι η τυχαία απενεργοποίηση ποσοστού των νευρώνων για να αποτρέψει τυχών υπερπροσαρμογή. Τέλος, στο τελευταίο επίπεδο dense υπάρχει μια συνάρτηση ενεργοποίησης, η οποία στην προκειμένη είναι η ReLU, η οποία ενεργοποιεί μέγιστο μεταξύ του 0 και της εισόδου που δέχεται και προβλέπει την κλάση της κάθε είδησης. Η απόδοση του FakeBERT ήταν σαφέστατα καλύτερη από όλα τα προαναφερθέντα μοντέλα.

Στο [33], η καινοτομία βρίσκεται στο γεγονός ότι συνδυάζονται διάφοροι Transformers, με σκοπό τον συνδυασμό των διανυσμάτων τους, για τη δημιουργία μιας πιο αντιπροσωπευτικής αναπαράστασης του κειμένου. Το μοντέλο ονομάζεται MisRoBERTa, το οποίο δέχεται σαν εισόδους διανύσματα προτάσεων που δημιουργήθηκαν μέσω του RoBERTa και διανύσματα προτάσεων μέσω του BART. Ο λόγος επιλογής των δύο αυτών transformers είναι το γεγονός ότι προηγήθηκε μια σύγκριση μεταξύ διαφόρων transformers χρησιμοποιώντας ένα απλό νευρωνικό δίκτυο αποτελούμενο από ένα επίπεδο διανυσμάτων, ένα dropout επίπεδο και ένα κρυφό επίπεδο, για ταξινόμηση του συνόλου δεδομένων που επιλέχθηκε μέσω σιγμοειδούς συνάρτησης, με τους προαναφερθέντες, να έχουν την καλύτερη απόδοση. Το RoBERTa είναι ουσιαστικά μια βελτιστοποίηση του BERT, η οποία χρησιμοποιεί τις βέλτιστες υπερπαραμέτρους, αφαιρώντας την πρόβλεψη επόμενης πρότασης και αυξάνοντας τα mini-batches και τα learning rates και το BART είναι μια γενίκευση του BERT, η οποία χρησιμοποιεί ακολουθία προς ακολουθία για μετάφραση μέσω ενός αμφίδρομου κωδικοποιητή και αποκωδικοποιητή από αριστερά προς τα δεξιά. Η διαδικασία που ακολουθείται είναι η εξής: τα διανύσματα ως εισόδοι σε νέο μπλοκ χρησιμοποιώντας ένα νευρωνικό δίκτυο που αποτελείται από επίπεδα BiLSTM, ανασχημάτισης, CNN, max pooling και flatten. Η έξοδος τους συνενώνεται, ανασχηματίζεται και τροφοδοτείται σε ένα μπλοκ που έχει ακριβώς την ίδια αρχιτεκτονική. Τέλος, χρησιμοποιήθηκε ένα dense επίπεδο για την ταξινόμηση και παρατηρείται η αποδοτικότητα του συνδυασμού των δύο Transformers.

Στο [34], χρησιμοποιούνται κείμενα τα οποία δεν έχουν το τυπικό περιεχόμενο μιας είδησης και αποτελούνται από τα tweets και τις απαντήσεις σε αυτά. Αρχικά, μαζεύονται tweets μέσω του Twitter API και μαζεύονται τα χαρακτηριστικά που χρειάστηκαν για την έρευνα όπως ο αριθμός των tweet ανά δευτερόλεπτο, ο αριθμός των ευαίσθητων tweet με ετικέτα, ο αριθμός των επαληθευμένων λογαριασμών που συμμετέχουν στη συνομιλία, ο μέσος αριθμός αναφορών ανά tweet, η μέση αναλογία ακολούθων των χρηστών που ακολουθεί ο λογαριασμός και ο μέσος tweet του χρήστη ανά δευτερόλεπτο. Έπειτα ανατίθεται στο κάθε tweet μια ετικέτα μέσω ενός ημιαυτόματου τρόπου, ο οποίος λαμβάνει υπόψη πολλά πεδία των μεταδεδομένων που περιέχονται στο κείμενο και δημιουργεί μετρήσεις με βάση αυτούς τους δείκτες. Προκειμένου να δημιουργηθεί η ετικέτα, καθένα από τα πεδία μεταδεδομένων κανονικοποιείται στο διάστημα (0, 1), το οποίο οδηγεί σε έναν τύπο τιμής πιθανότητας να έχουμε ψευδή είδηση ή αληθινή. Το νευρωνικό δίκτυο, αποτελείται από 5 επίπεδα: Αρχικά, έχουμε το επίπεδο text2vec – το οποίο είναι το αντίστοιχο του word2vec σε κείμενα - που μετατρέπει κείμενα σε διανύσματα στο χώρο και χρησιμοποιούνται ως είσοδος στο νευρωνικό δίκτυο. Έπειτα έχουμε το επίπεδο διανυσμάτων, όπου δημιουργεί διανύσματα τα οποία αναπαρίστανται σε ένα κοινό διανυσματικό χώρο. Ακολουθεί από επίπεδο max pooling, το οποίο αποτρέπει την μνήμη από το να υπερχειλίσει και παρατηρεί τις πιο μεγάλες τιμές στα διανύσματα. Ακολουθεί ένα επίπεδο BiGRU το οποίο δημιουργεί διανύσματα τα οποία αντιλαμβάνονται τη σημασιολογία της κάθε ακολουθίας που δέχονται. Τέλος, υπάρχει ένα επίπεδο dense το οποίο ακολουθεί ένα επίπεδο dropout, το οποίο χρησιμοποιεί μια σιγμοειδή συνάρτηση, ως συνάρτηση ενεργοποίησης και δείχνει την απόδοση που μπορεί να επιτευχθεί σε κείμενα, στα οποία αξίζει να αναφερθεί ότι δεν αφαιρέθηκαν ειδικοί χαρακτήρες όπως emoji και hashtags.

Στο [35], παρατίθεται ένα μοντέλο που ονομάζεται dEFEND, το οποίο επιτελεί ταξινόμηση ειδήσεων σε άρθρα ειδήσεων και τα αντίστοιχα σχόλιά τους. Χρησιμοποιείται BiGRU για την δημιουργία του σημασιολογικού διανύσματος του κειμένου ειδήσεων, ενώ χρησιμοποιείται και ο μηχανισμός προσοχής, ο οποίος εξακριβώνει πόσο σημαντική είναι η κάθε λέξη ως προς την είδηση, πληροφορία που συμπεριλαμβάνεται στο διάνυσμα αναπαράστασής της. Έπειτα μέσω BiGRU σε συνδυασμό με τον μηχανισμό προσοχής, δημιουργείται το διάνυσμα για κάθε σχόλιο. Μέσω του μηχανισμού συν-προσοχής, προσπαθεί να συνδέσει την εκάστοτε είδηση με τα αντίστοιχα σχόλια που την αφορούν, λαμβάνοντας υπόψη στο διάνυσμα ειδήσεων το διάνυσμα σχολίων και στο διάνυσμα σχολίων το διάνυσμα ειδήσεων κατά τον μετασχηματισμό τους, χρησιμοποιώντας μετέπειτα εκ νέου τον μηχανισμό προσοχής για κάθε είδηση και για κάθε σχόλιο. Αργότερα τα διανύσματα αυτά συνενώνονται και μέσω της softmax γίνεται η ταξινόμηση των ειδήσεων, με πολύ καλά αποτελέσματα.

Στο [36], υλοποιείται ένα μοντέλο που ονομάζεται MAC(Multi-head Attentive Network for Fact-Checking) το οποίο αποτελείται από σύνολα ισχυρισμών και των αντίστοιχων speakers τους, σύνολα κειμένων και των αντίστοιχων publishers τους και χρησιμοποιεί το μηχανισμό προσοχής τόσο σε επίπεδο λέξης, όσο και σε κείμενα. Ξεκινώντας, τόσο οι ισχυρισμοί, όσο και τα κείμενα, μοντελοποιούνται ως ένα σύνολο λέξεων. Κάθε λέξη αναπαρίσταται ως διάνυσμα σε ένα κοινό διανυσματικό χώρο, ενώ ανατίθεται διανύσματα και στους αντίστοιχους sharers και publishers. Τα διανύσματα λέξεων τροφοδοτούνται σε BiLSTM και ο μέσω mean pooling, δημιουργείται ένα διάνυσμα για ολόκληρο τον ισχυρισμό. Επίσης μέσω BiLSTM, δημιουργούνται σημασιολογικά διανύσματα για κάθε λέξη και δημιουργούν ένα διάνυσμα για κάθε κείμενο. Οι αναπαραστάσεις των κειμένων συνενώνονται και μέσω χρήσης του μηχανισμού προσοχής, έχει ως σκοπό να δοθεί έμφαση στις πιο σημαντικές πληροφορίες, γνωστές και ως λέξεις-κλειδιά, για να συμπεριληφθεί αυτή η πληροφορία στα ανανεωμένα τους διανύσματα. Ακολουθεί η συνένωση του διανύσματος του ισχυρισμού, με το διάνυσμα των speakers και του διανύσματος του κειμένου, με το διάνυσμα των publishers. Τα ανανεωμένα διανύσματα κειμένων σε συνδυασμό με τους publishers, συνενώνονται σε μια ενιαία αναπαράσταση, η οποία θα περιέχει όλα τα κείμενα και μέσω του μηχανισμού προσοχής, εξακριβώνεται ποια κείμενα περιέχουν τις πιο σημαντικές πληροφορίες, για να ληφθούν υπόψη στα ανανεωμένα τους διανύσματα. Τέλος, συνενώνεται το τελικό διάνυσμα κειμένων, το οποίο έχει συνενωθεί με τους publishers και περιλαμβάνει πληροφορίες ως προς τη σημαντικότητα κάθε κειμένου, με το τελικό διάνυσμα ισχυρισμών το οποίο το οποίο έχει συνενωθεί με τους sharers και περιλαμβάνει πληροφορίες ως προς τη σημαντικότητα κάθε λέξης και μέσω της εισόδου τους σε MLP και μέσω σιγμοειδούς συνάρτησης ταξινομούνται οι ειδήσεις.

Στο [37], παρατίθεται η πιο αποδοτική προσέγγιση που βρέθηκε σε ολόκληρη τη βιβλιογραφία, όπου αρχικά συγκρίθηκαν τα N-Gram και TF-IDF για εξαγωγή των χαρακτηριστικών από τα δεδομένα εισόδου, με το GloVe. Επιλέγονται τα διανύσματα μέσω GloVe, ενός για τις λέξεις που δεν παρέχει διανύσματα, χρησιμοποιείται η βιβλιοθήκη του keras, που δημιουργεί διανύσματα ίδιων διαστάσεων με αυτά του GloVe. Το μοντέλο ονομάζεται OPCNN-FAKE και αποτελείται από έξι επίπεδα: α) Το επίπεδο διανυσμάτων, όπου κάθε είδηση αναπαρίσταται ως

διανύσματα των λέξεων που περιέχει σε ένα κοινό διανυσματικό χώρο, β) Το επίπεδο που επιτελείται το dropout, γ) Το συνελκτικό επίπεδο, όπου χρησιμοποιούνται τα πιο σημαντικά χαρακτηριστικά, δ) Το επίπεδο pooling, όπου επιλέγεται η μεγαλύτερη τιμή, έτσι ώστε να πάρουμε τα πιο σημαντικά χαρακτηριστικά, ε) Το επίπεδο flatten όπου το κείμενο πλέον μετατρέπεται σε ένα μονοδιάστατο πίνακα, ο οποίος με τη σειρά του τροφοδοτείται στο τελευταίο στ) επίπεδο εξόδου dense, το οποίο μέσω σιγμοειδούς συνάρτησης αποφασίζει κατά πόσο η εκάστοτε είδηση είναι αληθής ή ψευδής. Τα αποτελέσματα δείχνουν ότι το OPCNN-FAKE έχει καλύτερη απόδοση σε σχέση με όλα τα άλλα μοντέλα, ενώ σε κάποιες περιπτώσεις αγγίζει μέχρι και το τέλειο (99,99%).

6.3. Εντοπισμός Ψευδών Ειδήσεων μέσω Δικτύου Διάδοσης

Ο κύκλος διάδοσης των ειδήσεων, συμπεριλαμβάνει το στάδιο κατά το οποίο οι ψευδείς ειδήσεις αναπαράγονται. Η διαδικασία αυτή, συνήθως επιτελείται μεταξύ ατόμων που έχουν κοινές απόψεις και έτσι δημιουργούνται μεταξύ τους κοινωνικοί δεσμοί, που τους επιτρέπουν να αλληλοεπιδρούν μεταξύ τους, ανταλλάζοντας ειδήσεις σύμφωνα με τα ενδιαφέροντά τους. Μαζί, σχηματίζουν κοινότητες ανταλλαγής πληροφοριών, όπου είτε άθελα είτε ηθελημένα, συνδέουν τους χρήστες μεταξύ τους με βάση το περιεχόμενο των ειδήσεων, καθώς και τις πηγές από τις οποίες προέρχεται το περιεχόμενο. Ακόμα και σε αυτές τις μικρές και απομονωμένες κοινωνίες, υπάρχει η δυνατότητα διάδοσης ψευδών ειδήσεων, οι οποίες μπορούν να διαδοθούν και να πληροφορούν λανθασμένα τους αναγνώστες. Οι μέθοδοι, οι οποίες βασίζονται στη διάδοση, ταξινομούν τις ειδήσεις σε ψευδείς ή αληθείς και μέσω αυτών, βασίζεται και το πρότυπο διάδοσης. Ακολουθούν πανομοιότυπη διαδικασία με τις προαναφερθείσες μεθόδους, αλλά διαφοροποιούνται στο γεγονός ότι αναπαριστούν τα δεδομένα σε μορφή γράφου, είτε κατευθυνόμενου είτε μη κατευθυνόμενου. Οι πληροφορίες που αναπαρίστανται στον γράφο είναι οι διαδρομές που έχουν ταξιδέψει οι ειδήσεις ή αλλιώς τα άλματα που έχουν κάνει για να διαδοθούν. Μπορούν επίσης να είναι ο αριθμός των χρηστών που διαδίδουν τις ειδήσεις ή ακόμα και η διάρκεια διάδοσης της εκάστοτε είδησης. Στους μη κατευθυνόμενους γράφους, ενδέχεται να περιέχονται πληροφορίες όπως τα δημογραφικά στοιχεία χρηστών, η πρόθεση διάδοσης ψευδών ειδήσεων και η αλληλεπίδραση με άλλους χρήστες. Τέτοιες μέθοδοι, βασίζονται στο φαινόμενο της κοινότητας, στην οποία περιέχονται χαρακτηριστικά που περιγράφουν το μοτίβο και την κυκλοφορία των ψευδών ειδήσεων και επίσης βασίζονται στην αλληλεπίδραση των χρηστών με τις ψευδείς ειδήσεις. Από τεχνικής απόψεως, χρησιμοποιούμε τα στοιχεία που παρέχονται από τις υπάρχουσες γνώσεις, που αποτελούν περιεχόμενου χαμηλού επιπέδου και μέσω επεξεργασίας, τα αναγάγουμε σε υψηλό επίπεδο, παίρνοντας έτσι τα δεδομένα εκπαίδευσης, για να γίνει αξιολόγηση ως προς τις σχέσεις που έχουν παρατηρηθεί, καθώς και αυτές οι οποίες που προβλέφθηκαν. Η διαδικασία γίνεται χρησιμοποιώντας ένα νευρωνικό – ενδεχομένως και σχεσιακό – γράφημα, το οποίο δημιουργεί καταναμημένες αναπαραστάσεις κόμβων που διαμορφώνονται από τη δομή του γραφήματος, επιτρέποντάς μας διαδώσουμε πληροφορίες από τους προϋπάρχοντες κόμβους, σε κόμβους άγνωστης πηγής χρησιμοποιώντας τα δια-

νύσματα του γραφήματος. Τέτοια δίκτυα, θεωρούνται τα GNN, τα οποία είναι ικανά να επιτελέσουν διάφορες εργασίες, από αναγνώριση ομιλίας, έως και δυαδική ταξινόμηση ειδήσεων. Μια ειδική περίπτωση GNNs είναι τα Συνελικτικά Δίκτυα Γράφων (GCN) και τα Δίκτυα Γράφων Προσοχής (GAT), τα οποία έχουν οριστεί πιο πάνω και παρέχουν τη δυνατότητα δυαδικής ταξινόμησης και δημιουργίας αναπαραστάσεων που λαμβάνουν υπόψη το περιεχόμενο των γειτονικών κόμβων ή στην περίπτωση του GAT, λαμβάνουν υπόψη πόσο σημαντικός είναι ο εκάστοτε γειτονικός κόμβος.

Βλέποντας την πιο πρόσφατη βιβλιογραφία, στο [38] χρησιμοποιείται ένα σύνολο δεδομένων, γνωστό ως FakeNewsNet, το οποίο πέραν των ψευδών και αληθινών ειδήσεων, περιέχει πληροφορίες των χρηστών που τις διέδωσαν – στην προκειμένη στο twitter – και από πόσους αναπαρήχθησαν, λόγω της αρχικής διάδοσης. Αυτό που κάνει μοναδική αυτή την προσέγγιση είναι ότι δεν χρησιμοποιεί το κείμενο των ειδήσεων, αλλά μελετά τα χαρακτηριστικά των χρηστών που τις διέδωσαν. Κάθε είδηση μετατρέπεται ως κόμβος που αποτελεί μέρος του γράφου και συνδέεται με άλλους κόμβους, που αποτελούν τα tweets που αναπαρήγαγαν την είδηση. Τα tweets συνδέονται με τα retweets, τα οποία επίσης αναπαρήγαγαν τα tweets και κατ' επέκταση την είδηση. Υπάρχουν ακμές μόνο μεταξύ των αρχικών κειμένων ειδήσεων και των χρηστών που τις έκαναν tweet και των χρηστών που τις έκαναν tweet με τους retweeters, σχέσεις οι οποίες αναπαρίστανται ως ένας πίνακας γειννίας. Ο πίνακας χαρακτηριστικών που χρησιμοποιείται για τη δημιουργία του γράφου, περιλαμβάνει πληροφορίες που αφορούν αν ο χρήστης είναι verified, τον αριθμό μηνών που ο χρήστης είναι ενεργός, τον αριθμό ακολούθων, τον αριθμό φίλων, οι λίστες, τα αγαπημένα, τα statuses και το timestamp του tweet. Τα χαρακτηριστικά που χρησιμοποιούνται για τον κάθε κόμβο είναι το in-degree, δηλαδή ο αριθμός χρηστών που έκαναν mention τον χρήστη i , το out-degree, δηλαδή ο αριθμός των χρηστών που έγιναν mentioned από τον χρήστη i , το weighted in-degree δηλαδή ο αριθμός όλων των χρηστών που έκαναν mention τον χρήστη i και το weighted out-degree, δηλαδή τον συνολικό αριθμό χρηστών που ανέφερε ο χρήστης i . Στη συνέχεια, χρησιμοποιούνται οι 2-hop γείτονες, οι οποίοι δηλαδή βρίσκονται δύο ακμές μακριά από τον κόμβο που μελετάται και αφορούν τον αριθμό των χρηστών αυτών που κάνουν mention τον χρήστη i και τον αριθμό των χρηστών που έγιναν mentioned από τον χρήστη i και χρησιμοποιείται και ο αριθμός των tweets. Ο λόγος επιλογής αυτών των χαρακτηριστικών είναι το γεγονός ότι οι λιγότερο αξιόπιστοι ή γενικότερα οι αναξιόπιστοι χρήστες, πιθανόν να αλληλεπιδρούν μεταξύ τους. Χρησιμοποιείται το DiffPool για την αναπαράσταση των κόμβων σε διανύσματα στον χώρο και μέσω του GraphSage, οι κόμβοι αναπαρίστανται ως διανύσματα, τα οποία χρησιμοποιούνται για να ταξινομήσουν τις ειδήσεις σε ψευδείς και αληθείς.

Στο [39] δημιουργείται ένα μοντέλο γνωστό ως MVAN το οποίο έχει σαν σκοπό μέσω της αξιοποίησης του κειμένου των ειδήσεων και της δομής του γράφου που αναπαριστά τη διαδικασία διάδοσης ειδήσεων στο twitter από χρήστες, να προβλέψει κατά πόσο το εκάστοτε tweet περιέχει ψευδείς ή αληθείς ειδήσεις και αποτελείται από τα εξής στοιχεία: Τη χρήση word2vec για την αναπαράσταση κάθε λέξης σε κάθε tweet. Έπειτα, τη χρήση BiGRU, η οποία δημιουρ-

γεί το σημασιολογικό διάνυσμα της είδησης και στη συνέχεια μέσω δικτύου κειμένου προσοχής χρησιμοποιώντας τον μηχανισμό προσοχής, έχει σαν σκοπό τη σημασιολογική αναπαράσταση του κειμένου – είδησης, λαμβάνοντας υπόψη πόσο σημαντική είναι η εκάστοτε λέξη στον κείμενο. Επίσης δημιουργείται ένας γράφος, η ρίζα του οποίου είναι τα χαρακτηριστικά του χρήστη που έκανε tweet την είδηση και οι άλλοι κόμβοι τα χαρακτηριστικά των χρηστών που το έκαναν retweet. Ο λόγος που συμβαίνει αυτό είναι για να αναδείξει το γεγονός ότι σύμφωνα με αρκετές έρευνες, υπάρχει διαφορετικό σχήμα διάδοσης μεταξύ αληθινών και ψευδών ειδήσεων. Έπειτα γίνεται χρήση του μηχανισμού προσοχής για να πάρουμε την αναπαράσταση των κόμβων που αναπαράγουν τις ειδήσεις και να γίνει αντιληπτός ο ρόλος των χαρακτηριστικών του κάθε retweeter. Αργότερα, το διάνυσμα ως προς τα χαρακτηριστικά και το διάνυσμα ως προς το κείμενο συνενώνονται και μέσω σιγμοειδούς συνάρτησης, γίνεται η ταξινόμηση των ειδήσεων με πάρα πολύ καλά αποτελέσματα.

Στο [40], χρησιμοποιούνται και πάλι τα χαρακτηριστικά χρηστών twitter, σε συνδυασμό με το κείμενο του source tweet, το οποίο ουσιαστικά είναι η πηγή της είδησης. Τα χαρακτηριστικά του χρήστη αναπαρίστανται ως ένα διάνυσμα και έπειτα τα tweets γίνονται pad για να έχουν το ίδιο μήκος και μέσω GRU γίνεται η κωδικοποίηση του tweet, η οποία παράγει ένα διάνυσμα το οποίο περιλαμβάνει τη σημασιολογία του κειμένου. Αργότερα, γίνεται η μοντελοποίηση της διάδοσης του source tweet, καθώς στο άρθρο αυτό υποστηρίζεται ότι η διάδοση ειδήσεων μεταξύ των χρηστών που διαδίδουν ψευδείς ειδήσεις, έχει διαφορετικό σχήμα από αυτούς που διαδίδουν ψευδείς. Χρησιμοποιούνται τα χαρακτηριστικά των χρηστών που κάνουν retweet και με τη βοήθεια της GRU αναπαρίσταται η διάδοση και μέσω CNN, εξακριβώνεται η συσχέτιση μεταξύ των χαρακτηριστικών των χρηστών. Δημιουργείται ένας γράφος που αναπαριστά τη διάδοση των ειδήσεων στο twitter και βασίζεται στο γεγονός ότι η αλληλεπίδραση μεταξύ των χρηστών, δίνουν την πιθανότητα μια είδηση να είναι ψευδής ή αληθής. Σε κάθε ακμή στο γράφο ανατίθεται ένα βάρος, το οποίο προκύπτει μέσω της ομοιότητας συνημιτόνου μεταξύ των χαρακτηριστικών των χρηστών, τα οποία αποτελούν τους κόμβους και μέσω GCN δημιουργείται η αναπαράσταση διάδοσης. Ακολουθεί ένας διττός μηχανισμός προσοχής – dual co-attention mechanism, ο οποίος βρίσκει τη σχέση μεταξύ source tweet και διανύσματος διάδοσης μέσω CNN και τη σχέση μεταξύ source tweet και μεταξύ διάδοσης GCN, δηλαδή της αλληλεπίδρασης μεταξύ των χρηστών. Ο μηχανισμός προσοχής λαμβάνει υπόψη τα διανύσματα της διάδοσης μέσω CNN στην αναπαράσταση του source tweet και αντίθετα, ενώ στο δεύτερο κομμάτι λαμβάνει υπόψη τα διανύσματα της διάδοσης μέσω GCN στην αναπαράσταση του source tweet και αντίθετα. Αργότερα μέσω του μηχανισμού προσοχής, δημιουργούνται νέα διανύσματα που λαμβάνουν υπόψη τα πιο σημαντικά χαρακτηριστικά ως προς την κάθε αναπαράσταση, συνδέοντας τα χαρακτηριστικά των χρηστών, με το αντίστοιχο κείμενο που διαδίδουν και τη γενικότερη διάδοση μεταξύ των χρηστών. Έτσι παράγονται τα τελικά διανύσματα για κάθε κομμάτι του διττού μηχανισμού, τα οποία συνενώνονται με το διάνυσμα διάδοσης μέσω GRU και αργότερα με τη χρήση της σιγμοειδούς συνάρτησης softmax, ταξινομούνται οι ειδήσεις σε ψευδείς ή αληθείς, με πολύ καλά αποτελέσματα.

