



ΔΙΕΘΝΕΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΗΣ ΕΛΛΑΔΟΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ
ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
«Ανίχνευση Ψευδών Ειδήσεων στα
Ελληνικά»



Της φοιτήτριας
Μακρυγιώτης Μαρίας
Αρ. Μητρώου: 516079

Επιβλέπων
Ονοματεπώνυμο Τζέκης
Παναγιώτης
Βαθμίδα Καθηγητής

Τίτλος Δ.Ε. "Έλεγχος ψευδών ειδήσεων στην Ελληνική γλώσσα μέσω λογισμικού"

Κωδικός Δ.Ε. 22122

Όνοματεπώνυμο φοιτητή/ών Μακρυνιώτη Μαρία

Όνοματεπώνυμο εισηγητή Παναγιώτης Τζέκης

Ημερομηνία ανάληψης Δ.Ε. 02-03-2022

Ημερομηνία περάτωσης Δ.Ε. 17-01-2023

Βεβαιώνω ότι είμαι ο συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως διπλωματική εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία της φοιτήτριας Μακρυνιώτης Μαρίας που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης, ο συγγραφέας/δημιουργός εκχωρεί στο Διεθνές Πανεπιστήμιο της Ελλάδος άδεια χρήσης του δικαιώματος αναπαραγωγής, δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσης της εργασίας διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος. Η ανοικτή πρόσβαση στο πλήρες κείμενο της εργασίας, δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού, ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, πώληση, εμπορική χρήση, διανομή, έκδοση, μεταφόρτωση (downloading), ανάρτηση (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος, δεν υποδηλώνει απαραίτητα και αποδοχή των απόψεων του συγγραφέα, εκ μέρους του Τμήματος.

Πρόλογος

Το πεδίο της εργασίας επιλέχθηκε καθώς ήθελα με κάποιο τρόπο να αποκτήσω γνώσεις πάνω στο κομμάτι του λογισμικού. Αυτή η ανάγκη προέκυψε διότι έχοντας φοιτήσει στο τμήμα της Ηλεκτρονικής και στοχεύοντας στο νέο πτυχίο της συνένωσης με το τμήμα της Πληροφορικής, ήθελα να έχω μια ευρεία γκάμα γνώσεων που θα αντικατοπτρίζουν τον τίτλο του πτυχίου. Το θέμα προέκυψε αμιγώς από την δική μου ανησυχία πάνω στο φαινόμενο των ψευδών ειδήσεων, το οποίο αυξήθηκε δραματικά την περίοδο του κορονοϊού. Ήθελα το θέμα να ασχολείται με κάτι που με ενδιαφέρει άμεσα και σαν μηχανικός αλλά και σαν προσωπικότητα.

Περίληψη

Η παρούσα εργασία παρουσιάζει μια λεπτομερή μελέτη του προβλήματος της ανίχνευσης ψευδών ειδήσεων στην ελληνική γλώσσα. Η εξάπλωση των ψευδών πληροφοριών στο διαδίκτυο έχει γίνει μείζον θέμα τα τελευταία χρόνια, καθώς μπορεί να έχει σημαντικό αντίκτυπο στην κοινή γνώμη και στη λήψη αποφάσεων. Η ικανότητα αυτόματης ανίχνευσης ψευδών ειδήσεων μπορεί να συμβάλει στον μετριασμό της εξάπλωσης της παραπληροφόρησης.

Αναπτύχθηκε ένα νευρωνικό δίκτυο για δυαδική ταξινόμηση με στόχο την ανίχνευση ψευδών ειδήσεων στα ελληνικά. Το μοντέλο χρησιμοποιεί τεχνικές επεξεργασίας φυσικής γλώσσας και αλγορίθμους μηχανικής μάθησης για την ανάλυση δεδομένων κειμένου. Δημιουργήθηκαν δύο νέα σύνολα δεδομένων με σκοπό την αξιολόγηση της απόδοσης του προτεινόμενου αλγορίθμου. Το πρώτο σύνολο δεδομένων δημιουργήθηκε με τη μετάφραση ενός υπάρχοντος αγγλικού συνόλου δεδομένων, ενώ το δεύτερο σύνολο δεδομένων δημιουργήθηκε με τη χρήση τεχνικών απόξεσης ιστού για τη συγκέντρωση μεγάλου αριθμού ελληνικών ειδησεογραφικών άρθρων από πηγές.

Εκτός από την αξιολόγηση της απόδοσης του προτεινόμενου αλγορίθμου, δοκιμάστηκαν επίσης διάφοροι υπάρχοντες αλγόριθμοι ταξινόμησης χρησιμοποιώντας τα νέα σύνολα δεδομένων. Τα αποτελέσματα των πειραμάτων καταδεικνύουν ότι ο προτεινόμενος αλγόριθμος είναι σε θέση να επιτύχει ικανοποιητική ακρίβεια και βαθμολογία fscore στον εντοπισμό ψευδών ειδήσεων στην ελληνική γλώσσα.

Στη εργασία συζητούνται επίσης οι περιορισμοί του προτεινόμενου αλγορίθμου και προτείνονται πιθανές κατευθύνσεις για μελλοντική έρευνα στον τομέα αυτό. Συνολικά, η παρούσα έρευνα συμβάλλει στον τομέα της ανίχνευσης ψευδών ειδήσεων, καθώς παρέχει μια ολοκληρωμένη μελέτη του προβλήματος στην ελληνική γλώσσα.

«Fake News Detection In Greek Language»

«Maria Makrynioti»

Abstract

This thesis presents a detailed study of the problem of false news detection in the Greek language. The spread of false information on the Internet has become a major issue in recent years, as it can have a significant impact on public opinion and decision making. The ability to automatically detect fake news can help to mitigate the spread of misinformation.

A neural network for binary classification was developed to detect fake news in Greek. The model uses natural language processing techniques and machine learning algorithms to analyse textual data. Two new datasets were created to evaluate the performance of the proposed algorithm. The first dataset was created by translating an existing English dataset, while the second dataset was created by using web scraping techniques to collect a large number of Greek news articles from sources.

In addition to evaluating the performance of the proposed algorithm, several existing classification algorithms were also tested using the new datasets. The experimental results demonstrate that the proposed algorithm is able to achieve satisfactory accuracy and F1 score in detecting fake news in Greek.

The thesis also discusses the limitations of the proposed algorithm and suggests possible directions for future research in this area. Overall, this research contributes to the field of fake news detection by providing a comprehensive study of the problem in the Greek language.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή κ. Παναγιώτη Τζέκη που μου έδωσε την δυνατότητα να αναλάβω αυτή τη διπλωματική και να ασχοληθώ με αυτό το πολύ ενδιαφέρον αντικείμενο. Σε προσωπικό επίπεδο θα ήθελα να ευχαριστήσω την οικογένεια και τους φίλους μου για την αμέριστη στήριξή τους.

Περιεχόμενα

Πρόλογος	ii
Περίληψη	iii
Abstract	iv
Ευχαριστίες	v
Περιεχόμενα	vi
Κατάλογος Σχημάτων	viii
Κατάλογος Πινάκων	ix
Συντομογραφίες	x
1 Εισαγωγή	1
1.1 Εισαγωγή	1
1.2 Πρόλογος	1
1.3 Πρόβλημα	1
1.4 Σκοπός Διπλωματικής	2
1.5 Δομή	3
1.6 Επίλογος	3
2 Βιβλιογραφική Αναφορά	4
2.1 Εισαγωγή	4
2.2 Βιβλιογραφική Αναφορά	4
2.3 Επίλογος	5
3 Θεωρητικό Υπόβαθρο	6
3.1 Εισαγωγή	6
3.2 Μάθηση Με Επίβλεψη	6
3.3 Ταξινόμηση	8
3.3.1 Γενικά	8
3.3.2 Ταξινομητές	8
3.4 Νευρωνικά Δίκτυα	11
3.4.1 Γενικά	11
3.4.2 Μοντέλο Τεχνητού Νευρώνα	11
3.5 Τεχνητά Νευρωνικά Δίκτυα Πολλών Επιπέδων (Multi-Layer Perceptron MLP)	12
3.5.1 Βαθύ Νευρωνικό Δίκτυο (Deep Neural Network DNN)	15
3.5.2 Πλήρως Συνδεδεμένο Στρώμα	17
3.5.3 Κανονικοποίηση και Dropout	18
3.6 Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing NLP)	19
3.6.1 Γενικά	19
3.7 Επίλογος	21
4 Τεχνολογικό Υπόβαθρο	22
4.1 Εισαγωγή	22
4.2 Python	22
4.3 PyTorch	23
4.4 Scikit Learn	23
4.5 Beautiful Soup	23
4.6 DeepL	24
4.7 NLTK	24
4.8 FastAPI	24
4.9 Unicorn	25
4.10 Επίλογος	25

5	Σύνολο Δεδομένων	26
5.1	Εισαγωγή	26
5.2	Γενικά	26
5.3	Διαδικασία Απόξεσης Ιστού	26
5.4	Ελληνικό Σύνολο Δεδομένων	27
5.4.1	Greek Hoaxes	27
5.4.2	Η Καθημερινή	28
5.5	Μεταφρασμένο Σύνολο Δεδομένων	28
5.5.1	Liar Dataset	28
5.5.2	Μέθοδος Μετάφρασης	29
5.6	Επίλογος	30
6	Μεθοδολογία	31
6.1	Εισαγωγή	31
6.2	Δημιουργία Μοντέλου	31
6.3	Διαδικτυακή Εφαρμογή	36
6.4	Επίλογος	38
7	Αποτελέσματα	39
7.1	Εισαγωγή	39
7.2	Κριτήρια	39
7.3	Αποτελέσματα Ταξινόμησης	42
7.4	Αποτελέσματα Μοντέλου	45
8	Συμπεράσματα	47
	ΒΙΒΛΙΟΓΡΑΦΙΑ	48

Κατάλογος Σχημάτων

3.1	Υποκατηγορίες Μηχανικής Μάθησης	7
3.2	Random Forest	9
3.3	Υπερ-επιφάνεια Μέγιστου Περιθωρίου	10
3.4	Μοντέλο Τεχνητού Νευρώνα	12
3.5	Νευρωνικό Δίκτυο Πολλών Επιπέδων	13
3.6	Σιγμοειδής Συνάρτηση, Συνάρτηση ReLu	14
3.7	Dropout	19
3.8	Μοντέλο GreekBERT	21
5.1	Ιστοσελίδα Greek Hoaxes	28
6.1	Διάγραμμα Οπτικοποίησης Μοντέλου	36
6.2	Πρόβλεψη ψευδής είδησης	37
6.3	Πρόβλεψη αληθούς είδησης	37
7.1	Πίνακας Σύγκρισης	39
7.2	Περιοχή κάτω την καμπύλη ROC	41
7.3	Greek: Passive Aggressive	44
7.4	Greek: Random Forest	44
7.5	Greek: SVM	44
7.6	Liar: Passive Aggressive	44
7.7	Liar: Random Forest	44
7.8	Liar: SVM	44
7.9	Mixed: Passive Aggressive	44
7.10	Mixed: Random Forest	44
7.11	Mixed: SVM	44
7.12	Πίνακας Σύγκρισης Μοντέλου	46
7.13	Roc Curve	46

Κατάλογος Πινάκων

7.1	Σύγκριση ταξινομητών	42
7.2	Παράμετροι Μοντέλου	45
7.3	Μετρικές Μοντέλου	45

Συνομογραφίες

Δ.Ε.	Διπλωματική Εργασία
ΔΙΠΙΑΕ	Διεθνές Πανεπιστήμιο Ελλάδος
CNN	Convolutional Neural Network
URL	Uniform Resource Locator
CART	Classification and Regression Tree
SVM	Support Vector Machine
PA	Passive Aggressive
TNΔ	Τεχνητά Νευρωνικά Δίκτυα
MLP	Multi-Layered Perceptrons
ANN	Artificial Neural Network
ReLU	Rectified Linear Unit
DNN	Deep Neural Network
NLP	Natural Language Processing
BERT	Bidirectional Encoder Representations from Transformers
MLM	Masked Language Modeling
API	Application Programming Interface
NN	Neural Network
HTML	Hyper-Text Markup Language
XML	Extensible Markup Language
JSON	JavaScript Object Notation
ROC	Receiver Operating Characteristics
TF-IDF	Term Frequency-Inverse Document Frequency

Κεφάλαιο 1ο: Εισαγωγή

1.1 Εισαγωγή

Σε αυτό το κεφάλαιο θα γίνει μια εισαγωγική αναφορά στο θέμα της διπλωματικής, στον στόχο που έχει τεθεί να βγει εις πέρας, και τη η δομή που ακολουθεί αυτή η αναφορά.

1.2 Πρόλογος

Η έμφυτη ανάγκη του ανθρώπου για σωστή ενημέρωση στην σημερινή κοινωνία είναι δύσκολο να καλυφθεί για ποικίλους λόγους. Όπως γράφει και η Δρ. Φιλοσοφίας Μάγδα Ρουμανέα –Μαλάμη σε άρθρο της "Η πληροφορία αλλάζει τον τρόπο που «βλέπουμε» τον κόσμο, τους συνανθρώπους μας, την κοινωνία και τον εαυτό μας. Η πληροφορία περιορίζει την αβεβαιότητα που έχει ο άνθρωπος για γεγονότα και καταστάσεις που τον απασχολούν και τον βοηθούν να πάρει τις σωστές αποφάσεις και να αντιμετωπίζει τα διάφορα προβλήματα που τον απασχολούν." [1]

Την τελευταία δεκαετία, η εξέλιξη της τεχνητής νοημοσύνης στον κόσμο της τεχνολογίας έφερε σημαντικές αλλαγές στον τρόπο που ζούμε. Μερικά από τα πιο εμφανή επιτεύγματα είναι η ανάπτυξη ευφών συστημάτων αναζήτησης, η χρήση εικονικών προσωπικών βοηθών, η ρομποτική χειρουργική και ίσως αυτό με την μεγαλύτερη επιρροή στην ζωή μας, η επεξεργασία της φυσικής γλώσσας [2].

1.3 Πρόβλημα

Στη σημερινή εποχή, με την συμβολή του διαδικτύου, η ποιότητα και η ποσότητα των ειδήσεων αυξάνεται μέρα με τη μέρα. Ταυτόχρονα αλλάζει συνεχώς και ο τρόπος με τον οποίο ο καταναλωτής έχει πρόσβαση και διαχείριση των διαδικτυακών πληροφοριών. Μερικοί χρήστες, ιδιαίτερα οι άπειροι, χρησιμοποιούν εφαρμογές, δυναμικές ιστοσελίδες και πλατφόρμες κοινωνικής δικτύωσης για να αντλήσουν τις απαραίτητες πληροφορίες γρήγορα και εύκολα, συχνά χωρίς διάκριση. Απόψεις, νοοτροπίες ή ακόμα και η συμμετοχή σε δημοκρατικές εκλογές μπορούν να επηρεαστούν ή ακόμα και να αλλάξουν με την ποικιλομορφία των ειδήσεων που κυκλοφορούν. Ωστόσο, η τεχνολογία εκτός από αγαθό, μπορεί να χρησιμοποιηθεί και με δόλο, να διαδώσει όγκο ψευδών πληροφοριών μέσω των κοινωνικών δικτύων και να γίνει ένα αποτελεσματικό όργανο παραπληροφόρησης. Ορισμένοι ερευνητές [3], προειδοποιούν πως μερικοί πολιτικοί, σφετεριζόμενοι την εξουσία που διαθέτουν, χρησιμοποιούν τις ψευδείς ειδήσεις για προσωπικό ή κομματικό όφελος. Από τους τομείς που έχουν δεχτεί μεγαλύτερο πλήγμα, είναι αυτή της επιστήμης και της δημοσιογραφίας, διότι πολλές ψευδείς ειδήσεις εστιάζουν στην αμφισβήτηση πηγών που στο παρελθόν θεωρούνταν αξιόπιστες. Συνεπώς, οι αξιόπιστες και αξιόλογες πηγές πληροφόρησης είναι απαραίτητες, ώστε το κοινό να μην πέσει θύμα κακόβουλων προθέσεων.

Σε αυτό σημαντικός παράγοντας είναι αλγόριθμοι και πλατφόρμες που εντοπίζουν ψευδείς ειδήσεις και προβλέπουν σε πραγματικό χρόνο την ορθότητα μιας ειδήσης.

Με μια γρήγορη πλοήγηση στο διαδίκτυο, μπορούν εύκολα να βρεθούν υπηρεσίες είτε δωρεάν κυρίως όμως επί πληρωμή, οι οποίες προσφέρουν προγράμματα για την αντιμετώπιση του προβλήματος της διάδοσης ψευδών ειδήσεων στην Αγγλική γλώσσα. Εάν εξεταστεί το ίδιο θέμα στην ελληνική πραγματικότητα, παρατηρείται ότι το κομμάτι της επεξεργασίας της φυσικής γλώσσας, βελτιώνεται καθημερινά. Παρόλα αυτά υπάρχει μεγάλη έλλειψη σε ελληνικά σύνολα δεδομένων που θα μπορούσαν να προσεγγιστούν προβλήματα όπως ο εντοπισμός των ψευδών ειδήσεων. Αυτό συμβαίνει κυρίως γιατί δεν υπάρχουν έμπιστες πηγές από τις οποίες θα αντληθούν αυτά τα δεδομένα μαζικά ώστε να εκπαιδευτούν τα μοντέλα. Επομένως, το πρόβλημα εντοπισμού και πρόβλεψης ψευδών ειδήσεων στην Ελλάδα είναι επίκαιρο και δεν έχει επιλυθεί.

Για την κατανόηση της εργασίας θα πρέπει να καταστεί ξεκάθαρος ο όρος ψευδής είδησης και οι μορφές της: Η EAVI (European Association for Viewers Interests) [4] [4] διαχωρίζει τις παραπλανητικές ειδήσεις σε 10 κατηγορίες:

- Προπαγάνδα
- Δόλωμα Κλικ - Clickbait
- Χορηγούμενο περιεχόμενο
- Σάτιρα και φάρσα
- Σφάλμα
- Κομματικό
- Θεωρία συνωμοσίας
- Ψευδοεπιστήμη
- Παραπληροφόρηση
- Ψεύτικη

Υπάρχουν επίσης 4 συμπληρωματικές κατηγορίες:

- Λανθασμένη απόδοση
- Πλαστογραφία
- Παραπλάνηση
- Παραποιημένο περιεχόμενο

1.4 Σκοπός Διπλωματικής

Ο πρώτος στόχος της συγκεκριμένης εργασίας είναι η δημιουργία ενός μοντέλου μηχανικής μάθησης το οποίο θα εκπαιδεύεται στην ελληνική γλώσσα και θα προβλέπει εάν ειδησεογραφικοί τίτλοι ανήκουν σε ψευδές ή σε αληθές φάσμα.

Ο δεύτερος στόχος είναι η δημιουργία ενός αντιπροσωπευτικού συνόλου δεδομένων το οποίο θα μπορεί να χρησιμοποιηθεί στο προαναφερθέν μοντέλο. Αυτό προϋποθέτει την συλλογή ειδήσεων γραμμένων στην ελληνική γλώσσα που είτε θα προέρχονται από μεταφρασμένες ξένες ειδήσεις είτε από καθαρά ελληνικές ειδήσεις προερχόμενες από ελληνικές πηγές.

Ο τρίτος στόχος είναι μια σύντομη μελέτη κατά την οποία γίνεται πρακτικός έλεγχος του μοντέλου που δημιουργήθηκε και εφάμιλλων αλγορίθμων που προσπαθούν να πετύχουν τον ίδιο στόχο.

