



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ
ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Ανίχνευση Ψυχικών Διαταραχών σε Social Media με Τεχνικές Μηχανικής Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μάριος Α. Κερασιώτης

Επιβλέπων: Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2023



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ
ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Ανίχνευση Ψυχικών Διαταραχών σε Social Media με Τεχνικές Μηχανικής Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μάριος Α. Κερασιώτης

Επιβλέπων: Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 19^η Οκτωβρίου 2023.

.....
Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

.....
Ιωάννης Ψαρράς
Καθηγητής Ε.Μ.Π.

.....
Ευάγγελος Μαρινάκης
Επίκουρος Καθηγητής
Ε.Μ.Π.

Αθήνα, Οκτώβριος 2023

.....
Μάριος Α. Κερασιώτης
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Μάριος Α. Κερασιώτης, 2023.
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περίληψη

Η παρούσα διπλωματική εργασία επικεντρώνεται στην ανίχνευση διαταραχών ψυχικής υγείας, συγκεκριμένα της κατάθλιψης, μέσω ανάλυσης αναρτήσεων στο Reddit. Χρησιμοποιώντας ένα σύνολο δεδομένων που αντιπροσωπεύει την κατάθλιψη με κλάσεις σοβαρότητας (ελάχιστη, ήπια, μέτρια, σοβαρή), η μελέτη χρησιμοποιεί ένα μοντέλο βασισμένο σε μετασχηματιστές, προερχόμενοι από τον BERT, μαζί με μεταδεδομένα, για την επεξεργασία των δεδομένων κειμένου. Τα τέσσερα τελευταία στρώματα των εξόδων του μετασχηματιστή συνενώνονται, πολλαπλασιάζονται με βάρη και συγχωνεύονται, προτού τροφοδοτηθούν σε αρχιτεκτονικές νευρωνικών δικτύων, όπως Δίκτυα Εμπρόσθιας Τροφοδότησης Πολλών Επιπέδων (Multi Layer Perceptron MLP), Δίκτυα Μακράς Βραχύχρονης Μνήμης (Long ShortTerm Memory, LSTM) και Πολυτροπικών Κυττάρων με Πύλες (Gated Multimodal Unit, GMU). Στην ανάλυση ενσωματώνονται μεταδεδομένα, συμπεριλαμβανομένων γλωσσικών μεταδεδομένων, αναφορών φαρμάκων κατάθλιψης, αποτελέσματα μοντέλου emoRoberta (συναισθηματική ανάλυση) και ανάλυσης συναισθήματος (πολικότητα). Χρησιμοποιούνται τεχνικές επαύξησης κειμένου για την αντιμετώπιση των περιορισμένων δεδομένων κατάθλιψης. Η έρευνα περιλαμβάνει επίσης τη σύγκριση των μετρικών που προκύπτουν από την εκτέλεση των μοντέλων νευρωνικών δικτύων για την αξιολόγηση της αποτελεσματικότητας των μοντέλων στην ανίχνευση κατάθλιψης σε αναρτήσεις στα μέσα κοινωνικής δικτύωσης.

Λέξεις Κλειδιά: ψυχική υγεία, ψυχικές διαταραχές, κατάθλιψη , μέσα κοινωνικής δικτύωσης, νευρωνικά δίκτυα, μετασχηματιστές, BERT, πολυεπίπεδο Perceptron, δίκτυα Μακράς Βραχύχρονης Μνήμης, επεξεργασία φυσικής γλώσσας, ανίχνευση συναισθήματος, κατηγοριοποίηση κειμένου, ανάλυση συναισθήματος, δίκτυα προσοχής, βαθειά μάθηση

Abstract

This diploma thesis focuses on the early detection of mental health disorders, specifically depression, through an analysis of Reddit posts obtained from public sources. Utilizing a dataset representing depression with severity classes (minimum, mild, moderate, severe), the study employs a transformer-based model, Bidirectional Encoder Representations from Transformers (BERT), integrated with metadata, to process the text data. The last four layers of the transformer outputs are concatenated, weighted, and fused, before being introduced to neural network architectures such as Multi Layer Perceptron (MLP), Long ShortTerm Memory (LSTM), and Gated Multimodal Unit (GMU). Metadata elements considered include temporal information, linguistic metadata, references to depression medication, emoRoberta outputs (emotional analysis), and sentiment analysis (polarity). Text augmentation techniques are also employed to counterbalance the limited depression data available. A primary objective of this research is to compare the metrics obtained from running these neural network models, assessing their effectiveness in identifying signs of depression in social media posts.

Keywords: mental health, mental disorders, depression, social media, neural networks, transformers, BERT, multi-layer perceptron, long-short term memory, natural language processing, emotion detection, text classification, sentiment analysis, attention networks, deep learning

Ευχαριστίες

Σε αυτό το σημείο θα ήθελα να εκφράσω την βαθιά μου ευγνωμοσύνη και εκτίμηση στον Καθηγητή Δημήτρη Ασκούνη για την πολύτιμη ευκαιρία που μου παρείχε να ασχοληθώ ενδελεχώς με το πρόβλημα της ανίχνευσης κατάθλιψης στα μέσα κοινωνικής δικτύωσης. Επίσης, θα ήθελα να ευχαριστήσω θερμά τον υποψήφιο διδάκτορα Λουκά Ηλία, ο οποίος με καθοδήγησε και με υποστήριξε σε κάθε βήμα της εκπόνησης αυτής της διπλωματικής.

Θέλω να ευχαριστήσω την οικογένειά μου, τη Λαμπρινή, τον Θανάση και τον Γιάννη, για την ανελλιπή τους αγάπη, την υποστήριξη και την πίστη που είχαν σε μένα καθ' όλη τη διάρκεια των σπουδών μου. Ένα ιδιαίτερο ευχαριστώ θέλω να απευθύνω στην κοπέλα μου, Κωνσταντίνα, για την αγάπη, την υποστήριξη και την υπομονή της.

Τέλος ευχαριστώ τους φίλους μου, αυτούς που απέκτησα κατά τη διάρκεια των σπουδών μου στη σχολή, αλλά και εκτός αυτής. Η παρουσία σας, η στήριξη και οι στιγμές που μοιραστήκαμε έκαναν αυτό το ταξίδι ακόμα πιο ξεχωριστό. Σας ευχαριστώ που ήσασταν πάντα εκεί για μένα, στις χαρές και στις δυσκολίες, και που με βοηθούσατε να διατηρώ το ηθικό μου ψηλά.

Σε όλους σας, ευχαριστώ από τα βάθη της καρδιάς μου.

Μάριος Κερασιώτης,
Αθήνα, Οκτώβριος 2023

Περιεχόμενα

1	Εισαγωγή	10
1.1	Ψυχικές διαταραχές	10
1.2	Μέσα κοινωνικής δικτύωσης	11
1.2.1	Reddit	12
1.3	Πρόσφατη έρευνα και περιορισμοί	13
1.4	Συνεισφορά της διπλωματικής	14
1.5	Δομή της εργασίας	15
2	Βιβλιογραφική Ανασκόπηση	16
2.1	Έρευνες που χρησιμοποιούν μόνο μοντέλα μηχανικής μάθησης	16
2.2	Έρευνες που χρησιμοποιούν νευρωνικά δίκτυα	18
2.3	Έρευνες που χρησιμοποιούν μετασχηματιστές	19
3	Τεχνητή Νοημοσύνη	21
3.1	Μηχανική Μάθηση	21
3.2	Επιβλεπόμενη Μάθηση	22
3.2.1	Λογιστική Παλινδρόμηση (Logistic Regression)	22
3.2.2	Log-Linear	22
3.2.3	Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)	23
3.2.4	Naive Bayes	24
3.2.5	K-Nearest Neighbors	25
3.2.6	Decision Trees	26
3.3	Μη Επιβλεπόμενη Μάθηση	28
3.3.1	K-Means	28
3.4	Νευρωνικά Δίκτυα	29
3.4.1	Στρώμα Dropout	30
3.4.2	Στρώμα Πλήρους Σύνδεσης	31
3.4.3	Δίκτυα Εμπρόσθια Τροφοδότησης Πολλών Επιπέδων	31
3.4.4	Δίκτυα Μακράς Βραχύχρονης Μνήμης	31
3.5	Μετασχηματιστές	32
3.5.1	BERT (Bidirectional Encoder Representations from Transformers)	36
3.5.2	DistilBERT	38
3.5.3	EmoRoBERTa	38
3.5.4	Twitter-RoBERTa για Ανάλυση συναισθημάτων	40
3.6	Προ επεξεργασία κειμένου	41
3.6.1	Tokenization	41
3.6.2	Stemming	43

3.6.3	Λημματοποίηση	43
3.7	Διανύσματα αναπαράστασης κειμένου	44
3.7.1	Bag of Words	44
3.7.2	N-grams	44
3.7.3	TF-IDF	45
3.8	Αναπαραστάσεις λέξεων (Word Embeddings)	46
3.9	Μοντελοποίηση Θεμάτων	46
3.9.1	Latent Dirichlet Allocation	46
3.10	Επαύξησή δεδομένων κειμένου	47
3.11	Συναρτήσεις ενεργοποίησης	48
3.11.1	Σιγμοειδής συνάρτηση	48
3.11.2	Softmax συνάρτηση	48
3.11.3	Υπερβολική Εφαπτομένη	48
3.11.4	Ανορθωμένη Γραμμική Μονάδα	49
3.12	Μετρικές	51
3.12.1	Πίνακας Σύγχυσης	51
3.12.2	Ακρίβεια (Accuracy)	51
3.12.3	Ακρίβεια (Precision) και Ανάκληση (Recall)	52
3.12.4	F1 score	53
4	Μεθοδολογία	54
4.1	Σύνολο Δεδομένων και Προεπεξεργασία	54
4.1.1	Σύνολο δεδομένων	54
4.1.2	Επαύξηση Δεδομένων	57
4.1.3	Προεπεξεργασία Δεδομένων	58
4.2	Αρχιτεκτονική Μοντέλου	59
4.2.1	Εξαγωγή Χαρακτηριστικών	60
4.2.2	Συγχώνευση Χαρακτηριστικών με τον Μετασχηματιστή	61
4.3	Πειραματική Διάταξη	65
4.3.1	Εργαστηριακό Περιβάλλον	65
4.3.2	Μετρικές αξιολόγησης	66
4.3.3	Μοντέλα αναφοράς	66
4.4	Μέλετη Διαφορετικών Διάταξεων	67
4.4.1	Κεφαλές Κατηγοριοποίησης	67
4.4.2	Μετασχηματιστές	67
4.4.3	Στρώματα του μετασχηματιστή	67
4.4.4	Επαύξηση Δεδομένων	67
5	Αποτελέσματα	68
5.1	Αποτελέσματα προτεινόμενου μοντέλου	68
5.2	Αποτελέσματα μελέτης διαφορετικών διατάξεων	69
5.2.1	Κεφαλές κατηγοριοποίησης	69
5.2.2	Μετασχηματιστές	70
5.2.3	Στρώματα του μετασχηματιστή	72
5.2.4	Επαύξηση δεδομένων	74
6	Συμπεράσματα και μελλοντικές προεκτάσεις	76
6.1	Συμπεράσματα	76
6.2	Μελλοντικές Προεκτάσεις	77

Κατάλογος σχημάτων

3.1	Μηχανή Διανυσμάτων Υποστήριξης(SVM) [54].	23
3.2	K-nearest neighbors (K-nn) [54].	25
3.3	Decision Tree του Iris dataset[55].	26
3.4	Παράδειγμα Random Forest[56].	27
3.5	Παράδειγμα K-Means[57].	29
3.6	Νευρώνας εγκεφάλου[59].	30
3.7	Τεχνητός νευρώνας.	30
3.8	Multi-Layer Perceptron (MLP) [62].	32
3.9	Κύτταρο Δικτύου Μακράς Βραχύχρονης Μνήμης (LSTM)[65].	33
3.10	Δομή ενός μετασχηματιστή [66].	34
3.11	Αυτό-προσοχή με κλιμακούμενο εσωτερικό διάνυσμα [66].	35
3.12	Το Multi-Head Attention αποτελείται από πολλά επίπεδα προσοχής που εκτελούνται παράλληλα [66].	36
3.13	Αποτελέσματα του EmoRoBERTa από τυχαίο κείμενο του συνόλου δεδομένων.	39
3.14	Αποτελέσματα του Twitter-RoBERTa από τυχαίο κείμενο του συνόλου δεδομένων.	40
3.15	Η Σιγμοειδής συνάρτηση.	49
3.16	Η συνάρτηση Softmax.	49
3.17	Η Υπερβολική Εφαπτομένη.	50
3.18	Η RELU.	50
4.1	Αναπαράσταση της προτεινόμενης μεθοδολογίας.	60
4.2	Παράδειγμα στρωμάτων BERT [95].	64
5.1	Διάγραμμα αποτελεσμάτων μεταξύ προτεινόμενου μοντέλου και των βάσεων αναφοράς.	69
5.2	Διάγραμμα αποτελεσμάτων μεταξύ των κεφαλών ταξινόμησης.	71
5.3	Διάγραμμα αποτελεσμάτων ανά στρώμα για το MLP με κρυφό μέγεθος 512.	72
5.4	Διάγραμμα αποτελεσμάτων ανά στρώμα για το LSTM με κρυφό μέγεθος 512.	73

Κατάλογος πινάκων

3.1	Πίνακας Σύγχυσης (Confusion Matrix).	51
4.1	Δείγμα του συνόλου δεδομένων.	55
4.2	Τεχνικές Προεπεξεργασίας Κειμένου.	59
4.3	Παράδειγμα του συνόλου δεδομένων.	62
4.4	Φάρμακα που χρησιμοποιούνται σε περιπτώσεις κατάθλιψης.	63
5.1	Σύγκριση απόδοσης βάσεων αναφοράς και του προτεινόμενου μοντέλου.	68
5.2	Χρήση διαφορετικών κεφαλών ταξινόμησης.	69
5.3	Χρήση διάφορων μετασχηματιστών.	70
5.4	Αποτελέσματα με την επιλογή διαφορετικών στρωμάτων για MLP κρυφού μεγέθους 512.	72
5.5	Αποτελέσματα με την επιλογή διαφορετικών στρωμάτων για LSTM κρυφού μεγέθους 512.	73
5.6	Σύγκριση των αποτελεσμάτων επαύξησης για μοντέλο DistilBERT base uncased και διάφορες κεφαλές ταξινόμησης χρησιμοποιώντας 4 στρώματα.	74
5.7	Σύγκριση των αποτελεσμάτων επαύξησης για διαφορετικούς μετασχηματιστές.	74
5.8	Σύγκριση των αποτελεσμάτων επαύξησης για μοντέλο DistilBERT base uncased και κεφαλή ταξινόμησης MLP με κρυφό μέγεθος 512.	74
5.9	Σύγκριση των αποτελεσμάτων επαύξησης για μοντέλο BERT base uncased και κεφαλή ταξινόμησης LSTM με κρυφό μέγεθος 512.	75

Κεφάλαιο 1

Εισαγωγή

1.1 Ψυχικές διαταραχές

Σύμφωνα με τα πιο πρόσφατα στατιστικά στοιχεία από τον παγκόσμιο οργανισμό υγείας (Worldwide Health Organization, WHO) [1], 1 σε κάθε 8 ανθρώπους στον κόσμο ζει με μια ψυχική διαταραχή. Οι ψυχικές διαταραχές περιλαμβάνουν σημαντικές διαταραχές στη σκέψη, τη ρύθμιση των συναισθημάτων ή τη συμπεριφορά. Υπάρχουν πολλοί διαφορετικοί τύποι ψυχικών διαταραχών, και η πρόληψη και η θεραπεία είναι αποτελεσματικές για πολλούς από αυτούς. Ωστόσο, πολλοί άνθρωποι δεν έχουν πρόσβαση σε αποτελεσματική φροντίδα. Οι ψυχικές διαταραχές μπορεί να περιγραφούν ως σημαντικές παρεκκλίσεις στη γνωστική, συναισθηματική ή συμπεριφορική λειτουργία του ατόμου. Αν και υπάρχουν πολλές κατηγορίες, οι πλέον διαδεδομένες είναι:

- η αγχώδης διαταραχή
- η κατάθλιψη
- η διπολική διαταραχή
- η διαταραχή μετατραυματικού στρες (PTSD)
- η σχιζοφρένεια
- οι διατροφικές διαταραχές
- η διαταρακτική συμπεριφορά
- οι αποκοινωνικές διαταραχές και
- οι νευροαναπτυξιακές διαταραχές

Το 2019, 301 εκατομμύρια άνθρωποι ζούσαν με διαταραχή αγχούς, συμπεριλαμβανομένων 58 εκατομμυρίων παιδιών και εφήβων. Οι διαταραχές αγχούς χαρακτηρίζονται από υπερβολικό φόβο και ανησυχία και σχετικές συμπεριφορικές διαταραχές. Τα συμπτώματα είναι αρκετά σοβαρά ώστε να προκαλέσουν σημαντική ανησυχία ή σημαντική δυσλειτουργία. Ταυτόχρονα, όσον αφορά την κατάθλιψη, το 2019, σχεδόν 280 εκατομμύρια άνθρωποι ανά τον κόσμο διαγνώστηκαν με αυτήν την ασθένεια, ενώ πάνω από 700.000 άνθρωποι επέλεξαν την αυτοκτονία [2].

Αργότερα, το 2020, ο αριθμός των ανθρώπων που ζούσαν με διαταραχές αγχούς και κατάθλιψης αυξήθηκε σημαντικά λόγω της πανδημίας COVID-19. Οι αρχικές εκτιμήσεις δείχνουν αύξηση κατά 26% και 28% αντίστοιχα για τις διαταραχές αγχούς και την κατάθλιψη μέσα σε ένα έτος. Παρά τις διαθέσιμες επιλογές πρόληψης και θεραπείας, η πλειονότητα των ανθρώπων με ψυχικές διαταραχές δεν έχει πρόσβαση σε αποτελεσματική φροντίδα. Πολλοί αντιμετωπίζουν επίσης στιγματισμό, διακρίσεις και παραβιάσεις των ανθρωπίνων δικαιωμάτων.

Η ψυχική υγεία είναι ένα κρίσιμο στοιχείο της συνολικής υγείας του ατόμου. Επηρεάζει την ικανότητα του ατόμου να συνδέεται με τους άλλους, να αντιμετωπίζει τις προκλήσεις της ζωής και να βρίσκει νόημα και σκοπό. Οι ψυχικές διαταραχές μπορούν να επηρεάσουν αυτές τις βασικές ανθρωπίνες λειτουργίες, καθιστώντας τη ζωή δύσκολη και πολλές φορές ανυπόφορη.

Είναι σημαντικό να κατανοούμε ότι οι ψυχικές διαταραχές δεν είναι αποτέλεσμα αδυναμίας ή χαρακτηριστικού του ατόμου. Πολλοί παράγοντες, όπως η γενετική, οι βιολογικοί και οι περιβαλλοντικοί παράγοντες, μπορούν να συμβάλλουν στην εμφάνιση των διαταραχών αυτών. Παρά την αυξημένη ευαισθητοποίηση γύρω από τη ψυχική υγεία και τη διαθεσιμότητα προληπτικών και θεραπευτικών επιλογών, πολλοί ασθενείς αντιμετωπίζουν δυσκολίες στην πρόσβαση σε κατάλληλες υπηρεσίες για αυτούς. Αυτό συχνά οφείλεται στον στιγματισμό που σχετίζεται με τα ψυχικά νοσήματα, τις διακρίσεις και τις παραβιάσεις των ανθρωπίνων δικαιωμάτων που αντιμετωπίζουν οι πάσχοντες.

Είναι ζωτικής σημασίας να στηρίξουμε τους ανθρώπους που αντιμετωπίζουν ψυχικές διαταραχές, να παρέχουμε την απαραίτητη εκπαίδευση και να αλλάξουμε τις αρνητικές αντιλήψεις που συνδέονται με αυτές τις καταστάσεις.

1.2 Μέσα κοινωνικής δικτύωσης

Κατά τη διάρκεια των τελευταίων δεκαετιών, τα κοινωνικά μέσα έχουν μεταμορφωθεί από απλές ψηφιακές καινοτομίες σε αναπόσπαστα μέρη της καθημερινότητας δισεκατομμυρίων ανθρώπων σε όλο τον κόσμο. Αυτές οι πλατφόρμες παρέχουν έναν μεγάλο αριθμό ψηφιακών εργαλείων και κοινοτήτων, προωθώντας τη δημιουργία περιεχομένου από τους χρήστες. Όπως υπογράμμισαν οι Kaplan και Haenlein (2010) [3], τα κοινωνικά μέσα αντιπροσωπεύουν "μια ομάδα εφαρμογών βασισμένων στο Διαδίκτυο που οικοδομούνται πάνω στις ιδεολογικές και τεχνολογικές βάσεις του Web 2.0, και που επιτρέπουν τη δημιουργία και ανταλλαγή περιεχομένου που παράγεται από τους χρήστες". Ξεκινώντας με τις απλές πλατφόρμες του Διαδικτύου, όπως τα chat rooms και τα bulletin boards, η πορεία των κοινωνικών μέσων έχει καταστεί μετεωρική. Η εμφάνιση των Friendster και MySpace στα αρχικά στάδια της δεκαετίας του 2000 σηματοδότησε μια μετάβαση προς προσωποποιημένα διαδικτυακά προφίλ. Αυτοί οι ιστότοποι δημιούργησαν τις βάσεις για το φαινόμενο που θα γινόταν γνωστό ως Facebook, μια πλατφόρμα που δημοκρατικοποίησε γρήγορα την παγκόσμια ψηφιακή σφαίρα μέχρι τα μέσα της δεκαετίας του 2000, θέτοντας νέα πρότυπα στην εξάπλωση πληροφοριών και μετασχηματίζοντας τα παραδοσιακά πρότυπα της κοινωνικής σύνδεσης.

Από κοινωνιολογική άποψη, η δύναμη των κοινωνικών μέσων έχει αποδειχθεί

περισσότερο στην ικανότητά τους να δημιουργούν κοινότητες. Αντί να είναι δεσμευμένες από γεωγραφικούς ή οικογενειακούς δεσμούς, αυτές οι εικονικές κοινότητες ανθίζουν με βάση τα κοινά πάθη, τα χόμπι και τις ιδεολογίες. Ωστόσο, αυτή η παγκόσμια διασύνδεση έχει αποκαλύψει κάποια παράδοξα αποτελέσματα. Η Sherry Turkle [4] το διερεύνησε αυτό στο "Alone Together", δείχνοντας πώς, ενώ περισσότεροι άνθρωποι είναι συνδεδεμένοι εικονικά, τα πραγματικά αισθήματα της απομόνωσης και της μοναξιάς έχουν εκτοξευτεί. Η παρουσίαση και η συνεχής κατανάλωση των ιδανικοποιημένων ζώων σε αυτές τις πλατφόρμες επιδεινώνουν τις κοινωνικές πιέσεις, δημιουργώντας ένα περιβάλλον συνεχούς σύγκρισης.

Από ψυχολογική άποψη, οι μηχανισμοί των κοινωνικών μέσων, ιδιαίτερα οι βρόχοι ανάδρασης άμεσης ικανοποίησης που σχηματίζονται από "likes", "shares" και σχόλια, εξυπηρετούν τις εγγενείς μας ανάγκες για αναγνώριση. Αυτοί οι βρόχοι, όπως τεκμηρίωσαν οι Twenge και Campbell [5], σχετίζονται στενά με τις αυξανόμενες ναρκισσιστικές τάσεις στους χρήστες και με μια συνεχή μείωση των αυθεντικών ενσυναίσθητων (empathetic) αλληλεπιδράσεων. Επιπλέον, η παράδοση περιεχομένου με βάση αλγόριθμους περιπλέκει το ψυχολογικό τοπίο καθώς συχνά περιορίζει τους χρήστες εντός "echo chambers", τους φοράει παρωπίδες και ενισχύει τις υπάρχουσες πεποιθήσεις τους ενώ περιορίζει την έκθεση σε διαφορετικές προοπτικές. Ο Pariser [6] ονόμασε αυτό το φαινόμενο "Filter Bubble", επισημαίνοντας την απομόνωση που μπορεί να προωθήσουν τέτοια ψηφιακά περιβάλλοντα. Στον πολιτιστικό τομέα, η παντοδυναμία των κοινωνικών μέσων δεν έχει μόνο επηρεάσει τον τρόπο με τον οποίο παράγουμε και καταναλώνουμε περιεχόμενο, αλλά έχει επίσης αλληλεπιδράσει στον τρόπο με τον οποίο αντιλαμβανόμαστε τον εαυτό μας και τους άλλους. Η σταδιακή μετάβαση προς την ψηφιακή παραγωγή περιεχομένου και η δυνατότητα αμεσότητας στη διάδοση και στην αλληλεπίδραση, έχει προκαλέσει ένα θόρυβο στην καθημερινή ζωή μας, καθορίζοντας ένα νέο κοινωνικό και πολιτιστικό τοπίο που συνεχίζει να εξελίσσεται καθημερινά.

Στον ακαδημαϊκό τομέα, τα κοινωνικά μέσα ενθαρρύνουν μια νέα μορφή συλλογικής μάθησης και ερευνητικής συνεργασίας. Οι ερευνητές χρησιμοποιούν συχνά πλατφόρμες όπως το Twitter και το Reddit για την προώθηση των εργασιών τους, την ανταλλαγή ιδεών, και τη συνάντηση συνεργατών για νέες έρευνες. Τα κοινωνικά μέσα επίσης προσφέρουν πλατφόρμες για την άμεση ανταλλαγή προτάσεων, διευρύνοντας τον ορίζοντα της επιστημονικής συζήτησης πέρα από τα παραδοσιακά ακαδημαϊκά περιοδικά. Ακόμη διεξάγονται πάρα πολλές έρευνες στα κοινωνικά μέσα όπως η ανίχνευση bots [7] [8], ψευδών ειδήσεων και πολλές άλλες.

1.2.1 Reddit

Μία ιδιαίτερα ενδιαφέρουσα πλατφόρμα είναι το Reddit. Παρότι δεν είναι ένα ακαδημαϊκό εργαλείο, πολλά φόρουμ ή "subreddits" όπως το r/AskScience ή το r/AskHistorians προσφέρουν μια πλατφόρμα για ερευνητές να ανταλλάξουν γνώσεις, να διευκρινίσουν παρεξηγήσεις και να συμμετάσχουν σε πλούσιες συζητήσεις με κοινό εκτός του ακαδημαϊκού περιβάλλοντος. Τα μέλη της κοινότητας του Reddit συχνά εκφράζουν απορίες και ερωτήματα που μπορεί να οδηγήσουν τους ερευνητές σε νέες προοπτικές ή να επιβεβαιώσουν τη σημασία και την κατεύθυνση των τρεχόντων εργασιών τους. Εκτός από το ρόλο του Reddit ως πλατφόρμα για την ανταλλαγή επιστημονικών πληροφοριών, η αξία του για τα νευρωνικά μοντέλα είναι εξί-

σου σημαντική. Το Reddit, λόγω της τεράστιας ποικιλίας των δημοσιεύσεων, των συζητήσεων, και των αλληλεπιδράσεων των χρηστών του, προσφέρει μια πλούσια πηγή δεδομένων που μπορεί να χρησιμοποιηθεί για την εκπαίδευση και τη βελτίωση νευρωνικών δικτύων. Σε συγκεκριμένες περιπτώσεις, τα δεδομένα από το Reddit χρησιμοποιούνται για να εκπαιδεύσουν μοντέλα στον εντοπισμό των διαθέσεων των χρηστών, στην κατανόηση των πολυπλοκότητων της ανθρώπινης γλώσσας, και ακόμη και στην παραγωγή κειμένου. Δεδομένα από φόρουμ που καλύπτουν ειδικά θέματα μπορεί να χρησιμοποιηθούν για την εκπαίδευση εξειδικευμένων μοντέλων, ενώ οι συζητήσεις ανάμεσα στα μέλη του Reddit μπορούν να παράσχουν δεδομένα για την ανάλυση των σχέσεων μεταξύ των χρηστών.

Συνοψίζοντας, τα κοινωνικά μέσα δεν είναι απλώς εργαλεία ή πλατφόρμες για την ανταλλαγή περιεχομένου. Είναι κάτι πολύ περισσότερο. Είναι το πρίσμα μέσω του οποίου πολλοί βλέπουν τον κόσμο, αντιλαμβάνονται τον εαυτό τους και συνδέονται με τους άλλους. Ως εκ τούτου, η κατανόηση της δυναμικής τους, της επίδρασής τους και των προκλήσεων τους είναι ουσιώδης για την εξερεύνηση της σύγχρονης κοινωνίας μας.

1.3 Πρόσφατη έρευνα και περιορισμοί

Τα τελευταία χρόνια, το πεδίο της ανίχνευσης της κατάθλιψης σε αναρτήσεις στα Μέσα Κοινωνικής Δικτύωσης έχει γνωρίσει σημαντικές εξελίξεις με την εμφάνιση προηγμένων μοντέλων μηχανικής μάθησης, που φαίνονται πολλά υποσχόμενα. Ωστόσο, αυτά τα μοντέλα έχουν σημαντικούς περιορισμούς. Καταρχάς, πολλά εξαρτώνται από παραδοσιακούς αλγόριθμους μηχανικής μάθησης, όπως η Λογιστική Παλινδρόμηση (LR) [9] [10], οι Μηχανές Υποστήριξης (SVM) [11] [12] [13], τα Δέντρα Απόφασης [11] [14] [15], τα Τυχαία Δάση (RF) [16] [11] [17], το AdaBoost [18] [19], και το k-Κοντινότεροι Γείτονες (k-NN) [13] [11], τα οποία, αν και χρήσιμα, μπορεί να ελλείπουν την εξελιγμένη πολυπλοκότητα που απαιτείται για την πλήρη κατανόηση των πολυποίκιλων δεδομένων κειμένου που σχετίζονται με την ψυχική υγεία, οδηγώντας σε μη βέλτιστες επιδόσεις. Για να αντιμετωπίσουν αυτόν τον περιορισμό, οι ερευνητές έχουν στραφεί σε πιο προηγμένες τεχνικές, συμπεριλαμβανομένων μεθόδων βαθιάς μάθησης όπως το Long Short-Term Memory (LSTM) [20] [21] [22], τα Συνελικτικά Νευρωνικά Δίκτυα (CNN) [23] [24] [25] [26], και στους μετασχηματιστές (transformers) [27] [14] [28] [29] [30] [31] όπως το BERT[32]. Αυτές οι μέθοδοι βαθιάς μάθησης έχουν δείξει πολλά υποσχόμενες στον καταγραφή περίπλοκων προτύπων στο κείμενο και στην επίτευξη υψηλότερων επιπέδων κατανόησης του νοήματος του κειμένου. Ωστόσο, απλούστερα νευρωνικά δίκτυα που δεν χρησιμοποιούν μετασχηματιστές ενδέχεται να μην αποτυπώνουν πλήρως αυτά τα περίπλοκα πρότυπα και, συνεπώς, να μην αποδίδουν τόσο αποτελεσματικά όσο τα μοντέλα μετασχηματιστών, τα οποία προεκπαιδεύονται με μεγάλα σώματα κειμένων [33] [34]. Από την άλλη πλευρά, πολλοί ερευνητές έχουν ενσωματώσει μετασχηματιστές στην εργασία τους, αλλά συχνά παραμελούν τη χρήση άλλων πολύτιμων μεταδεδομένων και γλωσσικών δεικτών που θα μπορούσαν ενδεχομένως να βελτιώσουν την κατάταξη των ατόμων ως καταθλιπτικών ή όχι.

1.4 Συνεισφορά της διπλωματικής

Για να ξεπεραστούν οι περιορισμοί των παραδοσιακών αλγορίθμων μηχανικής μάθησης και να εκμεταλλευτούν οι δυνατότητες των μοντέλων μετασχηματιστών, αυτή η έρευνα παρουσιάζει μια νέα αρχιτεκτονική νευρωνικού δικτύου που έχει σχεδιαστεί ειδικά για την ανίχνευση της κατάθλιψης σε αναρτήσεις σε κοινωνικά μέσα ενημέρωσης. Το προτεινόμενο μοντέλο βασίζεται στη δύναμη του DistilBERT base uncased [35], ένα δημοφιλές μοντέλο μετασχηματιστή, με την εξαγωγή πληροφοριών από τα τελευταία τέσσερα στρώματα του. Αυτά τα στρώματα πολλαπλασιάζονται στη συνέχεια με βάρη που εκπαιδεύονται, με αποτέλεσμα να δημιουργείται μια πλούσια αναπαράσταση που αποτυπώνει τα ουσιώδη γλωσσικά πρότυπα του κειμένου. Οι μονάδες CLS (classification) που προκύπτουν από το BERT, ως αναπαραστάσεις της περίληψης των φράσεων, συνενώνονται στη συνέχεια με πρόσθετα μεταδεδομένα και γλωσσικούς δείκτες. Αυτή η στρατηγική συγχώνευση των συστηματικών αναπαραστάσεων και των βοηθητικών χαρακτηριστικών ενισχύει την ικανότητα του μοντέλου να κατανοήσει πλήρως κάθε ανάρτηση, επιτρέποντάς του να εντοπίσει πιθανές ενδείξεις κατάθλιψης με μεγαλύτερη ακρίβεια. Για να βελτιώσουμε περαιτέρω τη γενίκευση και την ανθεκτικότητα του μοντέλου, εισάγονται στρώματα απόρριψης για τον προληπτικό περιορισμό της υπερεκπαίδευσης κατά την εκπαίδευση. Μετά την ενσωμάτωση των στρωμάτων απόρριψης, το αποτέλεσμα περνά από ένα πολυεπίπεδο perceptron (MLP), το οποίο λειτουργεί ως κεφαλή ταξινόμησης. Το MLP σχεδιάστηκε για να λαμβάνει την τελική απόφαση σχετικά με την κατάσταση της ανάρτησης ως προς την κατάθλιψη. Συνδυάζοντας τις αναπαραστάσεις που βασίζονται σε μετασχηματιστές με προσεκτικά επιλεγμένα χαρακτηριστικά και ενσωματώντας στρώματα απόρριψης, το προτεινόμενο νευρωνικό δίκτυο επιτυγχάνει ένα υψηλότερο επίπεδο πολυπλοκότητας σε σύγκριση με τους παραδοσιακούς αλγορίθμους, αποφεύγοντας ταυτόχρονα την υπερεξάρτηση από μεγάλα προεκπαιδευτικά σώματα. Αυτό επιτρέπει στο μοντέλο να αντιμετωπίσει αποτελεσματικά τις περιπλοκότητες των δεδομένων κειμένου που σχετίζονται με την ψυχική υγεία, με αποτέλεσμα τη βελτιωμένη απόδοση στον εντοπισμό κατάθλιψης σε αναρτήσεις σε κοινωνικά μέσα ενημέρωσης. Η ενσωμάτωση επιπρόσθετων μεταδεδομένων και γλωσσικών δεικτών συμβάλλει επίσης στην ικανότητα του μοντέλου να προσαρμόζεται στην κάθε ανάρτηση, αποτυπώνοντας σημαντικές λεπτομέρειες που ίσως διαφεύγουν αλλιώς, κάνοντας έτσι μια πολύτιμη συνεισφορά στον τομέα της ανίχνευσης της κατάθλιψης στο περιεχόμενο κοινωνικών μέσων ενημέρωσης.

Οι κύριες συνεισφορές μας μπορούν να συνοψιστούν ως εξής:

- Δυνατότητα συγχώνευσης των λεξικών αναπαραστάσεων από το BERT με πολύτιμα μεταδεδομένα και γλωσσικούς δείκτες.
- Εισαγωγή προσέγγισης μέσης τιμής για την εξαγωγή κρυφών καταστάσεων από διάφορες στρώσεις μετασχηματιστών, αποτυπώνοντας αποτελεσματικά απλές και πολύπλοκες σημασιολογικές πληροφορίες σε μικρό μέγεθος διανύσματος.
- Παρουσίαση για το πώς μια υβριδική προσέγγιση, συνδυάζοντας ένα μοντέλο μετασχηματιστή με ένα πολυεπίπεδο perceptron, μπορεί να χρησιμοποιηθεί αποτελεσματικά σε εργασίες ταξινόμησης κειμένου για την ψυχική υγεία.

1.5 Δομή της εργασίας

Η παρούσα διπλωματική εργασία στοχεύει στην ανάλυση, παρουσίαση και εφαρμογή μεθόδων τεχνητής νοημοσύνης σε ένα σύνολο δεδομένων, με σκοπό την αναγνώριση περιπτώσεων κατάθλιψης. Η εργασία έχει δομηθεί σε κεφάλαια που οδηγούν τον αναγνώστη καθ' όλη τη διάρκεια της ερευνητικής διαδικασίας.

Συγκεκριμένα:

- **Κεφάλαιο 1:** Παρουσιάζεται μια γενική εισαγωγή στο θέμα της εργασίας, τις ψυχικές διαταραχές, την κατάθλιψη και τα μέσα κοινωνικής δικτύωσης, καθώς και τους κύριους στόχους της. Ακόμα παρουσιάζεται με μικρή επισκόπηση της πρόσφατης έρευνας, οι περιορισμοί της και ποιές είναι οι συνεισφορές της παρούσας διπλωματικής.
- **Κεφάλαιο 2:** Πραγματοποιείται μια βιβλιογραφική ανασκόπηση των τεχνικών και μεθόδων που χρησιμοποιούνται στην εργασία.
- **Κεφάλαιο 3:** Εξετάζεται το θεωρητικό πλαίσιο των τεχνικών και των μεθόδων που θα χρησιμοποιηθούν και χρησιμοποιήθηκαν στην βιβλιογραφία. Ακόμη Εξετάζεται το θεωρητικό πλαίσιο πίσω από την επεξεργασία φυσικής γλώσσας.
- **Κεφάλαιο 4:** Παρουσιάζεται η μεθοδολογία που προτείνεται, από το σύνολο δεδομένων και την προεπεξεργασία του έως και την εφαρμογή των τεχνικών. Ακόμα περιγράφονται οι πειραματικές διαδικασίες, το περιβάλλον που τρέχουν τα μοντέλα και διάφορες διατάξεις του μοντέλου
- **Κεφάλαιο 5:** Περιγράφονται τα αποτελέσματα που προκύπτουν από την εφαρμογή της προτεινόμενης μεθοδολογίας και των διάφορων διατάξεων.
- **Κεφάλαιο 6:** Συνοψίζονται τα κύρια συμπεράσματα της εργασίας και προτείνονται πιθανές μελλοντικές κατευθύνσεις.

Η εργασία στοχεύει στην παροχή μιας ολοκληρωμένης εικόνας της έρευνας, των μεθόδων, των τεχνικών και των αποτελεσμάτων που προέκυψαν.

Κεφάλαιο 2

Βιβλιογραφική Ανασκόπηση

Ο εντοπισμός της κατάθλιψης μέσω του περιεχομένου των μέσων κοινωνικής δικτύωσης είναι ένας αυξανόμενος τομέας ερευνητικού ενδιαφέροντος, που επιβάλλεται από τον αυξανόμενο ρυθμό της παγκόσμιας κατάθλιψης και την αύξηση της χρήσης των μέσων κοινωνικής δικτύωσης. Αυτό το διεπιστημονικό ερευνητικό πεδίο συνδυάζει τη μηχανική μάθηση, την επεξεργασία φυσικής γλώσσας και τις μελέτες ψυχικής υγείας για να δημιουργήσει νέους και αποτελεσματικούς τρόπους αναγνώρισης πιθανών προβλημάτων ψυχικής υγείας. Αυτή η ενότητα είναι αφιερωμένη στην παρουσίαση μιας επισκόπησης της σχετικής έρευνας σε αυτόν τον τομέα. Οργανώσαμε την ανασκόπηση της βιβλιογραφίας σε τρία κύρια θέματα με βάση τις μεθόδους που χρησιμοποιήθηκαν: Μοντέλα μηχανικής μάθησης και στατιστική ανάλυση, νευρωνικά δίκτυα και τέλος μετασχηματιστές. Αυτές οι υποενότητες παρουσιάζουν τη ανάπτυξη μεθοδολογιών σε αυτόν τον τομέα, από τα πιο παραδοσιακά μοντέλα μηχανικής μάθησης έως τις πιο πρόσφατες τεχνικές που βασίζονται σε μετασχηματιστές.

2.1 Έρευνες που χρησιμοποιούν μόνο μοντέλα μηχανικής μάθησης

Πολλαπλές μελέτες τα τελευταία χρόνια έχουν προσεγγίσει το πρόβλημα της ανίχνευσης της κατάθλιψης σε αναρτήσεις στα μέσα κοινωνικής δικτύωσης χρησιμοποιώντας μια ποικιλία μοντέλων μηχανικής μάθησης (ML) και μεθόδων στατιστικής ανάλυσης. Αυτή η υπό ενότητα εξετάζει τις βασικές τεχνικές και τα σύνολα δεδομένων που χρησιμοποιούνται παρέχοντας μια βάση για την κατανόηση της συμβολής τους σε αυτήν την διπλωματική εργασία.