Στο [41], το οποίο βασίζεται σε μια διαφορετική προσέγγιση για το [36], μελετάται ο εντοπισμός ψευδών ειδήσεων με τη χρήση βοηθητικών κείμενων, χρησιμοποιώντας ως εισόδους ισχυρισμούς, τις αποδείξεις που υποστηρίζουν ή διαψεύδουν τον ισχυρισμό και την αντίστοιχη πηγή που διαδίδει τους ισχυρισμούς (sharers) και τις αποδείξεις (publishers). Για να εξακριβωθούν οι σχέσεις και οι εξαρτήσεις μεταξύ των πληροφοριών, οι ισχυρισμοί και οι αποδείξεις τους, μετατρέπονται σε γράφο με τεχνική συρόμενου παραθύρου, όπου επιλέγεται ένας συγκεκριμένος αριθμός λέξεων και στο κέντρο κάθε παραθύρου είναι μια λέξη που συνδέεται με όλες τις λέξεις που περιέχονται στο παράθυρο, τεχνική η οποία βοηθάει στην αντίληψη του τοπικού περιεχομένου της εκάστοτε γειτονιάς της κεντρικής λέξης. Για να εξακριβωθούν οι εξαρτήσεις μεταξύ των πληροφοριών, δημιουργείται ένας ενιαίος γράφος, ο οποίος περιέχει όλες τις λέξεις που αφορούν τόσο πληροφορίες ως προς τους ισχυρισμούς, όσο και πληροφορίες ως προς τις αποδείξεις, για να αντιληφθεί όλες τις πληροφορίες γύρω από την κάθε λέξη. Για να μπορούν οι κόμβοι να αντιληφθούν τις σημασιολογικές πληροφορίες αλλά ταυτοχρόνως στο διάνυσμά τους να συμπεριλαμβάνεται η διάδοση των ειδήσεων, χρησιμοποιούνται νευρωνικά δίκτυα γράφων πύλης (Gated Graph Neural Networks), τα οποία ουσιαστικά είναι ένας συνδυασμός GRU με GNNs. Παρακάτω θεωρήθηκε σημαντικό να εντοπιστεί ο πλεονασμός στους γράφους των αποδείξεων και μέσω ενός σκορ με τη χρήση και πάλι νευρωνικών δικτύων γράφων πύλης ενός επιπέδου για κάθε κόμβο, οι κόμβοι με το πιο ψηλό σκορ, αφαιρούνται από τον γράφο, με σκοπό την μείωση των δεδομένων και την αφαίρεση των δεδομένων που επαναλαμβάνουν την ίδια πληροφορία. Ο λόγος υλοποίησης αυτής της ενέργειας είναι το γεγονός ότι οι επαναλαμβανόμενες πληροφορίες, ενδέχονται να παραπλανήσουν το μοντέλο. Έπειτα μέσω χρήσης των GGNNs τα διάνυσματά πλέον συμπεριλαμβάνουν όλες τις χρήσιμες πληροφορίες, τόσο ως προς τους ισχυρισμούς, όσο και ως προς τις αποδείξεις. Στη συνέχεια, χρησιμοποιώντας ουσιαστικά τον μηχανισμό προσοχής, σε ένα ενιαίο γράφο, λαμβάνεται υπόψη πόσο σημαντική είναι η κάθε λέξη ισχυρισμού ως προς την εκάστοτε απόδειξη, συνδυάζοντας στην ανανεωμένη αναπαράσταση της απόδειξης, το διάνυσμα του γεγονότος. Το διάνυσμα ισχυρισμού συνενώνεται με το διάνυσμα του sharer του και το διάνυσμα απόδειξης συνενώνεται με το διάνυσμα του δικού του publisher. Χρησιμοποιείται και πάλι ο μηχανισμός προσοχής για να εξακριβωθεί πόσο σημαντικές είναι οι λέξεις του ισχυρισμού ως προς την απόδειξη, δημιουργώντας ένα διάνυσμα το οποίο πλέον συμπεριλαμβάνει πληροφορίες από τους ισχυρισμούς, τους publishers και τη σημασία τους. Τέλος, το διάνυσμα των ισχυρισμών συνενώνεται με το διάνυσμα αποδείξεων για μια τελευταία φορά και μέσω MLP και της συνάρτησης softmax, γίνεται η ταξινόμηση των ειδήσεων.

6.4. Εντοπισμός Ψευδών Ειδήσεων με βάση τη Γνώση

Στην περίπτωση που θέλουμε να βασίσουμε στον εντοπισμό των ψευδών ειδήσεων, στην γνώση που περιέχει η είδηση, μια είδηση χαρακτηρίζεται ως ψευδής, όταν το κείμενο δεν αποτελείται από γεγονότα. Η διαδικασία αυτή, χωρίζεται σε δύο μέρη: Στον χειροκίνητο έλεγχο δεδομένων (επεξηγείται πλήρως στο 1^ο Κεφάλαιο) και στον αυτόματο έλεγχο δεδομένων. Λόγω της ραγδαίας κλιμάκωσης της δημιουργίας ειδήσεων, ο χειροκίνητος έλεγχος είναι ένα

χρονοβόρο μέσο, το οποίο δεν μπορεί να εξυπηρετήσει πλήρως τις συνεχώς αυξανόμενες ανάγκες της αναζήτησης ψευδών ειδήσεων. Έτσι, στρεφόμαστε στον αυτόματο έλεγχο, όπου βασίζομαστε σε διάφορες τεχνικές επεξεργασίας γλώσσας και μηχανικής μάθησης. Για την αξιοποίηση της γνώσης είναι απαραίτητη η δημιουργία ενός γράφου γνώσης. Ένας γράφος γνώσης, αναφέρεται σε ένα γράφο σημασιολογικού δικτύου, ο οποίος αποτελείται από διαφορετικές οντότητες, έννοιες και σχέσεις, που αφορούν δεδομένα της καθημερινότητας. Οι γράφοι γνώσης αναπαρίστανται σε λεγόμενα triplets, τα οποία αποτελούνται από το σύνολο οντοτήτων, το σύνολο σχέσεων και το σύνολο γεγονότων (όπου τα γεγονότα ουσιαστικά είναι και πάλι οντότητες). Ουσιαστικά, τα triplets αποτελούνται από το υποκείμενο – ρήμα – αντικείμενο μιας πρότασης, τα οποία μετατρέπουν το περιεχόμενο της εκάστοτε πρότασης στην απαραίτητη μορφή, ώστε να μπορεί να μοντελοποιηθεί ως γράφος.

Βλέποντας την πρόσφατη βιβλιογραφία, στο [42] χρησιμοποιείται ένας συνδυασμός της αναπαράστασης της διάδοσης με την αναπαράσταση της γνώσης, με βάση άρθρα που δίνονται. Αρχικά, για κάθε άρθρο χρησιμοποιείται το εργαλείο OpenIE για την εξαγωγή triplets από την κάθε πρόταση, τα οποία αποτελούνται από υποκείμενο-ρήμα-αντικείμενο. Με τη χρήση του εργαλείου NeuralCoref αποσαφηνίζονται οι ασαφείς οντότητες και με τη χρήση μιας εξωτερικής βάσης γνώσης, συγκεντρώνονται όλες οι πληροφορίες που αφορούν την κάθε οντότητα. Όσο αφορά τη διαδικασία προ-επεξεργασίας, δεν μπορούν να αφαιρεθούν συγκεκριμένα σημεία του κειμένου, καθώς ενδέχεται να επηρεάσουν τη σημασιολογία του, γι' αυτό το λόγο γίνεται lemmatization μέσω SpaCy, για να μειωθεί ο αριθμός τόσο των οντοτήτων, όσο και των μεταξύ τους σχέσεων. Τα triplets που δημιουργήθηκαν, χρησιμοποιήθηκαν για τη δημιουργία του γράφου γνώσης. Γίνεται χρήση του Compositional GCN, όπου σε πρώτη φάση δημιουργούνται διανύσματα για κάθε κόμβο-οντότητα του γράφου και κάθε σχέση και σε δεύτερη φάση, μέσω της συνάρτησης ConVE, υπολογίζεται η ομοιότητα του διανύσματος που προκύπτει για το αντικείμενο, χρησιμοποιώντας τα διανύσματα του υποκειμένου και της μεταξύ τους σχέσης και το αρχικό διανύσματός του αντικειμένου. Έπειτα, μέσω χρήσης GCN, αναπαρίσταται η διάδοση των πληροφοριών στον γράφο γνώσης, με τη δημιουργία διανύσματος, το οποίο λαμβάνει υπόψη τις πληροφορίες των γειτονικών κόμβων-οντοτήτων και μέσω pooling, δημιουργείται το διάνυσμα για κάθε άρθρο. Με τη χρήση του μηχανισμού προσοχής, εξακριβώνεται πόσο σημαντική είναι η κάθε οντότητα ως προς το εκάστοτε κείμενο και δημιουργείται ένα διάνυσμα που συμπεριλαμβάνει αυτή την πληροφορία. Τέλος, συνενώνονται τα διανύσματα διάδοσης και το σχεσιακό διάνυσμα ειδήσεων που δημιουργήθηκε μέσω μηχανισμού προσοχής και μέσω MLP γίνεται η ταξινόμηση των ειδήσεων.

Στο [43] χρησιμοποιείται άρθρο ειδήσεων, σε συνδυασμό με επιπρόσθετες πληροφορίες ως προς αυτό. Για κάθε άρθρο ειδήσεων, δημιουργείται ένας ετερογενής κατευθυνόμενος γράφος κειμένου, χρησιμοποιώντας τρεις τύπους οντοτήτων: προτάσεις, θέματα και οι γνωστές οντότητες – ουσιαστικά. Τα σετ των ακμών αναπαριστούν τις σχέσεις μεταξύ προτάσεων, θεμάτων και οντοτήτων. Το κάθε άρθρο είδησης, χωρίζεται σε προτάσεις και εξακριβώνεται η αλληλεπίδραση μεταξύ τους. Για την εξαγωγή του θέματος κάθε πρότασης του κάθε άρθρου, εφαρμόζεται το unsupervised LDA, όπου είναι ένα εργαλείο που εντοπίζει αυτόματα το θέμα του

κειμένου που μελετάται. Αφού βρεθούν οι οντότητες μέσω NER, αντιστοιχίζονται στο Wikipedia μέσω του TAGME και μαζεύονται όλες οι πληροφορίες που τις αφορούν. Αν μια πρόταση έχει μια οντότητα, δημιουργείται μια ακμή μεταξύ πρότασης και οντότητας. Για κάθε πρόταση, μέσω LSTM δημιουργείται το σημασιολογικό διάνυσμα. Για να αναπαρασταθεί η διάδοση πληροφοριών μεταξύ των κόμβων του ετερογενή γράφου, χρησιμοποιείται ένα HGNC, όπου γίνεται η διάδοση γνώσης μεταξύ κόμβων με ίδιο θέμα και έτσι δημιουργείται διάνυσμα για κάθε οντότητα. Για τον ορισμό των επιπρόσθετων πληροφοριών, χρησιμοποιείται η Wikipedia ως βάση γνώσης και σε πρώτη φάση μέσω TransE δημιουργείται ένα δομικό διάνυσμα για κάθε triplet της βάσης, υπολογίζοντας τιμές για τα διανύσματα υποκειμένων h , σχέσεων r , αντικειμένων t , ως $h+r=t$. Για κάθε οντότητα, χρησιμοποιείται επίσης η πρώτη παράγραφος της στην Wikipedia, της οποίας το διάνυσμα προκύπτει μέσω LSTM. Μέσω ενός παράγοντα πύλης, συνδυάζονται τα δύο διανύσματα και δημιουργούν ένα ενιαίο διάνυσμα της εξωτερικής γνώσης, για τις οντότητες της βάσης. Μέσω συνάρτησης σύγκρισης, συγκρίνεται το διάνυσμα από τη βάση γνώσης, με το διάνυσμα των οντοτήτων από τον ετερογενή γράφο και δημιουργείται ένας συντελεστής σύγκρισης, που δείχνει κατά πόσο εμπεριέχουν κοντινά διανύσματα και κατ' επέκταση παρόμοιο νοήματος. Ο συντελεστής αυτός, σε συνδυασμό με το σημασιολογικό διάνυσμα ανά πρόταση μέσω LSTM χρησιμοποιούνται για την τελική ταξινόμηση των ειδήσεων.

Στο [44], γίνεται χρήση μόνο των τίτλων των ειδήσεων και αποτελείται από δύο κωδικοποιητές: έναν ειδήσεων και ένα οντοτήτων. Για την κωδικοποίηση των ειδήσεων, για τον κάθε τίτλο, δημιουργείται ένα διάνυσμα που εμπεριέχει τη σημασιολογία του μέσω χρήσης δύο BiLSTM. Για την κωδικοποίηση των οντοτήτων κάθε είδησης, λαμβάνεται υπόψη το περιεχόμενο των ειδήσεων, καθώς και η συσχέτιση των οντοτήτων. Αρχικά, εφαρμόζεται εντοπισμός οντοτήτων (NER) μέσω του RoBERTa, μέσω του NeuralCoref αποσαφηνίζονται οι οντότητες και μαζεύονται όλες οι πληροφορίες που τις αφορούν μέσω του γράφου γνώσης WikidataKG. Με τη χρήση του ComplEx, παράγονται διανύσματα για κάθε οντότητα του γράφου γνώσης και με τη χρήση mean pooling, δημιουργείται ένα ολικό διάνυσμα για κάθε τίτλο. Τέλος, χρησιμοποιείται ένα τελευταίο επίπεδο που ενώνει τα διανύσματα των οντοτήτων και των ειδήσεων και τροφοδοτούνται σε μια σιγμοειδή συνάρτηση, όπου ταξινομούνται οι ειδήσεις σε ψευδείς και αληθείς.

Στο [45] το οποίο βασίζεται στο [46], έχουμε το μοντέλο γνωστό ως DETERRENT, το οποίο χρησιμοποιεί διανύσματα που αφορούν τη σημασιολογία του κειμένου που δίνεται, σε συνδυασμό με διανύσματα που αφορούν τη διάδοση της πληροφορίας σε γράφο γνώσης που περιέχει χρήσιμες πληροφορίες ως προς το θέμα που μελετάται. Αρχικά, χρησιμοποιείται ένας γράφος γνώσης, γνωστός και σαν MedicalKG, τα triplets του οποίου έχουν εξαχθεί μέσω άρθρων που αφορούν αποκλειστικά θέματα υγείας. Ο γράφος περιέχει θετικές και αρνητικές σχέσεις και κατ' επέκταση triplets (Diabetes – Causes – Kidney Diseases => θετικό triplet, Herbal Supplement – DoesNotHeal – Obesity => αρνητικό triplet). Έπειτα χρησιμοποιείται ένα σύνολο άρθρων, των οποίων οι πληροφορίες συγκεντρώνονται, μέσω εντοπισμού οντοτήτων που περιέχονται στο MedicalKG και προσαρτώνται σε ένα ολικό γράφο με τις πληροφορίες του

MedicalKG. Για το διάνυσμα της διάδοσης των πληροφοριών στον γράφο γνώσης, μέσω του μηχανισμού προσοχής, εντοπίζονται οι πιο σημαντικές σχέσεις και λαμβάνονται υπόψη στο ανανεωμένο διάνυσμα κάθε κόμβου, που περιέχει πληροφορίες των γειτονικών του, μέσω R-GAT. Όσο αφορά, τη σημασιολογία του κειμένου, μέσω BiGRU δημιουργούνται διανύσματα για κάθε κείμενο, το οποίο μέσω μηχανισμού προσοχής, λαμβάνει υπόψη πόσο σημαντική είναι η κάθε λέξη ως προς το κείμενο. Τέλος, συνδυάζονται τα διανύσματα διάδοσης στον γράφο γνώσης και περιεχομένου του κειμένου και μέσω σιγμοειδούς συνάρτησης, ταξινομούνται οι ειδήσεις. Αξίζει να αναφερθεί ότι το μοντέλο του DETERRENT είχε την καλύτερη απόδοση από κάθε άλλο μοντέλο με βάση τη γνώση, σε ολόκληρη τη βιβλιογραφία.

Στο [47], χρησιμοποιείται ένα πολυτροπικό μοντέλο το KMGCN (Knowledge-driven Multimodal Graph Convolutional Networks), το οποίο αποσκοπεί στον εντοπισμό ψευδών ειδήσεων, σε ένα σύνολο αναρτήσεων ειδήσεων, σε συνδυασμό με τις αντίστοιχες τους εικόνες. Αρχικά γίνεται εξόρυξη της χρήσιμης γνώσης από τις αναρτήσεις που μελετώνται, μέσω εντοπισμού των οντοτήτων και αφού εξαχθούν όλες οι πληροφορίες που τις αφορούν από άλλες βάσεις γνώσης. Στη συνέχεια μέσω της χρήσης ενός σκορ γνωστού ως PMI (Point-wise Mutual Information), υπολογίζεται πόσο συχνά εμφανίζεται μια λέξη κοντά σε μια άλλη, δηλαδή την συνεμφάνισή τους, για εξακρίβωση της μεταξύ τους σχέσης, κάτι το οποίο βοηθάει στο να συγκεντρωθεί η σημαντική πληροφορία. Έπειτα μέσω του YoLoV3, εξακριβώνονται οι πληροφορίες που περιέχει κάθε εικόνα, σε μορφή λέξεων και αφού ολοκληρωθεί αυτή η διαδικασία, συμπεριλαμβάνονται ως μέρος του κειμένου είδησης. Επίσης, οποιαδήποτε πληροφορία έχει εξαχθεί από άλλες βάσεις γνώσης, χρησιμοποιείται ως κομμάτι της είδησης. Με τη χρήση του word2vec δημιουργούνται αρχικά διανύσματα για κάθε κόμβο του γράφου που δημιουργήθηκε, ο οποίος περιέχεται μόνο από λέξεις, που είναι οι πληροφορίες για κάθε είδηση. Μέσω GCN, δημιουργούνται διανύσματα που περιέχουν πληροφορίες των γειτονικών του κόμβων – λέξεων και μέσω pooling δημιουργείται ένα ολικό διάνυσμα για την είδηση, η οποία ταξινομείται μέσω σιγμοειδούς συνάρτησης.

Στο [48], δημιουργείται ένα μοντέλο γνωστό ως HC-COVID. Το πρώτο επίπεδο ονομάζεται ως CKGC (Claim-guided Knowledge Graph Completion), το οποίο χωρίζεται σε δύο φάσεις: στη δημιουργία γράφων ισχυρισμών και γενικών πληροφοριών για τη νόσο. Για την εξαγωγή πληροφοριών, σε πρώτη φάση μια ομάδα επιστημόνων, προτείνει συγκεκριμένα θέματα που αφορούν τον COVID-19 και έπειτα μαζεύει γενικές πληροφορίες σε μορφή triplets που αφορούν τη νόσο, σύμφωνα με τα θέματα που προτάθηκαν. Σύμφωνα με τα θέματα αυτά, μια ομάδα από απλούς πολίτες εξάγουν triplets από διάφορα άρθρα που αφορούν ισχυρισμούς με βάση τη νόσο. Δημιουργούνται οι δύο προαναφερθέντες γράφοι και ένας γενικός που περιέχει τόσο γεγονότα, όσο και γενικές πληροφορίες τα οποία συνδέονται αν αφορούν το ίδιο θέμα. Έπειτα ακολουθεί το επίπεδο που ονομάζεται CKSP (Claim-guided Specific Knowledge Propagator) όπου σκοπός είναι η αναπαράσταση της διάδοσης των ισχυρισμών και η κωδικοποίησή τους σύμφωνα με τα συμφραζόμενα. Οι αναπαραστάσεις των ισχυρισμών με βάση τα συμφραζόμενα, δημιουργούνται μέσω BiGRU, αφού διαχωριστεί ο κάθε ισχυρισμός σε λέξεις. Μέσω R-GCN διαδίδονται οι πληροφορίες μεταξύ των κόμβων, οι οποίες περιέχουν τις εξόδους της

BiGRU. Το R-GCN ουσιαστικά είναι ένα GCN, το οποίο πραγματοποιεί διαφορετικούς υπολογισμούς για κάθε σχέση. Έπειτα μέσω του TGKI (Topic-based Generalized Knowledge Integrator) χρησιμοποιείται ένας μηχανισμός συν-προσοχής, ο οποίος υπολογίζει τη συσχέτιση του εκάστοτε ισχυρισμού με όλους τους άλλους και τη συσχέτιση μεταξύ ισχυρισμών και γενικών γεγονότων. Επίσης, μέσω του μηχανισμού προσοχής, εξακριβώνεται πόσο σημαντικό είναι το κάθε triplet ως προς τον κάθε γράφο ισχυρισμών και γενικών γεγονότων αντίστοιχα. Το τελευταίο επίπεδο CGMD (Joint Claim-Graph-based Misinformation Detector), μέσω συνένωσης των διανυσμάτων συσχέτισης από το TGKI, των σημασιολογικών διανυσμάτων κειμένου μέσω BiGRU και των διανυσμάτων διάδοσης μέσω R-GCN που προαναφέρθηκαν από το δεύτερο επίπεδο CSKP, ταξινομεί τα γεγονότα σε ψευδή και αληθή, έχοντας πολύ καλά αποτελέσματα.

Στο [49], χρησιμοποιούνται κείμενα των ειδήσεων και μέσω συνδυασμού των διανυσμάτων του κειμένου, των χαρακτηριστικών του κειμένου και της εικόνας που το συνοδεύει, γίνεται η ταξινόμηση των ειδήσεων σε ψευδείς και αληθείς. Αφού εντοπιστούν οι οντότητες σε κάθε κείμενο, χρησιμοποιείται η τεχνική entity linking κατά την οποία συγκεντρώνονται όλες οι διαθέσιμες πληροφορίες για κάθε οντότητα. Έπειτα δημιουργείται ένας γράφος γνώσης που περιέχει όλες τις πληροφορίες για το εκάστοτε κείμενο. Ακολουθεί η δημιουργία τριών ξεχωριστών γράφων, οι οποίοι βρίσκουν: α) την συνεμφάνιση όλων των ζευγών λέξεων του κειμένου, σε όλο το σύνολο δεδομένων και κατ' επέκταση τη μεταξύ τους συσχέτιση, β) τη συνεμφάνιση των ζευγών λέξεων στο κάθε κείμενο ξεχωριστά και γ) την ομοιότητα μεταξύ των λέξεων του κειμένου. Οι γράφοι χρησιμοποιούνται ως ο πίνακας γειτνίασης του γράφου γνώσης και ο πίνακας χαρακτηριστικών που χρειάζεται, αρχικοποιείται ως τα διανύσματα κάθε λέξης, μέσω word2vec. Οι δύο πίνακες χρησιμοποιούνται για την εφαρμογή του GCN, η οποία αποσκοπεί στη διάδοση των πληροφοριών μεταξύ των λέξεων – κόμβων και το σύνολο λέξεων χρησιμοποιείται για την τελική αναπαράσταση του κειμένου ειδήσεων. Μέσω VGG-19 και ενός πλήρως-συνδεδεμένου επιπέδου, εξάγονται πληροφορίες από την εκάστοτε εικόνα σε μορφή λέξεων. Μέσω του μηχανισμού προσοχής εξακριβώνεται πόσο σημαντική είναι η κάθε λέξη του κειμένου και αυτό λαμβάνεται υπόψη στη συσχέτιση μεταξύ της αναπαράστασης της εικόνας με την εκάστοτε λέξη του κειμένου. Δημιουργείται ένα διάνυσμα για την εικόνα που λαμβάνει υπόψη πόσο σημαντική είναι η κάθε λέξη του κειμένου. Η αναπαράσταση του κειμένου συνενώνεται με την αναπαράσταση της εικόνας και γίνεται η ταξινόμηση των ειδήσεων, έχοντας πολύ καλά αποτελέσματα.

Στο [50], συνδυάζονται διανύσματα ως προς το στυλ γραφής και διανύσματα ως προς τον γράφο γνώσης, για τον εμπλουτισμό της αναπαράστασης της κάθε είδησης. Αρχικά, γίνεται η προ-επεξεργασία του κειμένου, κατά την οποία όλα τα γράμματα γίνονται πεζά, αφαιρούνται τα hashtags, τα punctuations και τα stopwords. Έπειτα, χρησιμοποιείται η τεχνική n-gram, όπου το n παίρνει τιμές 1,2 για λέξεις και 1,2,3 για χαρακτήρες και μέσω TF-IDF βρίσκεται πόσο σημαντική είναι το κάθε n-gram, ως προς το κάθε κείμενο. Μέσω DistilBert, RoBERTa και XLM, δημιουργούνται σημασιολογικά διανύσματα για κάθε κείμενο και ο μέσος όρος των

διανυσμάτων, δημιουργεί την τελική αναπαράσταση. Όσον αφορά τον γράφο γνώσης, χρησιμοποιείται ο WikiData5m ως εξωτερική βάση γνώσης, από την οποία και εντοπίζονται πληροφορίες για τις οντότητές του περιέχονται στο κείμενο που μελετάται και αφού προηγηθεί η προ-επεξεργασία κειμένου όπως στο πρώτο κομμάτι, χρησιμοποιούνται τα παρακάτω εργαλεία με τον σκοπό δημιουργίας αριθμητικών αναπαραστάσεων των οντοτήτων και των περιεχομένων του γράφου: Αρχικά χρησιμοποιείται το TransE, η λειτουργία του οποίου επεξηγήθηκε πιο πάνω. Έπειτα χρησιμοποιείται το RotatE, το οποίο ουσιαστικά θεωρεί την απόσταση μεταξύ του διανύσματος του υποκειμένου και του αντικειμένου, ως μια περιστροφή, δηλαδή ουσιαστικά μια γωνία, την οποία και υπολογίζει και ισούται με τη σχέση. Το επόμενο εργαλείο που χρησιμοποιείται είναι το ComplEx, το οποίο δημιουργεί διανύσματα των υποκειμένων, σχέσεων του γράφου γνώσης, ενώ το αντικείμενο πέραν από το πραγματικό κομμάτι του διανύσματος έχει και φανταστικό και η αναπαράσταση του triplet, γίνεται μέσω του αθροίσματος των τριών προαναφερθέντων διανυσμάτων. Ακολουθεί το QuatE, το οποίο ουσιαστικά κάνει ακριβώς το ίδιο με το ComplEx, αλλά τα διανύσματα αναπαρίστανται σε τετραδιάστατο χώρο. Στη συνέχεια, χρησιμοποιείται το DistMult, το οποίο χρησιμοποιεί ένα ενιαίο διάνυσμα για το υποκείμενο και την σχέση, μέσω υπολογισμού ομοιότητας συνημιτόνου μεταξύ υποκειμένου και αντικειμένου. Τέλος, χρησιμοποιείται το Simple, το οποίο μέσω συνάρτησης scoring δημιουργεί διανύσματα για τη σχέση και ξεχωριστά διανύσματα για τις οντότητες, βασισμένα στο μέρος λόγου που αποτελούν (υποκείμενο, αντικείμενο), παρατηρώντας επίσης και την επιρροή του υποκειμένου όταν χρησιμοποιείται ως αντικείμενο και το διάνυσμα σχέσης αντιστρέφεται, δηλαδή υπολογίζει το μέσο όρο: $(h,r,t) + (t,r^{-1},h)$. Το τελικό διάνυσμα για κάθε κείμενο προκύπτει από το μέσο όρο των τιμών όλων των προηγούμενων εργαλείων. Αφού συνδυαστούν τα διανύσματα του γράφου γνώσης και του στυλ γραφής, χρησιμοποιείται ένα βαθύ νευρωνικό δίκτυο, το οποίο έπειτα μέσω σιγμοειδούς συνάρτησης, ταξινομεί τις ειδήσεις σε αληθείς και ψευδείς, με πολύ καλά αποτελέσματα.

