Σχολή Θετικών Επιστημών και Τεχνολογίας

Μεταπτυχιακή Εξειδίκευση στα Πληροφοριακά Συστήματα

Διπλωματική Εργασία

Αναγνώριση τάσεων αυτοκτονίας σε κοινωνικά δίκτυα

Γεώργιος Μυρίσας

Επιβλέπων καθηγητής: Ανδρέας Καναβός

ΠΑΤΡΑ, Ιούλιος 2021

Αναγνώριση τάσεων αυτοκτονίας σε κοινωνικά δίκτυα

Γεώργιος Μυρίσας

Επιτροπή Επίβλεψης Διπλωματικής Εργασίας

Επιβλέπων Καθηγητής	Συν-Επιβλέπων Καθηγητής
Ανδρέας Καναβός	Αναστάσιος Σαλής
Συνεργαζόμενο Επιστημονικό Προσωπικό	Συνεργαζόμενο Επιστημονικό Προσωπικό
Ε.Α.Π.	Ε.Α.Π.

Πάτρα, Ιούλιος 2021

Στην οικογένειά μου και κυρίως στη σύζυγό μου Αθηνά, η οποία υπόμεινε αδιαμαρτύρητα τις ατελείωτες ώρες απομόνωσης μου στον ηλεκτρονικό υπολογιστή, καθώς και στα παιδιά μου, τον οχτώ ετών υιό μου Κωνσταντίνο και την επτά ετών κόρη μου Μαρία, τα οποία στερήθηκαν πολλές βόλτες στην παιδική χαρά.

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Ανδρέα Καναβό ο οποίος με την καθοδήγηση και τις συμβουλές του συνέβαλε στην ολοκλήρωση της Διπλωματικής μου Εργασίας, καθώς και τον επιβλέποντα καθηγητή κ. Αναστάσιο Σαλή.

Θα ήθελα επίσης να ευχαριστήσω θερμά την οικογένεια μου για την στήριξη που μου προσέφερε καθ' όλη τη διάρκεια των σπουδών μου.

Τέλος, ευχαριστώ όλους τους φίλους και συναδέλφους με τους οποίους συμπορευτήκαμε όλα αυτά τα χρόνια των σπουδών μας.

Περίληψη

Η αυτοκτονία έχει γίνει ένα σοβαρό ζήτημα κοινωνικής υγείας στη σύγχρονη κοινωνία. Ο αυτοκτονικός ιδεασμός είναι οι σκέψεις των ανθρώπων για αυτοκτονία. Πολλοί παράγοντες όπως η μακροχρόνια έκθεση σε αρνητικά συναισθήματα ή γεγονότα ζωής μπορούν να οδηγήσουν σε αυτοκτονικό ιδεασμό και απόπειρες αυτοκτονίας. Μεταξύ όλων των προσεγγίσεων της πρόληψης αυτοκτονίας, η έγκαιρη ανίχνευση αυτοκτονικού ιδεασμού είναι ένας από τους πιο αποτελεσματικούς τρόπους. Οι εξελίξεις των διαδικτυακών υπηρεσιών επικοινωνίας και κοινωνικής δικτύωσης παρέχουν επίσης μια πλατφόρμα για τους ανθρώπους να εκφράζουν τα βάσανα και τα συναισθήματά τους στον πραγματικό κόσμο, η οποία παρέχει μια πηγή για τον εντοπισμό αυτοκτονικών τάσεων και ιδεών. Η παρούσα διπλωματική εργασία διερευνά το διαδικτυακό κοινωνικό περιεχόμενο για έγκαιρη ανίχνευση αυτοκτονικού ιδεασμού.

Το περιεχόμενο που δημιουργείται από τους χρήστες, ειδικά το κείμενο που δημοσιεύεται από τους χρήστες, περιέχει πλούσιες πληροφορίες σχετικά με την κατάσταση των ανθρώπων και αντικατοπτρίζει τις νοητικές τους καταστάσεις. Σε αυτή τη διπλωματική, πρώτα θα έχουμε μια περιεκτική ανάλυση περιεχομένου για να ανακαλύψουμε γνώσεις από κείμενο που σχετίζεται με αυτοκτονίες και διαμορφώνει ένα σημείο αναφοράς για τη δυαδική ταξινόμηση του αυτοκτονικού ιδεασμού, συμπεριλαμβανομένης της χρήσης χαρακτηριστικών που βασίζονται στην εξαγωγή χαρακτηριστικών και των βαθιών νευρωνικών δικτύων. Οι λόγοι αυτοκτονίας είναι περίπλοκοι και οι αυτοκτονικοί παράγοντες διαφέρουν από άτομο σε άτομο. Για να ενσωματώσουμε παράγοντες αυτοκτονίας για κατανόηση πρόθεσης αυτοκτονίας, εξετάζουμε συναισθηματικές ενδείξεις και θέματα στις διαδικτυακές αναρτήσεις των ανθρώπων και προτείνουμε να αιτιολογήσουμε τις σχέσεις μεταξύ αυτών των παραγόντων και των αναρτήσεων με δίκτυα σχέσης προσοχής για την ακριβή ανίχνευση ιδεών αυτοκτονίας.

Συνολικά, ο έγκαιρος εντοπισμός αυτοκτονικού ιδεασμού απαιτείται επειγόντως για την πρόληψη της αυτοκτονίας. Αυτή η διπλωματική αναπτύσσει μεθόδους με ανάλυση περιεχομένου, τεχνική χαρακτηριστικών γνώσεων και τεχνικές βαθιάς μάθησης, όπως βαθιά νευρωνικά δίκτυα και προσεκτικά δίκτυα σχέσεων με την ελπίδα της αποτελεσματικής ανίχνευσης αυτοκτονικού ιδεασμού για την πρόληψη αυτοκτονιών και τη διάσωση της ανθρώπινης ζωής.

Λέξεις – Κλειδιά

Ανίχνευση αυτοκτονικού ιδεασμού, Τάσεις αυτοκτονίας, Διαδικτυακό περιεχόμενο, Μηχανική χαρακτηριστικών, Δίκτυα σχέσεων, Νευρωνικά δίκτυα, Μηχανική μάθηση, Βαθιά μάθηση, Εξόρυξη δεδομένων, Κοινωνικά δίκτυα.

IDENTIFICATION OF SUICIDAL TENDENCIES IN SOCIAL NETWORKS

George Myrisas

Abstract

Suicide has become a serious social health issue in the modern society. Suicidal ideation is people's thoughts about committing suicide. Many factors such as long-term exposure to negative feelings or life events can lead to suicidal ideation and suicide attempts. Among all the approaches of suicide prevention, early detection of suicidal ideation is one of the most effective ways. The advances of online communication and social networking services also provide a platform for people to express their sufferings and feelings in the real world, which provides a source for suicidal ideation detection. This thesis investigates the online social content for early detection of suicidal ideation.

User-generated content, especially text posted by users, contains rich information about people's status and reflects their mental states. In this thesis, we firstly have a comprehensive content analysis to discover knowledge from suicide-related text and preforms a benchmarking on binary classification of suicidal ideation including using feature extraction based classifiers and deep neural networks. The reasons of committing suicide are complicated, and suicidal factors vary from individuals. To incorporate suicidal factors for suicidal intention understanding, we consider sentimental clues and topics in people's posts and propose to reason the relations between those factors and posts with attention relation networks for fine-grained suicidal ideation detection.

Overall, early detection suicidal ideation is urgently in demand for suicide prevention. This thesis develops methods with content analysis, feature engineering, and deep learning techniques including deep neural networks, attentive relation and networks in the hope of using effective suicidal ideation detection to prevent suicide and save people's life.

Keywords

Suicidal ideation detection, Suicidal tendencies, online content, feature engineering, relation networks, Neural networks, Machine Learning, Deep Learning, Data Mining, Social Networks.

Πίνακας περιεχομένων

Κατάλογος Πινάκων

Κατάλογος Εικόνων / Σχημάτων

Κεφάλαιο 1

1. Εισαγωγή

Ζητήματα ψυχικής υγείας, όπως το άγχος και η κατάθλιψη, γίνονται όλο και πιο ανησυχητικά στη σύγχρονη κοινωνία, καθώς αποδεικνύονται ιδιαίτερα σοβαρά στις ανεπτυγμένες χώρες και τις αναδυόμενες αγορές. Οι σοβαρές ψυχικές διαταραχές χωρίς αποτελεσματική θεραπεία μπορούν να μετατραπούν σε αυτοκτονικό ιδεασμό ή ακόμη και απόπειρες αυτοκτονίας. Ορισμένες διαδικτυακές αναρτήσεις περιέχουν μεγάλο αριθμό αρνητικών πληροφοριών και δημιουργούν προβληματικά φαινόμενα, όπως cyberstalking και cyber bullying. Οι συνέπειες μπορεί να είναι σοβαρές και επικίνδυνες, καθώς τέτοιες κακές πληροφορίες εμπλέκονται συχνά σε κάποια μορφή κοινωνικής σκληρότητας, οδηγώντας σε φήμες ή ακόμη και σε ψυχικές βλάβες. Η έρευνα δείχνει ότι υπάρχει σχέση μεταξύ του διαδικτυακού εκφοβισμού και της αυτοκτονίας [1]. Τα θύματα που υπερεκτίθενται σε υπερβολικά αρνητικά μηνύματα ή γεγονότα μπορεί να γίνουν καταθλιπτικά και απελπισμένα, ακόμη χειρότερα, μερικά μπορεί να αυτοκτονήσουν.