Οι Burdisso et al.[36] ανέπτυξαν ένα καινοτόμο μοντέλο ταξινόμησης κειμένου που ονομάζεται SS3, το οποίο χρησιμοποιείται για την ανίχνευση της κατάθλιψης μέσω των ροών των μέσων κοινωνικής δικτύωσης (social media feed). Χρησιμοποίησαν το σύνολο δεδομένων eRisk 2017 [37], ένα διάσημο και συχνά χρησιμοποιημένο σύνολο δεδομένων. Το μοντέλο SS3 περιλαμβάνει τη δημιουργία ενός διανύσματος εμπιστοσύνης για την αξιολόγηση του πόσο μια λέξη ανήκει σε μια συγκεκριμένη κατηγορία και στη συνέχεια συνοψίζει το παρόν διάνυσμα σε ανώτερα επίπεδα κειμένου (πρόταση, παράγραφος και άλλα).

Σε μια διαφορετική προσέγγιση, οι Asad et al. [38] χρησιμοποίησαν δεδομένα από το Twitter και το Facebook για να εκπαιδεύσουν έναν ταξινομητή Naive Bayes. Ακόμα χρησιμοποίησαν term frequency–inverse document frequency (TF-IDF) για εξαγωγή χαρακτηριστικών μετά από προ επεξεργασία που περιλάμβανε την αφαίρεση μη αλφαριθμητικών χαρακτήρων, retweet, διευθύνσεων URL και των @αναφορών (@user).

Οι Islam et al. [13] βασίστηκαν σε σχόλια από μια σελίδα στο Facebook για να ανιχνεύσουν σημάδια κατάθλιψης χρησιμοποιώντας τεχνικές μηχανικής μάθησης. Χρησιμοποίησαν το λεξικό Linguistic Inquiry and Word Count (LIWC) [39] για εξαγωγή χαρακτηριστικών και μια σειρά από αλγόριθμους ML, συμπεριλαμβανομένων των μηχανών διανυσμάτων υποστήριξης (Support Vector Machine, SVM), δέντρων αποφάσεων (decision trees) και k-πλησιέστερων γειτόνων (kNN).

Λαμβάνοντας μια πιο εστιασμένη προσέγγιση, οι Aragón et al. [40] στόχευσαν στην ανίχνευση της ανορεξίας και της κατάθλιψης μέσω συναισθηματικών μοτίβων στα κείμενα των μέσων κοινωνικής δικτύωσης. Χρησιμοποίησαν το μοντέλο Bag-of-Emotions-Space (BOSE) στο σύνολο δεδομένων eRisk. Αναφέρονται όμως στον περιορισμό στη δυσκολία ερμηνείας των αποτελεσμάτων από την ανάλυση κειμένου και εξακρίβωσης της ακρίβειας τους.

Οι Kamite και Kamble [16] συνέλεξαν χειροκίνητα δεδομένα χρησιμοποιώντας το API του Twitter και εφάρμοσαν μοντέλα Random Forest και Naive Bayes για ανίχνευση κατάθλιψης. Τόνισαν την ανάγκη για συλλογή περισσότερων δεδομένων, καθώς η ποιότητα και η ποσότητα τους επηρεάζουν άμεσα τα αποτελέσματα.

Μια μελέτη των Eichstaedt et al. [9] χρησιμοποίησε δεδομένα από το Facebook, από συναινούντες ασθενείς που συμπλήρωσαν ερωτηματολόγια, και εφάρμοσαν Logistic Regression, χρησιμοποιώντας χαρακτηριστικά όπως το μέγεθος των προτάσεων, τη συχνότητα που δημοσιεύουν στο Facebook και άλλα χρησιμοποιώντας το LIWC.

Εν τω μεταξύ, οι Iavarone και Monrealle [41] βασίστηκαν στα δεδομένα του Reddit για την ανίχνευση της κατάθλιψης, χρησιμοποιώντας k-means συσταδοποίηση σε χαρακτηριστικά από το LIWC.

Ταυτόχρονα οι Aldarwish και Ahmad [12] εκπαιδεύσαν μοντέλα SVM και Naive Bayes σε δεδομένα που συλλέχθηκαν χειροκίνητα και που προ επεξεργάστηκαν με την αφαίρεση stop-words και stemming.

Οι Coppersmith et al. [42] ανέλυσαν τα δεδομένα του Twitter χρησιμοποιώντας έναν log-linear classifier και αξιοποιώντας μια πληθώρα χαρακτηριστικών όπως το LIWC, το μοντέλο γλώσσας 1 και 5 λέξεων (n-gram), το ποσοστό tweets που έγραφαν μέσα στην μέρα και τον αριθμό των @αναφορών.

Οι Huang et al. [10] μελέτησαν την ανίχνευση αυτοκτονικού ιδεασμού μέσω ανάλυσεων κοινωνικών μέσων, βασιζόμενοι σε χαρακτηριστικά n-gram και LIWC σε δεδομένα από την πλατφόρμα Sina Weibo, μια κινεζική πλατφόρμα κοινωνικών μέσων. Συγκεκριμένα χρησιμοποίησαν Logistic Regression και Decision Trees ως τα μοντέλα τους.

Τέλος, οι Leiva και Freire [11] χρησιμοποίησαν έναν συνδυασμό μοντέλων (Logistic Regression, SVM, kNN, Random Forests) στο γνωστό σύνολο δεδομένων eRisk. Εξήγα-

γαν χαρακτηριστικά TF-IDF και χρησιμοποίησαν VADER [43] για ανάλυση συναισθήματος ώστε να αποκτήσουν μια αναπαράσταση των δεδομένων. Αυτές οι μελέτες αποκαλύπτουν ένα ευρύ φάσμα μεθόδων και συνόλων δεδομένων που χρησιμοποιούνται για την ανίχνευση της κατάθλιψης σε αναρτήσεις στα μέσα κοινωνικής δικτύωσης. Υπογραμμίζουν τη δυνατότητα συνδυασμού της μηχανικής μάθησης και της στατιστικής ανάλυσης για την παροχή πληροφοριών και έγκαιρων προειδοποιήσεων σχετικά με τις καταστάσεις ψυχικής υγείας.

2.2 Έρευνες που χρησιμοποιούν νευρωνικά δίκτυα

Τα νευρωνικά δίκτυα έχουν κερδίσει σημαντική προσοχή στην ανίχνευση της κατάθλιψης και χρησιμοποιούνται ευρέως. Αυτά τα συστήματα μαθαίνουν πολύπλοκα μοτίβα από τεράστιες ποσότητες δεδομένων, επιτρέποντάς τους να κάνουν πιο ακριβείς προβλέψεις.

Οι Wu et al. [21] χρησιμοποίησαν μοντέλα Long Short-Term Memory (LSTM) σε δεδομένα καιρού, αναρτήσεις Facebook και δοκιμές ελέγχου CES-D. Το CES-D (Center for Epidemiologic Studies Depression Scale) [44] είναι ένα εργαλείο προσυμπτωματικού ελέγχου που αναπτύχθηκε για τη μέτρηση της καταθλιπτικής συμπτωματολογίας στον γενικό πληθυσμό. Μετράει 20 στοιχεία και αξιολογεί τη συχνότητα των συμπτωμάτων που σχετίζονται με την κατάθλιψη την περασμένη εβδομάδα, όπου υψηλότερες βαθμολογίες που υποδεικνύουν μεγαλύτερα συμπτώματα κατάθλιψης. Οι Wu et al. πρότειναν τη χρήση LSTM για τις αναρτήσεις και ύστερα την συνένωση των υπόλοιπων δεδομένων χρησιμοποιώντας δύο πλήρως συνδεδεμένα επίπεδα (fully connected layers). Η μελλοντική εργασία τους στοχεύει να χρησιμοποιήσει Συνελικτικά Νευρωνικά Δίκτυα (CNN) για να ενισχύσει την απόδοση και να ανιχνεύσει άλλες κρίσιμες ψυχικές διαταραχές όπως το άγχος και οι διπολικές διαταραχές και επίσης να αναζητήσει διάφορες θεραπείες με χρήση νευρωνικών δικτύων.

Σε μια άλλη μελέτη, οι Chiong et al. [45] κατέδειξαν την αποτελεσματικότητα των νευρωνικών δικτύων στην ανίχνευση της κατάθλιψης, χρησιμοποιώντας ταξινομητές μηχανικής μάθησης σε κείμενα κοινωνικών μέσων. Χρησιμοποίησαν νευρωνικά δίκτυα και μια σειρά αλγορίθμων, συμπεριλαμβανομένων των LR, SVM, MLP, Decision Trees, Random Forest, AdaBoost, Bag-Of-Words και gradiend boosting. Ενδιαφέρον όμως είναι ότι χρησιμοποίησαν τεχνολογίες όπως SentiWordNet και SenticNet, ενώ συμπεριέλαβαν σκορ του πόσο ευανάγνωστες είναι οι αναρτήσεις με χρήση του TextStat. Ωστόσο, αναγνώρισαν τον περιορισμό της προσέγγισής τους στα επισημασμένα σύνολα δεδομένων, προτείνοντας τη συμπερίληψη μη εποπτευόμενων ταξινομητών ως μελλοντική έρευνα.

Οι Chen και Wang [27] πρότειναν ένα συνδυασμένο μοντέλο LSTM-CNN για ανάλυση συναισθήματος στα δεδομένα Twitter. Πρότειναν επίσης να συνδέσουν το μοντέλο τους με ετικέτες τμήματος ομιλίας (POS) ώστε να βελτιώσουν την απόδοση του μοντέλου.

Σε μια διαφορετική προσέγγιση, οι Safa et al. [19] χρησιμοποίησαν μια ποικιλία αλγορίθμων μηχανικής μάθησης, συμπεριλαμβανομένων Decision Trees, SVM, Random Forest, Ridge Classifier, AdaBoost, CatBoost και MLP.

2.3 Έρευνες που χρησιμοποιούν μετασχηματιστές

Μια πρόσφατη τάση στις εργασίες επεξεργασίας φυσικής γλώσσας, συμπεριλαμβανομένης της ανίχνευσης κατάθλιψης στα μέσα κοινωνικής δικτύωσης, είναι η χρήση μετασχηματιστών (transformers), οι οποίοι είναι πιο προηγμένα μοντέλα νευρωνικών δικτύων που βασίζονται σε μεγάλο βαθμό στους μηχανισμούς προσοχής. Αυτά τα μοντέλα έχουν δείξει αξιοσημείωτα αποτελέσματα, όπως αποδεικνύεται από τις ακόλουθες μελέτες.

Στη μελέτη τους, οι Haque et al. [46] διερεύνησαν την εφαρμογή μοντέλων που βασίζονται σε μετασχηματιστές, συγκεκριμένα των BERT, ALBERT, ROBERTA και XLNET, για την ανίχνευση αυτοκτονικού ιδεασμού. Επισήμαναν τους περιορισμούς ενός μικρού και προκατειλημμένου συνόλου δεδομένων και πρότειναν τη χρήση περισσότερων δεδομένων και εστίαση σε γλωσσικά χαρακτηριστικά για μελλοντικές μελέτες.

Ομοίως, οι Malviya et al. [47] χρησιμοποίησαν μοντέλα μετασχηματιστών όπως BERT, DistilBert, Roberta, Electra και XLNet για ανίχνευση κατάθλιψης στα μέσα κοινωνικής δικτύωσης. Τόνισαν την ανάγκη για ένα μεγαλύτερο σύνολο δεδομένων από πολλαπλές πηγές για καλύτερα αποτελέσματα.

Οι Uban et al. [14] ενσωμάτωσαν επίσης τα μοντέλα BERT, RoBERTa και ALBERT στη συναισθηματική και γνωστική ανάλυση των διαταραχών ψυχικής υγείας από δεδομένα κοινωνικών μέσων. Πρότειναν μια πιο εις βάθος ανάλυση των χαρακτηριστικών σε διαφορετικά επίπεδα για ένα αξιόπιστο συμπέρασμα σχετικά με τους γλωσσικούς δείκτες των διαταραχών αυτών.

Οι Wang et al. [48] χρησιμοποίησαν μοντέλα μετασχηματιστών όπως BERT, RoBERTa και XLNET για να προβλέψουν τον κίνδυνο κατάθλιψης στα κινεζικά Microblogs, υποδεικνύοντας μελλοντικές προεκτάσεις για την επέκταση του συνόλου δεδομένων και την συμβολή της ιατρικής γνώσης σχετικά με την κατάθλιψη.

Οι Kabir et al. [30] χρησιμοποίησαν μοντέλα μετασχηματιστών, συμπεριλαμβανομένων των BERT και DistilBERT για τον εντοπισμό της σοβαρότητας της κατάθλιψης σε δεδομένα Twitter. Η μελλοντική τους εργασία περιλαμβάνει τον καλύτερο καθαρισμό δεδομένων πριν από την εκπαίδευση, την χρήση περισσότερων μετα-δεδομένων, την αύξηση των εκπαιδευσιμων δεδομένων με χρήση τεχνικών επαύξησης και την αύξηση των ετικετών.

Οι Chen et al. [27] χρησιμοποίησαν μια υβριδική προσέγγιση μεταξύ του SBERT και του CNN για την ανίχνευση χρηστών του Reddit με κατάθλιψη.

Οι Murarka et al. [31] εφάρμοσαν τα μοντέλα BERT και RoBERTa για τον εντοπισμό και την ταξινόμηση ψυχικών ασθενειών στα μέσα κοινωνικής δικτύωσης. Εξέφρασαν την πρόθεσή τους να εμπλέξουν ιατρούς-ειδικούς του τομέα σε μελλοντική έρευνα και να δημιουργήσουν έναν ταξινομητή πολλαπλών ετικετών για να αναγνωρίσουν χρήστες που μπορεί να υποφέρουν από πολλαπλές ψυχικές ασθένειες ταυτόχρονα.

Οι Ameer et al. [15] χρησιμοποίησαν διάφορα μοντέλα μετασχηματιστών, συμπεριλαμβανομένων των BERT, XLNET και ROBERTA, για την ταξινόμηση ψυχικών ασθενειών στα μέσα κοινωνικής δικτύωσης. Πρότειναν τη δημιουργία ενός συνόλου δεδομένων πολλαπλών ετικετών για προβλήματα ψυχικών ασθενειών και τη διερεύ-

νηση καλύτερων μοντέλων για τη βελτίωση της απόδοσης ταξινόμησης.

Οι Rao et al. [29] ενσωμάτωσαν το BERT στο μοντέλο αναπαράστασης ιεραρχικών θέσεων, και δημιούργησαν το MGL-CNN, για την αναγνώριση ατόμων με κατάθλιψη σε διαδικτυακά φόρουμ, προτείνοντας τη χρήση αυτού του μοντέλου για ανάλυση συναισθήματος σε επίπεδο χρήστη και όχι μόνο ανάρτησης.

Στη μελέτη των Ilias και Askounis [49], εξετάζεται η ανάγκη για την πρόωρη ανίχνευση του στρες και της κατάθλιψης με τη χρήση των κοινωνικών μέσων. Οι ερευνητές προτείνουν δύο πλαίσια πολλαπλών εργασιών μάθησης που χρησιμοποιούν την κατάθλιψη ως την κύρια εργασία και το στρες ως βοηθητική εργασία. Η πρώτη προσέγγιση χρησιμοποιεί ένα κοινό επίπεδο BERT για την επεξεργασία των δημοσιεύσεων, ενώ η δεύτερη προσέγγιση χρησιμοποιεί συνδυαστικά επίπεδα με εξισορρόπηση προσοχής. Η μελέτη διεξάγει πειράματα σύγκρισης με υπάρχουσες έρευνες, μεμονωμένη εκπαίδευση και μεταφορά μάθησης, και δείχνει πολλαπλά πλεονεκτήματα των προτεινόμενων προσεγγίσεων έναντι των υφιστάμενων τεχνικών.

Τέλος, οι Lin et al. [28] χρησιμοποίησαν το BERT για να ενσωματώσουν προτάσεις στο σύστημα ανίχνευσης κατάθλιψης, SenseMood, από δεδομένα Twitter. Η πρόκλησή τους ήταν να αντλήσουν το μοντέλο αναπαράστασης λέξεων από το BERT και πρότειναν μια μέθοδο πολυτροπικής σύντηξης χρησιμοποιώντας τανυστή χαμηλής κατάταξης για μελλοντική έρευνα. Μπορεί να συναχθεί το συμπέρασμα ότι τα μοντέλα μετασχηματιστών, παρά ορισμένους περιορισμούς, έχουν δείξει πολλά υποσχόμενα αποτελέσματα για την ανίχνευση της κατάθλιψης από τα δεδομένα των μέσων κοινωνικής δικτύωσης. Οι μελλοντικές εργασίες σε αυτόν τον τομέα ενδέχεται να περιλαμβάνουν τη βελτίωση αυτών των μοντέλων, τη βελτίωση της ποιότητας του συνόλου δεδομένων και την εστίαση σε προσεγγίσεις ταξινόμησης πολλαπλών ετικετών.

Συμπερασματικά, η μελέτη της ανίχνευσης της κατάθλιψης στα μέσα κοινωνικής δικτύωσης είναι ένα ταχέως εξελισσόμενο πεδίο, με προόδους στις μεθόδους από τα παραδοσιακά μοντέλα ML έως τα νευρωνικά δίκτυα και τα πιο πρόσφατα, τα μοντέλα μετασχηματιστών. Κάθε μία από αυτές τις προσεγγίσεις έχει πλεονεκτήματα αλλά και τους περιορισμούς της, όπως αποδεικνύεται στο εύρος των μελετών που συζητούνται σε αυτήν την ενότητα. Παρά την πρόοδο που έχει σημειωθεί, εξακολουθούν να υπάρχουν προκλήσεις, συμπεριλαμβανομένης της ανάγκης για μεγαλύτερα και πιο διαφορετικά σύνολα δεδομένων και καλύτερο χειρισμό μη ισορροπημένων δεδομένων στα μέσα κοινωνικής δικτύωσης. Αυτή η ανασκόπηση της βιβλιογραφίας υπογραμμίζει τη σημασία της συνεχούς έρευνας για τη βελτίωση της ακρίβειας και της αποτελεσματικότητας στον εντοπισμό ενδείξεων κατάθλιψης και άλλων ψυχικών ασθενειών μέσω αναρτήσεων στα μέσα κοινωνικής δικτύωσης, οι οποίες μπορούν ενδεχομένως να σώσουν ζωές παρέχοντας έγκαιρες προειδοποιήσεις. Με τις συνεχείς εξελίξεις στη μηχανική εκμάθηση και την επεξεργασία φυσικής γλώσσας, αναμένουμε να αναδυθούν πιο καινοτόμες προσεγγίσεις σε αυτόν τον κρίσιμο ερευνητικό τομέα.

Κεφάλαιο 3

Τεχνητή Νοημοσύνη

3.1 Μηχανική Μάθηση

Η μηχανική μάθηση (Machine learning, ML) είναι πεδίο της τεχνητής νοημοσύνης που ασχολείται με την ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν σε υπολογιστές να μαθαίνουν και να προσαρμόζονται από δεδομένα, χωρίς να χρειάζεται κάποιος άνθρωπος να προγραμματίσει ρητά έναν αλγόριθμο για να λύσει ένα συγκεκριμένο πρόβλημα [50] [51]. Τέτοιοι αλγόριθμοι τροφοδοτούνται από δεδομένα και ανακαλύπτοντας μοτίβα κάνουν τις κατάλληλες προβλέψεις ώστε το αποτέλεσμα τους να είναι όσο πιο κοντά γίνεται στα αποτελέσματα των δεδομένων.

Η μηχανική μάθηση έχει πάρα πολλές εφαρμογές σε προβλήματα που το να προγραμματίσουμε ρητά τον αλγόριθμο είναι ανέφικτο. Τέτοια προβλήματα αποτελούν την εύρεση ανωμαλιών, την ανίχνευση προτύπων, την επεξεργασία φυσικής γλώσσας και πολλά άλλα. Μπορούμε να την κατηγοριοποιήσουμε ως επιβλεπόμενη μάθηση, μη επιβλεπόμενη και ενισχυτική. Συγκεκριμένα:

- Επιβλεπόμενη (supervised) μάθηση είναι ο πιο κοινός τύπος μάθησης. Για κάθε δεδομένο εισόδου έχουμε και μια ετικέτα που υποδηλώνει σε πιο κατηγορία βρίσκεται. Το μοντέλο μαθαίνει από αυτά τα δεδομένα και μετά εφαρμόζεται σε άλλα, άγνωστα, δεδομένα. Τέτοιοι αλγόριθμοι μπορεί να είναι η γραμμική παλινδρόμηση (linear regression), η λογιστική παλινδρόμηση (logistic regression), οι μηχανές υποστήριξης διανυσμάτων (support vector machines, SVM) και τα νευρωνικά δίκτυα.
- Μη Επιβλεπόμενη (Unsupervised) μάθηση η οποία δεν χρειάζεται ετικέτες. Οι αλγόριθμοι ομαδοποιούν δεδομένα με βάση κοινά μοτίβα και σχέσεις μεταξύ των δεδομένων. Η συσταδοποίηση (clustering) και η μείωση διαστατικότητας (dimensionality reduction) είναι οι πιο συνήθεις αλγόριθμοι της.
- Ενισχυτική Μάθηση (Reinforcement learning) η οποία αφορά ένα πράκτορα που μαθαίνει από το να λαμβάνει αποφάσεις αλληλεπιδρώντας με το περιβάλλον του. Ανάλογα με τις ενέργειες του λαμβάνει ποινές ή ανταμοιβές και μαθαίνει θέτοντας ως στόχο την μεγιστοποίηση της συνολικής ανταμοιβής.

3.2 Επιβλεπόμενη Μάθηση

Σε αυτήν την υπό ενότητα θα αναλύσουμε τους αλγόριθμους επιβλεπόμενης μάθησης που χρησιμοποιήθηκαν στην βιβλιογραφία. Οι πληροφορίες πάρθηκαν από πολλές πηγές όπως τα έργα των Kelleher [52], Patterson [53] και Alpaydin [50][51].

3.2.1 Λογιστική Παλινδρόμηση (Logistic Regression)

Η λογιστική παλινδρόμηση είναι μια στατιστική μέθοδος που χρησιμοποιείται κυρίως σε προβλήματα δυαδικής ταξινόμησης, όπου η μεταβλητή αποτελέσματος μπορεί να λάβει ένα από τα δύο πιθανά αποτελέσματα. Σε αντίθεση με τη γραμμική παλινδρόμηση (linear regression), η οποία προβλέπει μια συνεχή μεταβλητή αποτέλεσμα, η λογιστική παλινδρόμηση προβλέπει την πιθανότητα ότι μια δεδομένη περίπτωση ανήκει σε μια συγκεκριμένη κατηγορία. Η βασική αρχή πίσω από τη λογιστική παλινδρόμηση είναι η λογιστική συνάρτηση, γνωστή και ως συνάρτηση σιγμοειδούς, η οποία μπορεί να λάβει οποιοδήποτε πραγματικό αριθμό και να τον μετατρέψει σε μια τιμή μεταξύ 0 και 1. Αυτό το χαρακτηριστικό την καθιστά ιδιαίτερα κατάλληλη για την εκτίμηση πιθανοτήτων.

Η λογιστική συνάρτηση ορίζεται ως $P(Y = 1) = \frac{1}{1+e^{-z}}$, όπου z είναι ένας γραμμικός συνδυασμός των εισαγόμενων χαρακτηριστικών και των σχετικών βαρών. Τα βάρη στη λογιστική παλινδρόμηση συνήθως εκτιμώνται με τη μέθοδο της μέγιστης πιθανοφάνειας. Καθώς ο γραμμικός συνδυασμός z γίνεται μεγάλος στη θετική ή αρνητική κατεύθυνση, η εκτιμώμενη πιθανότητα πλησιάζει το 1 ή το 0, αντίστοιχα. Αυτό διασφαλίζει ότι οι προβλεπόμενες πιθανότητες είναι εντός του εύρους $[0,1]$, καθιστώντας τις ερμηνεύσιμες ως πραγματικές πιθανότητες.

3.2.2 Log-Linear

Το μοντέλο Log-linear είναι ένας τύπος στατιστικού μοντέλου που χρησιμοποιείται κυρίως στην ανάλυση κατηγορικών δεδομένων. Αυτό το μοντέλο είναι μια επέκταση του γενικού γραμμικού μοντέλου, όπου ο λογάριθμος της μεταβλητής απόκρισης μοντελοποιείται ως γραμμικός συνδυασμός των μεταβλητών του προβλέπτη. Η γενική του μορφή είναι η εξής:

$$\exp(c + \sum_i w_i f_i(X))$$

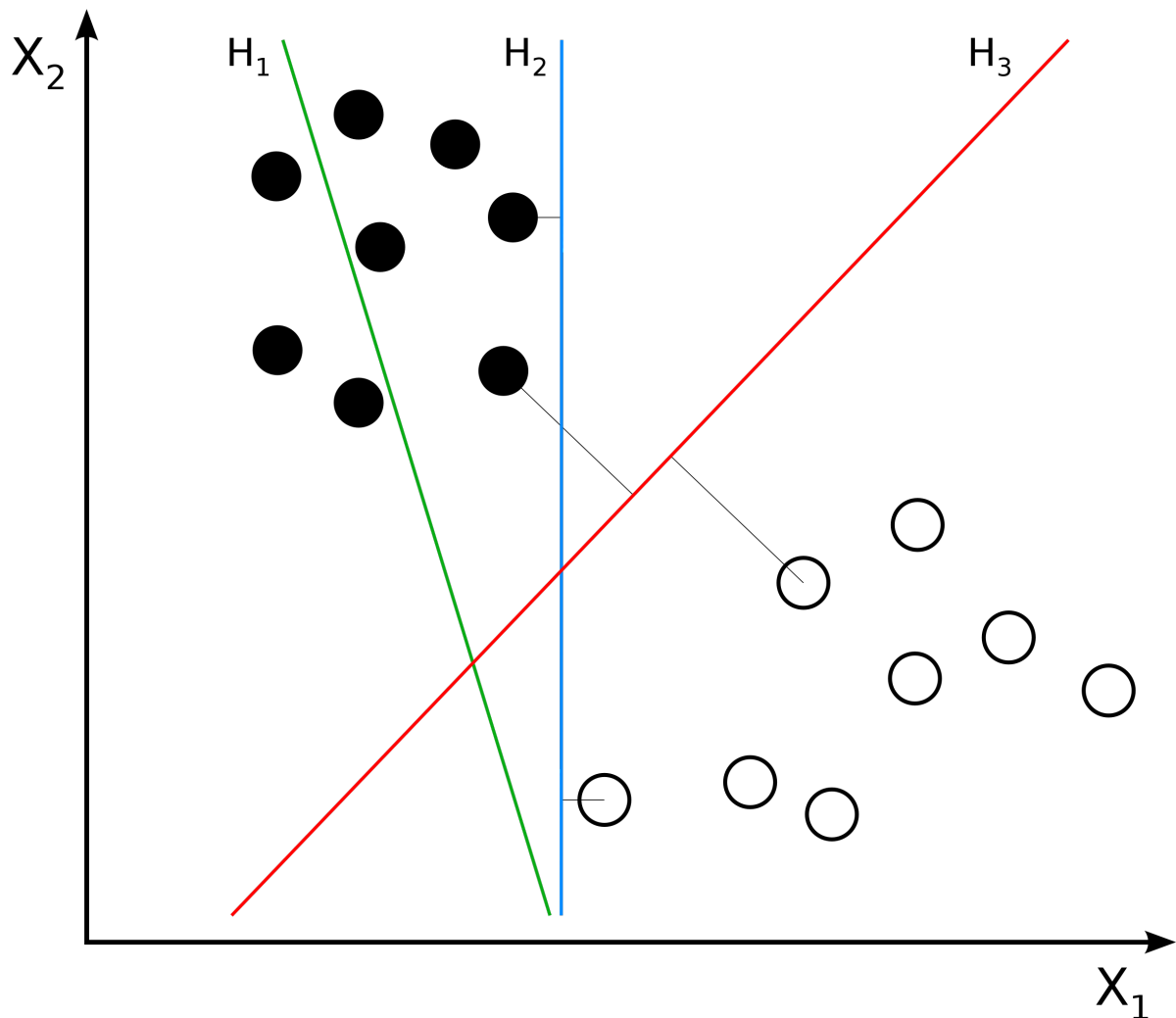
Το κύριο πλεονέκτημα της χρήσης ενός μοντέλου log-linear είναι η ικανότητά του να χειρίζεται πολλαπλασιαστικές σχέσεις μεταξύ των αναμενόμενων συχνοτήτων και των μεταβλητών προβλέπτη.

Όσον αφορά του πίνακες ενδεχομένων αποτελεσμάτων, τα μοντέλα log-linear χρησιμοποιούνται για την εξέταση των σχέσεων μεταξύ πολλαπλών κατηγορικών μεταβλητών. Το μοντέλο εξετάζει τις αλληλεπιδράσεις μεταξύ αυτών των μεταβλητών προσαρμόζοντας τις παρατηρούμενες συχνότητες στον πίνακα στις αναμενόμενες συχνότητες που προέρχονται από το μοντέλο.

Στην πράξη, τα μοντέλα log-linear είναι ευέλικτα και μπορούν να χρησιμοποιηθούν σε διάφορες εφαρμογές, από τη φυσική γλωσσική επεξεργασία έως τη βιοπληροφορική. Μπορούν να μοντελοποιήσουν αλληλεπιδράσεις οποιασδήποτε τάξης και

μπορούν να επεκταθούν για το χειρισμό δεδομένων μέτρησης που έχουν μηδενική περικοπή και υπερδιασπορά. Η ερμηνευτικότητα των συντελεστών του μοντέλου, όταν εκθετοποιούνται, παρέχει πολλαπλασιαστικές επιδράσεις στη μεταβλητή απόκρισης, προσφέροντας πληροφορίες σχετικά με τις σχέσεις μεταξύ των μεταβλητών στο μοντέλο.

3.2.3 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)



Σχήμα 3.1: Μηχανή Διανυσμάτων Υποστήριξης(SVM) [54].

Οι Μηχανές Διανυσμάτων Υποστήριξης (SVM) είναι μια κατηγορία αλγορίθμων εποπτευόμενης μηχανικής μάθησης που χρησιμοποιούνται κυρίως για εργασίες ταξινόμησης και παλινδρόμησης. Στον πυρήνα τους, τα SVM στοχεύουν στην εύρεση του βέλτιστου υπερεπιπέδου που διαχωρίζει καλύτερα τα δεδομένα σε διακριτές κατηγορίες. Σε έναν δισδιάστατο χώρο, αυτό το υπερεπίπεδο είναι μια γραμμή, αλλά σε υψηλότερες διαστάσεις, μπορεί να είναι ένα επίπεδο. Τα "διανύσματα υποστήριξης" είναι τα σημεία δεδομένων που βρίσκονται πλησιέστερα σε αυτό το υπερεπίπεδο και είναι καθοριστικά για τη θέση και τον προσανατολισμό του. Αυτά τα διανύσματα είναι τα πιο δύσκολα στην ταξινόμηση και βρίσκονται στην άκρη των αντίστοιχων κατηγοριών τους.

Η δύναμη των SVM βρίσκεται στην ικανότητά τους να χειρίζονται δεδομένα, που δεν είναι γραμμικά διαχωρίσιμα, μέσω της χρήσης συναρτήσεων πυρήνα. Μια συνάρτηση πυρήνα χαρτογραφεί τα εισαγόμενα δεδομένα σε έναν χώρο υψηλότερης διάστασης, καθιστώντας δυνατή την εύρεση ενός διαχωριστικού υπερεπιπέδου ακόμη και όταν αυτό δεν είναι εφικτό στον αρχικό χώρο. Οι συνήθως χρησιμοποιούμενοι πυρήνες περιλαμβάνουν τους γραμμικούς, πολυωνυμικούς, της ριζικής βάσης (RBF) και τους σιγμοειδείς πυρήνες. Μετατρέποντας έτσι τα δεδομένα, τα SVM μπορούν να αποτυπώσουν περίπλοκες σχέσεις και σύνορα μεταξύ των κατηγοριών χωρίς την ανάγκη για ρητή μηχανική χαρακτηριστικών (feature engineering).

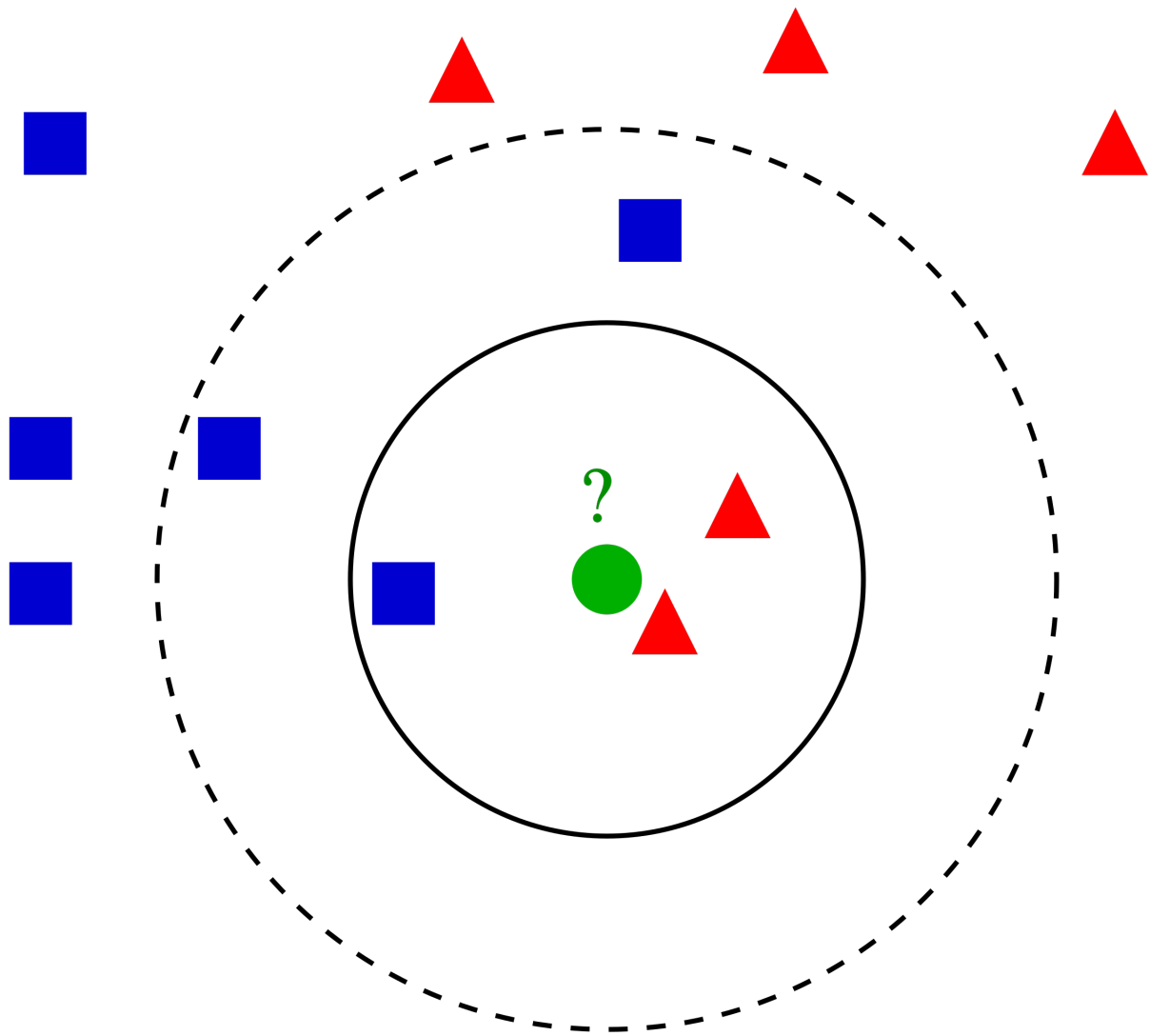
Στην πράξη, τα SVM προσφέρουν ανθεκτικότητα στην παρουσία θορυβώδων δεδομένων και είναι λιγότερο επιρρεπή στην υπερταίριαση (overfitting), ιδιαίτερα όταν το περιθώριο, που είναι η απόσταση μεταξύ του υπερεπιπέδου και των διανυσμάτων υποστήριξης, μεγιστοποιείται. Ωστόσο, μπορούν να είναι υπολογιστικά ακριβά, ιδιαίτερα με μεγάλα σύνολα δεδομένων. Παρά το γεγονός αυτό, η πολυμορφία τους, σε συνδυασμό με την ικανότητά τους να παρέχουν υψηλή ακρίβεια σε διάφορες εφαρμογές, έχει εδραιώσει τα SVM ως ένα βασικό εργαλείο στον τομέα της μηχανικής μάθησης.

3.2.4 Naive Bayes

Ο αλγόριθμος Naive Bayes είναι μια πιθανοτική μέθοδος ταξινόμησης βασισμένη στο θεώρημα του Bayes. Στην ουσία, ένας ταξινομητής Naive Bayes υποθέτει ότι η παρουσία ενός συγκεκριμένου χαρακτηριστικού σε μια κατηγορία είναι ασχετη με την παρουσία οποιουδήποτε άλλου χαρακτηριστικού, εξ ου και ο όρος "αφελής" ("naive"). Παρά την απλότητά του και την αφελή φύση των υποθέσεών του, ο αλγόριθμος συχνά επιδεικνύει εξαιρετικά καλή απόδοση σε διάφορες εφαρμογές, ιδιαίτερα σε εργασίες ταξινόμησης κειμένου όπως η ανίχνευση spam και η ανάλυση συναισθήματος.

Το θεώρημα του Bayes υπολογίζει την πιθανότητα ενός συμβάντος βασισμένο σε προηγούμενες γνώσεις συνθηκών που ενδέχεται να σχετίζονται με το συμβάν. Στο πλαίσιο της ταξινόμησης, ο αλγόριθμος υπολογίζει την πιθανότητα μιας κατηγορίας δεδομένων σε ένα σετ χαρακτηριστικών. Η κατηγορία με την υψηλότερη πιθανότητα επιλέγεται στη συνέχεια ως πρόβλεψη. Ο αλγόριθμος είναι ιδιαίτερα αποδοτικός επειδή απαιτεί μόνο μια μόνο διέλευση των δεδομένων εκπαίδευσης για να εκτιμήσει τις απαραίτητες παραμέτρους, καθιστώντας τον ιδιαίτερα κλιμακούμενο (scalable) και κατάλληλο για μεγάλα σύνολα δεδομένων.

Στην πράξη, υπάρχουν διάφορες εκδοχές του Naive Bayes, όπως ο Gaussian, Multinomial και Bernoulli, η καθεμία ειδικευμένη για συγκεκριμένους τύπους δεδομένων. Ενώ η υπόθεση ανεξαρτησίας χαρακτηριστικών μπορεί να είναι περιορισμένη σε ορισμένα σενάρια, η απλότητα, η αποδοτικότητα και η αποτελεσματικότητα του αλγορίθμου σε χώρους υψηλής διάστασης έχουν καταστήσει τον Naive Bayes ένα βασικό εργαλείο στην κοινότητα της μηχανικής μάθησης.



Σχήμα 3.2: K-nearest neighbors (K-nn) [54].

3.2.5 K-Nearest Neighbors

Ο αλγόριθμος K-nearest neighbors (K-NN) είναι μια μη παραμετρική, μέθοδος μάθησης που χρησιμοποιείται για εργασίες ταξινόμησης και παλινδρόμησης. Η βασική της αρχή είναι ότι παρόμοια σημεία δεδομένων υπάρχουν σε στενή εγγύτητα μεταξύ τους στον χώρο των χαρακτηριστικών. Δεδομένου ενός νέου, μη ταξινομημένου σημείου δεδομένων, το K-NN προσδιορίζει 'k' παραδείγματα εκπαίδευσης που είναι τα πλησιέστερα στο σημείο και επιστρέφει την πιο κοινή τιμή εξόδου μεταξύ τους για ταξινόμηση, ή έναν μέσο όρο σε περίπτωση παλινδρόμησης. Η απόσταση μεταξύ των σημείων δεδομένων, υπολογίζεται με μεθόδους όπως η Ευκλείδεια ή η Μανχάταν απόσταση.

Η επιλογή του 'k' είναι κρίσιμη για την απόδοση του αλγορίθμου K-NN. Μια μικρή τιμή του 'k' μπορεί να κάνει τον αλγόριθμο ευαίσθητο στον θόρυβο, ενώ μια μεγάλη τιμή μπορεί να εξομαλύνει τα όρια απόφασης, μειώνοντας την ακρίβεια στα δεδομένα εκπαίδευσης. Είναι συνηθισμένη πρακτική να επιλέγεται το 'k' μέσω διασταυρούμενης επικύρωσης (cross-validation) για να διασφαλίζεται μια ισορροπία μεταξύ

προκατάληψης (bias) και διακύμανσης (variance). Επιπλέον, μπορεί να εφαρμοστούν βάρη στους γείτονες βάσει της απόστασής τους, δίνοντας στους πλησιέστερους γείτονες περισσότερη επιρροή στην πρόβλεψη.