Εικόνα 16: Πίνακας Βιβλιογραφίας

| <u>Εντοπισμός Ψευδών Ειδήσεων βασιζόμενοι αποκλειστικά στο περιεχόμενο</u> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A. B. Presidio, R. R. Isnanto, D. Eridian, Y. A. A. Soetrisno, M. Arfan, and A. Sofwan, "Hoax detection system on Indonesian news sites-based text classification using SVM and SGD," in Proc. 4th Int. Conf. Inf. Technol., Compute., Electra. Eng. (ICITACEE), Oct. 2017, pp. 45–49 |
| Kaur S, Kumar P, Kumaraguru P (2020) Automating fake news detection system using multi-level voting model. Soft Comput 24:9049–9069 |
| Patil, Dharmaraj R. "Fake News Detection Using Majority Voting Technique." arXiv preprint arXiv:2203.09936 (2022). |

Εντοπισμός Ψευδών Ειδήσεων βασίζόμενοι στο περιεχόμενο, λαμβάνοντας υπόψη τα συμφοραζόμενα

Mehta D, Dwivedi A, Patra A, Anand Kumar M (2021) A transformer-based architecture for fake news classification, *Social Network Anal Mining*, 11(39)

Saleh H, Alharbi A, Alsamhi SH (2021) OPCNN-FAKE: optimized convolutional neural network for fake news detection. *IEEE Access* 9:129471–129489

Kaliyar, R.K., Goswami, A. & Narang, P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimed Tools Appl* 80, 11765–11788 (2021). <https://doi.org/10.1007/s11042-020-10183-2>

Defend: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM Conferences from <https://dl.acm.org/doi/10.1145/3292500.3330935>

Vo, N., & Lee, K. Hierarchical multi-head Attentive Network for evidence-aware fake news detection

Truică, C.-O.; Apostol, E.-S. MisRoBÆRTa: Transformers versus Misinformation. *Mathematics* 2022

Εντοπισμός Ψευδών Ειδήσεων μέσω δικτύου διάδοσης

G. Gadek and P. Guélorget, "An interpretable model to measure fakeness and emotion in news," *Procedia Computer Science*, Volume 176, 2020. Graph Neural Networks with Continual Learning for Fake News Detection from social media [2020]

GCAN: Graph-aware Co-Attention Networks for Explainable Fake News Detection on social media (2020)

Ni S, Li J, Kao H-Y (2021) MVAN: multi-view attention networks for fake news detection on social media. *IEEE Access* 9:106907–106917

Galende, B.A.; Hernández-Peñaloza, G.; Uribe, S.; García, F.Á. Conspiracy or Not? A Deep Learning Approach to Spot It on Twitter. *IEEE Access* 2022

Weizhi Xu, Junfei Wu, Qiang Liu, Shu Wu, and Liang Wang. 2022. Evidence-aware Fake News Detection with Graph Neural Networks

Εντοπισμός Ψευδών Ειδήσεων μέσω γνώσης

DETERRENT: Knowledge Guided Graph Attention Network for Detecting Healthcare Misinformation (2020)

Y. Wang, S. Qian, J. Hu, Q. Fang and C. Xu, "Fake News Detection via Knowledge-driven Multimodal Graph Convolutional Networks," ICMR '20: Proceedings of the 2020 International Conference on Multimedia Retrieval, June 2020.

Incorporating Relational Knowledge in Explainable Fake News Detection (2021)

Compare to The Knowledge: Graph Neural Fake News Detection with External Knowledge (ACL 2021)

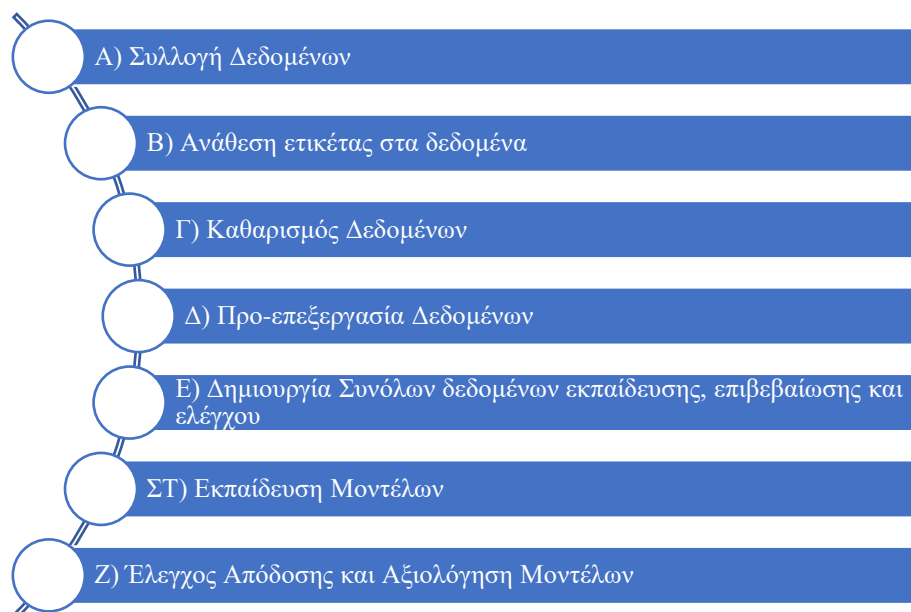
DEAP-FAKED: Knowledge Graph based Approach for Fake News Detection (2022)

HC-COVID: A Hierarchical Crowdsourced Knowledge Graph Approach to Explainable COVID-19 Misinformation Detection (2022)

Knowledge-aware multi-modal adaptive graph convolutional networks for fake news detection. ACM Transactions on Multimedia Computing, Communications, and Applications

Knowledge graph informed fake news classification via heterogeneous representation ensembles

7. Πειράματα



7.1. Συγκέντρωση Συνόλου Δεδομένων

Για τη συγκέντρωση του συνόλου δεδομένων (dataset), χρησιμοποιήθηκαν εργαλεία εξόρυξης κειμένου από το διαδίκτυο, για τη δημιουργία ενός ολοκληρωμένου dataset με ψευδείς και αληθείς ειδήσεις στην ελληνική γλώσσα. Για την εξόρυξη των ειδήσεων, χρησιμοποιήθηκαν το BeautifulSoup, το οποίο είναι μια βιβλιοθήκη της python, η οποία παίρνει html και xml αρχεία, τα οποία αναλύονται για να εξορυχθούν τα απαραίτητα κείμενα και το trafilatura, το οποίο επίσης είναι ένα πακέτο της python και επιτελεί παρόμοιες λειτουργίες, έχοντας επιπλέον σαν χαρακτηριστικό την οργάνωση κειμένου.

Οι επιβεβαιωμένες ψευδείς ειδήσεις εξορύχθηκαν από την ιστοσελίδα των ellinikahoaxes.gr, όπου εξειδικεύονται στον εντοπισμό ψευδών ειδήσεων. Τα [ellinikahoaxes](http://ellinikahoaxes.gr) στα άρθρα τους, συνήθως παραθέτουν την πρωτογενή πηγή της είδησης και αναλύουν τους λόγους μη εγκυρότητάς του. Αυτό που γίνεται είναι ότι μαζεύονται οι πρωτογενείς πηγές άρθρων, δηλαδή τα άρθρα τα οποία αναγράφουν την εκάστοτε είδηση και από αυτά εξορύσσουμε τους τίτλους τους, το κείμενό τους και τον αρθρογράφο (αν αναγράφεται). Αξίζει να σημειωθεί ότι χρησιμοποιήθηκαν ιστοσελίδες που περιέχουν τόσο αληθείς, όσο και ψευδείς ειδήσεις και τονίζεται ότι οι ειδήσεις που επιλέχθηκαν δεν ανήκουν στην σάτιρα – δηλαδή σε ιστοσελίδες όπως το koulouri.gr – καθώς ο σκοπός μας δεν είναι η ανίχνευση των σατιρικών ειδήσεων, αλλά των ειδήσεων που περιέχουν ψευδείς πληροφορίες. Για τις αληθινές ειδήσεις, χρησιμοποιήθηκαν οι ιστοσελίδες, οι οποίες σύμφωνα με πρόσφατη έρευνα του Amazon, θεωρούνται οι πιο αξιόπιστες στην Ελλάδα δηλαδή το in.gr, lifo.gr, dikaiologitika.gr, enimerwtiko.gr και το page-news.gr. Για το κάθε άρθρο ειδήσεων, πήραμε τον τίτλο του, το κείμενο του και τον αρθρογράφο. Εάν δεν υπάρχει αρθρογράφος, κρατάμε το domain της ιστοσελίδας που το φιλοξενεί.

Αφού μαζευτούν τα δεδομένα σε ένα dataframe, προστίθεται μια στήλη, η οποία τιτλοφορείται ως label και καθορίζει το κατά πόσο μια είδηση είναι αληθής ή ψευδής, με 1 να συμβολίζει την ψευδή είδηση και 0 την αληθή. Το σύνολο δεδομένων που δημιουργήθηκε με ψευδείς και αληθείς ειδήσεις είναι μοναδικό, καθώς σύμφωνα με την έρευνα που έγινε, δεν υπάρχει κάτι παρόμοιο.

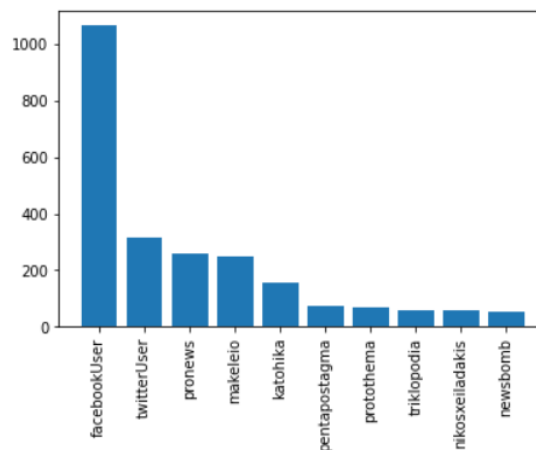
Εικόνα 17: Απεικόνιση συνόλου δεδομένων

| | title | text | domain-author | url | label | |
|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|---|
| 0 | Έβγαλαν τα μάτια από αδέσποτο σκυλί στο Μεσολόγγι | Ενώπιον ενός φρικτού και βάνουσου θεάματος βρέθηκαν το Σάββατο οι κάτοικοι του Μεσολογγίου , όταν εντόπισαν ένα αδέσποτο σκύλο , από τον οποίο είχαν αφαιρέσει τα μάτια . Το άτυχο τετράποδο εντοπίσ... | altsantiri | | altsantiri | 1 |
| 1 | Η δημοσιογράφος Χρύσα Ρουμελιώτη παραπήθηκε από την ΕΡΤ. Η παραίτησή της σχετίζεται με την ομιλία που παρέθεσε η Πρόεδρος της Δημοκρατίας κατά τη διάρκεια της Δεξίωσης για την 48η επέτειο αποκατά... | Η αποχώρηση της κ. Χρύσας Ρουμελιώτη από το ρεπορτάζ της Προεδρίας της Δημοκρατίας δημιούργησε δεκάδες άρθρα και δημοσιεύσεις στα μέσα κοινωνικής δικτύωσης που αναφέρουν ότι η δημοσιογράφος παραι... | fthis | | fthis | 1 |
| 2 | ΣΥΡΙΖΑ: Επιτέλους, ανοίγει το Μετρό στον Πειραιά | Στην ανακοίνωσή τους αναφέρουν ότι η νέα Γραμμή 3 του Μετρό που συνδέει τον Πειραιά με τη Νίκαια , τον Κορυθαλλό , την Αγία Βαρβάρα , το Αιγάλεω και με πολλούς σταθμούς της Αθήνας με τελική κατάλη... | newsroom | https://www.dikaiogitika.gr/eidhseis/politikes-eidhseis/403995/syriza-epitelous-anoigei-to-metro-ston-peiraia | | 0 |
| 3 | Είναι επίσημο: Ο Στέφανος Χίος κατεβαίνει στις εκλογές με δικό του κόμμα – «Εμείς δεν ήμαστε από σόγια» | Είναι επίσημο Ο Στέφανος Χίος κατεβαίνει στις εκλογές με δικό του κόμμα Εμείς δεν ήμαστε από σόγια Ο Στέφανος Χίος ανακοίνωσε πως θα κατέβει υποψήφιος στις επόμενες εθνικές εκλογές | Niki M. | https://www.enimerotiko.gr/politiki/einai-episimo-o-stefanos-chios-katevainei-stis-ekloges-me-diko-toy-komma-emeis-den-imaste-apo-sogia/ | | 0 |

7.2. Μελέτη Συνόλου Δεδομένων

Προτού αναλύσουμε τα πειράματα που εκτελέστηκαν στο σύνολο δεδομένων ελληνικών ειδήσεων είναι αρκετά σημαντικό να μελετήσουμε τα πιθανά μοτίβα που περιέχει. Αρχικά, παραθέτουμε το top10 των πηγών ψευδών ειδήσεων. Στην κορυφή, βρίσκονται οι χρήστες του κοινωνικού δικτύου facebook, ενώ αξιοσημείωτο είναι το γεγονός ότι στη δεύτερη θέση βρίσκονται οι χρήστες από άλλο ένα κοινωνικό δίκτυο, το twitter. Το top10 συμπληρώνουν ελληνικές ιστοσελίδες, αρκετές εκ των οποίων είναι γνωστές πηγές ψευδών ειδήσεων.

Εικόνα 18: Top10 πηγών ψευδών ειδήσεων



των αληθινών ειδήσεων είναι λιγότερο σε σχέση με τα ψευδή, με έκπληξη παρατηρήθηκε ότι το πλήθος λέξεων στις ψευδείς ελληνικές ειδήσεις είναι μικρότερο σε σχέση με τις αληθινές.

Εικόνα 21: Μέσος όρος λέξεων σε αληθινές και ψευδείς ειδήσεις

The average number of words in a real news article is 437.46110524463217
The average number of words in a fake news article is 320.36444352673965

7.3. Μετρικές Αξιολόγησης

Η ταξινόμηση των ειδήσεων σε ψευδείς και αληθείς, θεωρείται ένα binary task, δηλαδή υπάρχουν μόνο δύο τιμές, αληθής ή ψευδή είδηση. Για την ταξινόμηση των ειδήσεων, χρησιμοποιούνται οι ακόλουθες μετρικές: Accuracy, Precision, Recall και F1-Score, τα οποία αναλύονται στην Εικόνα 22.

Εικόνα 22: Ορισμός Παραμέτρων Μετρικών Αξιολόγησης

| | Predicted Positive | Predicted Negative |
|-------------------|---------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| Labelled Positive | <u>True Positive (TP):</u> Το μοντέλο που χρησιμοποιήσαμε προέβλεψε σωστά τις ειδήσεις ως ψευδείς | <u>False Negative (FN):</u> Το μοντέλο που χρησιμοποιήσαμε προέβλεψε λανθασμένα τις ειδήσεις ως αληθείς |
| Labelled Negative | <u>False Positive (FP):</u> Το μοντέλο που χρησιμοποιήσαμε προέβλεψε λανθασμένα τις ειδήσεις ως ψευδείς | <u>True Negative (TN):</u> Το μοντέλο που χρησιμοποιήσαμε προέβλεψε σωστά τις ειδήσεις ως αληθείς |

Το precision [51], μετράει την ικανότητα του μοντέλου να προβλέπει σωστά τα θετικά δείγματα, υπολογίζοντάς τα ως το πηλίκο των δειγμάτων που είναι πραγματικά θετικά δια τα δείγματα που προέβλεψε το μοντέλο ως θετικά. Το recall, μετράει την ικανότητα του μοντέλου να προβλέψει τις θετικές κλάσεις, υπολογίζοντας την αναλογία των πραγματικά θετικών δειγμάτων ως προς τα δείγματα που προέβλεψε το μοντέλο ως θετικά. Το accuracy μας δίνει την απόδοση του εκάστοτε μοντέλου, δηλαδή την ικανότητα του να κάνει σωστές προβλέψεις θετικών και αρνητικών δειγμάτων, ως προς το ολικό σύνολο δεδομένων, έχοντας υπόψη όμως ότι είναι πλήρως αξιόπιστο όταν το σύνολο δεδομένων είναι ισορροπημένο, δηλαδή στην περίπτωση μας, όταν το πλήθος των αληθινών και των ψευδών ειδήσεων είναι το ίδιο. Στην περίπτωση που το σύνολο δεδομένων δεν είναι ισορροπημένο, το F1-Score θεωρείται η πιο κα-

τάλληλη μετρική, καθώς απεικονίζει την ικανότητα του μοντέλου να προβλέψει σωστά τα θετικά και αρνητικά δείγματα, βλέποντας τις λανθασμένες και σωστές προβλέψεις θετικών και αρνητικών δειγμάτων, δηλαδή σωστά ταξινομημένων ψευδών και αληθών ειδήσεων.

Εικόνα 23: Ορισμός Μετρικών Αξιολόγησης

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, Precision = \frac{TP}{TP+FP}$$
$$Recall = \frac{TP}{TP+FN}, F1\ score = 2 * \frac{Precision*Recall}{Precision+Recall}$$

7.4. Διεκπεραίωση Πειραμάτων

Η γλώσσα η οποία επιλέχθηκε για την εκτέλεση των πειραμάτων είναι η python και τα πειράματα εκτελέστηκαν στο περιβάλλον του Google Colab. Για τη διαδικασία της προ-επεξεργασίας δεδομένων χρησιμοποιείται το spaCy, εάν δεν έχει το εκάστοτε μοντέλο, τη δική του διαδικασία προ-επεξεργασίας. Παρακάτω παρατίθενται τόσο τα baseline πειράματα που έγιναν και η πλήρης εξήγηση των μοντέλων τους, αλλά και τα δικά μας πειράματα, επίσης με την πλήρη εξήγηση της μεθοδολογίας που ακολουθήθηκε.

Η διεκπεραίωση πειραμάτων για εντοπισμό ψευδών ειδήσεων θα κινηθεί γύρω από τρεις πυρήνες: Την ανίχνευση με βάση το περιεχόμενο μιας είδησης, μελετώντας αποκλειστικά ειδήσεις σύμφωνα με το περιεχόμενο και το περιεχόμενο με βάση τα συμφραζόμενα, χρησιμοποιώντας μεθόδους Μηχανικής και Βαθιάς Μάθησης, την ανίχνευση με βάση την διάδοση των ειδήσεων και ανίχνευση με βάση την γνώση που περιέχουν οι ειδήσεις. Για την ανίχνευση σύμφωνα με το περιεχόμενο με βάση τη Μηχανική Μάθηση, εκτελούνται πειράματα με βάση baseline μεθόδους όπως FastText, Word2Vec, TF-IDF, GloVe και διανύσματα μέσω SpaCy. Για την ανίχνευση σύμφωνα με μεθόδους Βαθιάς Μάθησης, εκτελούνται πειράματα με BERT και η δική μας υλοποίηση, μέσω ενός νευρωνικού δικτύου που συμπεριλαμβάνει BiGRU. Για την ανίχνευση σύμφωνα με τη διάδοση, εκτελούνται πειράματα με δύο δικές μας υλοποιήσεις, που χρησιμοποιούν GAT και GCN για μέτρο σύγκρισης. Τέλος για την ανίχνευση ειδήσεων σύμφωνα με τη γνώση, εκτελείται πείραμα που χρησιμοποιεί τα πειράματα με την καλύτερη απόδοση, από τους προαναφερθέντες τομείς και συνδυάζει BiGRU και R-GAT, πάνω σε δεδομένα γράφου γνώσης, για την ανίχνευση ψευδών ειδήσεων.

Η συνεισφορά αυτής της διπλωματικής εργασίας, ως προς τις υπάρχουσες υλοποιήσεις ανίχνευσης ψευδών ειδήσεων είναι η εξής:

- Δημιουργία ενός συνόλου δεδομένων αποκλειστικά στην ελληνική γλώσσα, το οποίο περιλαμβάνει επιβεβαιωμένες ψευδείς και αληθείς ειδήσεις, το οποίο σύμφωνα με την έρευνα που προηγήθηκε, δεν έχει ξαναγίνει σε κάποια άλλη εργασία.
- Μελέτη με βάση όλους τους προϋπάρχοντες πυρήνες εντοπισμού ψευδών ειδήσεων, με βάση το περιεχόμενο μιας είδησης, το περιεχόμενο μιας είδησης λαμβάνοντας υπόψη

τα συμφραζόμενα, τη διάδοση μιας είδησης και την γνώση που περιέχει μια είδηση και εξαγωγή συμπερασμάτων ως προς την αποδοτικότητά τους και την πιο αποτελεσματική μέθοδο.

- Σύγκριση με διάφορες baseline μεθόδους και εκτέλεση πειραμάτων σε άλλα σύνολα δεδομένων, για να εξακριβωθεί η αποδοτικότητα των μοντέλων που δημιουργήθηκαν.

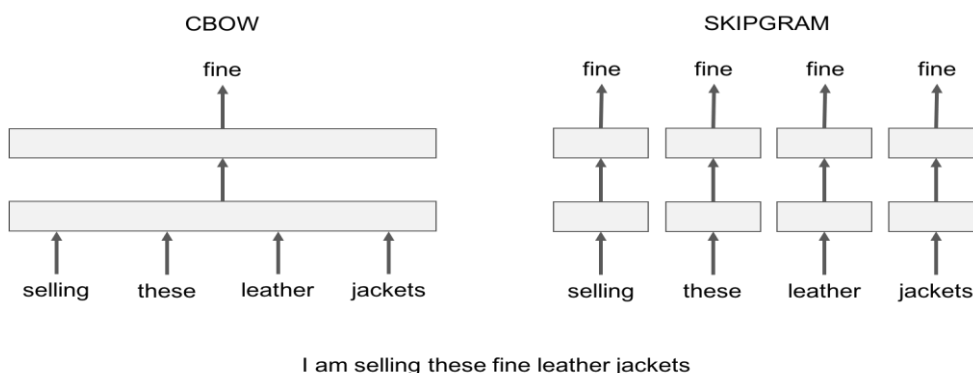
7.4.1. Πειράματα βασισμένα αποκλειστικά στο περιεχόμενο

Παρακάτω, παρατίθενται baseline μέθοδοι, οι οποίες είναι προϋπάρχουσες μέθοδοι κυρίαρχες στη βιβλιογραφία, όσο αφορά τη μηχανική μάθηση, για να εξακριβωθεί η απόδοσή τους στο σύνολο δεδομένων μας, καθώς και να διερευνηθεί η πιθανή χρήση τους σε κατοπινό στάδιο.

7.4.1.1. FastText

Το FastText είναι μια βιβλιοθήκη για εκμάθηση διανυσμάτων λέξεων και ταξινόμησης κειμένου, που δημιουργήθηκε από το εργαστήριο AI Research (FAIR) του Facebook, το οποίο διαθέτει προεκπαιδευμένα μοντέλα για 294 γλώσσες [52]. Το μοντέλο επιτρέπει σε κάποιον να δημιουργήσει έναν αλγόριθμο μάθησης με επίβλεψη ή χωρίς, για τη λήψη διανυσματικών αναπαραστάσεων για λέξεις. Αυτά τα διανύσματα, εμπεριέχουν κρυμμένες πληροφορίες για την γλώσσα, αναλογίες ή και σημασιολογικές πληροφορίες. Το FastText παρέχει δύο προεκπαιδευμένα μοντέλα για τον υπολογισμό των αναπαραστάσεων των λέξεων: Το Skipgram και το CBOW (Continuous Bag of Words). Το μοντέλο Skipgram μαθαίνει να προβλέπει μια λέξη-στόχο, με βάση μια κοντινή λέξη. Το μοντέλο CBOW προβλέπει τη λέξη-στόχο, σύμφωνα με το περιεχόμενό της. Η σημασιολογική πληροφορία της λέξης, αναπαρίσταται ως μια «τσάντα» με τις λέξεις – ο αριθμός των οποίων προκαθορίζεται - γύρω από τη λέξη-στόχο. Εάν παραδείγματος χάριν, δοθεί μια πρόταση και μια λέξη-στόχος, το Skipgram προσπαθεί να προβλέψει τη λέξη χρησιμοποιώντας μια τυχαία κοντινή της, ενώ το CBOW παίρνει κάποιες λέξεις που βρίσκονται γύρω από τη λέξη-στόχο και χρησιμοποιεί το άθροισμα των διανυσμάτων τους για να προβλέψει τον στόχο

Εικόνα 24: Σχηματική Αναπαράσταση των μοντέλων CBOW και SKIPGRAM



Το FastText αποτελείται από παραμέτρους, οι οποίες τροποποιούνται με βάση τα δεδομένα εισόδου. Αρχικά, υπάρχει το `dim`, το οποίο καθορίζει το μέγεθος των διανυσμάτων, έχοντας υπόψη ότι όσο πιο μεγάλο είναι τόσο πιο πολλές πληροφορίες θα μπορεί να αποθηκεύσει, έχοντας όμως σαν μειονέκτημα ότι θα απαιτείται περισσότερος χρόνος κατά την εκπαίδευση. Έπειτα έχουμε τα `subwords`, όπου μέσω των μεταβλητών `minn` (minimum size) και `maxn` (maximal size), καθορίζεται ο αριθμός των substrings που περιέχονται σε μια λέξη. Η παράμετρος `epoch`, καθορίζει πόσες φορές το μοντέλο θα διαβάσει όλα τα δεδομένα. Το `lr` (learning rate) είναι πόσο γρήγορα το μοντέλο έχει τη δυνατότητα να μάθει τα δεδομένα, έχοντας όμως σαν ρίσκο το γεγονός ότι υπερβολικά μεγάλη τιμή ενδέχεται να οδηγήσει στην υπερπροσαρμογή δεδομένων. Το `thread` καταδεικνύει τον αριθμό των πυρήνων που θα χρησιμοποιηθούν από τη CPU. Καθορίζοντας τις πιο πάνω κύριες μεταβλητές, παίρνουμε σαν έξοδο τα διανύσματα των λέξεων.

Ακολουθώντας τον επίσημο οδηγό της βιβλιοθήκης του FastText, προσαρμόζουμε κατάλληλα τα δεδομένα μας, για να ταξινομήσουμε σωστά τα κείμενα – άρθρα. Αρχικά, φορτώνουμε τη βιβλιοθήκη του FastText, η οποία για να μπορέσει να λειτουργήσει σωστά, θα πρέπει και το σύνολο δεδομένων μας να πάρει την κατάλληλη μορφή. Τα FastText σύνολα δεδομένων έχουν την μορφή: `__label__`(real/fake) κείμενο εισόδου. Έτσι, μέσω του κώδικα στην Εικόνα 25, προβαίνουμε σε αυτή ακριβώς την ενέργεια μορφοποίησης του συνόλου δεδομένων, όπου ως `shuffled_df`, ορίζεται το σύνολο δεδομένων μας, `shuffled_df.label` η ετικέτα της εκάστοτε είδησης και `shuffled_df.content` το κείμενο του κάθε άρθρου, στο οποίο συμπεριλαμβάνεται και ο τίτλος.