1.1. Στατιστικές αυτοκτονιών

Κάθε χρόνο χιλιάδες άνθρωποι σε όλο τον κόσμο πέφτουν θύματα αυτοκτονίας, καθιστώντας την πρόληψη των αυτοκτονιών μια κρίσιμη παγκόσμια αποστολή δημόσιας υγείας. Σύμφωνα με μια έκθεση του ΠΟΥ(Παγκόσμιος οργανισμός Υγείας)[1], 1 στα 4 άτομα παγκοσμίως πάσχουν από ψυχικές διαταραχές σε κάποιο βαθμό. Χωρίς αποτελεσματική θεραπεία, σοβαρές ψυχικές διαταραχές αλλά χωρίς αποτελεσματική θεραπεία είναι πολύ πιθανό να στραφούν σε αυτοκτονία. Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας τα τελευταία 45 χρόνια οι αυτοκτονίες αυξήθηκαν σε όλο τον κόσμο κατά 60%. Ένας άνθρωπος αυτοκτονεί κάθε 40 δευτερόλεπτα στον κόσμο και περίπου 900.000 με 1.000.000 άνθρωποι σε όλο τον κόσμο αυτοκτονούν κάθε χρόνο (IASP) , με τις χώρες γύρω από τη Βαλτική, τις χώρες της πρώην

[1] Τα ποσοστά αυτοκτονιών, δεδομένα του Παγκόσμιου Παρατηρητηρίου Υγείας (GHO) το 2015 από τον ΠΟΥ, διατίθενται στη διεύθυνση http://www.who.int/ gho/m ental_health /icide_rates/el/

ΕΣΣΔ και την Ουγγαρία να έχουν τα περισσότερα θύματα. Μια έρευνα από το Suicide Prevention Australia (SPA) σε συνεργασία με το Πανεπιστήμιο της Νέας Αγγλίας παρείχε μια κατανόηση για την έκθεση και τον αντίκτυπο της αυτοκτονίας στην Αυστραλία βάσει μιας έρευνας με 3.000 ερωτηθέντες[2]. Ανέφερε ότι το 89% των ερωτηθέντων εκτέθηκαν σε τουλάχιστον μία απόπειρα αυτοκτονίας, το 85% εκτέθηκαν σε τουλάχιστον έναν θάνατο αυτοκτονίας και το 80% εκτέθηκαν τόσο σε απόπειρα αυτοκτονίας όσο και σε θάνατο. Όσον αφορά την πρόσβασή τους στην υγειονομική περίθαλψη έξι μήνες πριν από το θάνατο, υποστηρίχθηκε μόνο το 36% των ερωτηθέντων , με το 19% να μην υποστηρίζεται και το 26% δεν το γνωρίζουν, ανέφεραν με βάση τις γνώσεις τους για τη δική τους χρήση της υγειονομικής περίθαλψης. Σύμφωνα με προηγούμενη έκθεση στις ΗΠΑ, 2,2 εκατομμύρια άτομα που εκτιμήθηκαν σε εθνικό επίπεδο είχαν κάνει σχέδια αυτοκτονίας κατά την περίοδο 2008-2009 [2] . Στις ΗΠΑ περίπου 30.000 άτομα αυτοκτονούν κάθε χρόνο και ένας μεγάλος αριθμός ανθρώπων, ειδικά εφήβων, αναφέρθηκε ότι είχε τάσεις αυτοκτονίας. Η αυτοκτονία αποτελεί την 3η κυριότερη αιτία θανάτου Αμερικανών ηλικίας 15-24 ετών. Επίσης οι αυτοκτονίες αποτελούν το 12% των συνολικών θανάτων στην εφηβεία και 7-10% των εφήβων 15-17 ετών αποπειρώνται ≥1 φορά στη ζωή τους(Shaffer, Columbia University, 1996). Στην Ελλάδα, οι αυτοκτονίες από 2,5/100.000 που ήταν παλαιότερα, διπλασιάστηκαν την τελευταία δεκαετία και ανέρχονται κατά μέσο όρο σε περίπου τις 500 κάθε χρόνο. Το ποσοστό επιτυχημένων απόπειρών αυτοκτονίας μεταξύ ανδρών και γυναικών στην Ελλάδα είναι ♂:♀ 6:1 (ΕΣΥΕ). Το 12,5% των μαθητών (6,6% ♂ , 17,6% ♀) αναφέρουν απόπειρα αυτοκτονίας (Α. Τσίτσικα).

1.2. Παράγοντες αυτοκτονίας

Οι λόγοι αυτοκτονίας των ανθρώπων είναι περίπλοκοι. Τα άτομα με κατάθλιψη είναι πολύ πιθανό να αυτοκτονήσουν, αλλά πολλοί χωρίς κατάθλιψη μπορούν επίσης να έχουν αυτοκτονικές σκέψεις [3]. Οι Nock et al. [4] ανέφεραν παράγοντες επιπολασμού και αυτοκτονίας σε 17 χώρες και διαπίστωσαν ότι οι παράγοντες κινδύνου συνίστανται στο να είναι γυναίκες, νεότεροι, λιγότερο μορφωμένοι, άγαμοι και να έχουν προβλήματα ψυχικής

[2] Πρόληψη αυτοκτονίας Αυστραλία (SPA). Ευρήματα από την έκθεση και τον αντίκτυπο της αυτοκτονίας στην έρευνα της Αυστραλίας, διατίθενται στη διεύθυνση https://www.suicidepreventionaust.org/exposure-and-impact-survey. Ανακτήθηκε το Σεπτέμβριο του 2018.

υγείας. Σύμφωνα με το αμερικανικό ίδρυμα για την πρόληψη αυτοκτονιών (AFSP), οι παράγοντες αυτοκτονίας εμπίπτουν σε τρεις κατηγορίες: παράγοντες υγείας, περιβαλλοντικοί παράγοντες και ιστορικοί παράγοντες [5]. Οι Ferrari et al. [6] διαπίστωσαν ότι τα ζητήματα ψυχικής υγείας και οι διαταραχές από τη χρήση ναρκωτικών και ψυχοτρόπων ουσιών αποτελούν σημαντικούς παράγοντες οι οποίοι μπορεί να οδηγήσουν σε αυτοκτονία. Οι O'Connor και Nock [7] διεξήγαγαν μια εμπεριστατωμένη ανασκόπηση σχετικά με την ψυχολογία της αυτοκτονίας, και συνόψισαν τους ψυχολογικούς κινδύνους ως προσωπικότητα και ατομικές διαφορές, γνωστικούς παράγοντες, κοινωνικούς παράγοντες και αρνητικά συμβάντα ζωής.

1.3. Αυτοκτονία και Διαδίκτυο

Λόγω της προόδου των κοινωνικών μέσων μαζικής ενημέρωσης και της ανωνυμίας στο διαδίκτυο, ένας αυξανόμενος αριθμός ατόμων στρέφεται στο να αλληλοεπιδρά με άλλους ανθρώπους στο Διαδίκτυο. Τα διαδικτυακά κανάλια επικοινωνίας γίνονται ένας νέος τρόπος για τους ανθρώπους να εκφράσουν τις αισθήσεις τους , τα βάσανα και τις τάσεις αυτοκτονίας. Ως εκ τούτου, τα διαδικτυακά κανάλια έχουν αρχίσει φυσικά να λειτουργούν ως εργαλείο επιτήρησης για αυτοκτονικό ιδεασμό και η εξόρυξη κοινωνικού περιεχομένου μπορεί να βελτιώσει την πρόληψη αυτοκτονιών [8]. Επιπλέον, εμφανίζονται παράξενα κοινωνικά φαινόμενα, π.χ. , διαδικτυακές κοινότητες που καταλήγουν σε συμφωνία για αυτοακρωτηριασμό και αυτοκτονία. Για παράδειγμα, ένα φαινόμενο κοινωνικού δικτύου το οποίο αποκαλείται «Blue Whale Game»[3] , το 2016 χρησιμοποιεί πολλές δοκιμασίες (όπως αυτοτραυματισμός) και οδηγεί τα μέλη του παιχνιδιού να αυτοκτονήσουν στο τέλος. Η αυτοκτονία είναι ένα κρίσιμο κοινωνικό ζήτημα και κοστίζει χιλιάδες ζωές κάθε χρόνο. Επομένως, είναι απαραίτητο να εντοπιστούν έγκαιρα τα άτομα που έχουν αυτοκτονικές τάσεις και να αποφευχθεί η αυτοκτονία τους προτού αυτά βάλουν τέλος στη ζωή τους. Η έγκαιρη ανίχνευση και θεραπεία θεωρούνται ως οι πιο αποτελεσματικοί τρόποι για την αποτροπή πιθανών προσπαθειών αυτοκτονίας.