Θα πρέπει να καταστεί σαφές πως το μοντέλο έχει ως μόνη είσοδο τον τίτλο, και η κατηγοριοποίηση έγκειται καθαρά στην δομή της πρότασης. Ο έλεγχος των στοιχείων (fact-checking), ένα προς ένα, είναι μια διαφορετική προσέγγιση για την καταπολέμηση των ψευδών ειδήσεων που δεν έχει επιλεχθεί στην συγκεκριμένη διπλωματική. Για την καταληκτική επιλογή του ιδανικότερου μοντέλου, είναι προαπαιτούμενη η δοκιμή διαφορετικών μεθόδων, τεχνολογιών και δεδομένων τα οποία αναλύονται στην εργασία αυτή.

1.5 Δομή

Το κεφάλαιο 1 αποτελεί το εισαγωγικό μέρος της εργασίας, στο οποίο παρουσιάζεται το πρόβλημα που προσπαθεί να επιλύσει καθώς και ο στόχος της. Στο κεφάλαιο 2 γίνεται μία βιβλιογραφική επισκόπηση όπου αναφέρονται διαφορετικές προσεγγίσεις από άλλες εργασίες που προσπάθησαν να αντιμετωπίσουν προβλήματα σχετικά με την επεξεργασία της φυσικής γλώσσας. Στο κεφάλαιο 3 αναπτύσσεται το θεωρητικό υπόβαθρο στο οποίο βασίστηκε η εργασία. Στο κεφάλαιο 4 αναπτύσσεται το τεχνολογικό υπόβαθρο το οποίο χρησιμοποιήθηκε για την εκπόνηση της εργασίας. Στο κεφάλαιο 5 γίνεται αναφορά στα σύνολα δεδομένων που δημιουργήθηκαν και που χρησιμοποιήθηκαν. Στο κεφάλαιο 6 αναλύεται η μεθοδολογία που ακολουθήθηκε για την δημιουργία του μοντέλου. Στο κεφάλαιο 7 παρουσιάζονται τα αποτελέσματα των μοντέλων που αναπτύχθηκαν. Στο κεφάλαιο 8 αναφέρονται τα συμπεράσματα που προέκυψαν.

1.6 Επίλογος

Η ενότητα αυτή ήταν μια εισαγωγή πάνω στο πρόβλημα το οποίο καλείται να επιλυθεί σε αυτή την εργασία. Τέθηκαν οι στόχοι της διπλωματικής και η δομή της αναφοράς.

Κεφάλαιο 2ο: Βιβλιογραφική Αναφορά

2.1 Εισαγωγή

Σε αυτή την ενότητα θα γίνει προσπάθεια αναφοράς προηγούμενων εργασιών και ερευνών που αφορούν την επεξεργασία της φυσικής γλώσσας σε συνδυασμό με την ανίχνευση ψευδών ειδήσεων.

2.2 Βιβλιογραφική Αναφορά

Οι Rohit Kumar Kaliyar, Anurag Goswami και Pratik Narang [5] προτείνουν, χρησιμοποιώντας το διαδομένο στην επεξεργασία της φυσικής γλώσσας μοντέλο BERT [6], μια προτεραιότητα για τη μοντελοποίηση των σχετικών πληροφοριών των ψευδών ειδήσεων, η οποία είναι ικανή να βελτιώσει την απόδοση ταξινόμησης με την ικανότητα να καταγράφει τις σημασιολογικές εξαρτήσεις και τις εξαρτήσεις μεγάλων αποστάσεων στις προτάσεις. Είναι μια προσέγγιση βαθιάς μάθησης που συνδυάζει διαφορετικά παράλληλα μπλοκ του βαθιού συνελκτικού νευρωνικού δικτύου ενός στρώματος (Convolutional Neural Network-CNN) με διαφορετικά μεγέθη πυρήνα και φίλτρα. Στην δημοσίευσή τους αναφέρουν ότι πετυχαίνουν ακρίβεια 98.9%.

Ταυτόχρονα, υπάρχουν και μελέτες οι οποίες έφτασαν στο στάδιο του προϊόντος όπως το The Factual [7] μια υπηρεσία η οποία με αμοιβή προσφέρει επίβλεψη των ειδήσεων που παρουσιάζονται και τις βαθμολογεί με βάση κάποια κριτήρια τα οποία προκύπτουν από ένα μοντέλο τεχνητής νοημοσύνης. Η υπηρεσία επιτρέπει στους χρήστες της να αποφασίζουν τι θεωρούν ψευδές και τι αληθές με βάση τη βαθμολογία των κριτηρίων που εκθέτουν.

Στο ελληνικό διαδίκτυο συναντάμε στην ιστοσελίδα GreekHoaxes [8] η οποία θα αναφερθεί και αργότερα στην συγκεκριμένη εργασία. Τα «ellinikahoaxes.gr» είναι η πρώτη πιστοποιημένη προσπάθεια κατάρριψης αναληθών δημοσιευμάτων στο διαδίκτυο, αλλά και κάθε πηγή ενημέρωσης στην Ελλάδα. Ανήκουν στην Αστική Μη Κερδοσκοπική Εταιρία, με την επωνυμία “ΑΣΤΙΚΗ ΜΗ ΚΕΡΔΟΣΚΟΠΙΚΗ ΕΤΑΙΡΕΙΑ ΚΕΝΤΡΟ ΚΑΤΑΠΟΛΕΜΗΣΗΣ ΤΗΣ ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗΣ” με έδρα την Θεσσαλονίκη. Τα Greek Hoaxes είναι μια ανεξάρτητη πρωτοβουλία στο χώρο του ελέγχου ειδήσεων και δεν υπάγονται σε άλλο ΜΜΕ ή άλλη εταιρία. Η μεθοδολογία που χρησιμοποιείται από την ομάδα δεν βασίζεται σε μια συγκεκριμένη μεθοδολογία παράλα αυτά απαριθμούν κάποια βασικά βήματα ώστε να καταστεί πιο απλή η διαδικασία επαλήθευσης.

- Βήμα 1ο: Εντοπισμός δυνητικά ύποπτου υλικού
- Βήμα 2ο: Ανάλυση του περιεχομένου
- Βήμα 3ο: Έρευνα στο οπτικοακουστικό υλικό
- Βήμα 4ο: Εξέταση επιστημονικών μελετών σε περιπτώσεις ψευδο-επιστημονικών ισχυρισμών

- Βήμα 5ο: Επικοινωνία με άλλες ομάδες ελέγχου ειδήσεων

Πριν τέσσερα έτη αυτή η ομάδα δημιούργησε μια επέκταση φυλλομετρητή όπου εμφανίζει προειδοποιήσεις σχετικά με την αξιοπιστία των ιστοσελίδων που επισκέπτονται οι χρήστες, η οποία αλλάζει ανάλογα με την κατηγορία της εκάστοτε ιστοσελίδας.

Όπως όλα τα προβλήματα, έτσι και αυτό, μπορεί να δεχτεί διαφορετικές προσεγγίσεις. Οι Alonso M.A, Vilares D, Gómez-Rodríguez C, Vilares [9] εστίασαν στο γεγονός πως οι δημιουργοί των ψευδών ειδήσεων χρησιμοποιούν διάφορα υφολογικά τεχνάσματα για να προωθήσουν την επιτυχία των δημιουργιών τους, με ένα από αυτά να είναι η διέγερση των συναισθημάτων των αποδεκτών. Αυτό οδήγησε στην ανάλυση συναισθήματος, το τμήμα της ανάλυσης κειμένου που είναι υπεύθυνο για τον προσδιορισμό της πολικότητας και της ισχύος των συναισθημάτων που εκφράζονται σε ένα κείμενο, να χρησιμοποιείται σε προσεγγίσεις ανίχνευσης ψευδών ειδήσεων, είτε ως βάση του συστήματος είτε ως συμπληρωματικό στοιχείο.

Αντίθετα, οι Nguyen Vo, Kyumin Lee [10] εστίασαν στο γεγονός ότι ακόμα και η ύπαρξη μοντέλων εντοπισμού των ψευδών ειδήσεων, το φαινόμενο συνεχίζει να υπάρχει σε ίδια συχνότητα. Αυτό που έκαναν είναι να συλλέξουν και να αναλύσουν διαδικτυακούς χρήστες που ονομάζονται κηδεμόνες, οι οποίοι διορθώνουν την παραπληροφόρηση και τις ψευδείς ειδήσεις σε διαδικτυακές συζητήσεις παραπέμποντας σε διευθύνσεις URL που ελέγχουν τα γεγονότα. Προτείνουν ουσιαστικά, ένα νέο μοντέλο σύστασης διευθύνσεων URL που ελέγχονται τα γεγονότα για να ενθαρρύνονται οι κηδεμόνες να συμμετέχουν περισσότερο σε δραστηριότητες ελέγχου των ειδήσεων. Όπως αναφέρουν το μοντέλο τους ξεπερνάει παρόμοια state of the art μοντέλα κατά 11%-33%.

Μια ακόμα προσέγγιση που προτείνεται από τους Benjamin Riedel, Isabelle Augenstein, Georgios P. Spithourakis, Sebastian Riedel [11] είναι ο έλεγχος της αληθοφάνειας ενός συγκεκριμένου ισχυρισμού και η αξιολόγηση της στάσης που τηρούν οι διάφορες ειδησεογραφικές πηγές απέναντι στον ισχυρισμό. Η αυτόματη αξιολόγηση της στάσης, δηλαδή η ανίχνευση της στάσης αυτής, διευκόλυνε αναμφισβήτητα τη διαδικασία ελέγχου των γεγονότων. Η έρευνα αυτή όπου έγινε το σύστημα ανίχνευσης στάσης κατέλαβε την τρίτη θέση στο πρώτο στάδιο του διαγωνισμού Fake News Challenge [12].

2.3 Επίλογος

Σε αυτή την ενότητα έγινε μια αναφορά σε προηγούμενες έρευνες, προϊόντα και εργαλεία τα οποία σχετίζονται με το θέμα αυτής της διπλωματικής εργασίας.

Κεφάλαιο 3ο: Θεωρητικό Υπόβαθρο

3.1 Εισαγωγή

Σε αυτήν την ενότητα, θα γίνει μια αναλυτική επισκόπηση, πάνω στην θεωρία που χρησιμοποιήθηκε για την εκπόνηση της εργασίας αυτής. Θα γίνει αναφορά σε εστιασμένο κομμάτι της Τεχνητής Νοημοσύνης και πιο συγκεκριμένα στο κομμάτι της μηχανικής εκπαίδευσης. Στο συγκεκριμένο πρόβλημα, η εκπαίδευση γίνεται με επίβλεψη καθώς η επίλυση του συγκεκριμένου προβλήματος επιτυγχάνεται με την ταξινόμηση. Παρακάτω αναλύεται και η θεωρία των νευρωνικών δικτύων, καθώς και η επεξεργασία της φυσικής γλώσσας.

3.2 Μάθηση Με Επίβλεψη

Στην μάθηση με επίβλεψη το σύστημα πρέπει να “μάθει” επαγωγικά από κάποια δεδομένα, μία συνάρτηση που ονομάζεται συνάρτηση στόχος (target function) και αποτελεί την έκφραση του μοντέλου που περιγράφει τα δεδομένα. Η συνάρτηση στόχος χρησιμοποιείται στη συνέχεια για την πρόβλεψη της τιμής μιας μεταβλητής, που ονομάζεται εξαρτημένη μεταβλητή ή μεταβλητή εξόδου, βάσει των τιμών ενός συνόλου μεταβλητών που ονομάζονται ανεξάρτητες μεταβλητές ή μεταβλητές εισόδου ή χαρακτηριστικά.

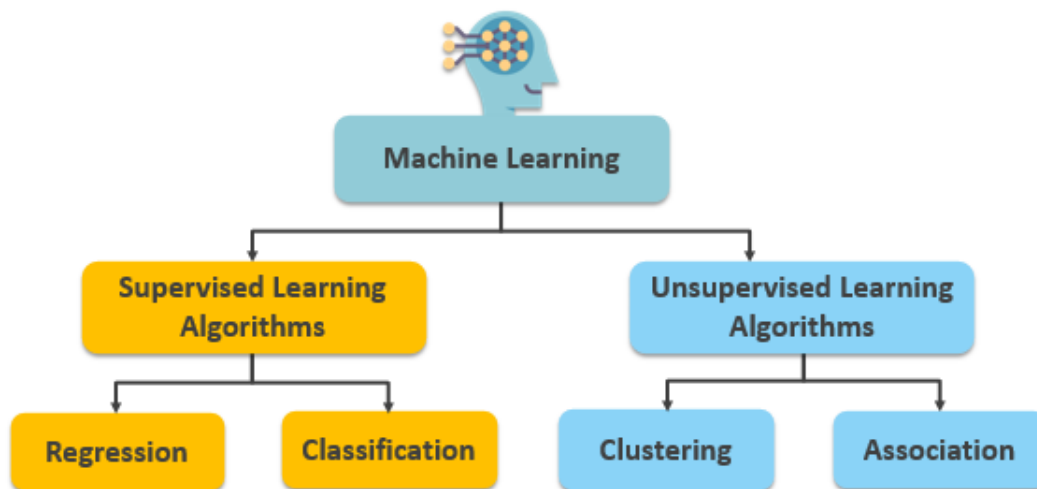
Η πιο απλή λύση στο πρόβλημα της μάθησης συνάρτησης είναι η αποστήθιση. Το σύστημα μάθησης απλά αποθηκεύει το σύνολο των περιπτώσεων και μπορεί να επιστρέψει την τιμή της εξαρτημένης μεταβλητής για μία νέα περίπτωση μόνο αν υπάρχει μια ολόκληρη αποθηκευμένη. Είναι ευνόητο ότι μία τέτοια προσέγγιση δεν περιέχει στοιχεία μάθησης και δεν μπορεί να είναι αποτελεσματική όταν το σύνολο δεδομένων περιλαμβάνει ένα μικρό μόνο υποσύνολο των περιπτώσεων.

Σε πολύ μεγάλα πεδία ορισμού είναι απαραίτητο να γίνει γενίκευση ώστε πολλές παρόμοιες περιπτώσεις να κωδικοποιηθούν πίσω από λίγες γενικευμένες περιπτώσεις. Η μάθηση με επίβλεψη διακρίνεται σε δύο είδη προβλημάτων, τα προβλήματα ταξινόμησης και τα προβλήματα παρεμβολής. Η ταξινόμηση (classification) αφορά στην δημιουργία μοντέλων πρόβλεψης διακριτών τάξεων (κλάσεων/κατηγοριών), ονομαστικών (nominal) ή βαθμωτών (ordinal) τιμών. Η παρεμβολή αφορά στην δημιουργία μοντέλων πρόβλεψης συνεχών αριθμητικών τιμών.

Στην προκειμένη περίπτωση μας αφορά η περίπτωση της δυαδικής ταξινόμησης, όπου η διακριτή συνάρτηση εκφυλίζεται σε: $h: X \rightarrow \{0,1\}$ Με h συμβολίζονται οι υποθέσεις. Με X συμβολίζεται το σύνολο των περιπτώσεων. Το 0 και 1 είναι οι δύο πιθανές κλάσεις καθώς έχουμε

δυναμική ταξινόμηση, ψευδές ή αληθές. Στα μη γραμμικά μοντέλα, η αναμενόμενη τιμή εξόδου συνδέεται με τις παραμέτρους εισόδου με πιο πολύπλοκο τρόπο καθώς εισέρχονται εκθετικές, λογαριθμικές κτλ. εκφράσεις. Σε τέτοιες περιπτώσεις, όπως και σε περιπτώσεις που ο τύπος της συνάρτησης $f(x)$ δεν είναι γνωστός αλλά είναι γνωστό ένα σύνολο ζευγών $(x_i, f(x_i))$, χρησιμοποιούνται πολυωνυμικές συναρτήσεις n -οστού βαθμού της μορφής:

$$y = f(x) = x_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n \quad (1.1)$$



Σχήμα 3.1: Υποκατηγορίες Μηχανικής Μάθησης

3.3 Ταξινόμηση

3.3.1 Γενικά

Συνήθως οι άνθρωποι στην καθημερινή τους πρακτική κατατάσσουν ένα πρόβλημα σε κάποια κατηγορία της οποίας η λύση είναι γνωστή. Συνεπώς η κατηγοριοποίηση αποτελεί πολλές φορές το πρώτο βήμα για την επίλυση πολλών ειδών προβλημάτων.

Η ταξινόμηση (classification) είναι ο προσδιορισμός της κατηγορίας στην οποία ανήκει ένα αντικείμενο, φαινόμενο, πρότυπο, μέτρηση, κτλ. Τα προβλήματα ταξινόμησης έχουν ως είσοδο ένα σύνολο δεδομένων. Αυτά τα δεδομένα αποτελούν τη βασική γνώση η οποία θα χρησιμοποιηθεί για την επίλυση του προβλήματος, δηλαδή την έξοδο.

Το κυριότερο χαρακτηριστικό της ταξινόμησης είναι ότι η επιλογή της κατηγορίας στην οποία κατατάσσεται ένα αντικείμενο γίνεται από ένα προκαθορισμένο σύνολο κατηγοριών. Χωρίς αυτό να σημαίνει απαραίτητα ότι το κάθε αντικείμενο ανήκει σίγουρα σε κάποια κατηγορία. Στην προκειμένη περίπτωση το πρόβλημα ταξινόμησης είναι η κατηγοριοποίηση των ειδήσεων σε ψευδή (α' κατηγορία) ή αληθής (β' κατηγορία).

3.3.2 Ταξινομητές

Με βάση έρευνα [13] που συντάχθηκε στο πανεπιστήμιο του Κατοβίτσε της Πολωνίας, πάνω στην "Ταχεία ανίχνευση ψευδών ειδήσεων με βάση μεθόδους μηχανικής μάθησης", η πρώτη από τις προσεγγίσεις που αναλύθηκαν αφορούσε τη μάθηση με βάση τον τίτλο των ειδήσεων.

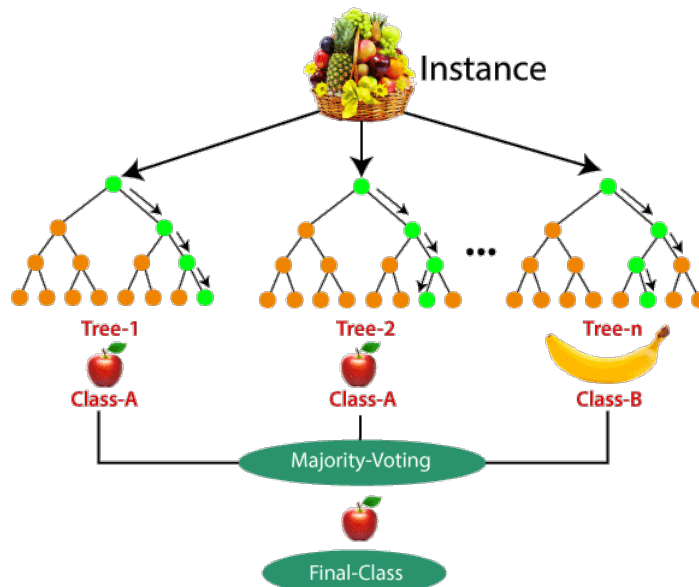
Η προσέγγιση αυτή επιτρέπει την ταξινόμηση ειδήσεων χωρίς την ανάγκη μακροχρόνιας ανάλυσης. Τα αποτελέσματα που σχετίζονται με αυτό το πείραμα αναλύονται στην έρευνα με βάση την ποιότητα της ταξινόμησης και τον χρόνο που απαιτείται για την εκμάθηση του αλγορίθμου.

Από τους πέντε επιλεγμένους αλγορίθμους CART, SVM, Random Forest, AdaBoost και Bagging, ξεχώρισαν με βάση τα παραπάνω κριτήρια, πρώτα ο αλγόριθμος SVM με ακρίβεια 0.94, με δεύτερο τον αλγόριθμο Random Forest με ακρίβεια 0.92. Αυτοί που ξεχώρισαν επιλέχθηκαν σε αυτή τη διπλωματική για την εφαρμογή των καινούριων συνόλων δεδομένων που δημιουργήθηκαν για την συγκεκριμένη εργασία. Επίσης, επιλέχθηκε και ο αλγόριθμος Passive Aggressive καθώς τα χαρακτηριστικά του, που αναλύονται παρακάτω, ταιριάζουν στο πρόβλημα και έχει χρησιμοποιηθεί για παρόμοιου είδους θέματα.