Στην πράξη, το K-NN είναι απλό στην υλοποίηση και κατανόηση, καθιστώντας το δημοφιλής επιλογή για εφαρμογές όπου η ερμηνευτικότητα είναι ουσιαστική. Ωστόσο, μπορεί να είναι υπολογιστικά έντονο για μεγάλα σύνολα δεδομένων, καθώς απαιτεί τον υπολογισμό των αποστάσεων μεταξύ του νέου σημείου δεδομένων και κάθε σημείου στο σύνολο εκπαίδευσης. Παρά αυτό, η διαισθητική του φύση και η ικανότητα να κάνει προβλέψεις χωρίς a priori υποθέσεις σχετικά με την κατανομή των δεδομένων έχουν εδραιώσει τη θέση του K-NN ως βασικού αλγορίθμου στο εργαλειοθήκη της μηχανικής μάθησης.

3.2.6 Decision Trees

Decision tree trained on all the iris features



Σχήμα 3.3: Decision Tree του Iris dataset[55].

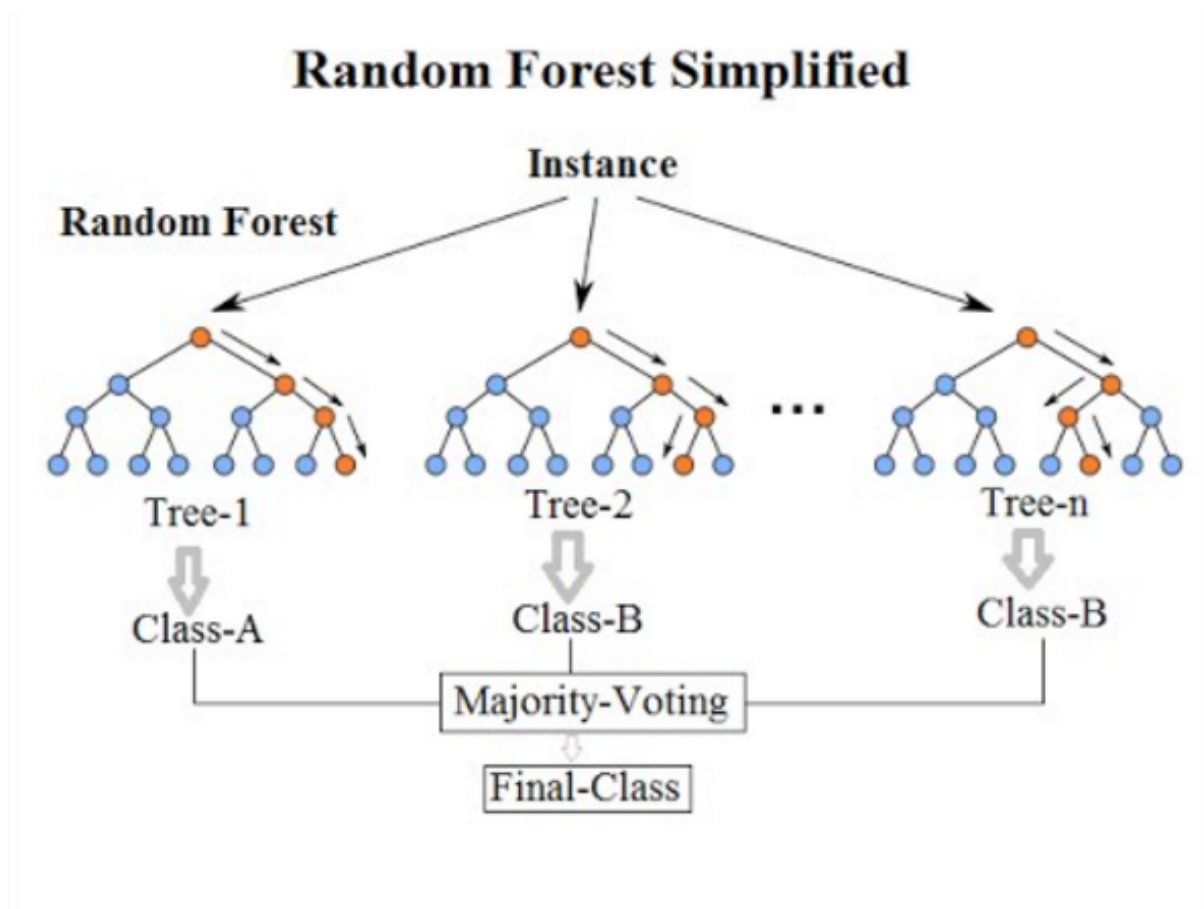
Τα Δέντρα Αποφάσεων είναι ένας δημοφιλής αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται τόσο για εργασίες ταξινόμησης όσο και παλινδρόμησης. Λειτουργούν διαιρώντας αναδρομικά τα δεδομένα σε υποσύνολα βάσει των τιμών των χαρακτηριστικών εισόδου, οδηγώντας σε ένα δενδροειδές μοντέλο αποφάσεων. Κάθε κόμβος στο δέντρο αντιπροσωπεύει ένα χαρακτηριστικό στο σύνολο δεδομένων, κάθε κλάδος αντιπροσωπεύει έναν κανόνα απόφασης, και κάθε φύλλο κόμβος αντιπροσωπεύει ένα αποτέλεσμα ή ετικέτα κλάσης. Η διαδρομή από τη ρίζα σε ένα

φύλλο παρέχει έναν κανόνα ταξινόμησης.

Ένα από τα κύρια πλεονεκτήματα των Δέντρων Αποφάσεων είναι η ερμηνευτικότητα τους. Η δομή του δέντρου είναι διαισθητική και μπορεί να οπτικοποιηθεί, καθιστώντας εύκολο τον προσδιορισμό και την εξήγηση της διαδικασίας λήψης αποφάσεων. Ωστόσο, τα Δέντρα Αποφάσεων μπορούν να είναι ευάλωτα στην υπερταίριαση (overfitting), ιδιαίτερα όταν το δέντρο είναι βαθύ, καταγράφοντας θόρυβο στα δεδομένα. Για να καταπολεμηθεί αυτό, χρησιμοποιούνται τεχνικές όπως το κλάδεμα (pruning) για την αφαίρεση κλάδων που έχουν λίγη ισχύ στην πρόβλεψη της στοχευόμενης μεταβλητής. Επιπλέον, οι μέθοδοι συνόλου όπως τα Random Forests και τα Gradient Boosted Trees βασίζονται στο βασικό αλγόριθμο δέντρου αποφάσεων για τη βελτίωση της ακρίβειας και της ανθεκτικότητας.

Στην πράξη, τα Δέντρα Αποφάσεων είναι ευέλικτα και μπορούν να χειριστούν τόσο αριθμητικά όσο και κατηγορικά δεδομένα. Είναι επίσης μη παραμετρικά, που σημαίνει ότι δεν κάνουν υποκείμενες υποθέσεις σχετικά με την κατανομή των δεδομένων. Αυτή η ευελιξία, σε συνδυασμό με την ευκολία ερμηνείας τους, έχει καταστήσει τα Δέντρα Αποφάσεων έναν βασικό αλγόριθμο στον τομέα της μηχανικής μάθησης.

Random Forests



Σχήμα 3.4: Παράδειγμα Random Forest[56].

Τα Random Forests είναι μια μέθοδος μάθησης συνόλου που βασίζεται στον αλγό-

ριθμο Decision Tree. Λειτουργούν κατασκευάζοντας πολλαπλά δέντρα αποφάσεων κατά τη διάρκεια της εκπαίδευσης και εξάγοντας τις πιο συνήθεις κλάσεις για εργασίες ταξινόμησης ή τη μέση πρόβλεψη των ατομικών δέντρων για εργασίες παλινδρόμησης. Ο κύριος στόχος πίσω από τα Random Forests είναι να συνδυάσουν τις προβλέψεις αρκετών βασικών εκτιμητών που έχουν κατασκευαστεί με έναν δεδομένο αλγόριθμο μάθησης για να βελτιώσουν τη γενίκευση και την ανθεκτικότητα έναντι ενός μόνο εκτιμητή.

Το "τυχαίο" στα Random Forests προέρχεται από δύο κύριες πηγές: το bootstrapping των δεδομένων εκπαίδευσης και την τυχαία επιλογή χαρακτηριστικών. Κατά την κατασκευή κάθε δέντρου, ένα υποσύνολο των δεδομένων εκπαίδευσης δειγματοληπτείται με αντικατάσταση (bootstrapping), και μόνο ένα τυχαίο υποσύνολο των χαρακτηριστικών λαμβάνεται υπόψη για διαίρεση σε κάθε κόμβο. Αυτή η διαδικασία εισάγει ποικιλία μεταξύ των δέντρων, που βοηθά στη μείωση της υπερταίριασης και προσθέτει ένα στοιχείο αποσυσχέτισης μεταξύ των ατομικών δέντρων.

Στην πράξη, τα Random Forests είναι γνωστά για την υψηλή τους ακρίβεια, την ικανότητά τους να χειρίζονται μεγάλα σύνολα δεδομένων με υψηλότερη διαστατικότητα, και την ικανότητά τους να χειρίζονται κενές τιμές. Παρέχουν επίσης βαθμολογίες σημαντικότητας χαρακτηριστικών, οι οποίες μπορούν να είναι διαφωτιστικές στην κατανόηση της υποκείμενης δομής των δεδομένων. Ενώ είναι υπολογιστικά πιο ακριβά από τα ατομικά δέντρα αποφάσεων, η απόδοσή τους συχνά δικαιολογεί το πρόσθετο υπολογιστικό κόστος. Η εγγενής τυχειότητα που εισάγεται κατά την κατασκευή του δέντρου, σε συνδυασμό με την προσέγγιση του συνόλου, καθιστά τα Random Forests ένα ισχυρό εργαλείο στο οπλοστάσιο της μηχανικής μάθησης.

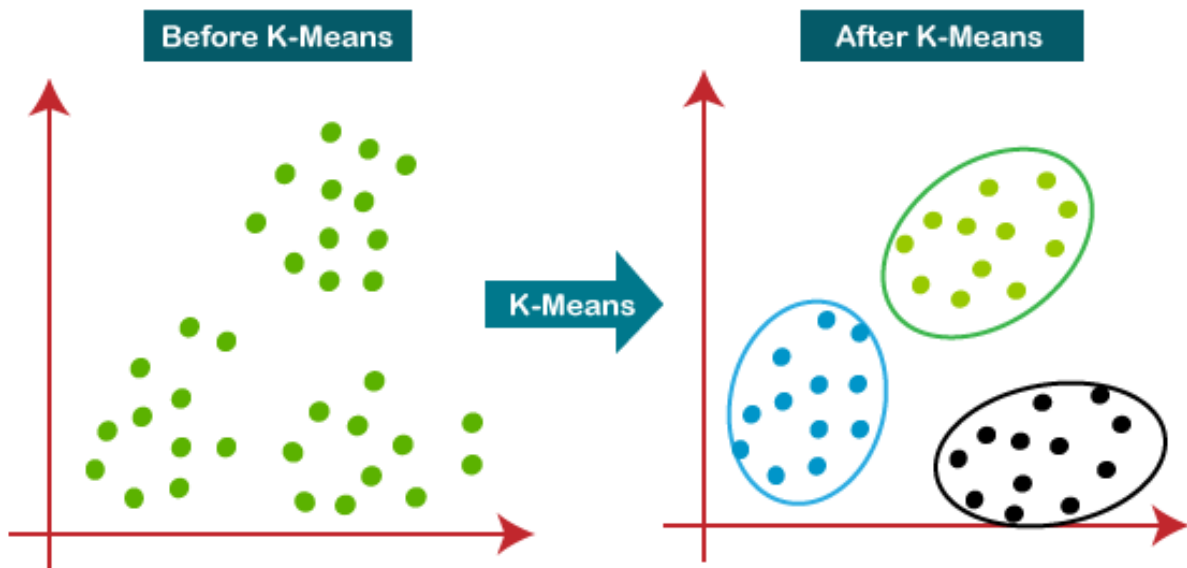
3.3 Μη Επιβλεπόμενη Μάθηση

Σε αυτήν την υπό ενότητα θα αναλύσουμε τον μοναδικό αλγόριθμο μη επιβλεπόμενης μάθησης που χρησιμοποιείται στην βιβλιογραφία. Οι πληροφορίες πάρθηκαν από πολλές πηγές όπως τα έργα των Kelleher [52], Patterson [53] και Alpaydin [50][51].

3.3.1 K-Means

Ο K-means είναι ένας ευρέως χρησιμοποιούμενος αλγόριθμος ομαδοποίησης που στοχεύει στην διαμέριση ενός συνόλου δεδομένων σε K διακριτές ομάδες, όπου κάθε σημείο δεδομένων ανήκει στην ομάδα με τον κοντινότερο μέσο όρο. Ο αλγόριθμος λειτουργεί επαναληπτικά, προσαρμόζοντας τα κεντροειδή των ομάδων μέχρι να επιτευχθεί μια βέλτιστη λύση ή να πληρούνται κάποια κριτήρια διακοπής. Ο κύριος στόχος του K-means είναι να ελαχιστοποιήσει το άθροισμα των τετραγώνων εντός της ομάδας, το οποίο ποσοτικοποιεί τη διακύμανση ή την εξάπλωση των σημείων δεδομένων εντός κάθε ομάδας.

Ο αλγόριθμος K-means ξεκινά με την τυχαία αρχικοποίηση των K κεντροειδών. Σε κάθε επανάληψη, τα σημεία δεδομένων ανατίθενται στο κοντινότερο κεντροειδές, σχηματίζοντας K ομάδες. Στη συνέχεια, τα κεντροειδή υπολογίζονται ξανά ως ο μέσος όρος όλων των σημείων δεδομένων εντός κάθε ομάδας. Αυτή η διαδικασία



Σχήμα 3.5: Παράδειγμα K-Means[57].

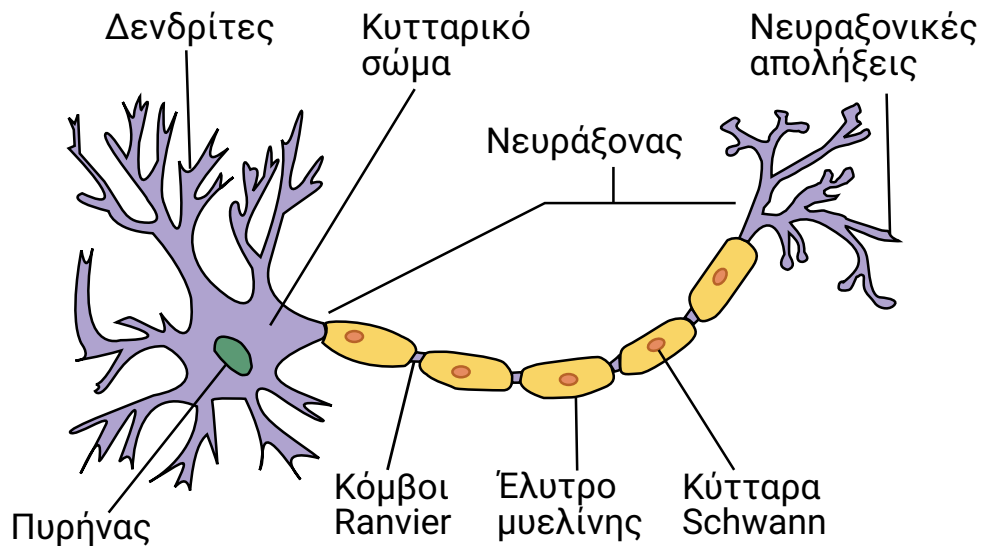
ανάθεσης και επανυπολογισμού συνεχίζεται μέχρι τα κεντροειδή να μην αλλάζουν σημαντικά ή να επιτευχθεί ένας προκαθορισμένος αριθμός επαναλήψεων. Μία από τις κύριες προκλήσεις με το K-means είναι η επιλογή του κατάλληλου αριθμού των ομάδων, K . Διάφορες μέθοδοι, όπως η μέθοδος Forgy, χρησιμοποιούνται για τον καθορισμό της βέλτιστης τιμής του K .

Στην πράξη, το K-means εκτιμάται για την απλότητα και την αποτελεσματικότητά του, ιδιαίτερα για μεγάλα σύνολα δεδομένων. Ωστόσο, είναι ευαίσθητο στην αρχική τοποθέτηση των κεντροειδών και μπορεί να συγκλίνει σε τοπικά ελάχιστα. Για να αντιμετωπιστεί αυτό, ο αλγόριθμος συχνά εκτελείται πολλές φορές με διαφορετικές αρχικοποιήσεις. Το K-means υποθέτει επίσης σφαιρικές ομάδες με παρόμοια μεγέθη και πυκνότητες, πράγμα που ενδέχεται να μην ισχύει πάντα για τα πραγματικά δεδομένα. Παρά τους περιορισμούς του, το K-means παραμένει ένας βασικός αλγόριθμος στη μη επιβλεπόμενη μάθηση και έχει βρει εφαρμογές σε διάφορους τομείς όπως για παράδειγμα τη συμπίεση εικόνας.

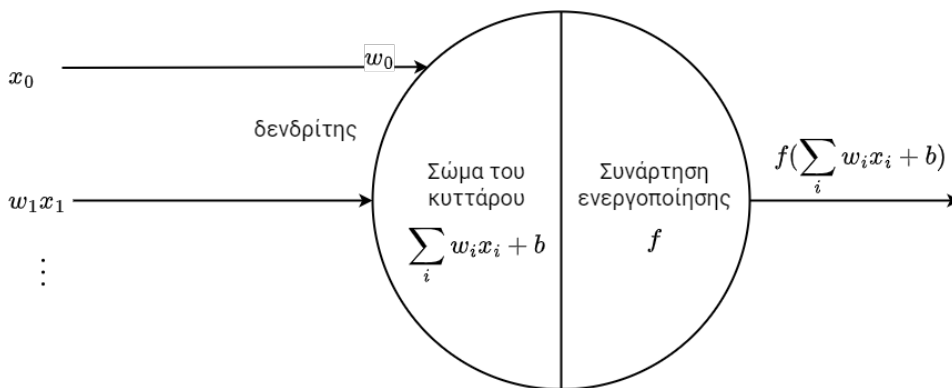
3.4 Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα είναι μοντέλα εμπνευσμένα από τον ανθρώπινο εγκέφαλο και την δικτύωση του. Ο εγκέφαλος είναι όργανο που επεξεργάζεται δεδομένα με μεγάλη ταχύτητα και απόδοση και προς το παρόν ξεπερνά πολλές τεχνολογίες [50]. Πιο συγκεκριμένα αποτελείται από περίπου 10^{11} νευρώνες, μονάδες επεξεργασίας, που δουλεύουν ταυτόχρονα και συνδέονται μεταξύ τους με συνάψεις. Υπάρχουν γύρω στις $10^{14} - 10^{15}$ συνάψεις και ο κάθε νευρώνας μπορεί να έχει από 1000 μέχρι και 200000 συνάψεις [58]. Μπορούμε να συγκρίνουμε τα δύο είδη νευρώνων, εγκεφάλου και τεχνητού, στα Σχήματα 3.6 και 3.7.

Αυτό το τεχνητό νευρωνικό δίκτυο (Perceptron) αποτελεί μια συνάρτηση με είσοδο ένα διάνυσμα x και έξοδο την $f(x)$. Στο ενδιάμεσο κάθε είσοδος x_i πολλαπλα-



Σχήμα 3.6: Νευρώνας εγκεφάλου[59].



Σχήμα 3.7: Τεχνητός νευρώνας.

σιάζεται με ένα βάρος w_i . Τελικά αθροίζονται τα γινόμενα και λειτουργούν ως είσοδος για την συνάρτηση ενεργοποίησης. Στους βιολογικούς νευρώνες, όταν το τελικό άθροισμα είναι μεγαλύτερο από ένα όριο, τότε ο νευρώνας πυροδοτείται δημιουργώντας ένα δυναμικό. Έτσι η συνάρτηση ενεργοποίησης λαμβάνει το άθροισμα και αποφασίζει αν θα "πυροδοτήσει" ή όχι, δηλαδή τι αποτέλεσμα θα παράγει ο νευρώνας [60]. Υπάρχουν πολλές συναρτήσεις ενεργοποίησης για τα νευρωνικά δίκτυα και παρακάτω θα μελετήσουμε κάποιες.

3.4.1 Στρώμα Dropout

Το Dropout είναι μια τεχνική κανονικοποίησης που χρησιμοποιείται συχνά στα νευρωνικά δίκτυα για την πρόληψη της υπερπροσαρμογής. Εισήχθη από τους Srivastava et al. το 2014 [61] και λειτουργεί θέτοντας τυχαία ένα μέρος των εισαγόμενων μονάδων σε μηδέν κατά τη διάρκεια κάθε επανάληψης εκπαίδευσης. Αυτή η διαδικασία μπορεί να θεωρηθεί ως προσωρινή "απόρριψη" ορισμένων νευρώνων.

Κατά τη διάρκεια της εκπαίδευσης, κάθε νευρώνας έχει μια πιθανότητα p να αφαιρεθεί προσωρινά ή να "απορριφθεί". Αν $p = 0$, το δίκτυο γίνεται ένα τυπικό νευρωνικό δίκτυο χωρίς dropout. Αν $p = 1$, όλες οι μονάδες εισόδου απορρίπτονται.

Ωστόσο, οι τυπικές τιμές για το p κυμαίνονται μεταξύ 0.2 έως 0.5.

Ως οφέλη λαμβάνουμε ότι το dropout λειτουργεί ως μορφή κανονικοποίησης, μειώνοντας τον κίνδυνο υπερπροσαρμογής και επιπλέον να θεωρηθεί ως η εκπαίδευση ενός ψευδο-συνόλου νευρωνικών δικτύων. Με το dropout, το δίκτυο γίνεται λιγότερο ευαίσθητο στα ειδικά βάρη των μεμονωμένων νευρώνων, οδηγώντας σε ένα πιο γενικευμένο μοντέλο. Ταυτόχρονα Το dropout μπορεί να θεωρηθεί ως εκπαίδευση ενός ψευδο-συνόλου νευρωνικών δικτύων. Κάθε επανάληψη εκπαίδευσης ενημερώνει μια μοναδική μικρή έκδοση του δικτύου. Κατά την εξαγωγή συμπερασμάτων, χρησιμοποιούνται όλοι οι νευρώνες, κάτι που μπορεί να ερμηνευθεί ως μέσο όρο των προβλέψεων όλων αυτών των αραιωμένων δικτύων.

3.4.2 Στρώμα Πλήρους Σύνδεσης

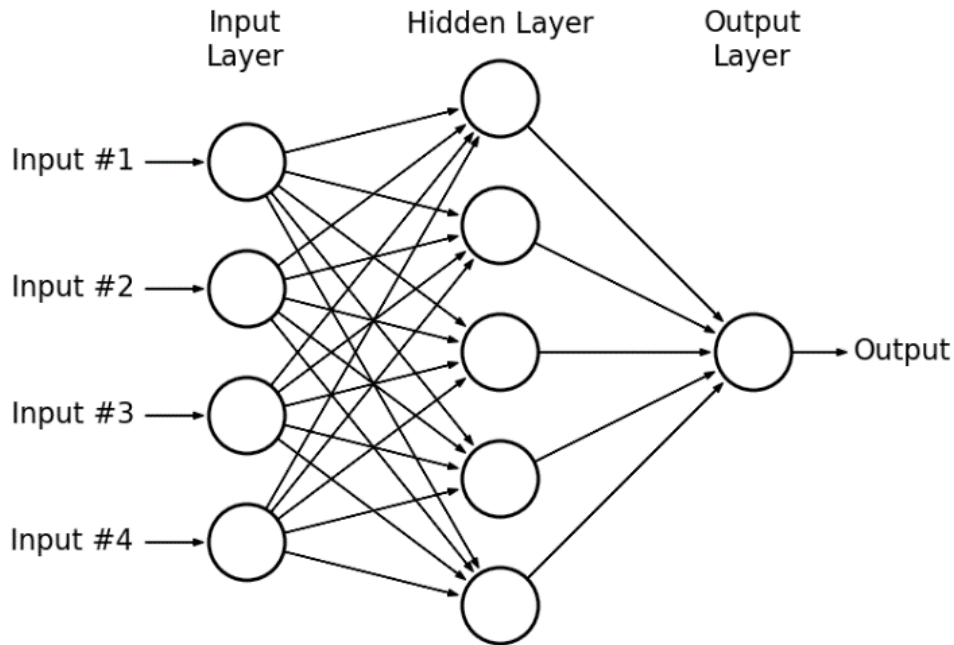
Ένα γραμμικό στρώμα, που αναφέρεται επίσης ως στρώμα πλήρως συνδεδεμένο (FC) ή πυκνό στρώμα, είναι ένα θεμελιώδες στοιχείο στα νευρωνικά δίκτυα. Σε αυτό το επίπεδο, κάθε νευρώνας συνδέεται με κάθε νευρώνα του προηγούμενου στρώματος, με κάθε σύνδεση να έχει το σχετικό βάρος της. Δεδομένης μιας εισόδου X και του στρώματος βάρους W και διανύσματος πόλωσης b , η έξοδος Y του στρώματος μπορεί να υπολογιστεί ως: $Y = XW + b$ Τα οφέλη της είναι πως διασφαλίζει ότι το στρώμα μπορεί να μάθει μια ευρεία γκάμα μετασχηματισμών από την είσοδό του και ότι τα γραμμικά στρώματα μπορούν να στοιβαχθούν εύκολα, επιτρέποντας τη δημιουργία βαθιών νευρωνικών δικτύων.

3.4.3 Δίκτυα Εμπρόσθιας Τροφοδότησης Πολλών Επιπέδων

Για να μπορέσει να υλοποιήσει ένα νευρωνικό δίκτυο πιο πολύπλοκες εργασίες συνδυάζουμε νευρώνες και τους τοποθετούμε σε στρώσεις έτσι ώστε το αποτέλεσμα του ενός να είναι είσοδος του άλλου. Έτσι δημιουργούμε τα δίκτυα πολλών επιπέδων (Multi Layer Perceptron, MLP). Εάν αυτά δεν δημιουργούν κύκλους ονομάζονται Εμπρός Τροφοδοτούμενα Νευρωνικά Δίκτυα (Feedforward Artificial Neural Networks, FFNN). Ένα MLP αποτελείται από τουλάχιστον τρία στρώματα (Διάγραμμα 3.8), από την είσοδο, το στρώμα που δέχεται τα δεδομένα, και όσα είναι τα δεδομένα τόσο μεγάλο είναι το στρώμα αυτό (αριθμός νευρώνων), το τελευταίο στρώμα που είναι η έξοδος του νευρωνικού δικτύου όπου δημιουργείται το αποτέλεσμα και ανάμεσα υπάρχουν όσα στρώματα χρειάζονται για να επεξεργαστούν κατάλληλα την είσοδο[53].

3.4.4 Δίκτυα Μακράς Βραχύχρονης Μνήμης

Τα αναδρομικά δίκτυα (Recurrent Neural Networks, RNN) προσεγγίζουν τον ανθρώπινο εγκέφαλο ακόμα περισσότερο. Ειδικότερα εκτός από τις συνδέσεις προς τα εμπρός στρώματα, οι νευρώνες συνδέονται και με τον εαυτό τους και ακόμα και με τα προηγούμενα στρώματα. Αυτές οι συνδέσεις τους προσφέρουν ένα είδος μνήμης [50] [63]. Τέτοιου τύπου νευρωνικά δίκτυα, λόγω της πολυπλοκότητάς τους, εκπαιδεύονται αρκετά δύσκολα και είναι χρονοβόρα. Όμως με βελτιστοποιήσεις και με την χρήση καρτών γραφικών γίνονται πιο προσιτά [53].



Σχήμα 3.8: Multi-Layer Perceptron (MLP) [62].

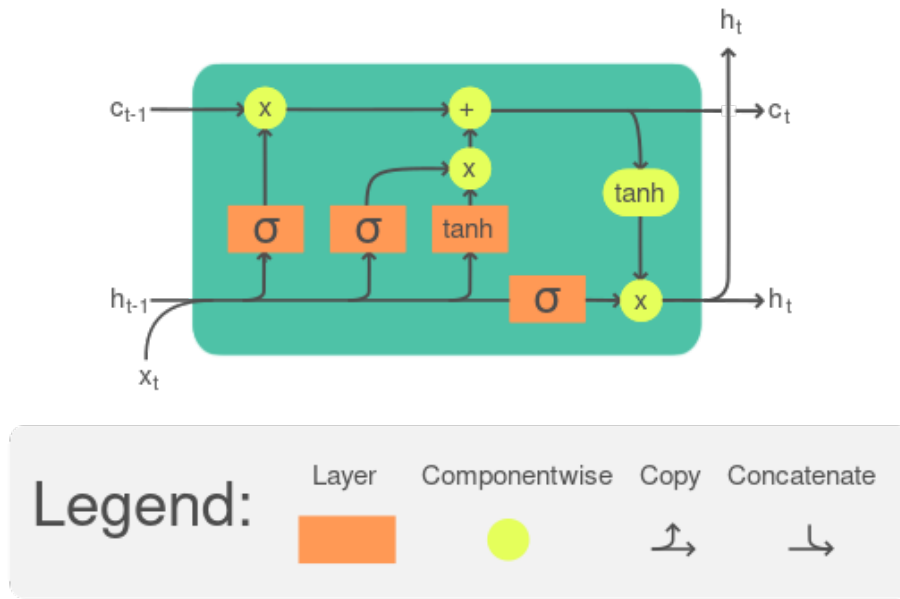
Τα Δίκτυα Μακράς Βραχύχρονης Μνήμης (Long Short-Term Memory, LSTM) εισήχθησαν το 1997 από τους Hochreiter και Schmidhuber[64] και είναι τα πιο γνωστά αναδρομικά δίκτυα. Τα LSTM έχουν τη δυνατότητα να μαθαίνουν και να θυμούνται παρτίδες πληροφοριών για μεγάλα χρονικά διαστήματα, κάτι που τα κάνει ιδιαίτερα κατάλληλα για εφαρμογές όπως η αναγνώριση φωνής, η μετάφραση γλώσσας και η πρόβλεψη σειρών χρόνου.

Το LSTM είναι ένα εξειδικευμένο RNN που λύνει το πρόβλημα της εξαφάνισης και της έκρηξης του κλίμακας στα παραδοσιακά RNN, κάτι που διευκολύνει την εκπαίδευση σε μεγάλες σειρές. Διαθέτει μια δομή (Διάγραμμα 3.9) με τρεις ειδικές πύλες (πύλη εισόδου, πύλη ξεχασμού και πύλη εξόδου) που ελέγχουν τη ροή των πληροφοριών στη μνήμη του κελιού. Αυτό επιτρέπει στο LSTM να φιλτράρει, αποθηκεύει και προσπελάσει τις πληροφορίες κατά τρόπο ελεγχόμενο και αποτελεσματικό.

3.5 Μετασχηματιστές

Οι μετασχηματιστές έχουν γίνει ακρογωνιαίος λίθος στον κόσμο της βαθιάς μάθησης από την εισαγωγή τους στην εργασία του 2017 από τους Vaswani et al. [66]. Χαρακτηριστικό γνώρισμα αυτής της αρχιτεκτονικής είναι ο μηχανισμός αυτοπροσοχής (Διάγραμμα 3.10), μια νέα προσέγγιση που επιτρέπει στα μοντέλα να ζυγίζουν διαφορετικά τα χαρακτηριστικά εισόδου, αποτυπώνοντας έμπειρα εξαρτήσεις μεγάλης εμβέλειας χωρίς να βασίζονται στην επανάληψη.

Στον πυρήνα τους, οι μετασχηματιστές ξεκινούν μετατρέποντας τα μονάδες (tokens) εισόδου σε διανύσματα χρησιμοποιώντας αναπαραστάσεις κειμένου (embeddings). Για να ληφθεί υπόψη η σειρά των tokens - ένα χαρακτηριστικό που οι μετα-



Σχήμα 3.9: Κύτταρο Δικτύου Μακράς Βραχύχρονης Μνήμης (LSTM)[65].

σχηματιστές δεν αναγνωρίζουν εγγενώς - αυτές οι ενσωματώσεις εμπλουτίζονται με κωδικοποιήσεις θέσης. Ο μηχανισμός αυτοπροσοχής αναλαμβάνει τότε δουλειά. Για κάθε token, υπολογίζει ένα σταθμισμένο άθροισμα όλων των άλλων tokens στην είσοδο. Τα βάρη, που καθορίζονται από τις βαθμολογίες προσοχής, δίνουν τη δυνατότητα σε κάθε token να "συμμετέχει" σε κάθε άλλο token, ένα χαρακτηριστικό που έχει καθοριστική σημασία για την καταγραφή εξαρτήσεων μεγάλης εμβέλειας. Τα παραπάνω γίνονται με τις πράξεις κλιμακούμενο εσωτερικό διάνυσμα και multi-head attention (δες Διαγράμματα 3.11 3.12) με τις εξής εξισώσεις:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

όπου υπολογίζουμε τη συνάρτηση προσοχής σε ένα σύνολο ερωτημάτων ταυτόχρονα, σε έναν πίνακα Q . Τα κλειδιά και οι τιμές συνενώνονται επίσης μαζί στους πίνακες K και V .

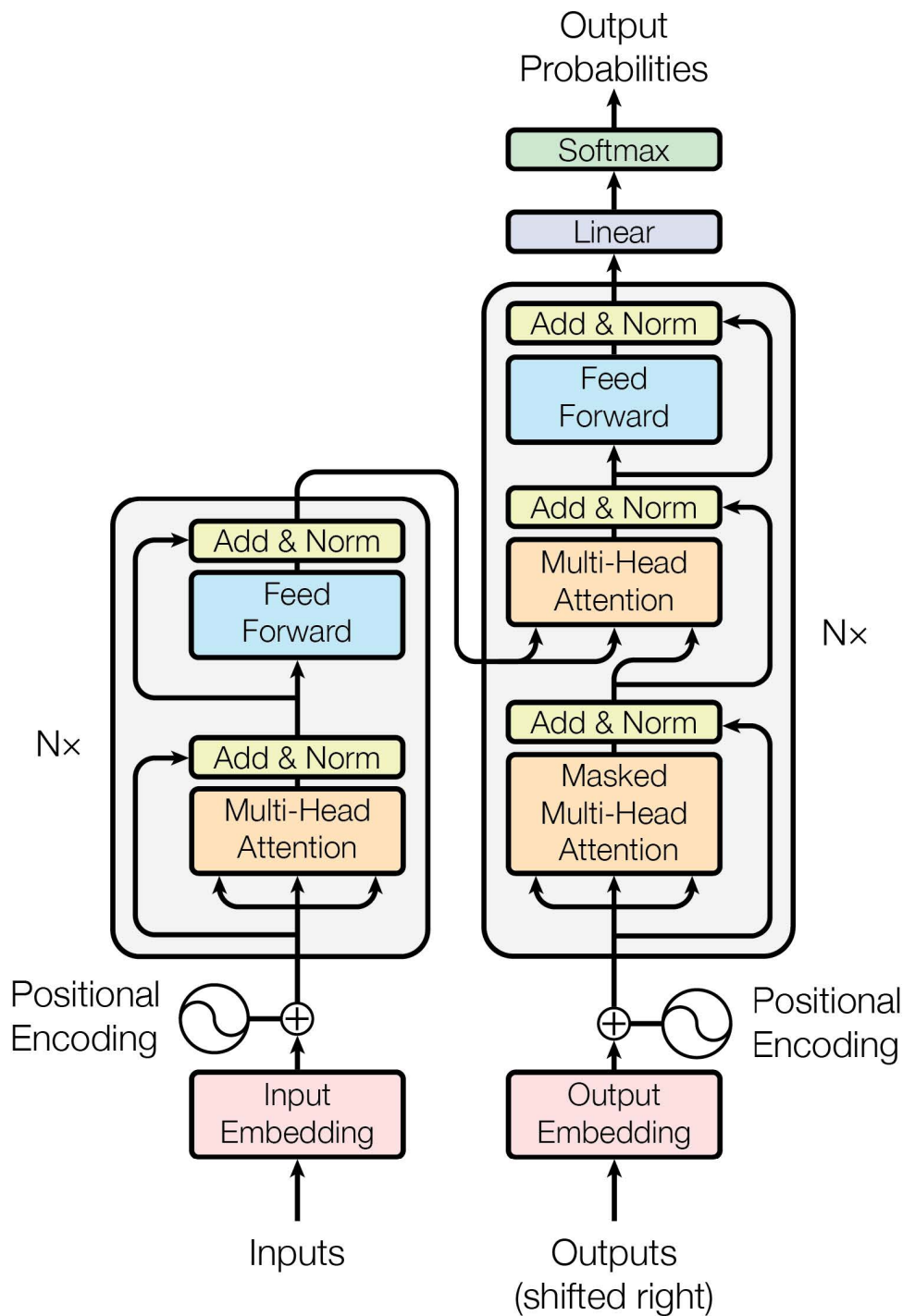
$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O$$

where

$$head_i = Attention(QW_i^Q, K(W_i)^K, VW_i^V)$$

Όπου οι προβολές είναι πίνακες παραμέτρων $W_i^Q \in R^{d_{model} \times d_k}$, $W_i \in R^{d_{model} \times d_k}$, $W_i^V \in R^{d_{model} \times d_v}$ και $W^O \in R^{d_u \times d_{model}}$.

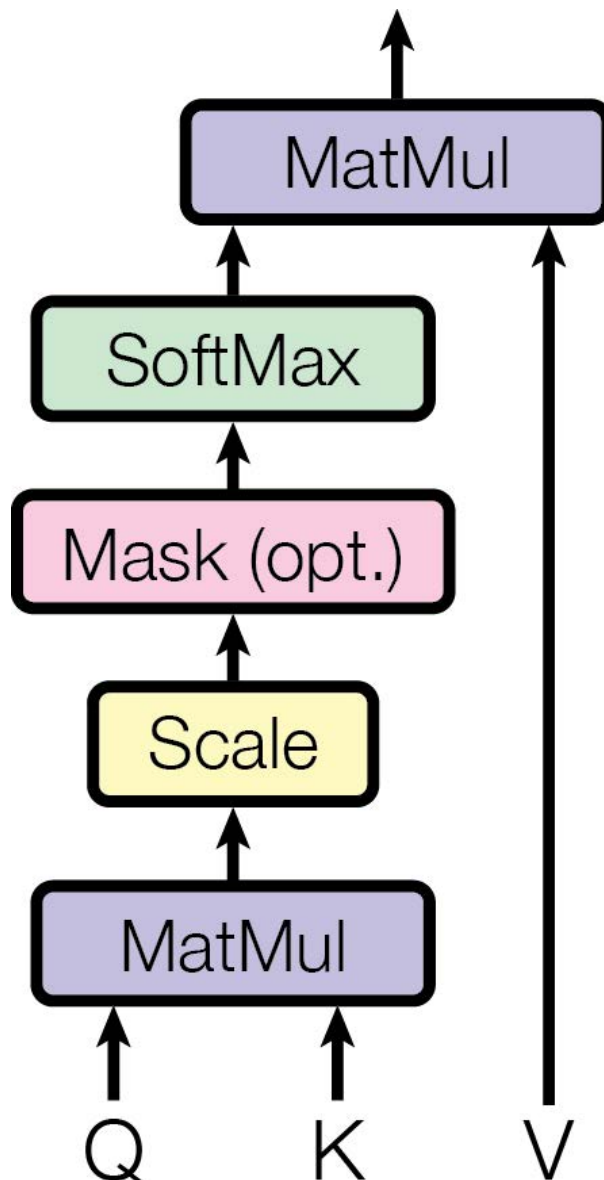
Ο υπολογισμός αυτών των βαθμολογιών προσοχής περιλαμβάνει τρία διανύσματα για κάθε διακριτικό: Ερώτημα (Query), Κλειδί (Key) και Τιμή (Value). Οι βαθμολογίες προέρχονται από το εσωτερικό γινόμενο του διανύσματος ερωτήματος ενός token με το διάνυσμα κλειδιού όλων των άλλων token. Στη συνέχεια, αυτές οι βαθμολογίες ζυγίζουν τα διανύσματα τιμής σε μια διαδικασία που ονομάζεται προσοχή σε



Σχήμα 3.10: Δομή ενός μετασχηματιστή [66].

κλιμακωτό εσωτερικό διάνυσμα. Εδώ, οι βαθμολογίες προσοχής κλιμακώνονται (συνήθως με βάση την τετραγωνική ρίζα του βάθους του μοντέλου) και περνούν από μια συνάρτηση softmax για να οριστικοποιηθούν τα βάρη.