[3] https://thesun.co.uk/news/worldnews/3003805

1.4. Ανίχνευση αυτοκτονικού ιδεασμού

Τα πιθανά θύματα με αυτοκτονικό ιδεασμό μπορούν να εκφράσουν τις σκέψεις τους για αυτοκτονία με τη μορφή φευγαλέων σκέψεων, σχεδίων αυτοκτονίας και παιχνιδιού ρόλων. Η ανίχνευση αυτοκτονικού ιδεασμού είναι να ανακαλύψουμε αυτές τις επικίνδυνες προθέσεις ή συμπεριφορές πριν χτυπήσει η τραγωδία. Οι λόγοι της αυτοκτονίας είναι περίπλοκοι και αποδίδονται σε μια πολύπλοκη αλληλεπίδραση πολλών παραγόντων [7]. Για την ανίχνευση του αυτοκτονικού ιδεασμού, πολλοί ερευνητές πραγματοποίησαν ψυχολογικές και κλινικές μελέτες [9] και ταξινόμησαν τις απαντήσεις των ερωτηματολογίων [10]. Με βάση τα δεδομένα των κοινωνικών μέσων, η τεχνητή νοημοσύνη (AI) και οι τεχνικές μηχανικής μάθησης μπορούν να προβλέψουν την πιθανότητα αυτοκτονίας των ανθρώπων [11], η οποία ανοίγει το δρόμο για έγκαιρη παρέμβαση. Η ανίχνευση σε κοινωνικό περιεχόμενο εστιάζεται στη μηχανική χαρακτηριστικών [12,13], στην ανάλυση συναισθημάτων [14,15] και στη βαθιά μάθηση [16–18].

Οι κινητές τεχνολογίες έχουν μελετηθεί και εφαρμοστεί στην πρόληψη αυτοκτονιών, για παράδειγμα, η εφαρμογή κινητής επέμβασης αυτοκτονίας iBobbly [19] που αναπτύχθηκε από το Black Dog Institute[4] . Πολλά άλλα εργαλεία πρόληψης αυτοκτονιών ενσωματωμένα σε υπηρεσίες κοινωνικής δικτύωσης έχουν επίσης αναπτυχθεί, συμπεριλαμβανομένου του Samaritans Radar[5] και Woebot[6]. Το πρώτο ήταν ένα πρόσθετο Twitter για την παρακολούθηση ανησυχητικών αναρτήσεων, οι οποίες διακόπηκαν εξαιτίας ζητημάτων απορρήτου. Το τελευταίο είναι ένα Facebook chatbot που βασίζεται σε τεχνικές γνωστικής συμπεριφορικής θεραπείας και επεξεργασίας φυσικής γλώσσας (NLP) για την ανακούφιση της κατάθλιψης και του άγχους των ανθρώπων.

Είναι αναπόφευκτο ότι η εφαρμογή προηγμένων τεχνολογιών AI για την ανίχνευση αυτοκτονικών ιδεών συνοδεύεται από ζητήματα απορρήτου [20] και ηθικά ζητήματα [21]. Οι Linthicum et al. [22] υπέβαλαν τρία ηθικά ζητήματα, συμπεριλαμβανομένης της επιρροής της προκατάληψης στους αλγορίθμους μηχανικής μάθησης, την πρόβλεψη για την ώρα της αυτοκτονικής πράξης, και ηθικά και νομικά ζητήματα που εγείρονται από ψευδείς θετικές και

[4] https://blackdoginstitute.org.au/research/digital-dog/programs/ibobbly-app

[5] https://samaritans.org/about-samaritans/research-policy/internet-suicide/samaritans-radar

[6] https: //woebot.i o

ψευδείς αρνητικές προβλέψεις. Δεν είναι εύκολο να απαντήσουμε σε ηθικές ερωτήσεις για την τεχνητή νοημοσύνη, καθώς απαιτούν αλγόριθμους για την επίτευξη ισορροπίας μεταξύ ανταγωνιστικών αξιών, ζητημάτων και ενδιαφερόντων [20].

Μία πιθανή προσέγγιση για την αποτελεσματική πρόληψη της αυτοκτονίας είναι η έγκαιρη ανίχνευση αυτοκτονικών ιδεών για αποτελεσματική παρέμβαση. Έτσι, η ανάπτυξη μεθόδων ανίχνευσης αυτοκτονικών ιδεών γίνεται μια σημαντική αποστολή. Ωστόσο, εξακολουθούν να υπάρχουν αρκετές προκλήσεις όπως:

- Υπάρχει περιορισμένος αριθμός κριτηρίων για την εκπαίδευση και την αξιολόγηση της ανίχνευσης αυτοκτονικών ιδεών.
- Τα δεδομένα κειμένου είναι θορυβώδη για αποτελεσματική ανίχνευση αυτοκτονικών ιδεών.
- Το κείμενο με αυτοκτονικό ιδεασμό και κείμενο με μικρές ψυχικές διαταραχές, έχουν παρόμοιες χρήσεις γλώσσας, καθιστώντας δύσκολη την κατανόηση της πρόθεσης αυτοκτονίας.
- Σε ορισμένα σενάρια όπως το chat room, η απομόνωση των δεδομένων βλάπτει την απόδοση των εποπτευόμενων μοντέλων μάθησης .

Αυτή η διπλωματική εργασία επικεντρώνεται σε τεχνικές μηχανικής μάθησης, ιδίως σε μοντέλα βαθιάς μάθησης για αποτελεσματική ανίχνευση αυτοκτονικών ιδεών σε διαδικτυακό κοινωνικό περιεχόμενο. Σκοπεύει να επιλύσει δύο καθήκοντα, δηλαδή, τη συγκριτική αξιολόγηση για την ανίχνευση αυτοκτονικού ιδεασμού, τη λεπτομερή ανίχνευση αυτοκτονικών ιδεών λαμβάνοντας υπόψη τους παράγοντες κινδύνου αυτοκτονίας

1.5. Κίνητρο και στόχοι της παρούσας έρευνας

Σαν Αξιωματικός της Ελληνικής Αστυνομίας, κατά τη διάρκεια εκτέλεσης των καθηκόντων μου έχω επιληφθεί σε πάρα πολλά περιστατικά όπου άνθρωποι είτε απειλούσαν να αυτοκτονήσουν είτε τελικά πραγματοποίησαν απόπειρα ή τετελεσμένη αυτοκτονία. Το γεγονός αυτό πάντα με σόκαρε, με έκανε να σκέφτομαι τον πόνο και τη θλίψη που προκαλούνταν στον οικογενειακό και κοινωνικό περίγυρο του θύματος αυτοκτονίας και με

έβαλε σε μια διαδικασία σκέψης για το τι θα μπορούσαμε να κάνουμε σαν αστυνομικοί και σαν οργανωμένη πολιτεία, ώστε να μη βρισκόμαστε προ τετελεσμένων γεγονότων, αλλά να υπήρχε κάποιος τρόπος ώστε να μπορούμε να έρθουμε σε επαφή με τα άτομα τα οποία έχουν αυτοκτονικές τάσεις και να τους βοηθήσουμε να ξεπεράσουν αυτό τον σκόπελο, πριν φτάσουν στο απονενοημένο διάβημα. Γυρίζοντας πίσω αρκετά χρόνια πριν, σε ένα από τα μεγαλύτερα επιστημονικά συνέδρια που διοργανώθηκαν στην Αμερική για το ζήτημα των αυτοκτονιών, μου έκανε ιδιαίτερα μεγάλη εντύπωση η οπτική απεικόνιση της έννοιας «πρόληψη της αυτοκτονίας». Η εικόνα ενός αυτιού! Σήμερα, αρκετά χρόνια μετά, και μέσα από μια πολύχρονη διαδρομή επιστημονικής διερεύνησης, μελέτης, καταγραφής, προσέγγισης, από την αφετηρία της πρόληψης έως την αποκατάσταση και ό,τι άλλο εννοεί ή επιβάλλει η αντιμετώπιση του προβλήματος της αυτοκτονικότητας, καταλήγουμε και εμείς ότι η εικόνα ενός αυτιού είναι η πιο άμεση, σύντομη και κατανοητή οπτική απεικόνιση σε ότι αφορά την προσπάθεια πρόληψης της αυτοκτονίας. Τα άτομα που συμβάλλουν στην πρόληψη της αυτοκτονίας θα πρέπει να είναι σε θέση να αντιληφθούν, να κατανοήσουν, να εκπαιδευτούν, να ενεργήσουν ή απλά να ακούσουν υποστηρικτικά οποιονδήποτε διπλανό τους και να αισθάνονται ότι βρίσκεται σε αδιέξοδο με αυτοκτονικές ιδέες και συμπεριφορές. Οπότε σύμφωνα με τα παραπάνω ο έγκαιρος εντοπισμός των αυτοκτονικών τάσεων στα on-line κοινωνικά δίκτυα θα μπορούσε να σώσει από την αυτοκτονία έναν πάρα πολύ μεγάλο αριθμό συνανθρώπων μας.