Random Forest Ο Random Forest είναι ένας τύπος μεθόδου συλλογικής μάθησης για ταξινόμηση και παλινδρόμηση. Συνδυάζει πολλαπλά δέντρα απόφασης για τη δημιουργία ενός πιο ισχυρού μοντέλου που είναι λιγότερο επιρρεπές στην υπερπροσαρμογή σε σύγκριση με ένα μεμονωμένο δέντρο απόφασης.

Η βασική ιδέα πίσω από το Random Forest είναι ότι ένα μεγάλο πλήθος από μη συσχετιζόμενα δέντρα που αποφασίζουν από κοινού (σαν επιτροπή), θα πάρουν καλύτερη απόφαση από κάθε δέντρο ξεχωριστά. Αν κάποια δέντρα αποφασίσουν λάθος πρόβλεψη, πολλά άλλα θα υπολογίσουν την σωστή οπότε λογικά η πρόβλεψη θα κινηθεί προς την σωστή κατεύθυνση. Με τον μέσο όρο των προβλέψεων όλων των δέντρων απόφασης, ο Random Forest μειώνει τη διακύμανση των προβλέψεων. Αυτές οι τεχνικές επιλογής υποσυνόλων είναι γνωστές ως bootstrap aggregating (bagging) και μέθοδος τυχαίου υποσυνόλου (feature randomness) αντίστοιχα.

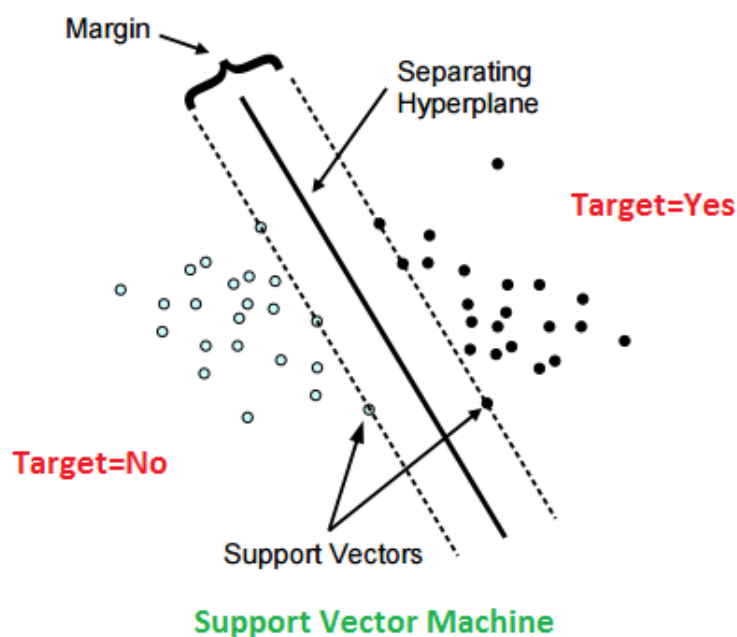
Ο random Forest κατατάσσεται στους κορυφαίους αλγορίθμους ταξινόμησης. Ακολουθεί την φιλοσοφία της σοφίας του πλήθους τα οποία από πολύ αδύναμα μοντέλα συνεργαζόμενα δημιουργούν ισχυρά μοντέλα. Ένα σημαντικό πλεονέκτημα τους είναι ότι δεν υπερπροσαρμόζονται σε καλά σύνολα δεδομένων (μείωση της διακύμανσης χωρίς αύξηση της μεροληψίας-bias) ενώ παράλληλα μπορούν να χειριστούν μεγάλα σύνολα εκπαίδευσης με πολλά χαρακτηριστικά. Επιπλέον, η κατασκευή τους δεν είναι υπολογιστικά πολύπλοκη. [14]



Σχήμα 3.2: Random Forest

SVM Οι Μηχανές Διανυσμάτων Υποστήριξης - Support Vector Machines (SVM) είναι ένας τύπος αλγορίθμων μάθησης με επίβλεψη που μπορεί να χρησιμοποιηθεί τόσο για εργασίες ταξινόμησης όσο και παλινδρόμησης. Βασίζονται στη θεωρία στατιστικής μάθησης.

Η μέθοδος SVM επιδιώκει να βρει το σύνορο που απέχει όσο το δυνατόν περισσότερο από τα παραδείγματα των κλάσεων που διαχωρίζει. Η υπερ-επιφάνεια αυτή ονομάζεται υπερ-επιφάνεια μέγιστου περιθωρίου και σε γραμμικώς διαχωρίσιμα προβλήματα ορίζεται από έναν πεπερασμένο αριθμό παραδειγμάτων του συνόλου εκπαίδευσης που ονομάζονται διανύσματα υποστήριξης. Επιπλέον, μέσω των συναρτήσεων πυρήνα, οι SVM μπορούν να μετασχηματίσουν τον αρχικό χώρο υποθέσεων έτσι ώστε τα μη γραμμικά διαχωρίσιμα προβλήματα να μετατραπούν σε γραμμικά διαχωρίσιμα και τελικά να λυθούν με την ίδια μεθοδολογία. [14]



Σχήμα 3.3: Υπερ-επιφάνεια Μέγιστου Περιθωρίου

Οι SVM είναι ιδιαίτερα αποτελεσματικές σε περιπτώσεις όπου ο αριθμός των χαρακτηριστικών είναι πολύ μεγαλύτερος από τον αριθμό των δειγμάτων και χρησιμοποιούνται συχνά σε προβλήματα ταξινόμησης εικόνων και κειμένων.

Passive Aggressive Ο αλγόριθμος Passive Aggressive (PA) είναι ένας διαδικτυακός αλγόριθμος μάθησης που μπορεί να χρησιμοποιηθεί για εργασίες δυαδικής ταξινόμησης. Αυτός ο αλγόριθμος βασίζεται στον αλγόριθμο Perceptron, αλλά περιλαμβάνει μια παράμετρο κανονικοποίησης, C , η οποία ελέγχει τον συμβιβασμό μεταξύ της αύξησης του μεγέθους του περιθωρίου και της διασφάλισης ότι ο ταξινομητής δεν ταξινομεί λανθασμένα παραδείγματα εκπαίδευσης. Επιπλέον, οι passive aggressive ταξινομητές ενημερώνουν το μοντέλο τους μόνο όταν υπάρχει λάθος και όχι μετά από κάθε παράδειγμα όπως ο Perceptron. Οι passive aggressive ταξινομητές έχουν βρεθεί να αποδίδουν καλά σε εργασίες κατηγοριοποίησης κειμένου και συνεργατικού

φιλτραρίσματος σε πραγματικές εφαρμογές. Στην παραλλαγή PA-I, οι παράμετροι του μοντέλου ενημερώνονται με πιο επιθετικό τρόπο με βάση το σφάλμα της πρόβλεψης. Συγκεκριμένα, οι παράμετροι του μοντέλου ενημερώνονται προς την κατεύθυνση που ελαχιστοποιεί την απώλεια άρθρωσης, η οποία ορίζεται ως: $loss = \max(0, 1 - \text{πρόβλεψη}(\text{prediction}))$ (3.1)

$$loss = \max(0, 1 - (\text{prediction})) \quad (3.1)$$

όπου y είναι η πραγματική ετικέτα (true label) (1 ή -1) και prediction είναι η πρόβλεψη που γίνεται από το τρέχον μοντέλο.

3.4 Νευρωνικά Δίκτυα

3.4.1 Γενικά

Τα τεχνητά νευρωνικά δίκτυα (ΤΝΔ) είναι μία ιδιαίτερη προσέγγιση στην δημιουργία συστημάτων με στοιχεία νοημοσύνης, καθώς αποφεύγουν να αναπαραστήσουν ρητά την γνώση και να υιοθετήσουν ειδικά σχεδιασμένους αλγόριθμους αναζήτησης. Αντίθετα, βασίζονται σε βιολογικά πρότυπα καθώς χρησιμοποιούν δομές και διαδικασίες που μιμούνται τις αντίστοιχες του ανθρώπινου εγκεφάλου.

Στην κλασική τους θεώρηση, τα ΤΝΔ μαθαίνουν να εκτελούν δεδομένη λειτουργία μελετώντας κάποια παραδείγματα, χωρίς δηλαδή να προγραμματίζονται ειδικά για αυτό, με την κλασική έννοια του όρου. Το εντυπωσιακό είναι ότι μαθαίνουν χωρίς κάποια άλλη ιδιαίτερη γνώση για αυτά τα αντικείμενα.

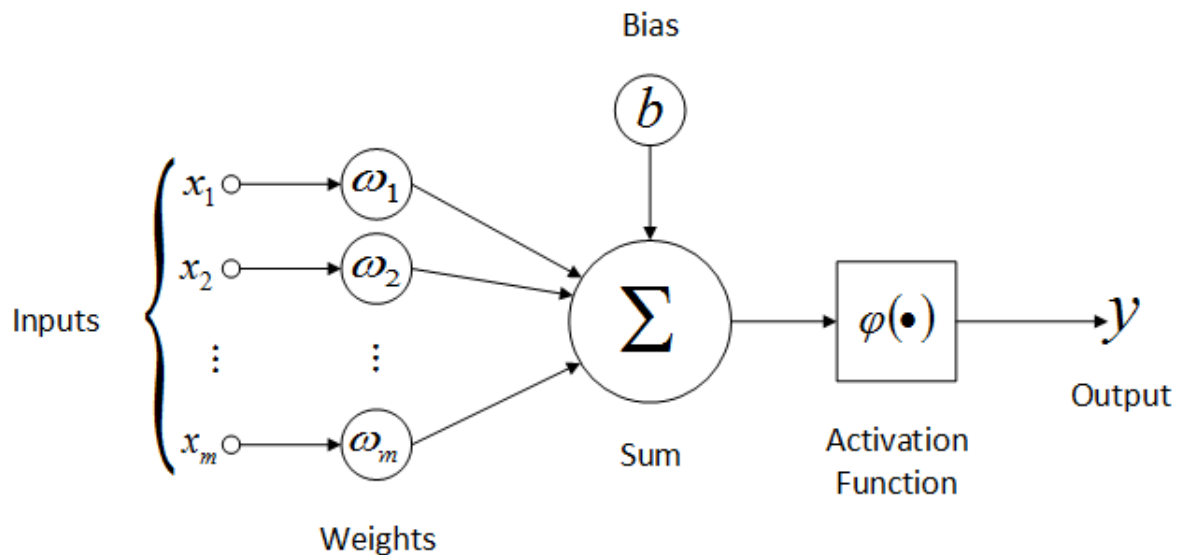
3.4.2 Μοντέλο Τεχνητού Νευρώνα

Ο τεχνητός νευρώνας είναι ένα υπολογιστικό μοντέλο τα μέρη του οποίου μπορεί να αντιστοιχιστούν άμεσα με αυτά του βιολογικού νευρώνα. Ένας τεχνητός νευρώνας δέχεται κάποια σήματα εισόδου τα οποία, σε αντίθεση με τους ηλεκτρικούς παλμούς του εγκεφάλου, αντιστοιχούν σε συνεχείς μεταβλητές. Κάθε τέτοιο σήμα εισόδου επηρεάζεται από μία τιμή βάρους ο ρόλος της οποίας είναι αντίστοιχος του ρόλου της σύναψης στο βιολογικό νευρώνα. Η τιμή βάρους μπορεί να είναι θετική ή αρνητική, σε αντιστοιχία με την επιταχυντική ή επιβραδυντική λειτουργία της σύναψης.

Εκτός από τα εισερχόμενα σήματα x_i που διαμορφώνονται (πολλαπλασιάζονται) από τα αντίστοιχα βάρη w_1 , ο νευρώνας δέχεται ως είσοδο και μία σταθερή τιμή $w_0=1$ που διαμορφώνεται

από το βάρος w_0 . Αυτό το επιπλέον σήμα εισόδου ονομάζεται πόλωση παρένθεση (bias) του νευρώνα.

Το σώμα του τεχνητού νευρώνα χωρίζεται σε δύο μέρη, τον αθροιστή (sum) που προσθέτει τα διαμορφωμένα από τα βάρη σήματα εισόδου παράγοντας στην ποσότητα S και την συνάρτηση ενεργοποίησης (activation function), ένα είδος φίλτρο που δέχεται ως όρισμα την ποσότητα S και διαμορφώνει την τελική τιμή y του σήματος εξόδου.

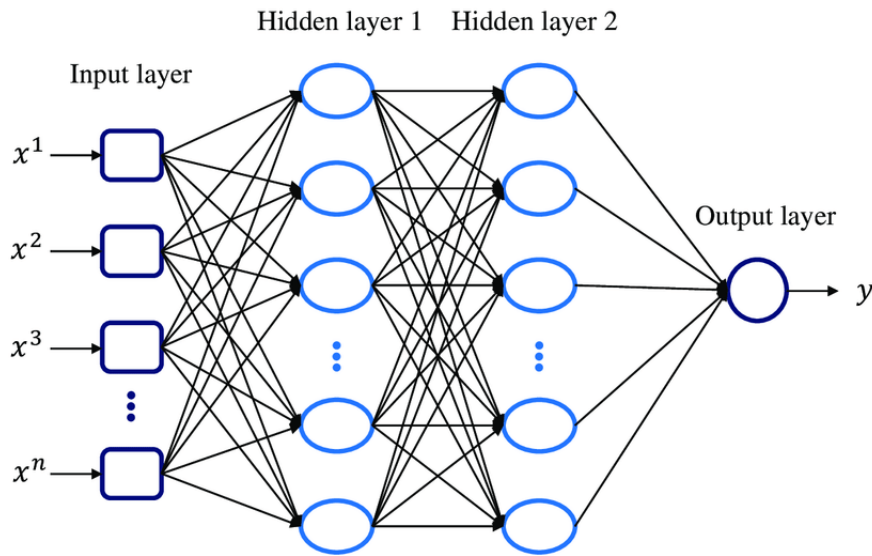


Σχήμα 3.4: Μοντέλο Τεχνητού Νευρώνα

3.5 Τεχνητά Νευρωνικά Δίκτυα Πολλών Επιπέδων (Multi-Layer Perceptron MLP)

Το Perceptron είναι ένας τύπος μοντέλου νευρωνικού δικτύου που αναπτύχθηκε τη δεκαετία του 1950 και θεωρείται ο πρόδρομος των πιο σύνθετων μοντέλων νευρωνικών δικτύων, όπως τα Multi-Layer Perceptrons (MLP). Το perceptron είναι ένα δίκτυο ενός στρώματος που αποτελείται από έναν γραμμικό συνδυασμό εισόδων και μια δυαδική συνάρτηση κατωφλίου. Χρησιμοποιείται για εργασίες δυαδικής ταξινόμησης όπου ο στόχος είναι να διαχωριστούν δύο κλάσεις με ένα γραμμικό όριο. Ο αλγόριθμος perceptron ενημερώνει επαναληπτικά τα βάρη του μοντέλου προκειμένου να ελαχιστοποιηθεί ο αριθμός των λανθασμένων ταξινομήσεων στο σύνολο της εκπαίδευσης. Ωστόσο, τα perceptron έχουν έναν θεμελιώδη περιορισμό ότι μπορούν να επιλύσουν μόνο γραμμικά διαχωρίσιμα προβλήματα, πράγμα που σημαίνει ότι δεν μπορούν να επιλύσουν προβλήματα όπου οι κλάσεις δεν είναι γραμμικά διαχωρίσιμες. Αυτός ο περιορισμός ξεπεράστηκε με την ανάπτυξη των πολυεπίπεδων perceptrons (MLP) τα οποία είναι ικανά να επιλύουν μη γραμμικά προβλήματα. Ο αλγόριθμος MLP χρησιμοποιεί πολλαπλά στρώμα-

τα perceptrons και η προσθήκη μη γραμμικών συναρτήσεων ενεργοποίησης τους επιτρέπει να μαθαίνουν πολύπλοκες σχέσεις στα δεδομένα.



Σχήμα 3.5: Νευρωνικό Δίκτυο Πολλών Επιπέδων

Η αρχιτεκτονική ενός MLP περιλαμβάνει συνήθως ένα στρώμα εισόδου, ένα ή περισσότερα κρυφά στρώματα και ένα στρώμα εξόδου. Το στρώμα εισόδου δέχεται δεδομένα εισόδου, τα οποία στη συνέχεια επεξεργάζονται και μετασχηματίζονται από τα κρυφά στρώματα προτού φτάσουν στο στρώμα εξόδου. Κάθε στρώμα ενός MLP αποτελείται από πολλαπλά perceptrons, εκτελούν έναν υπολογισμό σε αυτά και παράγουν μια ενιαία έξοδο. Ο υπολογισμός που εκτελείται από ένα perceptron είναι συχνά μια απλή μαθηματική πράξη, όπως μια μη γραμμική συνάρτηση ενεργοποίησης, η οποία εφαρμόζεται στις εισόδους.

Τα κρυφά στρώματα ενός MLP είναι υπεύθυνα για την εξαγωγή των σχετικών χαρακτηριστικών από τα δεδομένα εισόδου και τη μετατροπή τους σε μια αναπαράσταση που είναι χρήσιμη για την εκάστοτε εργασία. Αυτή η διαδικασία είναι γνωστή ως εξαγωγή χαρακτηριστικών και είναι κρίσιμη για την επιτυχία ενός MLP. Το στρώμα εξόδου, από την άλλη πλευρά, παράγει την τελική πρόβλεψη ή έξοδο του δικτύου με βάση τα χαρακτηριστικά που εξάγονται από τα κρυφά στρώματα.

Το MLP είναι ένα μοντέλο μάθησης με επίβλεψη, πράγμα που σημαίνει ότι για την εκπαίδευσή του απαιτείται ένα σύνολο δεδομένων με ετικέτες. Η διαδικασία εκπαίδευσης ενός MLP περιλαμβάνει την προσαρμογή των βαρών των συνδέσεων του δικτύου, έτσι ώστε το δίκτυο να μπορεί να παράγει τη σωστή έξοδο για μια δεδομένη είσοδο. Αυτό γίνεται με την επαναληπτική παρουσίαση στο δίκτυο ζευγών εισόδου/εξόδου από το σύνολο δεδομένων εκπαίδευσης και την προσαρμογή των βαρών με βάση το σφάλμα μεταξύ της εξόδου του δικτύου και της σωστής

εξόδου.

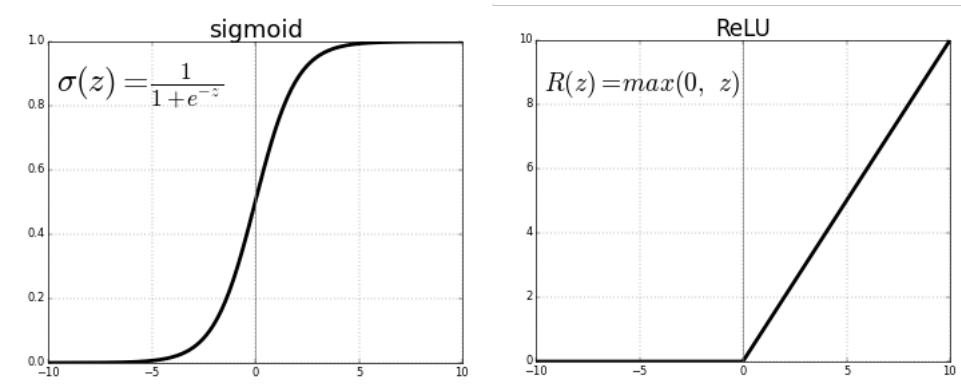
Τα βασικά μαθηματικά πίσω από ένα MLP βασίζονται στα μαθηματικά των τεχνητών νευρωνικών δικτύων (ANN) γενικά. Ένας τεχνητός νευρώνας, επίσης γνωστός ως perceptron, είναι μια απλή μονάδα επεξεργασίας που δέχεται ένα σύνολο εισόδων, εφαρμόζει μια μαθηματική πράξη σε αυτές και παράγει μια μοναδική έξοδο. Η μαθηματική πράξη που εφαρμόζεται στις εισόδους είναι γνωστή ως συνάρτηση ενεργοποίησης. Μια κοινή συνάρτηση ενεργοποίησης που χρησιμοποιείται στα MLP είναι η σιγμοειδής συνάρτηση, η οποία ορίζεται ως εξής:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3.2)$$

Η σιγμοειδής συνάρτηση αντιστοιχίζει κάθε είσοδο x σε μια τιμή μεταξύ 0 και 1. Αυτό είναι χρήσιμο για το επίπεδο εξόδου ενός MLP όταν η εργασία είναι ένα δυαδικό πρόβλημα ταξινόμησης. Μια άλλη κοινή συνάρτηση ενεργοποίησης είναι η συνάρτηση ReLU (rectified linear unit), η οποία ορίζεται ως εξής:

$$f(x) = \max(0, x) \quad (3.3)$$

Η συνάρτηση ReLU αντιστοιχίζει κάθε είσοδο x είτε στο 0 είτε στο x , ανάλογα με την τιμή του x . Αυτή η συνάρτηση χρησιμοποιείται συνήθως στα κρυφά στρώματα των MLP, επειδή βοηθά στη μείωση του προβλήματος της εξαφανισμένης κλίσης.