Μετά τον μηχανισμό αυτοπροσοχής, οι μετασχηματιστές χρησιμοποιούν νευρωνικά δίκτυα εμπρόσθιας τροφοδοσίας για κάθε θέση token. Για την καταπολέμηση του προβλήματος των εξαφανιζόμενων κλίσεων (vanishing gradient), μια πρόκληση στα βαθιά δίκτυα, οι μετασχηματιστές ενσωματώνουν υπολειμματικές συνδέσεις

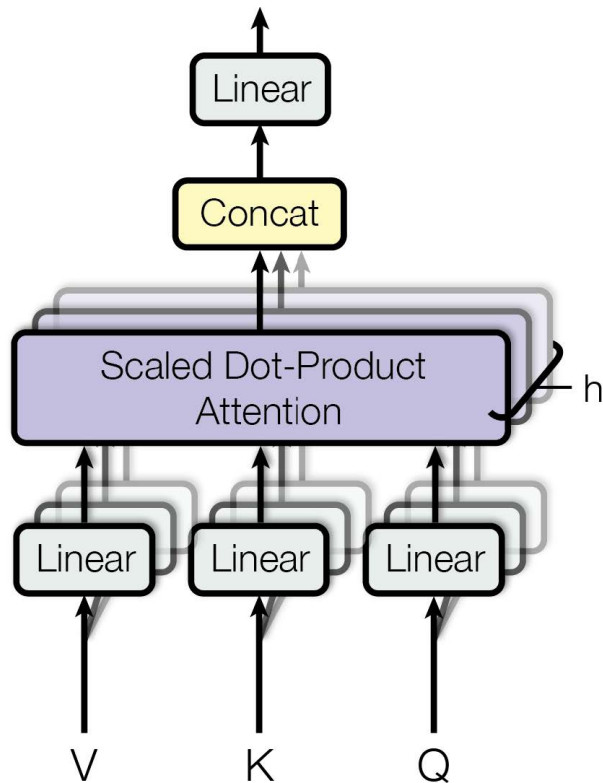


Σχήμα 3.11: Αυτό-προσοχή με κλιμακούμενο εσωτερικό διάνυσμα [66].

γύρω από κάθε υπο-στρώμα. Αυτό περιλαμβάνει τα επίπεδα νευρωνικών δικτύων αυτοπροσοχής και τροφοδοσίας. Επιπλέον, η κανονικοποίηση στρώματος εφαρμόζεται με συνέπεια πριν από κάθε υπο-στρώμα, με τα υπολείμματα να προστίθενται μετά από αυτήν την κανονικοποίηση.

Η αρχιτεκτονική του μετασχηματιστή περιλαμβάνει αυτά τα στρώματα στοιβαγμένα πολλές φορές, συχνά έξι ή περισσότερες, τόσο στο τμήμα κωδικοποιητή όσο και στο τμήμα του αποκωδικοποιητή. Η επιτυχία του αρχικού μοντέλου μετασχηματιστή ώθησε την ανάπτυξη πολλών παραλλαγών όπως το BERT [67], το GPT [68] και το T5 [69]. Κάθε προσαρμογή είτε επεκτείνει είτε βελτιώνει τον αρχικό σχεδιασμό για να επιτύχει συγκεκριμένες εργασίες ή να βελτιώσει την απόδοση.

Σχεδιασμένο αρχικά για μετάφραση κειμένων, η ικανότητα των μετασχηματιστών να διακρίνουν γρήγορα τις έννοιες των συμφραζομένων των λέξεων τα έκανε βασικά μοντέλο για ένα ευρύ φάσμα εργασιών NLP. Αυτό περιλαμβάνει ταξινόμηση



Σχήμα 3.12: Το Multi-Head Attention αποτελείται από πολλά επίπεδα προσοχής που εκτελούνται παράλληλα [66].

κειμένου, ανάλυση συναισθημάτων και απάντηση ερωτήσεων. Η επιρροή τους δεν έχει περιοριστεί μόνο στο NLP. Παραλλαγές μετασχηματιστών εισχωρούν τώρα σε άλλους τομείς, κυρίως στην όραση υπολογιστών.

Εν ολίγοις, οι μετασχηματιστές εγκαινίασαν μια αλλαγή παραδείγματος στο σχεδιασμό μοντέλων βαθιάς μάθησης για το NLP. Δίνοντας έμφαση στην παράλληλη επεξεργασία και αποτυπώνοντας έμπειρα τις εξαρτήσεις μεγάλης εμβέλειας, έχουν θέσει νέα σημεία αναφοράς στην απόδοση. Η προσαρμοστικότητα και η επιτυχία τους έχουν τροφοδοτήσει εκτεταμένη έρευνα και ανάπτυξη, εδραιώνοντας τη θέση τους ως θεμελιώδης αρχιτεκτονική στη σύγχρονη τεχνητή νοημοσύνη.

3.5.1 BERT (Bidirectional Encoder Representations from Transformers)

Το BERT (Bidirectional Encoder Representations from Transformers) [67] είναι ένα νέο μοντέλο αναπαράστασης γλώσσας που εισήχθη από τον Jacob Devlin και την ομάδα του στο Google AI Language. Ο πρωταρχικός στόχος του BERT είναι να εκπαιδεύσει εκ των προτέρων βαθιές αμφίδρομες αναπαραστάσεις από κείμενα χωρίς ετικέτα, χωρίς να λαμβάνει την σειρά των λέξεων εκατέρωθεν. Αυτή η προσέγγιση έρχεται σε αντίθεση με προηγούμενες μεθόδους που είτε εκπαίδευαν μονοκατευθυντικά μοντέλα είτε συνέδεαν τα αποτελέσματα ανεξάρτητων εκπαιδευμένων μο-

ντέλων από αριστερά προς τα δεξιά και από τα δεξιά προς τα αριστερά.

Ο σχεδιασμός του BERT είναι εννοιολογικά απλός αλλά εμπειρικά ισχυρός. Έχει επιτύχει αποτελέσματα αιχμής σε μια ποικιλία εργασιών επεξεργασίας φυσικής γλώσσας. Για παράδειγμα, ο BERT έχει αυξήσει τη βαθμολογία GLUE στο 80,5%, βελτίωσε την ακρίβεια MultiNLI στο 86,7% και βελτίωσε τις βαθμολογίες SQuAD v1.1 και v2.0 Test F1 σε 93,2% και 83,1%, αντίστοιχα.

Η προεκπαιδευτική στρατηγική του BERT βασίζεται σε δύο τεχνικές: το «μοντέλο γλώσσας με μάσκα» (masked language model, MLM) και την εργασία «πρόβλεψη επόμενης πρότασης» (next sentence prediction). Στην προσέγγιση MLM, ορισμένα από τα διακριτικά από την είσοδο καλύπτονται τυχαία και το μοντέλο εκπαιδεύεται να προβλέπει το αρχικό αναγνωριστικό λεξιλογίου της καλυμμένης λέξης με βάση αποκλειστικά το περιεχόμενό της. Αυτή η μέθοδος επιτρέπει στον BERT να εκπαιδεύει εκ των προτέρων έναν μετασχηματιστή βαθιάς διπλής κατεύθυνσης, σε αντίθεση με τις παραδοσιακές προσεγγίσεις από αριστερά προς τα δεξιά ή από τα δεξιά προς τα αριστερά. Η εργασία "πρόβλεψη επόμενης πρότασης" χρησιμοποιείται για την εκ των προτέρων εκπαίδευση αναπαραστάσεων για ζεύγη κειμένων, η οποία είναι ιδιαίτερα χρήσιμη για εργασίες που απαιτούν κατανόηση της σχέσης μεταξύ δύο προτάσεων ή παραγράφων.

Ένα σημαντικό πλεονέκτημα του BERT είναι η ικανότητά του να μειώνει την ανάγκη για βαριά σχεδιασμένες αρχιτεκτονικές για συγκεκριμένες εργασίες. Μόλις προεκπαιδευτεί, το μοντέλο BERT μπορεί να ρυθμιστεί με ακρίβεια με ένα μόνο επιπλέον επίπεδο εξόδου, επιτρέποντάς του να δημιουργεί μοντέλα τελευταίας τεχνολογίας για ένα ευρύ φάσμα εργασιών χωρίς ουσιαστικές τροποποιήσεις στην υποκείμενη αρχιτεκτονική. Αυτή η προσέγγιση έχει αποδειχθεί εξαιρετικά αποτελεσματική, ξεπερνώντας σε απόδοση πολλά μοντέλα που αφορούν συγκεκριμένες εργασίες.

Η υλοποίηση του BERT χρησιμοποιεί πολλαπλά επίπεδα των Transformers. Το μοντέλο διατίθεται σε δύο μεγέθη: BERTBASE, που έχει 12 επίπεδα, 768 κρυφές μονάδες και 12 κεφαλές αυτοπροσοχής (συνολικά 110 εκατομμύρια παραμέτρους) και BERTLARGE, που έχει 24 επίπεδα, 1024 κρυφές μονάδες και 16 κεφαλές αυτοπροσοχής (συνολικά 340M παράμετροι). Συγκεκριμένα, το BERTBASE έχει το ίδιο μέγεθος μοντέλου με το GPT του OpenAI για λόγους σύγκρισης. Ωστόσο, ενώ το Transformer του GPT χρησιμοποιεί μονοκατευθυντική αυτοπροσοχή, το Transformer του BERT χρησιμοποιεί αμφίδρομη αυτοπροσοχή. Ένα μοναδικό χαρακτηριστικό στην αναπαράσταση εισόδου του BERT είναι η χρήση ειδικών διακριτικών. Ένα από αυτά είναι το διακριτικό [CLS], το οποίο σημαίνει "ταξινόμηση". Όταν το BERT προσαρμόζεται για συγκεκριμένα καθήκοντα ταξινόμησης, όπως η ανάλυση συναισθήματος ή κατηγοριοποίηση θεμάτων, η τελική κρυφή κατάσταση που αντιστοιχεί στο CLS εξάγεται και χρησιμοποιείται ως είσοδος για το κεφάλι ταξινόμησης (Classification head). Το κεφάλι ταξινόμησης είναι συνήθως ένα νευρωνικό δίκτυο προώθησης, που παίρνει αυτές τις πληροφορίες και εξάγει τις τελικές ετικέτες κλάσης. Είναι προσαρμοσμένο στη συγκεκριμένη εργασία ταξινόμησης, μεταφράζοντας τις πλούσιες πληροφορίες που αποτυπώνονται από το CLS σε μια μορφή κατάλληλη για ταξινόμηση. Μαζί, το διάκριτο CLS και το κεφάλι ταξινόμησης επιτρέπουν στο BERT να προσαρμοστεί και να προσαρμοστεί για μια μεγάλη ποικιλία καθηκόντων ταξινόμησης, εκμεταλλευόμενοι την προπονημένη γνώση του μοντέλου.

Τώρα, σχετικά με τα δεδομένα εκπαίδευσης για το BERT, ιδιαίτερα το BERT base model (uncased) που είναι διαθέσιμο από το Hugging Face. Το BERT base model (uncased) εκπαιδεύεται σε έναν συνδυασμό δύο εκτεταμένων συνόλων δεδομένων: Wikipedia [70] και BookCorpus [71]. Το σύνολο δεδομένων Wikipedia αποτελείται από το σύνολο των άρθρων της αγγλικής Wikipedia, συνολικά περίπου 2500 εκατομμύρια λέξεις. Το σύνολο δεδομένων BookCorpus περιλαμβάνει κείμενα από 11038 βιβλία, περίπου 800 εκατομμύρια λέξεις. Μαζί, αυτά τα σύνολα δεδομένων παρέχουν ένα πλούσιο και ποικίλο σύνολο δεδομένων εκπαίδευσης περίπου 3300 εκατομμύρια λέξεις, επιτρέποντας στο μοντέλο να μάθει μια ευρεία γκάμα μοτίβων και δομών γλώσσας. Αυτή η εκτεταμένη προπόνηση σε τέτοιο μεγάλο σύνολο δεδομένων είναι ένας καίριος παράγοντας για την επιτυχία του BERT σε διάφορα καθήκοντα επεξεργασίας φυσικής γλώσσας.

Συμπερασματικά, το BERT αντιπροσωπεύει μια σημαντική πρόοδο στον τομέα της επεξεργασίας φυσικής γλώσσας. Η αμφίδρομη προσέγγιση εκπαίδευσης, σε συνδυασμό με τη δύναμη της αρχιτεκτονικής του Transformer, του επιτρέπει να επιτυγχάνει κορυφαίες επιδόσεις σε ένα ευρύ φάσμα εργασιών.

3.5.2 DistilBERT

Το DistilBERT [35] είναι μια παραλλαγή του BERT (Bidirectional Encoder Representations from Transformers) που δημιουργήθηκε από την Hugging Face με στόχο να προσφέρει μια πιο ελαφριά και αποδοτική επιλογή στον χώρο της επεξεργασίας φυσικής γλώσσας. Μια από τις κύριες αλλαγές που εισήχθησαν είναι η μείωση του αριθμού των επιπέδων του μετασχηματιστή και των κρυφών μονάδων, χωρίς να θυσιάζεται όμως σημαντικά η απόδοση.

Ένα σημαντικό πλεονέκτημα του DistilBERT είναι η δυνατότητά του να διατηρεί υψηλή απόδοση ενώ ταυτόχρονα απαιτεί λιγότερους υπολογιστικούς πόρους. Αυτό το καθιστά ιδανική επιλογή για περιβάλλοντα με περιορισμένους υπολογιστικούς πόρους ή για εφαρμογές που απαιτούν χρήση μικρότερου μοντέλου. Συγκεκριμένα χρησιμοποιεί 40% λιγότερες παραμέτρους και τρέχει 60% πιο γρήγορα από το BERT, διατηρώντας όμως κατά 95% την ακρίβεια του [72].

Το DistilBERT ακολουθεί την προσέγγιση της μεταφοράς γνώσης από το BERT, κάνοντας το αρχικό BERT ως δάσκαλο και μεταφέροντας τη γνώση και τις αναπαραστάσεις σε ένα απλούστερο μοντέλο, το οποίο είναι το DistilBERT ως μαθητής. Αυτή η συμπίεση και μεταφορά γνώσης διατηρεί την ικανότητα του DistilBERT να κατανοεί τη γλώσσα, ενώ το καθιστά ελαφρύτερο.

Συνοψίζοντας, το DistilBERT είναι μια εξαιρετική επιλογή για εφαρμογές που απαιτούν υψηλή απόδοση σε συνδυασμό με χαμηλότερη απαίτηση σε υπολογιστικούς πόρους, και αποτελεί μια επιτυχημένη προσπάθεια για τη βελτιστοποίηση του BERT μοντέλου.

3.5.3 EmoRoBERTa

Το EmoRoBERTa είναι μια προσαρμογή του μοντέλου RoBERTa, ειδικά προσαρμοσμένη για το έργο της ταξινόμησης συναισθημάτων. Το μοντέλο RoBERTa, το οποίο

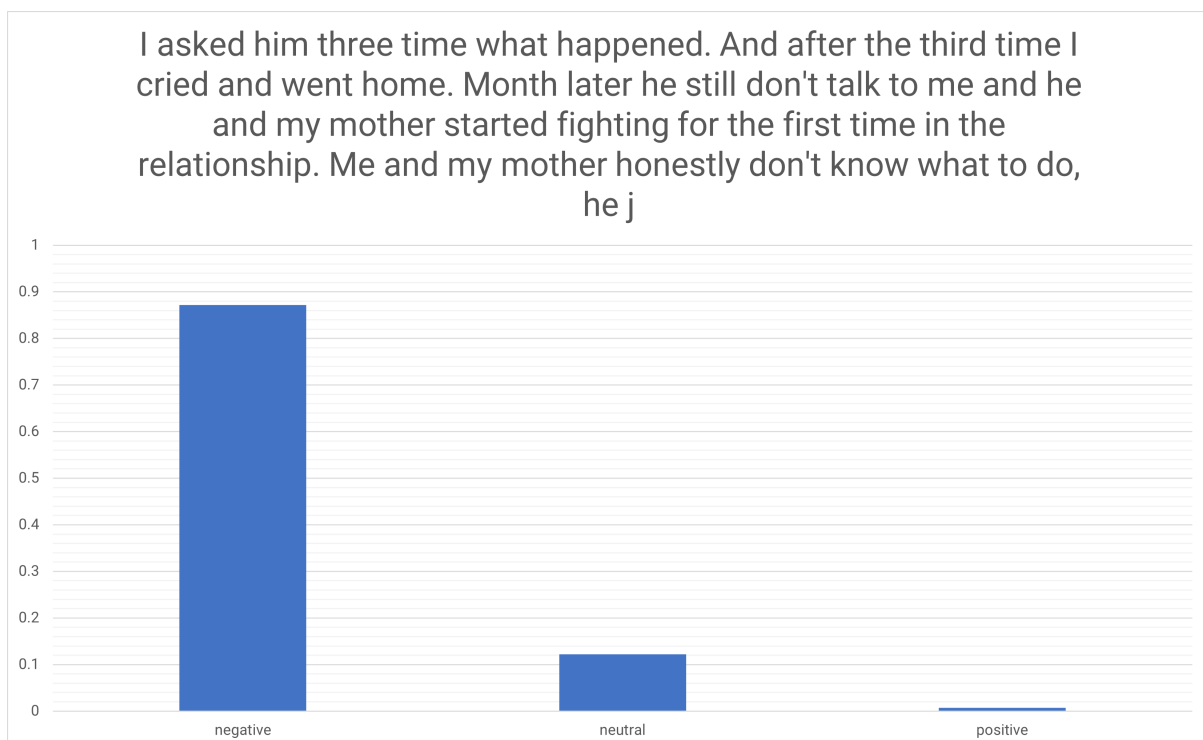


Σχήμα 3.13: Αποτελέσματα του EmoRoBERTa από τυχαίο κείμενο του συνόλου δεδομένων.

σημαίνει "A Robustly Optimized BERT Pretraining Approach", είναι μια παραλλαγή του BERT (Αμφίδρομες Αναπαραστάσεις Κωδικοποιητή από Μετασχηματιστές). Το BERT και οι παραλλαγές του, συμπεριλαμβανομένου του RoBERTa, έφεραν επανάσταση στον τομέα της επεξεργασίας φυσικής γλώσσας (NLP) παρέχοντας αποτελέσματα τελευταίας τεχνολογίας σε ένα ευρύ φάσμα εργασιών. Το μοντέλο RoBERTa διαφέρει από το BERT σε πολλές βασικές πτυχές. Πρώτον, χρησιμοποιεί περισσότερα δεδομένα για την προεκπαίδευσή του, αξιοποιώντας μεγάλες ποσότητες κειμένων από βιβλία και ιστότοπους. Δεύτερον, καταργεί τον στόχο πρόβλεψης επόμενης πρότασης, τον οποίο χρησιμοποιεί ο BERT κατά την προεκπαίδευση, και αντ' αυτού εστιάζει αποκλειστικά στον στόχο του μοντέλου της γλώσσας με μάσκα. Αυτή η αλλαγή στους στόχους εκπαίδευσης, σε συνδυασμό με μεγαλύτερους χρόνους εκπαίδευσης και μεγαλύτερα μεγέθη παρτίδων, οδήγησε τη RoBERTa να ξεπεράσει τον BERT σε διάφορες συγκρίσεις. Το EmoRoBERTa παίρνει τη δύναμη του RoBERTa και το εξειδικεύει στην ταξινόμηση συναισθημάτων. Είναι εκπαιδευμένο στο σύνολο δεδομένων "Go Emotions", ένα σύνολο δεδομένων που περιέχει σχόλια του Reddit. Κάθε σχόλιο σε αυτό το σύνολο δεδομένων επισημαίνεται με μία από τις 28 κατηγορίες συναισθημάτων, οι οποίες περιλαμβάνουν: θαυμασμό, διασκέδαση, θυμό, ενόχληση, έγκριση, Φροντίδα, σύγχυση, περιέργεια, επιθυμία, απογοήτευση, αποδοκιμα-

σία, αηδία, αμηχανία, ενθουσιασμός, Φόβος, ευγνωμοσύνη, Θλίψη, χαρά, αγάπη, νευρικήτητα, αισιοδοξία, απαισιοδοξία, συνειδητοποίηση, ανακούφιση, τύψεις, θλίψη, έκπληξη και ουδέτερο. Αυτό το σύνολο δεδομένων παρέχει μια πλούσια πηγή διαφορετικών συναισθηματικών εκφράσεων, καθιστώντας το ιδανική επιλογή για την εκπαίδευση ενός μοντέλου να αναγνωρίζει και να ταξινομεί τα συναισθήματα στο κείμενο. Το μοντέλο EmoRoBERTa φιλοξενείται στο Model Hub του Hugging Face, μια πλατφόρμα που διευκολύνει την κοινή χρήση και τη συνεργασία μοντέλων μηχανικής μάθησης. Το Model Hub παρέχει εύκολη πρόσβαση σε προεκπαιδευμένα μοντέλα, επιτρέποντας στους ερευνητές και τους προγραμματιστές να αξιοποιήσουν αυτά τα μοντέλα για τις συγκεκριμένες εργασίες τους χωρίς την ανάγκη εκτεταμένης εκπαίδευσης. Το EmoRoBERTa, ως μέρος του, επωφελείται από αυτό το οικοσύστημα, επιτρέποντας στους χρήστες να ενσωματώνουν εύκολα τις δυνατότητες ταξινόμησης συναισθημάτων στις εφαρμογές τους. Συμπερασματικά, το EmoRoBERTa αποτελεί απόδειξη της προσαρμοστικότητας και της ισχύος μοντέλων που βασίζονται σε μετασχηματιστές όπως το RoBERTa. Με τη λεπτομερή ρύθμιση σε ένα εξειδικευμένο σύνολο δεδομένων, το EmoRoBERTa παρέχει κορυφαίες επιδόσεις στην ταξινόμηση συναισθημάτων, καθιστώντας το ένα πολύτιμο εργαλείο για ερευνητές και προγραμματιστές που εργάζονται στον τομέα της ανάλυσης συναισθημάτων και της ανίχνευσης συναισθημάτων.

3.5.4 Twitter-RoBERTa για Ανάλυση συναισθημάτων



Σχήμα 3.14: Αποτελέσματα του Twitter-RoBERTa από τυχαίο κείμενο του συνόλου δεδομένων.

Η ταχεία ανάπτυξη των πλατφορμών μέσων κοινωνικής δικτύωσης, ιδιαίτερα του Twitter, έχει οδηγήσει σε αυξανόμενη ζήτηση για εργαλεία ικανά να κατανοούν και να αναλύουν τα συναισθήματα που εκφράζονται στα tweets. Αναγνωρίζοντας

αυτή την ανάγκη, το μοντέλο `twitter-roberta-base-sentiment` αναπτύχθηκε ως μια εξειδικευμένη προσαρμογή του μοντέλου βάσης `RoBERTa`, προσαρμοσμένο ειδικά για την ανάλυση συναισθήματος στα tweets. Προερχόμενη από το μοντέλο `RoBERTa`, το οποίο είναι μια παραλλαγή του `BERT` (Αμφίδρομες Αναπαραστάσεις Κωδικοποιητή από τους `Transformers`), αυτή η προσαρμογή που εστιάζει στο συναίσθημα εκπαιδεύτηκε στην εργασία συναισθήματος του συνόλου δεδομένων `TweetEval`. Το `TweetEval` είναι ένα σύνολο δεδομένων σχεδιασμένο για ταξινόμηση tweet, που περιλαμβάνει επτά διαφορετικές εργασίες, όπως συναίσθημα, ρητορική μίσους, ειρωνεία, προσβλητική γλώσσα και σαρκασμό. Για την εργασία του συναισθήματος, το μοντέλο εκπαιδεύεται να προβλέπει τρεις διακριτές κατηγορίες: αρνητική, ουδέτερη και θετική. Το μοντέλο `twitter-roberta-base-sentiment` παρουσιάζει την προσαρμοστικότητα μοντέλων που βασίζονται σε μετασχηματιστές όπως το `RoBERTa`. Αξιοποιώντας την τεράστια και ποικιλόμορφη φύση των tweets, το μοντέλο είναι βελτιστοποιημένο ώστε να αποτυπώνει τις λεπτές αποχρώσεις των συναισθημάτων που εκφράζονται στη συνοπτική μορφή των tweets. Αυτό το καθιστά ένα ανεκτίμητο εργαλείο για ερευνητές και προγραμματιστές που στοχεύουν να εμβαθύνουν στην ανάλυση συναισθημάτων για ανεπίσημα κείμενα.

3.6 Προ επεξεργασία κειμένου

3.6.1 Tokenization

Το `tokenization` [73] είναι μια βασική διαδικασία στην Επεξεργασία Φυσικής Γλώσσας (NLP) και στην υπολογιστική γλωσσολογία. Περιλαμβάνει τον διαχωρισμό ενός κειμένου σε ατομικές μονάδες, που συνήθως αναφέρονται ως "tokens". Συνηθέστερα αυτά τα tokens είναι λέξεις, αλλά μπορούν επίσης να είναι προτάσεις, παράγραφοι ή ακόμη και χαρακτήρες, ανάλογα με τη λεπτομέρεια της ανάλυσης. Ο κύριος στόχος του `tokenization` είναι να μετατρέψει μια συνεχή ροή κειμένου σε διακριτά, διαχειρίσιμα κομμάτια, καθιστώντας το πιο εύκολο για επόμενες διαδικασίες.

Η διαδικασία του `tokenization` μπορεί να φαίνεται απλή, αλλά μπορεί να είναι και πολύ περίπλοκη, ιδιαίτερα όταν αντιμετωπίζουμε διαφορετικές γλώσσες και συστήματα γραφής. Για παράδειγμα, ενώ τα κενά διαχωρίζουν σαφώς τις λέξεις σε γλώσσες όπως τα Αγγλικά, γλώσσες όπως τα Κινεζικά ή τα Ταϊλανδικά δεν χρησιμοποιούν τα κενά με τον ίδιο τρόπο, καθιστώντας το λεξιλογικό `tokenization` πιο δύσκολη. Επιπλέον τα σημεία στίξης, οι συστολές και οι ειδικοί χαρακτήρες περιπλέκουν περαιτέρω τη διαδικασία.

Παρά τις προκλήσεις της, το `tokenization` είναι ένα αναπόσπαστο βήμα στην ανάλυση κειμένου. Διαιρώντας το κείμενο σε ατομικά tokens, θέτει την αρχή για πιο προηγμένες εργασίες NLP, από συντακτικό parsing έως την ανάλυση συναισθήματος και την μετάφραση. Λειτουργεί ως γέφυρα μεταξύ του ακατέργαστου κειμένου και των δομημένων δεδομένων, επιτρέποντας στις μηχανές να κατανοούν και να επεξεργάζονται την ανθρώπινη γλώσσα με μεγαλύτερη αποτελεσματικότητα.

BERT και DistilBERT Tokenizers

Ο BERT tokenizer είναι ένα βασικό συστατικό του μετασχηματιστή BERT [74]. Χρησιμοποιεί την μέθοδο WordPiece [75] όπου δημιουργεί άπληστα ένα λεξιλόγιο σταθερού μεγέθους που περιέχει χαρακτήρες, υπολέξεις και λέξεις. Για παράδειγμα μια λέξη όπως η "unffable" μπορεί να παράγει το εξής λεξιλόγιο ["un", "aff", "able"]. Από τα πιο σπουδαία χαρακτηριστικά του είναι ότι μπορεί να διαχειρίζεται λέξεις που δεν βρίσκονται στο λεξιλόγιο του. Όταν έρχεται αντιμέτωπος με μια τέτοια λέξη, ο tokenizer την αναλύει σε μικρότερες υπολέξεις. Εάν αυτές οι υπολέξεις παραμείνουν μη αναγνωρισμένες, η διαδικασία συνεχίζεται, αναλύοντάς τις περαιτέρω έως ότου φτάσουν την ευαισθησία των μεμονωμένων χαρακτήρων. Αυτή η σχολαστική ανάλυση διασφαλίζει την ικανότητα του tokenizer να επεξεργάζεται οποιαδήποτε λέξη, ακόμη και εκείνες που βρίσκονται εκτός του προκαθορισμένου λεξιλογίου του.

Ο DistilBERT Tokenizer αποτελεί ένα βασικό στοιχείο του μετασχηματιστή DistilBERT [35]. Χρησιμοποιεί την μέθοδο SentencePiece [76], η οποία διαφέρει από τη μέθοδο WordPiece που χρησιμοποιεί ο BERT Tokenizer. Το SentencePiece διαχειρίζεται δυναμικά το κείμενο, διασπώντας το σε υπολέξεις, επιτρέποντας την χρήση λέξεων που δεν υπάρχουν στο αρχικό λεξιλόγιο. Αυτό εξασφαλίζει την ικανότητα του tokenizer να επεξεργάζεται πρακτικά οποιαδήποτε λέξη, ακόμη και αυτές που είναι εκτός του προκαθορισμένου λεξιλογίου του, χωρίς την ανάγκη για ένα εξαιρετικά μεγάλο και προκαθορισμένο λεξιλόγιο.

Και οι δύο μετασχηματιστές ενσωματώνουν επίσης ένα σύνολο ειδικών tokens σχεδιασμένων για διαφορετικούς σκοπούς. Το token [CLS], που τοποθετείται στην αρχή κάθε πρότασης, είναι ζωτικής σημασίας για τις εργασίες ταξινόμησης. Το token [SEP] χρησιμεύει ως διαχωριστικό μεταξύ προτάσεων, ιδιαίτερα σημαντικό όταν χειρίζεται πολλές προτάσεις σε εργασίες όπως η απάντηση ερωτήσεων. Επιπλέον, το token [MASK] παίζει κεντρικό ρόλο στη μοντελοποίηση γλώσσας με μάσκα, όπου ένα συγκεκριμένο token αντικαθίσταται από αυτό το ειδικό token, προκαλώντας το μοντέλο να προβλέψει το κείμενο. Κάθε token που αναγνωρίζεται από τον tokenizer αντιστοιχεί σε ένα συγκεκριμένο αναγνωριστικό μέσα στο λεξιλόγιό του. Αυτά τα tokens χρησιμεύουν ως την κύρια είσοδο στα μοντέλα. Σε σενάρια που περιλαμβάνουν δύο προτάσεις, οι μετασχηματιστές χρησιμοποιούν αναγνωριστικά τμημάτων για τη διαφοροποίηση μεταξύ τους, εκχωρώντας ID 0 στην πρώτη και ID 1 στην επόμενη πρόταση. Επιπλέον, για να διατηρήσουν την επίγνωση της θέσης κάθε token μέσα σε μια ακολουθία, οι μετασχηματιστές χρησιμοποιούν αναγνωριστικά θέσης. Αυτά συγχωνεύονται με τα embeddings πριν από την εισαγωγή τους στο μοντέλο BERT.

Οι ακολουθίες που εισάγονται στον μετασχηματιστή, πρέπει να έχουν σε ένα ομοιόμορφο μήκος όταν συλλέγονται. Αυτό απαιτεί τις διαδικασίες πλήρωσης (padding) και περικοπής (truncation). Οι ακολουθίες που υπολείπονται του μέγιστου μήκους συμπληρώνονται με μηδενικά, εξασφαλίζοντας ομοιομορφία. Αντίθετα, οι ακολουθίες που υπερβαίνουν αυτό το όριο περικόπτονται. Αξίζει να σημειωθεί ότι το τυπικό μέγιστο μήκος για περικοπή στο BERT είναι 512 tokens.

Τέλος, ο μηχανισμός προσοχής στους μετασχηματιστές καθοδηγείται από τη μάσκα προσοχής. Αυτή η μάσκα διασφαλίζει ότι η προσοχή του μοντέλου κατευθύνε-

ται αποκλειστικά σε σημαντικά διακριτικά και ουσιαστικά αγνοεί τα padding tokens. Ορίζει tokens που δικαιολογούν την προσοχή με τιμή 1 και εκείνα που πρέπει να παραβλεφθούν με τιμή 0.

Είναι σημαντικό να επισημάνουμε ότι ο κάθε μετασχηματιστής χρειάζεται το αντίστοιχο tokenizer του για λόγους συμβατότητας και λεξιλογίου. Συγκεκριμένα το λεξιλόγιο του BERT είναι πιο μεγάλο από του DistilBERT και για να χρησιμοποιηθεί από τον DistilBERT θα χρειαστούν πολλές προσαρμογές. Ακόμη το λεξιλόγιο του BERT περιέχει επίσης πιο ιδιαίτερους χαρακτήρες από του DistilBERT όπως αυτός της δίεσης ("#").

3.6.2 Stemming

Το stemming είναι μια διαδικασία στην επεξεργασία φυσικής γλώσσας, όπου οι λέξεις μετατρέπονται στη βασική τους μορφή, αφαιρώντας τις καταλήξεις. Στόχος της διαδικασίας αυτής είναι η συμπίεση των λέξεων σε μια κοινή ρίζα, ώστε να μπορούν να αντιμετωπιστούν ως ίδιες από τα συστήματα ανάλυσης κειμένου. Για παράδειγμα, στην ελληνική γλώσσα, οι λέξεις "τρέχω", "τρέχεις" και "τρέχαμενος" έχουν την ίδια ρίζα, "τρεχ-". Μέσω του stemming, όλες αυτές οι λέξεις θα μετατραπούν στην ίδια βασική μορφή, δηλαδή "τρεχ".

Η διαδικασία του stemming είναι ιδιαίτερα χρήσιμη σε εφαρμογές όπως η αναζήτηση πληροφοριών, η ανάλυση συναισθημάτων και άλλες εφαρμογές επεξεργασίας κειμένου. Επιτρέπει στα συστήματα να αναγνωρίζουν ότι διαφορετικές μορφές μιας λέξης σχετίζονται με την ίδια βασική έννοια. Ωστόσο, το stemming δεν είναι πάντα τέλειο. Σε ορισμένες περιπτώσεις, μπορεί να αφαιρέσει καταλήξεις που αλλάζουν σημαντικά το νόημα μιας λέξης ή να μην αφαιρέσει καταλήξεις όταν είναι απαραίτητο. Επιπλέον, σε γλώσσες με πολύπλοκη μορφολογία, όπως η ελληνική, η διαδικασία του stemming μπορεί να είναι πιο πολύπλοκη.

Παρά τις προκλήσεις, το stemming είναι μια βασική τεχνική στην επεξεργασία φυσικής γλώσσας. Σε συνδυασμό με άλλες τεχνικές, όπως το lemmatization (η διαδικασία μετατροπής μιας λέξης στη βασική της λεξιλογική μορφή), το stemming μπορεί να βοηθήσει στη βελτίωση της ακρίβειας και της απόδοσης των συστημάτων επεξεργασίας κειμένου.

3.6.3 Λημματοποίηση

Η λημματοποίηση (lemmatization) αποτελεί μια βασική διαδικασία στην επεξεργασία φυσικής γλώσσας, με στόχο τη μετατροπή των λέξεων στην βασική τους λεξιλογική μορφή ή λήμμα. Ενώ το stemming αποσκοπεί στην αφαίρεση των καταλήξεων των λέξεων, η λημματοποίηση επιδιώκει την αναγνώριση της "κανονικής" μορφής μιας λέξης, λαμβάνοντας υπόψη τη γραμματική της σημασία. Για παράδειγμα, στην ελληνική γλώσσα, οι λέξεις "τρέχω", "τρέχοντας" και "έτρεξα" έχουν το ίδιο λήμμα, το οποίο είναι "τρέχω". Η λημματοποίηση θα μετατρέψει όλες αυτές τις λέξεις στη βασική τους μορφή "τρέχω", ανεξάρτητα από τον χρόνο ή τη μορφή τους.

Η λημματοποίηση είναι ιδιαίτερα χρήσιμη σε εφαρμογές όπως η αναζήτηση πληροφοριών, η ανάλυση συναισθημάτων και η αυτόματη μετάφραση. Επιτρέπει την αναγνώριση της ίδιας λέξης σε διάφορες μορφές και την κατανόηση της πραγματικής

της σημασίας μέσα στο κείμενο. Ωστόσο, η λημματοποίηση μπορεί να είναι μια πολύπλοκη διαδικασία, ιδιαίτερα σε γλώσσες με πλούσια μορφολογία όπως η ελληνική. Απαιτεί γνώση της γραμματικής της γλώσσας και συχνά χρησιμοποιεί λεξικά ή βάσεις δεδομένων για να αναγνωρίσει το κατάλληλο λήμμα.

Σε αντίθεση με το stemming, που επικεντρώνεται στην αφαίρεση καταλήξεων και μπορεί να παράγει λέξεις που δεν υπάρχουν πραγματικά, η λημματοποίηση παράγει πάντα πραγματικές λέξεις και είναι πιο ακριβής στην αναγνώριση της βασικής μορφής μιας λέξης. Ωστόσο, είναι συνήθως πιο χρονοβόρα και απαιτητική από το stemming.

3.7 Διανύσματα αναπαράστασης κειμένου

3.7.1 Bag of Words

Το μοντέλο Bag of Words (BoW) είναι μια βασική τεχνική στην Επεξεργασία Φυσικής Γλώσσας (NLP) και την ανάλυση κειμένου [77]. Στην ουσία του, το BoW μετατρέπει τα κειμενικά δεδομένα σε αριθμητικά διανύσματα αναπαριστώντας τη συχνότητα κάθε λέξης μέσα σε ένα έγγραφο, αγνοώντας τη σειρά ή τη δομή των λέξεων. Ουσιαστικά, ένα κείμενο αντιλαμβάνεται ως μια μη ταξινομημένη "τσάντα" που περιέχει λέξεις, εξ ου και το όνομα.

Στο μοντέλο BoW, πρώτα δημιουργείται ένα λεξιλόγιο από ολόκληρο το σώμα κειμένου, και κάθε έγγραφο αναπαρίσταται ως ένα διάνυσμα με το μήκος αυτού του λεξιλογίου. Κάθε στοιχείο στο διάνυσμα αντιστοιχεί σε μια λέξη στο λεξιλόγιο, και η τιμή του αντιπροσωπεύει τη συχνότητα ή την παρουσία (δυναμική αναπαράσταση) αυτής της λέξης στο έγγραφο. Παρόλο που το BoW προσφέρει απλότητα και αποτελεσματικότητα, ιδιαίτερα για μεγάλα σύνολα δεδομένων, έχει τους περιορισμούς του. Το μοντέλο αγνοεί τις σημασιολογικές σχέσεις, το πλαίσιο και τη σειρά των λέξεων, οι οποίες μπορούν να είναι κρίσιμες για την κατανόηση των λεπτομερειών της γλώσσας.

Παρά αυτούς τους περιορισμούς, το μοντέλο BoW έχει χρησιμοποιηθεί σε πολλά καθήκοντα NLP, όπως η ταξινόμηση κειμένου, η ανάλυση συναισθήματος και η ομαδοποίηση εγγράφων. Η απλότητά του έχει δρομολογήσει πιο προηγμένες τεχνικές, αλλά το BoW παραμένει ένας βασικός λίθος στην εξέλιξη των μεθοδολογιών επεξεργασίας κειμένου.

3.7.2 N-grams

Τα N-grams [78], μια βασική έννοια στην Επεξεργασία Φυσικής Γλώσσας (NLP) και την υπολογιστική γλωσσολογία, αναφέρονται σε συνεχείς ακολουθίες n στοιχείων από ένα δεδομένο δείγμα κειμένου ή ομιλίας. Συνήθως, αυτά τα στοιχεία είναι λέξεις, αλλά μπορούν επίσης να είναι χαρακτήρες ή συλλαβές. Ο κύριος σκοπός των n-grams είναι να καταγράφουν τη γλωσσική δομή, όπως οι φραστικές εκφράσεις και τα μοτίβα συν-εμφάνισης λέξεων, από ένα κείμενο. Για παράδειγμα, ένα bigram (2-gram) καταγράφει ζεύγη διαδοχικών λέξεων, ενώ ένα trigram (3-gram) καταγράφει ακολουθίες τριών λέξεων.

Η χρήση των n-grams βοηθά στη διατήρηση των τοπικών διαδοχικών πληροφοριών, που συχνά χάνονται σε άλλες τεχνικές αναπαράστασης κειμένου όπως το Bag of Words (BoW). Λαμβάνοντας υπόψη ακολουθίες λέξεων, τα n-grams μπορούν να προσφέρουν περισσότερες πληροφορίες με χρήση της συναφής έκφρασης, καθιστώντας τα πολύτιμα για καθήκοντα όπως η πρόβλεψη κειμένου, η διόρθωση ορθογραφίας και η ανάλυση συναισθήματος. Ωστόσο, καθώς το 'n' αυξάνεται, η διαστατικότητα των δεδομένων μπορεί να αυξηθεί εκθετικά, οδηγώντας σε προβλήματα αραιότητας δεδομένων και αυξημένο υπολογιστικό κόστος.

Παρά αυτές τις προκλήσεις, τα n-grams έχουν χρησιμοποιηθεί σε διάφορες εφαρμογές NLP, από παραδοσιακά στατιστικά μοντέλα γλώσσας έως σύγχρονες αρχιτεκτονικές βαθιάς μάθησης. Η ικανότητά τους να ενσωματώνουν το τοπικό πλαίσιο στα κειμενικά δεδομένα έχει καθιερώσει τη θέση τους ως βασική τεχνική στον τομέα της ανάλυσης κειμένου.

3.7.3 TF-IDF

Το TF-IDF, ακρωνύμιο για Term Frequency-Inverse Document Frequency, είναι μια ευρέως υιοθετημένη στατιστική μέτρηση στην Επεξεργασία Φυσικής Γλώσσας (NLP) και την ανάκτηση πληροφοριών. Ο κύριος στόχος του είναι να ποσοτικοποιήσει τη σημασία ενός όρου μέσα σε ένα έγγραφο σε σχέση με μια συλλογή εγγράφων, που συχνά αναφέρεται ως σώμα. Η μέτρηση είναι ένα προϊόν δύο συστατικών: Term Frequency (TF) και Inverse Document Frequency (IDF) [79].