1.6. Εισαγωγικοί Ορισμοί

BIG DATA

Με τον όρο Big Data περιγράφονται τα σύνολα δεδομένων που είναι τόσο μεγάλα και περίπλοκα που οι παραδοσιακοί τρόποι και εφαρμογές επεξεργασίας δεν επαρκούν για την ανάλυση και τον χειρισμό τους. Οι προκλήσεις συμπεριλαμβάνουν την ανάλυση, τη σύλληψη, επιμέλεια δεδομένων, αναζήτηση, κοινή χρήση, αποθήκευση, μεταφορά, απεικόνιση, και το ζήτημα της ιδιωτικότητας της πληροφορίας. Ο όρος αυτός αναφέρεται συνήθως στην χρήση predictive analytics ή άλλων εξελιγμένων δεδομένων εξαγωγής πληροφορίας και αξίας από δεδομένα και σπάνια στα δεδομένα καθ’ αυτά . Η ακρίβεια στα Big Data, μπορεί να οδηγήσει στη λήψη πιο σωστών αποφάσεων και κατ’ επέκταση σε καλύτερης ποιότητας αποφάσεις, που οδηγούν σε καλύτερη λειτουργική απόδοση, μείωση του κόστους και του ρίσκου.

ΚΟΙΝΩΝΙΚΑ ΔΙΚΤΥΑ – SOCIAL MEDIA

Ο όρος Social media είναι δύσκολο να οριστεί διότι με την έννοια αυτή , πολλές φορές αναφερόμαστε σε μια πλατφόρμα, άλλες φορές σε ένα λογισμικό και άλλες σε μια δραστηριότητα. Επιπλέον συνεχώς εξελίσσεται και αλλάζει, με αποτέλεσμα να προστίθενται συνέχεια καινούργιοι ορισμοί και έννοιες. Παρόλα αυτά υπάρχουν πολλοί ορισμοί που αποσαφηνίζουν την έννοια αυτή καθ΄ αυτή. Ένας πολύ σημαντικός ορισμός είναι αυτός των Kaplan και Haelnlein, οι οποίοι ορίζουν τα μέσα κοινωνικής δικτύωσης, ως μια ομάδα εφαρμογών βασισμένες στο διαδίκτυο, φτιαγμένες στα ιδεολογικά και τεχνολογικά θεμέλια του Web 2.0. Οι συγκεκριμένες εφαρμογές επιτρέπουν τη δημιουργία και την ανταλλαγή περιεχομένου από τους χρήστες. Ένας άλλος ορισμός είναι ότι τα μέσα κοινωνικής δικτύωσης αποτελούν εργαλεία που αυξάνουν την ικανότητά μας να μοιραζόμαστε, να συνεργαζόμαστε και να πραγματοποιούμε συλλογικές δράσεις. Ένας επιπλέον ορισμός είναι των Lon Safko και David K. Brake στο βιβλίο τους The Social Media Bible _ Tactics, Tools, and Strategies for Business Success, όπου χαρακτηριστικά λένε ότι τα Social Media αναφέρονται σε δραστηριότητες, πρακτικές και συμπεριφορές ανάμεσα σε κοινότητες ανθρώπων που συγκεντρώνονται στο διαδίκτυο, με σκοπό να μοιραστούν πληροφορίες, γνώσεις και απόψεις χρησιμοποιώντας μέσα συζήτησης (conversational media). Conversational media είναι εφαρμογές που καθιστούν δυνατή τη δημιουργία και την εύκολη μετάδοση περιεχομένου με τη μορφή λέξεων, εικόνων, βίντεο ,και ηχογραφημένων μηνυμάτων. Συμπερασματικά ως Κοινωνικά Δίκτυα (Social Media) χαρακτηρίζονται σε γενικές γραμμές τα πολλά και σχετικά φθηνά και ευρέως προσβάσιμα ηλεκτρονικά εργαλεία που δίνουν την δυνατότητα σε οποιονδήποτε να δημοσιεύσει πληροφορίες, να έχει πρόσβαση σε πληροφορίες, να συνεργαστεί για ένα κοινό σκοπό ή να δημιουργήσει σχέσεις. Κάποια από τα πιο δημοφιλή κοινωνικά δίκτυα είναι το Facebook, το Twitter, το Instagram και το Reddit.

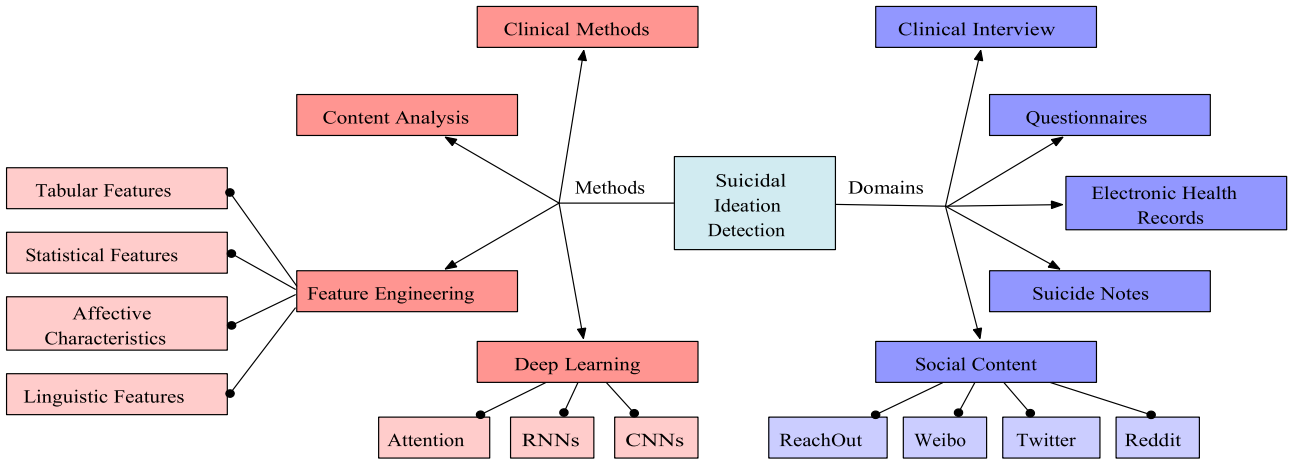
Κεφάλαιο 2

2. Βιβλιογραφική Ανασκόπηση

Η ΤΝ(Τεχνητή Νοημοσύνη)-AI(Artificial Intelligence) έχει εφαρμοστεί για την επίλυση πολλών προκλήσεων κοινωνικών προβλημάτων. Η ανίχνευση του αυτοκτονικού ιδεασμού με τεχνικές τεχνητής νοημοσύνης είναι μια από τις πιθανές εφαρμογές για κοινωνικό καλό και πρέπει να αντιμετωπιστεί για τη σημαντική βελτίωση της ευημερίας των ανθρώπων. Το κεφάλαιο εξετάζει τις μεθόδους ανίχνευσης επαρκούς ιδεολογίας από την οπτική γωνία της ΤΝ και της μηχανικής μάθησης και συγκεκριμένων εφαρμογών τομέα με κοινωνικό αντίκτυπο. Η κατηγοριοποίηση από αυτές τις δύο προοπτικές φαίνεται στο Σχήμα 1. Ο ορισμός της ανίχνευσης αυτοκτονικού ιδεασμού είναι η περιγραφή στον ορισμό 1.

Ορισμός 1: (Ανίχνευση αυτοκτονικών ιδεών) . Δοθέντων πινάκων δεδομένων ενός ατόμου ή περιεχομένου κειμένου που έχει γραφτεί από ένα άτομο, η ανίχνευση αυτοκτονικού ιδεασμού είναι να προσδιοριστεί εάν το άτομο έχει αυτοκτονικό ιδεασμό ή σκέψεις.

Παρουσιάζουμε και συζητάμε τόσο την κλασική ανάλυση περιεχομένου όσο και τις πρόσφατες τεχνικές μηχανικής μάθησης, καθώς και την εφαρμογή τους σε ερωτηματολόγια, δεδομένα EHR, σημειώσεις αυτοκτονίας και διαδικτυακό κοινωνικό περιεχόμενο.



Σχήμα 1:Η κατηγοριοποίηση της ανίχνευσης αυτοκτονικών ιδεών: μέθοδοι και τομείς

2.1. Μέθοδοι και Κατηγοριοποίηση

Η ανίχνευση αυτοκτονιών έχει τραβήξει την προσοχή πολλών ερευνητών λόγω του αυξανόμενου ποσοστού αυτοκτονιών τα τελευταία χρόνια και έχει μελετηθεί εκτενώς από πολλές οπτικές γωνίες. Οι ερευνητικές τεχνικές που χρησιμοποιήθηκαν για την εξέταση της αυτοκτονίας καλύπτουν επίσης πολλά πεδία και μεθόδους, για παράδειγμα , κλινικές μεθόδους με αλληλεπίδραση ασθενούς-κλινικής [9] και αυτόματη ανίχνευση από περιεχόμενο που δημιουργείται από χρήστες (κυρίως κείμενο) [12,17]. Οι τεχνικές μηχανικής μάθησης εφαρμόζονται ευρέως για την αυτόματη ανίχνευση αυτοκτονικών τάσεων.