Σχήμα 3.6: Σιγμοειδής Συνάρτηση, Συνάρτηση ReLU

Η έξοδος ενός τεχνητού νευρώνα υπολογίζεται ως το τετραγωνικό γινόμενο των εισόδων και των βαρών συν τη μεροληψία. Μαθηματικά, μπορεί να αναπαρασταθεί ως εξής:

$$output = f(w_1x_1 + w_2x_2 + \dots + w_nx_n + b) \quad (3.4)$$

Όπου w_1, w_2, \dots, w_n είναι τα βάρη των συνδέσεων μεταξύ των εισόδων και του perceptron, x_1, x_2, \dots, x_n είναι οι εισοδοί, b είναι η προκατάληψη και f είναι η συνάρτηση ενεργοποίησης.

3.5.1 Βαθύ Νευρωνικό Δίκτυο (Deep Neural Network DNN)

Το βαθύ νευρωνικό δίκτυο (DNN) είναι ένας τύπος τεχνητού νευρωνικού δικτύου (ANN) που αποτελείται από πολλαπλά στρώματα διασυνδεδεμένων κόμβων ή "νευρώνων". Τα DNN χρησιμοποιούνται για ένα ευρύ φάσμα εργασιών, συμπεριλαμβανομένης της αναγνώρισης εικόνων, της επεξεργασίας φυσικής γλώσσας και της αναγνώρισης ομιλίας.

Το δομικό στοιχείο ενός DNN είναι ο νευρώνας, ο οποίος είναι μια απλή μαθηματική συνάρτηση που δέχεται ένα σύνολο εισόδων, εκτελεί έναν υπολογισμό σε αυτές και παράγει μια μοναδική έξοδο. Σε ένα DNN, οι εισοδοί είναι συνήθως τα χαρακτηριστικά ενός δεδομένου εισόδου, όπως τα εικονοστοιχεία μιας εικόνας, οι λέξεις σε μια πρόταση ή τα δείγματα ήχου σε ένα σήμα ομιλίας. Οι έξοδοι των νευρώνων σε ένα επίπεδο χρησιμοποιούνται στη συνέχεια ως εισοδοί για το επόμενο επίπεδο, και ούτω καθεξής, μέχρι να παραχθεί η τελική έξοδος.

Αυτό που καθιστά τα DNN ισχυρά είναι ο μεγάλος αριθμός παραμέτρων που μπορούν να μάθουν, γεγονός που τους επιτρέπει να μοντελοποιούν πολύπλοκα μοτίβα και σχέσεις στα δεδομένα. Αυτές οι παράμετροι μαθαίνονται συνήθως με τη χρήση ενός αλγορίθμου μάθησης με επίβλεψη, όπως ο αλγόριθμος backpropagation, ο οποίος ρυθμίζει επαναληπτικά τις παραμέτρους ώστε να ελαχιστοποιείται η διαφορά μεταξύ των προβλεπόμενων εξόδων και των πραγματικών εξόδων.

Ο αριθμός των επιπέδων σε ένα DNN μπορεί να ποικίλλει ανάλογα με την εργασία, με ορισμένα DNN να έχουν μόνο μερικά επίπεδα, ενώ άλλα μπορεί να έχουν εκατοντάδες ή ακόμη και χιλιάδες επίπεδα. Η αρχιτεκτονική ενός DNN μπορεί επίσης να ποικίλλει, με ορισμένα DNN να χρησιμοποιούν πλήρως συνδεδεμένα στρώματα, όπου κάθε νευρώνας σε ένα στρώμα συνδέεται με κάθε νευρώνα στο επόμενο στρώμα, ενώ άλλα χρησιμοποιούν συνελκτικά στρώματα, όπου οι συνδέσεις είναι πιο εντοπισμένες.

Τα βασικά μαθηματικά πίσω από ένα βαθύ νευρωνικό δίκτυο (DNN) περιλαμβάνουν γραμμική άλγεβρα και αριθμητική. Σε υψηλό επίπεδο, ένα DNN είναι μια συνάρτηση που αντιστοιχίζει τις εισόδους στις εξόδους, όπου οι εισοδοί είναι συνήθως τα χαρακτηριστικά των δεδομένων εισόδου και οι έξοδοι είναι οι προβλέψεις ή οι ταξινομήσεις.

1.Γραμμικός μετασχηματισμός στο πρώτο στρώμα: w_1 είναι ο πίνακας βαρών του πρώτου στρώματος με διάσταση $m \times n$, όπου m είναι ο αριθμός των νευρώνων στο κρυφό στρώμα και n είναι ο αριθμός των νευρώνων στο στρώμα εισόδου.

$$z_1 = w_1x + b_1 \quad (3.6)$$

όπου x είναι το διάνυσμα εισόδου και b_1 είναι το διάνυσμα πόλωσης. Αυτός είναι ο γραμμικός μετασχηματισμός του διανύσματος εισόδου, όπου W_1 και b_1 είναι οι παράμετροι του DNN.

2.Μη γραμμική συνάρτηση ενεργοποίησης στο πρώτο επίπεδο:

$$a_1 = g(z_1) \quad (3.7)$$

όπου g είναι η συνάρτηση ενεργοποίησης, όπως η σιγμοειδής συνάρτηση ή η συνάρτηση ReLU.

3.Γραμμικός μετασχηματισμός στο δεύτερο στρώμα: w_2 είναι ο πίνακας βαρών του δεύτερου στρώματος με διάσταση $k \times m$, όπου k είναι ο αριθμός των νευρώνων στο στρώμα εξόδου και m είναι ο αριθμός των νευρώνων στο κρυφό στρώμα.

$$z_2 = w_2a_1 + b_2 \quad (3.8)$$

όπου a_1 είναι η έξοδος της ενεργοποίησης του πρώτου στρώματος και b_2 είναι το διάνυσμα προκατάληψης.

4.Μη γραμμική συνάρτηση ενεργοποίησης στο δεύτερο στρώμα:

$$y = h(z_2) \quad (3.9)$$

όπου h είναι η τελική συνάρτηση ενεργοποίησης, η οποία είναι συνήθως softmax για πρόβλημα ταξινόμησης πολλαπλών κλάσεων.

5.Συνάρτηση απώλειας:

$$L(y, t) = Loss(y, t) \quad (3.10)$$

όπου y είναι η έξοδος του DNN, t είναι η πραγματική ετικέτα και Loss μπορεί να είναι οποιαδήποτε συνάρτηση απώλειας (π.χ. απώλεια cross-entropy).

6.Οπισθοδρόμηση:

Οι πίνακες βαρών W_1 και W_2 ενημερώνονται με τη χρήση αλγορίθμου βαθμωτής καθόδου με την κλίση της συνάρτησης απώλειας ως προς τους πίνακες βαρών.

Αυτοί οι τύποι βασίζονται σε ένα μοντέλο με ένα στρώμα εισόδου, ένα κρυμμένο στρώμα και ένα στρώμα εξόδου. Στην εργασία χρησιμοποιούνται πολλά παραπάνω στρώματα οπότε οι μαθηματικοί τύποι που εφαρμόζονται αλλάζουν ως προς τους πίνακες των βαρών και τα διανύσματα προκατάληψης (bias).

3.5.2 Πλήρως Συνδεδεμένο Στρώμα

Ένα πλήρως συνδεδεμένο στρώμα, επίσης γνωστό ως πυκνό στρώμα (dense layer), είναι ένας τύπος στρώματος νευρωνικού δικτύου στο οποίο κάθε νευρώνας στο στρώμα συνδέεται με κάθε νευρώνα στο προηγούμενο στρώμα. Αυτό σημαίνει ότι η έξοδος κάθε νευρώνα στο στρώμα υπολογίζεται με τη λήψη ενός σταθμισμένου αθροίσματος των εξόδων όλων των νευρώνων στο προηγούμενο στρώμα, την εφαρμογή μιας συνάρτησης ενεργοποίησης και, στη συνέχεια, τη διαβίβαση του αποτελέσματος στο επόμενο στρώμα. Ο αριθμός των νευρώνων στο πλήρως συνδεδεμένο στρώμα καθορίζει τον αριθμό των χαρακτηριστικών εξόδου και τα βάρη των συνδέσεων μεταξύ των νευρώνων μαθαίνονται κατά τη διάρκεια της διαδικασίας εκπαίδευσης μέσω της οπισθοδρόμησης. Σε ένα πλήρως συνδεδεμένο νευρωνικό δίκτυο, η έξοδος από ένα επίπεδο τροφοδοτείται ως είσοδος στο επόμενο επίπεδο. Αυτό γίνεται έτσι ώστε το δίκτυο να μπορεί να μάθει μοτίβα στα δεδομένα εισόδου που δεν περιορίζονται από τη χωρική δομή των δεδομένων. Αυτό επιτρέπει στο δίκτυο να μαθαίνει τόσο τοπικά όσο και παγκόσμια μοτίβα στα δεδομένα και καθιστά τα πλήρως συνδεδεμένα στρώματα πολύ ισχυρά για εργασίες όπως η ταξινόμηση εικόνων, η επεξεργασία φυσικής γλώσσας και άλλα.

Τα μαθηματικά πίσω από ένα πλήρως συνδεδεμένο επίπεδο μπορούν να αναλυθούν σε δύο μέρη: τον υπολογισμό του σταθμισμένου αθροίσματος των εισόδων και την εφαρμογή της συνάρτησης ενεργοποίησης. Το σταθμισμένο άθροισμα των εισόδων για έναν μεμονωμένο νευρώνα στο πλήρως συνδεδεμένο στρώμα υπολογίζεται ως εξής:

$$z = (w_i x_i) + b \quad (3.11)$$

όπου z είναι το σταθμισμένο άθροισμα, w_i είναι το βάρος της σύνδεσης μεταξύ του i -οστού νευρώνα στο προηγούμενο στρώμα και του τρέχοντος νευρώνα, x_i είναι η έξοδος του i -οστού νευρώνα στο προηγούμενο στρώμα και b είναι ο όρος πόλωσης για τον τρέχοντα νευρώνα.

3.5.3 Κανονικοποίηση και Dropout

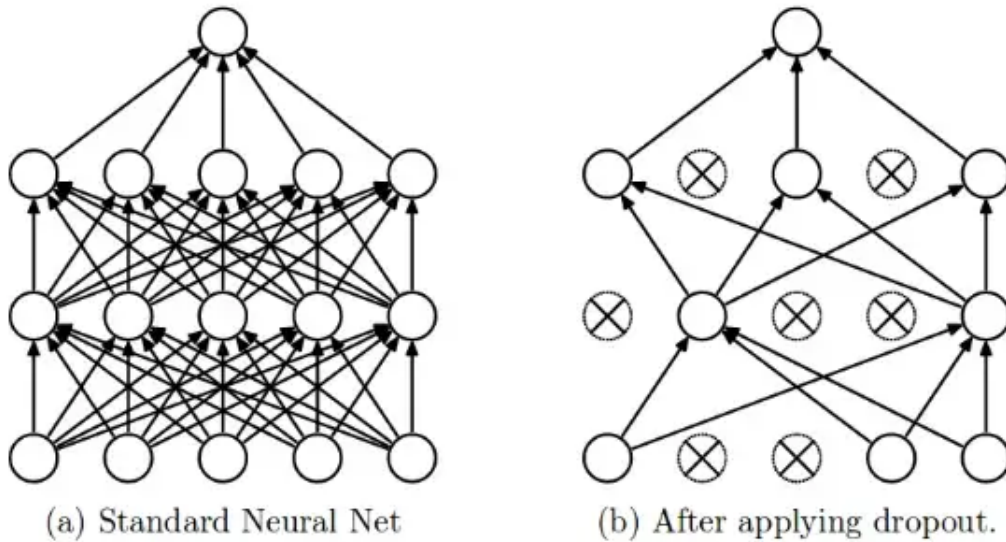
A. Batch Normalisation

Η ομαλοποίηση παρτίδων είναι μια τεχνική που χρησιμοποιείται για την ομαλοποίηση των εισόδων για κάθε μίνι-παρτίδα δεδομένων κατά τη διάρκεια της εκπαίδευσης. Βοηθά στη σταθεροποίηση της εκπαίδευσης των βαθιών νευρωνικών δικτύων μειώνοντας την εσωτερική μετατόπιση των διακυμάνσεων, δηλαδή την αλλαγή στην κατανομή των εισόδων σε ένα επίπεδο καθώς οι παράμετροι των προηγούμενων επιπέδων αλλάζουν κατά τη διάρκεια της εκπαίδευσης. Η βασική ιδέα πίσω από την ομαλοποίηση δέσμης είναι η ομαλοποίηση των ενεργοποιήσεων ενός στρώματος αφαιρώντας τη μέση τιμή και διαιρώντας με την τυπική απόκλιση. Αυτή η κανονικοποίηση γίνεται ξεχωριστά για κάθε μίνι-πακέτο δεδομένων, εξ ου και η ονομασία "κανονικοποίηση παρτίδας". Επιπλέον, η ομαλοποίηση παρτίδας έχει επίσης ένα αποτέλεσμα κανονικοποίησης προσθέτοντας θόρυβο στις ενεργοποιήσεις, το οποίο συμβάλλει στη μείωση της υπερ-προσαρμογής. Στην πράξη, η κανονικοποίηση παρτίδας εφαρμόζεται συνήθως στα πλήρως συνδεδεμένα στρώματα ενός νευρωνικού δικτύου και η κανονικοποίηση γίνεται συνήθως με τη χρήση του μέσου όρου και της τυπικής απόκλισης των ενεργοποιήσεων για κάθε μίνι-παρτίδα δεδομένων. Η ομαλοποίηση παρτίδας έχει αποδειχθεί πολύ αποτελεσματική στην πράξη και χρησιμοποιείται συνήθως στη βαθιά μάθηση και στα νευρωνικά δίκτυα, χρησιμοποιείται επίσης στα υπολειμματικά δίκτυα (ResNets) και είναι ένα σημαντικό συστατικό των γεννητικών αντιθετικών δικτύων (Generative Adversarial Networks - GANs)

B. Dropout

Το Dropout είναι μια τεχνική που χρησιμοποιείται για την αποφυγή της υπερ-προσαρμογής στα νευρωνικά δίκτυα. Λειτουργεί θέτοντας τυχαία ένα ποσοστό των νευρώνων του δικτύου στο μηδέν κατά τη διάρκεια κάθε επανάληψης εκπαίδευσης. Αυτό αναγκάζει το δίκτυο να μάθει πολλαπλές ανεξάρτητες αναπαραστάσεις των δεδομένων εισόδου, γεγονός που συμβάλλει στη μείωση της υπερ-προσαρμογής. Η βασική ιδέα πίσω από το dropout είναι η τυχαία απόρριψη ορισμένων νευρώνων κατά τη διάρκεια του forward pass, γεγονός που μειώνει τη συν-προσαρμογή των νευρώνων και ενθαρρύνει το δίκτυο να μάθει πολλαπλές ανεξάρτητες αναπαραστάσεις των δεδομένων εισόδου. Το επίπεδο Dropout θέτει τυχαία τις μονάδες εισόδου στο 0 με συχνότητα ρυθμού σε κάθε βήμα κατά τη διάρκεια της εκπαίδευσης, γεγονός που βο-

θά στην αποφυγή της υπερ-προσαρμογής. Οι εισόδοι που δεν τίθενται στο 0 αναβαθμίζονται κατά $1/(1 - \text{rate})$ έτσι ώστε το άθροισμα σε όλες τις εισόδους να παραμένει αμετάβλητο. Η προτεινόμενη τιμή για το dropout σε ένα κρυφό στρώμα είναι μεταξύ 0,5 και 0,8. Τα στρώματα εισόδου χρησιμοποιούν μεγαλύτερο ποσοστό διακοπής, όπως 0,8.



Σχήμα 3.7: Dropout

3.6 Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing NLP)

3.6.1 Γενικά

Η Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP) είναι ένας τομέας της επιστήμης των υπολογιστών, της τεχνητής νοημοσύνης και της υπολογιστικής γλωσσολογίας που επικεντρώνεται στην αλληλεπίδραση μεταξύ υπολογιστών και ανθρώπων μέσω της χρήσης της φυσικής γλώσσας. Περιλαμβάνει τεχνικές για την ανάλυση, την παραγωγή και την κατανόηση της ανθρώπινης γλώσσας, ώστε να μπορούν οι υπολογιστές να επεξεργάζονται και να αναλύουν μεγάλες ποσότητες δεδομένων φυσικής γλώσσας. Οι τεχνολογίες NLP χρησιμοποιούνται σε διάφορες εφαρμογές, όπως η αναγνώριση κειμένου και ομιλίας, η μηχανική μετάφραση, η ανάλυση συναισθήματος και πολλές ακόμα. Η NLP έχει σημειώσει σημαντική πρόοδο τα τελευταία χρόνια με την πρόοδο της μηχανικής μάθησης και των τεχνικών βαθιάς μάθησης. Αυτές οι τεχνικές έχουν επιτρέψει στους υπολογιστές να εκτελούν διάφορες εργασίες NLP, όπως γλωσσική μετάφραση, περίληψη και αναγνώριση ονοματικών οντοτήτων με υψηλό επίπεδο ακρίβειας. Ωστόσο, η NLP εξακολουθεί να αποτελεί ένα πεδίο προκλήσεων, καθώς η ανθρώπινη γλώσσα είναι πολύπλοκη, διαφορούμενη και συνεχώς εξελισσόμενη. Ως εκ

τούτου, υπάρχει ακόμη πολλή δουλειά που πρέπει να γίνει για να μπορέσουν οι υπολογιστές να κατανοήσουν και να επεξεργαστούν πλήρως την ανθρώπινη γλώσσα.