Η Term Frequency (TF) αντιπροσωπεύει τη συχνότητα ενός όρου σε ένα συγκεκριμένο έγγραφο. Καταγράφει την ακατέργαστη εμφάνιση του όρου, συχνά κανονικοποιημένη από τον συνολικό αριθμό των όρων στο έγγραφο. Από την άλλη πλευρά, το Inverse Document Frequency (IDF) μετρά τη σπανιότητα ενός όρου σε ολόκληρο το σώμα. Μειώνει το βάρος των όρων που είναι διαδεδομένοι σε πολλά έγγραφα, όπως οι συνηθισμένες λέξεις-σταματήματα. Ο πολλαπλασιασμός του TF και του IDF παρέχει ένα βάρος που τονίζει τους όρους που είναι συχνοί σε ένα συγκεκριμένο έγγραφο αλλά σπάνιοι σε ολόκληρο το σώμα. Οι συναρτήσεις αυτών των δύο για μια λέξη t σε ένα αρχείο d είναι:

$$tf(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}$$

όπου $f_{t,d}$ αριθμός των εμφανίσεων του t στο d και

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

όπου N ο αριθμός των εγγράφων και $|d \in D : t \in d|$ ο αριθμός των εγγράφων όπου η λέξη t εμφανίζεται. Συνολικά

$$tfidf(t, d, D) = tf(t, d) \cdot idf(t, D)$$

Η κομψότητα του TF-IDF βρίσκεται στην ικανότητά του να τονίζει όρους που είναι πιθανώς πιο σημαντικοί για ένα συγκεκριμένο έγγραφο. Έχει χρησιμοποιηθεί σε διάφορα καθήκοντα NLP, όπως η ταξινόμηση εγγράφων, η ομαδοποίηση και η ανάκτηση πληροφοριών. Εξισορροπώντας τη συχνότητα του όρου με την αντίστροφη συχνότητα του εγγράφου, το TF-IDF διασφαλίζει ότι λαμβάνονται υπόψη και οι τοπικές και οι παγκόσμιες χαρακτηριστικές ιδιότητες του κειμένου.

3.8 Αναπαραστάσεις λέξεων (Word Embeddings)

Το word embedding [80] είναι μια μετασχηματιστική τεχνική στην Επεξεργασία Φυσικής Γλώσσας (NLP) που αντιστοιχίζει λέξεις ή φράσεις σε διανύσματα πραγματικών αριθμών. Σε αντίθεση με τις παραδοσιακές αναπαραστάσεις που αντιμετωπίζουν τις λέξεις ως διακριτά σύμβολα, τα word embeddings καταγράφουν τις σημασιολογικές σχέσεις μεταξύ των λέξεων αντιπροσωπεύοντάς τες σε ένα συνεχή διανυσματικό χώρο. Η βασική διαίσθηση πίσω από αυτήν την προσέγγιση είναι ότι οι λέξεις που εμφανίζονται σε παρόμοιες φράσεις τείνουν να έχουν παρόμοιες σημασίες. Εκπαιδεύοντας σε μεγάλα σώματα, τα μοντέλα embeddings μπορούν να καταγράψουν περίπλοκες σημασιολογικές σχέσεις, όπως συνώνυμα, αντώνυμα και αναλογίες.

Δημοφιλή μοντέλα όπως το Word2Vec, το GloVe και το FastText έχουν συμβάλει σημαντικά στην πρόοδο του τομέα των word embeddings. Αυτά τα μοντέλα χρησιμοποιούν νευρωνικά δίκτυα ή τεχνικές παραγοντοποίησης πινάκων για να μάθουν πυκνές διανυσματικές αναπαραστάσεις για λέξεις. Τα προκύπτοντα διανύσματα μπορούν να καταγράψουν σημασιολογικές λεπτομέρειες, επιτρέποντας λειτουργίες όπως η διανυσματική αριθμητική για να αποκαλύψουν περίπλοκες σχέσεις.

Η υιοθέτηση των word embeddings έχει επαναστατίσει διάφορες εργασίες NLP, όπως η ανάλυση συναισθήματος, η μετάφραση και η ανάκτηση πληροφοριών. Παρέχοντας μια πλουσιότερη αναπαράσταση του κειμένου, τα embeddings έχουν γεφυρώσει το χάσμα μεταξύ των ακατέργαστων γλωσσικών δεδομένων και της υψηλότερης σημασιολογικής κατανόησης, επιτρέποντας στις μηχανές να επεξεργάζονται τη γλώσσα με βάθος και λεπτομέρεια άνευ προηγουμένου.

3.9 Μοντελοποίηση Θεμάτων

3.9.1 Latent Dirichlet Allocation

Το Latent Dirichlet Allocation (LDA) [81] είναι ένα πιθανοτικό μοντέλο που χρησιμοποιείται ευρέως στην Επεξεργασία Φυσικής Γλώσσας (NLP) και την εξόρυξη κειμένου για να ανακαλύψει αφηρημένα θέματα εντός μιας συλλογής εγγράφων. Το LDA υποθέτει ότι τα έγγραφα είναι μείγματα θεμάτων και ότι τα θέματα είναι κατανομές πάνω σε λέξεις. Ο κύριος στόχος του LDA είναι να καθορίσει το μείγμα των θεμάτων που περιέχει ένα έγγραφο και την κατανομή των λέξεων για κάθε θέμα, επιτρέποντας την ταυτοποίηση μοτίβων και θεμάτων εντός μεγάλων κειμενικών συνόλων δεδομένων.

Η βασική αρχή του LDA βασίζεται στην κατανομή Dirichlet. Σύμφωνα με το LDA, κάθε έγγραφο σε ένα σώμα αντιπροσωπεύεται ως μια κατανομή θεμάτων, και κάθε θέμα είναι σχεδιασμένο ως μια κατανομή λέξεων. Εκμεταλλευόμενο την κατανομή Dirichlet και τις επαναλαμβανόμενες τεχνικές δειγματοληψίας, το LDA μπορεί να συμπεράνει αποτελεσματικά τη δομή των θεμάτων ενός σώματος.

Η εφαρμογή του LDA έχει συμβάλει σημαντικά σε διάφορους τομείς, από συστήματα συστάσεων περιεχομένου έως την ανάκτηση πληροφοριών. Μετατρέποντας το μη δομημένο κείμενο σε δομημένες αναπαραστάσεις θεμάτων, το LDA παρέχει

μια βαθύτερη κατανόηση της κρυφής σημασιολογικής δομής εντός μεγάλων όγκων κειμένου, επιτρέποντας στους ερευνητές και τους επαγγελματίες να αντλούν πληροφορίες, να κατηγοριοποιούν έγγραφα και να ανιχνεύουν εμφανιζόμενες τάσεις.

3.10 Επαύξησή δεδομένων κειμένου

Η επαύξησή δεδομένων κειμένου (text augmentation) είναι μια τεχνική επέκτασης δεδομένων ειδικά προσαρμοσμένη για εργασίες επεξεργασίας φυσικής γλώσσας (NLP) [82]. Δεδομένης της σημασίας των δεδομένων στην εκπαίδευση των μοντέλων μηχανικής μάθησης, η διαθεσιμότητα ενός περιορισμένου ή ανισορροπημένου συνόλου δεδομένων μπορεί να εμποδίσει την απόδοση του μοντέλου. Η επαύξησή αυτή στοχεύει στην αντιμετώπιση των παραπάνω, αυξάνοντας τεχνητά το μέγεθος και την ποικιλομορφία των συνόλων δεδομένων εκπαίδευσης. Δημιουργώντας διαφορετικές εκδοχές των αρχικών κειμενικών δεδομένων, διασφαλίζει ότι τα μοντέλα εκτίθενται σε ευρύτερο φάσμα γλωσσικών κατασκευών, οδηγώντας σε βελτιωμένη γενίκευση και ανθεκτικότητα.

Οι μέθοδοι για την επαύξησή είναι πολλαπλές. Κοινές τεχνικές περιλαμβάνουν την "πίσω μετάφραση" (back translation) (μεταφράζοντας μια πρόταση σε άλλη γλώσσα και στη συνέχεια πίσω στην αρχική γλώσσα), την αντικατάσταση συνωνύμων (αντικαθιστώντας λέξεις στο κείμενο με τα συνώνυμά τους) και την ανακατάταξη προτάσεων. Πιο προηγμένες μέθοδοι εκμεταλλεύονται νευρωνικά μοντέλα, όπως τα variational autoencoders και τα transformers, για να δημιουργήσουν σημασιολογικά συνεπείς επεκτατικά κείμενα. Η επιλογή της μεθόδου συχνά εξαρτάται από το συγκεκριμένο εργασία NLP και τη φύση των διαθέσιμων δεδομένων. Μια πολύ ενδιαφέρουσα μέθοδος είναι αυτή που χρησιμοποιεί transformers και συγκεκριμένα το BERT. Εκμεταλλευόμενοι τα προ-εκπαιδευμένα embeddings του, οι ερευνητές μπορούν να δημιουργήσουν σημασιολογικά συνεπείς και διαφοροποιημένες κειμενικές παραλλαγές. Συγκεκριμένα, λέξεις ή φράσεις εντός μιας πρότασης μπορούν να κρυφτούν και στη συνέχεια να αντικατασταθούν με προβλέψεις από το BERT, διασφαλίζοντας ότι το επεκτατικό κείμενο διατηρεί σημασιολογική ακεραιότητα ενώ εισάγει ποικιλομορφία. Αυτή η προσέγγιση όχι μόνο εμπλουτίζει τα δεδομένα εκπαίδευσης, αλλά επίσης εκμεταλλεύεται την τεράστια γνώση που περιέχεται στα προ-εκπαιδευμένα βάρη του BERT, γεφυρώνοντας το χάσμα μεταξύ της έλλειψης δεδομένων και της ανθεκτικότητας του μοντέλου.

Η σημασία της επαύξης δεδομένων, ειδικά στην ταξινόμηση κειμένου, τονίζεται περαιτέρω από την εργασία των Wei και Zhou [83]. Η εργασία εξετάζει τα πολλαπλά οφέλη της επαύξης δεδομένων:

- Ευρωστία σε μικρότερα σύνολα δεδομένων: Η ταξινόμηση κειμένων είναι ζωτικής σημασίας στο NLP και η απόδοσή της εξαρτάται συχνά από το μέγεθος και την ποιότητα των δεδομένων εκπαίδευσης. Η επαύξηση, ιδιαίτερα σε μικρότερα σύνολα δεδομένων, μπορεί να ενισχύσει την εκπαίδευση πιο εύρωστων μοντέλων. Με τη δημιουργία δεδομένων παρόμοιων με το πρωτότυπο, αλλά με μικρές παραλλαγές, εισάγεται ένα επίπεδο θορύβου, περιορίζοντας την υπερπροσαρμογή (overfitting).
- Εισαγωγή νέου λεξιλογίου: Οι τεχνικές EDA, όπως η αντικατάσταση συνωνύ-

μων και η τυχαία εισαγωγή, μπορούν να εισάγουν νέο λεξιλόγιο. Αυτό εξοπλίζει τα μοντέλα να γενικεύονται σε λέξεις που απουσίαζαν στο σετ εκπαίδευσης.

- Αποτροπή υπερπροσαρμογής: Ορισμένες λειτουργίες EDA μπορεί να φαίνονται αντιφατικές, όπως τυχαίες εναλλαγές ή εισαγωγές. Ωστόσο, ακόμα κι αν το επαυξημένο κείμενο φαίνεται παράξενο, διατηρεί το μεγαλύτερο μέρος των αρχικών λέξεων και τη σειρά τους με κάποιον επιπρόσθετο θόρυβο. Αυτός ο θόρυβος είναι καθοριστικός για την αποφυγή της υπερπροσαρμογής.

Η εφαρμογή της επαύξησής έχει αποδειχθεί ωφέλιμη σε διάφορες εργασίες NLP, από την ανάλυση συναισθήματος έως την αναγνώριση ονοματισμένων οντοτήτων. Με τη μετρίαση των ζητημάτων που σχετίζονται με τη σπανιότητα και την ανισορροπία των δεδομένων, η επαύξησή διαδραματίζει κεντρικό ρόλο στην ενίσχυση της απόδοσης και της αξιοπιστίας των μοντέλων NLP, καθιστώντας τα πιο προσαρμοστικά σε γλωσσικές παραλλαγές.

3.11 Συναρτήσεις ενεργοποίησης

Εδώ θα παραθέσουμε τις πιο συνήθεις συναρτήσεις ενεργοποίησης που χρησιμοποιούνται:

3.11.1 Σιγμοειδής συνάρτηση

Η σιγμοειδής (sigmoid) συνάρτηση έχει την εξής εξίσωση:

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

Όπως φαίνεται και στο Διάγραμμα 3.15 η έξοδος της βρίσκεται στο διάστημα [0,1] παρουσιάζοντας την έξοδο ως πιθανότητα. Σε γενικές γραμμές η σιγμοειδής συνάρτηση χρησιμοποιείται σε νευρωνικά δίκτυα με λίγα στρώματα [84].

3.11.2 Softmax συνάρτηση

Η συνάρτηση softmax έχει την ακόλουθη εξίσωση:

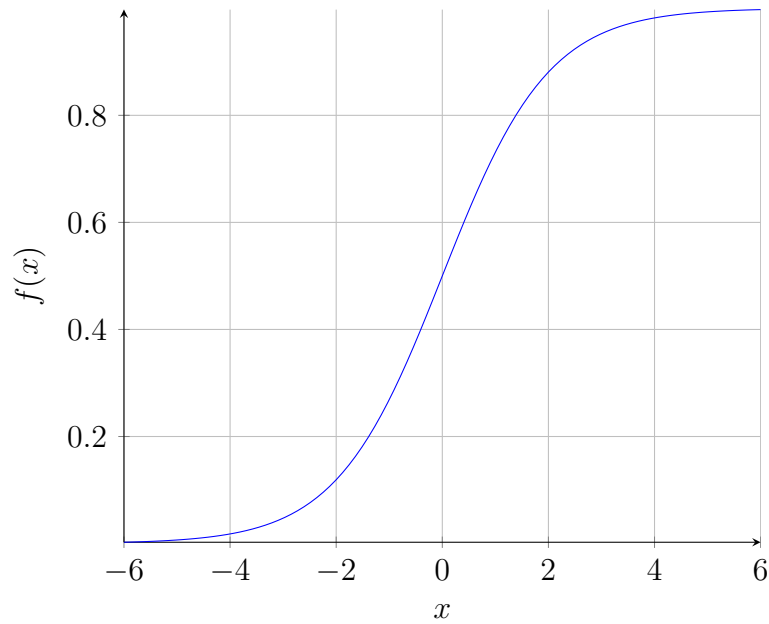
$$f(x) = \frac{e^{x_i}}{\sum_j e^j}$$

Είναι αντίστοιχη της σιγμοειδούς συνάρτησης, όπως φαίνεται και στο Διάγραμμα 3.16 αλλά χρησιμοποιείται κυρίως για την ταξινόμηση πολλών κλάσεων καθώς επιστρέφει ένα διάνυσμα με τις πιθανότητες της κάθε κλάσης[84].

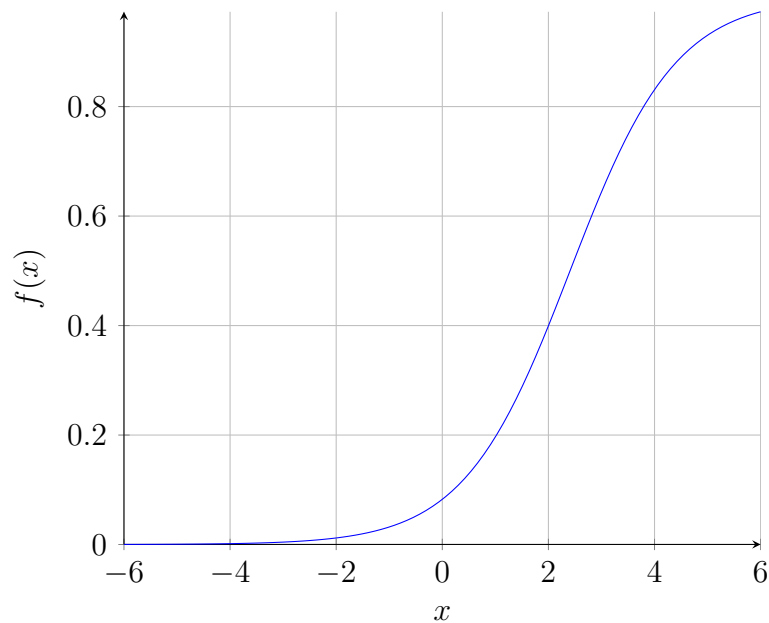
3.11.3 Υπερβολική Εφαπτομένη

Η συνάρτηση υπερβολικής εφαπτομένης (tanh) έχει την ακόλουθη εξίσωση:

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



Σχήμα 3.15: Η Σιγμοειδής συνάρτηση.

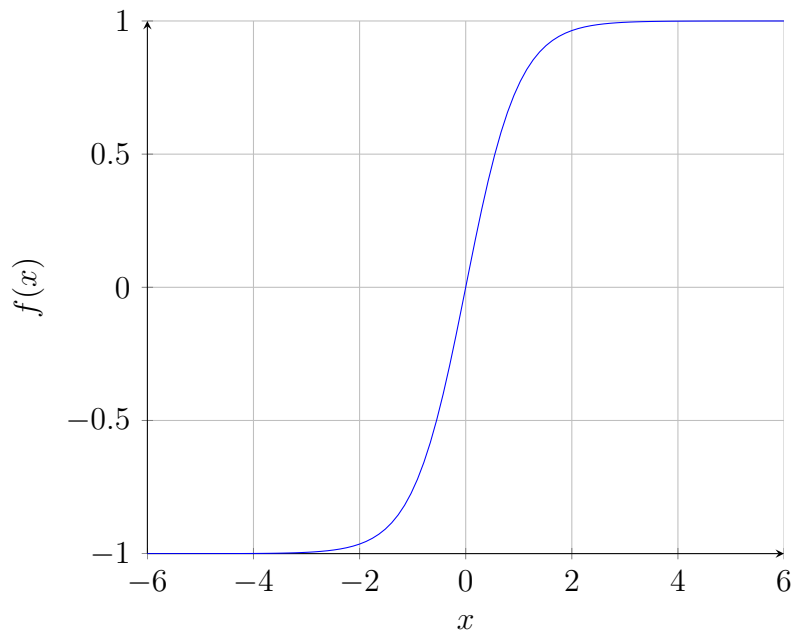


Σχήμα 3.16: Η συνάρτηση Softmax.

Η υπερβολική εφαπτομένη, όπως φαίνεται και στο Διάγραμμα 3.17, είναι ομαλή και περιορίζεται στο διάστημα $[-1, 1]$. Η συνάρτηση αυτή χρησιμοποιείται ευρέως για την επεξεργασία φυσικής γλώσσας και κυρίως σε αναδρομικά νευρωνικά δίκτυα[84].

3.11.4 Ανορθωμένη Γραμμική Μονάδα

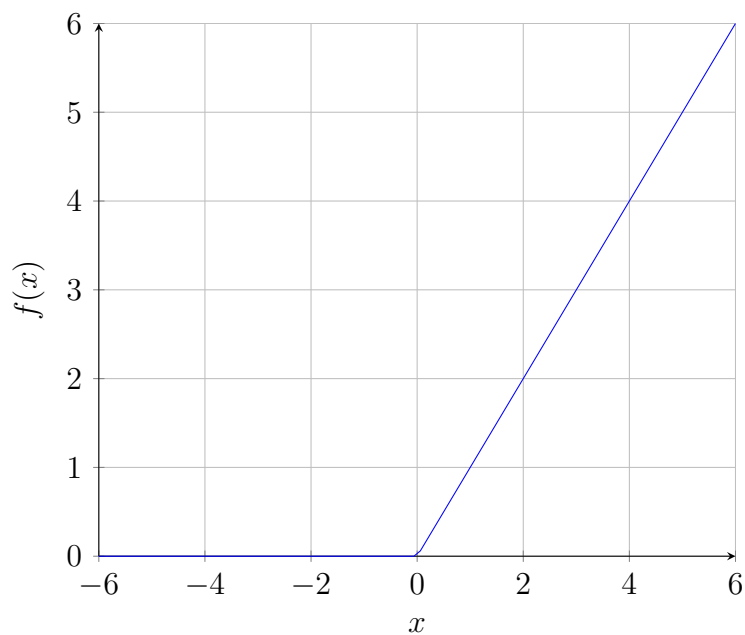
Η ανορθωμένη γραμμική μονάδα (Rectified Linear Unit, ReLU) έχει την παρακάτω εξίσωση και η γραφική της παράσταση φαίνεται στο Διάγραμμα 3.18:



Σχήμα 3.17: Η Υπερβολική Εφαπτομένη.

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

Επιστρέφει το x αν είναι μη αρνητικό αλλιώς μηδέν. Είναι η πιο χρησιμοποιημένη σε δίκτυα βαθιάς μάθησης και προσφέρει γρήγορο υπολογισμό των αποτελεσμάτων [84].



Σχήμα 3.18: Η RELU.

3.12 Μετρικές

Για να συγκρίνουμε τα μοντέλα και τις παραμέτρους τους και για να δούμε ποια λειτουργούν καλύτερα πρέπει να χρησιμοποιήσουμε κάποιες μετρικές (συναρτήσεις) που έχουν ως παραμέτρους τις προβλέψεις του μοντέλου και τα πραγματικά αποτελέσματα. Για να μετρήσουμε την ορθότητα των προβλέψεων σε μια δυαδική ταξινόμηση χρειάζεται να γνωρίζουμε το σύνολο των δεδομένων που ταξινομήθηκαν σωστά στην θετική κλάση (αληθώς θετικό, true positive, TP), το σύνολο των δεδομένων που ταξινομήθηκαν λάθος στην θετική κλάση (ψευδώς θετικό, false positive, FP), το σύνολο των δεδομένων που ταξινομήθηκαν σωστά στην αρνητική κλάση (αληθώς αρνητικό, true negative, TN) και το σύνολο των δεδομένων που ταξινομήθηκαν λάθος στην αρνητική κλάση (ψευδώς αρνητικό, false negative, FN). Όμοια αν έχουμε περισσότερες κλάσεις επεκτείνουμε τις παραπάνω μετρικές για την κάθε κλάση [85].

3.12.1 Πίνακας Σύγχυσης

Ο πίνακας σύγχυσης (confusion matrix) είναι η πιο απλή και πολύ δημοφιλής τεχνική και λειτουργεί τόσο για δυαδικές ταξινομήσεις όσο και για όταν υπάρχουν περισσότερες κλάσεις. Ο πίνακας σύγχυσης ουσιαστικά παρέχει μια εποπτική αξιολόγησή του μοντέλου αναγράφοντας τις τιμές που υπολογίσαμε παραπάνω [86]. Ένα παράδειγμα πίνακα σύγχυσης σε δυαδική ταξινόμηση είναι ο Πίνακας 3.1.

		Πρόβλεψη (Predicted value)	
		Θετικό (Positive)	Αρνητικό (Negative)
Πραγματική τιμή (Actual value)	Θετικό (Positive)	Αληθώς Θετικό (True Positive, TP)	Ψευδώς Αρνητικό (False Negative, FN)
	Αρνητικό (Negative)	Ψευδώς Θετικό (False Positive, FP)	Αληθώς Αρνητικό (True Negative, TN)

Πίνακας 3.1: Πίνακας Σύγχυσης (Confusion Matrix).

3.12.2 Ακρίβεια (Accuracy)

Η πιο χρησιμοποιημένη μετρική για δεδομένα με δύο τάξεις είναι η ακρίβεια. Η εξίσωση της είναι η εξής:

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

Ενώ είναι μια πολύ καλή μετρική για να έχουμε μια γενική εικόνα του μοντέλου μας, σε περίπτωση που τα δεδομένα είναι μη ισορροπημένα τότε το accuracy θα είναι πολύ υψηλό και θα μας δίνει την λάθος άποψη ότι το μοντέλο μας είναι καλό [86].

Σε περίπτωση που έχουμε άνω των δύο κλάσεων μπορούμε να υπολογίσουμε την μέση ακρίβεια ως εξής [85]:

$$AverageAccuracy = \frac{\sum_{i=1}^l \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{l}$$

3.12.3 Ακρίβεια (Precision) και Ανάκληση (Recall)

Η ακρίβεια (precision) μετράει το ποσοστό των αληθών θετικών προβλέψεων ως προς τις θετικές προβλέψεις ενώ η ανάκληση (recall) μετράει το ποσοστό των αληθών θετικών προβλέψεων ως προς τις θετικές τιμές. Και οι δύο είναι πολύ σημαντικές μετρικές και σε γενικές γραμμές επιζητούμε να αυξάνουμε το recall, κρατώντας σταθερό το precision[86][87]. Οι εξισώσεις για τις δύο αυτές μετρικές δίνονται παρακάτω.

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

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

Για περισσότερες κλάσεις μπορούμε να επεκτείνουμε τις παραπάνω εξισώσεις για να βρούμε τον μέσο όρο micro και macro (micro average, macro average). Ο micro average υπολογίζεται λαμβάνοντας το άθροισμα όλων των μετρικών των κλάσεων, ενώ ο macro average υπολογίζει την μετρική για κάθε κλάση ξεχωριστά και μετά βρίσκει τον μέσο όρο τους. Στον macro average εάν προσθέσουμε στην εξίσωση βάρη, τέτοια ώστε να αναλογούν στο πλήθος των στοιχείων για κάθε κλάση λαμβάνουμε το weighted average που λαμβάνει υπόψη την ανισορροπία των δεδομένων[86][87].

$$Precision_{\mu} = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FP_i}$$

$$Precision_M = \frac{\sum_{i=1}^l \frac{TP_i}{TP_i + FP_i}}{l}$$

$$Recall_{\mu} = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FN_i}$$

$$Recall_M = \frac{\sum_{i=1}^l \frac{TP_i}{TP_i + FN_i}}{l}$$

3.12.4 F1 score

Η μετρική F1 για να υπολογίσει το αποτέλεσμα χρησιμοποιεί την ακρίβεια (precision) και την ανάκληση. Συγκεκριμένα είναι ο αρμονικός μέσος όρος μεταξύ των δύο αυτών μετρικών. Η εξίσωση της είναι:

$$F_1 = \frac{2TP}{2TP + FP + FN}$$

όπως και πριν μπορούμε να ορίσουμε τα micro, macro, weighted average της μετρικής[86][87].

Κεφάλαιο 4

Μεθοδολογία

Στο παρόν κεφάλαιο θα παρουσιάσουμε την μεθοδολογία της διπλωματικής. Αναλυτικότερα στην Ενότητα 4.1 θα αναλύσουμε το σύνολο δεδομένων, την επαύξηση δεδομένων που του εφαρμόσαμε και την προεπεξεργασία του. Μετά στην Ενότητα 4.2 θα μελετήσουμε τα χαρακτηριστικά που εξάγαμε από τα δεδομένα και θα καταγράψουμε λεπτομερώς το υβριδικό μοντέλο που προτείνουμε. Αργότερα στην Ενότητα 4.3 θα αναφέρουμε το πειραματικό περιβάλλον και τέλος στην Ενότητα 4.4 θα αναλύσουμε παραλλαγές του μοντέλου μας.

4.1 Σύνολο Δεδομένων και Προεπεξεργασία

4.1.1 Σύνολο δεδομένων

Τα δεδομένα που χρησιμοποιήθηκαν προέρχονται από το αποθετήριο `usmaann/Depression_Severity_Dataset`¹ [88]. Τα δεδομένα αυτά δημιουργήθηκαν για να εξυπηρετήσουν τους σκοπούς του έργου από τους Naseem et al., που παρουσιάστηκε στο συνέδριο "The Web Conference '22". Ο κύριος στόχος του συνόλου δεδομένων είναι να κατηγοριοποιήσει τα δεδομένα κειμένου από το Reddit σε τέσσερα διαφορετικά επίπεδα σοβαρότητας κατάθλιψης: Ελάχιστο (minimum), Ήπιο (mild), Μέτριο (moderate) και Σοβαρό (severe).

Το σύνολο δεδομένων αποτελείται από δύο στήλες. Η πρώτη στήλη που ονομάζεται "text" περιέχει τα δεδομένα κειμένου που εξάγονται από αναρτήσεις στο Reddit ενώ η δεύτερη στήλη που ονομάζεται "label" κατηγοριοποιεί την κάθε καταχώρηση κειμένου σε ένα από τα τέσσερα επίπεδα σοβαρότητας της κατάθλιψης. Πιο συγκεκριμένα περιέχει 3553 στοιχεία εκ των οποίων η κατανομή της στήλης label είναι ως εξής:

- **Minimum:** 72.68%
- **Mild:** 11.16%
- **Moderate:** 8.21%
- **Severe:** 7.96%

¹https://github.com/usmaann/Depression_Severity_Dataset

Ο Πίνακας 4.1 περιέχει παραδείγματα των αναρτήσεων.

Κατά τη διάρκεια της ανακάλυψης παρεμφερούς βιβλιογραφίας βρήκαμε πολλά σύνολα δεδομένων. Τέτοια είναι τα Reddit Self-reported Depression Diagnosis (RSDD) [89] dataset, το Self-reported Mental Health Diagnoses (SMHD) [90] dataset, το eRisk 2017 [37] dataset. Ακόμα θα μπορούσε να γίνει χρήση του Reddit API ή ακόμα καλύτερα του pushshift API [91] το οποίο είναι δημιουργημένο για να επιστρέφει δεδομένα μεγάλου όγκου από το Reddit, προσφέροντας περισσότερες λειτουργίες από το επίσημο API του Reddit. Ακόμη θα μπορούσε να γίνει χρήση και άλλων μέσων κοινωνικής δικτύωσης, όπως το Facebook και το Twitter. Ο λόγος που δεν έγινε αφορά ότι το Reddit προσφέρει μεγαλύτερη ευκολία στους χρήστες να εκφράσουν τα συναισθήματα τους με όσες λέξεις θέλουν (σε αντίθεση με το Twitter που το περιορίζει σε λίγους χαρακτήρες) και ταυτόχρονα επειδή αποτελείται από πολλά φόρα συζητήσεων μπορούμε ευκολότερα να αποσπάσουμε δημοσιεύσεις που αφορούν την κατάθλιψη ή κάποια άλλη ψυχική ασθένεια. Ακόμη επιλέχθηκε αυτό το σύνολο δεδομένων διότι είναι πιο εύκολο προσβάσιμο μιας και τα υπόλοιπα είτε απαιτούν έγκριση από τους δημιουργούς τους, είτε έχουν μεγάλο μέγεθος για τους πόρους που έχουμε.

Πίνακας 4.1: Δείγμα του συνόλου δεδομένων.

Text	Label
He agreed. I told him a million times, when we were together and after we broke up,that loyalty is my 1 in relationship, period. I hate cheaters more than anything in theworld. He knew that from the very beginning. Yet recently found out that I was cheatedon for more than 6 months before we broke up.	minimum
Well, the warm season is here which means higher levels of anxiety for me.Thing is, Im so used to this pattern that I donteven care anymore. Ive been through the worse of anxiety attacksand now its just another coh here we goagaind kind of mood. I like the warm weather and whatnot buttheres too much activity going on since people enjoy going out.I dont blame them.	minimum
My father is an alcoholic, and he has been for about 34 years. My mom has had todeal with this the entire time, and lately, he has been having parties at their houseand drinking up until 3-4am almost every weekend. He drinks everyday, and mymother usually comes to me to vent about it. Ive honestlylost my patience because I feel she should do something about it, but shewont. Shes afraid to leave becausehe says hell accuse her of abandoning the home.	minimum
Fast foward almost a year later.. My mom begs me to let stepfather move backin with us (Im 16, paying a lot of the bills while my mom parties, leaving meto watch my 2 little brothers.). Well, I say f*ck it because she'll probably do itanyway. Two months after he moves in, I was raped by the person I wasdating at the time.. I was a virgin. I cried and begged him to stop.	mild
Συνεχίζεται παρακάτω . . .	

Πίνακας 4.1: Δείγμα του συνόλου δεδομένων.
(Συνέχεια)

Συνέχεια πίνακα	
Text	Label
Sorry bad English Since almost 2 weeks now I've cut myself off every social interaction because I've been confronted to doubting and all that shit and I thought that was a necessary measure to take to reborn (through introspection, my whole life since now is involved) and stop burst out of rage and being depressed from nothing And now I keep myself at the strict minimum of social interaction, 10 days that my daily conversation is limited at some words My mental health is decreasing I'm starting to struggle breath by the mouth, now I breath and speak with the nose (probably because I don't open it enough)	mild
Good day. I am from the Philippines and I was recently diagnosed with Depression and Anxiety almost a month ago (July 10th). Since then it has been a very challenging and difficult time for me and my family. Work required me to secure a fit to work order before I can get back to work. My medical insurance does not cover mental health conditions and it took me over 3 weeks to secure one.	mild
this has been happening for the longest time, and it's very hard to describe so bear with me. once in a while when i'm trying to sleep, suddenly my racing mind starts experiencing opposites at the same time. the most terrifying one is feeling like the world is expanding infinitely but also crushingly contracting in on itself. then there's also times when i focus so much on the silence in the environment that i end up feeling like the room sounds extremely loud. and a more physical one is feeling hot and cold at the same time, which prevents me from being comfortable enough to fall asleep.	moderate
But, for now, and until I'm on the proper road to recovery, I just feel noneed to have sex and it worries me more than anything. He says sex is natural and in-the-moment, but I find myself having to plan everything down to what to say and do in advance. Sex just doesn't come naturally to me. This has been posing a major threat to my sexual identity, and I fear if I initiate tomorrow I'll just burst into tears in the middle of it all. I'm so, so scared.	moderate
But I have been on about a dozen times, and have been hung up on nine. It is always the same issue, sadly. I get depressed, it is late, I cannot sleep or stop thinking about past trauma, and I would just like to talk to someone. But the bulk of them have hung up, some almost immediately, and some after longer. It is the longer ones that hurt more.	moderate
Συνεχίζεται παρακάτω . . .	

Πίνακας 4.1: Δείγμα του συνόλου δεδομένων.
(Συνέχεια)

Συνέχεια πίνακα	
Text	Label
He loves them and treats them like gold. I love the type of father he is but I hate the way he treats me. I have no one to open up to or go to if I even tried to leave him. I'm tired of him abusing me mentally and physically. He has cheated on me several times and each time it knocks my self-esteem lower and lower.	severe
I've been in therapy (and will go back soon now) but I've never had to practice harm reduction around self-harm—I need practical advice. I know that because I've never felt so out of control than last night when I was hitting myself. TL;DR I got drunk last night, destroyed my room, and hit myself on the head hard enough to give myself welts for the first time in my adult life. I need advice on not doing this self-harm crap.	severe
Where do I go with this kind of shit? I'm not violent or suicidal, so the hospital is a no-go. But I don't really know where you seek help outside the ER. I dunno the process, how you set up organized appointments, how you pay (got insurance however), don't know any of it. All I know is the shit I've seen really affects me to the point where I have a home and a job yet still feel so disconnected from normal society.	severe

4.1.2 Επαύξηση Δεδομένων

Η επαύξηση κειμένου είναι μια ισχυρή στρατηγική που χρησιμοποιείται για την βελτίωση του μεγέθους και της ποικιλομορφίας ενός συνόλου δεδομένων. Στο πλαίσιο αυτής της μελέτης, η αύξηση ήταν απαραίτητη για τη διασφάλιση μιας ισορροπημένης αναπαράστασης των διαφορετικών επιπέδων σοβαρότητας της κατάθλιψης και για την ενίσχυση της ευρωστίας των μοντέλων μηχανικής μάθησης που εκπαιδεύτηκαν στο σύνολο δεδομένων.

Η βιβλιοθήκη `nlpaug` επιλέχθηκε για τη διαδικασία αύξησης, μια επιλογή που επηρεάστηκε από την ευρεία χρήση και την αποδεδειγμένη αποτελεσματικότητά της στη βελτίωση των συνόλων δεδομένων κειμένου. Χρησιμοποιήθηκαν δύο συγκεκριμένες μέθοδοι από αυτή τη βιβλιοθήκη: το `BERT Substitute` και το `BERT Random Insert`. Η μέθοδος `BERT Substitute` αξιοποιεί το μοντέλο `BERT` για να αναγνωρίσει και να αντικαταστήσει λέξεις στο κείμενο με άλλες σημασιολογικά παρόμοιες. Αντίθετα, η μέθοδος `BERT Random Insert` εισάγει τυχαία λέξεις που σχετίζονται με σημασιολογία στο κείμενο, αυξάνοντας το μήκος και την ποικιλομορφία του.

Οι τεχνικές επαύξησης εμπνεύστηκαν από την εργασία των Wei και Zou [83], η οποία υπογραμμίζει τη σημασία απλών αλλά αποτελεσματικών στρατηγικών αύξησης στις εργασίες ταξινόμησης κειμένου.

Ο αντίκτυπος της διαδικασίας αύξησης στο σύνολο δεδομένων ήταν αξιοσημείωτος. Το αρχικό σύνολο δεδομένων, που αποτελείται από 3553 σειρές, επεκτάθηκε

σε 4353 σειρές μετά την αύξηση, σημειώνοντας αύξηση 800 καταχωρήσεων. Αυτή η αύξηση εξασφάλισε μια πιο ισορροπημένη αναπαράσταση των διαφορετικών επιπέδων σοβαρότητας. Μετά την αύξηση, η κατανομή των ετικετών σοβαρότητας ήταν η εξής:

- **Minimum:** Μείωση από 72,68% σε 61,53%.
- **Mild:** Αύξηση από 11,16% σε 14,24%.
- **Moderate:** Αύξηση από 8,21% σε 12,22%.
- **Severe:** Αύξηση από 7,96% σε 12,01%.

Το σκεπτικό πίσω από τη διαδικασία αύξησης ήταν να διασφαλιστεί ότι κάθε επίπεδο σοβαρότητας αντιπροσωπεύεται επαρκώς, μειώνοντας τις πιθανές προκαταλήψεις στο σύνολο δεδομένων και ενισχύοντας τη γενίκευση των εκπαιδευμένων μοντέλων.

4.1.3 Προεπεξεργασία Δεδομένων

Η προεπεξεργασία κειμένου είναι μια απαραίτητη φάση στη επεξεργασία φυσικής γλώσσας (NLP). Συνεπάγεται με την βελτίωση και την εξυγίανση των ακατέργαστων κειμενικών δεδομένων για τη βελτίωση της ποιότητάς τους, καθιστώντας τα πιο κατάλληλα για μεταγενέστερες αναλύσεις και λειτουργίες. Ακολουθεί μια λεπτομερής ανάλυση και ο Πίνακας 4.2 που περιέχει σύντομα παραδείγματα:

- **Μετατροπή σε πεζά γράμματα:** Η μετατροπή όλου του κειμένου σε πεζά είναι ένα θεμελιώδες βήμα στην προεπεξεργασία κειμένου. Εξασφαλίζει μια συνεπή μορφή σε όλο το σύνολο δεδομένων, εξαλείφοντας πιθανές αποκλίσεις ή προκαταλήψεις που μπορεί να προκύψουν λόγω ποικίλης κεφαλοποίησης. Αυτή η ομοιομορφία απλοποιεί τις επόμενες εργασίες επεξεργασίας κειμένου και διασφαλίζει ότι οι λέξεις αντιμετωπίζονται ισότιμα ανεξάρτητα από την τοποθέτησή τους.
- **Αφαίρεση ετικετών HTML:** Τα δεδομένα κειμένου που προέρχονται από ιστοσελίδες συχνά περιέχουν υπολειπόμενες ετικέτες HTML. Αυτές οι ετικέτες, αν και είναι απαραίτητες για τις ιστοσελίδες, είναι αχρείαστες για την ανάλυση κειμένου. Η κατάργηση αυτών των ετικετών διασφαλίζει ότι η ανάλυση εστιάζει αποκλειστικά στο περιεχόμενο κειμένου.
- **Κατάργηση διεύθυνσης URL:** Δεδομένης της διαδικτυακής φύσης πολλών συνόλων δεδομένων, ειδικά εκείνων που προέρχονται από πλατφόρμες όπως το Reddit, οι διευθύνσεις URL μπορούν συχνά να παρεμβάλλονται στο κείμενο. Αυτές οι διευθύνσεις URL, αν και μπορούν χρησιμοποιηθούν ως επιπλέον τροπικότητες, θεωρούνται θόρυβος στην παρούσα εργασία. Η αφαίρεσή τους διασφαλίζει ένα καθαρότερο σύνολο δεδομένων, χωρίς συνδέσμους που ενδέχεται να παραμορφώσουν την ανάλυση.
- **Κατάργηση Αναφορών Χρήστη:** Οι αναφορές χρηστών, που συνήθως αντιπροσωπεύονται σε μορφές όπως u/user, μπορεί να είναι διαδεδομένες σε διαδικτυακά δεδομένα κειμένου. Η κατάργηση αυτών των αναφορών εξυπηρετεί

έναν διπλό σκοπό: διασφαλίζει το απόρρητο των χρηστών και επίσης καθαρίζει το σύνολο δεδομένων από στοιχεία που δεν συμβάλλουν στην ανάλυση του βασικού περιεχομένου.