Η παραδοσιακή ανίχνευση αυτοκτονίας βασίζεται σε κλινικές μεθόδους, συμπεριλαμβανομένων των αυτοαναφορών και των προσωπικών συνεντεύξεων. Οι Venek et al. [9] σχεδίασαν ένα πανταχού παρόν ερωτηματολόγιο για την αξιολόγηση των αυτοκτονικών κινδύνων, και εφάρμοσαν έναν ιεραρχικό ταξινομητή στην απόκριση των ασθενών για να προσδιορίσει τις αυτοκτονικές τους έννοιες. Μέσω προσωπικής αλληλεπίδρασης, μπορούν να χρησιμοποιηθούν λεκτικές και ακουστικές πληροφορίες. Ο Scherer [23] διερεύνησε τα χαρακτηριστικά της προσωδιακής ομιλίας και την ποιότητα της φωνής σε μια δυαδική συνέντευξη για να εντοπίσει αυτοκτονικούς και μη αυτοκτονικούς ανηλίκους. Άλλες κλινικές μέθοδοι εξετάζουν τον καρδιακό ρυθμό ηρεμίας από μετατραπέντα αισθητήρια σήματα [24], ταξινομούν λειτουργικές απεικονίσεις μαγνητικού συντονισμού βασισμένες σε νευρικές αναπαραστάσεις των λέξεων που σχετίζονται με το θάνατο και τη ζωή [25], και υποκινητές που σχετίζονται με συμβάντα που μετατρέπονται από σήματα EEG [26]. Μια άλλη πτυχή της κλινικής θεραπείας είναι η κατανόηση της ψυχολογίας πίσω από την αυτοκτονική συμπεριφορά [7]. Αυτό, ωστόσο, βασίζεται σε μεγάλο βαθμό στη γνώση του γιατρού και στην αλληλεπίδραση, πρόσωπο με πρόσωπο. Οι κλίμακες εκτίμησης κινδύνου αυτοκτονίας με κλινική συνέντευξη μπορούν να αποκαλύψουν πληροφοριακά στοιχεία για την πρόβλεψη αυτοκτονίας [27]. Οι Tan et al. [28] πραγματοποίησαν μια συνέντευξη και μια μελέτη έρευνας στο Weibo, μια υπηρεσία που μοιάζει με το Twitter στην Κίνα, για να διερευνήσουν τη δέσμευση των απόπειρών αυτοκτονίας με επέμβαση μέσω άμεσων μηνυμάτων.

2.1.1. Ανάλυση Περιεχομένου

Οι δημοσιεύσεις των χρηστών στις ιστοσελίδες κοινωνικής δικτύωσης αποκαλύπτουν πλούσιες πληροφορίες και τις προτιμήσεις γλώσσας τους. Μέσω της διερευνητικής ανάλυσης δεδομένων, το περιεχόμενο που δημιουργείται από τον χρήστη μπορεί να έχει μια εικόνα για τη χρήση της γλώσσας και τις γλωσσικές ενδείξεις των απόπειρών αυτοκτονίας. Η σχετική ανάλυση περιλαμβάνει φιλτράρισμα με βάση λεξικά, στατιστικά γλωσσικά χαρακτηριστικά και θεματική μοντελοποίηση σε θέσεις που σχετίζονται με αυτοκτονίες.

Το λεξικό και το λεξικό λέξεων-κλειδιών που σχετίζονται με αυτοκτονίες έχουν δημιουργηθεί με μη αυτόματο τρόπο για να επιτρέπουν το φιλτράρισμα λέξεων-κλειδιών [29, 30] και το φιλτράρισμα φράσεων [31]. Οι λέξεις – κλειδιά και οι φράσεις που σχετίζονται με τις αυτοκτονίες περιλαμβάνουν «σκότωσε», «αυτοκτονία», «νιώθω μόνος», «κατάθλιψη» και «κόβοντας τον εαυτό μου». Οι Vioules et al. [5] δημιούργησαν ένα λεξικό συμπτωμάτων αμοιβαίας πληροφόρησης με χρήση σχολιασμένου συνόλου δεδομένων Twitter. Ο Gunn και ο Lester [32] ανέλυσαν αναρτήσεις από το Twitter 24 ώρες πριν από το θάνατο ενός ατόμου το οποίο έχει διαπράξει αυτοκτονία. Οι Coppersmith et al. [33] ανέλυσαν τη χρήση γλώσσας των δεδομένων από την ίδια πλατφόρμα. Οι αυτοκτονικές σκέψεις μπορεί να περιλαμβάνουν έντονα αρνητικά συναισθήματα, άγχος και απελπισία ή άλλους κοινωνικούς παράγοντες όπως η οικογένεια και οι φίλοι. Ο Ji et al. [17] πραγματοποίησαν οπτικοποίηση νέφους λέξεων και θεματική μοντελοποίηση για περιεχόμενο που σχετίζεται με αυτοκτονίες και διαπίστωσαν ότι η συζήτηση που σχετίζεται με αυτοκτονίες καλύπτει τόσο προσωπικά όσο και κοινωνικά ζητήματα. Οι Colombo et al. [34] ανέλυσαν τα γραφικά χαρακτηριστικά της συνδεσιμότητας και της επικοινωνίας στο κοινωνικό δίκτυο Twitter. Οι Coppersmith et al. [35] παρείχαν μια διερευνητική ανάλυση σχετικά με τα γλωσσικά πρότυπα και τα συναισθήματα στο Twitter. Άλλες μέθοδοι και τεχνικές περιλαμβάνουν την ανάλυση των Google Trends για την παρακολούθηση του κινδύνου αυτοκτονίας [36] , την ανίχνευση του περιεχομένου των κοινωνικών μέσων και την ανάλυση προτύπων ομιλίας [37], την αξιολόγηση της απόκλισης απόκρισης μέσω γλωσσικών ενδείξεων [38], την άνθρωπος-μηχανή υβριδική μέθοδο για την ανάλυση της επίδρασης της γλώσσας κοινωνικής υποστήριξης στον κίνδυνο αυτοκτονικών ιδεών [39].

2.1.2. Μηχανική Χαρακτηριστικών

Ο στόχος της ταξινόμησης αυτοκτονιών βάσει κειμένου είναι να προσδιορίσει εάν οι υποψήφιοι, μέσω των αναρτήσεων τους, έχουν αυτοκτονικές ιδέες. Οι μέθοδοι της μηχανικής μάθησης και της επεξεργασίας φυσικής γλώσσας(NLP) έχουν επίσης εφαρμοστεί σε αυτόν τον τομέα.

Χαρακτηριστικά Πίνακα

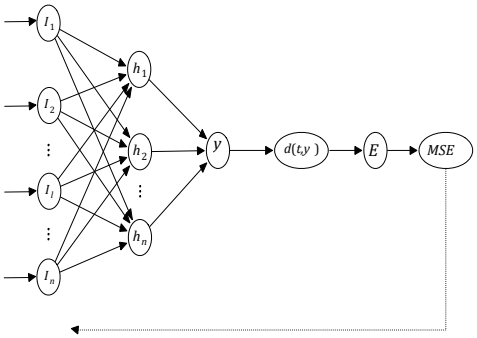
Τα δεδομένα πίνακα για την ανίχνευση αυτοκτονικού ιδεασμού αποτελούνται από απαντήσεις στο ερωτηματολόγιο και δομημένες στατιστικές πληροφορίες που εξάγονται από ιστότοπους. Τέτοια δομημένα δεδομένα μπορούν να χρησιμοποιηθούν άμεσα ως χαρακτηριστικά για ταξινόμηση ή παλινδρόμηση. Οι Masuda et al. [40] Εφάρμοσαν λογιστική παλινδρόμηση για την ταξινόμηση των ομάδων αυτοκτονίας και ελέγχου με βάση τα χαρακτηριστικά των χρηστών και των μεταβλητών κοινωνικής συμπεριφοράς και διαπίστωσαν ότι μεταβλητές όπως ο αριθμός της κοινότητας, ο τοπικός συντελεστής ομαδοποίησης και η ομοφυλοφιλία έχουν μεγαλύτερη επιρροή στην αυτοκτονική ιδεολογία σε ένα SNS της Ιαπωνίας. Ο Chattopadhyay [41] εφάρμοσε το Pierce Suicidal Intent Scale (PSIS) για να αξιολογήσει τους παράγοντες αυτοκτονίας και πραγματοποίησαν μια ανάλυση παλινδρόμησης. Τα ερωτηματολόγια λειτουργούν ως καλή πηγή χαρακτηριστικών πίνακα. Οι Delgado-Gomez et al. [42] χρησιμοποίησαν το διεθνές ερωτηματολόγιο διαλογής για την εξέταση προσωπικής διαταραχής και την κλίμακα αξιολόγησης της κοινωνικής προσαρμογής Holmes-Rahe. Ο Chattopadhyay [43] πρότεινε να εφαρμόσει ένα νευρωνικό δίκτυο πολλαπλών στρώσεων τροφοδοσίας όπως φαίνεται στο Σχήμα 2α για να ταξινομήσει τους δείκτες πρόθεσης αυτοκτονίας σύμφωνα με την κλίμακα πρόθεσης αυτοκτονίας του Beck.