A. BERT

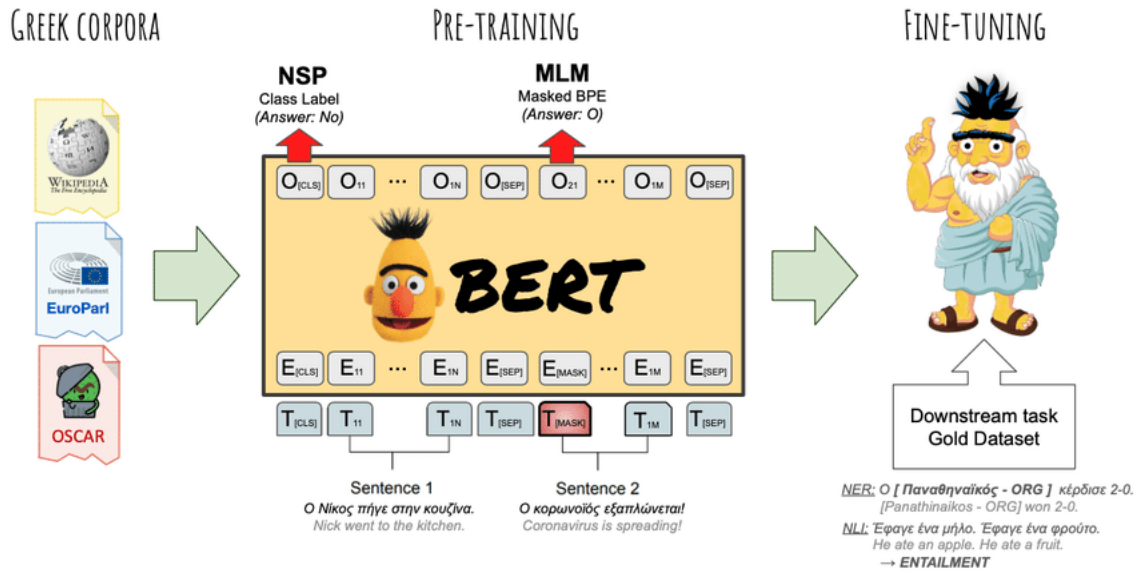
Η BERT (Bidirectional Encoder Representations from Transformers) είναι μια υπερσύγχρονη τεχνική επεξεργασίας φυσικής γλώσσας (NLP) που περιλαμβάνει την εκπαίδευση ενός μοντέλου μετασχηματιστή σε ένα μεγάλο σύνολο δεδομένων για την κατανόηση του γλωσσικού πλαισίου με τρόπο παρόμοιο με τον τρόπο που το κάνουν οι άνθρωποι. Το BERT μπορεί να ρυθμιστεί λεπτομερώς για ένα ευρύ φάσμα εργασιών επεξεργασίας φυσικής γλώσσας, όπως η απάντηση ερωτήσεων και η γλωσσική μετάφραση, με την προσθήκη ειδικών για την εργασία επιπέδων πάνω στο προ-εκπαιδευμένο μοντέλο. Έχει αποδειχθεί ότι αποδίδει πολύ καλά σε μια ποικιλία εργασιών NLP και έχει γίνει δημοφιλής επιλογή για τους επαγγελματίες NLP.

B. GreekBERT

Το GREEK-BERT [15] είναι ένα μονόγλωσσο γλωσσικό μοντέλο βασισμένο στο BERT για τη νέα ελληνική γλώσσα. Τα προ-εκπαιδευμένα γλωσσικά μοντέλα που βασίζονται σε μετασχηματιστές, και οι παραλλαγές τους, έχουν πετύχει κορυφαία αποτελέσματα σε διάφορες μεταγενέστερες εργασίες της επεξεργασίας της φυσικής γλώσσας (π.χ. ταξινόμηση κειμένου, φυσική γλώσσα, εξαγωγή συμπερασμάτων). Ενώ τα πολύγλωσσα μοντέλα παρέχουν εκπληκτικά καλές επιδόσεις στις zero-shot διαμορφώσεις (π.χ. Fine-tuning ενός προ-εκπαιδευμένου μοντέλου σε μια άλλη γλώσσα για την ίδια διαδικασία χωρίς περαιτέρω εκπαίδευση), τα μονόγλωσσα μοντέλα, όταν είναι διαθέσιμα εξακολουθούν να υπερτερούν στις περισσότερες μεταγενέστερες εργασίες, με εξαίρεση την μηχανική μετάφραση.

Το σύνολο των δεδομένων που χρησιμοποιήθηκαν στην δημιουργία του ελληνικού μοντέλου BERT προέρχεται από την Wikipedia, έγγραφα του Ευρωπαϊκού κοινοβουλίου και του Ελληνικού Oscar της Common Crawl [16]. Το Greek Bert εκπαιδεύεται εκ του μηδενός, γι' αυτό τα βάρη του αρχικοποιούνται τυχαία. Το μοντέλο προ-εκπαιδεύεται σε ζεύγος προτάσεων. Κάθε πρόταση, αναπαρίσταται ως μια ακολουθία μονάδων υποστήριξης (0-1) tokens, ενώ υπάρχουν δύο σύμβολα τα οποία υποδεικνύουν την αρχή και το τέλος της ακολουθίας. Μερικά Tokens και στις δύο προτάσεις αποκρύπτονται επί τούτου. Το BERT χωρίζεται σε δύο εργασίες με αυτό-επίβλεψη. Η πρώτη ονομάζεται Masked Language Modeling (MLM), όπου το μοντέλο προσπαθεί να μαντέψει τα tokens που έχουν αποκρυφτεί, με βάση μια αναπαράσταση που παράγεται στα τελευταία 12 επίπεδα. Η δεύτερη εργασία ονομάζεται 'Πρόβλεψη επόμενης πρότασης' και σε αυτή το μοντέλο χρησιμοποιεί την τελική ακολουθία του classification token για να προβλέψει αν η δεύτερη πρόταση ακολουθεί νοηματικά την πρώτη. Σε τελικό στάδιο, το

BERT μπορεί να χρησιμοποιηθεί και να εκπαιδευτεί σε διάφορες εργασίες, όπως η ταξινόμηση και η επισήμανση (labelling).



Σχήμα 3.8: Μοντέλο GreekBERT

3.7 Επίλογος

Σε αυτή την ενότητα έγινε μια αναφορά στην θεωρία πάνω στην οποία βασίστηκε η διπλωματική αυτή. αρχικά έγινε μια εισαγωγή στην μηχανική μάθηση και συγκεκριμένα στην μηχανική μάθηση με επίβλεψη. Έγινε αναλυτική περιγραφή της διαδικασίας της ταξινόμησης καθώς και η περιγραφή συγκεκριμένων αλγορίθμων ταξινόμησης που χρησιμοποιήθηκαν. Έπειτα, έγινε αναφορά στα νευρωνικά δίκτυα, και συγκεκριμένα με αρχιτεκτονική ενός στρώματος και πολλών στρωμάτων. Καταγράφηκαν επίσης κάποιες τεχνικές των νευρωνικών δικτύων που χρησιμοποιήθηκαν όπως τα πυκνά στρώματα, η κανονικοποίηση και το dropout. Τέλος, αναλύθηκε ο τομέας της επεξεργασίας της φυσικής γλώσσας και τα μοντέλα της.

Κεφάλαιο 4ο: Τεχνολογικό Υπόβαθρο

4.1 Εισαγωγή

Οι βιβλιοθήκες είναι σημαντικές για τις εφαρμογές μηχανικής μάθησης για διάφορους λόγους:

1. Παρέχουν προσχεδιασμένο και βελτιστοποιημένο κώδικα: Οι βιβλιοθήκες μηχανικής μάθησης παρέχουν προ-γραμμένο και βελτιστοποιημένο κώδικα για κοινές εργασίες, όπως πράξεις γραμμικής άλγεβρας, αλγόριθμους βελτιστοποίησης και επίπεδα νευρωνικών δικτύων, γεγονός που μπορεί να εξοικονομήσει πολύ χρόνο και προσπάθεια στους προγραμματιστές.
2. Βελτιώνουν τις επιδόσεις: Οι βιβλιοθήκες μηχανικής μάθησης είναι συνήθως βελτιστοποιημένες για απόδοση, πράγμα που σημαίνει ότι ο κώδικας εκτελείται ταχύτερα και αποτελεσματικότερα. Αυτό είναι σημαντικό για έργα μηχανικής μάθησης μεγάλης κλίμακας που απαιτούν την επεξεργασία μεγάλου όγκου δεδομένων.
3. Διευκολύνουν τον πειραματισμό: Οι βιβλιοθήκες μηχανικής μάθησης παρέχουν ένα ευρύ φάσμα εργαλείων και λειτουργιών για τη δημιουργία, την εκπαίδευση και την ανάπτυξη μοντέλων μηχανικής μάθησης, γεγονός που διευκολύνει τους προγραμματιστές να πειραματιστούν με διαφορετικές προσεγγίσεις και τεχνικές.
4. Παρέχουν μια κοινότητα: Οι βιβλιοθήκες μηχανικής μάθησης είναι συχνά ανοικτού κώδικα και διαθέτουν μεγάλες κοινότητες προγραμματιστών και χρηστών. Αυτό παρέχει πρόσβαση σε πληθώρα γνώσεων και πόρων, όπως τεκμηρίωση, σεμινάρια και δείγματα κώδικα.
5. Επιτρέπουν την αναπαραγωγιμότητα: Αυτό είναι σημαντικό για την επιστημονική διαδικασία και θα μπορούσε να συμβάλει στην αύξηση της αξιοπιστίας των ευρημάτων.

4.2 Python

Ως γλώσσα προγραμματισμού στην εργασία αυτή επιλέχθηκε η Python 3 [17]. Η πρακτική εφαρμογή της Python σε έργα και εργασίες μηχανικής μάθησης έχει διευκολύνει την εργασία των προγραμματιστών, των επιστημόνων δεδομένων και των τεχνικών μηχανικής μάθησης. Η Python μπορεί εύκολα να χρησιμοποιηθεί για την ανάλυση και τη σύνθεση των διαθέσιμων δεδομένων, γεγονός που την καθιστά επίσης μία από τις πιο δημοφιλείς γλώσσες στην επιστήμη των δεδομένων. Η πλούσια εγγενής επέκταση ενισχύει επίσης τα πλεονεκτήματα της Python, η οποία είναι πιο κατάλληλη για μηχανική μάθηση, λογιστική δεδομένων κ.λπ. [18] Τα πλεονεκτήματα της χρήσης της Python σε εφαρμογές της μηχανικής μάθησης:

1. Ποικιλία σε βιβλιοθήκες και frameworks
2. Σταθερότητα και απλότητα
3. Ανεξαρτησία μεταξύ πλατφορμών
4. Υποστηρικτική κοινότητα

4.3 PyTorch

Η PyTorch [19] είναι μια δημοφιλής βιβλιοθήκη μηχανικής μάθησης ανοικτού κώδικα για την Python που παρέχει API υψηλού επιπέδου για την κατασκευή και την εκπαίδευση μοντέλων βαθιάς μάθησης. Έχει σχεδιαστεί ώστε να είναι ευέλικτη και αρθρωτή, επιτρέποντας στους προγραμματιστές να δημιουργούν και να πειραματίζονται εύκολα με διαφορετικά μοντέλα. Το PyTorch περιλαμβάνει μια ποικιλία προ-δημιουργημένων ενοτήτων για εργασίες όπως η επεξεργασία φυσικής γλώσσας, η όραση υπολογιστών και η πρόβλεψη χρονοσειρών, καθιστώντας το ένα ισχυρό εργαλείο για ένα ευρύ φάσμα εφαρμογών μηχανικής μάθησης. Επιπλέον, το PyTorch διαθέτει ισχυρή υποστήριξη για κατανεμημένη εκπαίδευση, επιτρέποντας στους προγραμματιστές να εκπαιδεύουν μεγάλα μοντέλα σε πολλαπλές GPU και στο cloud.

4.4 Scikit Learn

Η Scikit-learn [20] είναι μια δημοφιλής βιβλιοθήκη μηχανικής μάθησης για την Python που παρέχει μια σειρά εργαλείων για εργασίες όπως η ταξινόμηση, η παλινδρόμηση, η ομαδοποίηση και η μείωση διαστάσεων. Είναι χτισμένη πάνω στις NumPy [21] και SciPy [22], δύο βιβλιοθήκες για επιστημονικούς υπολογισμούς στην Python, και έχει σχεδιαστεί για να είναι εύκολη στη χρήση και την ενσωμάτωση με άλλες βιβλιοθήκες. Η Scikit-learn διαθέτει έναν αριθμό καλά τεκμηριωμένων APIs και μια μεγάλη κοινότητα χρηστών, καθιστώντας το μια καλή επιλογή για πολλές εργασίες μηχανικής μάθησης. Είναι επίσης ικανή να χειρίζεται αποτελεσματικά μεγάλα σύνολα δεδομένων και διαθέτει υποστήριξη για κατανεμημένους υπολογισμούς, καθιστώντας τη ένα ισχυρό εργαλείο για την εργασία με μεγάλα δεδομένα.

4.5 Beautiful Soup

Η Beautiful Soup [21] είναι μια βιβλιοθήκη της Python για την εξαγωγή δεδομένων από αρχεία HTML και XML. Δημιουργεί δέντρα ανάλυσης που είναι χρήσιμα για την εύκολη εξαγωγή των δεδομένων. Βρίσκεται πάνω σε έναν αναλυτή HTML ή XML, παρέχοντας Pythonic idioms

για την επανάληψη, την αναζήτηση και την τροποποίηση του parse tree. Η Beautiful Soup παρέχει αρκετές απλές μεθόδους και ιδιωματισμούς για την περιήγηση, την αναζήτηση και την τροποποίηση ενός δέντρου ανάλυσης, και στηρίζεται πάνω στις δημοφιλείς lxml και html5lib, επιτρέποντάς να δοκιμαστούν διαφορετικές στρατηγικές ανάλυσης ή να επιλεγθεί είτε η ταχύτητα ή η ευελιξία ανάλογα με το εκάστοτε πρόβλημα.

4.6 DeepL

Η DeepL [22] είναι μια υπηρεσία μετάφρασης. Από το αρχικό λανσάρισμα της DeepL το 2017, αναπτύσσεται μια νέα γενιά νευρωνικών δικτύων (NN). Χρησιμοποιώντας έναν καινοτόμο σχεδιασμό νευρωνικών δικτύων, τα δίκτυα DeepL μαθαίνουν να αντιλαμβάνονται τις λεπτές έννοιες των προτάσεων και να τις μεταφράζουν σε μια γλώσσα-στόχο με πρωτοφανή τρόπο. Αυτό έχει οδηγήσει σε μια παγκοσμίως γνωστή ποιότητα μηχανικής μετάφρασης που ξεπερνά όλες τις μεγάλες εταιρείες τεχνολογίας. Τα νευρωνικά δίκτυα του DeepL είναι σε θέση να συλλάβουν ακόμη και τις παραμικρές διαφορές και να τις αναπαράγουν στη μετάφραση σε αντίθεση με οποιαδήποτε άλλη υπηρεσία. Για να γίνει αξιολόγηση της ποιότητας των μοντέλων αυτόματης μετάφρασης, διεξάγονται τακτικά τυφλές δοκιμές. Στις τυφλές δοκιμές, οι επαγγελματίες μεταφραστές επιλέγουν την πιο ακριβή μετάφραση χωρίς να γνωρίζουν ποια εταιρεία την παρήγαγε. Το DeepL ξεπερνά τον ανταγωνισμό κατά 3:1.

4.7 NLTK

Το Natural Language Toolkit (NLTK) είναι μια βιβλιοθήκη Python για την εργασία με δεδομένα ανθρώπινης γλώσσας. Παρέχει ένα ευρύ φάσμα εργαλείων για εργασίες όπως tokenization, stemming και tagging, καθώς και εργαλεία για την επεξεργασία και ανάλυση γλωσσικών δεδομένων. Το NLTK περιλαμβάνει επίσης μια σειρά από σύνολα δεδομένων και πόρους που βασίζονται σε κείμενο, όπως το σώμα δεδομένων του Πανεπιστημίου Brown, τα οποία μπορούν να χρησιμοποιηθούν για την εκπαίδευση και την αξιολόγηση γλωσσικών μοντέλων. Συνολικά, το NLTK είναι ένα ισχυρό εργαλείο για την επεξεργασία φυσικής γλώσσας και χρησιμοποιείται ευρέως στον τομέα αυτό.

4.8 FastAPI

Το FastAPI είναι ένα σύγχρονο, γρήγορο, framework ιστού για την δημιουργία API μέσω της γλώσσας Python 3.7+ που βασίζεται σε τυποποιημένες υποδείξεις τύπου Python. Είναι χτισμένο πάνω στο framework Starlette για τα τμήματα του ιστού και στο framework Pydantic για

τα τμήματα των δεδομένων. Ουσιαστικά, το FastAPI μπορεί να δημιουργήσει αυτόματη τεκμηρίωση για το API και να επικυρώσει τα εισερχόμενα δεδομένα. Αυτό καθιστά εύκολη τη δημιουργία API που είναι προβλέψιμα, αξιόπιστα και εύχρηστα.

Χρησιμοποιείται κυρίως ως ένας τρόπος επικοινωνίας μεταξύ διαφορετικών εφαρμογών. Τα API χρησιμοποιούνται συνήθως στην ανάπτυξη ιστοσελίδων για να επιτρέπουν σε διαφορετικά μέρη ενός ιστότοπου ή μιας εφαρμογής ιστού να μοιράζονται δεδομένα και λειτουργίες.

4.9 Unicorn

Το Unicorn είναι ένα ασύγχρονο πλαίσιο διακομιστή ιστού υψηλής απόδοσης για την Python 3.7+ που βασίζεται στην πρότυπη βιβλιοθήκη `asyncio`. Είναι χτισμένο πάνω στο `Starlette` για τα τμήματα ιστού και το `Pydantic` για την επικύρωση δεδομένων. Παρέχει μια γρήγορη, εύχρηστη και αποτελεσματική μέθοδο για την εξυπηρέτηση APIs και εφαρμογών ιστού. Το Unicorn προσφέρει πρώτης τάξεως υποστήριξη για `WebSockets`, γρήγορο χειρισμό αιτήσεων/απαντήσεων και ένα μοντέλο ασύγχρονης εκτέλεσης που το καθιστά κατάλληλο για εφαρμογές ιστού και μικροπηρεσίες υψηλής απόδοσης.

Επιπροσθέτως, το Unicorn έχει σχεδιαστεί για να είναι εύκολο στη χρήση και την ενσωμάτωση με άλλες βιβλιοθήκες Python. Διαθέτει ένα ελάχιστο και απλό API που διευκολύνει τη συγγραφή και τη συντήρηση της εφαρμογής σας. Το πλαίσιο περιλαμβάνει επίσης χαρακτηριστικά όπως η αυτόματη δημιουργία τεκμηρίωσης, η ανίχνευση αιτήσεων και η υποστήριξη για δοκιμές και αποσφαλμάτωση. Το Unicorn μπορεί να εκτελεστεί αυτόνομα ή πίσω από έναν αντίστροφο διακομιστή μεσολάβησης, καθιστώντας το μια ευέλικτη επιλογή για ένα ευρύ φάσμα σεναρίων ανάπτυξης. Διαθέτει μια αυξανόμενη κοινότητα χρηστών και συνεργατών και έχει υιοθετηθεί ευρέως από τους προγραμματιστές λόγω της απλότητας, της ταχύτητας και της αξιοπιστίας του.

4.10 Επίλογος

Σε αυτή την ενότητα θα αναφέρθηκαν επιγραμματικά οι βιβλιοθήκες και τα frameworks τα οποία αποτέλεσαν τα εργαλεία για την πραγματοποίηση αυτής της διπλωματικής.

Κεφάλαιο 5ο: Σύνολο Δεδομένων

5.1 Εισαγωγή

Σε αυτό το κεφάλαιο θα γίνει αναφορά στο σύνολο δεδομένων που δημιουργήθηκε και χρησιμοποιήθηκε για την εξαγωγή των αποτελεσμάτων, καθώς και τα προβλήματα τα οποία βρέθηκαν κατά την επεξεργασία τους και οι τρόποι αντιμετώπισης τους.