Εικόνα 25: Κώδικας μορφοποίησης του συνόλου δεδομένων

```
[ ] fasttext_dataset = []
for i in range(len(shuffled_df)):
    if shuffled_df.label.iloc[i] == 1:
        fasttext_dataset.append(f"__label__fake {shuffled_df.content[i]}")
    elif shuffled_df.label.iloc[i] == 0:
        fasttext_dataset.append(f"__label__real {shuffled_df.content[i]}")
```

Για την ταξινόμηση, χρησιμοποιούμε την εντολή `fasttext.train_supervised`, η οποία μας επιστρέφει ένα αντικείμενο – μοντέλο, πάνω στο οποίο μπορούμε να κάνουμε δοκιμές και προβλέψεις. Για να γίνει η διαδικασία της ταξινόμησης των κειμένων, χωρίζουμε το σύνολο δεδομένων σε σύνολο εκπαίδευσης, επιβεβαίωσης και πειραμάτων, με το καθένα να περιέχει 8605, 1000, 1000 ειδήσεις αντίστοιχα. Μέσω της εντολής `fasttext.train_supervised` γίνεται η εκπαίδευση, η οποία μας παρέχει και το μοντέλο. Έπειτα, μέσω της εντολής `model.predict`, μας δίνεται η πρόβλεψη της ετικέτας και η τιμή του διανύσματος του κειμένου.

Εικόνα 26: Απεικόνιση του μοντέλου εκπαίδευσης

```
import fasttext
model = fasttext.train_supervised(input="fasttext_dataset.train")

model.predict("__label__fake Η δημοσιογράφος",
              ('__label__real',), array([0.53657532]))
```

Παρατηρούμε ότι το μοντέλο δεν είναι τόσο ακριβές στις προβλέψεις του, αφού στην παραπάνω Εικόνα 26, αν και δίνεται ένα απλό παράδειγμα κειμένου, το ταξινομεί λανθασμένα. Για να αξιολογήσουμε καλύτερα το μοντέλο, θα δοκιμάσουμε την αξιοπιστία του στο σύνολο επιβεβαίωσης. Μέσω της εντολής `model.test`, μας δίνεται το σκορ του `precision` και του `recall`.

Εικόνα 27: Συνάρτηση Πειράματος

```
model.test("fasttext_dataset.valid")

(1000, 0.937, 0.937)
```

Για να βελτιώσουμε την αξιοπιστία του μοντέλου, εκτελέσαμε κάποια πειράματα προσθέτοντας διάφορες παραμέτρους κατά την εκπαίδευση του συνόλου δεδομένων, χωρίς να έχει υποστεί κάποια προεπεξεργασία. Οι παραμέτροι που ορίζονται πιο πάνω, χρησιμοποιήθηκαν η κάθε μια ξεχωριστά, καθώς και ως συνδυασμοί με το κάθε ένα: Αρχικά, έχουμε την προσθήκη εποχών στο στάδιο της εκπαίδευσης, όπου είναι ο αριθμός των επαναλήψεων – φορών που παίρνουμε τα δεδομένα. Έπειτα, τον ρυθμό με τον οποίο μαθαίνει κάποιο μοντέλο (`learning rate`) και είναι μια παράμετρος, η οποία καθορίζει το μέγεθος του βήματος σε κάθε επανάληψη και καθορίζει κατά πόσο αλλάζει το μοντέλο μας μετά από την επεξεργασία κάθε εισόδου. Δοκιμάζουμε επίσης και τη χρήση `Bigrams – N-Grams`, όπου χρησιμοποιούνται οι λέξεις ανά ζεύγη ή ακόμα περισσότερες συνεχόμενες λέξεις. Προστίθεται και η ιεραρχική `softmax`, όπου η κάθε είσοδος, μετατρέπεται σε τιμές που αντιπροσωπεύουν την πιθανότητα να έχουν μια ετικέτα και βοηθά επίσης στο να γίνει η εκπαίδευση γρηγορότερα. Χρησιμοποιούμε επίσης και τα λεγόμενα `καλάθια`, όπου χωρίζεται το σύνολο δεδομένων σε διαφορετικές ομάδες. Οι ίδιες εργασίες και τα ίδια πειράματα, έγιναν και με το σύνολο δεδομένων να έχει δεχθεί προεπεξεργασία. Τέλος, εξαγάγαμε το βέλτιστο μοντέλο, που εξήγαγε το μεγαλύτερο σκορ `recall` και `precision`, που καταδεικνύουν την ποιότητα της απόδοσης του `FastText`. Με βάση το μοντέλο αυτό, υπολογίσαμε την ακρίβεια και το `f1-score`, τα οποία υπολογίστηκαν στο 94% και 94% αντίστοιχα.

Εικόνα 28: Έκθεση ταξινόμησης μέσω FastText

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Real | 0.94 | 0.96 | 0.95 | 533 |
| Fake | 0.95 | 0.93 | 0.94 | 467 |
| accuracy | | | 0.94 | 1000 |
| macro avg | 0.94 | 0.94 | 0.94 | 1000 |
| weighted avg | 0.94 | 0.94 | 0.94 | 1000 |

7.4.1.2. Word2Vec

Το Word2Vec είναι ένα ρηχό νευρωνικό δίκτυο δύο επιπέδων, το οποίο είναι εκπαιδευμένο να αναδομεί γλωσσικά πλαίσια λέξεων [53]. Λαμβάνει ως είσοδο πολλές λέξεις και τις αναπαράγει σε ένα διανυσματικό χώρο, με κάθε μοναδική λέξη να έχει ένα αντίστοιχο διάνυσμα στο χώρο. Τα διανύσματα λέξεων είναι τοποθετημένα στο διανυσματικό χώρο έτσι ώστε οι παρόμοιες λέξεις να βρίσκονται σε κοντινή απόσταση μεταξύ τους στο χώρο. Το Word2Vec είναι ένα ιδιαίτερα αποδοτικό υπολογιστικά μοντέλο πρόβλεψης για την εκμάθηση διανυσμάτων λέξεων από ακατέργαστο κείμενο. Για την υλοποίησή του συνήθως χρησιμοποιούνται, το μοντέλο CBOW και το μοντέλο Skip-Gram. Το Word2Vec είναι ένα απλό νευρωνικό δίκτυο με ένα μόνο κρυφό επίπεδο, και όπως όλα τα νευρωνικά δίκτυα, έχει βάρη και κατά τη διάρκεια της προπόνησης, στόχος του είναι να προσαρμόσει αυτά τα βάρη για να μειώσει μια συνάρτηση απώλειας. Οι λέξεις οι οποίες μελετώνται, εισάγονται στο δίκτυο σαν one-hot vectors, όπου αποτελούνται από το μήκος των διαφορετικών λέξεων που υπάρχουν όπου όλες οι τιμές είναι μηδενικές, πλην της θέσης που αντιπροσωπεύει τη λέξη που μελετάται και θα ισούται με 1. Το κρυμμένο επίπεδο είναι πλήρως συνδεδεμένο και τα βάρη του, αποτελούν τα διανύσματα των λέξεων. Ως έξοδο, παίρνουμε την πιθανότητα η λέξη που μελετάται να είναι αυτή που ψάχνουμε.

Εικόνα 29: Εξίσωση Πιθανότητας Εντοπισμού λέξης

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$$

όπου, e^{z_j} το διάνυσμα της λέξης που μελετάται και e^{z_k} τα διανύσματα των υπόλοιπων λέξεων. Εάν διαφορετικές λέξεις, έχουν παρόμοιο νόημα, τότε το Word2Vec θα έχει εξόδους, παρόμοια διανύσματα για τις λέξεις αυτές, ώστε να αναπαρίστανται κοντά στο διανυσματικό χώρο. Το Word2Vec είναι σε θέση να καταγράψει πολλούς διαφορετικούς βαθμούς ομοιότητας μεταξύ των λέξεων, έτσι ώστε σημασιολογικά και συντακτικά μοτίβα να μπορούν να αναπαραχθούν χρησιμοποιώντας διανυσματική αριθμητική.

Όπως λέει και το όνομα της τεχνικής, η αποστολή του Word2Vec είναι να θέσει κάθε λέξη ίση με ένα διάνυσμα – vector, η οποία θα αναπαριστά λέξεις ή ακόμα και φράσεις στον διανυσματικό χώρο. Αφού γίνει προεπεξεργασία του συνόλου δεδομένων, επιλέγουμε ένα διαχωρισμό 80:20 για την εκπαίδευση και τα πειράματα αντίστοιχα. Φορτώνοντας την βιβλιοθήκη του gensim, καλούμε το μοντέλο Word2Vec για εκπαίδευση και καθορίζουμε τις εξής παραμέτρους: Size είναι το μέγεθος των διανυσμάτων, Window ουσιαστικά είναι το περιθώριο των λέξεων που θα συγκεντρωθούμε για το νόημα της λέξης, πριν και μετά από αυτή και Min_count ο ελάχιστος αριθμός εμφανίσεων των λέξεων, για να δημιουργηθεί ένα διάνυσμα.

Εικόνα 30: Μοντέλο Word2Vec

```
w2v_model = gensim.models.Word2Vec(X_train,
                                   vector_size=100,
                                   window=5,
                                   min_count=5)
```

Μέσω του μοντέλου αυτού, μπορούμε να βρούμε ποιες λέξεις έχουν αποκτήσει διανύσματα, ενώ ακόμα μπορούμε να βρούμε και παρόμοιες λέξεις όπως φαίνεται π.χ. με την λέξη φωτογραφία, που παρατηρούμε όμως ότι πέραν της πρώτης λέξης, οι υπόλοιπες είναι τελείως άσχετες με την αρχική λέξη. Το γεγονός ότι το word2vec, μπορεί να αντιληφθεί το περιεχόμενο των λέξεων, αλλά από ότι παρατηρούμε αυτό δεν γίνεται και τόσο αποδοτικά στην ελληνική γλώσσα, έχει αρνητική επιρροή και στην απόδοση του μοντέλου.

Εικόνα 31: Παρόμοιες λέξεις, με τη λέξη εισόδου

```
w2v_model.wv.most_similar('φωτογραφία')
[('βίντεο', 0.9996368885040283),
 ('στα', 0.9995592832565308),
 ('ένα', 0.9995366930961609),
 ('πως', 0.9995344877243042),
 ('γυναίκα', 0.9995225071907043),
 ('σε', 0.9995060563087463),
 ('ουκρανία', 0.9995025396347046),
```

Έπειτα με το άθροισμα των διανυσμάτων ανά λέξη, υπολογίζεται το ολικό διάνυσμα του κειμένου. Όμως, κάθε κείμενο έχει διαφορετικό αριθμό λέξεων, πράγμα που σημαίνει ότι ενδέχεται να προκαλέσει σφάλματα στη διαδικασία της εκπαίδευσης. Έτσι μετατρέπουμε τα διανύσματα, από ένα για κάθε λέξη σε ένα ολικό για το κείμενο προβαίνοντας στις εξής ενέργειες: η διάσταση των διανυσμάτων έχουμε καθορίσει να είναι ίση με 100 και αυτό που θα κάνουμε είναι να πάρουμε τον μέσο όρο των διανυσμάτων, ο οποίος προκύπτει μέσω του αθροίσματος των διανυσμάτων σε της κάθε λέξης, δια των αριθμό των λέξεων ανά κείμενο. Καταλήγουμε

να έχουμε ένα διάνυσμα μήκους 100, το οποίο αναπαριστά κάθε κείμενο. Για να γίνει η ταξινόμηση του κειμένου, χρησιμοποιούμε ένα τον Naïve Bayes ταξινομητή, ο οποίος σύμφωνα με την βιβλιογραφία, υπερτερεί του SVM σε μεγάλα σύνολα δεδομένων. Η απόδοση που παίρνουμε είναι 75.80% για ακρίβεια και 75% για f1-score.

Εικόνα 32: Έκθεση ταξινόμησης μέσω word2vec

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Real | 0.74 | 0.86 | 0.79 | 1140 |
| Fake | 0.79 | 0.64 | 0.71 | 965 |
| accuracy | | | 0.76 | 2105 |
| macro avg | 0.77 | 0.75 | 0.75 | 2105 |
| weighted avg | 0.76 | 0.76 | 0.75 | 2105 |

7.4.1.3. TF-IDF

Το TF-IDF (Term Frequency – Inverse Document Frequency) εξακριβώνει την σημασία ενός string, είτε αυτό είναι κείμενο, λέξη, κτλ., σε ένα κείμενο, το οποίο περιέχεται σε μια ομάδα κειμένων [54]. Το Term Frequency όπως λέει και ο όρος, έχει τη δυνατότητα να καταμετράει τις φορές που εμφανίζεται ο όρος που ψάχνουμε στο εκάστοτε κείμενο ή τη συχνότητα εμφάνισης αναλόγως με το μήκος του κειμένου - δηλαδή τον αριθμό εμφανίσεων δια τον αριθμό των λέξεων του εκάστοτε κειμένου - ή τη λογαριθμική συχνότητα (όπως π.χ. $\log(1 + \text{αριθμός εμφανίσεων})$) ή ακόμα και την Boolean συχνότητα, όπου είναι 0 αν δεν εμφανίζεται η λέξη ή 1 αν εμφανίζεται. Το Inverse Document Frequency, εξετάζει πόσο συνηθισμένη είναι μια λέξη ως προς το ολικό σύνολο κειμένων, όπου θέτει σαν t την λέξη που θέλουμε να γίνει η μέτρηση και N , ο αριθμός των κειμένων στο ολικό σύνολο κειμένων. Ο παρονομαστής είναι απλώς ο αριθμός των κειμένων στα οποία εμφανίζεται ο όρος t .

Εικόνα 33: Εξίσωση Υπολογισμού Inverse Document Frequency

$$idf(t, D) = \log\left(\frac{N}{count(d \in D: t \in d)}\right)$$

Ο λόγος χρήσης του IDF είναι για να περιορίσουμε το βάρος που θα έχουν «ασήμαντες» λέξεις και να δοθεί έμφαση στις πιο σημαντικές λέξεις για το σύνολο κειμένων. Η βασική ιδέα του TF-IDF είναι ότι η σημασία ενός όρου σχετίζεται αντιστρόφως με τη συχνότητά του σε έγγραφα. Το TF δίνει πληροφορίες πόσο συχνά εμφανίζεται ένας όρος στο έγγραφο αυτό και το IDF μας δίνει πληροφορίες σχετικά με τη σπανιότητα του όρου στη συλλογή των κειμένων. Ο πολλαπλασιασμός αυτών των τιμών μας δίνει την τελική αναπαράσταση.

Το TF-IDF μπορεί να χρησιμοποιηθεί σε διάφορα ζητήματα για να καθορίσει πόσο σημαντικός είναι ένας όρος ως προς το κείμενο, όπως σε ζητήματα φυσικής επεξεργασίας γλώσσας (NLP) μετατρέποντας τις τιμές για κάθε λέξη σε διανύσματα και αναζητώντας την ομοιότητά τους μέσω της ομοιότητας συνημίτονου, σε ζητήματα μηχανών αναζήτησης, θέλοντας αποτελέσματα που να συσχετίζονται όσο το δυνατόν περισσότερο με τον όρο που αναζητείται ή ακόμα και στην εξακρίβωση των πιο σημαντικών όρων σε ένα κείμενο. Η τεχνική του TF-IDF έχει και τα μειονεκτήματά της: Αρχικά δεν μπορεί να εντοπίσει την σημασιολογική σημασία των λέξεων. Εν συνεχεία, αγνοώντας την σειρά των λέξεων, χωρίζει εκφράσεις η οποίες πρέπει να είναι αλληλένδετες όπως π.χ. Βασίλισσα Ελισάβετ. Τέλος, μπορεί να υστερεί στην αποτελεσματικότητα ως προς την μνήμη, ειδικότερα όταν αφορά πολύ μεγάλα κείμενα.

Όπως έχει προαναφερθεί, το TF-IDF μπορεί να καθορίσει πόσο σημαντική είναι μια λέξη ως προς το ολικό κείμενο που μελετάται, αλλά δεν μπορεί να αντιληφθεί τα συμφραζόμενα σε ένα κείμενο, ούτε και την συντακτική του σημασία (π.χ. συνώνυμο ή συναισθηματικά φορτισμένη γλώσσα). Αφού γίνει η προεπεξεργασία του συνόλου δεδομένων, ξεκινούμε τη διαδικασία υλοποίησης του TF-IDF. Παίρνουμε το TF-IDF Vectorizer, το οποίο όπως αναφέραμε πιο πάνω, μετράει ουσιαστικά την αυθεντικότητα μιας λέξης, δηλαδή συγκρίνει τον αριθμός των φορών που παρουσιάζεται μια λέξη σε ένα κείμενο με τον αριθμό των κειμένων στα οποία εμφανίζεται. Έχοντας χωρίσει το σύνολο δεδομένων με τα ίδια ποσοστά που προαναφέρθηκαν, προσαρμόζουμε τον Vectorizer για να πάρουμε τα διανύσματα. Έπειτα με τη χρήση του πολυωνυμικού ταξινομητή Naïve Bayes, κάνουμε την ταξινόμηση του κειμένου, τα αποτελέσματα του οποίου ανέρχονται στο για ακρίβεια στο 86% και για f1-score στο 86%.

Εικόνα 34: Έκθεση ταξινόμησης μέσω TF-IDF

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Real | 0.80 | 1.00 | 0.89 | 551 |
| Fake | 1.00 | 0.69 | 0.82 | 449 |
| accuracy | | | 0.86 | 1000 |
| macro avg | 0.90 | 0.84 | 0.85 | 1000 |
| weighted avg | 0.89 | 0.86 | 0.86 | 1000 |

7.4.1.4. GloVe

Τα διανύσματα λέξεων GloVe με βάση προ-εκπαιδευμένα μοντέλα, προσπαθούν να βρουν παρόμοιες λέξεις ως προς την εκάστοτε λέξη που μελετάται. Το GloVe (Global Vectors) είναι ένας αλγόριθμος εκμάθησης του Stanford χωρίς επίβλεψη, με στόχο τη δημιουργία διανυσμάτων, συγκεντρώνοντας πίνακες συν-εμφάνισης λέξεων από ένα δεδομένο σώμα κειμένων. Η βασική ιδέα πίσω από τα διανύσματα λέξεων GloVe είναι η εξαγωγή της σχέσης μεταξύ των λέξεων από στατιστικές. Σε αντίθεση με τον πίνακα εμφάνισης, ο πίνακας συν-εμφάνισης, μας λέει πόσο συχνά ένα συγκεκριμένο ζεύγος λέξεων εμφανίζεται μαζί. Κάθε τιμή στον πίνακα

συν-εμφάνισης αντιπροσωπεύει ένα ζεύγος λέξεων που εμφανίζονται μαζί. Χρησιμοποιώντας δύο κείμενα, ο πίνακας συν-εμφάνισης σχηματίζεται ως εξής: Λαμβάνει υπόψη τις μοναδικές λέξεις από τα δύο κείμενα και ο πίνακας είναι τετραγωνικός με τις διαστάσεις του να είναι ο αριθμός των λέξεων. Οι φορές που μια λέξη εμφανίζεται και στα δύο κείμενα, αναγράφεται κάτω από την διαγώνιο, ενώ πάνω από την διαγώνιο δεν βάζουμε τιμές. Τα διανύσματα υπολογίζονται ως εξής:

Εικόνα 35: Εξίσωση Συχνότητας Εμφάνισης κάθε ζεύγους λέξεων στο σώμα κειμένου

$$F(w_i, w_j, w_k) = \frac{P_{ik}}{P_{jk}}$$

Αφού, επιτελέσουμε τον καθαρισμό του συνόλου δεδομένων, δημιουργούμε ένα λεξικό το οποίο περιέχει μια φορά την κάθε λέξη που έχει απομείνει στο προ-επεξεργασμένο σύνολο δεδομένων. Ο στόχος της μεθόδου είναι να δημιουργήσουμε διανύσματα για την κάθε λέξη, οι οποίες θα αναπαραστήσουν παρόμοιες λέξεις κοντά στο διανυσματικό χώρο, ενώ λέξεις που δεν συσχετίζονται μακριά. Επιλέγουμε την έκδοση διανυσμάτων Glove του Stanford, η οποία περιέχει 840 δισεκατομμύρια λέξεις και οι ενσωματώσεις που περιέχει είναι διανύσματα διάστασης 300. Αντιστοιχίζουμε την εκάστοτε λέξη με το κατάλληλο διάνυσμα, λαμβάνοντας πάντοτε υπόψη ότι αναπαρίστανται με διανύσματα μόνο οι λέξεις που περιέχονται στην έκδοση του Glove που επιλέξαμε. Λόγω του ότι έχουμε ένα μεγάλο σύνολο λέξεων, το οποίο όπως είναι αναμενόμενο μας δίνει τεράστιες τιμές διανυσμάτων, καθώς κάθε λέξη έχει διαφορετικό νόημα αναλόγως της θέσης της στο κείμενο και κατ' επέκταση διαφορετικό διάνυσμα κάθε φορά, επιλέγουμε τον μέσο όρο των διανυσμάτων που αναπαριστούν την εκάστοτε λέξη, καταλήγοντας με ένα διάνυσμα ανά λέξη. Έπειτα μέσω του ταξινομητή Naïve Bayes, πραγματοποιούμε την ταξινόμηση, τα αποτελέσματα της οποίας παρουσιάζονται στην Εικόνα 36. Για να μπορέσουμε να χρησιμοποιήσουμε τον ταξινομητή, βεβαιωνόμαστε ότι η εκάστοτε τιμή είναι μεταξύ 0 και 1. Η απόδοση του GloVe είναι για ακρίβεια στο 64% και για f1-score στο 64%.

Εικόνα 36: Έκθεση ταξινόμησης μέσω GloVe

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Real | 0.65 | 0.70 | 0.68 | 1132 |
| Fake | 0.62 | 0.57 | 0.59 | 973 |
| accuracy | | | 0.64 | 2105 |
| macro avg | 0.64 | 0.63 | 0.63 | 2105 |
| weighted avg | 0.64 | 0.64 | 0.64 | 2105 |

7.4.1.5. Διανύσματα μέσω SpaCy

Το SpaCy αποτελεί μια βιβλιοθήκη, η οποία περιέχει ειδικό πακέτο για την ελληνική γλώσσα, το οποίο έχει αποδειχτεί να είναι πιο αποδοτικό, σε σχέση με άλλων βιβλιοθηκών όπως το Nltk. Το SpaCy ειδικεύεται στην φυσική επεξεργασία γλώσσας και οι δυνατότητες του ποικίλουν από την εξακρίβωση του μέρους λόγου μιας λέξης, μέχρι τον χωρισμό κειμένου σε προτάσεις και λέξεις, μέχρι την δυνατότητα επεξεργασία κειμένου και εξαγωγής διανυσμάτων ως προς αυτό. Με αυτό το κομμάτι θα ασχοληθούμε και εμείς, προσπαθώντας μέσω των διανυσμάτων να εξακριβώσουμε την αξιοπιστία του κειμένου.

Αφού προβούμε στην προεπεξεργασία κειμένου, ακολουθούμε παρόμοια διαδικασία με την τεχνική του TF-IDF. Έπειτα, μετατρέπουμε το κείμενο σε διανύσματα, μέσω των οποίων θέλουμε με τη χρήση ενός ταξινομητή, να εξάγουμε κατά πόσο μια είδηση είναι αληθής ή ψευδής. Αρχικά, χωρίζουμε το σύνολο δεδομένων σε σύνολο εκπαίδευσης και πειραμάτων και μετατρέπουμε τα παραγόμενα διανύσματα σε δυσδιάστατους πίνακες. Για να μπορέσουμε να χρησιμοποιήσουμε τον ταξινομητή του Naïve Bayes, μετατρέπουμε τα διανύσματα σε τιμές μεταξύ 0 και 1. Τα αποτελέσματα που εξήχθησαν, ανέρχονται στο 74% για ακρίβεια και στο 73% για f1-score.

Εικόνα 37: Έκθεση ταξινόμησης μέσω διανυσμάτων Spacy

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Fake | 0.72 | 0.83 | 0.77 | 1132 |
| Real | 0.76 | 0.63 | 0.69 | 973 |
| accuracy | | | 0.74 | 2105 |
| macro avg | 0.74 | 0.73 | 0.73 | 2105 |
| weighted avg | 0.74 | 0.74 | 0.73 | 2105 |

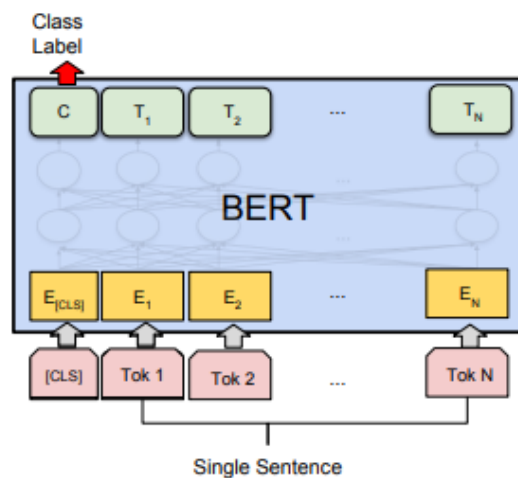
7.4.2. Πειράματα βασισμένα στο περιεχόμενο, λαμβάνοντας υπόψη τα συμφραζόμενα

7.4.2.1. BERT

Ο BERT (Bidirectional Encoder Representations from Transformers) είναι ένα μοντέλο αναπαράστασης φυσικής γλώσσας, το οποίο χρησιμοποιείται για την προ-εκπαίδευση (pre-training) μοντέλων NLP. Ο BERT αναπαριστά τις λέξεις, λαμβάνοντας υπόψη τα συμφραζόμενα (context-sensitive) και διαθέτει σταθερή αρχιτεκτονική δομή. Η σημασία του BERT γίνεται άμεσα αισθητή, όταν αναλογιστεί κανείς ότι μια λέξη μπορεί να είναι αμφίσημη και έτσι κρίνεται απαραίτητο, κατά την αναπαράστασή της σε διάνυσμα, να λαμβάνονται υπόψη τα συμφραζόμενά της. Αυτό που ξεχωρίζει τον BERT σε σχέση με άλλα μοντέλα, τα οποία επίσης λαμβάνουν υπόψη τα συμφραζόμενα είναι το γεγονός ότι λαμβάνει υπόψη τόσο προηγούμενες, όσο και επόμενες λέξεις της πρότασης, εξού και το Bidirectional εξ ορισμού του. Η αμφίδρομη

χρήση των συμφραζόμενων, μας δίνει καλύτερες αναπαραστάσεις των λέξεων ως προς τις έννοιές τους. Σχετικά με τη δομή ενός μοντέλου αναπαράστασης φυσικής γλώσσας υπάρχουν δύο προσεγγίσεις: Η πρώτη είναι η feature-based προσέγγιση, όπου η δομή του μοντέλου διαφοροποιείται πλήρως ανάλογα με το πρόβλημα και η δεύτερη είναι η fine-tuning προσέγγιση, όπου η δομή του μοντέλου παραμένει σταθερή, ανεξαρτήτως του προβλήματος. Αυτό που αλλάζει είναι το τελευταίο επίπεδο - layer του μοντέλου και γίνονται βελτιστοποιήσεις, μεταβάλλοντας τις τιμές κάποιων παραμέτρων. Ο BERT χρησιμοποιεί τη fine-tuning προσέγγιση και η εκπαίδευσή του χωρίζεται στην αρχική φάση της προ-εκπαίδευσης (pre-training) και την τελική φάση της προσαρμογής (fine-tuning). Κατά τη φάση της προ-εκπαίδευσης γίνονται δύο εργασίες: Η πρώτη είναι η Μοντελοποίηση Κρυμμένων Λέξεων (Masked Language Modeling) όπου επιλέγονται τυχαία κάποιες λέξεις και αποκρύπτονται. Στη συνέχεια, ο BERT προσπαθεί να τις «μαντέψει» - προβλέψει βάσει των υπόλοιπων λέξεων των προτάσεων στις οποίες ανήκουν. Η δεύτερη είναι η Πρόβλεψη Επόμενης Πρότασης (Next Sentence Prediction). Σε αυτή ο BERT λαμβάνει ένα ζεύγος προτάσεων και προσπαθεί να μαντέψει αν η δεύτερη πρόταση που έλαβε είναι η αμέσως επόμενη της πρώτης στο αρχικό κείμενο. Έτσι το μοντέλο BERT εκπαιδεύεται λαμβάνοντας υπόψη τα συμφραζόμενα των λέξεων.