- **Αφαίρεση Emoji:** Στη σημερινή ψηφιακή επικοινωνία, τα emoji έχουν γίνει πανταχού παρόντα. Ενώ προσφέρουν έναν συνοπτικό τρόπο μετάδοσης συναισθημάτων, η συμβολική τους φύση μπορεί να δημιουργήσει ασάφεια κατά την επεξεργασία του κειμένου. Η κατάργηση των emoji διασφαλίζει ότι η ανάλυση παραμένει εστιασμένη στο περιεχόμενο του κειμένου, παρέχοντας σαφήνεια και μειώνοντας πιθανές παρερμηνείες.
- **Χειρισμός σημείων στίξης:** Τα σημεία στίξης διαδραματίζουν κεντρικό ρόλο στη μετάδοση της δομής και του τόνου των προτάσεων. Ωστόσο, για την υπολογιστική ανάλυση, είναι σημαντικό να επεξεργαστούν κατάλληλα. Ανάλογα με το γενικό πλαίσιο, τα σημεία στίξης μπορούν να αφαιρεθούν ή να διατηρηθούν.
- **Διόρθωση συντομογραφίας και συνηρημένων λέξεων:** Οι συντομογραφίες και οι συνηρημένες λέξεις είναι κοινές στη γλώσσα, ιδιαιτέρως στην ανεπίσημη. Η επέκτασή τους στην πλήρη μορφή τους εξυπηρετεί δύο σκοπούς: διασφαλίζει μια τυποποιημένη αναπαράσταση λέξεων και φράσεων και βελτιώνει την αναγνωσιμότητα και την κατανόηση του κειμένου, καθιστώντας το πιο προσιτό τόσο για ανθρώπους όσο και για μηχανές.

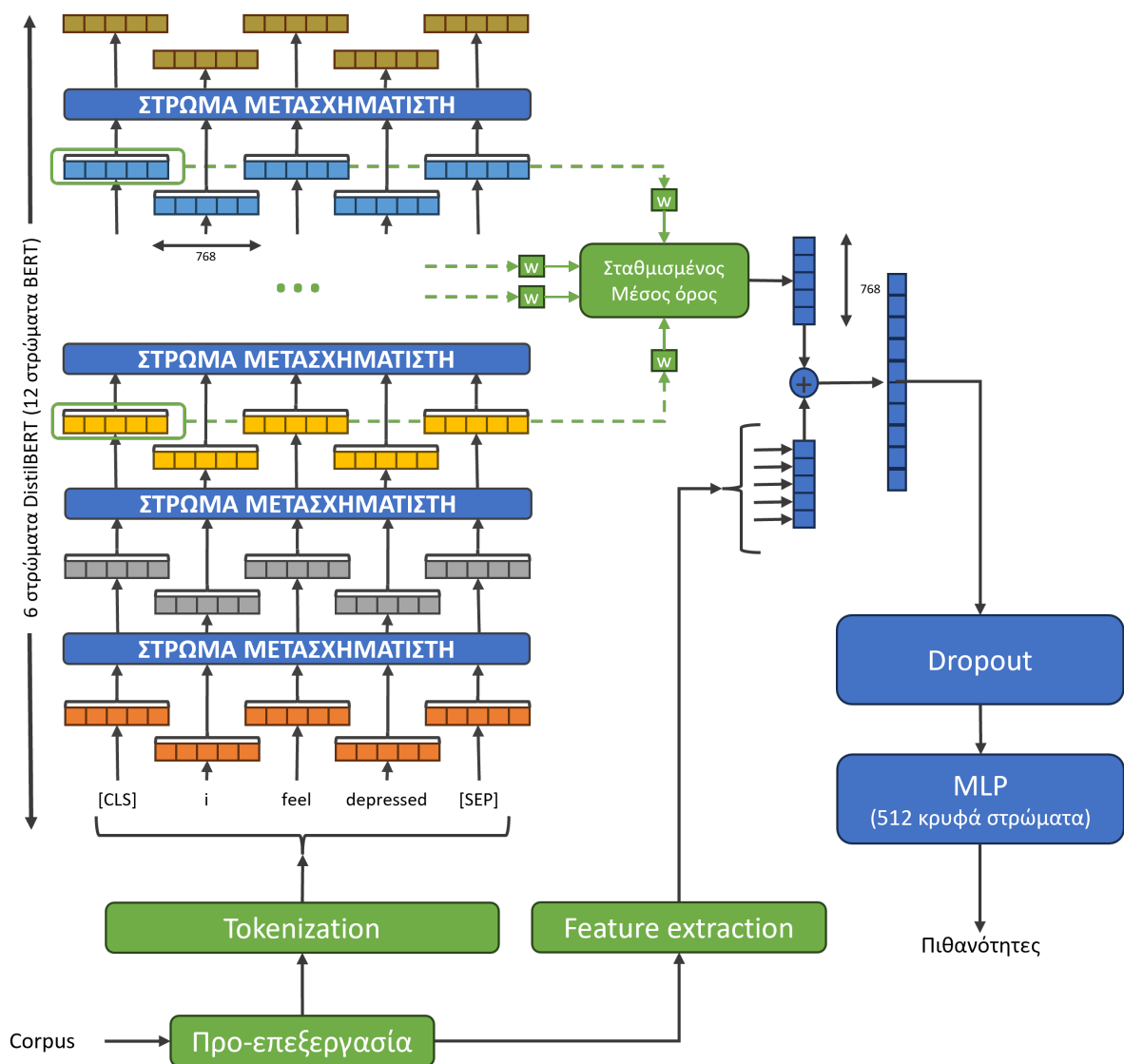
Συνοπτικά, τα βήματα προεπεξεργασίας που υιοθετήθηκαν σε αυτή τη μελέτη επιλέχθηκαν και εκτελέστηκαν σχολαστικά για να διασφαλιστεί η υψηλότερη ποιότητα του συνόλου δεδομένων. Η σωστή προεπεξεργασία όχι μόνο καθαρίζει τα δεδομένα, αλλά θέτει επίσης μια γερή βάση για όλα τα επόμενα στάδια της ανάλυσης, διασφαλίζοντας ότι οι πληροφορίες που προκύπτουν είναι ακριβείς και ουσιαστικές.

Πίνακας 4.2: Τεχνικές Προεπεξεργασίας Κειμένου.

Τεχνική	Περιγραφή	Πριν	Μετά
Μετατροπή σε πεζά	Μετατρέπει το κείμενο σε πεζά.	Text Analysis	text analysis
Αφαίρεση HTML	Αφαιρεί τις ετικέτες HTML από το κείμενο.	<p>Text</p>	Text
Αφαίρεση URL	Αφαιρεί τα URL.	Check out https://example.com	Check out
Αφαίρεση Αναφορών Χρηστών	Αφαιρεί τις αναφορές χρηστών.	Thanks u/user!	Thanks!
Αφαίρεση Emoji	Διαγράφει τα emojis για σαφήνεια.	I'm happy :-)	I'm happy
Χειρισμός Στίξης	Επεξεργάζεται τα σημεία στίξης.	Hello, world!	Hello world
Διόρθωση Συντομοφύσεων	Επεκτείνει τις λέξεις στις πλήρεις τους μορφές.	It's a sample.	It is a sample.

4.2 Αρχιτεκτονική Μοντέλου

Η ενότητα αυτή αποτελεί μια καθοριστική πτυχή της διαδικασίας ανάπτυξης του προτεινόμενου μοντέλου νευρωνικού μας. Στόχος της είναι να εξετάσει και να παρουσιάσει εκτενώς τις διάφορες φάσεις που απαιτούνται για την επιτυχή εφαρμογή του μοντέλου αυτού. Το μοντέλο που προτείνουμε παρουσιάζεται στο Διάγραμμα 4.1.



Σχήμα 4.1: Αναπαράσταση της προτεινόμενης μεθοδολογίας.

4.2.1 Εξαγωγή Χαρακτηριστικών

Η εξαγωγή χαρακτηριστικών είναι ένα θεμελιώδες βήμα στον τομέα της μηχανικής μάθησης και της επεξεργασίας φυσικής γλώσσας (NLP). Περιλαμβάνει τον μετασχηματισμό των δεδομένων κειμένου σε μια μορφή που είναι ταυτόχρονα δομημένη και ποσοτικοποιήσιμη, καθιστώντας την κατάλληλη για εισαγωγή σε αλγόριθμους μηχανικής μάθησης. Με τον απόσταγμα των λεπτομερειών σε ένα σύνολο διακριτών χαρακτηριστικών, μπορούμε να αποτυπώσουμε αποτελεσματικά τα εγγενή του μοτίβα, διευκολύνοντας έτσι την υπολογιστική ανάλυση.

Για αυτή τη μελέτη, χρησιμοποιήθηκαν οι παρακάτω τεχνικές εξαγωγής χαρακτηριστικών:

- **Ανάλυση Συναισθήματος με το EmoRoBERTa:** Το μοντέλο arpanghoshal/EmoRoBERTa [92] επιλέχθηκε για την ικανότητά του στη διεξαγωγή λεπτομερούς ανάλυσης συναισθήματος. Αυτό το μοντέλο έχει εκπαιδευτεί στο σύνολο δεδομένων GoEmotions, μια περιεκτική συλλογή που ετικετοποιεί τα σχόλια του

Reddit με 28 διακριτά συναισθήματα. Τα συναισθήματα αυτά κυμαίνονται από θαυμασμό και διασκέδαση έως θυμό και λύπη, μεταξύ άλλων. Το μοντέλο EmoRoBERTa είναι μια επανάληψη της αρχιτεκτονικής RoBERTa, μια βελτιωμένη έκδοση του BERT που επωφελείται από την εκπαίδευση σε ένα εκτεταμένο σύνολο δεδομένων και βελτιστοποιημένες υπερπαραμέτρους. Χρησιμοποιώντας αυτό το μοντέλο, μπορούμε να εισέλθουμε βαθιά στο κείμενο για να αναγνωρίσουμε και να κατανοήσουμε τα συγκεκριμένα συναισθήματα που μεταφέρει, παρέχοντας μια πολυδιάστατη άποψη του συναισθηματικού του τοπίου.

- **Ανάλυση Συναισθήματος με το Twitter RoBERTa:** Το μοντέλο cardiffnlp/twitter-roberta-base-sentiment-latest [93] [94] ξεχωρίζει για την ικανότητά του στην ανάλυση συναισθήματος. Αυτό το μοντέλο έχει εκπαιδευτεί σε ένα τεράστιο σώμα περίπου 124 εκατομμυρίων tweets που καλύπτουν την περίοδο από τον Ιανουάριο 2018 έως τον Δεκέμβριο 2021. Έχει περαιτέρω βελτιστοποιηθεί για ανάλυση συναισθήματος χρησιμοποιώντας το διάσημο σημείο αναφοράς TweetEval. Το μοντέλο κατηγοριοποιεί τα συναισθήματα σε τρεις κύριες κλάσεις: Αρνητικό, Ουδέτερο και Θετικό. Δεδομένης της ένταξής του στο πλαίσιο TweetNLP και της εξειδίκευσής του στην ανάλυση αγγλικών tweets, προσφέρει μια αξιόπιστη λύση για τον προσδιορισμό του κυρίαρχου συναισθήματος του κειμενικού περιεχομένου.
- **Αναφορές Φαρμάκων:** Αναγνωρίζοντας τη σημασία του ιατρικού πλαισίου στα κειμενικά δεδομένα, σχεδιάστηκε μια προσαρμοσμένη μέθοδος εξαγωγής χαρακτηριστικών. Αυτή η μέθοδος σαρώνει προσεκτικά το κείμενο για αναφορές σε οποιοδήποτε φάρμακο που αναφέρεται στο drugs.com (Πίνακας 4.4). Η αναγνώριση συγκεκριμένων αναφορών φαρμάκων μπορεί να παράσχει πολύτιμες πληροφορίες για τις πιθανές θεραπείες που σχετίζονται με τον συγγραφέα, προσθέτοντας ένα ακόμη επίπεδο βάθους στην ανάλυση.

Περίληπτικά, η εξαγωγή χαρακτηριστικών λειτουργεί ως η κρίσιμη γέφυρα που συνδέει αρμονικά τα κειμενικά δεδομένα με τους αλγόριθμους μηχανικής μάθησης. Μετατρέποντας με επιδεξιότητα το κείμενο σε μια δομημένη και αναλύσιμη μορφή, εξοπλίζει τους αλγόριθμους με τη δυνατότητα να αναγνωρίζουν μοτίβα, να καθιερώνουν σχέσεις και να αντλούν σημαντικές πληροφορίες, εξασφαλίζοντας έτσι την επιτυχία των επόμενων αναλυτικών διαδικασιών.

4.2.2 Συγχώνευση Χαρακτηριστικών με τον Μετασχηματιστή

Το μοντέλο που προτείνουμε αντιπροσωπεύει μια καινοτόμα προσέγγιση στην αρχιτεκτονική των νευρωνικών δικτύων για την ταξινόμηση δημοσιεύσεων κατάθλιψης, ενσωματώνοντας τις δυνατότητες των μοντέλων μετασχηματιστών με τα ειδικά χαρακτηριστικά (features) των δεδομένων. Αυτός ο υβριδικός σχεδιασμός βασίζεται στην ιδέα της μεγιστοποίησης των οφελών των προ-εκπαιδευμένων μετασχηματιστικών μοντέλων, ενώ ταυτόχρονα εκμεταλλεύεται επιπλέον χαρακτηριστικά που μπορεί να προσφέρουν περισσότερες πληροφορίες που δεν είναι απαραίτητο ότι ο μετασχηματιστής θα μπορέσει να συλλάβει.

Βασίζεται σε ένα υβριδικό μοντέλο [95] (Διάγραμμα 4.1) που συνδυάζει ένα μοντέλο transformer, συγκεκριμένα το μοντέλο DistilBERT base uncased, με ένα Πολυεπίπεδο Perceptron (MLP). Το μοντέλο DistilBERT base uncased [35] είναι ένα προεκ-

Πίνακας 4.3: Παράδειγμα του συνόλου δεδομένων.

label	text	fear	sadness	disapp.	positive	annoyance	love	negative	mentions med.
minimum	these were not the only punishments though sometimes he would make us choose a favorite toy or plushie to be thrown out if wed made him mad and one time when he caught me rescuing ...	5.82E-04	2.21E-02	4.74E-03	1.10E-02	7.09E-03	1.32E-03	8.52E-01	FALSE
	i am not doing this for people to make money but to help those struggling to feed themselves andor family my days of that are over if any supermarkets or stores offer similar promotions in ...	4.09E-04	5.46E-02	3.97E-03	9.58E-02	2.89E-04	9.15E-04	3.52E-01	FALSE
	my boyfriend constantly states that we are no longer in a relationship and i need to be constantly working towards showing him that he will want to be with me again i do 90 of the cooking ...	3.99E-05	9.63E-04	1.28E-02	8.02E-02	5.13E-03	1.63E-04	4.82E-01	FALSE
mild	they have lives outside the group and are often coming from far away to go to this group should i try and be friends with them i feel like they are not going to want to hang out with a dorky 23...	2.98E-05	8.16E-04	9.06E-01	8.73E-03	7.55E-02	2.17E-04	7.85E-01	FALSE
	i am not exaggerating in that i legitimately 100 thought i was going to die i do not think i have ever been so scared i keep a diary and reading back on what i wrote at that time makes me cry at...	5.25E-01	4.08E-01	6.89E-03	2.09E-02	6.61E-04	1.22E-03	8.63E-01	FALSE
	it was the same therapist id been seeing since i was little when next i was first diagnosed as asd at the time termed pdd nos as i did not out fit the diagnostic requirements for aspergers and she ...	2.48E-05	1.64E-04	2.39E-04	5.89E-03	5.33E-05	4.46E-05	8.89E-01	FALSE
moderate	usually it was i literally had to take breathing and exercise therapy thingys in because tonight i felt what i was on pass myself out my damn heart is racing now not many notifications find me else like here so i will not really be too concerned but still ...	1.56E-02	1.16E-02	1.52E-03	3.89E-02	5.03E-03	1.24E-04	7.30E-01	FALSE
	its sick and twisted i have only been getting old nauseas and ire seen getting more dissociative flashbacks my nightmares kept coming crashing back but they are usually mostly of his past so iam afraid to actually see them although my dreams ...	9.73E-01	2.37E-03	2.88E-03	5.81E-03	5.16E-03	1.36E-04	9.13E-01	FALSE
	also the headaches loads more of headaches happening all the time i ' ' m almost so done i hate this almost getting as bad as my brain constantly constantly telling me me i already ' m a pos however anxiety itself is fun	1.07E-03	3.27E-03	1.14E-03	1.36E-02	3.98E-01	1.10E-03	9.22E-01	FALSE
severe	i feel so worthless on those days i guess i just needed to vent to a community that will hopefully understand and see if anyone else had any similar experiences i had ptsd for almost a year before going on benzos but it was not until i was put on benzos and ...	2.78E-05	2.44E-04	5.70E-04	1.04E-02	4.77E-03	3.12E-04	8.82E-01	TRUE
	i never felt so uncomfortable and scared other than the night he did put his hands on me and it was exact same feeling today that was the scariest part it felt like that night never ended and i was just at the peak of emotion i know most of you will say 'you need ...	9.76E-01	2.79E-04	4.43E-04	1.05E-02	1.21E-03	6.86E-05	8.98E-01	FALSE
	i am fat unattractive unmotivated and best of all six figures in student loan debt and not even graduated yet ill never be able to afford living on my own so there goes the last little bit of hope for dating though relationships do not seem great to me anyway i ...	1.85E-03	8.82E-03	2.41E-03	7.25E-03	8.61E-03	9.54E-04	9.42E-01	FALSE

παιδευμένο μοντέλο transformer που έχει εκπαιδευτεί σε ένα μεγάλο σώμα αγγλικών δεδομένων (3300 εκατομμύρια λέξεις) με αυτο-εποπτευόμενο τρόπο. Αυτό το μοντέλο είναι uncased, πράγμα που σημαίνει ότι δεν διακρίνει μεταξύ κεφαλαίων και πεζών αγγλικών γραμμάτων.

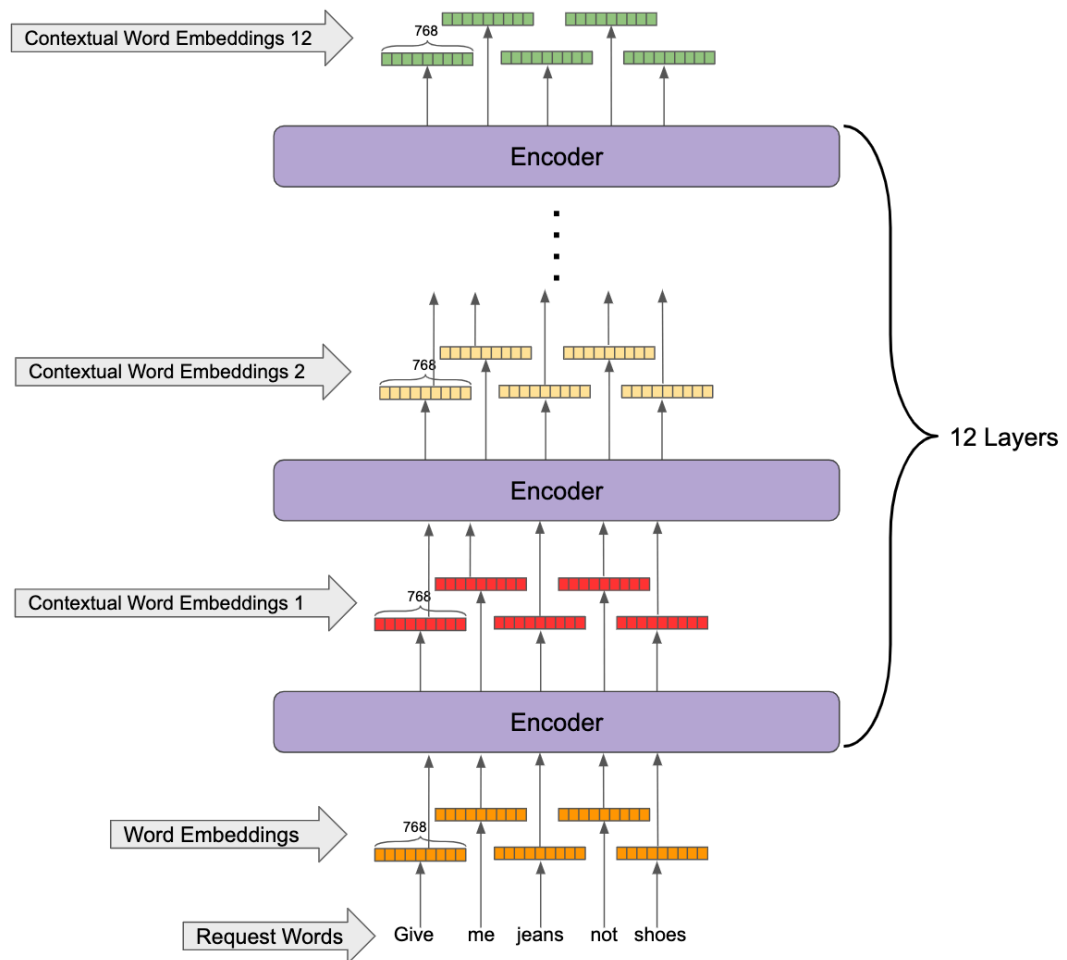
Στην αρχιτεκτονική μας, χρησιμοποιούμε τις κρυφές καταστάσεις από τα τελευταία τέσσερα επίπεδα του transformer. Αυτά τα επίπεδα επιλέγονται επειδή τα βαθύ-

Πίνακας 4.4: Φάρμακα που χρησιμοποιούνται σε περιπτώσεις κατάθλιψης.

Φάρμακα για την κατάθλιψη		
bupropion	quetiapine	Aventyl Hydrochloride
Cymbalta	Deplin	clomipramine
sertraline	nortriptyline	Emsam
Zoloft	Zyprexa	fluoxetine / olanzapine
Lexapro	Budeprion SR	ketamine
citalopram	desvenlafaxine	lisdexamfetamine
fluoxetine	Desyrel	Norpramin
Prozac	Forfivo XL	Pamelor
Wellbutrin XL	methylphenidate	Pexeva
Celexa	Aplenzin	phenelzine
escitalopram	Desyrel Dividose	Tofranil
venlafaxine	doxepin	tranylcypromine
Trintellix	Fetzima	vilazodone
duloxetine	lithium	Zyprexa Zydis
Abilify	olanzapine	amitriptyline / perphenazine
Effexor XR	risperidone	armodafinil
Rexulti	desipramine	atomoxetine
vortioxetine	Irenka	brexpiprazole
trazodone	modafinil	isocarboxazid
mirtazapine	nefazodone	levomilnacipran
Pristiq	niacin	Marplan
Effexor	fluvoxamine	paliperidone
Paxil	imipramine	protriptyline
Wellbutrin SR	l-methylfolate	selegiline
Remeron	Methylin ER	trimipramine
paroxetine	Nardil	Vivactil
amitriptyline	Parnate	XaQuil XR
lamotrigine	Paxil CR	esketamine
Viibryd	Remeron SolTab	L-Methylfolate Forte
alprazolam	Symbyax	Spravato
tramadol	Abilify MyCite	Surmontil
Seroquel XR	amitriptyline / chlordiazepoxide	thyroid desiccated
aripiprazole	amoxapine	

τερα επίπεδα στο μοντέλο BERT αποτυπώνουν πιο περίπλοκες σημασιολογικές πληροφορίες, ενώ σύμφωνα με τους δημιουργούς του BERT τα τέσσερα επίπεδα είναι ένας ικανοποιητικός αριθμός στρωμάτων. Υπολογίζουμε έναν σταθμισμένο μέσο όρο αυτών των κρυφών καταστάσεων, όπου τα βάρη εκπαιδεύονται και αντιπροσωπεύουν πόσο σημαντικές είναι διανυσματικές αναπαραστάσεις των λέξεων σε κάθε επίπεδο για την ταξινόμηση. Αυτός δηλαδή επιτρέπει στο μοντέλο να αποφασίσει τη σημαντικότητα της εξόδου από κάθε επίπεδο. Το αποτέλεσμα αυτής της διαδικασίας είναι ένας τανυστής που αντιπροσωπεύει τις σταθμισμένες μέσες κρυφές

καταστάσεις του μοντέλου transformer. Συγκεκριμένα, χρησιμοποιούμε τα διακριτά CLS, τα οποία είναι τα πρώτα tokens σε κάθε ακολουθία και περιέχουν μια περίληψη του νοήματος της ακολουθίας. Το transformer δέχεται μια είσοδο 512 tokens και έχει κρυφό μέγεθος 768 (Διάγραμμα 4.1 4.2). Τα διακριτά CLS είναι ιδιαίτερα σημαντικά καθώς παρέχουν μια συνοπτική αναπαράσταση ολόκληρης της εισαγόμενης ακολουθίας, καθιστώντας τα ιδανικά για εργασίες όπως η ταξινόμηση.



Σχήμα 4.2: Παράδειγμα στρωμάτων BERT [95].

Το μέσο διακριτό CLS στη συνέχεια συνδυάζεται με τα επιπλέον χαρακτηριστικά για να σχηματίσει την είσοδο για το MLP. Πριν περάσει η είσοδος στο MLP, υπόκειται σε ένα επίπεδο απόρριψης (dropout layer) με ένα ποσοστό απόρριψης 0,1 για την κοινωνικοποίησή. Αυτό βοηθά στην πρόληψη της υπερπροσαρμογής με το να θέτει τυχαία ένα μέρος των εισαγόμενων μονάδων σε 0 σε κάθε ενημέρωση κατά τη διάρκεια του χρόνου εκπαίδευσης. Η απόρριψη (dropout) είναι μια ευρέως χρησιμοποιούμενη τεχνική στη βαθιά μάθηση για να διασφαλίσει ότι το μοντέλο γενικεύει καλά σε δεδομένα. Εισάγοντας τυχαιότητα κατά τη διάρκεια της εκπαίδευσης, το μοντέλο γίνεται πιο ανθεκτικό και λιγότερο πιθανό να βασίζεται σε συγκεκριμένα χαρακτηριστικά των δεδομένων εκπαίδευσης.

Το MLP είναι ένα απλό νευρωνικό δίκτυο προς τα εμπρός που αποτελείται από ένα γραμμικό επίπεδο, μια συνάρτηση ενεργοποίησης ReLU και ένα άλλο γραμμικό επίπεδο. Η είσοδος στο MLP είναι ένας συνδυασμός των σταθμισμένων μέσω κρυ-

φών καταστάσεων από το μοντέλο DistilBERT base uncased και τα επιπλέον χαρακτηριστικά (features). Το MLP έχει κρυφό μέγεθος 512, και η διάσταση εξόδου του MLP καθορίζεται από τον αριθμό των ετικετών, που στο εν λόγω έργο σε αυτή την περίπτωση είναι 4. Η επιλογή του ReLU ως συνάρτηση ενεργοποίησης διασφαλίζει ότι το δίκτυο μπορεί να αποτυπώσει μη γραμμικές σχέσεις στα δεδομένα, ενισχύοντας τις προβλεπτικές του δυνατότητες. Η έξοδος του MLP είναι η τελική έξοδος του υβριδικού μοντέλου, τα logits.

Τα logits είναι, στην ουσία, τα αποτελέσματα πριν την εφαρμογή της συνάρτησης ενεργοποίησης και μπορούν να θεωρηθούν ως βαθμοί πιθανότητας. Για να μετατραπούν αυτά τα logits σε πραγματικές πιθανότητες που αθροίζονται στο 1, χρησιμοποιούμε τη συνάρτηση softmax. Η συνάρτηση softmax λαμβάνει τα logits και τα "συμπιέζει" σε ένα εύρος μεταξύ 0 και 1, καθιστώντας τα ως πιθανότητες. Στη συνέχεια, για να προβλέψουμε την τελική κατηγορία της εισόδου, χρησιμοποιούμε τη συνάρτηση argmax, η οποία επιλέγει την κατηγορία με την υψηλότερη πιθανότητα από τις εξόδους της συνάρτησης softmax. Αυτή η διαδικασία εξασφαλίζει ότι το μοντέλο μας παράγει μια σαφή και ακριβή πρόβλεψη για κάθε δείγμα εισόδου.

4.3 Πειραματική Διάταξη

4.3.1 Εργαστηριακό Περιβάλλον

Τα πειράματά μας διεξήχθησαν κυρίως χρησιμοποιώντας Python. Συγκεκριμένα χρησιμοποιήσαμε Python notebooks με kernel Python 3.10.12, τα οποία διασφάλισαν την αναπαραγωγικότητα, την ευκολία κοινής χρήσης και άμεσων αποτελεσμάτων. Για υπολογιστικούς πόρους, καταφύγαμε σε δωρεάν cloud πλατφόρμες, συγκεκριμένα στο Google Colab εξοπλισμένο με ένα T4 GPU και στο Kaggle με έναν P100 GPU ή 2 T4 GPUs.

Σε ό,τι αφορά τις βιβλιοθήκες της Python, χρησιμοποιήσαμε το pandas για τη διαχείριση και την ανάλυση δεδομένων. Τα μοντέλα βαθιάς μάθησης κατασκευάστηκαν χρησιμοποιώντας το PyTorch [96]. Για την αξιολόγηση του μοντέλου, χρησιμοποιήσαμε μετρήσεις από τη βιβλιοθήκη scikit-learn [97]. Η προεπεξεργασία των δεδομένων κειμένου ήταν ένα κρίσιμο βήμα, και για αυτό, χρησιμοποιήσαμε τη βιβλιοθήκη emoji για τη διαχείριση των emojis και το spacy [98] για ένα προταρχικό tokenization, stemming και άλλα γλωσσικά χαρακτηριστικά. Ακόμα χρησιμοποιήθηκε η βιβλιοθήκη contractions για να επεκτείνει διάφορες συντομογραφίες. Η βιβλιοθήκη huggingface transformers [99] ήταν ανεκτίμητη, παρέχοντάς μας πρόσβαση σε προ εκπαιδευμένα μοντέλα κορυφαίας τεχνολογίας και τους αντίστοιχους tokenizers τους. Για να ενισχύσουμε την ποικιλομορφία των δεδομένων εκπαίδευσής μας, χρησιμοποιήσαμε το nlraug [100] για την επαύξηση του κειμένου.

Οι παράμετροι εκπαίδευσής μας ορίστηκαν με μέγεθος παρτίδας (batch size) 8, και διαθέσαμε το 80% των δεδομένων για εκπαίδευση, διατηρώντας το υπόλοιπο 20% για επικύρωση και δοκιμές. Τα μοντέλα υποβλήθηκαν σε εκπαίδευση για συνολικά 11 εποχές. Ορίσαμε συγκεκριμένους ρυθμούς μάθησης για διαφορετικά στοιχεία: ο ταξινομητής είχε έναν ρυθμό $1e - 5$, ενώ όλα τα άλλα στοιχεία ορίστηκαν στο $1e - 3$. Η διαδικασία βελτιστοποίησης καθοδηγήθηκε από την συνάρτηση Cross-

EntropyLoss, και χρησιμοποιήσαμε τον βελτιστοποιητή Adam με έναν ρυθμό μάθησης $1e - 5$.

4.3.2 Μετρικές αξιολόγησης

Για την αξιολόγηση των μοντέλων μας, χρησιμοποιήσαμε τη βαρυσταθμισμένη ακρίβεια (weighted precision), τη βαρυσταθμισμένη ανάκληση (weighted recall) και το βαρυσταθμισμένο σκορ F1 (weighted F1 score). Αυτές οι μετρικές είναι ιδιαίτερα σημαντικές σε προβλήματα ταξινόμησης όπου οι κατηγορίες δεν είναι ισορροπημένες, δηλαδή όταν ορισμένες κατηγορίες εμφανίζονται πιο συχνά από άλλες. Αυτές οι μετρήσεις μας επέτρεψαν να αξιολογήσουμε την απόδοση του μοντέλου μας με μεγαλύτερη ακρίβεια, λαμβάνοντας υπόψη την ισορροπία μεταξύ της ακρίβειας και της ανάκλησης και την ικανότητα του μοντέλου να ταξινομεί σωστά τις διάφορες κατηγορίες. Τέλος, το βαρυσταθμισμένο σκορ F1 αποτελεί μια συνδυαστική μετρική που παρέχει μια ολοκληρωμένη εικόνα της απόδοσης, συνδυάζοντας τόσο την ακρίβεια όσο και την ανάκληση σε έναν μόνο δείκτη.

4.3.3 Μοντέλα αναφοράς

Στην διπλωματική μας, συγκρίνουμε την προτεινόμενη μεθοδολογία μας με διάφορα μοντέλα-μέτρα σύγκρισης. Το πρώτο μοντέλο είναι η προσέγγιση που παρουσιάζεται από τους Ilias et al. [101]. Αυτό το μοντέλο είναι μια προσέγγιση βασισμένη σε μετασχηματιστές που ενσωματώνει επιπρόσθετες γλωσσικές πληροφορίες στα μοντέλα BERT και MentalBERT. Η μέθοδος ξεκινάει με την εξαγωγή διαφόρων γλωσσικών χαρακτηριστικών, όπως το NRC Sentiment Lexicon, χαρακτηριστικά που προκύπτουν από τα Latent Dirichlet Allocation (LDA), Top2vec και Linguistic Inquiry and Word Count (LIWC). Τα χαρακτηριστικά αυτά προβάλλονται στην ίδια διάσταση με τις εξόδους των μοντέλων transformer. Οι συγγραφείς συνδυάζουν τις αναπαραστάσεις που προκύπτουν από το BERT (ή MentalBERT) και τη γλωσσική πληροφορία, εφαρμόζοντας ένα Multimodal Adaptation Gate για να ελέγχουν τη σημασία κάθε αναπαραστάσης. Οι συνδυασμένες αναπαραστάσεις στη συνέχεια περνούν από ένα μοντέλο BERT (ή MentalBERT), με το διακριτικό σύμβολο ταξινόμησης [CLS] να τροφοδοτείται σε Dense layers για να παράγει την τελική πρόβλεψη. Για να αποτρέψουν τα μοντέλα από την υπερεκπαίδευση, οι συγγραφείς χρησιμοποιούν επίπεδα εξομάλυνσης ετικετών.

Το δεύτερο βασικό μοντέλο προέρχεται από το άρθρο των Yang et al. [102]. Σε αυτήν την προσέγγιση, οι συγγραφείς χρησιμοποίησαν τα BERT και MentalBERT για την ανίχνευση της κατάθλιψης.

Οι παραπάνω συγγραφείς παρέχουν ένα ολοκληρωμένο σύνολο αποτελεσμάτων για τις προσεγγίσεις τους, τόσο με όσο και χωρίς την εξομάλυνση ετικετών. Αυτά τα αποτελέσματα χρησιμεύουν ως σημείο αναφοράς για την προτεινόμενη μεθοδολογία μας, επιτρέποντάς μας να αξιολογήσουμε την αποτελεσματικότητά του μοντέλου σε σχέση με καθιερωμένες μεθόδους στο χώρο της αναγνώρισης κατάθλιψης.

4.4 Μέλετη Διαφορετικών Διάταξεων

4.4.1 Κεφαλές Κατηγοριοποίησης

Στην παρούσα έρευνα θα μελετήσουμε εναλλαγές στην κεφαλή ταξινόμησης του μοντέλου. Συγκεκριμένα θα αξιολογήσουμε την απόδοση των παρακάτω τεσσάρων διακριτών κεφαλών ταξινόμησης:

- Πολυεπίπεδο Perceptron (MLP) με κρυφό μέγεθος 512 στρώματα.
- Δίκτυο Μακράς Βραχύχρονης Μνήμης (LSTM) με κρυφό μέγεθος 512 στρώματα.
- Πολυτροπικές Πύλες (MM-Gate) με κρυφό μέγεθος 512 στρώματα.
- Πύλες διασταυρούμενης προσοχής (MM-Att) με κρυφό μέγεθος 512 στρώματα.

Οι κεφαλές MM-Gate και MM-Att εμπνεύστηκαν από το έργο των Sánchez Villegas και Aletras [103] αλλά έχουν χρησιμοποιηθεί σε πολλά νέα έργα με τεχνολογίες αιχμής [104] [7].

4.4.2 Μετασχηματιστές

Ταυτόχρονα η μελέτη μας περιέχει τη σύγκριση της απόδοσης δύο διάσημων μοντέλων μετασχηματιστών, του BERT base uncased και του DistilBERT base uncased. Ο λόγος που χρησιμοποιήσαμε αυτά τα δύο είναι διότι είναι παρεμφερή μοντέλα με την αλλαγή ότι το DistilBERT εκπαιδεύτηκε με λιγότερα δεδομένα και έχει λίγο χαμηλότερη ακρίβεια από το BERT. Εδώ πρέπει να τονίσουμε ότι εκτός του μετασχηματιστή αλλάζει και ο tokenizer του μοντέλου για λόγους συμβατότητας.

4.4.3 Στρώματα του μετασχηματιστή

Στην διπλωματική μας επίσης, αναλάβαμε μια εξερεύνηση στις επιδόσεις διαφορετικών βαθμίδων των στρωμάτων του μετασχηματιστή. Συγκεκριμένα, δοκιμάσαμε μια σειρά βαθμίδων, από 1 έως 6 στρώματα του μετασχηματιστή. Επιπλέον, αξιολογήσαμε αυτές τις αλλαγές με δύο διακριτές επικεφαλίδες ταξινόμησης: MLP και LSTM.

4.4.4 Επαύξηση Δεδομένων

Τέλος θα εξερευνήσουμε την επιρροή της επαύξησης των δεδομένων. Το ίδιο μοντέλο θα το τρέξουμε με τα αρχικά δεδομένα και με δεδομένα που τους εφαρμόσαμε επαύξηση με χρήση BERT μοντέλου.

Κεφάλαιο 5

Αποτελέσματα

5.1 Αποτελέσματα προτεινόμενου μοντέλου

Στη διάρκεια των πειραμάτων μας, εκπαιδεύσαμε και αξιολογήσαμε πολλαπλές διαμορφώσεις των μοντέλων μας. Παρατηρήσαμε ότι το μοντέλο όπου χρησιμοποιεί τα 4 τελευταία (9, 10, 11, 12) στρώματα με κεφαλή ταξινόμησης MLP με 512 κρυφά επίπεδα και DistilBERT base uncased μετασχηματιστή υπερέβη τις άλλες διαμορφώσεις. Αυτό το μοντέλο πέτυχε μια ζυγισμένη ακρίβεια της τάξης του 84,25%, ζυγισμένη ανάκληση της τάξης του 84,16% και ένα ζυγισμένο F1-score της τάξης του 84,10%.

Σε σύγκριση με τα μοντέλα που προτάθηκαν από τους Ilias et al. [101] και τους Yang et al. [102], το καλύτερο μοντέλο μας παρουσίασε ανταγωνιστικά αποτελέσματα.

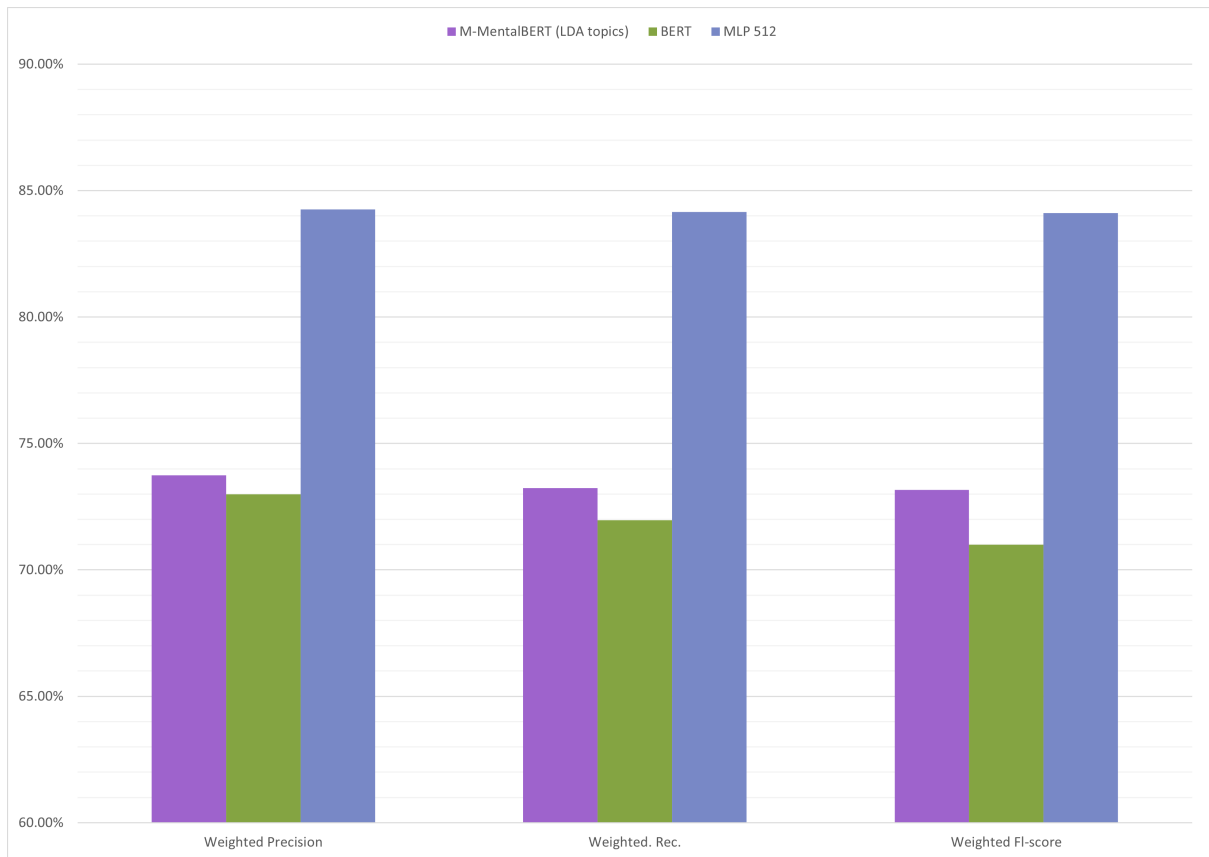
Το καλύτερο μοντέλο των Ilias et al. [101], το M-MentalBERT (LDA topics), είχε ζυγισμένη ακρίβεια 73,74%, ζυγισμένη ανάκληση 73,23% και ζυγισμένο F1-score 73,16%.