Γενικά χαρακτηριστικά κειμένου

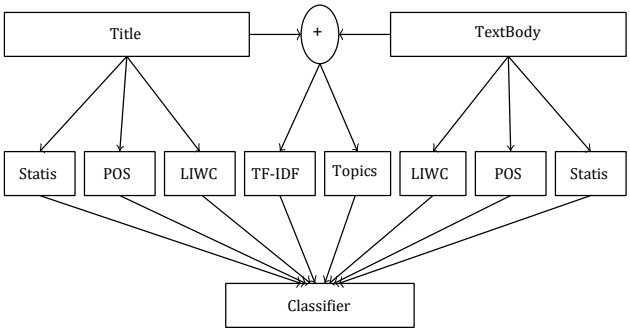
Μια άλλη κατεύθυνση της μηχανικής χαρακτηριστικών είναι η εξαγωγή χαρακτηριστικών από μη δομημένο κείμενο. Τα κύρια χαρακτηριστικά αποτελούνται από χαρακτηριστικά N-gram, χαρακτηριστικά βασισμένα στη γνώση, συντακτικά χαρακτηριστικά, χαρακτηριστικά περιβάλλοντος και χαρακτηριστικά ειδικής κατηγορίας [44]. Οι Abboute et al. [45] δημιούργησαν ένα σύνολο λέξεων-κλειδιών για εξαγωγή χαρακτηριστικών λεξιλογίου σε εννέα θέματα αυτοκτονίας. Οι Okhapkina et al. [46] δημιούργησαν ένα λεξικό όρων που

σχετίζονται με περιεχόμενο αυτοκτονίας και εισήγαγαν πίνακες συχνότητας όρων-αντίστροφης συχνότητας εγγράφου (TF-IDF) για μηνύματα και μια μοναδική αποσύνθεση τιμής (SVD) για πίνακες. Ο Mulholland και ο Quinn [47] εξήγαγαν λεξιλόγια και συντακτικά χαρακτηριστικά για να δημιουργήσουν έναν ταξινομητή για να προβλέψουν την πιθανότητα αυτοκτονίας ενός στιχουργού. Οι Huang et al. [48] δημιούργησαν ένα ψυχολογικό λεξικό λεξικολογικών λεξικών επεκτείνοντας το HowNet (μια λογική συλλογή λέξεων) και χρησιμοποίησαν μηχανές διανυσμάτων υποστήριξης (SVM) για την ανίχνευση των κυβερνοαυτοκτονιών(αυτοκτονιών μέσω διαδικτύου) στα κινεζικά microblogs. Η μοντελοποίηση θέματος [49] ενσωματώνεται με τις τεχνικές μηχανικής μάθησης για τον εντοπισμό αυτοκτονίας στο Sina Weibo. Οι Ji et al. [17] εξήγαγαν διάφορα ενημερωτικά σύνολα χαρακτηριστικών, συμπεριλαμβανομένων στατιστικών, συντακτικών, γλωσσικών ερευνών και καταμέτρησης λέξεων (LIWC), ενσωμάτωσης λέξεων και χαρακτηριστικών θέματος και στη συνέχεια, τοποθετήσαν το εξαγόμενο χαρακτηριστικό σε ταξινομητές όπως φαίνεται στην Εικόνα 2β, όπου συγκρίνονται τέσσερις παραδοσιακοί εποπτευόμενοι ταξινομητές. Οι Shing et al. [13] εξήγαγαν πολλά χαρακτηριστικά όπως σάκο λέξεων (BoWs), empath, αναγνωσιμότητα, συντακτικά χαρακτηριστικά, οπίσθια μοντέλα θεμάτων, ενσωματώσεις λέξεων, γλωσσική έρευνα και καταμέτρηση λέξεων, χαρακτηριστικά συναισθημάτων και λεξικό ψυχικών ασθενειών.

Τα μοντέλα για την ανίχνευση αυτοκτονικού ιδεασμού με μηχανική χαρακτηριστικών περιλαμβάνουν μηχανές διανυσμάτων υποστήριξης (SVM) [44], τεχνητά νευρωνικά δίκτυα (ANN) [50] και υποθετικό τυχαίο πεδίο (CRF) [51]. Οι Tai et al. [50] επέλεξαν διάφορα χαρακτηριστικά όπως το ιστορικό ιδεών αυτοκτονίας και συμπεριφοράς αυτοτραυματισμού, η θρησκευτική πίστη, η οικογενειακή κατάσταση, το ιστορικό ψυχικών διαταραχών των υποψηφίων και η οικογένειά τους. Οι Pestian et al. [52] σύγκριναν την απόδοση διαφορετικών τεχνικών πολλαπλών παραλλαγών με χαρακτηριστικά γνωρίσματα λέξεων, POS, έννοιες και βαθμολογία αναγνωσιμότητας. Ομοίως , Οι Ji et al. [17] σύγκριναν τέσσερις μεθόδους ταξινόμησης, της λογιστικής παλινδρόμησης (logistic regression), του τυχαίου δάσους (random forest), του δέντρου απόφασης βαθμιαίας κλίσης (gradient boosting decision tree) και της extreme gradient boosting (XGBoost). Οι Braithwaite et al. [53] συμπέραναν ότι, οι αλγόριθμοι μηχανικής μάθησης μπορούν αποτελεσματικά να αναγνωρίσουν τον υψηλό κίνδυνο αυτοκτονίας.



(α) Νευρωνικό δίκτυο με μηχανική χαρακτηριστικών



(β) Ταξινομητής με μηχανική χαρακτηριστικών

Σχήμα 2: Απεικονίσεις μεθόδων με μηχανική χαρακτηριστικών

Συναισθηματικά χαρακτηριστικά

Τα συναισθηματικά χαρακτηριστικά είναι μια από τις πιο διακριτές διαφορές μεταξύ εκείνων που επιχειρούν αυτοκτονία και φυσιολογικών ατόμων, η οποία έχει τραβήξει μεγάλη προσοχή τόσο από τους επιστήμονες υπολογιστών όσο και από τους ερευνητές ψυχικής υγείας. Για να ανιχνεύσουν τα συναισθήματα στις σημειώσεις αυτοκτονίας, οι Liakata et al. [51] χρησιμοποίησαν χειροκίνητες κατηγορίες συναισθημάτων όπως θυμός, θλίψη, ελπίδα , ευτυχία / γαλήνη, φόβος, υπερηφάνεια, κακοποίηση και συγχώρεση. Οι Wang et al. [44] χρησιμοποίησαν συνδυασμένα χαρακτηριστικά τόσο των πραγματικών (2 κατηγοριών) όσο και των συναισθηματικών πτυχών (13 κατηγορίες) για να ανακαλύψουν μια λεπτομερή ανάλυση συναισθημάτων. Ομοίως, οι Pestian et al. [52] εντοπίσαν συναισθήματα κακοποίησης, θυμού, κατηγορίας, φόβου, ενοχής, απελπισίας, θλίψης, συγχώρεσης, ευτυχίας, γαλήνης, ελπίδας, αγάπης, υπερηφάνειας, ευγνωμοσύνης, οδηγιών και πληροφοριών. Οι Ren et al. [14] πρότειναν ένα σύνθετο μοντέλο θεμάτων συναισθημάτων και το εφάρμοσαν για να αναλύσουν συσσωρευμένα συναισθηματικά χαρακτηριστικά σε ιστολόγια αυτοκτονίας και να εντοπίσουν αυτοκτονικές προθέσεις από μια ροή ιστολογίου(blog stream). Συγκεκριμένα, οι συγγραφείς μελέτησαν συσσωρευμένα συναισθηματικά χαρακτηριστικά, όπως συσσώρευση συναισθημάτων, συνδιαλλαγή συναισθημάτων και μετάβαση συναισθημάτων μεταξύ οκτώ βασικών συναισθημάτων χαράς, αγάπης, προσδοκίας, έκπληξης, άγχους, θλίψης, θυμού και μίσους με μια ένταση πέντε επιπέδων.

2.1.3. Τεχνητή Νοημοσύνη

Ο όρος **Τεχνητή Νοημοσύνη (ΤΝ – Artificial Intelligence)** επινοήθηκε από τον John McCarthy για να δώσει όνομα σε έναν τομέα ερευνών που είχαν αρχίσει αρκετό καιρό πριν.

Η Τεχνητή Νοημοσύνη αποτελεί τομέα της πληροφορικής που ασχολείται με την υλοποίηση «ευφυών» υπολογιστικών συστημάτων, τα οποία εκτελούν εργασίες που προσομοιάζουν με την ανθρώπινη συμπεριφορά, όπως μάθηση, εξαγωγή συμπερασμάτων, λήψη αποφάσεων και επίλυση προβλημάτων κ.α.. Ο συγκεκριμένος επιστημονικός τομέας διαιρείται σε δύο υποτομείς τη μηχανική μάθηση (Machine learning) και τη βαθιά μάθηση (Deep learning), οι οποίοι αλληλοκαλύπτονται μεταξύ τους. (Τεχνητή Νοημοσύνη, 2019)

Ο Άγγλος μαθηματικός Alan Turing, που θεωρείται ένας από τους πατέρες της ΤΝ, ήταν ο πρώτος που διατύπωσε την έννοια της ευφυούς υπολογιστικής μηχανής και προσδιόρισε τη δοκιμή με την οποία αποδεικνύεται η ύπαρξη ή όχι ευφυΐας σε ένα υπολογιστή.