Εικόνα 38: Προ-εκπαίδευση και fine-tuning του BERT



Στη δική μας υλοποίηση, η οποία είναι εμπνευσμένη από το TensorFlow, ισχύουν τα παραπάνω, έχοντας επίσης: το pooled output, το οποίο αναπαριστά τις ακολουθίες εισόδου (batch size, Hidden State) και είναι ουσιαστικά τα διανύσματα για κάθε είδηση, το sequence output το οποίο αναπαριστά κάθε token αναλόγως με το context του σε κάθε τίτλο (batch size, sequence length, H) και τα encoder outputs, τα οποία είναι tensors (batch size, sequence length, length) και αποτελούνται από τα outputs των L Transformer blocks. Το τελευταίο στοιχείο κάθε λίστας αποτελεί την ακολουθία εξόδου. Για το Fine-tuning θα χρησιμοποιήσουμε το

pooled output. Η διαδικασία θα αποτελείται από το μοντέλο προ-εκπαίδευσης Bert που επιλέξαμε, μια dense layer και μια dropout layer. Κάθε layer δέχεται σαν είσοδο την έξοδο του προηγούμενου layer, με εξαίρεση το πρώτο, το οποίο παίρνει σαν είσοδο τους τίτλους. Το κάθε μοντέλο Bert διαφέρει από το άλλο στα L,H,A(L=Layer, H=Hidden state, A=Attention). Εάν τα σκορ που παράγει το μοντέλο είναι <0,5, τότε γίνεται 0 και σε αντίθετη περίπτωση 1. Έτσι διαχωρίζονται τα ψευδή και αληθινά γεγονότα.

Μέσω του TensorFlow [55], χρησιμοποιούμε αρκετά μοντέλα για να βρούμε αυτό με την καλύτερη απόδοση. Τα μοντέλα που περιέχει είναι το BERT-Base, BERT-Uncased και επτά άλλα μοντέλα με προ-εκπαιδευμένα βάρη, Small-BERTs οι οποίοι έχουν την ίδια αρχιτεκτονική με τα προαναφερθέντα, αλλά με διαφορετικό αριθμό Transformer blocks, τα οποία μας δίνουν την ικανότητα να μελετήσουμε τα πλεονεκτήματα/μειονεκτήματα ταχύτητας, μεγέθους και ποιότητας, ALBERT το οποίο είναι ένας πιο απλός BERT που μειώνει το μέγεθος του μοντέλου μέσω διάδοσης παραμέτρων μεταξύ στρωμάτων, BERT Experts όπου έχουμε οκτώ μοντέλα με BERT-base αρχιτεκτονική τα οποία προσφέρουν διαφορετικά pre-training domains τα οποία πλησιάζουν πιο πολύ στην ταξινόμηση που θέλουμε να επιτελέσουμε, Electra το οποίο έχει επίσης την ίδια αρχιτεκτονική με BERT και εκπαιδεύεται εκ των προτέρων ως παράγοντας διάκρισης σε μια εγκατάσταση που μοιάζει με GAN και BERT με προσοχή talking-heads και gated GELU που βελτιώνει τον κορμό της αρχιτεκτονικής του Transformer. Έπειτα κάνουμε fine-tuned τον εκάστοτε BERT στα δεδομένα μας, χρησιμοποιώντας και τον Greek-BERT, ο οποίος γίνεται fine-tuned μέσω hugging-face. Αφού επιτελέσουμε προ-επεξεργασία των δεδομένων, ταξινομούμε μέσω του εκάστοτε μοντέλου, για να δούμε ποιο μοντέλο είναι το βέλτιστο. Αφού τα δοκιμάσουμε όλα, συμπεραίνουμε ότι το βέλτιστο μοντέλο είναι το BERT-multi-cased L-12 H-768, με 80% σκορ ακρίβειας και 80% f1-score.

Εικόνα 39: Έκθεση ταξινόμησης μέσω BERT multi-cased L-12 H-768

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.77 | 0.89 | 0.83 | 1421 |
| 1 | 0.84 | 0.69 | 0.76 | 1211 |
| accuracy | | | 0.80 | 2632 |
| macro avg | 0.81 | 0.79 | 0.79 | 2632 |
| weighted avg | 0.80 | 0.80 | 0.80 | 2632 |

7.4.2.2. GRU-BiGRU

Για να μπορέσουμε να ορίσουμε την GRU [56], αρχικά πρέπει να προσδιορίσουμε τί είναι τα RNNs και τα LSTM. Τα RNNs (Recurrent Neural Networks), που όπως προαναφέρθηκε είναι γνωστά και ως αναδρομικά νευρωνικά δίκτυα και δέχονται ως εισόδους ακολουθιακά δεδομένα. Η διαφορά τους σε σχέση με απλά νευρωνικά δίκτυα είναι το γεγονός ότι εκτός από την

τρέχουσα είσοδο, εξαρτώνται και από προηγούμενες εισόδους και μέσω της μνήμης που διαθέτει, διατηρεί πληροφορίες οι οποίες θα χρησιμοποιηθούν για να συνδυαστούν με την τρέχουσα είσοδο. Παρακάτω θα γίνει η περιγραφή λειτουργίας των κλασικών RNNs, γνωστά και ως vanilla RNNs: Τα RNNs αποτελούνται από τρεις υπολογισμούς: την προηγούμενη κρυφή κατάσταση (hidden state) h_{t-1} , την τρέχουσα κρυφή κατάσταση h_t και την τρέχουσα είσοδο x_t . Για την τρέχουσα κρυφή κατάσταση, γίνεται ο παρακάτω υπολογισμός:

$$h_t = f(W_{hh}h_{t-1} + W_{hx}x_t + b_h)$$

Αρχικά, υπάρχει το x_t , το οποίο λειτουργεί ως η τρέχουσα είσοδος του RNN και είναι το διάλυμα της εκάστοτε ακολουθιακής εισόδου, η οποία στην περίπτωση μας είναι λέξεις, το σύνολο των οποίων αποτελεί το εκάστοτε κείμενο στο οποίο γίνεται η κάθε εργασία. Το h_t είναι η τρέχουσα κρυφή κατάσταση, η οποία λειτουργεί ως η μνήμη του δικτύου και υπολογίζεται με βάση την προηγούμενη κρυφή κατάσταση h_{t-1} και την τρέχουσα είσοδο. W_{hh} και W_{hx} ορίζονται ως οι εκπαιδευσιμες παράμετροι, ενώ b_h είναι το bias. Οι συναρτήσεις f είναι οι συναρτήσεις ενεργοποίησης, όπου συνήθως γίνεται η χρήση του tanh.

Για την έξοδο του RNN, γίνεται ο παρακάτω υπολογισμός:

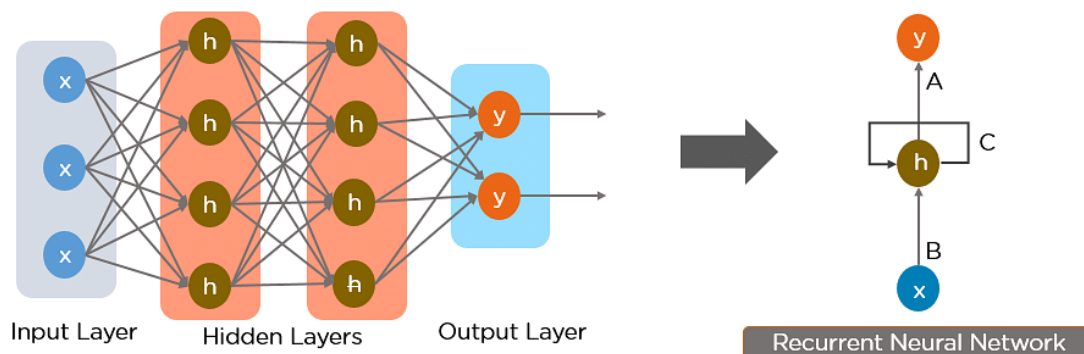
$$y_t = g(W_{yh}h_t + b_y)$$

Αρχικά, το W_{yh} είναι η εκπαιδευσιμη παράμετρος του συστήματος, όπως ήταν τα W_{hx} και W_{hh} για την κρυφή κατάσταση. Το h_t αποτελεί την τρέχουσα κρυφή κατάσταση που αναλύθηκε πιο πάνω, ενώ το b_y είναι το bias. Οι συναρτήσεις g είναι οι συναρτήσεις ενεργοποίησης για τον υπολογισμό της εξόδου, όπου χρησιμοποιούνται η σιγμοειδής, η softmax και η ReLU.

Όπως παρατηρείται, τα RNNs χρησιμοποιούν τις ίδιες εκπαιδευσιμες παραμέτρους σε όλα τα βήματα, ενώ αξίζει να αναφερθεί ότι στη μνήμη, παρέχονται πληροφορίες από όλα τα προηγούμενα βήματα. Κατά την εκπαίδευση των RNNs, χρησιμοποιείται η μέθοδος της διάδοσης προς τα πίσω, διαδικασία γνωστή και ως back-propagation. Σε μεγάλες ακολουθίες κειμένου, όπου προκύπτουν πολλά επίπεδα, η λειτουργία του RNN τίθεται σε κίνδυνο, καθώς πέραν του γεγονότος ότι είναι υπολογιστικά ακριβό, εμφανίζεται ένα πρόβλημα γνωστό και σαν exploding ή vanishing gradient. Τα gradients είναι κλίσεις, οι οποίες χρησιμοποιούνται για προσαρμογή των βαρών κατά την εκπαίδευση. Κατά τη διάρκεια του back-propagation τα gradients υπολογίζονται αθροιστικά για κάθε προηγούμενο επίπεδο ως $\frac{\partial L^T}{\partial w} = \sum_{t=1}^T \frac{\partial L^T}{\partial w} \Big|_t$ και για κάθε επίπεδο οι τιμές αλλάζουν εκθετικά, αφού χρησιμοποιούν τα βάρη των προηγούμενων επιπέδων. Έτσι σε περίπτωση που προκύπτουν μεγάλες τιμές, θα μεγαλώνουν εκθετικά, ενώ στην περίπτωση μικρών, θα μικραίνουν εκθετικά, με αποτέλεσμα η κλίση να προκαλεί αναλόγως τεράστιες ή αμελητέες αλλαγές στα βάρη, κάτι το οποίο σημαίνει ότι η εκπαίδευση του δικτύου δεν γίνεται αποδοτικά. Μια λύση του προβλήματος της κλίσης, θα ήταν να οριστεί μια ανώτερη και αντίστοιχα κατώτερη τιμή, περιορίζοντας ουσιαστικά την κλίση στις επιθυμητές τιμές. Όταν έχουμε ένα μια μικρή ακολουθία λέξεων, έχουμε κατ' επέκταση ένα μικρό αριθμό επιπέδων RNN, άρα θα μπορούσαμε να χρησιμοποιήσουμε τα RNNs, καθώς δεν υπάρχει το

πρόβλημα εξαφάνισης ή έκρηξης των κλίσεων. Η πλήρης αναπαράσταση των RNNs απεικονίζεται πιο κάτω στην Εικόνα 40.

Εικόνα 40: Αρχιτεκτονική RNN



Για το πρόβλημα που αντιμετωπίζουν τα RNNs στις μεγάλες ακολουθίες δεδομένων με την έκρηξη ή την εξαφάνιση κλίσης, έχει προταθεί μια πιο σύνθετη δομή, γνωστή και ως δίκτυα μακροπρόθεσμης και βραχυπρόθεσμης μνήμης (LSTM – Long Short Term Memory). Τα LSTM διαφέρουν ως προς το γεγονός ότι οι μηχανισμοί πύλης από τους οποίους αποτελείται, ελέγχουν την ροή πληροφοριών, δηλαδή κατά πόσο μια πληροφορία είναι χρήσιμη ή άχρηστη. Στα παραδοσιακά RNNs, η μοναδική μνήμη που διαθέτουν είναι η κρυφή κατάσταση, η οποία περιέχει τις πληροφορίες, από όλα τα προηγούμενα βήματα. Από την άλλη τα LSTM περιλαμβάνουν δύο καταστάσεις που αναπαριστούν διαφορετικές καταστάσεις της μνήμης: Αρχικά, υπάρχει η κατάσταση κελιού (cell state) c_t , η οποία αποτελεί την μακρά μνήμη και περιέχει πληροφορίες τόσο από τρέχουσες εισόδους, αλλά και εισόδους από προηγούμενα επίπεδα, κάτι το οποίο δικαιολογεί και το όνομα της μακράς μνήμης, καθώς μπορεί να περιλαμβάνει πληροφορίες από προηγούμενα βήματα, στα οποία η κρυφή κατάσταση δεν θα είχε πρόσβαση. Η δεύτερη κατάσταση είναι η γνωστή και από τα RNNs κρυφή κατάσταση, η οποία διατηρεί πληροφορίες από τα αμέσως προηγούμενα επίπεδα – λέξεις. Η συνήθης αρχιτεκτονική των μονάδων πύλης των LSTM περιγράφεται ως εξής: Υπάρχουν τρεις μονάδες πύλης, η πύλη εισόδου (input gate) i_t , η πύλη λήθης (forget gate) f_t και η πύλη εξόδου (output gate) o_t . Η πύλη εισόδου ελέγχει τη ροή των νέων πληροφοριών, η πύλη λήθης αποφασίζει ποιες πληροφορίες είναι χρήσιμες για να παραμείνουν στο LSTM και η πύλη εξόδου καθορίζει την τιμή της επόμενης κρυφής κατάστασης, καθορίζοντας πόσο σημαντική είναι η τρέχουσα πληροφορία που βρίσκεται στο LSTM. Οι υπολογισμοί που πραγματοποιεί το LSTM για τις καταστάσεις που προαναφέρθηκαν, ορίζονται ως εξής:

$$f_t = \sigma_g (W_f x_t + U_f h_{t-1} + b_f)$$

$$i_t = \sigma_g (W_i x_t + U_i h_{t-1} + b_i)$$

$$o_t = \sigma_g (W_o x_t + U_o h_{t-1} + b_o)$$

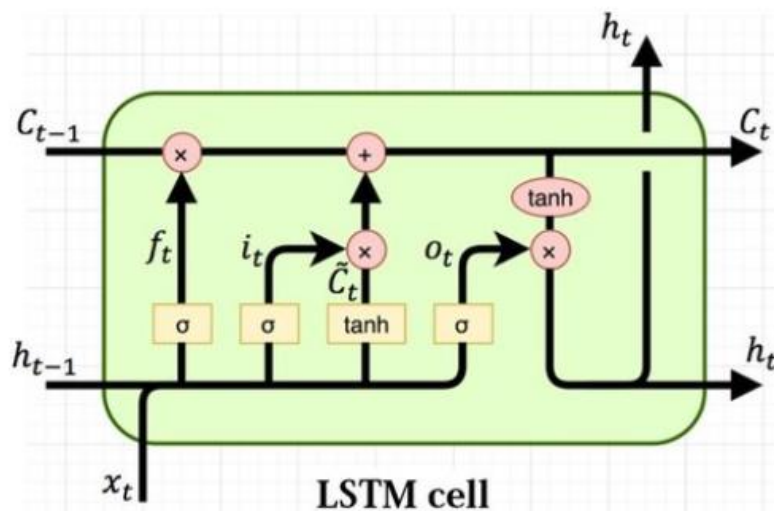
$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c (W_c x_t + U_c h_{t-1} + b_c)$$

$$h_t = o_t \circ \sigma_h(c_t)$$

Αρχικά, υπάρχει το x_t που λειτουργεί ως η τρέχουσα εισόδος του LSTM και είναι το διάνυσμα της εκάστοτε ακολουθιακής εισόδου. Το h_{t-1} είναι η κρυφή κατάσταση του προηγούμενου βήματος, δηλαδή της προηγούμενης εισόδου, η οποία λειτουργεί ως η βραχυπρόθεσμη μνήμη του δικτύου, ενώ το c_{t-1} το κελί μνήμης του προηγούμενου βήματος, δηλαδή των πληροφοριών από τις προηγούμενες εισόδους, το οποίο λειτουργεί ως η μακροπρόθεσμη μνήμη του δικτύου. Τα b_f , b_i , b_o , b_c ορίζονται ως τα biases και τα W , U ορίζονται ως οι εκπαιδευσιμες παράμετροι. Τέλος, οι συναρτήσεις σ_g είναι οι συναρτήσεις ενεργοποίησης, οι οποίες ορίζονται ως σιγμοειδείς συναρτήσεις και οι σ_c , σ_h είναι συναρτήσεις υπερβολικής εφαπτομένης - \tanh .

Στην Εικόνα 41, βλέπουμε το LSTM, του οποίου η λειτουργία εξηγείται αναλυτικά. Αρχικά, δέχεται ως εισόδους την μνήμη κελιού του προηγούμενου βήματος c_{t-1} , η οποία όπως προαναφέρθηκε αποτελεί την μακροπρόθεσμη μνήμη του δικτύου, την κρυφή μνήμη του προηγούμενου βήματος h_{t-1} , η οποία αποτελεί την βραχυπρόθεσμη μνήμη του δικτύου και την τρέχουσα είσοδο (λέξη στην περίπτωση κειμένου) x_t . Στο κάτω μέρος του δικτύου, ακολουθεί η πρόσθεση της τρέχουσας εισόδου x_t και της προηγούμενης κρυφής μνήμης h_{t-1} και έπειτα ακολουθούν οι υλοποιήσεις των πυλών. Ξεκινώντας από τα αριστερά προς τα δεξιά, έχουμε την πύλη λήθης f_t , η οποία τροφοδοτεί την προηγούμενη πρόσθεση, σε μια σιγμοειδή συνάρτηση και την πολλαπλασιάζει με την μνήμη κελιού του προηγούμενου βήματος και αναλόγως του αποτελέσματος αποφασίζεται ποιες πληροφορίες είναι χρήσιμες. Όσο πιο κοντά στο μηδέν βρίσκεται το αποτέλεσμα, η πληροφορία θεωρείται άχρηστη, ενώ όσο πιο κοντά στο ένα βρίσκεται το αποτέλεσμα, η πληροφορία θεωρείται χρήσιμη. Έπειτα ακολουθεί η πύλη εισόδου i_t , η οποία ουσιαστικά εκτελεί την ίδια πράξη για την πρόσθεση που προαναφέρθηκε και την πολλαπλασιάζει με την τροφοδότηση της ίδιας πρόσθεσης σε μια συνάρτηση \tanh , η οποία αποτελεί την υποψήφια μνήμη κελιού \hat{c}_t , η οποία έχει σκοπό να καθορίσει πόση από την τρέχουσα πληροφορία θα προστεθεί στην μακροπρόθεσμη μνήμη. Αυτό γίνεται αφού η πύλη εισόδου, ουσιαστικά μέσω της σιγμοειδούς συνάρτησης, καθορίζει ποιες πληροφορίες της υποψήφιας μνήμης είναι χρήσιμες, με πανομοιότυπο τρόπο με την πύλη λήθης. Η πρόσθεση των παραπάνω δίνει το αποτέλεσμα της τρέχουσας μνήμης κελιού, η οποία και περνά στην έξοδο. Η τρέχουσα μνήμη κελιού, τροφοδοτείται σε μια συνάρτηση \tanh και πολλαπλασιάζεται με την πύλη εξόδου o_t , η οποία υπολογίζεται με τον ίδιο τρόπο με τις άλλες δύο πύλες και χρησιμοποιείται για να καθορίσει ποιες από τις πληροφορίες που βρίσκονται στο LSTM είναι χρήσιμες για να αποθηκευτούν στην κρυφή μνήμη, δηλαδή την βραχυπρόθεσμη μνήμη του δικτύου. Στα LSTMs, η πύλη λήθης δίνει τη δυνατότητα πιο αποδοτικής κανονικοποίησης των κλίσεων – gradients σε κάθε βήμα, ανανεώνοντας τις παραμέτρους της, καθώς καθορίζει κατά πόσο πληροφορίες πρέπει να διατηρούνται σε κάθε βήμα. Έτσι αντιμετωπίζεται το πρόβλημα κλίσεων που υπήρχε στα RNNs.

Εικόνα 41: Αρχιτεκτονική LSTM



Αφού έγινε η επεξήγηση των εννοιών των RNNs και των LSTMs, μπορούμε να εισαγάγουμε την έννοια της αναδρομικής μονάδας πύλης (GRU - Gated Recurrent Unit), η οποία έχει εφευρεθεί το 2014 και είναι ένας τρόπος κατά τον οποίο περιορίζεται ο αριθμός των εισόδων και των παραμέτρων σε σχέση με το LSTM. Όπως φαίνεται στην Εικόνα 42, οι GRUs δέχονται σαν εισόδους την τρέχουσα είσοδο και την προηγούμενη κρυφή κατάσταση. Οι μηχανισμοί πύλης, οι οποίοι ελέγχουν τη ροή πληροφοριών είναι μόλις δύο και αποτελούνται από την πύλη επαναφοράς (reset gate) r_t και την πύλη ενημέρωσης (update gate) z_t . Η πύλη ενημέρωσης καθορίζει πόσες από τις πληροφορίες των προηγούμενων βημάτων είναι χρήσιμες, για να αξιοποιηθούν στο τρέχον βήμα, ενώ η πύλη επαναφοράς χρησιμοποιείται για να καθορίσει πόσες από τις πληροφορίες των προηγούμενων βημάτων δεν είναι χρήσιμες. Η GRU παράγει μόνο μια έξοδο, η οποία αποτελεί και την νέα κρυφή κατάσταση h_t , για τον υπολογισμό της οποίας χρησιμοποιείται μια κατάσταση αντίστοιχη της υποψήφιας μνήμης κελιού των LSTM \hat{c}_t που ορίζεται ως η υποψήφια κρυφή μνήμη \hat{h}_t . Στην υποψήφια κρυφή μνήμη, καθορίζεται η επίδραση που θα έχει η προηγούμενη κρυφή μνήμη h_{t-1} , στον υπολογισμό της τρέχουσας h_t . Οι υπολογισμοί που πραγματοποιεί η GRU για τις καταστάσεις που προαναφέρθηκαν, ορίζονται ως εξής:

$$z_t = \sigma_g (W_z x_t + U_z h_{t-1} + b_z)$$

$$r_t = \sigma_g (W_r x_t + U_r h_{t-1} + b_r)$$

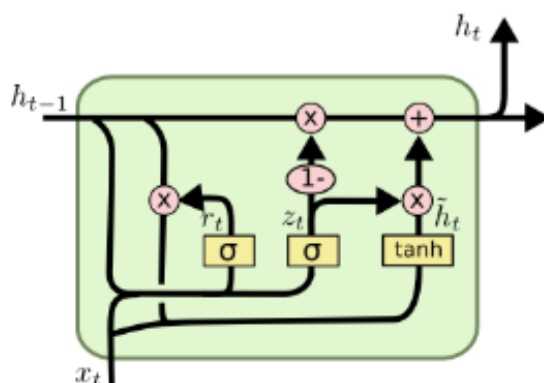
$$\hat{h}_t = \sigma_h (W_h x_t + U_h (r_t \circ h_{t-1}) + b)$$

$$h_t = z_t \circ h_{t-1} + (1 - z_t) \hat{h}_t$$

όπου σ_g σιγμοειδής συνάρτηση, σ_h συνάρτηση tanh, W και U οι εκπαιδευσιμες παράμετροι και b τα biases. Η λειτουργία που επιτελούν οι συναρτήσεις είναι παρόμοια με τις αντίστοιχες των LSTM. Λόγω του ότι η GRU έχει λιγότερες παραμέτρους είναι πιο γρήγορη στους υπολογισμούς, αλλά στην περίπτωση που έχουμε πολύ μεγάλα κείμενα και η GRU δεν είναι αρκετή αποδοτική, η εναλλακτική επιλογή είναι να πάμε πίσω στα LSTM. Όπως διαφαίνεται από την

μελέτη της βιβλιογραφίας, τα σημασιολογικά διανύσματα που δημιουργούνται είναι πιο αποδοτικά εάν επεξεργάζονται πληροφορίες και από τις δύο κατευθύνσεις του κειμένου, καθώς αναπαριστούν τις σχέσεις μεταξύ των λέξεων ή των φράσεων, καθ' όλο το μήκος της ακολουθίας. Δηλαδή αντιλαμβάνονται πληροφορίες πριν και μετά από κάθε λέξη – φράση. Εξού και η εκτεταμένη χρήση LSTM και GRU δύο κατευθύνσεων, γνωστά και ως BiLSTM και BiGRU αντίστοιχα.

Εικόνα 42: Αρχιτεκτονική GRU



Όσο αφορά την εκτέλεση του πειράματος, προτιμήθηκε η χρήση της BiGRU για τους λόγους που προαναφέρθηκαν και σε συνδυασμό με τη διαδικασία της συνέλιξης μέσω CNN, γίνεται η ταξινόμηση των ειδήσεων. Γίνεται χρήση του CNN, αφού παρατηρείται ότι χρησιμοποιείται στις πιο αποδοτικές υλοποιήσεις της βιβλιογραφίας. Αφού γίνει προ-επεξεργασία δεδομένων, αναπαριστούμε κάθε λέξη με μια μοναδική ταυτότητά της, μέσω χρήσης του keras. Έπειτα μετατρέπουμε το μήκος της κάθε ειδήσης, ώστε όλες οι ειδήσεις να έχουν το ίδιο μήκος χαρακτήρων, εφαρμόζοντας την μέθοδο padding, δηλαδή βάζοντας επιπλέον χαρακτήρες που συμβολίζονται με το 0, που βοηθούν στην επαύξηση του μεγέθους της πρότασης. Αφού χωρίσουμε το πλέον μορφοποιημένο σύνολο δεδομένων σε σύνολο εκπαίδευσης και δοκιμών, τροφοδοτείται σε ένα επίπεδο διανυσμάτων, το οποίο ακολουθείται από ένα επίπεδο dropout, το οποίο ακολουθείται από ένα επίπεδο BiGRU, τα σημασιολογικά διανύσματα του οποίου τροφοδοτούνται σε ένα επίπεδο συνέλιξης. Μέσω global-Max pooling εξάγονται οι πιο μεγάλες τιμές και αφού γίνει και πάλι χρήση επιπέδου dropout για αποφυγή υπερπροσαρμογής – overfitting, τροφοδοτούνται σε ένα dense επίπεδο, στο οποίο γίνεται η ταξινόμηση των ειδήσεων, μέσω της συνάρτησης ενεργοποίησης softmax. Στην παρακάτω εικόνα, παρουσιάζονται τα βέλτιστα αποτελέσματα στην Εικόνα 43, με ακρίβεια και f1-score να ανέρχονται στο 95%.