5.2 Γενικά

Τα δεδομένα είναι μη οργανωμένα και μη επεξεργασμένα γεγονότα σχετικά με αντικείμενα η συμβάντα του πραγματικού κόσμου. Τα δεδομένα είναι στατικά, δηλαδή από τη στιγμή που θα καταγραφούν δεν αλλάζουν. Από μόνα τους τα δεδομένα, χωρίς ένα πλαίσιο αναφοράς, δεν είναι ιδιαίτερα χρήσιμα, δηλαδή δεν μεταφέρουν πληροφορία. Με τη σειρά τους, τα γεγονότα (facts) είναι δηλώσεις (statements) περί της αλήθειας σχετικά με τις ιδιότητες κάποιου αντικειμένου ή της συσχέτισης του με άλλα αντικείμενα. [14]

Η πληροφορία χρειάζεται ένα εννοιολογικό πεδίο που να επιτρέπει την ερμηνεία της. Γενικά, η πληροφορία αποτελείται από δεδομένα τα οποία όμως έχουν φιλτραριστεί και μορφοποιηθεί κατάλληλα. Η πληροφορία, σε σύγκριση με τα δεδομένα, έχει νόημα, σκοπό και συνάφεια, και έτσι μπορεί να ερμηνεύεται και να διευκολύνει στην λήψη αποφάσεων.

Η γνώση, με τη σειρά της, είναι πληροφορία η οποία έχει υποστεί μία σειρά ειδικών ελέγχων για την πιστοποίηση της, όπως για παράδειγμα η επιστημονική γνώση ή η γνώση που προέρχεται από την μακρόχρονη επιβεβαίωση των καθημερινών εμπειριών (ευρετική / εμπειρική γνώση). Τέλος, πολλοί ερευνητές προσθέτουμε και την έννοια της σοφίας, ως την ικανότητα να χρησιμοποιεί κάποιος την γνώση όσο το δυνατόν αποδοτικότερα, δηλαδή να γνωρίζει το τμήμα της γνώσης που θα πρέπει να χρησιμοποιηθεί, ανάλογα με την περίπτωση. Στην συγκεκριμένη διπλωματική τα δεδομένα αποτελούνται από 2 χαρακτηριστικά, τον τίτλο της είδησης, label: title, και το is_fake, όπου 1 - Ψευδές, και 0 - Αληθές

5.3 Διαδικασία Απόξεσης Ιστού

Μία από τις μεθόδους που χρησιμοποιήθηκαν για την συλλογή των δεδομένων που θα εκπαιδεύσουν το μοντέλο είναι η απόξεση ιστού (web scraping). Το web scraping αναφέρεται στη διαδικασία αυτόματης εξαγωγής δεδομένων από ιστότοπους με χρήση λογισμικού. Μια τεχνο-

λογία που μας επιτρέπει να εξάγουμε δομημένα δεδομένα από κείμενο σε μορφή HTML. Είναι εξαιρετικά χρήσιμη σε περιπτώσεις όπου τα δεδομένα δεν παρέχονται σε αναγνώσιμη μορφή, όπως JSON ή XML.

5.4 Ελληνικό Σύνολο Δεδομένων

Το πρώτο σύνολο δεδομένων "Greek News" δημιουργήθηκε εκ του μηδενός και περιλαμβάνει 3255 εγγραφές.

5.4.1 Greek Hoaxes

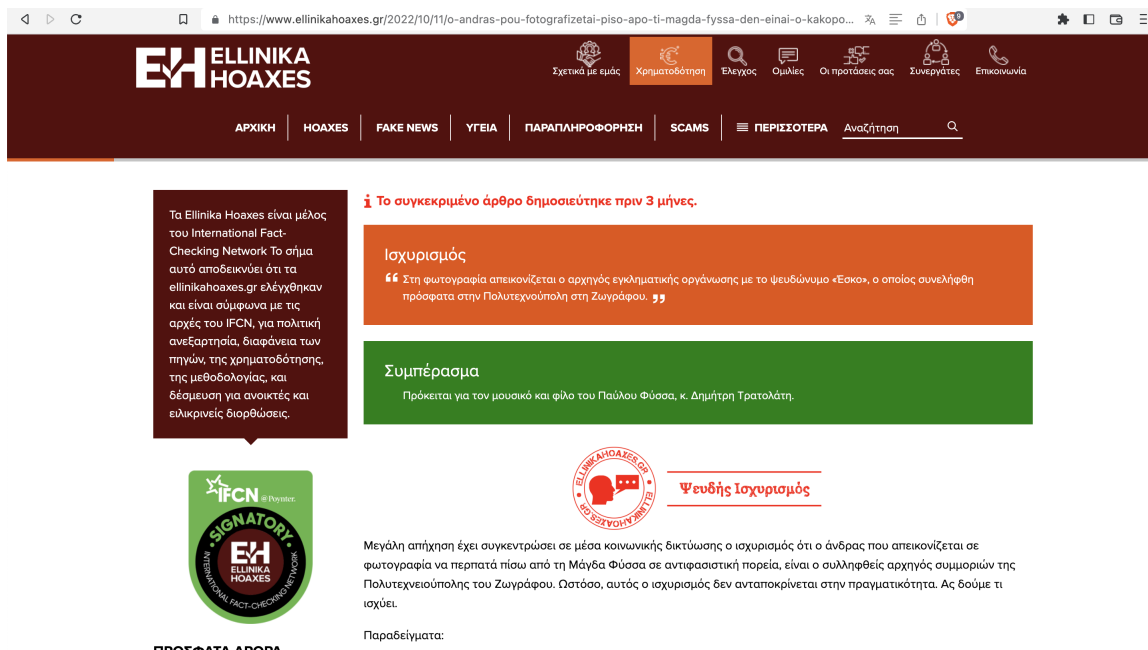
Η σελίδα που επιλέχθηκε πρώτη είναι αυτή των Greek Hoaxes (<https://www.ellinikahoaxes.gr/>) [8], από εδώ συλλέχθηκαν τα νέα που ανήκουν στην κατηγορία των ψευδών ειδήσεων.

Η πλατφόρμα επιλέχθηκε καθώς προσφέρει υλικό βασισμένο στην ελληνική ειδησεογραφία, είναι η μόνη με επαρκή στοιχεία για την εγκυρότητα της και διαθέτει δεδομένα συγκεντρωμένα έτσι ώστε να γίνει μαζική περισυλλογή.

Η διαδικασία που ακολουθήθηκε για την απόξεση των δεδομένων είναι η εξής. Αρχικά επιλέχθηκαν οι κατηγορίες της ιστοσελίδας hoaxes¹, fake news, παραπληροφόρηση και scam². Από εκεί ο γενικός τίτλος των άρθρων δεν είναι αντιπροσωπευτικός ως ψευδής είδηση, γιατί μεταφέρεται η άρνηση της ψευδής είδησης, με στόχο την καταγραφή της πραγματικής συνθήκης. Για παράδειγμα, ο τίτλος του άρθρου είναι "Όχι, το πρόσωπο του Μέσι ΔΕΝ θα τυπωθεί σε Αργεντίνικο χαρτονόμισμα" κάτι που δεν θα ήταν χρήσιμο στην δημιουργία του συνόλου δεδομένων. Αντίθετα, επιλέχθηκε εντός του άρθρου η παράγραφος "ισχυρισμός". Στο παράδειγμα που χρησιμοποιήθηκε ο ισχυρισμός είναι "Η Κεντρική Τράπεζα της Αργεντινής εξετάζει σοβαρά το ενδεχόμενο να τοποθετηθεί το πρόσωπο του ποδοσφαιριστή Λιονέλ Μέσι στο χαρτονόμισμα των χιλίων πέσος της χώρας."

¹η φάρσα, κάτι που γίνεται για να παραπλανησουμε τους άλλους και να θεωρήσουν ένα στημένο γεγονός ή μια πλαστή ιστορία ως αληθινά

²η απάτη



Σχήμα 5.1: Ιστοσελίδα Greek Hoaxes

5.4.2 Η Καθημερινή

Στη συνέχεια για την περισυλλογή των αληθών ειδήσεων χρησιμοποιήθηκε η online πλατφόρμα της εφημερίδας "Η Καθημερινή" [23], και η εξαγωγή έγινε με την μέθοδο του web scraping. Επιλέχθηκε διότι είναι η δεύτερη σε σειρά ηλεκτρονική σελίδα που εμπιστεύεται και επισκέπτεται η Ελληνική Κοινωνία με βάση τον δείκτη brand trust score στην έρευνα του Ινστιτούτου Reuters [24]. Για να υπάρχει μεγαλύτερη ποικιλία στο είδος των ειδήσεων οι κατηγορίες που επιλέχθηκαν για την συλλογή των τίτλων είναι οι εξής: Πολιτική, Οικονομία, Κοινωνία, Κόσμος, Κουλτούρα. Δόθηκε ιδιαίτερη προσοχή στην ημερομηνία δημοσίευσης των τίτλων, λήφθηκαν δείγματα σε βάθος τριετίας έτσι ώστε να αποφευχθεί η συγκέντρωση πολλών ειδήσεων με κοινή θεματική. Παράδειγμα προς αποφυγή ήταν πλήθος πρόσφατων δημοσιευμάτων που είχαν ως κύριο θέμα το "QatarGate".

Οι τίτλοι που συλλέχθηκαν από την Καθημερινή και από τα Ελληνικά Hoaxes συντάχθηκαν όλα μαζί και αποτελούν το σύνολο δεδομένων με όνομα "Greek News".

5.5 Μεταφρασμένο Σύνολο Δεδομένων

5.5.1 Liar Dataset

Ένα από τα ευρέως διαδεδομένα σύνολα δεδομένων πάνω στην κατηγορία των ψευδών ειδήσεων είναι το "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection"

[24], για συντομία "Liar".

Το LIAR είναι ένα δημόσια διαθέσιμο σύνολο δεδομένων για την ανίχνευση ψευδών ειδήσεων. Μια δεκαετία συλλογής 12,8 χιλιάδων χειροκίνητα επισημασμένων σύντομων δηλώσεων, τα οποία συλλέχθηκαν από το POLITIFACT.COM. Η πλατφόρμα αυτή παρέχει λεπτομερή έκθεση ανάλυσης και συνδέσμους σε έγγραφα πηγής για κάθε περίπτωση. Αυτό το σύνολο δεδομένων μπορεί επίσης να χρησιμοποιηθεί για έρευνα ελέγχου των πραγματικών γεγονότων. Αξίζει να σημειωθεί ότι αυτό το νέο σύνολο δεδομένων είναι μια τάξη μεγέθους μεγαλύτερο από τα προηγούμενα μεγαλύτερα δημόσια σύνολα δεδομένων ψευδών ειδήσεων παρόμοιου τύπου. Το σύνολο δεδομένων LIAR4 περιλαμβάνει 12,8 χιλιάδες σύντομες δηλώσεις με ανθρώπινη επισημάνση από το API του POLITIFACT.COM και κάθε δήλωση αξιολογείται από έναν συντάκτη του POLITIFACT.COM ως προς την αληθοφάνειά της.

Καθώς το dataset είναι γραμμένο στην Αγγλική γλώσσα, για την χρήση του στην συγκεκριμένη διπλωματική χρειάστηκε να μεταφραστούν όλοι οι τίτλοι από τα αγγλικά στα ελληνικά. Η διαδικασία αυτή έγινε με την βοήθεια του DeepL framework. Η κατηγοριοποίηση των ειδήσεων στο Liar: 'true': αληθές, 'mostly-true': κυρίως αληθές, 'half-true': μισή αλήθεια, 'mostly-false': κυρίως ψευδές, 'false': ψευδές and 'pants-fire': εντελώς ψευδές. Καθώς αυτό το φάσμα ήταν ευρύ και στην συγκεκριμένη διπλωματική ακολουθήθηκε δυαδική προσέγγιση (binary approach), δηλαδή κατηγοριοποίηση σε δύο καταστάσεις, ψευδής και αληθής, έγινε η αναγωγή των έξι καταστάσεων σε δυο.

Πιο συγκεκριμένα τα "αληθές", "κυρίως αληθές", "μισή αλήθεια" ως "αληθές", και τα "κυρίως ψευδές", "ψευδές" και "εντελώς ψευδές" ως "ψευδές".

5.5.2 Μέθοδος Μετάφρασης

Η DeepL είναι μια υπηρεσία μετάφρασης που προσφέρει και δυναμική ιστοσελίδα αλλά και διεπαφή προγραμματισμού εφαρμογών (API). Στην παρούσα διπλωματική η διεπαφή DeepL χρησιμοποιήθηκε ώστε να μεταφραστεί το σύνολο δεδομένων Liar από τα αγγλικά στα ελληνικά. Η διαδικασία περιπλέχθηκε καθώς η πλατφόρμα έβαζε πλαφόν δωρεάν μεταφρασμένων χαρακτήρων στους 500.000, από τους συνολικούς 1.642.874 του dataset. Για να μειωθεί ο αριθμός των χαρακτήρων, δημιουργήθηκε μια μέθοδος με την οποία αντικαταστάθηκαν μεγαλύτερου μήκους λέξεις, με κάτι μικρότερο

And σε &
Percentage σε %
Dollars σε \$

5.6 Επίλογος

Σε αυτή την ενότητα έγινε μια αναλυτική περιγραφή της συλλογής και της δημιουργίας δυο συνόλων δεδομένων. Το ελληνικό (greek.csv) και το μεταφρασμένο (translated.csv). Στο μοντέλο όμως φορτώνονται τρία σύνολα. Το κάθε ένα που αναφέρθηκε ξεχωριστά, και ένα ακόμα που δημιουργήθηκε από τον συνδυασμό τους (mixed.csv).

Κεφάλαιο 6ο: Μεθοδολογία

6.1 Εισαγωγή

Σε αυτό το κεφάλαιο γίνεται καταγραφή των βημάτων που ακολουθήθηκαν για την δημιουργία του μοντέλου νευρωνικού δικτύου, καθώς και η δημιουργία .

6.2 Δημιουργία Μοντέλου

Προεργασία

Καθώς το μοντέλο βασίζεται πάνω στο προ-εκπαιδευμένο (pretrained) του ελληνικού BERT, φορτώνεται στον κώδικα ο tokenizer και το μοντέλο του. Έπειτα, φορτώνεται το dataset greek.csv και αρχίζει η διαδικασία του pre-processing. Αυτό περιλαμβάνει την προετοιμασία των δεδομένων κειμένου. Πρόκειται για το προκαθορισμένο πρώτο βήμα των έργων πάνω στην επεξεργασία της φυσικής γλώσσας. Ορισμένα από τα βήματα προεπεξεργασίας είναι τα εξής:

1. Αφαίρεση των σημείων στίξης
2. Αφαίρεση διευθύνσεων URL
3. Μετατροπή σε πεζά γράμματα
4. Αφαίρεση λέξεων διακοπής (stop words)
5. Οι λέξεις διακοπής είναι ουσιαστικά μια λίστα με λέξεις οι οποίες αφαιρούνται καθώς δεν προσφέρουν ουσιαστική πληροφορία στον αλγόριθμο. Για παράδειγμα η πρόταση ” Ποιες είναι οι λέξεις διακοπής σε αυτή την πρόταση;” θα γίνει ”λέξεις διακοπής πρόταση”. Αυτή η διαδικασία βοηθάει έτσι ώστε να μην αποπροσανατολίζεται ο αλγόριθμος και να βρεθεί η σωστή αναλογία ανάμεσα στις λέξεις που χρειάζονται και η διατήρηση του νοήματος. Στην συγκεκριμένη περίπτωση, καθώς η βιβλιοθήκη nltk περιέχει περιορισμένες λέξεις που προέρχονται από την αρχαία ελληνική γλώσσα έγινε προσθήκη μιας λίστας με λέξεις της νέας ελληνικής που θεωρήθηκαν κατάλληλες.
6. Λημματοποίηση (Lemmatizer) - Χρησιμοποιήθηκε μόνο με τους γνωστούς ταξινομητές. Η λημματοποίηση είναι ένας γλωσσολογικός όρος που εκφράζει την ομαδοποίηση λέξεων με την ίδια ρίζα ή λήμμα αλλά με διαφορετικές κλίσεις ή παράγωγα της σημασίας, ώστε να μπορούν να αναλυθούν ως ένα στοιχείο. Ο στόχος είναι να αφαιρεθούν τα επιθέματα και τα προθέματα της κλίσης για να αναδειχθεί η λεξικογραφική μορφή της λέξης. Για παράδειγμα η πρόταση ”Τα μάτια της κουζίνας” μετά από λημματοποίηση θα γίνει ”Τα μάτι της κουζίνα”.

7. Μέτρηση Βαρύτητας Λέξεων - Χρησιμοποιήθηκε μόνο με τους γνωστούς ταξινομητές. Το TF-IDF (term frequency-inverse document frequency) είναι ένα στατιστικό μέτρο που χρησιμοποιείται για την αξιολόγηση της σημασίας μιας λέξης σε ένα έγγραφο ή σε μια συλλογή εγγράφων. Χρησιμοποιείται συχνά σε εργασίες επεξεργασίας φυσικής γλώσσας, ανάκτησης πληροφοριών και εξόρυξης κειμένου. Η ιδέα πίσω από το TF-IDF είναι ότι μια λέξη που εμφανίζεται συχνά σε ένα έγγραφο, αλλά όχι σε πολλά έγγραφα εντός του σώματος, είναι πιθανό να είναι πιο σημαντική για το νόημα του εγγράφου από μια λέξη που εμφανίζεται συχνά σε πολλά έγγραφα. Η συνιστώσα TF του TF-IDF είναι η συχνότητα των όρων, δηλαδή ο αριθμός των φορών που εμφανίζεται ένας συγκεκριμένος όρος σε ένα έγγραφο. Όσο υψηλότερη είναι η συχνότητα του όρου, τόσο πιο σημαντικός θεωρείται ο όρος στο συγκεκριμένο έγγραφο. Για παράδειγμα, εάν η λέξη "σκύλος" εμφανίζεται 10 φορές σε ένα έγγραφο, θα έχει υψηλότερη συχνότητα όρου από τη λέξη "γάτα" που εμφανίζεται μόνο 5 φορές. Η συνιστώσα IDF του TF-IDF είναι η αντίστροφη συχνότητα εγγράφου, η οποία είναι ένα μέτρο του πόσο σπάνιος είναι ο όρος σε ολόκληρο το σώμα κειμένων. Όσο υψηλότερη είναι η αντίστροφη συχνότητα εγγράφου, τόσο πιο σπάνιος θεωρείται ο όρος. Υπολογίζεται με τη διαίρεση του λογαρίθμου του αριθμού των εγγράφων στο σώμα κειμένων με τον αριθμό των εγγράφων που περιέχουν τον όρο.

Χωρισμός Συνόλου Δεδομένων

Η διαδικασία διαχωρισμού του συνόλου δεδομένων γίνεται στην προσπάθεια βελτιστοποίησης του μοντέλου. Ο διαχωρισμός σε εκπαίδευση, δοκιμή και επικύρωση (train, test, valid) είναι μια τεχνική που αξιολογεί την απόδοση του μοντέλου μηχανικής μάθησης, είτε ταξινόμησης είτε οπισθοδρόμησης. Το σύνολο της εκπαίδευσης (train), είναι τα δεδομένα με τα οποία εκπαιδεύεται το μοντέλο και δημιουργούνται οι αρχικές παράμετροι. Το σύνολο της επικύρωσης (valid), είναι τα δεδομένα που χρησιμοποιούνται για να δοθεί μια μη προκατειλημμένη αξιολόγηση που έχει δημιουργηθεί με το σετ της εκπαίδευσης, κάνοντας τελειοποιήσεις στις υπερ-παραμέτρους του μοντέλου. Ταυτόχρονα, παίζει ρόλο και σε άλλες ενέργειες του μοντέλου όπως η επιλογή χαρακτηριστικών, ή η επιλογή αποκοπής κατωφλίου. Το σύνολο της δοκιμής (test), είναι η τελική αξιολόγηση του μοντέλου καθώς με βάση αυτό δημιουργούνται τα στατιστικά που το καθορίζουν επιτυχές ή μη.

Στον κώδικα, για το διαχωρισμό των δύο στην συγκεκριμένη περίπτωση σετ δεδομένων χρησιμοποιείται η έτοιμη μέθοδος που δίνεται από την βιβλιοθήκη scikit-learn. Τα δεδομένα διαχωρίζονται τυχαία στα δυο σε ποσοστό 70% εκπαίδευση 30% δοκιμή. Με την παράμετρο stratify με βάση την λίστα labels βεβαιωνόμαστε ότι και τα δύο σύνολα που δημιουργήθηκαν θα έχουν το ίδιο ποσοστό ψευδών και αληθών ειδήσεων.