Κάποιοι από τους προτεινόμενους ορισμούς είναι :

- ΤΝ είναι ένας κλάδος της Πληροφορικής, ο οποίος ασχολείται με την αυτοματοποίηση ευφυούς συμπεριφοράς. (Luger και Stubblefield – 1998)
- ΤΝ είναι η μελέτη των μηχανισμών που διέπουν ευφυή συμπεριφορά, μέσω της κατασκευής και αξιολόγησης συστημάτων τα οποία παριστάνουν αυτούς τους μηχανισμούς (Luger και Stubblefield – τροποποιημένος ορισμός).
- ΤΝ είναι η ανάπτυξη υπολογιστικών συστημάτων για την επίλυση δύσκολων προβλημάτων, τα οποία δεν μπορούν να επιλυθούν με την εξαντλητική εξέταση όλων των πιθανών λύσεων μια και αυτές μπορεί να είναι πάρα πολλές.
- ΤΝ είναι η μελέτη του πώς να κάνουμε τον υπολογιστή να πράξει κάτι που επί του παρόντος ο άνθρωπος μπορεί να πράξει καλύτερα.

Οι παραπάνω ορισμοί έχουν τα δυνατά και αδύνατα σημεία τους όσον αφορά στην άμεση ή έμμεση αναφορά και την κατανόηση του όρου “ευφυής συμπεριφορά».

Οι Stuart Russell και Peter Norvig προτείνουν οκτώ ορισμούς που παρατίθενται σε δύο διαστάσεις:

Σκεπτόμενοι Ανθρώπινα	**Σκεπτόμενοι Λογικά**
«Η συναρπαστική νέα προσπάθεια για να κάνουμε του υπολογιστές σκεπτόμενους …μηχανές με νόηση με την πλήρη και κυριολεκτική σημασία» (Haugeland, 1985) «[Η αυτοματοποίηση σε] ενέργειες που συνδέουμε με την ανθρώπινη σκέψη, όπως η λήψη αποφάσεων, η επίλυση προβλημάτων, η μάθηση…» (Belman, 1978)	«Η μελέτη της νοητικής ικανότητας μέσω της χρήσης υπολογιστικών μοντέλων» (Charniac & McDermott, 1985) «Η μελέτη των υπολογισμών που καθιστά εφικτή την υλοποίηση της δράσης και αντίδρασης» (Winston, 1992)
Ενεργώντας Ανθρώπινα	**Ενεργώντας Λογικά**
«Η τέχνη της δημιουργίας μηχανών που εκτελούν πράξεις οι οποίες χρειάζονται ευφυΐα όταν εκτελούνται από άνθρωπο» (Kurzweil, 1990) «Η μελέτη του πώς να κατορθώσουμε να κάνουμε τους υπολογιστές να εκτελούν πράγματα στα οποία αυτή τη στιγμή οι άνθρωποι είναι καλύτεροι» (Rich & Knight, 1991)	«Υπολογιστική Νοημοσύνη είναι η μελέτη του σχεδιασμού ευφυών διαμεσολαβητών/πρακτόρων» (Poole at al., 1998) «Η Τεχνητή Νοημοσύνη…αφορά στην ευφυή συμπεριφορά στα ψευδή δεδομένα» (Nilsson, 1998)

Πίνακας 1: ορισμοί *ΤΝ σε δύο διαστάσεις*

Οι παραπάνω ορισμοί της 1ης γραμμής αφορούν στην διαδικασία σκέψης και συλλογισμού, της 2ης έχουν να κάνουν με τη συμπεριφορά. Αυτοί της 1ης στήλης μετρούν την επιτυχία με όρους πιστότητας στην ανθρώπινη συμπεριφορά, ενώ αυτοί της 2ης μετρούν με γνώμονα ένα ιδεατό

μέτρο απόδοσης που καλείται «ορθολογικότητα» . Ένα σύστημα είναι «ορθολογικό» όταν πράττει «το σωστό» δεδομένης της γνώσης που έχει μέχρι εκείνη τη στιγμή[7].

2.1.4. Μηχανική Μάθηση

Η **Μηχανική Μάθηση (Machine Learning ή ML)** είναι ένα πεδίο της τεχνητής νοημοσύνης (ΤΝ-AI) που ασχολείται με τη μελέτη, το σχεδιασμό και την ανάπτυξη μεθόδων και αλγορίθμων για την υλοποίηση υπολογιστικών συστημάτων που μαθαίνουν εμπειρικά από διαθέσιμα δεδομένα, προκειμένου να πραγματοποιήσουν προβλέψεις, να βελτιώσουν την απόδοσή της ή να εξάγουν αποφάσεις σχετικά με αυτά. (Machine Learning, 2019) και μελετά τον τρόπο με τον οποίο μπορούν να χρησιμοποιηθούν οι μηχανές ώστε να προσομοιώνουν τις ανθρώπινες μαθησιακές διεργασίες και να εντοπίζουν μεθόδους αυτό-βελτίωσης ώστε να κατακτήσουν νέα γνώση και ικανότητες, να αναγνωρίσουν υπάρχουσα γνώση και να βελτιώνουν συνεχώς την απόδοση και τα επιτεύγματά τους.

Σε σύγκριση με την ανθρώπινη μάθηση, η Μηχανική Μάθηση συντελείται με γρηγορότερους ρυθμούς ενώ η συσσώρευση γνώσης και τα αποτελέσματα της μάθησης διαδίδονται ευκολότερα.

Στην ουσία με τον όρο «μηχανική» εννοείται η χρήση κάποιου αλγόριθμου μέσω του οποίου αναλύονται δεδομένα. Ο αλγόριθμος «μαθαίνει» από τα δεδομένα με τα οποία τροφοδοτείται και στη συνέχεια αυτή η γνώση χρησιμοποιείται για να προβεί σε προβλέψεις σχετικές με τα δεδομένα αυτά. Αυτό το επιτυγχάνει εντοπίζοντας σχέσεις που διαφαίνονται να υπάρχουν από κοινού στα δεδομένα και με βάση αυτές προβαίνει σε προβλέψεις πάνω σε δεδομένα άγνωστα για αυτόν. Η «εμπειρία» που αποκτά βασίζεται στη χρήση περισσότερων δεδομένων για την εκπαίδευσή του. Χρησιμοποιώντας κάποια συνάρτηση κόστους, μεταβάλλει ένα πλήθος από

[7] Για να ξεχωρίσουμε ανάμεσα στην «*ανθρώπινη*» και «*λογική*» συμπεριφορά, δεν υποθέτουμε ότι τα ανθρώπινα όντα είναι υποχρεωτικά «*παράλογα*» με την έννοια του «*συναισθηματικά ασταθούς*» ή του «*αλλόκοτου*». Πρέπει να σημειωθεί ότι δεν είναι όλοι οι αθλητές πρωταθλητές ούτε όλοι οι σπουδαστές αριστούχοι. Κάποια συστηματικά λάθη στην ανθρώπινη συλλογιστική έχουν καταγραφεί από τους Kahneman et al. (1982).

παραμέτρους και προσπαθεί να εντοπίσει εκείνο το συνδυασμό από αυτές, ώστε οι προβλέψεις του να έχουν όσο το δυνατόν μεγαλύτερη επιτυχία.

Όταν οι άνθρωποι μαθαίνουν, αλλάζουν τον τρόπο που επιδρούν με το περιβάλλον (δεν θα ακουμπήσω ξανά μια φλόγα γιατί θα καεί το χέρι μου). Όταν οι αλγόριθμοι μαθαίνουν, απλά αλλάζουν τον τρόπο που επεξεργάζονται τα δεδομένα. Και ενώ ο ανθρώπινος νους είναι ικανός να μαθαίνει και να εκτελεί πολλές και διαφορετικές διαδικασίες, τελειοποιώντας τες στο πέρασμα του χρόνου, ο αλγόριθμος απλά μαθαίνει να εκτελεί μία μοναδική και πολύ συγκεκριμένη εργασία γρήγορα και αποτελεσματικά.

2.1.5. Βαθιά Μάθηση

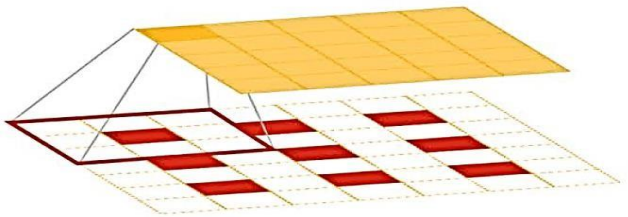
Η **Βαθιά Μάθηση (Deep Learning ή DL)** είναι πεδίο της Μηχανικής Μάθησης που χρησιμοποιεί Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) ώστε να επιτύχει μεγάλη ακρίβεια σε σημαντικά προβλήματα όπως για παράδειγμα αναγνώριση εικόνας και ήχου, επεξεργασία φυσικής γλώσσας κλπ. Γενικά με τη Βαθιά Μάθηση μπορούμε να λύσουμε προβλήματα που δεν είναι γραμμικά.