Εικόνα 43: Αποτελέσματα ταξινόμηση μέσω BiGRU

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| True | 0.96 | 0.96 | 0.96 | 1113 |
| Fake | 0.95 | 0.95 | 0.95 | 992 |
| accuracy | | | 0.95 | 2105 |
| macro avg | 0.95 | 0.95 | 0.95 | 2105 |
| weighted avg | 0.95 | 0.95 | 0.95 | 2105 |

Αξίζει να σημειωθεί ότι έγιναν πειράματα, στα οποία προστέθηκαν επίπεδα max pooling για να μειωθούν διαστάσεις των δεδομένων, γενικότερα οι υπολογιστικές πράξεις και παράγει σαν έξοδο τις μέγιστες τιμές των διανυσμάτων και flatten για μετατραπούν τα διανύσματα σε μονοδιάστατα διανύσματα, τα οποία θα τροφοδοτηθούν στο τελευταίο επίπεδο. Εκ του αποτελέσματος, η διαφορά ήταν μηδαμινή, καθώς και πάλι τα αποτελέσματα, ανέρχονταν στο 95%.

7.4.3. Πειράματα βάσει Δικτύου Διάδοσης Ειδήσεων

7.4.3.1. Δίκτυο Γράφων Προσοχής

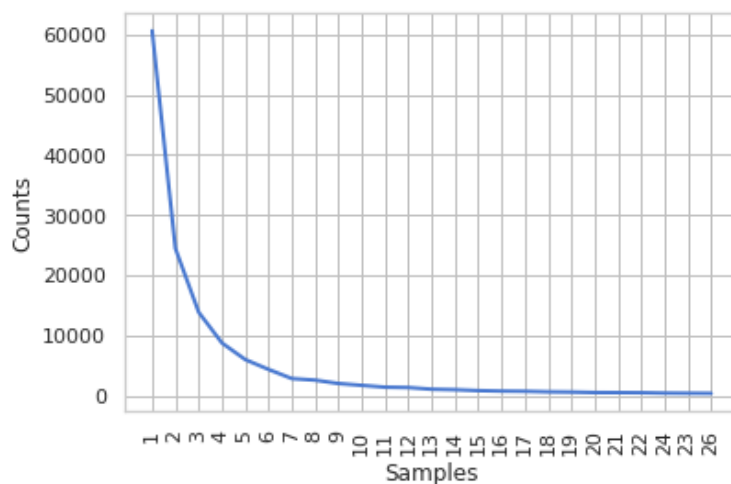
Αρχικά, ξεκινούμε από τις δομές που θα μας χρειαστούν για την υλοποίηση του δικτύου διάδοσης ειδήσεων: Το λεξικό γειτνίασης - adjacency dictionary, αποτελεί μια δομή κατά την οποία τα κλειδιά του λεξικού είναι τα indices των άρθρων και οι τιμές του εκάστοτε κλειδιού είναι τα indices των άρθρων που έχουν το ίδιο domain-author με το άρθρο, το index του οποίου βρίσκεται στο κλειδί. Έπειτα έχουμε τον πίνακα γειτνίασης, ο οποίος είναι ένας τετράγωνος πίνακας, του οποίου ο αριθμός των γραμμών όπως και ο αριθμός των στηλών είναι ίσος με τον αριθμό των ειδήσεων – κόμβων. Το κάθε κελί του πίνακα ισούται με 1 αν το άρθρο έχει κοινό domain-author με το άρθρο, του οποίου το index είναι η αντίστοιχη γραμμή στην οποία βρισκόμαστε. Σε αντίθετη περίπτωση, το κελί θα ισούται με 0. Δημιουργούμε μια συνάρτηση, η οποία δέχεται σαν είσοδο το λεξικό γειτνίασης και παράγει ως έξοδο, ζευγάρια από indices, τα οποία έχουν κοινό domain – author. Στη συνέχεια, επιτελούμε την προ-επεξεργασία δεδομένων, κατά την οποία επιλέγουμε την βιβλιοθήκη του SpaCy έναντι του Nltk, λόγω του ότι περιέχει περισσότερες λέξεις stop-words που αφορούν την ελληνική γλώσσα, αλλά επίσης έχει βρεθεί ότι είναι πιο αποδοτικό στην ελληνική γλώσσα. Κατά βήμα:

- Κάνουμε πεζούς όλους τους χαρακτήρες
- Χωρίζουμε το κάθε κείμενο σε λέξεις
- Αφαιρούμε τα stop words από κάθε κείμενο, που είναι οι πιο συχνά χρησιμοποιούμενες λέξεις
- Αφαιρούμε τα σημεία στίξης – punctuations
- Αφαιρούμε κάποια επιπρόσθετα punctuations, τα οποία έχουμε παρατηρήσει ότι δεν έχουν αφαιρεθεί
- Αφαιρούμε χαρακτήρες μηδενικού μήκους – κενά
- Παίρνουμε τα λήμματα των λέξεων που έχουν απομείνει

Δημιουργούμε ένα λεξικό, του οποίου τα κλειδιά είναι η κάθε λέξη που έχει παραμείνει στο λεξικό και οι τιμές είναι οι φορές που εμφανίζεται στα κείμενα ειδήσεων. Μέσω αυτής της ανάλυσης, ορίζουμε ένα κατώτατο όριο συχνότητας εμφάνισης λέξεων, μέσω του οποίου ουσιαστικά περιορίζουμε τις επαναλαμβανόμενες πληροφορίες. Ο αριθμός που επιλέγεται είναι το 8 και έχει εξαχθεί μέσω ενός ιστογράμματος στο σημείο που εξομαλύνεται, λαμβάνοντας

ταυτοχρόνως υπόψη και το κόστος σε πόρους. Το μέγεθος του λεξικού που μας έχει απομείνει είναι 19722 λέξεις.

Εικόνα 44: Ιστόγραμμα συχνότητας εμφάνισης λέξεων



Από τις εναπομείναντες λέξεις του λεξικού, δημιουργούμε τη δομή των χαρακτηριστικών των κόμβων – node features, έναν ορθογώνιο πίνακα, του οποίου οι γραμμές είναι ίσες με τον αριθμό των αρθρών και οι στήλες με το μέγεθος του λεξικού. Η κάθε γραμμή είναι και το index της αντίστοιχης είδησης και σε περίπτωση που η είδηση περιέχει τη λέξη – στήλη την οποία μελετάμε, θέτουμε την τιμή του κελιού ίση με 1 και σε αντίθετη περίπτωση με 0.

Μετατρέπουμε το σύνολο δεδομένων μας σε pytorch geometric dataset με εισόδους τις προαναφερθείσες δομές, όπου μετατρέπουμε τα δεδομένα σε μορφή tensors, για να είναι συμβατά με την προϋπάρχουσα υλοποίηση του Δικτύου Γράφων Προσοχής. Πέρα των tensors, τα χαρακτηριστικά που μπορούμε να εξάγουμε από ένα τέτοιας μορφής σύνολο δεδομένων είναι τα χαρακτηριστικά δεδομένων σε tensors, τα edge indices που είναι το σύνολο των ακμών σε μορφή tensors, τις ετικέτες επίσης σε μορφή tensors και τον αριθμό των κλάσεων που είναι δύο. Κάποια παραδείγματα εξόδων είναι:

Εικόνα 45: Χαρακτηριστικά του Pytorch Geometric συνόλου δεδομένων

```
data = dataset[0]
print(data.edge_index)
print(data.x)
print(data.y)

tensor([[ 0, 12, 0, ..., 5603, 4860, 5615],
        [4860, 5616, 4860, ..., 9647, 9685, 9672]])
tensor([[1., 0., 1., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        ...,
        [1., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.]])
tensor([0, 0, 1, ..., 0, 0, 0])
```

```
[ ] classes = data.num_classes = data.y.max().item() + 1

print(f"Number of Nodes in Greek News Dataset:", len(data.y))
print(f"Number of Node Features in Greek News Dataset:", dataset.num_node_features)
print(f"Number of Classes in Greek News Dataset:", classes)

Number of Nodes in Greek News Dataset: 9686
Number of Node Features in Greek News Dataset: 21681
Number of Classes in Greek News Dataset: 2
```

Με βάση το άρθρο των Kipf & Welling [57], ο διαχωρισμός των δεδομένων γίνεται ως εξής, με βάση τη μεταγωγική μάθηση: 20 ειδήσεις ανά κλάση για εκπαίδευση, 500 ειδήσεις για επαλήθευση και 1000 ειδήσεις για δοκιμές. Η υλοποίηση του δικτύου γράφων προσοχής παρατίθεται στην Εικόνα 46 και επιγραμματικά η κάθε γραμμή είναι: Ορισμός του πρώτου επίπεδου self.gat1, όπου έχουμε οκτώ κεφάλια, το οποίο έχει 8 χαρακτηριστικά το καθένα και αργότερα στο κομμάτι του forward, αφού τα χαρακτηριστικά των κόμβων περαστούν από ένα dropout επίπεδο, καλούν το επίπεδο του δικτύου προσοχής και παίρνουν τα ανανεωμένα βάρη, τα οποία πλέον έχουν μεταμορφωθεί λόγω του μηχανισμού προσοχής και ακολουθεί μια εκθετική γραμμική μονάδα ELU και ένα τελευταίο dropout επίπεδο (η δομή με τα δύο dropout επίπεδα, προτείνεται στο αντίστοιχο άρθρο του GAT, που αποτελεί το επίσημο άρθρο, για την εξαγωγή των τελικών βαρών. Το δεύτερο επίπεδο self.gat2, αποτελεί το στάδιο της ταξινόμησης, όπου έχει εισόδους της μεταμορφωμένες τιμές και πλέον έχουμε ένα κεφάλι, αριθμό χαρακτηριστικών ίσο με τον αριθμό των διαθέσιμων ετικετών του συνόλου δεδομένων, δηλαδή δύο (ψευδής, αληθής) και ακολουθείται από μια συνάρτηση softmax, η οποία μετατρέπει την έξοδο σε τιμές 0 έως 1, οι οποίες αποτελούν την πιθανότητα η εκάστοτε είδηση να ανήκει σε μια εκ των δύο κλάσεων. Πέραν την υλοποίησης αυτής, υπάρχει και μια δική μας υλοποίηση, η οποία έχει τα ίδια αποτελέσματα σε μεγαλύτερο όμως χρονικό διάστημα. Τόσο η υλοποίηση αυτής της συνάρτησης, όσο και του συνελκτικού δικτύου γράφων βασίζονται στο [58].

Εικόνα 46: Συνάρτηση Δικτύου Γράφων Προσοχής

```
class GAT(torch.nn.Module):
    def __init__(self, dim_in, dim_h, dim_out, heads=8):
        super().__init__()
        self.gat1 = GATConv(dim_in, dim_h, heads=heads)
        self.gat2 = GATConv(dim_h*heads, dim_out, heads=1)
        self.optimizer = torch.optim.Adam(self.parameters(),
                                           lr=0.005,
                                           weight_decay=5e-4)

    def forward(self, x, edge_index):
        h = F.dropout(x, p=0.6, training=self.training)
        h = self.gat1(x, edge_index)
        h = F.elu(h)
        h = F.dropout(h, p=0.6, training=self.training)
        h = self.gat2(h, edge_index)

        return h, F.log_softmax(h, dim=1)#.to("cuda")
```

Εκτελούμε πειράματα στο σύνολο δεδομένων με διάφορες παραμέτρους και τα βέλτιστα αποτελέσματα, τα οποία προέκυψαν, υπολογίζεται γύρω στο 89.39% ακρίβεια και 94.40% f1-score.

```
GAT(
  (gat1): GATConv(19722, 8, heads=8)
  (gat2): GATConv(64, 2, heads=1)
)
Epoch  0 | Train Loss: 0.692 | Train Acc: 57.16% | Val Loss: 0.69 | Val Acc: 56.30%
Epoch 10 | Train Loss: 0.056 | Train Acc: 99.40% | Val Loss: 0.52 | Val Acc: 94.33%
Epoch 20 | Train Loss: 0.045 | Train Acc: 99.40% | Val Loss: 0.38 | Val Acc: 94.33%
Epoch 30 | Train Loss: 0.042 | Train Acc: 99.40% | Val Loss: 0.29 | Val Acc: 94.33%
Epoch 40 | Train Loss: 0.040 | Train Acc: 99.40% | Val Loss: 0.32 | Val Acc: 94.33%
Epoch 50 | Train Loss: 0.039 | Train Acc: 99.40% | Val Loss: 0.32 | Val Acc: 94.33%
Epoch 60 | Train Loss: 0.038 | Train Acc: 99.40% | Val Loss: 0.31 | Val Acc: 94.33%
Epoch 70 | Train Loss: 0.036 | Train Acc: 99.40% | Val Loss: 0.31 | Val Acc: 94.33%
Epoch 80 | Train Loss: 0.034 | Train Acc: 99.40% | Val Loss: 0.31 | Val Acc: 94.33%
Epoch 90 | Train Loss: 0.034 | Train Acc: 99.40% | Val Loss: 0.31 | Val Acc: 94.31%
Epoch 100 | Train Loss: 0.031 | Train Acc: 99.40% | Val Loss: 0.31 | Val Acc: 94.31%

GAT test accuracy: 89.39%

GAT f1-score: 0.9439703822135925%
```

7.4.3.2. Συνελκτικό Δίκτυο Γράφων

Η ακριβώς ίδια διαδικασία ακολουθήθηκε και στα πειράματα του Συνελκτικού Δικτύου Γράφων. Επίσης έγινε, ακριβώς ο ίδιος διαχωρισμός δεδομένων και η υλοποίηση της συνάρτησης του Συνελκτικού Δικτύου Γράφων, παρατίθεται στην Εικόνα 47 και εξηγείται αναλυτικά στο Κεφάλαιο 4.4.2. και επιγραμματικά η κάθε γραμμή είναι: στο self.gcn καθορίζεται η διάσταση των χαρακτηριστικών που χρησιμοποιούνται ως εισόδοι και η διάσταση των κρυφών στρωμάτων, η οποία επιλέγεται να είναι 3 και στο self.out γίνεται η γραμμική μετατροπή των δεδομένων εισόδου. Ένα επίπεδο συνέλιξης γράφου, υπολογίζει τα νέα χαρακτηριστικά κάθε κόμβου, βασιζόμενο στα χαρακτηριστικά εισόδου, καθώς και στη δομή του γράφου και στη συνέχεια, επιλέγεται η ReLU ως συνάρτηση ενεργοποίησης και το τελικό αποτέλεσμα είναι να ανατίθενται μεγαλύτερα βάρη σε χαρακτηριστικά κόμβων, τα οποία δεν έχουν πολλούς γείτονες. Χρησιμοποιείται και πάλι ένα επίπεδο dropout για αποφυγή υπερπροσαρμογής και στη συνέχεια μέσω της σιγμοειδούς συνάρτησης softmax, γίνεται η πρόβλεψη της ετικέτας.

Εικόνα 47: Υλοποίηση συνάρτησης Συνελκτικού Δικτύου Γράφων

```
# GCN model with 2 layers
class Net(torch.nn.Module):
    def __init__(self):
        super(Net, self).__init__()
        self.conv1 = GCNConv(data.num_features, 16)
        self.conv2 = GCNConv(16, int(data.num_classes))

    def forward(self):
        x, edge_index = data.x, data.edge_index
        x = F.relu(self.conv1(x, edge_index))
        x = F.dropout(x, training=self.training)
        x = self.conv2(x, edge_index)
        return F.log_softmax(x, dim=1)
```

Τα βέλτιστα αποτελέσματα που προέκυψαν, παρατίθενται ανέρχονται στο 89.38% ακρίβεια και στο 94% f1-score.

7.3.4. Πειράματα με βάση τη Γνώση

Για τα πειράματα με βάση τη γνώση, η βιβλιογραφία ως προς τον εντοπισμό ψευδών ειδήσεων μέσω γράφων γνώσης είναι αρκετά περιορισμένη. Έπειτα από την ανασκόπησή της και τη μελέτη της, καταλήγουμε σε ένα μοντέλο, το οποίο ουσιαστικά συνδυάζει τις πιο αποδοτικές τεχνικές με βάση τα προηγούμενα πειράματα, μαζί με τη γνώση.

7.3.4.1 Πείραμα βασισμένο στο DETERRENT

Οι γράφοι γνώσης έχουν τη δυνατότητα να αναπαριστούν δομημένες σχέσεις μεταξύ οντοτήτων και γι' αυτό θεωρούνται ένα πολύ σημαντικό ερευνητικό πεδίο στη γνωστική και τεχνητή νοημοσύνη. Τα νευρωνικά δίκτυα γράφων, χρησιμοποιούν βαθιά νευρωνικά δίκτυα για να ενσωματώσουν τις πληροφορίες της τοπολογίας και τα χαρακτηριστικά των πληροφοριών στα δεδομένα του γράφου και στη συνέχεια παρέχουν ανανεωμένες αναπαραστάσεις των κόμβων. Η χρήση τους σε συνδυασμό με τη γνώση, εξυπηρετεί τις απαιτήσεις για εκμάθηση των δομικών χαρακτηριστικών των οντοτήτων και των σχέσεων και είναι πολύ αποδοτικά σε μια πληθώρα από τομείς, συμπεριλαμβανομένου και της ταξινόμησης ειδήσεων, που μας αφορά άμεσα.

Αρχικά, για να μπορέσουμε να φτιάξουμε τον γράφο γνώσης πρέπει να παράξουμε τα γνωστά ως triplets από το σύνολο δεδομένων. Ξεκινάμε χωρίζοντας το σύνολο δεδομένων σε προτάσεις. Έπειτα εκτελούμε την προ-επεξεργασία δεδομένων, αφαιρώντας τους άχρηστους χαρακτήρες που είναι θόρυβος των δεδομένων και εκτελώντας λημματοποίηση των λέξεων, για να αντιμετωπίσουμε την διαφορετική αντιμετώπιση της ίδιας λέξης και να περιορίσουμε την χρήση υπολογιστικού χρόνου και πόρων. Δεν μπορούμε να εκτελέσουμε την ολοκληρωμένη διαδικασία της προ-επεξεργασίας, λόγω του γεγονότος ότι χρειαζόμαστε τη γνώση καθαυτή για να μπορέσουμε να φτιάξουμε τα triplets και η αφαίρεση stopwords και αριθμών, επηρεάζει τη διαδικασία αυτή. Αργότερα, θέλουμε να πάρουμε τις οντότητες που περιέχονται στο σύνολο δεδομένων. Δυστυχώς, λόγω της ελληνικής γλώσσας το NER (Named Entity Recognition) δεν μπορεί να εντοπίσει όλες τις οντότητες που περιέχονται στο σύνολο δεδομένων, άρα πρέπει να το κάνουμε με δικό μας τρόπο. Έτσι, χωρίζουμε την κάθε πρόταση σε λέξεις και για κάθε λέξη, παίρνουμε την λέξη καθαυτή, το μέρος του λόγου που αποτελεί, την θέση – εξάρτησή της στην πρόταση και την λέξη από την οποία εξαρτάται. Η παραδοχή που γίνεται σε κάθε δουλειά κατασκευής γράφου γνώσης είναι ότι κάθε triplet αποτελείται από ουσιαστικό – ρήμα – ουσιαστικό, δηλαδή οι οντότητες είναι ουσιαστικά και η σχέση που τις ενώνει είναι ρήμα. Οι οντότητες είναι τα γνωστά μέρη της πρότασης ως το υποκείμενο και το αντικείμενο. Για αυτό ξεκινάμε με το να πάρουμε όλα τα ουσιαστικά που είναι οι υποψήφιοι οντότητες, που είναι η κάθε λέξη, μαζί με τα χαρακτηριστικά που προαναφέρθηκαν. Μετά, αφαιρούμε τις προτάσεις

που έχουν λιγότερο από δύο υποψήφιες οντότητες, αφού δεν μπορούν να μας παρέχουν την οποιαδήποτε γνώση λόγω του γεγονότος ότι για να φτιαχτούν triplets, χρειάζονται τουλάχιστον δύο οντότητες. Έτσι ανασυντάσσουμε το σύνολο δεδομένων ενώνοντας τις προτάσεις που είναι ικανές ή τουλάχιστον είναι πιθανόν να είναι ικανές να μας παρέχουν γνώση, έχοντας επίσης τις πιθανές οντότητες ανά πρόταση και την ετικέτα της είδησης. Ακολουθεί η μέθοδος σύνταξης triplets όπου παίρνουμε λέξεις ανά πρόταση και τα ενδεχόμενα που μελετάμε για τα δύο ουσιαστικά που παίρνουμε κάθε φορά για να δημιουργήσουμε triplets είναι: α) αν η λέξη από την οποία εξαρτώνται είναι η ίδια και είναι ρήμα, τότε τα δύο ουσιαστικά αποτελούν οντότητες και το ρήμα, αποτελεί τη σχέση τους, β) αν η λέξη από την οποία εξαρτώνται είναι ρήμα αλλά δεν είναι απαραίτητα η ίδια, τότε τα δύο ουσιαστικά αποτελούν οντότητες και το ένα ρήμα, αποτελεί τη σχέση τους, γ) αν τουλάχιστον μια εκ των δύο λέξεων εξαρτάται από ρήμα, τότε η λέξη που εξαρτάται από το ρήμα είναι το υποκειμένο, η άλλη το αντικείμενο και το ρήμα η σχέση, δ) μαζέψαμε όλους τους τύπους υποκειμένων, αντικειμένων και βλέπουμε αν το εκάστοτε ζεύγος λέξεων ανήκει σε αυτά και παίρνουμε το ρήμα ανάμεσά τους ως την σχέση. Αργότερα, ενώνουμε όλα τα triplets και αφού αναθέσουμε μια μοναδική ταυτότητα σε κάθε διαφορετική οντότητα και σε κάθε διαφορετική σχέση, αντικαθιστούμε τις λέξεις που αναπαριστούν τα triplets με τις αντίστοιχες ταυτότητες τους, διαδικασία γνωστή και ως entity linking, relationship linking. Κάθε λέξη που δεν είναι οντότητα και είναι παρούσα στο προ-επεξεργασμένο σύνολο δεδομένων, αποκτάει επίσης μια μοναδική ταυτότητα και για κάθε είδηση, κρατάμε τις ταυτότητες των λέξεων που περιέχει.

Εικόνα 48 : Πιθανές οντότητες, με το μέρος που είναι στην πρόταση, την λέξη από την οποία εξαρτώνται και το μέρος του λόγου που αποτελεί αυτή η λέξη

```
[['μάτι', 'obj', 'έβγαλαν', 'VERB'],
 ['αδέσποτο', 'obl', 'έβγαλαν', 'VERB'],
 ['σκυλί', 'flat', 'αδέσποτο', 'NOUN'],
 ['μεσολόγγι', 'nmod', 'αδέσποτο', 'NOUN'],
 ['φρικτού', 'nmod', 'μεσολόγγι', 'NOUN'],
 ['θεάματος', 'conj', 'φρικτού', 'NOUN'],
 ['σάββατο', 'nsubj', 'βρέθηκαν', 'VERB'],
 ['κάτοικοι', 'nsubj', 'βρέθηκαν', 'VERB'],
 ['μεσολογγίου', 'nmod', 'κάτοικοι', 'NOUN'],
 ['σκύλο', 'flat', 'αδέσποτο', 'X'],
 ['μάτι', 'obj', 'αφαίρσει', 'VERB']]
```

Εικόνα 49: Κείμενο με τα αντίστοιχά του triplets και ετικέτα

```
{'corpus': 'έβγαλαν τα μάτι από αδέσποτο σκυλί στο μεσολόγγι ενώπιον ενός φρικτού και βάνασον θεάματος βρέθηκαν το σάββατο οι κάτοικοι του μεσολογγίου όταν εντόπισαν ένα αδέσποτο σκύλο από τον οποίο είχαν αφαιρέσει τα μάτι . το άτυχο τετράποδο εντοπίστηκε νεκρό από κατοίκους της περιοχής οι οποίοι ζητούν σύμφωνα με την ιστοσελίδα agrinio site . gr τον εντοπισμό και τη σύλληψη του δολοφόνου και βασανιστής . αυτό το άτομο είναι επικίνδυνο για ζώα και άνθρωπος . οι αστυνομικές και δημοτικές αρχές διεξάγουν έρευνα προκειμένου να ταυτοποιήσουν τον δράστη ο οποίος εικάζουν ότι είναι κάτοικος της πόλη ',
'triplets': [['μάτι', 'έβγαλαν', 'αδέσποτο'],
 ['αδέσποτο', 'έβγαλαν', 'σκυλί'],
 ['θεάματος', 'βρέθηκαν', 'σάββατο'],
 ['σάββατο', 'βρέθηκαν', 'κάτοικοι'],
 ['κάτοικοι', 'βρέθηκαν', 'μεσολογγίου'],
 ['τετράποδο', 'εντοπίστηκε', 'κατοίκους'],
 ['αρχές', 'διεξάγουν', 'έρευνα'],
 ['έρευνα', 'ταυτοποιήσουν', 'δράστη'],
 ['δράστη', 'εικάζουν', 'κάτοικος']],
'label': 1}
```

Εικόνα 50: Πίνακας triplets

| | source | relation | target |
|--------|---------------|------------|-------------|
| 0 | μάπι | έβγαλαν | αδέσποτο |
| 1 | αδέσποτο | έβγαλαν | σκυλί |
| 2 | θεάματος | βρέθηκαν | σάββατο |
| 3 | σάββατο | βρέθηκαν | κάτοικοι |
| 4 | κάτοικοι | βρέθηκαν | μεσολογγίου |
| ... | ... | ... | ... |
| 295659 | περιστατικό | έγινε | φίλο |
| 295660 | δήμαρχος | εξέφρασε | πρόκληση |
| 295661 | πρόκληση | είχε | δήμαρχος |
| 295662 | agenda | διαφημίζει | σχέδιο |
| 295663 | συνομοσολόγοι | αφυπνίσουν | αναγνώστη |

Για να αποτυπωθεί πιο αποδοτικά η δομή του γράφου γνώσης, αρχικά έγινε η σκέψη χρήσης του Σχεσιακού-Συνελικτικού Δικτύου Γράφων (R-GCN) [59], το οποίο μοντελοποιεί σχεσιακά δεδομένα. Στο Σχεσιακό – Συνελικτικό Δίκτυο Γράφων, ο κάθε κόμβος αναπαριστάται αρχικά ως $h_i^{(l)}$. Η διάδοση σε επίπεδα, ενημερώνει τις αναπαραστάσεις των κόμβων, χρησιμοποιώντας τις αναπαραστάσεις των γειτονικών κόμβων, όπως παρουσιάζεται στην Εικόνα 51.