Όριο Μεγίστων Λέξεων

Καθώς στην διαδικασία της κωδικοποίησης λέξεων(tokenization) θα δημιουργηθούν ακολουθίες συγκεκριμένου μήκους χαρακτήρων θα πρέπει να ορισθεί ένα όριο μεγίστων λέξεων. Αυτό καθορίζεται χειροκίνητα με βάση το ιστόγραμμα αριθμού λέξεων ανά τίτλο.

Κωδικοποίηση λέξεων(Tokenization)

Το μοντέλο n - gram είναι ένα γλωσσικό μοντέλο για τον εντοπισμό και την ανάλυση των χαρακτηριστικών που χρησιμοποιούνται στη μοντελοποίηση και επεξεργασία φυσικής γλώσσας. Υπάρχουν δυο γνωστότερα είδη, κωδικοποίηση χαρακτήρων και κωδικοποίηση λέξεων [25]. Στην συγκεκριμένη περίπτωση χρησιμοποιούνται n-grams λέξεων. Σε πολύ απλοποιημένη μορφή "Αυτή η πρόταση" μετά από την διαδικασία της κωδικοποίησης θα αποθηκευτεί στο σύστημα ως "Αυτή", "η", "πρόταση". Έπειτα, ο κάθε τίτλος κωδικοποιείται και κάθε λέξη αντιστοιχίζεται και αντικαθίσταται από έναν αριθμό. Κάθε στοιχείο του συνόλου δεδομένων είναι πλέον μια ακολουθία αριθμών με μέγεθος το όριο που ορίσαμε παραπάνω. Στην περίπτωση που ο τίτλος έχει μικρότερο μέγεθος από το όριο τότε αυτός γεμίζεται με μηδενικά μέχρι να συμπληρωθεί ο απαιτούμενος αριθμός στοιχείων ώστε να φτάσει το όριο.

Τα Attention Mask είναι μια ακολουθία 0 και 1, όπου οι θέσεις που βρίσκονται οι 1, ταυτίζονται με τις θέσεις που προηγουμένως υπήρχαν λέξεις στις προτάσεις. Ενώ τα 0, είναι τα γεμίσματα που μπήκαν στην διαδικασία της κωδικοποίησης. Ιδανικά θα λέγαμε ότι είναι ένας χάρτης που βοηθάει το μοντέλο να εντοπίσει γρηγορότερα πιο μέρος της ακολουθίας περιέχει την χρήσιμη για αυτό πληροφορία. Ένα παράδειγμα:

```
sentences = ["It will rain in the",
              "I want to eat a big bowl of",
              "My dog is"]
```

Η πρόταση μετά την κωδικοποίηση και την συμπλήρωση των μηδενικών μέχρι το όριο.:

```
input_ids = [
    [50256, 50256, 50256, 1026, 481, 6290, 287, 262],
    [ 40, 765, 284, 4483, 257, 1263, 9396, 286],
    [50256, 50256, 50256, 50256, 50256, 3666, 3290, 318]
]
```

```
attention_mask = [  
    [0, 0, 0, 1, 1, 1, 1, 1],  
    [1, 1, 1, 1, 1, 1, 1, 1],  
    [0, 0, 0, 0, 0, 1, 1, 1]  
]
```

Καθώς δημιουργείται ένα νέο μοντέλο βασισμένο στο BERT, είναι επιθυμητή η δημιουργία καινούργιων παραμέτρων. Για το λόγο αυτό γίνεται πάγωμα των ήδη υπάρχοντων παραμέτρων του BERT. Χωρίς αυτό, το BERT κατά τη διάρκεια της εκπαίδευσης θα επανεκπαιδευόταν και θα άλλαζαν οι παράμετροι του, κάτι το οποίο δεν είναι επιθυμητό στην προκειμένη περίπτωση.

Πάγωμα παραμέτρων του GreekBERT

Είναι σύνηθες να "παγώνουν" οι παράμετροι του προ-εκπαιδευμένου μοντέλου κατά την αρχική διαδικασία εκπαίδευσης. Αυτό σημαίνει ότι τα βάρη του προ-εκπαιδευμένου μοντέλου δεν θα ενημερώνονται κατά τη διάρκεια της εκπαίδευσης, μόνο τα νεοπροστιθέμενα στρώματα θα έχουν εκπαιδευσιμες παραμέτρους που θα ενημερώνονται. Ο λόγος που γίνεται αυτό είναι για να διατηρηθεί η γνώση του προ-εκπαιδευμένου μοντέλου, ενώ παράλληλα να επιτραπεί στο νέο μοντέλο να μάθει πληροφορίες που αφορούν την εργασία.

Τένσορες

Καθώς η PyTorch λειτουργεί με τένσορες, οι κωδικοποιημένοι τίτλοι, οι μάσκες και οι κλάσεις μετατρέπονται από λίστες ακεραίων σε τένσορες ακεραίων. Αυτό συμβαίνει για να αξιοποιηθεί πλήρως η υπολογιστική ισχύς του εκάστοτε υπολογιστή.

Κορμός

Το μοντέλο που δημιουργήθηκε, χρησιμοποιεί σαν σημείο εισόδου το μοντέλο του GreekBERT και έχει δύο πλήρη ενσωματωμένα επίπεδα. Το πρώτο έχει 768 χαρακτηριστικά εισόδου, όσα είναι και τα χαρακτηριστικά εξόδου του GreekBERT και 512 χαρακτηριστικά εξόδου. Το δεύτερο πλήρως ενσωματωμένο επίπεδο έχει 512 χαρακτηριστικά εισόδου, και 1 μεταβλητή εξόδου, με τιμές από 0 έως 1 που είναι η πιθανότητα η είσοδος να αντιστοιχεί σε ψευδή είδηση.

Αμέσως μετά το πρώτο ενσωματωμένο επίπεδο, συναντάται η συνάρτηση ενεργοποίησης ReLu, γίνεται το batch normalization και το dropout με πιθανότητα 0,2. Ακολουθεί το δεύτερο ενσωματωμένο επίπεδο και στο τέλος μια σιγμοειδής συνάρτηση εξόδου. Η συνάρτηση αυτή χρησιμοποιείται κατά κόρον σε προβλήματα δυαδικής ταξινόμησης καθώς δίνει σαν έξοδο μία τιμή

από 0 έως 1, που αντιπροσωπεύει μία πιθανότητα.

Τέλος, για τον υπολογισμό του κόστους εκπαίδευσης επιλέχθηκε η binary cross entropy loss η οποία υπολογίζει μετρά τη διαφορά μεταξύ της προβλεπόμενης και της πραγματικής κατανομής πιθανοτήτων. Ο στόχος είναι η ελαχιστοποίηση αυτής της απώλειας, η οποία αντιστοιχεί στη μεγιστοποίηση της ομοιότητας μεταξύ της προβλεπόμενης και της πραγματικής κατανομής.

Το μοντέλο που αναπτύχθηκε με την βιβλιοθήκη της pytorch καθώς και ένα διάγραμμα επεξήγησης της αρχιτεκτονικής του παραθέτονται παρακάτω.

```
class BERT_Fake(nn.Module):

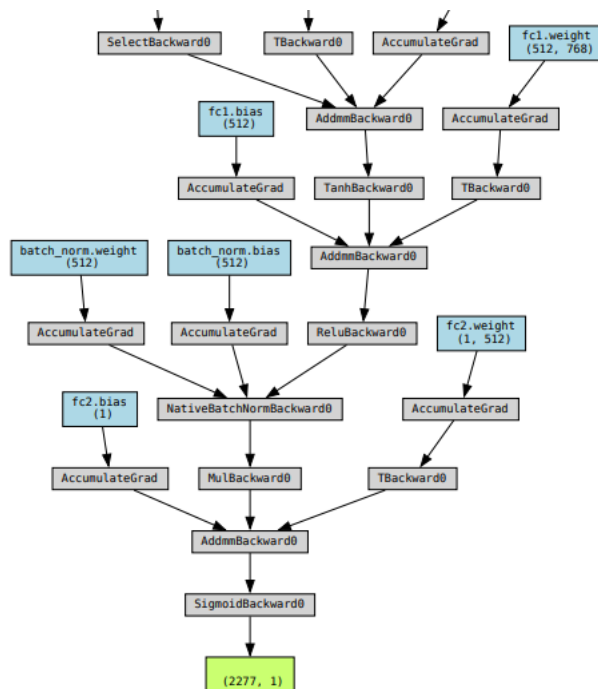
    def __init__(self,
                bert,
                dropout=DEFAULT_DROPOUT,
                hidden_size=DEFAULT_HIDDEN_SIZE) -> None:
        super(BERT_Fake, self).__init__()

        self.bert = bert
        # DropoutLayer
        self.dropout = nn.Dropout(dropout)
        # ReLU
        self.relu = nn.ReLU()
        # Dense Layers
        self.fc1 = nn.Linear(768, hidden_size)
        self.fc2 = nn.Linear(hidden_size, 1)
        # Batch normalization
        self.batch_norm = nn.BatchNorm1d(hidden_size)
        self.sigmoid = nn.Sigmoid()

    def forward(self, send_id, mask):
        _, cls_hs = self.bert(send_id, attention_mask=mask)
        x = self.fc1(cls_hs)
        x = self.relu(x)
        x = self.batch_norm(x)
        x = self.dropout(x)
        x = self.fc2(x)
```

```
x = self.sigmoid(x)
return x
```

Κώδικας Μοντέλου



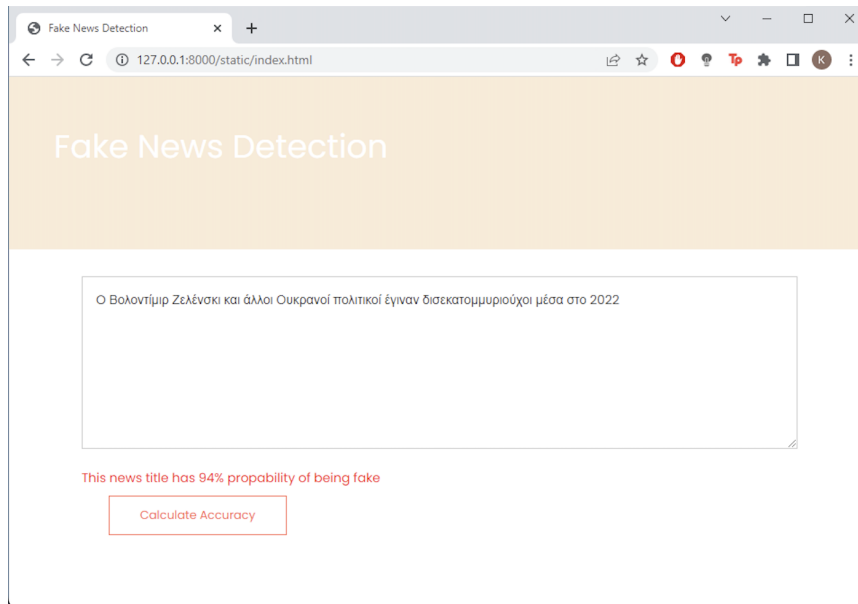
Σχήμα 6.1: Διάγραμμα Οπτικοποίησης Μοντέλου

6.3 Διαδικτυακή Εφαρμογή

Ταυτόχρονα, υλοποιήθηκε μία διαδικτυακή πλατφόρμα για την επίδειξη της λειτουργίας του μοντέλου μηχανικής μάθησης της εργασίας αυτής. Η πλατφόρμα αυτή δίνει την δυνατότητα στον χρήστη να εισάγει έναν τίτλο ειδήσεων, στην ελληνική γλώσσα, και με το πάτημα ενός κουμπιού να του εμφανίζει αν αυτή είναι ψευδής ή όχι. Η πληροφορία αυτή δίνεται στην μορφή ποσοστού για την πιθανότητα της είδησης να είναι ψευδής.

Με το πάτημα του κουμπιού, δημιουργείται ένα αίτημα POST στο /predict που περιέχει την είσοδο του χρήστη. Το αίτημα αυτό, το λαμβάνει το backend της πλατφόρμας που είναι υλοποιημένο με την βοήθεια του Framework ιστού FastAPI σε γλώσσα python, που σερβίρεται από τον διακομιστή ιστού unicorn. Ο τελευταίος φορτώνει κατά την έναρξή του το μοντέλο με τα βάρη που παράχθηκαν, ώστε οι προβλέψεις να γίνονται στιγμιαία χωρίς καθυστέρηση.

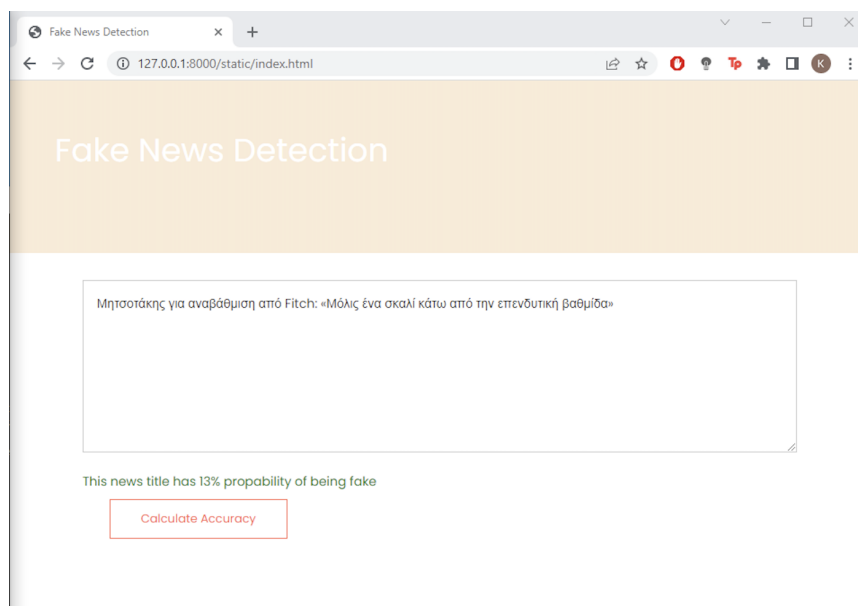
Το FastAPI, δέχεται την είσοδο του χρήστη, την εισάγει στο μοντέλο, παίρνει πρόβλεψη για την πιθανότητα ψευδής είδησης και την επιστρέφει σαν ποσοστό στο frontend της πλατφόρμας.



Σχήμα 6.2: Πρόβλεψη ψευδής είδησης

Αυτή με την σειρά της εμφανίζει το αποτέλεσμα με πράσινο χρώμα αν το ποσοστό είναι χαμηλό (κάτω από 50%) και με κόκκινο αν είναι υψηλό (πάνω από 50%).

Με αυτόν τον τρόπο, η ιστοσελίδα και το μοντέλο μηχανικής μάθησης ενσωματώνονται απρόσκοπτα, επιτρέποντας στους χρήστες να αλληλεπιδρούν εύκολα με το μοντέλο και να λαμβάνουν προβλέψεις σε πραγματικό χρόνο. Η χρήση του FastAPI διευκολύνει επίσης τη συντήρηση και την ενημέρωση της εφαρμογής, καθώς τυχόν αλλαγές στο μοντέλο μπορούν εύκολα να ενσωματωθούν στην ιστοσελίδα χωρίς να επηρεαστεί η εμπειρία του χρήστη.



Σχήμα 6.3: Πρόβλεψη αληθούς είδησης

6.4 Επίλογος

Σε αυτό το κεφάλαιο έγινε αναλυτική επεξήγηση του κώδικα πάνω στον οποίο γράφτηκε το νευρωνικό δίκτυο. Επίσης, παρουσιάστηκε και η διαδικτυακή εφαρμογή που δημιουργήθηκε για την προώθηση του μοντέλου σε χρήστες.

Κεφάλαιο 7ο: Αποτελέσματα

7.1 Εισαγωγή

Στο συγκεκριμένο κεφάλαιο, γίνεται καταγραφή των αποτελεσμάτων που προέκυψαν κατά τη διάρκεια των δοκιμών του μοντέλου, καθώς και τα αποτελέσματα που έτρεξαν για την δοκιμή των συνόλων δεδομένων με ήδη υπάρχοντες αλγόριθμους.

7.2 Κριτήρια

Πίνακας Σύγχυσης (Confusion Matrix) Μια από αυτές τις μεθόδους είναι η εξαγωγή του Πίνακα Σύγχυσης (Confusion Matrix). Αυτός ο πίνακας, είναι μια μέτρηση απόδοσης για το πρόβλημα ταξινόμησης μηχανικής μάθησης όπου η έξοδος μπορεί να είναι δύο ή περισσότερες κλάσεις. Χωρίζεται σε 4 διαφορετικούς συνδυασμούς προβλεπόμενων και πραγματικών τιμών.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Σχήμα 7.1: Πίνακας Σύγχυσης

TP: True Positive, **TN:** True Negative, **FP:** False Positive, **FN:** False Negative

Accuracy Το ποσοστό των σωστών προβλέψεων που κάνει ένα μοντέλο, συνήθως εκφρασμένο ως ποσοστό.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision Η αναλογία των αληθώς θετικών προβλέψεων προς όλες τις θετικές προβλέψεις που γίνονται από ένα μοντέλο. Μετράει πόσες από τις θετικές προβλέψεις ήταν σωστές.

$$Precision = \frac{TP}{TP + FP}$$

Sensitivity Το ποσοστό των αληθώς θετικών προβλέψεων προς όλες τις πραγματικές θετικές περιπτώσεις. Μετρά πόσο καλά ένα μοντέλο ανιχνεύει θετικές περιπτώσεις.

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity

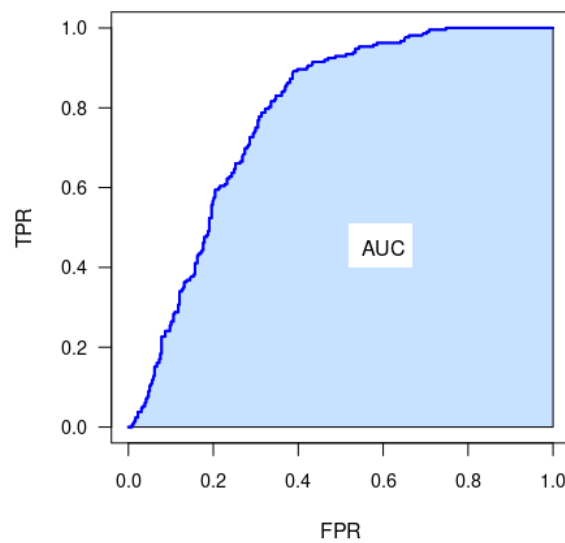
Η αναλογία των αληθών αρνητικών προβλέψεων προς όλες τις πραγματικές αρνητικές περιπτώσεις. Μετρά πόσο καλά ένα μοντέλο αποφεύγει τις ψευδώς θετικές περιπτώσεις.

$$Specificity = \frac{TN}{TN + FP}$$

Fscore

Ένα μέτρο που συνδυάζει την accuracy και την sensitivity, είναι ο αρμονικός μέσος όρος τους. Χρησιμοποιείται ως ενιαία μετρική για την αξιολόγηση της απόδοσης ενός μοντέλου.

$$f_{score} = \frac{Precision \times Sensitivity}{Precision + Sensitivity}$$

AUC

Σχήμα 7.2: Περιοχή κάτω την καμπύλη ROC

Η περιοχή κάτω από την καμπύλη ROC είναι μια μετρική που συγκρίνει την απόδοση ενός δυαδικού ταξινομητή μετρώντας την περιοχή κάτω από την καμπύλη ROC (Receiver Operating Characteristic). Η καμπύλη ROC είναι μια γραφική παράσταση του αληθώς θετικού ποσοστού (TPR) έναντι του ψευδώς θετικού ποσοστού (FPR) για διαφορετικά κατώφλια ταξινόμησης. Η AUC κυμαίνεται μεταξύ 0 και 1, με έναν τέλειο ταξινομητή να έχει AUC 1 και έναν τυχαίο ταξινομητή να έχει AUC 0,5.