Το μοντέλο BERT των Yang et al. [102] είχε ζυγισμένη ακρίβεια 72,99%, ζυγισμένη ανάκληση 71,97% και ζυγισμένο F1-score 71,00%, ενώ το μοντέλο τους MentalBERT είχε ζυγισμένη ακρίβεια 73,35%, ζυγισμένη ανάκληση 70,81% και ζυγισμένο F1-score 71,67%.

Το μοντέλο με τα 4 στρώματα και κεφαλής ταξινόμησης MLP με 512 κρυφά επίπεδα χρησιμοποιώντας το DistilBERT base uncased επιτυγχάνει βελτίωση περίπου 10,94% σε σύγκριση με το καλύτερο βασικό μοντέλο. Μια λεπτομερής σύγκριση της απόδοσης του μοντέλου μας σε σχέση με τα βασικά μοντέλα παρουσιάζεται στον Πίνακα 5.1.

Πίνακας 5.1: Σύγκριση απόδοσης βάσεων αναφοράς και του προτεινόμενου μοντέλου.

Μοντέλο	Weighted Precision (%)	Weighted Recall (%)	Weighted F1-score (%)
Προτεινόμενο Μοντέλο (MLP, 512 hidden layers, DistilBERT base uncased)	84.25%	84.16%	84.10%
M-MentalBERT - LDA topics (Ilias et al.)	73.74%	73.23%	73.16%
BERT (Yang et al.)	72.99%	71.97%	71.00%
MentalBERT (Yang et al.)	73.35%	70.81%	71.67%



Σχήμα 5.1: Διάγραμμα αποτελεσμάτων μεταξύ προτεινόμενου μοντέλου και των βάσεων αναφοράς.

5.2 Αποτελέσματα μελέτης διαφορετικών διατάξεων

5.2.1 Κεφαλές κατηγοριοποίησης

Πίνακας 5.2: Χρήση διαφορετικών κεφαλών ταξινόμησης.

Κεφαλή ταξινόμησης και κρυφό μέγεθος	Weighted Precision (%)	Weighted Recall (%)	Weighted F1-score (%)
MLP 512	84.25%	84.16%	84.10%
LSTM 512	76.92%	75.38%	74.88%
MM-Gate 512	80.43%	80.14%	79.88%
MM-Xatt 512	51.00%	61.65%	54.71%

Το Multi-Layer Perceptron (MLP) παρουσίασε την υψηλότερη απόδοση μεταξύ των κεφαλών ταξινόμησης, επιτυγχάνοντας σταθμισμένη ακρίβεια, ανάκληση και βαθμολογία F1 84,25%, 84,16% και 84,10% αντίστοιχα. Η υπεροχή του MLP σε αυτό το πλαίσιο μπορεί να αποδοθεί στην απλότητα και την αποτελεσματικότητά του. Όντας τροφοδοτικά νευρωνικά δίκτυα, τα MLP στερούνται επαναλαμβανόμενων λειτουργιών ή λειτουργιών πύλης, καθιστώντας τα εύκολα υλοποιήσιμα και υπολογιστικά αποδοτικά. Η ικανότητά τους να καταγράφουν μοτίβα στα δεδομένα είναι επίσης απόδειξη της δύναμης των αρχιτεκτονικών προώθησης τροφοδοσίας.

Τα Long Short-Term Memory (LSTM) είχαν την επόμενη καλύτερη επίδοση. Τα LSTM είναι γνωστά για την ικανότητά τους να διατηρούν μακροπρόθεσμες εξαρτήσεις σε διαδοχικά δεδομένα. Ωστόσο, η σταθμισμένη τους ακρίβεια, η ανάκληση και

το σκορ F1 ήταν ελαφρώς κατώτερα από το MLP, καταγράφοντας 76,92%, 75,38% και 74,88% αντίστοιχα. Αν και είναι πιο πολύπλοκα από τα MLP, η επαναλαμβανόμενη φύση τους μπορεί να μην ήταν απολύτως απαραίτητη για τη συγκεκριμένη εργασία ταξινόμησης, με αποτέλεσμα να μην δικαιολογείται η πρόσθετη υπολογιστική επιβάρυνση.

Ο μηχανισμός MM-Gate πέτυχε αξιοπρεπή αποτελέσματα με σκορ 80,43% (ακρίβεια), 80,14% (ανάκληση) και 79,88% (βαθμολογία F1). Δεδομένου ότι είναι ένας μηχανισμός πύλης όπου οι πύλες μπορούν να βοηθήσουν στο συνδυασμό χαρακτηριστικών ή αναπαραστάσεων, μπορεί να μην οδηγούν πάντα σε ανώτερη απόδοση. Πρέπει να επιτευχθεί ισορροπία μεταξύ πολυπλοκότητας και λειτουργικότητας για βέλτιστα αποτελέσματα.

Απο την άλλη ο μηχανισμός MM-Chart παρουσίασε μια έντονη αντίθεση μεταξύ της ακρίβειας και της ανάκλησής του, με σκορ 51,00% και 61,65% αντίστοιχα. Η σταθμισμένη βαθμολογία F1 ήταν 54,71%. Το σημαντικό χάσμα μεταξύ ακρίβειας και ανάκλησης υποδηλώνει μια πιθανή μεροληψία στις προβλέψεις του, όπου ο μηχανισμός μπορεί να υπερ-προβλέπει ορισμένες κατηγορίες, οδηγώντας σε υψηλότερα ψευδώς θετικά αποτελέσματα.

Ενώ οι μηχανισμοί πύλης είναι νέοι και υπόσχονται βελτιωμένη χωρητικότητα μοντέλου αξιοποιώντας πολύπλοκες αλληλεπιδράσεις, μπορεί να μην είναι πάντα ωφέλιμοι για κάθε εργασία. Αυτή η έρευνα καταδεικνύει τη σημασία της επιλογής μοντέλων για συγκεκριμένη εργασία, όπου απλούστερες αρχιτεκτονικές όπως τα MLP μπορούν να ξεπεράσουν τις πιο σύνθετες, ανάλογα με τη φύση των δεδομένων και την εργασία.

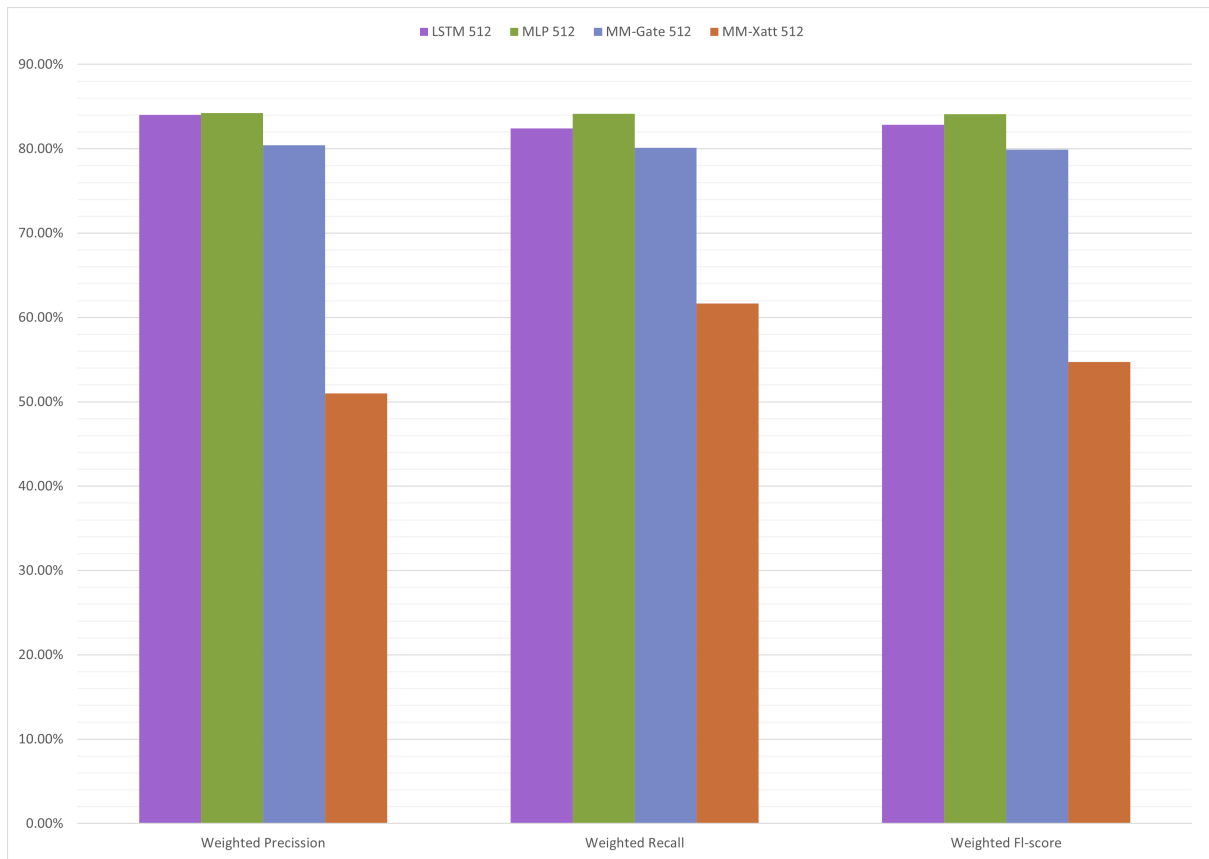
5.2.2 Μετασχηματιστές

Πίνακας 5.3: Χρήση διάφορων μετασχηματιστών.

Κεφαλή ταξινόμησης και κρυφό μέγεθος	models	Weighted Precision (%)	Weighted Recall (%)	Weighted F1-score (%)
MLP 512	BERT base uncased	84.37%	84.27%	84.30%
MLP 512	DistilBERT base uncased	84.25%	84.16%	84.10%
LSTM 512	BERT base uncased	84.01%	82.43%	82.86%
LSTM 512	DistilBERT base uncased	76.92%	75.38%	74.88%

Ο Πίνακας 5.3, παρουσιάζει συναρπαστικά αποτελέσματα, ότι δηλαδή η απόκλιση στην απόδοση μεταξύ αυτών των δύο μοντέλων είναι ελάχιστη. Συγκεκριμένα, κατά την εξέταση των σταθμισμένων βαθμολογιών F1 για την αρχιτεκτονική MLP 512, το BERT base uncased κατέγραψε 84,30% ενώ το DistilBERT base uncased υστερούσε στενά με 84,10%. Αυτή η διαφορά μικρότερη από 0,2% υπογραμμίζει περαιτέρω τη συγκρίσιμη αποτελεσματικότητά τους. Ωστόσο, μια αξιοσημείωτη διαφορά μπορεί να φανεί όταν χρησιμοποιείτε την αρχιτεκτονική LSTM 512. Εδώ, το BERT base uncased ξεπέρασε τις επιδόσεις του DistilBERT base uncased με βαθμολογία F1 από 82,86% έως 74,88%, το οποίο μπορεί υποδηλώνει πιθανές διαφορές στον τρόπο με τον οποίο αυτά τα μοντέλα αλληλεπιδρούν με διαφορετικές αρχιτεκτονικές.

Το DistilBERT, όπως υποδηλώνει το όνομά του, είναι μια αποσταγμένη έκδοση του BERT, σχεδιασμένη να είναι ελαφρύτερη και πιο αποτελεσματική. Έχοντας 40% λιγότερες παραμέτρους από το BERT σημαίνει ότι μπορεί να επιτύχει παρόμοια επίπεδα



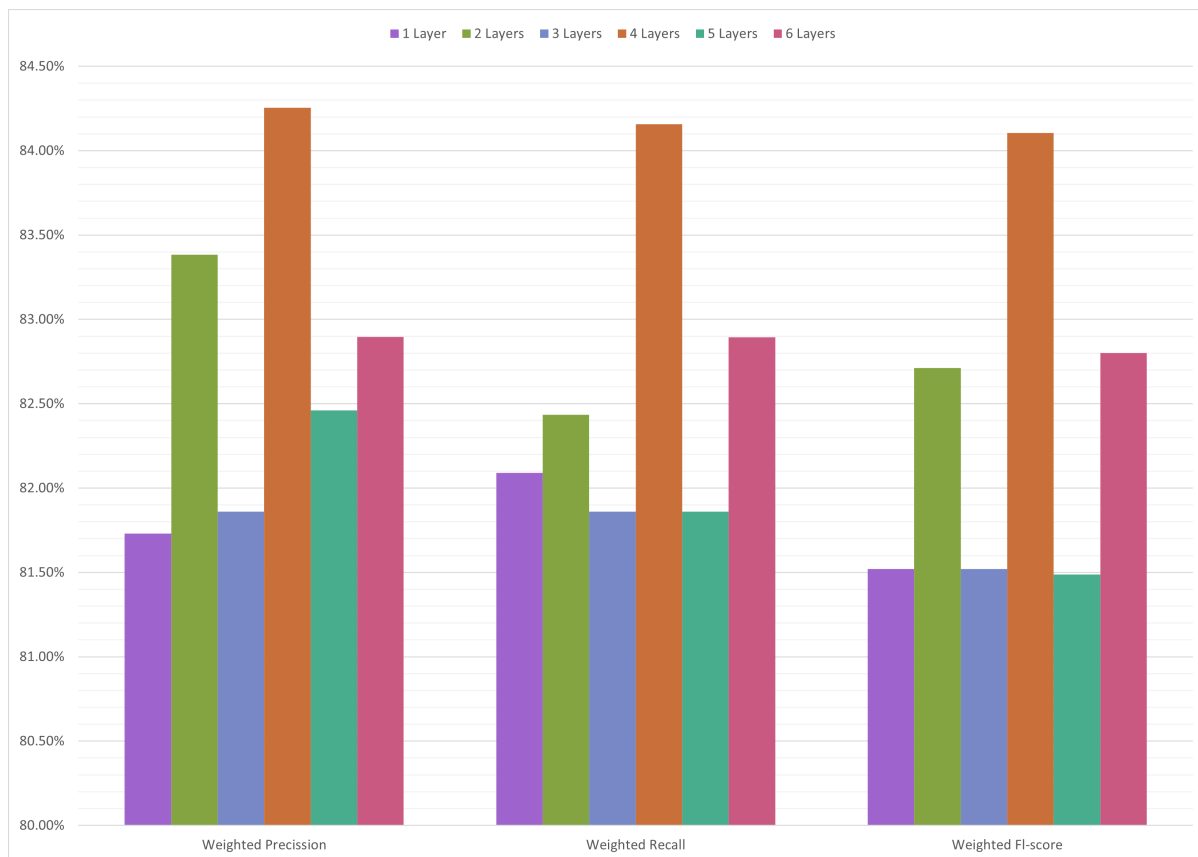
Σχήμα 5.2: Διάγραμμα αποτελεσμάτων μεταξύ των κεφαλών ταξινόμησης.

απόδοσης με σημαντικά μειωμένο υπολογιστικό αποτύπωμα. Αυτή η μείωση όχι μόνο επιταχύνει τον χρόνο προπόνησης κατά περίπου 60% αλλά προσφέρει και ταχύτερους χρόνους πρόβλεψης. Αυτός είναι ένας ουσιαστικός παράγοντας για εφαρμογές στον πραγματικό κόσμο όπου η ταχεία απόκριση είναι ζωτικής σημασίας. Επιπλέον, το μειωμένο μέγεθος του μοντέλου σημαίνει ότι απαιτεί χαμηλότερες απαιτήσεις μνήμης, καθιστώντας το μια ανώτερη επιλογή για ανάπτυξη σε περιβάλλοντα όπου οι πόροι είναι περιορισμένοι ή ακριβοί. Επιπλέον, όταν συγκρίνεται με την εργασία κατανόησης γλώσσας GLUE, το DistilBERT διατηρεί πάνω από το 95% της απόδοσης του BERT. Αυτό το αποτέλεσμα αναφοράς αποτελεί απόδειξη της αποτελεσματικότητας του DistilBERT, καθώς ανταλλάσσει ουσιαστικά μια μικρή πτώση στην απόδοση με σημαντικά κέρδη στην ταχύτητα και τη χρήση των πόρων.

Δεδομένων αυτών των ευρημάτων και των εμπειρικών στοιχείων από τα πειράματά μας, η απόφασή μας να υιοθετήσουμε το DistilBERT base uncased ως το μοντέλο μετασχηματιστή στην αρχιτεκτονική μας ήταν στρατηγική. Ενώ και τα δύο μοντέλα είναι ισχυρά και ικανά, τα εγγενή πλεονεκτήματα του DistilBERT όσον αφορά την ταχύτητα, το υπολογιστικό κόστος και τη χρήση μνήμης το καθιστούν ανεκτίμητο, ειδικά όταν εξετάζουμε περιβάλλοντα όπου οι υπολογιστικοί πόροι μπορεί να αποτελούν περιοριστικό παράγοντα.

Πίνακας 5.4: Αποτελέσματα με την επιλογή διαφορετικών στρωμάτων για MLP κρυφού μεγέθους 512.

Κεφαλή ταξινόμησης και κρυφό μέγεθος	number of transformer layers in use	Weighted Precision (%)	Weighted Recall (%)	Weighted F1-score (%)
MLP 512	1	81.73%	82.09%	81.52%
MLP 512	2	83.38%	82.43%	82.71%
MLP 512	3	81.86%	81.86%	81.52%
MLP 512	4	84.25%	84.16%	84.10%
MLP 512	5	82.46%	81.86%	81.49%
MLP 512	6	82.90%	82.89%	82.80%



Σχήμα 5.3: Διάγραμμα αποτελεσμάτων ανά στρώμα για το MLP με κρυφό μέγεθος 512.

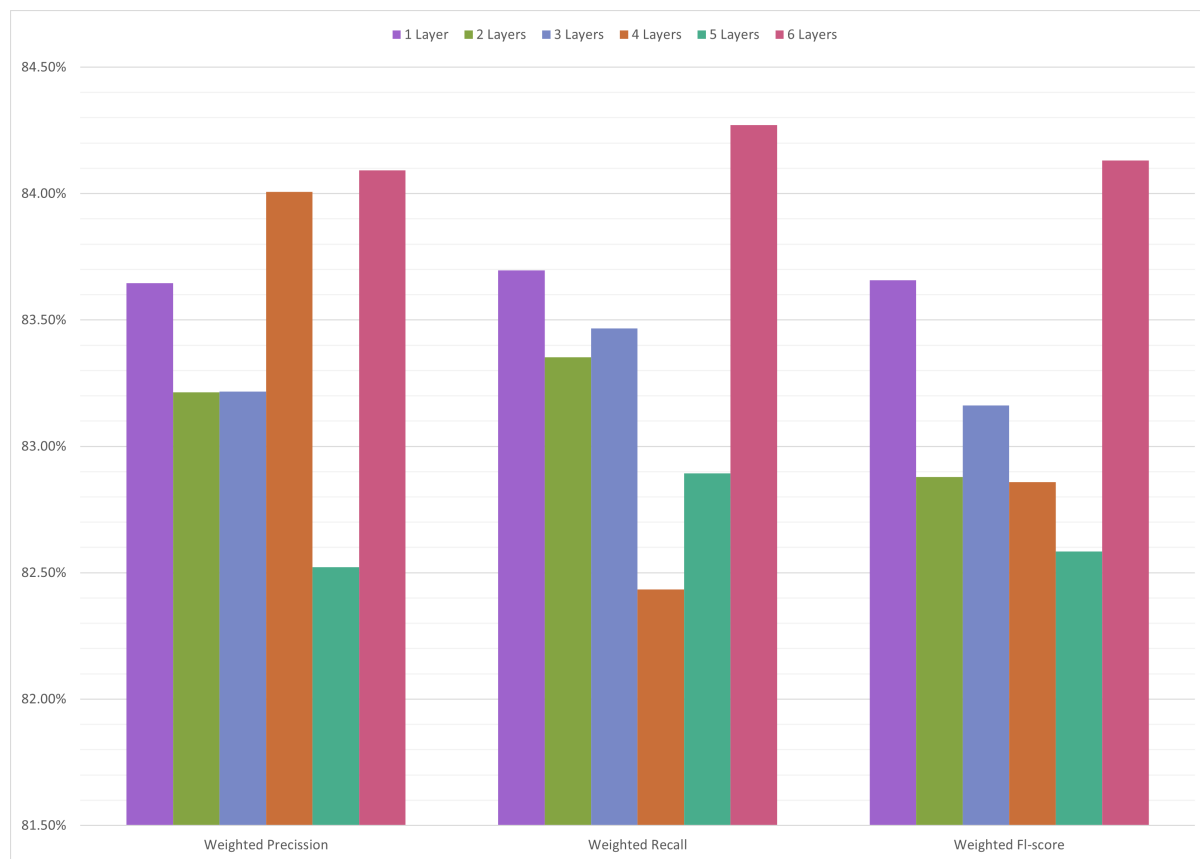
5.2.3 Στρώματα του μετασχηματιστή

Από τον Πίνακα 5.4 (και Διάγραμμα 5.3), ο οποίος λεπτομερεί τα αποτελέσματα χρησιμοποιώντας την κεφαλή ταξινόμησης MLP με κρυφό μέγεθος 512, μπορούμε να κάνουμε διάφορες παρατηρήσεις. Με ένα μόνο στρώμα του μετασχηματιστή, το μοντέλο παρουσίασε ένα βαρυκεντρωμένο σκορ F1 81,52%. Αυξάνοντας τα στρώματα στα δύο, η απόδοση βελτιώθηκε σημαντικά, φτάνοντας σε ένα σκορ F1 82,71%. Ενδιαφέρον είναι ότι με τρία στρώματα παρατηρήθηκε μια μείωση στην απόδοση, όπου το σκορ F1 έπεσε πίσω στο 81,52%. Η μέγιστη απόδοση επετεύχθη με τέσσερα στρώματα του μετασχηματιστή, φτάνοντας σε ένα βαρυκεντρωμένο σκορ F1 84,10%. Πέρα από αυτό το βάθος, η απόδοση του μοντέλου άρχισε να μειώνεται, με το σκορ F1 να πέφτει στο 81,49% και 82,80% για πέντε και έξι στρώματα αντίστοιχα.

Από την άλλη, ο Πίνακας 5.5 (και Διάγραμμα 5.4) παρέχει πληροφορίες για τα αποτελέσματα της κεφαλής ταξινόμησης LSTM, και πάλι με κρυφό μέγεθος 512. Ξε-

Πίνακας 5.5: Αποτελέσματα με την επιλογή διαφορετικών στρωμάτων για LSTM κρυφού μεγέθους 512.

Κεφαλή ταξινόμησης και κρυφό μέγεθος	number of transformer layers in use	Weighted Precision (%)	Weighted Recall (%)	Weighted F1-score (%)
LSTM 512	1	83.65%	83.70%	83.66%
LSTM 512	2	83.21%	83.35%	82.88%
LSTM 512	3	83.22%	83.47%	83.16%
LSTM 512	4	84.01%	82.43%	82.86%
LSTM 512	5	82.52%	82.89%	82.58%
LSTM 512	6	84.09%	84.27%	84.13%



Σχήμα 5.4: Διάγραμμα αποτελεσμάτων ανά στρώμα για το LSTM με κρυφό μέγεθος 512.

κινώντας με ένα μόνο στρώμα του μετασχηματιστή, το σκορ F1 καταγράφηκε στο 83,66%. Τα επόμενα στρώματα, μέχρι το τρίτο, παρουσίασαν σχετικά σταθερά σκορ, με τα σκορ F1 να περιστρέφονται γύρω από το εύρος του 82,88% έως 83,16%. Σε τέσσερα στρώματα, το σκορ F1 ήταν ελαφρώς χαμηλότερο στο 82,86%. Ωστόσο, η πιο σημαντική παρατήρηση ήταν ότι, αντίθετα με τα αποτελέσματα του MLP, το μοντέλο με βάση το LSTM έφτασε στο υψηλότερο σκορ F1 του 84,13% με έξι στρώματα του μετασχηματιστή.

Αυτά τα αποτελέσματα επισημαίνουν σαφώς την ευαίσθητη ισορροπία που απαιτείται για τον καθορισμό του βέλτιστου βάθους του μοντέλου που χρησιμοποιείτε. Ενώ για την κεφαλή MLP, η κορυφαία απόδοση βρέθηκε με τέσσερα στρώματα του μετασχηματιστή, η επικεφαλίδα LSTM απαιτεί ένα βάθος έξι στρωμάτων. Αυτό υπογραμμίζει τη σημασία της ειδικής ρύθμισης της αρχιτεκτονικής και παρουσιάζει τις λεπτές διαφορές μεταξύ των επικεφαλίδων ταξινόμησης MLP και LSTM στον χειρισμό των βαθμίδων των στρωμάτων του μετασχηματιστή.

5.2.4 Επαύξηση δεδομένων

Πίνακας 5.6: Σύγκριση των αποτελεσμάτων επαύξησης για μοντέλο DistillBERT base uncased και διάφορες κεφαλές ταξινόμησης χρησιμοποιώντας 4 στρώματα.

Κεφαλή ταξινόμησης και κρυφό μέγεθος	Weighted Score πριν την επαύξηση			Weighted Score μετά την επαύξηση		
	Precision	Recall	F1-score	Precision	Recall	F1-score
MLP 512	71.99%	73.41%	72.59%	84.25%	84.16%	84.10%
LSTM 512	73.56%	74.68%	73.75%	76.92%	75.38%	74.88%
MM-Xatt 512	58.16%	72.98%	63.54%	51.00%	61.65%	54.71%
MM-Gate 512	73.25%	71.57%	71.95%	80.43%	80.14%	79.88%

Στον Πίνακα 5.6 παρατηρούμε ότι μετά την επαύξηση τα μοντέλα MLP και LSTM και MM-Gate με κρυφό μέγεθος 512 στρωμάτων και μοντέλο μετασχηματιστή distillbert base uncased παρουσιάζουν αύξηση στα F1-score περίπου 12% και 1% και 8%. Το μοναδικό όμως μοντέλο σε κάθε συνδυασμό που παρουσιάζει μείωση μετά την επαύξηση είναι MM-Xatt με μείωση F1-score κατά 9%.

Πίνακας 5.7: Σύγκριση των αποτελεσμάτων επαύξησης για διαφορετικούς μετασχηματιστές.

Κεφαλή ταξινόμησης και κρυφό μέγεθος	Μετασχηματιστές	Weighted Score πριν την επαύξηση			Weighted Score μετά την επαύξηση		
		Precision	Recall	F1-score	Precision	Recall	F1-score
MLP 512	DistilBERT base uncased	71.99%	73.41%	72.59%	84.25%	84.16%	84.10%
MLP 512	BERT base uncased	72.52%	74.12%	73.16%	84.37%	84.27%	84.30%
LSTM 512	DistilBERT base uncased	73.56%	74.68%	73.75%	76.92%	75.38%	74.88%
LSTM 512	BERT base uncased	74.67%	72.98%	73.74%	84.01%	82.43%	82.86%

Βάσει του Πίνακα 5.7, φαίνεται ότι η επαύξηση δεδομένων βελτίωσε γενικά τις επιδόσεις των μοντέλων. Αν κοιτάξουμε τα σκορ της ακρίβειας (Weighted Precision), της ανάκλησης (Weighted Recall), και του F1 (Weighted F1-score), θα δούμε σημαντικές αυξήσεις σε πολλές περιπτώσεις γύρω στο 10%. Το μόνο μοντέλο που δεν αυξήθηκε σημαντικά είναι το LSTM με κρυφό μέγεθος 512 στρωμάτων και με distillbert base uncased μετασχηματιστή που αυξήθηκε μόνο κατά 6,5%.

Πίνακας 5.8: Σύγκριση των αποτελεσμάτων επαύξησης για μοντέλο DistilBERT base uncased και κεφαλή ταξινόμησης MLP με κρυφό μέγεθος 512.

Κεφαλή ταξινόμησης και κρυφό μέγεθος	Αριθμός στρωμάτων μετασχηματιστή	Weighted Score πριν την επαύξηση			Weighted Score μετά την επαύξηση		
		Precision	Recall	F1-score	Precision	Recall	F1-score
MLP 512	1	73.86%	75.81%	74.64%	81.73%	82.09%	81.52%
MLP 512	2	72.24%	67.47%	69.30%	83.38%	82.43%	82.71%
MLP 512	3	75.69%	72.28%	72.73%	81.86%	81.86%	81.52%
MLP 512	4	71.99%	73.41%	72.59%	84.25%	84.16%	84.10%
MLP 512	5	71.67%	74.82%	72.78%	82.46%	81.86%	81.49%
MLP 512	6	73.88%	74.68%	74.22%	82.90%	82.89%	82.80%

Για παράδειγμα, το μοντέλο MLP με 4 επίπεδα και βασισμένο στο distillbert base uncased (Πίνακας 5.8) είχε F1-score 72.59% χωρίς επαύξηση, και 84.10% με επαύξηση. Εδώ βλέπουμε μια βελτίωση, κοντά στο 12%. Το ίδιο βλέπουμε να γίνεται και για τα άλλα στρώματα.

Όμοια στο μοντέλο LSTM με 4 επίπεδα και βασισμένο στο bert base uncased (Πίνακας 5.9) είχε F1-score 73.74% χωρίς επαύξηση, και 82.86% με επαύξηση. Εδώ βλέπουμε μια βελτίωση, κοντά στο 10%. Το ίδιο βλέπουμε να γίνεται και για τα άλλα στρώματα όπου πριν την επαύξηση έχουμε τιμές γύρω στο 70% και μετά την επαύξηση τιμές κοντά στο 80%.

Πίνακας 5.9: Σύγκριση των αποτελεσμάτων επαύξησης για μοντέλο BERT base uncased και κεφαλή ταξινόμησης LSTM με κρυφό μέγεθος 512.

Κεφαλή ταξινόμησης και κρυφό μέγεθος	Αριθμός στρωμάτων μετασχηματιστή	Weighted Score πριν την επαύξηση			Weighted Score μετά την επαύξηση		
		Precision	Recall	F1-score	Precision	Recall	F1-score
LSTM 512	1	71.22%	73.41%	71.96%	83.65%	83.70%	83.66%
LSTM 512	2	74.38%	74.54%	73.88%	83.21%	83.35%	82.88%
LSTM 512	3	74.82%	74.26%	74.44%	83.22%	83.47%	83.16%
LSTM 512	4	74.67%	72.98%	73.74%	84.01%	82.43%	82.86%
LSTM 512	5	72.40%	75.39%	72.86%	82.52%	82.89%	82.58%
LSTM 512	6	72.86%	71.00%	72.19%	84.09%	84.27%	84.13%

Η επαύξηση δεδομένων προσφέρει στο μοντέλο περισσότερη "εμπειρία" και τη δυνατότητα να γενικεύει καλύτερα σε νέα δεδομένα. Προσθέτει ποικιλία και ανθεκτικότητα, επιτρέποντας στο μοντέλο να αντιμετωπίζει πιο ευέλικτα και αποτελεσματικά διάφορες συνθήκες. Συνολικά, η επαύξηση δεδομένων φαίνεται να έχει θετικό αντίκτυπο, αυξάνοντας τις μετρικές απόδοσης σημαντικά, σε κάποιες περιπτώσεις έως και 13%.

Κεφάλαιο 6

Συμπεράσματα και μελλοντικές προεκτάσεις

6.1 Συμπεράσματα

Στον συνεχώς εξελισσόμενο τομέα της Επεξεργασίας Φυσικής Γλώσσας (NLP), η χρήση προεκπαιδευμένων μοντέλων όπως το BERT έχει φέρει επανάσταση στον τρόπο με τον οποίο κατανοούμε και επεξεργαζόμαστε τεράστιες ποσότητες δεδομένων κειμένου. Αυτή η διπλωματική παρουσίασε μια νέα προσέγγιση για την κατηγοριοποίηση του επιπέδου της κατάθλιψης συγχωνεύοντας τα αποτελέσματα του BERT με συμπληρωματικά μεταδεδομένα. Μετά από επεξεργασία, το συνδυασμένο διάνυμα, όταν διήλθε μέσω ενός Perceptron πολλαπλών στρωμάτων (MLP), έδειξε πολλά υποσχόμενο στην ικανότητά του να κατηγοριοποιεί τα επίπεδα κατάθλιψης με αυξημένη ακρίβεια.

Ένα από τα κύρια κίνητρα πίσω από αυτήν την έρευνα ήταν η κατανόηση ότι ενώ το BERT είναι εξαιρετικά ικανό στην αποτύπωση γλωσσικών αποχρώσεων, υπάρχουν εξωτερικοί παράγοντες και μεταδεδομένα που μπορούν να παρέχουν πρόσθετες πληροφορίες, ειδικά όταν ασχολούμαστε με ένα ευαίσθητο και πολύπλοκο θέμα όπως η ψυχική υγεία. Τα επιπλέον μεταδεδομένα έπαιξαν καθοριστικό ρόλο σε αυτή την έρευνα. Οι αναφορές στη φαρμακευτική αγωγή στο κείμενο προέκυψαν ως ιδιαίτερα χρήσιμοι δείκτες, που ρίχνουν φως στη σοβαρότητα και την φύση της κατάστασης του ατόμου. Επιπλέον, ο αρνητικός τόνος του κειμένου και τα άλλα συναισθήματα που εκφράζονται σε αυτό παρείχαν πολύτιμες ενδείξεις, ενισχύοντας περαιτέρω την ικανότητα του μοντέλου να διακρίνει τα επίπεδα κατάθλιψης.

Με την ενσωμάτωση αυτών των εμπλουτισμένων μεταδεδομένων με τα embeddings του BERT, το μοντέλο όχι μόνο αποκτά μια πιο ολιστική εικόνα των δεδομένων, αλλά αντιμετωπίζει επίσης τους περιορισμούς της εξάρτησης αποκλειστικά από το περιεχόμενο του κειμένου. Αυτή η προσέγγιση συγχώνευσης υπογραμμίζει τη σημασία της εξέτασης τόσο γλωσσικών όσο και συμπραζομένων στοιχείων στις αξιολογήσεις ψυχικής υγείας.

Η έρευνα μας πάνω σε πολυτροπικές πύλες αποκάλυψε ότι μπορεί να μην είναι οι πιο αποτελεσματικές για το συγκεκριμένο θέμα. Ενώ αυτές οι πύλες έχουν δείξει πολλά υποσχόμενες σε άλλους τομείς, η εφαρμογή τους στο πλαίσιο της κατηγοριο-

ποίησης του επιπέδου κατάθλιψης απαιτεί περαιτέρω βελτίωση. Υποδηλώνει ότι αυτές οι πύλες μπορεί να χρειάζονται τροποποιήσεις ή ακόμα και συνδυασμό με άλλες τεχνικές για να αξιοποιήσουν πραγματικά τις δυνατότητές τους σε αυτόν τον συγκεκριμένο τομέα.

Μια άλλη κομβική πτυχή αυτής της έρευνας ήταν η ενσωμάτωση τεχνικών επαύξησης κειμένου, οι οποίες ενίσχυσαν σημαντικά την εργασία ταξινόμησης. Όπως επισημαίνεται από το έργο των Wei και Zou [83], τέτοιες μέθοδοι επαύξησης μπορούν να παρέχουν στα μοντέλα μια πλουσιότερη κατανόηση των παραλλαγών του κειμένου, οδηγώντας σε βελτιωμένη απόδοση. Αυτό το εύρημα υπογραμμίζει τη σημασία της ποικιλομορφίας των δεδομένων και της προσθήκης θορύβου στην εκπαίδευση ισχυρών μοντέλων NLP, ειδικά σε τομείς όπως η ψυχική υγεία και οι παθήσεις της.

Η προσέγγιση κατηγοριοποίησης πολλαπλών ετικετών που υιοθετήθηκε σε αυτήν την έρευνα υπογραμμίζει περαιτέρω την πολυπλοκότητα και την πολύπλευρη φύση της κατάθλιψης. Οι παραδοσιακές δυαδικές κατηγοριοποιήσεις συχνά υπεραπλουστεύουν την κατάσταση, αποτυγχάνοντας να καταγράψουν τις περίπλοκες διαβαθμίσεις της. Το προτεινόμενο μοντέλο MLP, εκπαιδευμένο στο συγχωνευμένο διάγραμμα, αναγνωρίζει αποτελεσματικά αυτές τις αποχρώσεις, ανοίγοντας το δρόμο για πιο εξατομικευμένες και ακριβείς αξιολογήσεις ψυχικής υγείας.

Για την ακαδημαϊκή κοινότητα, αυτή η έρευνα χρησιμεύει ως απόδειξη για τις δυνατότητες των υβριδικών μοντέλων στο NLP. Αξιοποιώντας τα πλεονεκτήματα τόσο των προεκπαιδευμένων μοντέλων, όπως το BERT όσο και των παραδοσιακών νευρωνικών δικτύων όπως το MLP, οι ερευνητές μπορούν να επιτύχουν μια συνέργεια που ξεπερνά τα μεμονωμένα μοντέλα μεμονωμένα. Αυτή η προσέγγιση συγχώνευσης, όπως αποδεικνύεται, μπορεί να είναι ιδιαίτερα επωφελής σε πολλούς τομείς μελέτης.

Συμπερασματικά, η συγχώνευση των αποτελεσμάτων BERT με πρόσθετα μεταδεδομένα, σε συνδυασμό με αποτελεσματικές τεχνικές αύξησης δεδομένων, αποτελεί μια πολλά υποσχόμενη οδό για την ενίσχυση της ακρίβειας και του βάθους των εφαρμογών NLP στον τομέα της ψυχικής υγείας. Καθώς συνεχίζουμε να βελτιώνουμε και να καινοτομούμε, ελπίζουμε ότι τέτοιες εξελίξεις θα οδηγήσουν σε μια βαθύτερη κατανόηση των συνθηκών ψυχικής υγείας, ωφελώντας τελικά όσους έχουν ανάγκη.

6.2 Μελλοντικές Προεκτάσεις

Εμβαθύνοντας στην αλληλεπίδραση μεταξύ γλωσσικών ενδείξεων και ανθρώπινου συναισθήματος, αυτή η διπλωματική άνοιξε δρόμους για τον εντοπισμό σημείων κατάθλιψης στις αναρτήσεις στα μέσα κοινωνικής δικτύωσης. Ενώ έχει σημειωθεί σημαντική πρόοδος, το έργο αυτό περαιτέρω εξερεύνησης:

1. **Χρήση άλλων συνόλων δεδομένων:** Τα μέσα κοινωνικής δικτύωσης, με την τεράστια παγκόσμια εμβέλειά τους, παρέχουν μια μεγάλη πηγή δεδομένων. Η αξιοποίηση πιο εκτεταμένων και διαφορετικών συνόλων δεδομένων μπορεί να βελτιώσει το μοντέλο μας, καθιστώντας το καθολικά ικανό και ευαίσθητο σε διάφορες κοινωνιογλωσσικές αποχρώσεις.
2. **Συγχώνευση πυλών και νευρωνικών δικτύων ταξινόμησης:** Η συγχώνευση των

λειτουργιών των Gates με νευρωνικά δίκτυα ταξινόμησης, όπως τα LSTM ή τα CNN, μπορεί να αποφέρει πλουσιότερες γλωσσικές γνώσεις. Αυτή η υβριδική προσέγγιση θα μπορούσε να καταγράψει πιο αποτελεσματικά περίπλοκα μοτίβα ενδεικτικά καταθλιπτικών τάσεων.