Εικόνα 51: Ενημέρωση αναπαραστάσεων κόμβων σε Σχεσιακό – Συνελικτικό Δίκτυο Γράφων

$$h_i^{(l+1)} = \sigma \sum_{r \in R'} \sum_{(j,r,i) \in T'} \frac{1}{c_{i,r}} w^r h_j^{(l)}$$

όπου $c_{i,r}$ παράγοντας κανονικοποίησης, που συνήθως τίθεται ίσος με τον αριθμό των γειτονικών κόμβων του εκάστοτε κόμβου i υπό τη σχέση r , w^r είναι μια παράμετρος που εξαρτάται από τις ακμές και επιστρέφει το αντίστοιχο βάρος και $\sigma()$, συνάρτηση ενεργοποίησης, η οποία στην προκειμένη περίπτωση, έχει επιλεγθεί να είναι η LeakyRelu. Όμως, ο κάθε γειτονικός κόμβος, έχει διαφορετική σημασία, ως προς τις αναπαραστάσεις των κόμβων, έτσι θέτουμε ένα μηχανισμό προσοχής, ο οποίος λειτουργεί με ακριβώς τον ίδιο τρόπο που είδαμε και στα δίκτυα γράφων προσοχής. Έτσι, αντί του Σχεσιακού – Συνελικτικού Δικτύου Γράφων, χρησιμοποιείται ένα Σχεσιακό – Δίκτυο Γράφων Προσοχής (R-GAT), το οποίο ουσιαστικά είναι ένας συνδυασμός του R-GCN και του GAT και η ενημέρωση των κόμβων, εκτελεί διαφορετικούς υπολογισμούς για κάθε σχέση και τροποποιείται, ως εξής:

Εικόνα 52: Ενημέρωση αναπαραστάσεων κόμβων σε Σχεσιακό – Δίκτυο Γράφων Προσοχής

$$h_i^{(l+1)} = \sigma \sum_{r \in R'} \sum_{(j,r,i) \in T'} \alpha_{ij}^r w^r h_j^{(l)}$$

όπου a_{ij}^r ο μηχανισμός προσοχής, που ορίζεται στην Εικόνα 53.

Εικόνα 53: Μηχανισμός Προσοχής

$$a_{ij}^r = \frac{e^{a^r u_{ij}^r}}{\sum_{(k,r,i) \in T^r} e^{a^r u_{ik}^r}} \quad \text{όπου } u_{ij}^r = W^r (h_i^{(l)} \parallel h_j^{(l)})$$

Επειδή έχουμε πολλές σχέσεις και ένα γενικό φαινόμενο το οποίο παρατηρείται στα σχεσιακά δίκτυα γράφων είναι η ανεξέλεγκτη αύξηση των παραμέτρων, για την αντιμετώπιση αυτού, κανονικοποιούμε τα δεδομένα μέσω της διαδικασίας basis decomposition, όπου γίνεται μείωση του αριθμού των παραμέτρων και ουσιαστικά αντί να χρησιμοποιείται ένας πίνακας βαρών για κάθε σχέση και κάθε επίπεδο, χρησιμοποιείται ένας γραμμικός συνδυασμός basis πινάκων, οι οποίοι είναι λιγότεροι των πινάκων ανά σχέση που χρησιμοποιούνταν πιο πριν.

Η βασική χρήση των GNNs είναι η συλλογή των πληροφοριών των γειτονικών κόμβων και αυτό που προτείνεται είναι η χρήση του μηχανισμού προσοχής για τη χρήση του κειμένου των ειδήσεων, το οποίο χρησιμοποιείται κατά το στάδιο της διάδοσης του εκάστοτε μηνύματος – κειμένου κατά μήκος της κάθε διαδρομής του γράφου προσοχής μέσα στο επίπεδο του R-GAT. Δηλαδή, γίνεται χρήση του κωδικοποιητή κειμένου BiGRU, ο οποίος κωδικοποιεί τις ακολουθίες λέξεων αμφίδρομα (μπροστά και πίσω ως προς την κάθε πρόταση) και αναπαριστά το εκάστοτε κείμενο ως μια ακολουθία αναπαραστάσεων της κάθε λέξης. Έχοντας υπολογίσει τα διανύσματα των λέξεων του κάθε άρθρου v_1, \dots, v_s , το διάνυσμα του ολικού άρθρου υπολογίζεται ως:

Εικόνα 54: Υπολογισμός διανύσματος του άρθρου

$$\begin{aligned} \vec{s}_t &= GRU(\vec{s}_{t-1}, v_t) \\ \overleftarrow{s}_t &= GRU(\overleftarrow{s}_{t-1}, v_t) \end{aligned}$$

Ενώνουμε τις δύο καταστάσεις, παίρνοντας ως $s_t = [\vec{s}_t, \overleftarrow{s}_t]$ τα διανύσματα που δημιουργήθηκαν σύμφωνα με τα συμφραζόμενα γύρω από την εκάστοτε λέξη που μελετάται. Για να βρούμε ποιες λέξεις είναι οι πιο σημαντικές, γίνεται και πάλι χρήση του μηχανισμού προσοχής για να δοθεί εν τέλη το διάνυσμα για το κάθε άρθρο, το οποίο πέραν της σημασιολογίας της κάθε λέξης, συμπεριλαμβάνει και το πόσο σημαντική είναι.

Εικόνα 55: Σημασία της εκάστοτε λέξης στο άρθρο

$$\beta_t = \frac{e^{u_t^T g'}}{\sum_{k=1}^{|S|} e^{u_k^T g'}} \quad \text{όπου} \quad u_t = \tanh(W_c s_t + b_c)$$

Εικόνα 56: Υπολογισμός διανύσματος κάθε άρθρου

$$c = \sum_{t=1}^{|S|} \beta_t s_t$$

όπου u_t η κρυφή αναπαράσταση της λέξης v_t . Ο πλήρης ορισμός του g δίνεται στην Εικόνα 57 και εξηγείται αναλυτικά παρακάτω.

Εικόνα 57: Χρήση γνώσης σε συνδυασμό με τις πληροφορίες κειμένου

$$g' = \gamma g + (1 - \gamma) W_k h^s$$

Το διάνυσμα g είναι η κωδικοποίηση του περιεχομένου του εκάστοτε άρθρου. Το h^s είναι τα διανύσματα των κόμβων του άρθρου S εμπλουτισμένες πλέον με τον μηχανισμό προσοχής, τόσο ως προς τη σημαντικότητα των λέξεων ως προς το κάθε άρθρο, όσο και ως προς τη σημαντικότητα των σχέσεων ως προς τον κάθε γράφο και W_k είναι ένας πίνακας μετασχηματισμού. Αν $\gamma = 1$, το g' ισούται με g και η ταξινόμηση γίνεται χωρίς πληροφορίες από τον γράφο γνώσης, ενώ σε διαφορετική περίπτωση, το g' λαμβάνει υπόψη BiGRU και τον γράφο γνώσης, καθώς και το σκορ προσοχής β_t για το διάνυσμα των αρθρών c .

Ουσιαστικά συνδυάζεται η διάδοση της γνώσης στο γράφο, άρα κατ' επέκταση η δομή του, με τη σημασιολογία κάθε είδησης, δηλαδή το νόημα που εμπεριέχει. Έπειτα τοποθετείται ένα επίπεδο dense, στο οποίο κάθε νευρώνας δέχεται την είσοδο από κάθε νευρώνα του προηγούμενου επιπέδου, στο οποίο ορίζεται το g' και έπειτα μέσω της χρήσης της σιγμοειδούς συνάρτησης, μας δίνει κατά πόσο ένα άρθρο είναι ψευδές ή αληθές. Για κάθε άρθρο, ο σκοπός είναι να μειωθεί όσο πιο πολύ γίνεται το binary cross entropy loss, όπου \hat{y} η ετικέτα του άρθρου, δηλαδή 0 ή 1, αληθές ή ψευδές αντίστοιχα.

Εικόνα 58: Μείωση του Binary Cross Entropy Loss

$$BCE = -\frac{1}{N} \sum_{i=0}^N y_i \log \hat{y}_i - (1 - y_i) \log (1 - \hat{y}_i)$$

Θα εφαρμόσουμε κάποιες συγκεκριμένες τεχνικές για να διασφαλίσουμε την αποδοτικότητα του μοντέλου. Εισάγουμε μια υπερπαραμέτρο γνωστή ως dropout rate, που καθορίζει την πιθανότητα κατά την οποία οι έξοδοι του κάθε επιπέδου απορρίπτονται ή αντιστρόφως, η πιθανότητα διατήρησης των εξόδων του κάθε επιπέδου, δηλαδή επιλέγει τυχαία νευρώνες οι οποίοι δεν θα συμβάλουν καθόλου κατά την εκπαίδευσης. Αυτό που προσφέρει το dropout rate είναι η αποτροπή του φαινομένου υπερπροσαρμογής. Χρησιμοποιούμε άλλη μια υπερπαραμέτρο

γνωστή ως learning rate, η οποία καθορίζει την αλλαγή των βαρών σε κάθε βήμα της εκπαίδευσης και τον ρυθμό εκμάθησης των δεδομένων του μοντέλου. Η τελευταία σημαντική υπερπαράμετρος που χρησιμοποιείται είναι το decay_factor, το οποίο είναι μια μορφή κανονικοποίησης των αποτελεσμάτων, με σκοπό και πάλι την αποτροπή της υπερπροσαρμογής. Δυστυχώς, η εύρεση των κατάλληλων τιμών μπορεί να βρεθεί μόνο μέσω trial and error, δηλαδή μόνο μέσω πειραμάτων, διαφόρων δοκιμών και σφαλμάτων. Αφού επιβεβαιώσουμε ότι το μοντέλο είναι κατάλληλο για το ζήτημα που μελετάται, κάνοντας overfit ένα μικρό μέρος των δεδομένων και βλέποντας ότι η απώλεια μηδενίζεται, σταδιακά εισάγουμε όλα τα δεδομένα και προσπαθούμε να παράξουμε τα βέλτιστα αποτελέσματα, τα οποία καταλήγουμε ότι ισούνται με 83,33% ακρίβεια και 85% f1-score.

7.4. Σύγκριση Απόδοσης Μοντέλων σε διάφορα σύνολα δεδομένων

Για τη σύγκριση της απόδοσης τόσο των μοντέλων, όσο και του συνόλου δεδομένων που δημιουργήθηκε, θα τρέξουμε τα πειράματα σε σύνολα δεδομένων, τα οποία χρησιμοποιήθηκαν εκτενώς στη βιβλιογραφία και θα συγκρίνουμε την απόδοσή τους, σε σχέση με τις υλοποιήσεις που προτείνονται στα αντίστοιχα papers. Για να συγκριθούν τα μοντέλα, πρέπει να συγκριθεί η απόδοσή τους, τόσο ως προς την ακρίβεια, όσο και ως προς το F1-Score.

Γίνεται εκπαίδευση των μοντέλων και αξιολόγηση τους ως προς τα σύνολα δεδομένων FA-KES, FakeNewsNet, ISOT Dataset, Cancer, Diabetes. Το FA-KES είναι ένα σύνολο δεδομένων, το οποίο περιέχει 804 ειδήσεις που αφορούν τον Συριακό πόλεμο και αποτελείται από μια σειρά από άρθρα που έχουν ετικέτες 0 (fake) ή 1 (real). Κάθε άρθρο έχει ημερομηνία κυκλοφορίας, τοποθεσία, τίτλο, καθώς και το κείμενο που περιγράφει λεπτομερώς την είδηση. Για τις ανάγκες της αξιολόγησης, συνδυάζονται οι τίτλοι με τα κείμενα, για να γίνει εξαγωγή όλης της πληροφορίας που περιέχει μια είδηση. Το FakeNewsNet, περιέχει ειδήσεις από δύο ιστοσελίδες εντοπισμού ψευδών ειδήσεων, το Gossipcop και το PolitiFact και περιέχουν tweets που διαδίδουν ψευδείς και αληθείς ειδήσεις. Επιπροσθέτως, περιέχει την ταυτότητα του κάθε tweet, την ιστοσελίδα προέλευσής του, ετικέτα ψευδούς (0) και αληθούς (1) είδησης και τον τίτλο του κάθε tweet. Σε αυτή την περίπτωση, χρησιμοποιούνται μόνο τίτλοι, οι οποίοι ουσιαστικά είναι επιγραμματικά το κείμενο που αναγράφεται στο tweet και παρατηρούμε την απόδοση, χρησιμοποιώντας μόνο μικρά κομμάτια κειμένου. Στο ISOT Dataset έχουμε ένα σύνολο ειδήσεων που συγκεντρώθηκαν οι αληθείς από το Reuters και οι ψευδείς από τη Wikipedia και το PolitiFact. Για τις ανάγκες του πειράματος, έχει προστεθεί μια στήλη ετικέτας που καταδεικνύει κατά πόσο μια είδηση είναι ψευδής ή αληθής. Το σύνολο δεδομένων Cancer περιέχει διάφορες πληροφορίες γύρω από τον καρκίνο, είτε έχει να κάνει με θεραπείες, με φάρμακα ή άλλες ιατρικές πληροφορίες, με κάποιες πληροφορίες να είναι αληθείς και άλλες ψευδείς. Περιέχει ήδη ετικέτες με 0 (αληθής) και 1 (ψευδής) είδηση και για αξιολόγηση, γίνεται χρήση των τίτλων σε συνδυασμό με το κείμενό τους. Τέλος το σύνολο δεδομένων Diabetes, περιέχει πληροφορίες που συνδέονται με τον διαβήτη, θεραπείες ή και φάρμακα και γενικότερα ιατρι-

κές πληροφορίες. Κάποιες πληροφορίες είναι αληθείς – 0 και άλλες ψευδείς – 1. Για την αξιολόγηση, χρησιμοποιούνται και πάλι οι τίτλοι των πληροφοριών σε συνδυασμό με τα κείμενα. Τα αναλυτικά στατιστικά κάθε συνόλου δεδομένων, παρατίθενται στον παρακάτω πίνακα.

Εικόνα 59: Στατιστικά Συνόλων Δεδομένων

| Σύνολο Δεδομένων | Τύπος Ειδήσεων | Αριθμός Ειδήσεων |
|------------------|----------------|------------------|
| FA-KES | Αληθείς | 426 |
| | Ψευδείς | 378 |
| FakeNewsNet* | Αληθείς | 33634 |
| | Ψευδείς | 10646 |
| ISOT Dataset | Αληθείς | 21417 |
| | Ψευδείς | 23481 |
| Diabetes | Αληθείς | 1661 |
| | Ψευδείς | 608 |
| Cancer | Αληθείς | 4623 |
| | Ψευδείς | 1476 |

* Το σύνολο δεδομένων FakeNewsNet, αποτελείται από δύο άλλα σύνολα δεδομένων, το PolitiFact και το Gossipcop.

Εικόνα 60: Baseline υλοποιήσεις ως προς το περιεχόμενο σε άλλα σύνολα δεδομένων

| Baseline μέθοδοι με βάση το περιεχόμενο στα υπόλοιπα σύνολα δεδομένων | | | | | | | | | | |
|-----------------------------------------------------------------------|--------------|--------------|----------|-------|--------|----|-------|----|-------|----|
| Σύνολα Δεδομένων | FastText | | Word2Vec | | TF-IDF | | GloVe | | SpaCy | |
| | Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 |
| FA-KES | 58.00 | 70.42 | 51.60 | 62.53 | 57 | 51 | 58 | 50 | 55 | 54 |
| Fake- NewsNet | 76.21 | 69.80 | 74.10 | 45.42 | 81 | 77 | 76 | 65 | 75 | 64 |
| ISOT Dataset | 98.07 | 98.14 | 89.10 | 90.72 | 94 | 94 | 87 | 87 | 88 | 88 |
| Cancer | 91.60 | 84.15 | 86.3 | 75.04 | 82 | 78 | 82 | 80 | 81 | 78 |
| Diabetes | 93.50 | 88.69 | 76.9 | 64.42 | 72 | 62 | 81 | 80 | 78 | 76 |

Εικόνα 61: : Baseline υλοποιήσεις ως προς το περιεχόμενο σύμφωνα με τα συμφραζόμενα σε άλλα σύνολα δεδομένων

| Μοντέλο BERT στα υπόλοιπα σύνολα δεδομένων | | |
|---------------------------------------------------|------------------------------------|------------------|
| Σύνολα Δεδομένων | BERT-multi-cased L-12 H-768 | |
| | Accuracy | F1-Score |
| FA-KES | 52.74 | 42 |
| FakeNewsNet | 78.06 | 74 |
| ISOT Dataset | <u>98.15</u> | <u>98</u> |
| Cancer | 77 | 74 |
| Diabetes | 76 | 72 |

Εικόνα 62: Σύγκριση απόδοσης συνόλων δεδομένων με BiGRU, με το βέλτιστο μοντέλο OPCNN

| Σύγκριση μοντέλου με BiGRU και μοντέλου OPCNN [32] | | | | |
|-----------------------------------------------------------|---------------------|---------------------|---------------------|---------------------|
| Σύνολα Δεδομένων | BiGRU | | OPCNN | |
| | Accuracy (%) | F1-Score (%) | Accuracy (%) | F1-Score (%) |
| FA-KES | 52 | 51 | 53.99 | 53.99 |
| FakeNewsNet | 85 | 77 | 95.26 | 95.27 |
| ISOT Dataset | <u>99.96</u> | <u>99.96</u> | 99.99 | 99.99 |

| Μοντέλο με BiGRU στα υπόλοιπα σύνολα δεδομένων | | |
|-------------------------------------------------------|---------------------|---------------------|
| Σύνολα Δεδομένων | Accuracy (%) | F1-Score (%) |
| Cancer | <u>97</u> | <u>97</u> |
| Diabetes | 96 | 96 |

Εικόνα 63: Απόδοση συνόλων δεδομένων με την υλοποίηση με χρήση GAT

| Μοντέλο με χρήση GAT στα υπόλοιπα σύνολα δεδομένων | | |
|----------------------------------------------------|---------------------|---------------------|
| Σύνολα Δεδομένων | GAT | |
| | Accuracy (%) | F1-Score (%) |
| FA-KES | 87.24 | 93.18 |
| FakeNewsNet | 86.89 | 92.98 |
| Cancer | 75.85 | 82.12 |
| Diabetes | <u>87.29</u> | <u>93.21</u> |

Εικόνα 64: Απόδοση συνόλων δεδομένων με την υλοποίηση με βάση το DETERRENT

| Σύγκριση KG με μοντέλο DETERRENT[43] | | | | |
|--------------------------------------|---------------------|---------------------|--------------|--------------|
| Σύνολα Δεδομένων | KG | | DETERRENT | |
| | Accuracy (%) | F1-Score (%) | Accuracy (%) | F1-Score (%) |
| Diabetes | <u>92.86</u> | <u>85.71</u> | 92.06 | 84.74 |
| Cancer | 83.87 | 84.85 | 96.52 | 93.09 |

Παρατηρούμε και πάλι ότι η υλοποίηση του νευρωνικού δικτύου που χρησιμοποιεί BiGRU είναι η πιο αποδοτική και σε αρκετές περιπτώσεις, ξεπερνά σε απόδοση και τις αρχικές υλοποιήσεις στις οποίες έχουν χρησιμοποιηθεί τα σύνολα δεδομένων. Όπως αναμενόταν και από το [60], εάν το μοντέλο που χρησιμοποιείται, ανταποκρίνεται στις ανάγκες του συνόλου δεδομένων και του θέματος που μελετάται, ακόμα και σε μια πολύ απλή δομή, μπορεί να πετύχει πολύ υψηλές αποδόσεις. Όσον αφορά την υλοποίηση του Δικτύου Γράφου Προσοχής (GATs), σημειώνει επίσης πολύ υψηλές επιδόσεις στον τομέα ανίχνευσης ψευδών ειδήσεων, ενώ ιδιαίτερα στο σύνολο δεδομένων FA-KES πετυχαίνει ακόμα πιο υψηλές επιδόσεις, σε σχέση με τη βέλτιστη υλοποίηση στη βιβλιογραφία. Το επίπεδο του μηχανισμού προσοχής που υλοποιείται στα δίκτυα αυτά, είναι υπολογιστικά αποδοτικό, αφού η χρήση του βοηθάει στην αποφυγή οι πολύπλοκων πράξεων με πίνακες και είθισται η εφαρμογή του σε κόμβους διαφορετικού είδους και με μεταβλητό αριθμό γειτόνων, ενώ δεν απαιτείται η γνώση ολόκληρου του γράφου εκ των προτέρων. Η υλοποίηση μέσω γράφου γνώσης (KG) παρατηρούμε ότι παρά το γεγονός ότι έχει αρκετά καλά αποτελέσματα στο σύνολο δεδομένων ελληνικής γλώσσας, αρχικά παρατηρούμε ότι αποδίδει καλύτερα σε σύνολα δεδομένων ξένης γλώσσας και εν συνεχεία τα αποτελέσματά της υλοποίησης που γίνεται στα πλαίσια της διπλωματικής, πλησιάζουν αρκετά

την βέλτιστη υλοποίηση που βρέθηκε στη βιβλιογραφία, αλλά και πάλι τα αποτελέσματα είναι πιο χαμηλά, εκτός ως προς το σύνολο δεδομένων Diabetes, όπου πετυχαίνει ακόμα καλύτερα αποτελέσματα από την υλοποίηση του DETERRENT.

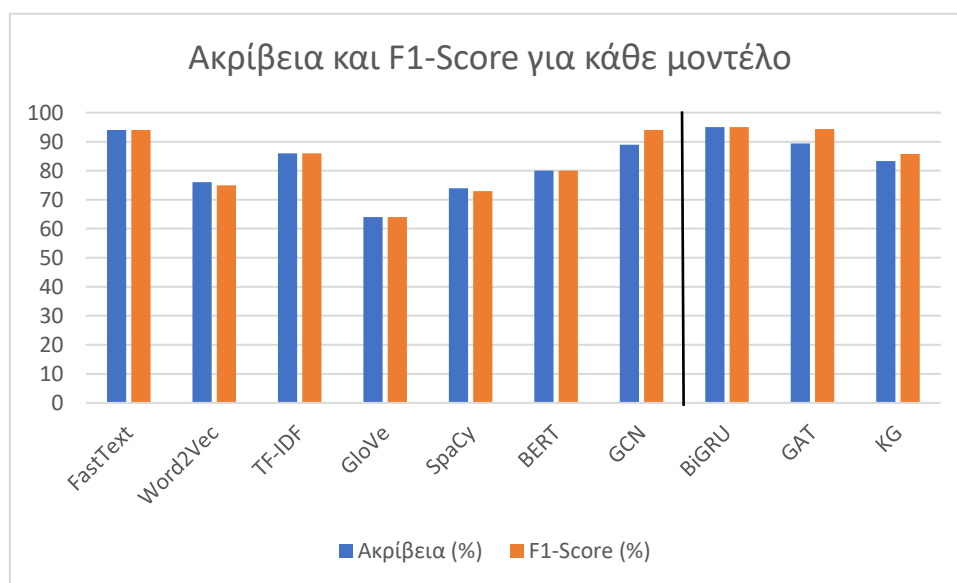
Εν κατακλείδι, παρατηρούμε ότι η βέλτιστη υλοποίηση που εντοπίζεται στην βιβλιογραφία και αφορά ξένα κείμενα είναι το μοντέλο OPCNN, το οποίο χρησιμοποιεί ένα απλό νευρωνικό δίκτυο, αποτελούμενο από ένα πίνακα διανυσμάτων λέξεων που έχουν προκύψει μέσω GloVe, ένα επίπεδο διανυσμάτων, ένα επίπεδο dropout, ένα συνελκτικό επίπεδο στο οποίο μετατρέπονται τα διανύσματα σε πίνακες μικρότερων διαστάσεων με τη χρήση φίλτρων, ένα pooling επίπεδο όπου μειώνονται οι υπολογιστικές πράξεις και επιστρέφει τις μεγαλύτερες τιμές διανυσμάτων που περιέχουν την πιο σημαντική πληροφορία, ένα flatten επίπεδο που μετατρέπει το κείμενο σε ένα μονοδιάστατο πίνακα και ένα επίπεδο εξόδου το οποίο μέσω σιγμοειδούς συνάρτησης ταξινομούνται οι ειδήσεις σε ψευδείς και αληθείς, με απόδοση που πλησιάζει το τέλειο (99,99%) τόσο σε ακρίβεια, όσο και σε f1-score.

Όσο αφορά το σύνολο δεδομένο μας, το οποίο είναι στην ελληνική γλώσσα, παρατηρούμε και πάλι ότι η υλοποίηση ως προς το περιεχόμενο με βάση τα συμφραζόμενα που χρησιμοποιεί διανύσματα BiGRU, σε συνδυασμό με CNN και ένα απλό νευρωνικό δίκτυο είναι η πιο αποδοτική, ως προς τον εντοπισμό ειδήσεων με αποτελέσματα τόσο σε ακρίβεια, όσο και σε f1-score να ανέρχονται στο 95%. Έπειτα ακολουθεί ο εντοπισμός μέσω δικτύου διάδοσης των ειδήσεων, με τη χρήση Δικτύου Γράφου Προσοχής, με απόδοση που ανέρχεται στο 94%. Τέλος, ακολουθεί ο εντοπισμός μέσω της γνώσης με 83,33%, πράγμα αναμενόμενο, καθώς ο τομέας εντοπισμού ψευδών ειδήσεων μέσω γνώσης είναι αρκετά καινούργιος και το υλικό που υπάρχει είναι αρκετά περιορισμένο, ιδιαίτερα σε μια γλώσσα που υπολείπεται βοηθητικά εργαλεία, όπως την ελληνική γλώσσα. Στις πλείστες των περιπτώσεων, οι προϋπάρχουσες υλοποιήσεις στη βιβλιογραφία, δίνουν τα καλύτερα αποτελέσματα στα σύνολα δεδομένων, με εξαίρεση τα σύνολα δεδομένων Diabetes και FA-KES, όπου οι δικές μας υλοποιήσεις είχαν καλύτερα αποτελέσματα.