7.3 Αποτελέσματα Ταξινόμησης

Classifier	Greek.csv			Translated.csv			Mixed.csv		
	PA	RF	SVM	PA	RF	SVM	PA	RF	SVM
<u>Accuracy</u>	87,6%	64,1%	65,7%	86,07%	68,54%	70,1%	87,8%	65,5%	68%
<u>Precision</u>	83,1%	67,3%	61,6%	91,7%	67,4%	72,72%	85,6%	66,4%	66,5%
<u>Sensitivity</u>	89,3%	66,2%	63,1%	74,3%	81,57%	53,9%	86,2%	73,9%	58,4%
<u>Specificity</u>	86,3%	61,6%	67,8%	94,9%	53%	83,3%	89%	55,5%	75,8%
<u>Fscore</u>	0,86	0,66	0,62	0,82	0,73	0,61	0,85	0,69	0,62

Πίνακας 7.1: Σύγκριση ταξινομητών

PA : *Passive Aggressive*, *RF*: *Random Forest*

Ένα γενικό συμπέρασμα που προκύπτει άμεσα κοιτώντας τα παραπάνω δεδομένα είναι πως το σύνολο δεδομένων greek.csv και με τους τρεις ταξινομητές φαίνεται να φέρνει τα καλύτερα αποτελέσματα. Καθώς το accuracy και το fscore που είναι τα πιο συμπεριληπτικά μεγέθη, είναι υψηλότερα σε σχέση με τα άλλα δύο σύνολα δεδομένων.

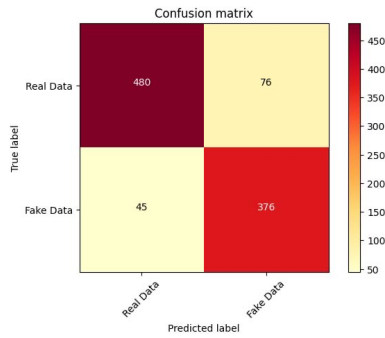
Έχοντας αυτή την γενική εικόνα για το ελληνικό σύνολο δεδομένων, βλέπουμε πως και οι τρεις ταξινομητές φέρνουν κοντινά σε ποσοστό αποτελέσματα. Παρόλα αυτά βλέπουμε πως ο passive aggressive και ο SVM έχουν πιο σταθερά αποτελέσματα. Ο Passive Aggressive φαίνεται να πετυχαίνει καλύτερα την κατηγοριοποίηση των ψευδών ειδήσεων ως ψευδής με sensitivity 89,3%, ενώ ο SVM έχει προκατάληψη προς τις αληθείς ειδήσεις με specificity 89%, πετυχαίνει δηλαδή καλύτερα να κατηγοριοποιήσει τις αληθείς ως αληθείς. Τέλος, ο αλγόριθμος random forest φαίνεται να έχει τις μεγαλύτερες διακυμάνσεις στα ποσοστά του, με τα καθολικά υψηλότερα ποσοστά σε precision και specificity, και τα καθολικά χαμηλότερα σε sensitivity και accuracy. Αυτό δείχνει μια αστάθεια προς την κατηγοριοποίηση των ψευδών ειδήσεων ως ψευδείς καθώς έχει την τάση να κατηγοριοποιήσει τις ψευδείς ειδήσεις ως αληθείς, σενάριο που φαίνεται το πιο επικίνδυνο.

Βλέποντας τα αποτελέσματα για το μεταφρασμένο σύνολο δεδομένων βλέπουμε πως λειτουργεί καλύτερα με τον ταξινομητή random forest αντίθετα με το ελληνικό σύνολο. Εκεί που φαίνεται να υστερεί περισσότερο ο συνδυασμός αυτός είναι στο specificity καθώς μια αληθής είδηση έχει σχεδόν 50% πιθανότητα να καταλήξει λανθασμένα ως ψευδής. Αυτό που καταφέρνει να βελτιώσει εντυπωσιακά τα συνολικά ποσοστά του είναι το υψηλό sensitivity στο 81,57%, ποσοστό που το κάνει συνολικά πιο πετυχημένο σε σχέση με τα άλλα δυο. Στο ίδιο κομμάτι υστερεί επίσης και ο SVM με το μεταφρασμένο σύνολο δεδομένων με specificity 53%. Ο passive aggressive

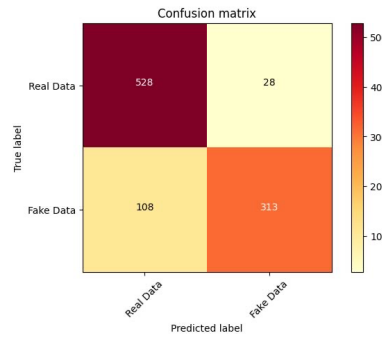
στην συγκεκριμένη περίπτωση φαίνεται να παρουσιάζει τα πιο σταθερά ποσοστά, με νούμερα όμως που τον κατατάσσουν ως μέτριο.

Τέλος, βλέπουμε πως το ενωμένο σύνολο δεδομένων έχει κοντινά ποσοστά επιτυχίας και με τους τρεις ταξινομητές. Αυτός όμως που ξεπερνά τους υπόλοιπους είναι για άλλη μια φορά ο random forest με accuracy 70,1%. Με το ενοποιημένο σύνολο, αντίθετα με το μεταφρασμένο φαίνεται να παρουσιάζει προκατάληψη προς τις αληθείς ειδήσεις, κατηγοριοποιώντας τες με μεγαλύτερη ποσοστό ως αληθείς. Ενώ σε δύο από τα τρία σύνολα δεδομένων έχει μεγαλύτερη επιτυχία ο αλγόριθμος random forest, είναι αυτός που έχει και τις μεγαλύτερες διακυμάνσεις στα sensitivity και specificity.

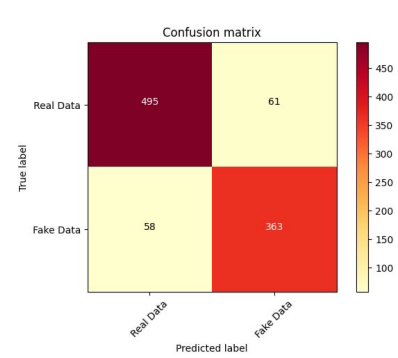
Κεφάλαιο 7



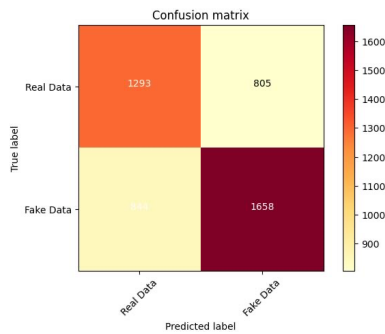
Σχήμα 7.3: Greek: Passive Aggressive



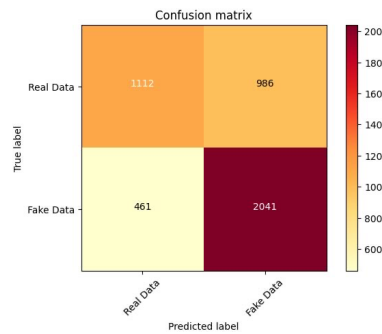
Σχήμα 7.4: Greek: Random Forest



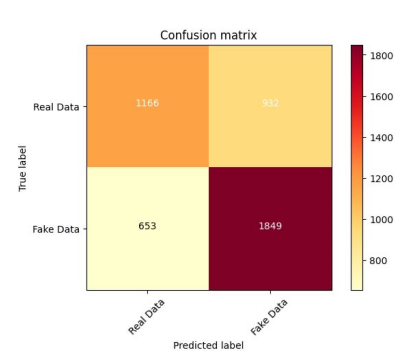
Σχήμα 7.5: Greek: SVM



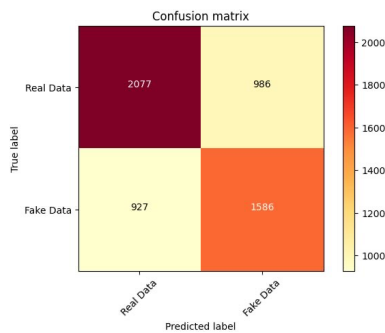
Σχήμα 7.6: Liar: Passive Aggressive



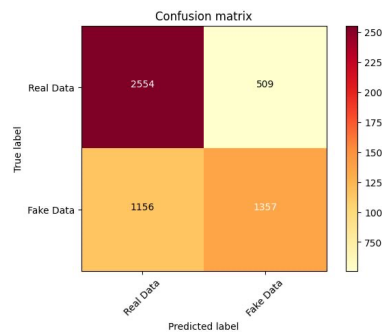
Σχήμα 7.7: Liar: Random Forest



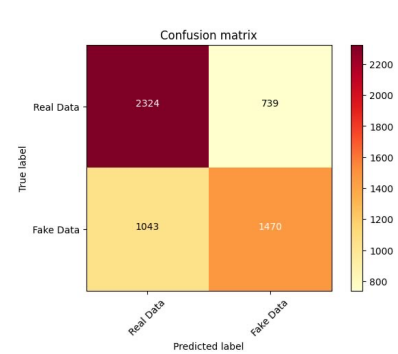
Σχήμα 7.8: Liar: SVM



Σχήμα 7.9: Mixed: Passive Aggressive



Σχήμα 7.10: Mixed: Random Forest



Σχήμα 7.11: Mixed: SVM

7.4 Αποτελέσματα Μοντέλου

Στο μοντέλο δοκιμάστηκαν και τα δύο σύνολα δεδομένων ελληνικό και ενωμένο, αλλά καθώς το ελληνικό είχε καλύτερα αποτελέσματα εν γένει, και στο νευρωνικό δίκτυο και στους γνωστούς ταξινομητές, επιλέχθηκε ως βασικό για την εκπαίδευση του μοντέλου. Για να βρεθεί η βέλτιστη εκδοχή του μοντέλου μεταβλήθηκαν οι εξής παράμετροι: αριθμός εποχών, πιθανότητα dropout, αριθμός παραμέτρων στο κρυφό επίπεδο, ρυθμός εκμάθησης και συνάρτηση κόστους. Κάνοντας πολλές δοκιμές, με διάφορους συνδυασμούς των παραπάνω μετρικών, βρέθηκαν οι παράμετροι με την μεγαλύτερη ακρίβεια. Αυτές οι τιμές παραθέτονται στον παρακάτω πίνακα:

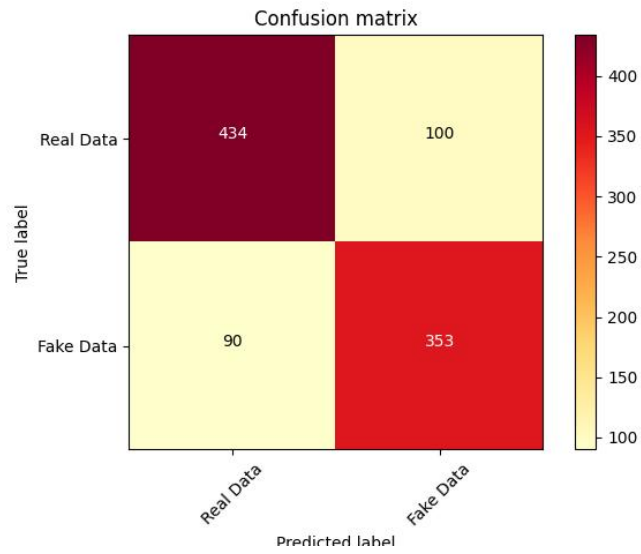
Παράμετροι	
Εποχές	15
Πιθανότητα Dropout	0.2
Αριθμός παραμέτρων κρυφού επιπέδου	512
Ρυθμός εκμάθησης	10^{-4}
Συνάρτηση κόστους	Binary Cross Entropy Loss

Πίνακας 7.2: Παράμετροι Μοντέλου

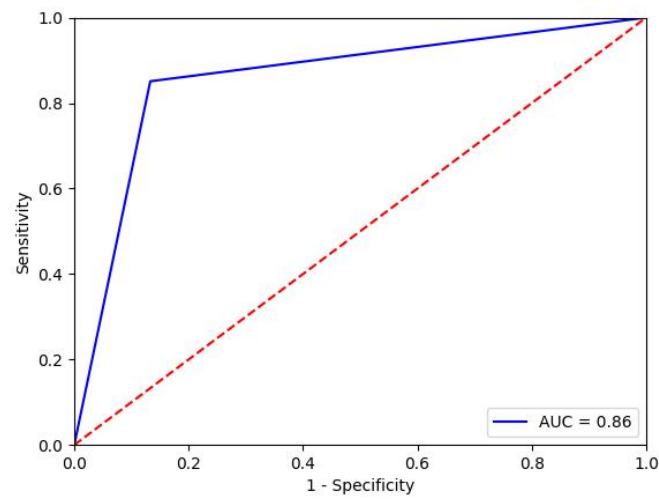
Με τις παραπάνω παραμέτρους το μοντέλο είχε την βέλτιστη απόδοση. Αυτή είναι η αξιολόγηση που προκύπτει με βάση τα κριτήρια:

accuracy	80,55%
precision	77,92%
sensitivity	79,68%
specificity	81,27%
fscore	78,79%

Πίνακας 7.3: Μετρικές Μοντέλου



Σχήμα 7.12: Πίνακας Σύγκρισης Μοντέλου



Σχήμα 7.13: Roc Curve

Κεφάλαιο 8ο: Συμπεράσματα

Για την επιλογή των κατάλληλων υπερπαραμέτρων για το μοντέλο που δημιουργήθηκε στα πλαίσια αυτής της εργασίας, πραγματοποιήθηκαν πειράματα με διαφορετικούς συνδυασμούς των παραμέτρων και στα 3 σύνολα δεδομένων. Σαν κριτήριο σύγκρισης μεταξύ των πειραμάτων επιλέχθηκε το ποσοστό ακρίβειας και το fscore. Θα μπορούσαν να επιλεγθούν άλλες μετρικές, όπως το precision ή η περιοχή κάτω από το ROC(Area Under Roc), ωστόσο στην περίπτωση ενός ισορροπημένου συνόλου δεδομένου και σε ένα πρόβλημα δυαδικής ταξινόμησης, τα 2 που επιλέχθηκαν έχουν το περισσότερο νόημα και δίνουν την καλύτερη πληροφορία για την επίδοση της ταξινόμησης. [26]

Σύνολο δεδομένων

Το σύνολο δεδομένων με την καλύτερη επίδοση σε όλα τα πειράματα και με όλους τους αλγορίθμους ήταν το ελληνικό που δημιουργήθηκε για την παρούσα εργασία. Η επιτυχία του οφείλεται σε μεγάλο βαθμό στην ποιότητα του, καθώς η επιλογή της πηγής των ειδήσεων έγινε προσεκτικά με σκοπό την αποφυγή εισαγωγής θορύβου σε αυτό. Ωστόσο, λόγω του μικρού αριθμού εγγραφών του σε σχέση με το Liar, για να υπάρχει μια πιο σαφή εικόνα για την επίδοση του θα πρέπει να ενημερωθεί με περισσότερες ειδήσεις. Αυτό αποτελεί από μόνο του μια πρόκληση καθώς δεν υπάρχει έως τώρα κάποιος αυτοματοποιημένος τρόπος για την καταγραφή των νέων ψευδών ειδήσεων.

Αλγόριθμοι

Όσον αφορά τους αλγορίθμους που χρησιμοποιήθηκαν για την επίλυση του προβλήματος στο οποίο αναφέρεται η εργασία, παρατηρούμε ότι το μοντέλο νευρωνικού δικτύου που δημιουργήθηκε καταφέρνει να έχει παρόμοια επίδοση με την καλύτερη που κατέγραψαν οι αλγόριθμοι ταξινόμησης. Πιο συγκεκριμένα, η καλύτερη επίδοση του νευρωνικού με το ελληνικό σύνολο δεδομένων απέχει περίπου 6% από την καλύτερη των ταξινομητών με το ίδιο σύνολο δεδομένων.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Μ. Μαλάμη, “Η σπουδαιότητα της πληροφορίας και το φαινόμενο των fake news.” <https://www.enainstitute.org/>. Accessed: 2010-09-30.
- [2] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language models are few-shot learners,” 2020.
- [3] H. L. Althuis L., “Fake News. A Roadmap.,” *Riga: NATO Strategic Communications Centre of Excellence.*, 2018.
- [4] E. A. for Viewers Interests, “Beyond fake news – 10 types of misleading news.” https://library-nd.libguides.com/ld.php?content_id=51257692, 2020.
- [5] K. Rohit, Kumar and G. Anurag, “Fakebert: Fake news detection in social media with a bert-based deep learning approach,” *Multimedia Tools and Applications volume*, vol. 80, no. issue, 2021.
- [6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” 2018.
- [7] “The Factual.” <https://www.thefactual.com>. Accessed: 19-01-2023.
- [8] “Greek Hoaxes.” <https://www.ellinikahoaxes.gr/>. Accessed: 19-01-2023.
- [9] M. A. Alonso, D. Vilares, C. Gómez-Rodríguez, and J. Vilares, “Sentiment analysis for fake news detection,” *Electronics*, vol. 10, no. 11, 2021.
- [10] N. Vo and K. Lee, “The rise of guardians: Fact-checking url recommendation to combat fake news,” SIGIR '18, (New York, NY, USA), p. 275–284, Association for Computing Machinery, 2018.
- [11] B. Riedel, I. Augenstein, G. P. Spithourakis, and S. Riedel, “A simple but tough-to-beat baseline for the fake news challenge stance detection task,” 2017.
- [12] “Fake News Challenge.” <http://www.fakenewschallenge.org/>. Accessed: 19-01-2023.
- [13] B. Probierz, P. Stefański, and J. Kozak, “Rapid detection of fake news based on machine learning methods,” *Procedia Computer Science*, vol. 192, pp. 2893–2902, 2021. Knowledge-Based and Intelligent Information Engineering Systems: Proceedings of the 25th International Conference KES2021.

- [14] P. K. Ioannis Vlahavas, *Artificial Intelligence*. University of Macedonia, 2020.
- [15] K. Koutsikakis and G. Tsatsaronis, “Greekbert: Pre-training of deep bidirectional transformers for greek natural language processing,” in *Proceedings of the 12th Language Resources and Evaluation Conference*, 2020.
- [16] “Common Crawl.” <https://commoncrawl.org/>. Accessed: 19-01-2023.
- [17] “Python.” <http://www.fakenewschallenge.org/>. Accessed: 19-01-2023.
- [18] R. Karan, “Why Python for Machine Learning is Important?,” 2022.
- [19] “Pytorch.” <https://www.python.org/>. Accessed: 19-01-2023.
- [20] “Scikit Learn.” <https://scikit-learn.org/stable/>. Accessed: 19-01-2023.
- [21] “Numpy.” <https://numpy.org/>. Accessed: 19-01-2023.
- [22] “Scipy.” <https://scipy.org/>. Accessed: 19-01-2023.
- [23] “Η Καθημερινή.” <https://www.kathimerini.gr/>. Accessed: 19-01-2023.
- [24] “Reuters Institute.” <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022/greece>. Accessed: 19-01-2023.
- [25] G. Sidorov, F. Velasquez, E. Stamatatos, A. Gelbukh, and L. Chanona-Hernández, “Syntactic n-grams as machine learning features for natural language processing,” *Expert Systems with Applications*, vol. 41, no. 3, pp. 853–860, 2014. *Methods and Applications of Artificial and Computational Intelligence*.
- [26] A. D’Agostino, “The explanation you need on binary classification metrics.” shorturl.at/BEKVX. Accessed: 2010-09-30.
- [27] R. Karan, “Why Python for Machine Learning is Important?,” 2022.
- [28] J. L. Spudich and B. H. Satir, *Sensory Receptors and Signal Transduction*. New York: Wiley-Liss, 2001.