3. **Χρήση πιο μεγάλων και πολύπλοκων μετασχηματιστών:** Η αξιοποίηση της των μοντέλων τελευταίας τεχνολογίας όπως το Chat GPT και το Llama 2 θα μπορούσε να ωθήσει το μοντέλο μας σε μεγαλύτερη ακρίβεια. Αυτοί οι μετασχηματιστές, με τις εκλεπτυσμένες ικανότητές τους επεξεργασίας κειμένου, θα μπορούσαν να αποκαλύψουν διακριτικούς δείκτες καταθλιπτικού συναισθήματος που συχνά παραβλέπονται.
4. **Πολυδιάστατη ανάλυση χαρακτηριστικών:** Πέρα από το κύριο κείμενο, χαρακτηριστικά όπως οι χρονικές σφραγίδες που έγινε η ανάρτηση μπορεί να προσφέρουν πρόσθετες ενδείξεις. Για παράδειγμα, οι νυχτερινές δημοσιεύσεις υποδηλώνουν αϋπνία; Εργαλεία όπως η προσθήκη ετικετών LIWC ή POS μπορούν να αναλύσουν περαιτέρω τις αναρτήσεις, ξετυλίγοντας τα συναισθηματικά και συντακτικά επίπεδα που είναι ενσωματωμένα μέσα.
5. **Χρήση ιστορικού χρήστη:** Η υιοθέτηση ενός ιστορικού θα μπορούσε να παρέχει πληροφορίες για την εξελισσόμενη ψυχική κατάσταση ενός χρήστη. Αυτή η προσέγγιση θα μπορούσε να είναι ανεκτίμητη για τη διάκριση της κυκλικής ή επεισοδιακής φύσης των καταθλιπτικών τάσεων, βοηθώντας ενδεχομένως σε πρώιμες παρεμβάσεις.
6. **Επέκταση των τροπικοτήτων:** Καθώς το περιεχόμενο πολυμέσων γίνεται αναπόσπαστο σε αναρτήσεις μέσω κοινωνικής δικτύωσης, το μοντέλο μας μπορεί να επωφεληθεί από την αποκρυπτογράφηση των συναισθημάτων των ενσωματωμένων εικόνων, βίντεο ή συνδέσμων. Τεχνικές από την όραση υπολογιστή ή την ανάλυση συναισθημάτων που βασίζεται στο περιεχόμενο βίντεο μπορεί να χρησιμοποιηθούν σε αυτές τις εκφράσεις.
7. **Ενσωμάτωση αισθητήρων κινητού:** Ενώ είναι πιο πολύπλοκη, η διασύνδεση δεδομένων φυσικής συμπεριφοράς όπως κινήσεις GPS με τις αναρτήσεις μπορεί να προσφέρει μια πιο στρογγυλεμένη άποψη της κατάστασης του χρήστη. Υπάρχουν σωματικές συμπεριφορές που σχετίζονται με διαδικτυακές εκφράσεις ερήμωσης ή απελπισίας;
8. **Επικύρωση αλήθειας:** Η συνεργασία με ιδρύματα ψυχικής υγείας θα μπορούσαν να προσφέρουν μια γέφυρα μεταξύ των προβλέψεων του μοντέλου μας και των κλινικών διαγνώσεων. Τέτοιες ευθυγραμμίσεις στον πραγματικό κόσμο θα μπορούσαν να αυξήσουν σημαντικά την αξιοπιστία και τη δυνατότητα εφαρμογής του μοντέλου.
9. **Διεύρυνση του φάσματος της ψυχικής υγείας:** Αν και η κύρια εστίασή μας ήταν η κατάθλιψη, το έργο θα μπορούσε να προσαρμοστεί για να ανιχνεύσει άλλες καταστάσεις ψυχικής υγείας που εκδηλώνονται στα μέσα κοινωνικής δικτύωσης, από το άγχος έως το PTSD ή ακόμα και την ανορεξία.
10. **Εξηγησιμότητα:** Προκειμένου να αυξήσουμε την εξηγησιμότητα του μοντέλου μας και να κατανοήσουμε καλύτερα τις αποφάσεις του, μπορούμε να εξετάσουμε τη χρήση μεθόδων εξηγησιμότητας [105] [106]. Αυτές οι μέθοδοι μπο-

ρούν να προσφέρουν επιπλέον ενδείξεις σχετικά με τον τρόπο με τον οποίο το μοντέλο κατανοεί και κατατάσσει τα συναισθήματα στις αναρτήσεις. Η ενσωμάτωση αυτών των μεθόδων στο μοντέλο μας μπορεί να βοηθήσει τους ερευνητές και τους επαγγελματίες της ψυχικής υγείας να κατανοήσουν καλύτερα τα αποτελέσματα και τις προβλέψεις του μοντέλου σχετικά με την κατάθλιψη και άλλες καταστάσεις ψυχικής υγείας στα μέσα κοινωνικής δικτύωσης.

11. **Χρήση του συνόλου δεδομένων DAIC-WOZ:** Σε μελλοντική έρευνα μπορούμε να αξιοποιήσουμε το σύνολο δεδομένων DAIC-WOZ για την ανίχνευση της κατάθλιψης, εμπνευσμένοι από έργα τελευταίας τεχνολογίας [107] [108] [109] και ενσωματώνοντας τις μεθοδολογίες που παρουσιάσαμε. Συγκεκριμένα, σχεδιάζουμε να αξιοποιήσουμε την πολυτροπική φύση του συνόλου δεδομένων, υιοθετώντας τεχνικές παρόμοιες με αυτές που παρουσιάστηκαν στις παραπάνω έρευνες. Αυτές οι μελέτες, όπως παρουσιάζονται, έχουν επιδείξει την αποτελεσματικότητα των πολυτροπικών μοντέλων που περιλαμβάνουν καινοτόμα στοιχεία, όπως τα Multimodal Shifting Gates και στρώματα προσοχής που λαμβάνουν υπόψη τις αλληλεπιδράσεις μεταξύ των τροπικοτήτων. Με τη διεύρυνση αυτών των μεθόδων, στοχεύουμε στη βελτίωση της ακρίβειας και της πληρότητας των προβλεπτικών μας μοντέλων.
12. **Χρήση μεθόδων αναζήτησης νευρικών αρχιτεκτονικών:** Ακόμη σκοπεύουμε να εφαρμόσουμε την χρήση μεθόδων αναζήτησης νευρικών αρχιτεκτονικών (Neural Architecture Search, NAS), οι οποίες χρησιμοποιούνται σε έργα τελευταίας τεχνολογίας [110] [111]. Εμπνευσμένοι από τις μεθοδολογίες τους, θα χρησιμοποιήσουμε το NAS για να ανακαλύψουμε αυτόματα τις βέλτιστες αρχιτεκτονικές νευρωνικών δικτύων προσαρμοσμένες στο έργο της διάγνωσης της κατάθλιψης. Για να το επιτύχουμε αυτό, θα προσαρμόσουμε και θα επεκτείνουμε τεχνικές όπως παρουσιάζονται στα έργα με σκοπό να βελτιώσουμε την αποτελεσματικότητα και την επεκτασιμότητα της διαδικασίας αναζήτησης αρχιτεκτονικής. Η έρευνά αυτή θα συνεισφέρει σημαντικά στον τομέα της διάγνωσης της κατάθλιψης, προσφέροντας πιο αποτελεσματικές προσεγγίσεις που βασίζονται στη βαθιά μάθηση για την αξιολόγηση της ψυχικής υγείας.

Σε αυτό το πλέγμα γλωσσολογίας, τεχνολογίας και ψυχικής ευεξίας, η τρέχουσα έρευνα είναι πολλά υποσχόμενη. Ο ορίζοντας μπροστά είναι τεράστιος, προσκαλώντας διεπιστημονική συνεργασία και συνεχή εξερεύνηση, προσπαθώντας για μια ολοκληρωμένη κατανόηση και υποστήριξη των ψηφιακών αντανakλάσεων της ανθρώπινης ψυχής.

Βιβλιογραφία

- [1] World Health Organization. "Mental disorders". (30 Ιούν. 2023), διεύθυν.: <https://www.who.int/news-room/fact-sheets/detail/mental-disorders> (επίσκεψη 05/07/2023).
- [2] World Health Organization. "Depressive disorder (depression)". (30 Ιούν. 2023), διεύθυν.: <https://www.who.int/news-room/fact-sheets/detail/depression> (επίσκεψη 05/07/2023).
- [3] A. M. Kaplan και M. Haenlein, "Users of the world, unite! the challenges and opportunities of social media", *Business Horizons*, τόμ. 53, αρθμ. 1, Ιαν. 2010, ISSN: 00076813. DOI: 10.1016/j.bushor.2009.09.003. διεύθυν.: <https://linkinghub.elsevier.com/retrieve/pii/S0007681309001232> (επίσκεψη 18/08/2023).
- [4] "Alone together: Why we expect more from technology and less from each other", *Choice Reviews Online*, τόμ. 48, αρθμ. 12, 1 Αύγ. 2011, ISSN: 0009-4978, 1523-8253. DOI: 10.5860/CHOICE.48-7239. διεύθυν.: <http://choicereviews.org/review/10.5860/CHOICE.48-7239> (επίσκεψη 18/08/2023).
- [5] J. M. Twenge και W. K. Campbell, "The narcissism epidemic: Living in the age of entitlement", Ιαν. 2009.
- [6] E. Pariser, "The filter bubble: how the new personalized Web is changing what we read and how we think", *Choice Reviews Online*, Οκτ. 2012. DOI: 10.5860/choice.50-0926.
- [7] L. Ilias, D. Askounis και J. Psarras, "Detecting dementia from speech and transcripts using transformers", *Computer Speech & Language*, τόμ. 79, 1 Απρ. 2023, ISSN: 0885-2308. DOI: 10.1016/j.csl.2023.101485. διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S0885230823000049> (επίσκεψη 07/09/2023).
- [8] L. Ilias, I. M. Kazelidis και D. Askounis, *Multimodal Detection of Social Spambots in Twitter using Transformers*, 28 Αύγ. 2023. DOI: 10.48550/arXiv.2308.14484. arXiv: 2308.14484[cs]. διεύθυν.: <http://arxiv.org/abs/2308.14484> (επίσκεψη 07/09/2023).
- [9] J. C. Eichstaedt, R. J. Smith, R. M. Merchant κ.ά., "Facebook language predicts depression in medical records", *Proceedings of the National Academy of Sciences*, τόμ. 115, αρθμ. 44, 30 Οκτ. 2018. DOI: 10.1073/pnas.1802331115. διεύθυν.: <https://www.pnas.org/doi/full/10.1073/pnas.1802331115> (επίσκεψη 29/03/2023).

- [10] Y. Huang, X. Liu και T. Zhu, "Suicidal ideation detection via social media analytics", στο *Human Centered Computing*, D. Milošević, Y. Tang και Q. Zu, επιμελητές, σειρά Lecture Notes in Computer Science, Cham: Springer International Publishing, 2019, ISBN: 978-3-030-37429-7. DOI: 10.1007/978-3-030-37429-7_17.
- [11] V. Leiva και A. Freire, "Towards suicide prevention: Early detection of depression on social media", στο *Internet Science*, I. Kompatsiaris, J. Cave, A. Satsiou κ.ά., επιμελητές, σειρά Lecture Notes in Computer Science, Cham: Springer International Publishing, 2017, ISBN: 978-3-319-70284-1. DOI: 10.1007/978-3-319-70284-1_34.
- [12] M. M. Aldarwish και H. F. Ahmad, "Predicting Depression Levels Using Social Media Posts", στο *2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS)*, Μαρ. 2017. DOI: 10.1109/ISADS.2017.41.
- [13] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang και A. Ulhaq, "Depression detection from social network data using machine learning techniques", *Health Information Science and Systems*, τόμ. 6, αρθμ. 1, 27 Αύγ. 2018, ISSN: 2047-2501. DOI: 10.1007/s13755-018-0046-0. διεύθυν.: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6111060/> (επίσκεψη 24/03/2023).
- [14] A.-S. Uban, B. Chulvi και P. Rosso, "An emotion and cognitive based analysis of mental health disorders from social media data", *Future Generation Computer Systems*, τόμ. 124, 1 Νοέ. 2021, ISSN: 0167-739X. DOI: 10.1016/j.future.2021.05.032. διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S0167739X21001825> (επίσκεψη 27/03/2023).
- [15] I. Ameer, M. Arif, G. Sidorov, H. Gómez-Adorno και A. Gelbukh, *Mental Illness Classification on Social Media Texts using Deep Learning and Transfer Learning*, 3 Ιουλ. 2022. DOI: 10.48550/arXiv.2207.01012. arXiv: 2207.01012 [cs]. διεύθυν.: <http://arxiv.org/abs/2207.01012> (επίσκεψη 16/04/2023).
- [16] S. R. Kamite και V. B. Kamble, "Detection of Depression in Social Media via Twitter Using Machine learning Approach", στο *2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)*, Οκτ. 2020. DOI: 10.1109/ICSIDEMPC49020.2020.9299641.
- [17] F. CACHEDA, D. Fernandez, F. J. Novoa και V. Carneiro, "Early Detection of Depression: Social Network Analysis and Random Forest Techniques", *Journal of Medical Internet Research*, τόμ. 21, αρθμ. 6, 14 Ιούν. 2019. DOI: 10.2196/12554. διεύθυν.: <https://www.jmir.org/2019/6/e12554> (επίσκεψη 29/03/2023).
- [18] R. Chiong, G. S. Budhi, S. Dhakal και F. Chiong, "A textual-based featuring approach for depression detection using machine learning classifiers and social media texts", *Computers in Biology and Medicine*, τόμ. 135, 1 Αύγ. 2021, ISSN: 0010-4825. DOI: 10.1016/j.combiomed.2021.104499. διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S0010482521002936> (επίσκεψη 27/03/2023).
- [19] R. Safa, P. Bayat και L. Moghtader, "Automatic detection of depression symptoms in twitter using multimodal analysis", *The Journal of Supercomputing*, τόμ. 78, αρθμ. 4, 1 Μαρ. 2022, ISSN: 1573-0484. DOI: 10.1007/s11227-021-04040-8. διεύθυν.: <https://doi.org/10.1007/s11227-021-04040-8> (επίσκεψη 27/03/2023).

- [20] F. M. Shah, F. Ahmed, S. K. Saha Joy κ.ά., “Early Depression Detection from Social Network Using Deep Learning Techniques”, στο *2020 IEEE Region 10 Symposium (TENSYP)*, Ιούν. 2020. DOI: 10.1109/TENSYP50017.2020.9231008.
- [21] M. Y. Wu, C.-Y. Shen, E. T. Wang και A. L. P. Chen, “A deep architecture for depression detection using posting, behavior, and living environment data”, *Journal of Intelligent Information Systems*, τόμ. 54, αρθμ. 2, 1 Απρ. 2020, ISSN: 1573-7675. DOI: 10.1007/s10844-018-0533-4. διεύθυν.: <https://doi.org/10.1007/s10844-018-0533-4> (επίσκεψη 06/04/2023).
- [22] Q. Cong, Z. Feng, F. Li, Y. Xiang, G. Rao και C. Tao, “X-A-BiLSTM: a Deep Learning Approach for Depression Detection in Imbalanced Data”, στο *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Δεκ. 2018. DOI: 10.1109/BIBM.2018.8621230.
- [23] H. Yao, S. Rashidian, X. Dong, H. Duanmu, R. N. Rosenthal και F. Wang, “Detection of Suicidality Among Opioid Users on Reddit: Machine Learning–Based Approach”, *Journal of Medical Internet Research*, τόμ. 22, αρθμ. 11, 27 Νοέ. 2020. DOI: 10.2196/15293. διεύθυν.: <https://www.jmir.org/2020/11/e15293> (επίσκεψη 27/03/2023).
- [24] M. Trotzek, S. Koitka και C. M. Friedrich, “Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences”, *IEEE Transactions on Knowledge and Data Engineering*, τόμ. 32, αρθμ. 3, Μαρ. 2020, ISSN: 1558-2191. DOI: 10.1109/TKDE.2018.2885515.
- [25] A. Yates, A. Cohan και N. Goharian, “Depression and Self-Harm Risk Assessment in Online Forums”, στο *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark: Association for Computational Linguistics, Σεπτ. 2017. DOI: 10.18653/v1/D17-1322. διεύθυν.: <https://aclanthology.org/D17-1322> (επίσκεψη 30/03/2023).
- [26] G. Gkotsis, A. Oellrich, S. Velupillai κ.ά., “Characterisation of mental health conditions in social media using informed deep learning”, *Scientific Reports*, τόμ. 7, αρθμ. 1, 22 Μαρ. 2017, ISSN: 2045-2322. DOI: 10.1038/srep45141. διεύθυν.: <https://www.nature.com/articles/srep45141> (επίσκεψη 30/03/2023).
- [27] Z. Chen, R. Yang, S. Fu, N. Zong, H. Liu και M. Huang, *Detecting Reddit Users with Depression Using a Hybrid Neural Network*, 3 Φεβ. 2023. DOI: 10.48550/arXiv.2302.02759. arXiv: 2302.02759[cs]. διεύθυν.: <http://arxiv.org/abs/2302.02759> (επίσκεψη 21/03/2023).
- [28] C. Lin, P. Hu, H. Su κ.ά., “SenseMood: Depression Detection on Social Media”, στο *Proceedings of the 2020 International Conference on Multimedia Retrieval*, σειρά ICMR’20, New York, NY, USA: Association for Computing Machinery, 8 Ιούν. 2020, ISBN: 978-1-4503-7087-5. DOI: 10.1145/3372278.3391932. διεύθυν.: <https://doi.org/10.1145/3372278.3391932> (επίσκεψη 05/04/2023).
- [29] G. Rao, Y. Zhang, L. Zhang, Q. Cong και Z. Feng, “MGL-CNN: A Hierarchical Posts Representations Model for Identifying Depressed Individuals in Online Forums”, *IEEE Access*, τόμ. 8, 2020, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2973737.

- [30] M. Kabir, T. Ahmed, M. B. Hasan κ.ά., “DEPTWEET: A typology for social media texts to detect depression severities”, *Computers in Human Behavior*, τόμ. 139, 1 Φεβ. 2023, ISSN: 0747-5632. DOI: 10.1016/j.chb.2022.107503. διεύθν.: <https://www.sciencedirect.com/science/article/pii/S0747563222003235> (επίσκεψη 15/04/2023).
- [31] A. Murarka, B. Radhakrishnan και S. Ravichandran, *Detection and Classification of mental illnesses on social media using RoBERTa*, 23 Νοέ. 2020. DOI: 10.48550/arXiv.2011.11226. διεύθν.: <http://arxiv.org/abs/2011.11226> (επίσκεψη 16/04/2023).
- [32] J. Devlin, M.-W. Chang, K. Lee και K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding”, *CoRR*, τόμ. abs/1810.04805, 2018. arXiv: 1810.04805. διεύθν.: <http://arxiv.org/abs/1810.04805>.
- [33] W. Foundation. “Wikimedia downloads”. (25 Ιαν. 2022), διεύθν.: <https://dumps.wikimedia.org>.
- [34] Y. Zhu, R. Kiros, R. Zemel κ.ά., “Aligning books and movies: Towards story-like visual explanations by watching movies and reading books”, στο *The IEEE international conference on computer vision (ICCV)*, Δεκ. 2015.
- [35] V. Sanh, L. Debut, J. Chaumond και T. Wolf, “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter”, *ArXiv*, τόμ. abs/1910.01108, 2019.
- [36] S. G. Burdisso, M. Errecalde και M. Montes-y-Gómez, “A text classification framework for simple and effective early depression detection over social media streams”, *Expert Systems with Applications*, τόμ. 133, 1 Νοέ. 2019, ISSN: 0957-4174. DOI: 10.1016/j.eswa.2019.05.023. διεύθν.: <https://www.sciencedirect.com/science/article/pii/S0957417419303525> (επίσκεψη 29/03/2023).
- [37] D. E. Losada, F. Crestani και J. Parapar, “eRISK 2017: CLEF lab on early risk prediction on the internet: Experimental foundations”, στο *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, G. J. Jones, S. Lawless, J. Gonzalo κ.ά., επιμελητές, σειρά Lecture Notes in Computer Science, Cham: Springer International Publishing, 2017, ISBN: 978-3-319-65813-1. DOI: 10.1007/978-3-319-65813-1_30.
- [38] N. A. Asad, M. A. Mahmud Pranto, S. Afreen και M. M. Islam, “Depression Detection by Analyzing Social Media Posts of User”, στο *2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON)*, Νοέ. 2019. DOI: 10.1109/SPICSCON48833.2019.9065101.
- [39] “Welcome to LIWC-22”. (1 Ιαν. 1970), διεύθν.: <https://www.liwc.app/> (επίσκεψη 06/08/2023).
- [40] M. E. Aragón, A. P. López-Monroy, L. C. González-Gurrola και M. Montes-y-Gómez, “Detecting Mental Disorders in Social Media Through Emotional Patterns - The Case of Anorexia and Depression”, *IEEE Transactions on Affective Computing*, τόμ. 14, αρθμ. 1, 1 Ιαν. 2023, ISSN: 1949-3045. DOI: 10.1109/TAFFC.2021.3075638. διεύθν.: <https://doi.org/10.1109/TAFFC.2021.3075638> (επίσκεψη 16/04/2023).
- [41] B. Iavarone και A. Monreale, “From depression to suicidal discourse on Reddit”, στο *2021 IEEE International Conference on Big Data (Big Data)*, Δεκ. 2021. DOI: 10.1109/BigData52589.2021.9671801.

- [42] G. Coppersmith, M. Dredze και C. Harman, “Quantifying Mental Health Signals in Twitter”, στο *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Baltimore, Maryland, USA: Association for Computational Linguistics, Ιούν. 2014. DOI: 10.3115/v1/W14-3207. διεύθυν.: <https://aclanthology.org/W14-3207> (επίσκεψη 29/03/2023).
- [43] C. Hutto και E. Gilbert, “VADER: A parsimonious rule-based model for sentiment analysis of social media text”, *Proceedings of the International AAAI Conference on Web and Social Media*, τόμ. 8, αρθμ. 1, 16 Μάι. 2014, ISSN: 2334-0770. DOI: 10.1609/icwsm.v8i1.14550. διεύθυν.: <https://ojs.aaai.org/index.php/ICWSM/article/view/14550> (επίσκεψη 06/08/2023).
- [44] “Center for epidemiological studies depression (CESD)”, <https://www.apa.org>. (1 Ιαν. 1970), διεύθυν.: <https://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/depression-scale> (επίσκεψη 19/08/2023).
- [45] R. Chiong, G. S. Budhi και S. Dhakal, “Combining Sentiment Lexicons and Content-Based Features for Depression Detection”, *IEEE Intelligent Systems*, τόμ. 36, αρθμ. 6, Νοέ. 2021, ISSN: 1941-1294. DOI: 10.1109/MIS.2021.3093660.
- [46] F. Haque, R. U. Nur, S. A. Jahan, Z. Mahmud και F. M. Shah, “A Transformer Based Approach To Detect Suicidal Ideation Using Pre-Trained Language Models”, στο *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, Δεκ. 2020. DOI: 10.1109/ICCIT51783.2020.9392692.
- [47] K. Malviya, B. Roy και S. Saritha, “A Transformers Approach to Detect Depression in Social Media”, στο *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, Μαρ. 2021. DOI: 10.1109/ICAIS50930.2021.9395943.
- [48] X. Wang, S. Chen, T. Li κ.ά., “Depression Risk Prediction for Chinese Microblogs via Deep-Learning Methods: Content Analysis”, *JMIR Medical Informatics*, τόμ. 8, αρθμ. 7, 29 Ιούλ. 2020. DOI: 10.2196/17958. διεύθυν.: <https://medinform.jmir.org/2020/7/e17958> (επίσκεψη 18/04/2023).
- [49] L. Ilias και D. Askounis, *Multitask learning for recognizing stress and depression in social media*, 30 Μάι. 2023. DOI: 10.48550/arXiv.2305.18907. arXiv: 2305.18907 [cs]. διεύθυν.: <http://arxiv.org/abs/2305.18907> (επίσκεψη 07/09/2023).
- [50] E. Alpaydin, *Introduction to machine learning* (Adaptive computation and machine learning series), Fourth edition. Cambridge, Massachusetts: The MIT Press, 2020, 682 **pagetotals**, ISBN: 978-0-262-04379-3.
- [51] E. Alpaydin, *Machine learning* (The MIT Press essential knowledge series), Revised and updated edition. Cambridge, Massachusetts: The MIT Press, 2021, 255 **pagetotals**, ISBN: 978-0-262-54252-4.
- [52] J. D. Kelleher, B. Mac Namee και A. D’Arcy, *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. Cambridge, Massachusetts: The MIT Press, 2015, 595 **pagetotals**, ISBN: 978-0-262-02944-5.
- [53] J. Patterson και A. Gibson, *Deep learning: a practitioner’s approach*, First edition. Sebastopol, CA: O’Reilly, 2017, 507 **pagetotals**, ISBN: 978-1-4919-1425-0.

- [54] U. User:Cyc based on PNG version by, *English: Graphic showing how a support vector machine would choose a separating hyperplane for two classes of points in 2D. H1 does not separate the classes. H2 does, but only with a small margin. H3 separates them with the maximum margin.* 26 Νοέ. 2012. διεύθν.: [https://commons.wikimedia.org/wiki/File:Svm_separating_hyperplanes_\(SVG\).svg](https://commons.wikimedia.org/wiki/File:Svm_separating_hyperplanes_(SVG).svg) (επίσκεψη 08/08/2023).
- [55] “1.10. decision trees”, scikit-learn. (), διεύθν.: <https://scikit-learn/stable/modules/tree.html> (επίσκεψη 11/08/2023).
- [56] “File:random forest diagram complete.png - wikipedia”. (24 Μαρ. 2017), διεύθν.: https://commons.wikimedia.org/wiki/File:Random_forest_diagram_complete.png (επίσκεψη 20/08/2023).
- [57] “K-Means Clustering Algorithm - Javatpoint”. (1 Ιαν. 1970), διεύθν.: <https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning> (επίσκεψη 20/08/2023).
- [58] <https://aiimpacts.org/author/katja>. “Scale of the human brain”, AI Impacts. (16 Απρ. 2015), διεύθν.: <https://aiimpacts.org/scale-of-the-human-brain/> (επίσκεψη 16/07/2023).
- [59] U. Badseed translated {and} modified by, *Ελληνικά: Τυπική δομή ενός Νευρώνα.* 24 Ιαν. 2013. διεύθν.: https://commons.wikimedia.org/wiki/File:Neuron_el.svg (επίσκεψη 16/07/2023).
- [60] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain”, *Psychological Review*, τόμ. 65, αρθμ. 6, 1958, ISSN: 1939-1471. DOI: 10.1037/h0042519.
- [61] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever και R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting”, *Journal of Machine Learning Research*, Ιαν. 2014.
- [62] H. Hassan, A. Negm, M. Zahran και O. Saavedra, “ASSESSMENT OF ARTIFICIAL NEURAL NETWORK FOR BATHYMETRY ESTIMATION USING HIGH RESOLUTION SATELLITE IMAGERY IN SHALLOW LAKES: CASE STUDY EL BURULLUS LAKE.”, *International Water Technology Journal*, τόμ. 5, 1 Δεκ. 2015.
- [63] A. Sherstinsky, “Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network”, *Physica D: Nonlinear Phenomena*, τόμ. 404, 1 Μαρ. 2020, ISSN: 0167-2789. DOI: 10.1016/j.physd.2019.132306. διεύθν.: <https://www.sciencedirect.com/science/article/pii/S0167278919305974> (επίσκεψη 20/07/2023).
- [64] S. Hochreiter και J. Schmidhuber, “Long Short-Term Memory”, *Neural Computation*, τόμ. 9, αρθμ. 8, Νοέ. 1997, ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735.
- [65] G. Chevalier, *English: Schematic of the Long-Short Term Memory cell, a component of recurrent neural networks*, 16 Μάι. 2018. διεύθν.: https://commons.wikimedia.org/wiki/File:LSTM_Cell.svg (επίσκεψη 05/08/2023).
- [66] A. Vaswani, N. Shazeer, N. Parmar κ.ά., *Attention Is All You Need*, 1 Αύγ. 2023. DOI: 10.48550/arXiv.1706.03762. arXiv: 1706.03762[cs]. διεύθν.: <http://arxiv.org/abs/1706.03762> (επίσκεψη 05/08/2023).

- [67] J. Devlin, M.-W. Chang, K. Lee και K. Toutanova, *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, 24 Μάι. 2019. DOI: 10.48550/arXiv.1810.04805. arXiv: 1810.04805[cs]. διεύθυν.: <http://arxiv.org/abs/1810.04805> (επίσκεψη 05/08/2023).
- [68] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever κ.ά., “Language models are unsupervised multitask learners”, *OpenAI blog*, τόμ. 1, αρθμ. 8, 2019.
- [69] C. Raffel, N. Shazeer, A. Roberts κ.ά., *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*, 28 Ιούλ. 2020. DOI: 10.48550/arXiv.1910.10683. arXiv: 1910.10683[cs, stat]. διεύθυν.: <http://arxiv.org/abs/1910.10683> (επίσκεψη 20/08/2023).
- [70] “wikipedia · Datasets at Hugging Face”. (30 Ιούν. 2023), διεύθυν.: <https://huggingface.co/datasets/wikipedia> (επίσκεψη 29/07/2023).
- [71] “bookcorpus · Datasets at Hugging Face”. (30 Ιούν. 2023), διεύθυν.: <https://huggingface.co/datasets/bookcorpus> (επίσκεψη 29/07/2023).
- [72] “DistilBERT”. (1 Ιαν. 1970), διεύθυν.: https://huggingface.co/docs/transformers/model_doc/distilbert (επίσκεψη 14/09/2023).
- [73] A. Rai και S. Borah, “Study of various methods for tokenization”, στο *Applications of Internet of Things*, J. K. Mandal, S. Mukhopadhyay και A. Roy, επιμελητές, σειρά Lecture Notes in Networks and Systems, Singapore: Springer, 2021, ISBN: 9789811561986. DOI: 10.1007/978-981-15-6198-6_18.
- [74] “BERT - tokenization and encoding | albert au yeung”. (5 Ιούν. 2022), διεύθυν.: <https://albertauyeung.github.io/2020/06/19/bert-tokenization.html/> (επίσκεψη 19/08/2023).
- [75] “WordPiece tokenization - Hugging Face NLP Course”. (1 Ιαν. 1970), διεύθυν.: <https://huggingface.co/learn/nlp-course/chapter6/6> (επίσκεψη 19/08/2023).
- [76] T. Kudo και J. Richardson, “SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing”, στο *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Brussels, Belgium: Association for Computational Linguistics, Νοέ. 2018. DOI: 10.18653/v1/D18-2012. διεύθυν.: <https://aclanthology.org/D18-2012> (επίσκεψη 14/09/2023).
- [77] Y. Zhang, R. Jin και Z.-H. Zhou, “Understanding bag-of-words model: A statistical framework”, *International Journal of Machine Learning and Cybernetics*, τόμ. 1, αρθμ. 1, 1 Δεκ. 2010, ISSN: 1868-808X. DOI: 10.1007/s13042-010-0001-0. διεύθυν.: <https://doi.org/10.1007/s13042-010-0001-0> (επίσκεψη 19/08/2023).
- [78] G. Sidorov, F. Velasquez, E. Stamatatos, A. Gelbukh και L. Chanona-Hernández, “Syntactic N-grams as machine learning features for natural language processing”, *Expert Systems with Applications, Methods and Applications of Artificial and Computational Intelligence*, τόμ. 41, αρθμ. 3, 15 Φεβ. 2014, ISSN: 0957-4174. DOI: 10.1016/j.eswa.2013.08.015. διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S0957417413006271> (επίσκεψη 19/08/2023).
- [79] A. Aizawa, “An information-theoretic perspective of tf-idf measures”, *Information Processing & Management*, τόμ. 39, αρθμ. 1, 2003. DOI: 10.1016/S0306-4573(02)00021-3.

- [80] F. Almeida και G. Xexéo, *Word Embeddings: A Survey*, 1 Μάι. 2023. DOI: 10.48550/arXiv.1901.09069. arXiv: 1901.09069[cs, stat]. διεύθυν.: <http://arxiv.org/abs/1901.09069> (επίσκεψη 19/08/2023).
- [81] D. M. Blei, A. Y. Ng και M. I. Jordan, "Latent dirichlet allocation", *Journal of machine Learning research*, τόμ. 3, Jan 2003.
- [82] C. Shorten, T. M. Khoshgoftaar και B. Furht, "Text data augmentation for deep learning", *Journal of Big Data*, τόμ. 8, αρθμ. 1, 19 Ιούλ. 2021, ISSN: 2196-1115. DOI: 10.1186/s40537-021-00492-0. διεύθυν.: <https://doi.org/10.1186/s40537-021-00492-0> (επίσκεψη 19/08/2023).
- [83] J. Wei και K. Zou, "EDA: Easy data augmentation techniques for boosting performance on text classification tasks", στο *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, Hong Kong, China: Association for Computational Linguistics, Νοέ. 2019. DOI: 10.18653/v1/D19-1670. διεύθυν.: <https://aclanthology.org/D19-1670>.
- [84] C. Nwankpa, W. Ijomah, A. Gachagan και S. Marshall, *Activation Functions: Comparison of trends in Practice and Research for Deep Learning*, 8 Νοέ. 2018. DOI: 10.48550/arXiv.1811.03378. arXiv: 1811.03378[cs]. διεύθυν.: <http://arxiv.org/abs/1811.03378> (επίσκεψη 19/07/2023).
- [85] M. Sokolova και G. Lapalme, "A systematic analysis of performance measures for classification tasks", *Information Processing & Management*, τόμ. 45, αρθμ. 4, 1 Ιούλ. 2009, ISSN: 0306-4573. DOI: 10.1016/j.ipm.2009.03.002. διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S0306457309000259> (επίσκεψη 16/07/2023).
- [86] A. Kulkarni, D. Chong και F. A. Batarseh, "5 - foundations of data imbalance and solutions for a data democracy", στο *Data Democracy*, F. A. Batarseh και R. Yang, επιμελητές, Academic Press, 1 Ιαν. 2020, ISBN: 978-0-12-818366-3. DOI: 10.1016/B978-0-12-818366-3.00005-8. διεύθυν.: <https://www.sciencedirect.com/science/article/pii/B9780128183663000058> (επίσκεψη 16/07/2023).
- [87] P. C. Bruce, A. Bruce και P. Gedeck, *Practical statistics for data scientists: 50+ essential concepts using R and Python*, Second edition. Sebastopol, CA: O'Reilly Media, Inc, 2020, 342 **pagetotals**, ISBN: 978-1-4920-7294-2.
- [88] U. Naseem, A. G. Dunn, J. Kim και M. Khushi, "Early Identification of Depression Severity Levels on Reddit Using Ordinal Classification", στο *Proceedings of the ACM Web Conference 2022*, σειρά WWW '22, New York, NY, USA: Association for Computing Machinery, 25 Απρ. 2022, ISBN: 978-1-4503-9096-5. DOI: 10.1145/3485447.3512128. διεύθυν.: <https://doi.org/10.1145/3485447.3512128> (επίσκεψη 27/03/2023).
- [89] "Reddit Self-reported Depression Diagnosis (RSDD) dataset". (10 Σεπτ. 2017), διεύθυν.: <https://georgetown-ir-lab.github.io/emnlp17-depression/> (επίσκεψη 13/08/2023).

- [90] A. Cohan, B. Desmet, A. Yates, L. Soldaini, S. MacAvaney και N. Goharian, "SMHD: a Large-Scale Resource for Exploring Online Language Usage for Multiple Mental Health Conditions", στο *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA: Association for Computational Linguistics, Αύγ. 2018. διεύθυν.: <https://aclanthology.org/C18-1126> (επίσκεψη 13/08/2023).
- [91] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire και J. Blackburn, *The Pushshift Reddit Dataset*, 23 Ιαν. 2020. DOI: 10.48550/arXiv.2001.08435. arXiv: 2001.08435[cs]. διεύθυν.: <http://arxiv.org/abs/2001.08435> (επίσκεψη 13/08/2023).
- [92] "arpanghoshal/EmoRoBERTa · Hugging Face". (25 Ιαν. 2023), διεύθυν.: <https://huggingface.co/arpanghoshal/EmoRoBERTa> (επίσκεψη 04/05/2023).
- [93] D. Loureiro, F. Barbieri, L. Neves, L. Espinosa Anke και J. Camacho-collados, "TimeLMs: Diachronic language models from Twitter", στο *Proceedings of the 60th annual meeting of the association for computational linguistics: System demonstrations*, Dublin, Ireland: Association for Computational Linguistics, Μάι. 2022. DOI: 10.18653/v1/2022.acl-demo.25. διεύθυν.: <https://aclanthology.org/2022.acl-demo.25>.
- [94] J. Camacho-collados, K. Rezaee, T. Riahi κ.ά., "TweetNLP: Cutting-edge natural language processing for social media", στο *Proceedings of the 2022 conference on empirical methods in natural language processing: System demonstrations*, Abu Dhabi, UAE: Association for Computational Linguistics, Δεκ. 2022. DOI: 10.18653/v1/2022.emnlp-demos.5. διεύθυν.: <https://aclanthology.org/2022.emnlp-demos.5>.
- [95] "Utilizing transformer representations efficiently". (1 Ιαν. 1970), διεύθυν.: <https://kaggle.com/code/rhtsingh/utilizing-transformer-representations-efficiently> (επίσκεψη 29/07/2023).
- [96] malteos, *PyTorch BERT Document Classification*, 28 Απρ. 2023. διεύθυν.: <https://github.com/malteos/pytorch-bert-document-classification> (επίσκεψη 14/06/2023).
- [97] F. Pedregosa, G. Varoquaux, A. Gramfort κ.ά., "Scikit-learn: Machine learning in Python", *Journal of Machine Learning Research*, τόμ. 12, 2011.
- [98] M. Honnibal και I. Montani, "spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing", 2017.
- [99] T. Wolf, L. Debut, V. Sanh κ.ά., "Transformers: State-of-the-art natural language processing", στο *Proceedings of the 2020 conference on empirical methods in natural language processing: System demonstrations*, Online: Association for Computational Linguistics, Οκτ. 2020. διεύθυν.: <https://www.aclweb.org/anthology/2020.emnlp-demos.6>.
- [100] E. Ma, *NLP augmentation*, 2019. διεύθυν.: <https://github.com/makcedward/nlpaug>.
- [101] L. Ilias, S. Mouzakis και D. Askounis, "Calibration of Transformer-Based Models for Identifying Stress and Depression in Social Media", *IEEE Transactions on Computational Social Systems*, 2023, ISSN: 2329-924X. DOI: 10.1109/TCSS.2023.3283009.

- [102] K. Yang, T. Zhang και S. Ananiadou, “A mental state knowledge-aware and contrastive network for early stress and depression detection on social media”, *Information Processing & Management*, τόμ. 59, αρθμ. 4, 1 Ιούλ. 2022, ISSN: 0306-4573. DOI: 10.1016/j.ipm.2022.102961. διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S0306457322000796> (επίσκεψη 05/08/2023).
- [103] D. Sánchez Villegas και N. Aletras, “Point-of-Interest Type Prediction using Text and Images”, στο *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online και Punta Cana, Dominican Republic: Association for Computational Linguistics, Νοέ. 2021. DOI: 10.18653/v1/2021.emnlp-main.614. διεύθυν.: <https://aclanthology.org/2021.emnlp-main.614> (επίσκεψη 23/06/2023).
- [104] L. Ilias, D. Askounis και J. Psarras, “Multimodal detection of epilepsy with deep neural networks”, *Expert Systems with Applications*, τόμ. 213, 1 Μαρ. 2023, ISSN: 0957-4174. DOI: 10.1016/j.eswa.2022.119010. διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S0957417422020280> (επίσκεψη 07/09/2023).
- [105] L. Ilias και D. Askounis, “Explainable Identification of Dementia From Transcripts Using Transformer Networks”, *IEEE Journal of Biomedical and Health Informatics*, τόμ. 26, αρθμ. 8, Αύγ. 2022, ISSN: 2168-2208. DOI: 10.1109/JBHI.2022.3172479.
- [106] L. Ilias, F. Soldner και B. Kleinberg, *Explainable Verbal Deception Detection using Transformers*, 6 Οκτ. 2022. DOI: 10.48550/arXiv.2210.03080. arXiv: 2210.03080[cs]. διεύθυν.: <http://arxiv.org/abs/2210.03080> (επίσκεψη 15/09/2023).
- [107] L. Ilias και D. Askounis, “Context-aware attention layers coupled with optimal transport domain adaptation and multimodal fusion methods for recognizing dementia from spontaneous speech”, *Knowledge-Based Systems*, τόμ. 277, 9 Οκτ. 2023, ISSN: 0950-7051. DOI: 10.1016/j.knsys.2023.110834. διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S0950705123005841> (επίσκεψη 15/09/2023).
- [108] L. Ilias, D. Askounis και J. Psarras, “A Multimodal Approach for Dementia Detection from Spontaneous Speech with Tensor Fusion Layer”, στο *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, Σεπτ. 2022. DOI: 10.1109/BHI56158.2022.9926818.
- [109] L. Ilias και D. Askounis, “Multimodal Deep Learning Models for Detecting Dementia From Speech and Transcripts”, *Frontiers in Aging Neuroscience*, τόμ. 14, 2022, ISSN: 1663-4365. διεύθυν.: <https://www.frontiersin.org/articles/10.3389/fnagi.2022.830943> (επίσκεψη 15/09/2023).
- [110] M. Chatzianastasis, L. Ilias, D. Askounis και M. Vazirgiannis, “Neural Architecture Search with Multimodal Fusion Methods for Diagnosing Dementia”, στο *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Ιούν. 2023. DOI: 10.1109/ICASSP49357.2023.10096579.
- [111] H. Liu, K. Simonyan και Y. Yang, *DARTS: Differentiable Architecture Search*, 23 Απρ. 2019. DOI: 10.48550/arXiv.1806.09055. arXiv: 1806.09055[cs, stat]. διεύθυν.: <http://arxiv.org/abs/1806.09055> (επίσκεψη 15/09/2023).