8. Συμπεράσματα και Μελλοντικές Βελτιώσεις

Εικόνα 65: Απόδοση baseline μοντέλων και δικών μας υλοποιήσεων στο ελληνικό σύνολο δεδομένων

| <u>Μοντέλο</u> | <u>Ακρίβεια (%)</u> | <u>F1-Score (%)</u> |
|----------------|---------------------|---------------------|
| FastText | 94 | 94 |
| Word2Vec | 76 | 75 |
| TF-IDF | 86 | 86 |
| GloVe | 64 | 64 |
| SpaCy | 74 | 73 |
| BERT | 80 | 80 |
| GCN | 89 | 94 |
| BiGRU | <u>95</u> | <u>95</u> |
| GAT | <u>89.39</u> | <u>94.40</u> |
| KG | <u>83.33</u> | <u>85.71</u> |



Λόγω του γεγονότος ότι έχουμε ανισόρροπο σύνολο δεδομένων, η απόδοση του μοντέλου παρατηρείται κυρίως μέσω του f1-score, καθώς παρέχει πληροφορίες τόσο ως προς τις σωστές, όσο και ως προς τις λανθασμένες προβλέψεις. Η ακρίβεια είναι μια αρκετά αξιόπιστη μετρική

και σε αυτή την περίπτωση, όμως δεν δείχνει πόσο αποδοτικό είναι το μοντέλο ως προς και τις 2 ετικέτες, αλλά γενικά ως προς τις προβλέψεις που έχει κάνει

Δεξιά της μαύρης κάθετης γραμμής είναι οι baseline υλοποιήσεις, δηλαδή υλοποιήσεις που ναι μεν πραγματοποιήθηκαν αλλά υπάρχουν ήδη στη βιβλιογραφία, ενώ αριστερά είναι τα πειράματα που υλοποιήθηκαν, με βάση τα συμπεράσματα που εξήχθησαν από την βιβλιογραφία. Όπως έχει παρατηρηθεί, με μικρή διαφορά τα βέλτιστα αποτελέσματα έχουν προκύψει μέσω του μοντέλου που χρησιμοποιεί BiGRU με 95%, τόσο ως προς τις υλοποιήσεις με βάση το περιεχόμενο – καθώς αποδεικνύεται ότι η χρήση του περιεχομένου που λαμβάνει υπόψη τα συμφραζόμενα δίνει καλύτερα αποτελέσματα – όσο και ως προς τις υλοποιήσεις με βάση τη διάδοση και τη γνώση και αποτελεί το βέλτιστο μοντέλο που δημιουργήσαμε. Στον τομέα εντοπισμού ψευδών ειδήσεων σύμφωνα με την διάδοση τους, το δίκτυο γράφων προσοχής είχε την καλύτερη απόδοση με 94%. Στον τομέα εντοπισμού με βάση τη γνώση, η υλοποίηση με χρήση της γνώσης που χρησιμοποιήθηκε, έφτασε μέχρι και το 83,33%. Άρα ο εντοπισμός των ειδήσεων ως προς το περιεχόμενο με βάση τα συμφραζόμενα, αποτελεί τον πιο αποδοτικό τρόπο, με τη διάδοση ειδήσεων μέσω δικτύου γράφου προσοχής να είναι πολύ κοντά σε απόδοση.

Το γεγονός ότι ο εντοπισμός μέσω των συμφραζομένων είχε έστω κατά πολύ λίγο τα καλύτερα αποτελέσματα, δεν αποτελεί έκπληξη, καθώς σύμφωνα και με το [60], οι τεχνικές που δημιουργούν διανύσματα με βάση τα συμφραζόμενα, γνωστά και ως contextual embeddings, έχουν αναπάντεχα καλές αποδόσεις, ακόμα και όταν χρησιμοποιούνται σε πολύ απλά νευρωνικά δίκτυα. Επίσης, σύμφωνα και με την βιβλιογραφία, οι τεχνικές σύμφωνα με τα συμφραζόμενα είναι μακράν οι πιο αποδοτικές και υπάρχουν περιπτώσεις στις οποίες αγγίζουν το τέλειο με 99,99%, όπως παρουσιάζεται και στο προηγούμενο κεφάλαιο. Οι τεχνικές αποκλειστικά με βάση το περιεχόμενο, στην πλειοψηφία τους έχουν καλή απόδοση, με τις περισσότερες όμως να έχουν τους περιορισμούς τους, όταν χρησιμοποιούνται απλά σε συνδυασμό με ένα ταξινομητή, όπως παρουσιάστηκε στο κεφάλαιο 6 και στην βιβλιογραφία, πράγμα το οποίο επιβεβαιώνεται και με το ελληνικό σύνολο δεδομένων. Το γεγονός αυτό υποστηρίζεται και στη σύγκριση με την προσέγγιση που χρησιμοποιείται BiGRU, όπου το μόνο μοντέλο που χρησιμοποιεί το περιεχόμενο με τη χρήση προ-εκπαιδευμένων ενσωματώσεων για κάθε λέξη χωρίς να αντιλαμβάνεται τη θέση της λέξης στο κείμενο το οποίο είχε συγκρίσιμα αποτελέσματα είναι αυτό του FastText. Ο λόγος αυτού του φαινομένου είναι το γεγονός ότι το FastText διαθέτει πρόγνωση σε πολύ μεγάλα σύνολα δεδομένων, τα οποία συμπεριλαμβάνουν και την ελληνική γλώσσα και παρέχουν επιπρόσθετη γνώση και εμπειρία στην εκπαίδευση, ενώ χρησιμοποιεί διάφορες παραμέτρους όπως το n-gram, οι οποίες δίνουν καλύτερα αποτελέσματα. Λόγω των διαφορών στα αποτελέσματα, φαίνεται για άλλη μια φορά ο ρόλος των συμφραζομένων στα διανύσματα και στην απόδοση του εκάστοτε μοντέλου, σε σχέση με όλες τις άλλες τεχνικές.

Όσο αφορά τον εντοπισμό με τη χρήση γνώσης, τα αποτελέσματα επιδιώξαμε να είναι ακόμα καλύτερα, από το μοντέλο που χρησιμοποιεί τα συμφραζόμενα, πράγμα το οποίο δεν φαίνεται στην πράξη. Κάποιοι πιθανοί λόγοι της διαφοράς αυτής είναι:

- Η ποιότητα των triplets, καθώς για την ελληνική γλώσσα, δεν υπάρχει ούτε εργαλείο αποσαφήνισης οντοτήτων όπως το NeuralCoref, ούτε κάποιο εργαλείο το οποίο μπορεί με μεγάλη ακρίβεια να μας δώσει τις οντότητες αυτόματα. Για αυτό το λόγο, έγινε ένας αυτοσχέδιος σχηματισμός triplets, ο οποίος να μην περιέχει όλα τα triplets που χρειάζονται, όμως από την άλλη περιέχει triplets τα οποία ενδεχομένως να μην ανταποκρίνονται πλήρως στον ορισμό υποκείμενο – ρήμα - αντικείμενο. Αυτό βάζει σε κίνδυνο την ποιότητα των triplets, τα οποία όπως είναι φυσιολογικό, επηρεάζουν την αποδοτικότητα του μοντέλου. Στο μοντέλο του DETTERENT στο οποίο βασίσαμε την υλοποίησή μας, χρησιμοποιείται μια προϋπάρχουσα βάση, που περιέχει οντότητες αγγλικής γλώσσας και είναι κατασκευασμένη ακριβώς για το λόγο δημιουργίας γράφου γνώσης, αφού περιέχει triplets στη μορφή υποκείμενο – ρήμα – αντικείμενο, πράγμα το οποίο διαφυλάσσει την ποιότητα των triplets και κατ' επέκταση την αποδοτικότητα του μοντέλου.

- Η ποσότητα των σχέσεων, καθώς στο δικό μας σύνολο δεδομένων έχουμε δεκάδες χιλιάδες σχέσεις, οι οποίες ενδεχομένως να μην χρησιμοποιούνται αρκετές φορές για να εκπαιδευτεί αποδοτικά πάνω τους το μοντέλο. Στο DETERRENT, υπάρχουν μόλις δεκατριές σχέσεις οι οποίες χρησιμοποιούνται από εκατοντάδες, έως χιλιάδες φορές, δίνοντας τη δυνατότητα στο μοντέλο να εκπαιδευτεί αποδοτικά σε αυτές.

- Το είδος των σχέσεων, όπου στο DETTERENT έχουμε θετικές και αρνητικές σχέσεις π.χ. heal, noheal, οι οποίες είναι πολύ συγκεκριμένες και σημασιολογικά είναι πολύ εύκολο να συσχετιστούν. Στο δικό μας σύνολο δεδομένων δεν έχουμε συγκεκριμένο τύπο σχέσεων, το οποίο σαφέστατα μπορεί να επηρεάζει την ποιότητα των αποτελεσμάτων.

- Το είδος των δεδομένων, καθώς στο δικό μας σύνολο δεδομένων δεν έχουμε κάποια συγκεκριμένη θεματολογία, ενώ στο DETERRENT, έχουμε ειδήσεις που αφορούν αποκλειστικά ιατρικά θέματα και έχουν συγκεκριμένη θεματολογία, το οποίο σημασιολογικά μπορεί να αντιληφθεί τα παρόμοια θέματα που μελετώνται, γεγονός που και πάλι συμβάλλει στην πιο εύκολη εκμάθηση των δεδομένων.

- Η ποσότητα της γνώσης που παρέχεται, καθώς ο ενδεχόμενος συνδυασμός με μια εξωτερική βάση γνώσης, ίσως να αναπαριστούσε καλύτερα το περιεχόμενο της κάθε είδησης, παίρνοντας όλες τις πληροφορίες που αφορούν κάθε οντότητα.

- Η ποσότητα των ειδήσεων, καθώς στο δικό μας σύνολο δεδομένων έχουμε μόλις πάνω από δέκα χιλιάδες ειδήσεις, ενώ στο DETERRENT υπάρχουν γύρω στις μισές.

- Η γλώσσα των ειδήσεων, καθώς στο DETERRENT οι ειδήσεις είναι στην αγγλική γλώσσα, η οποία διαθέτει αρκετά εργαλεία δημιουργίας triplets, αποσαφήνισης triplets, ενώ στη δική μας προσέγγιση οι ειδήσεις είναι στην ελληνική γλώσσα, στην οποία τα εργαλεία που υπάρχουν είναι μηδενικά και η διαδικασία δεν είναι αυτοματοποιημένη, με ορατό το ενδεχόμενο σφαλμάτων στα δεδομένων και την αρνητική επιρροή στην ποιότητά τους.

Τέλος, πριν ολοκληρωθεί αυτή η διπλωματική εργασία, είναι σημαντικό να προταθούν κάποιες βελτιώσεις που μπορούν να γίνουν μελλοντικά στον εντοπισμό ψευδών ειδήσεων στην ελληνική γλώσσα. Αρχικά ως προς τον εντοπισμό των ειδήσεων αποκλειστικά μέσω του περιεχομένου, αλλά και με βάση τα συμφραζόμενα, όπως έχει παρατηρηθεί από τη βιβλιογραφία παράγονται καλύτερα αποτελέσματα όταν συνδυάζονται οι δύο αυτές μέθοδοι. Όταν χρησιμοποιείται ένα νευρωνικό δίκτυο με ελάχιστα επίπεδα, σχεδιασμένα αποκλειστικά για το θέμα που μελετάται, ώστε να γίνεται σωστά η ταξινόμηση των ειδήσεων, παρατηρείται επίσης βελτίωση της απόδοσης, καθώς όπως προαναφέρθηκε και στο [60] είναι επίσης πολύ σημαντική η δομή του νευρωνικού δικτύου εάν χρησιμοποιείται. Άρα, θεμιτός θα ήταν ο συνδυασμός ενός απλού νευρωνικού που χρησιμοποιεί BiGRU, με μεθόδους που χρησιμοποιούν αποκλειστικά το περιεχόμενο όπως word2vec ή fasttext. Όσο αφορά τις μεθόδους που βασίζονται αποκλειστικά στο περιεχόμενο, όπως εξάχθηκε και από τη βιβλιογραφία, η χρήση ενός συστήματος ψήφισης, μπορεί να συμβάλει στην δραστηρίως στη βελτίωση της απόδοσης. Πιθανές επεκτάσεις που θα μπορούσαν να γίνουν σε όλους τους τομείς, θα είναι η εισαγωγή πολύ-τροπικών στοιχείων όπως εικόνες και βίντεο, καθώς και η ενδεχόμενη χρήση τους στην δημιουργία γράφων, με σκοπό τη δημιουργία ενσωματώσεων που θα αναπαριστούν ακόμα καλύτερα τη σημασιολογία της κάθε είδησης ή ακόμα και την αξιοποίηση περισσότερων δεδομένων. Όσον αφορά τον εντοπισμό ψευδών ειδήσεων με βάση τη γνώση, τα μοντέλα που υπάρχουν, ακόμα δεν φτάνουν σε απόδοση τα μοντέλα που χρησιμοποιούν τα συμφραζόμενα – τουλάχιστον ως εκ των πλείστων – καθώς είναι ένας σχετικά καινούργιος τομέας, ως προς τον εντοπισμό ψευδών ειδήσεων. Κάποιες βελτιώσεις που προτείνονται και αφορούν άμεσα τον εντοπισμό ως προς τη γνώση είναι: Το σύνολο δεδομένων να αποτελείται από δεδομένα συγκεκριμένης θεματολογίας (π.χ. αποκλειστικά πολιτικές ειδήσεις, αποκλειστικά ειδήσεις για την οικονομία), με σκοπό την πιο αποδοτική εκπαίδευση και την πιο εύκολη εκμάθηση την θεματολογίας των δεδομένων, κάτι το οποίο όπως είδαμε στο [45] και στο [46], βοηθά αρκετά στην απόδοση του μοντέλου. Η εύρεση πιο αποδοτικού τρόπου σχηματισμού triplets, με σκοπό τη διασφάλιση της ποιότητας των δεδομένων εκπαίδευσης, που θα βοηθήσει άμεσα στην εύρεση ειδήσεων με βάση τη γνώση και την πιο αποδοτική εκμάθηση των δεδομένων. Έπειτα, χρήσιμη θα ήταν και η τροποποίηση της εκπαίδευσης του μοντέλου, ώστε να μην χρειάζεται την ετικέτα για πρόβλεψη του κατά πόσο μια είδηση είναι ψευδής ή αληθής και να περάσουμε από εποπτευόμενη μάθηση, στην μη εποπτευόμενη και να γίνει πιο εύκολη η εφαρμογή του μοντέλου σε πιο περίπλοκα προβλήματα. Επίσης, η ενδεχόμενη χρήση συνωνύμων, με σκοπό τη μείωση του πλήθους των σχέσεων στα triplets, θα συνείφερε στην πιο εύκολη εκμάθηση των δεδομένων, η οποία κατ' επέκταση θα μας απέφερε καλύτερα αποτελέσματα. Τέλος, σημαντική θα ήταν και εξέλιξη του μοντέλου ώστε να ταξινομεί ολόκληρους γράφους και όχι μόνο τους κόμβους τους, βελτίωση η οποία μπορεί να συνδράμει τόσο στην υλοποίηση ταξινόμησης ειδήσεων με βάση τη διάδοση, όσο και με βάση τη γνώση.

9. Βιβλιογραφία

- [1] Infocom security – 6 & 7 Ιουλίου 2022. From <https://www.infocomsecurity.gr/presentations/2019/day2/4th-section.pdf>
- [2] Wikimedia Foundation. (2023, January 29). Fake news. Wikipedia. Retrieved February 3, 2023, from https://en.wikipedia.org/wiki/Fake_news
- [3] Fake news detection in social media – California State University. From https://csustan.edu/sites/default/files/groups/University%20Honors%20Program/Journals/02_stahl.pdf
- [4] (Group), R. S. A. (O. P. E. N. D. (2022, June 16). Στην τελευταία θέση τα ελληνικά ΜΜΕ ως προς την αίσθηση ανεξαρτησίας - τι δείχνει έρευνα. ΕΘΝΟΣ <https://www.ethnos.gr/greece/article/212370/sthnteleytaiatheshtaellhnikammeosprosthnaisthshshanexarthsiasideixneieryna>
- [5] Μεθοδολογία ελέγχου και επεξήγηση χαρακτηρισμών. ELLINIKA HOAXES. (2020, August 18) from <https://www.ellinikahoaxes.gr/transparency-and-methodology/>
- [6] Τι είναι η μηχανική μάθηση; Ορισμός, τύποι και παραδείγματα: SAP insights. SAP from <https://www.sap.com/greece/insights/what-is-machine-learning.html>
- [7] Wikimedia Foundation. (2022, August 5). Μηχανική μάθηση. Wikipedia from https://el.wikipedia.org/wiki/%CE%9C%CE%B7%CF%87%CE%B1%CE%BD%CE%B9%CE%BA%CE%AE_%CE%BC%CE%AC%CE%B8%CE%B7%CF%83%CE%B7
- [8] Postprocessing in machine learning and Data Mining from https://kdd.org/exploration_files/KD/D2000PostWkshp.pdf
- [9] Lavrač, N., Grobelnik, M. (2003). Data Mining. In: Mladenić, D., Lavrač, N., Bohanec, M., Moyle, S. (eds) Data Mining and Decision Support. The Springer International Series in Engineering and Computer Science, vol 745. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-0286-9_1
- [10] Data quality and systems theory – ACM Digital Library from <https://dl.acm.org/doi/pdf/10.1145/269012.269023>
- [11] Bertachas, I., Tzanakis, G., Michelaki, P., Pavlakis, K., Μπερταχάς, Ι., Τζανάκης, Γ., Μιχελάκη, Π. & Παυλάκης, Κ. (2016, January 19). Electronic Book Laboratory exercises in physics I. Kallipos. From <https://repository.kallipos.gr/handle/11419/1095>
- [12] Computer Science Department, University of Cyprus from <http://www.cs.ucy.ac.cy/~panic/teaching/2013S.EPL231.website/lectures/2013S.EPL231.18.graphs.pdf>

- [13] Sanchez-Lengeling, B., Reif, E., Pearce, A., & Wiltchko, A. B. (2021, September 8). A gentle introduction to graph neural networks. Distill from <https://distill.pub/2021/gnn-intro/>
- [14] Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018, February 4). Graph attention networks. arXiv.org from <https://arxiv.org/abs/1710.10903>
- [15] Rossi, A., Tiezzi, M., Dimitri, G.M., Bianchini, M., Maggini, M., Scarselli, F. (2018). Inductive–Transductive Learning with Graph Neural Networks. In: Pancioni, L., Schwenker, F., Trentin, E. (eds) Artificial Neural Networks in Pattern Recognition. ANNPR 2018. Lecture Notes in Computer Science (), vol 11081. Springer
- [16] Singhal, A. (2012, May 16). Introducing the knowledge graph: Things, not strings. Google from <https://blog.google/products/search/introducing-knowledge-graph-things-not/>
- [17] Piotr Bojanowski, Edouard Grave, Armand Joulin, Tomas Mikolov. 2016. Enriching Word Vectors with Subword Information. arXiv preprint arXiv:1607.0466 (2016)
- [18] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013).
- [19] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In Empirical Methods in Natural Language Processing (EMNLP). 1532–1543. <http://www.aclweb.org/anthology/D14-1162>
- [20] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759 (2016).
- [21] V. Agarwal, H. P. Sultana, S. Malhotra, and A. Sarkar, "Analysis of classifiers for fake news detection," Proc. Compute. Sci., vol. 165, pp.377–383, Jan. 2019.
- [22] A. B. Presidio, R. R. Isnanto, D. Eridian, Y. A. A. Soetrisno, M. Arfan, and A. Sofwan, "Hoax detection system on Indonesian news sites-based text classification using SVM and SGD," in Proc. 4th Int. Conf. Inf. Technol., Compute., Electra. Eng. (ICITACEE), Oct. 2017, pp. 45–49
- [23] Patil, Dharmaraj R. "Fake News Detection Using Majority Voting Technique." arXiv preprint arXiv:2203.09936 (2022).
- [24] Kaur S, Kumar P, Kumaraguru P (2020) Automating fake news detection system using multi-level voting model. Soft Comput 24:9049–9069
- [25] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations

- using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014).
- [26] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [27] Ramiah, M. (2021, January 10). Bidirectional RNN & Basics of LSTM and gru. Medium from <https://medium.com/analytics-vidhya/bi-directional-rnn-basics-of-lstm-and-gru-e114aa4779bb>
- [28] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Duajaili, A., Duan, Y., Al-Shamma, O., Santamaria, J., Fadhel, M. A., Al-Amidie, M., & Farhan L. (2021, March 31). Review of Deep Learning Concepts, CNN Architectures, challenges, applications, future directions – *Journal of Big Data*. SpringerOpen from <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8>
- [29] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017, December 6). Attention is all you need. arXiv.org from <https://arxiv.org/abs/1706.03762>
- [30] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, May 24). Bert: Pre-training of deep bidirectional Transformers for language understanding. arXiv.org from <https://arxiv.org/abs/1810.04805>
- [31] Santiago Gonzalez-Carvajal, Eduardo C. Garrido-Merchan. Comparing BERT against traditional machine learning text classification. arXiv preprint arXiv: 2005:13012 (2020)
- [32] Kaliyar, R.K., Goswami, A. & Narang, P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimed Tools Appl* 80, 11765–11788 (2021). <https://doi.org/10.1007/s11042-020-10183-2>
- [33] Truică, C.-O.; Apostol, E.-S. MisRoBÆRTa: Transformers versus Misinformation. *Mathematics* 2022, 10, 569. <https://doi.org/10.3390/math10040569>
- [34] Galende, B.A.; Hernández-Peñaloza, G.; Uribe, S.; García, F.Á. Conspiracy or Not? A Deep Learning Approach to Spot It on Twitter. *IEEE Access* 2022
- [35] University, K. S. A. S., Shu, K., University, A. S., University, L. C. P. S., Cui, L., University, P. S., University, S. W. P. S., Wang, S., University, D. L. P. S., Lee, D., University, H. L. A. S., Liu, H., KenSci, Minnesota, U. of, Corporation, E. V. A., LinkedIn, University, B., & Metrics, O. M. V. A. (2019, July 1). Defend: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining from <https://dl.acm.org/doi/10.1145/3292500.3330935>
- [36] Vo, N., & Lee, K. Hierarchical multi-head Attentive Network for evidence-aware fake news detection. ACL Anthology from <https://aclanthology.org/2021.eacl-main.83/>

- [37] Saleh H, Alharbi A, Alsamhi SH (2021) OPCNN-FAKE: optimized convolutional neural network for fake news detection. *IEEE Access* 9:129471–129489
- [38] G. Gadek and P. Guélorget, "An interpretable model to measure fakeness and emotion in news," *Procedia Computer Science*, Volume 176, 2020. Graph Neural Networks with Continual Learning for Fake News Detection from social media [2020]
- [39] Ni S, Li J, Kao H-Y (2021) MVAN: multi-view attention networks for fake news detection on social media. *IEEE Access* 9:106907–106917
- [40] Lu, Y.-J., & Li, C.-T. GCAN: Graph-aware co-attention networks for explainable fake news detection on social media. *ACL Anthology* from <https://aclanthology.org/2020.acl-main.48/>
- [41] Weizhi Xu, Junfei Wu, Qiang Liu, Shu Wu, and Liang Wang. 2022. Evidence-aware Fake News Detection with Graph Neural Networks. In *Proceedings of the ACM Web Conference 2022 (WWW '22)*. Association for Computing Machinery, New York, NY, USA, 2501–2510. <https://doi.org/10.1145/3485447.3512122>
- [42] Incorporating Relational Knowledge in Explainable Fake News Detection – Yue Ning (2021) from <https://yue-ning.github.io/docs/pakdd2021.pdf>
- [43] Hu, L., Yang, T., Zhang, L., Zhong, W., Tang, D., Shi, C., Duan, N., & Zhou, M. Compare to the knowledge: Graph neural fake news detection with external knowledge. *ACL Anthology* from <https://aclanthology.org/2021.acl-long.62/>
- [44] DEAP-faked: Knowledge graph-based approach for fake news detection (2022) from <https://arxiv.org/pdf/2107.10648.pdf>
- [45] Limeng Cui, Haeseung Seo, Cui, L., Seo, Maryam Tabar, Fenglong Ma, Suhang Wang, Dongwon Lee: The Pennsylvania State University, Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery* from <https://dl.acm.org/doi/10.1145/3394486.3403092>
- [46] Learning to update knowledge graphs by reading news - *ACL anthology* from <https://aclanthology.org/D19-1265.pdf>
- [47] Y. Wang, S. Qian, J. Hu, Q. Fang and C. Xu, "Fake News Detection via Knowledge-driven Multimodal Graph Convolutional Networks," *ICMR '20: Proceedings of the 2020 International Conference on Multimedia Retrieval*, June 2020.
- [48] Ziyi Kou, Lanyu Shang, Dong Wang: University of Illinois at Urbana-Champaign, Yang Zhang University of Notre Dame, O. M. V. A. (2022, January 1): HC-COVID: A hierarchical crowdsourced knowledge graph approach to explainable COVID-19 misinformation detection: *Proceedings of the ACM on Human-Computer Interaction: Vol 6, No Group. Proceedings of the ACM on Human-Computer Interaction* from <https://dl.acm.org/doi/10.1145/3492855>

- [49] Shengsheng Qian, Jun Hu, Quan Fang, Xu, C., Firenze, Changsheng, National Lab of Pattern Recognition, O. M. V. A. (2021, August 1): Knowledge-aware multi-modal adaptive graph convolutional networks for fake news detection. ACM Transactions on Multimedia Computing, Communications, and Applications. from <https://dl.acm.org/doi/10.1145/3451215>
- [50] Boshko Koloski, Timen Stepišnik-Perdih, Marko Robnik-Šikonja, Senja Pollak, Blaž Škrlić: Knowledge graph informed fake news classification via heterogeneous representation ensembles from <https://arxiv.org/pdf/2110.10457.pdf>
- [51] Naveen: What is precision, recall, accuracy and F1-score? Nomidl, from <https://www.nomidl.com/machine-learning/what-is-precision-recall-accuracy-and-f1-score/>
- [52] Word representations fasttext. FastText from <https://fasttext.cc/docs/en/unsupervised-tutorial.html>
- [53] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In Proc. of NAACL.
- [54] Liu, F. & Lu, X. (2011). Survey on text clustering algorithm. In Proceedings of 2nd International IEEE Conference on Software Engineering and Services Science (ICSESS), China, 901-904
- [55] Classify text with Bert: text: tensorflow. TensorFlow from https://www.tensorflow.org/text/tutorials/classify_text_with_bert
- [56] Chung, J., Gulcehre, C., Cho, K. H., & Bengio, Y. (2014, December 11). Empirical evaluation of gated recurrent neural networks on sequence modelling. arXiv.org. from <https://arxiv.org/abs/1412.3555>
- [57] Kipf, T. N., & Welling, M. (2017, February 22). Semi-supervised classification with graph Convolutional Networks. arXiv.org from <https://arxiv.org/abs/1609.02907>
- [58] Graph attention networks from <https://petar-v.com/GAT/>
- [59] Relational Graph Convolutional Networks: A Closer Look from <https://arxiv.org/pdf/2107.10015.pdf>
- [60] The surprising performance of simple baselines for misinformation detection from https://www.researchgate.net/publication/350875921_The_Surprising_Performance_of_Simple_Baselines_for_Misinformation_Detection