Συγκεκριμένα η Βαθιά Μάθηση(DL) περιλαμβάνει μεθόδους και αλγόριθμους μηχανικής μάθησης, βασιζόμενους σε υπολογιστικά συστήματα που προσομοιώνουν τη λειτουργία του ανθρώπινου εγκεφάλου, τα **Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks ή ANNs)**. Τα συγκεκριμένα υπολογιστικά μοντέλα εφαρμόζονται σε μεγάλα σύνολα δεδομένων και χρησιμοποιούν πολλαπλά στρώματα επεξεργασίας για τη σταδιακή μάθηση αναπαραστάσεων δεδομένων με πολλαπλά επίπεδα αφαίρεσης. Στόχος της βαθιάς μάθησης είναι η πρόβλεψη και η εξαγωγή πιο ολοκληρωμένων συμπερασμάτων, προκειμένου να ληφθούν αποφάσεις με μεγαλύτερη ακρίβεια για την αποτελεσματικότερη επίλυση των σύνθετων προβλημάτων. (Deep_learning, 2019)

Η Βαθιά Μάθηση(DL) είναι η μέθοδος κατά την οποία οι αλγόριθμοι κατασκευάζουν νευρωνικά δίκτυα με πολλαπλά στρώματα μεταξύ επιπέδων εξόδου. Κατά τη μάθηση αυτών των δικτύων τα δεδομένα που εισέρχονται στα δίκτυα αυτά υφίσταται επεξεργασία σε κάθε επίπεδο, από το οποίο διέρχονται, με σκοπό να υπολογιστεί στο τελικό επίπεδο η πιθανότητα για την επιθυμητή έξοδο. Παράδειγμα ομάδας συνόλου τεχνητών νευρωνικών δικτύων, το

οποίο χρησιμοποιείται και στη ανίχνευση αυτοκτονικού ιδεασμού αποτελεί το **συνελικτικό νευρωνικό δίκτυο (Convolutional neural network ή CNN)**. Τα CNN είναι νευρωνικά δίκτυα εμπρόσθιας τροφοδότησης. Περιλαμβάνουν μια ποικιλία από πολυεπίπεδα perceptron σχεδιασμένα να απαιτούν ελάχιστη προεπεξεργασία. Προσομοιάζουν το ζωτικό οπτικό φλοιό και εφαρμόζονται σε κείμενα προκειμένου να εντοπίσουν σε αυτά μοτίβα και να αποκτηθεί γνώση.

Κατά την εκπαίδευσή τους τα CNN δέχονται δεδομένα σε μορφή κειμένου που αντιπροσωπεύονται από ένα πίνακα, γνωστό και ως μήτρα. Κάθε γραμμή του πίνακα αντιπροσωπεύει μία λεκτική μονάδα, η οποία αποτελεί διάνυσμα. Αποτελούνται από αρκετά στρώματα συνέλιξης με λειτουργίες μη γραμμικής ενεργοποίησης που εφαρμόζονται σε δεδομένα. Δηλαδή σε κάθε στρώμα εσόδου χρησιμοποιούνται συνελίξεις για να υπολογίσουμε τις τιμές της εξόδου. Σε κάθε στρώμα εφαρμόζονται διαφορετικά φίλτρα, τα οποία συνδυάζουν τα αποτελέσματα τους. Ως εκ τούτου, τα CNN μαθαίνουν αυτόματα από τις τιμές που παράγουν τα φίλτρα. Γνωστές παραλλαγές CNN στη βιβλιογραφία που χρησιμοποιήθηκαν και στη ρητορική μίσους αποτελούν τα **Char CNN**, τα **Word CNN** και τα **Hybric CNN**. (Βλαχάβας, Ι., Κεφαλάς, Π., Βασιλειάδης, Ν., Κόκκορας, Φ., & Σακελλαρίου, Η. , 2006; Zhang, Z., & Luo, L.., 2018)



Σχήμα 3: Convolutional neural network ή CNN*:*

Για παράδειγμα Βαθιά Μάθηση χρησιμοποιήθηκε στον γνωστό αλγόριθμο AlphaGo ο οποίος νίκησε στο παιχνίδι Go τους πρωταθλητές Lee Sedol κατά το 2016 και Ke Jie κατά το 2017. Μην ξεχάσουμε βέβαια και την πρώτη ιστορική «νίκη» της μηχανής επί του ανθρώπου, όταν ο υπολογιστής Deep Blue νίκησε τον παγκόσμιο πρωταθλητή στο σκάκι Γκάρι Κασπάροφ το 1997. Βέβαια σε κάθε τέτοια επιτυχία τίθεται ο προβληματισμός κατά πόσο τέτοιου είδους νίκες οφείλονται στο γεγονός ότι οι μηχανές ανέπτυξαν «νοημοσύνη» ισάξια με την ανθρώπινη

ή αν είναι απλά θέμα μεγάλης και γρήγορης υπολογιστικής ικανότητας, την οποία απλά ο άνθρωπος δεν μπορεί να συναγωνιστεί.

Συμπερασματικά, μπορούμε να θεωρήσουμε το τρίπτυχο Τεχνητή Νοημοσύνη-Μηχανική Μάθηση-Βαθιά Μάθηση σαν τρία σύνολα, το καθένα γνήσιο υποσύνολο των προηγουμένων με την ΤΝ να αποτελεί το υπερσύνολο όλων.

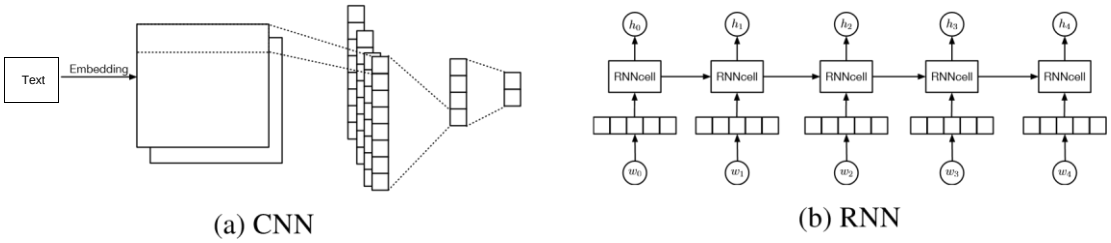
Η βαθιά μάθηση υπήρξε μεγάλη επιτυχία σε πολλές εφαρμογές, συμπεριλαμβανομένης της όρασης υπολογιστών, του NLP (Επεξεργασία Φυσικής Γλώσσας- Native Language Processing) και της ιατρικής διάγνωσης. Στον τομέα της έρευνας αυτοκτονίας, είναι επίσης μια σημαντική μέθοδος για την αυτόματη ανίχνευση αυτοκτονικών ιδεών και την πρόληψη αυτοκτονιών. Μπορεί να μάθει αποτελεσματικά τις λειτουργίες κειμένου αυτόματα χωρίς πολύπλοκες τεχνικές τεχνικών χαρακτηριστικών. Τα δημοφιλή δίκτυα deep neural (DNN) περιλαμβάνουν συνελικτικά νευρωνικά δίκτυα (CNNs) και επαναλαμβανόμενα νευρωνικά δίκτυα (RNNs) όπως φαίνεται στα σχήματα 4α και 4β, αντίστοιχα. Για την εφαρμογή DNNs, το κείμενο της φυσικής γλώσσας ενσωματώνεται συνήθως σε κατανεμημένο διανυσματικό χώρο με δημοφιλείς τεχνικές ενσωμάτωσης λέξεων, όπως word2vec [54] και GloVe [55]. Οι Shing et al. [13] εφάρμοσαν CNN σε επίπεδο χρήστη με μέγεθος φίλτρου 3, 4 και 5 για την κωδικοποίηση των αναρτήσεων των χρηστών. Το δίκτυο βραχυπρόθεσμης μνήμης (LSTM), μια δημοφιλής παραλλαγή του RNN, εφαρμόζεται για την κωδικοποίηση ακολουθιών κειμένου και στη συνέχεια υποβάλλεται σε επεξεργασία για ταξινόμηση με πλήρως συνδεδεμένα επίπεδα [17].

Οι πρόσφατες μέθοδοι εισάγουν άλλα προηγμένα πρότυπα μάθησης για ενσωμάτωση με DNNs για την ανίχνευση αυτοκτονικών ιδεών. Οι Ji et al. [56] πρότειναν το μοντέλο αθροιστικής για μεθόδους ενημέρωσης νευρωνικών δικτύων, δηλαδή CNN και LSTM, με στόχο την ανίχνευση αυτοκτονικού ιδεασμού σε ιδιωτικούς χώρους συνομιλίας. Ωστόσο, η αποκεντρωμένη εκπαίδευση βασίζεται σε συντονιστές στα δωμάτια συνομιλίας για την επισήμανση των θέσεων χρήστη για εποπτευόμενη εκπαίδευση, που μπορεί να εφαρμοστεί μόνο σε πολύ περιορισμένα σενάρια. Ένας πιθανόν καλύτερος τρόπος είναι να χρησιμοποιηθούν μη εποπτευόμενες ή ημι-εποπτευόμενες μέθοδοι μάθησης. Οι Benton et al. [16] προέβλεψαν απόπειρες αυτοκτονίας και ψυχική ασθένειες με νευρωνικά μοντέλα στο πλαίσιο της μάθησης πολλαπλών εργασιών, προβλέποντας το φύλο των χρηστών ως βοηθητική εργασία. Οι Gaur et al. [57] ενσωμάτωσαν εξωτερικές βάσεις γνώσεων και οντολογία που σχετίζεται με αυτοκτονίες στην αναπαράσταση

κειμένου και απέκτησε βελτιωμένη απόδοση με ένα μοντέλο CNN. Οι Coppersmith et al. [58] ανέπτυξαν ένα βαθύ μοντέλο μάθησης με το GloVe για ενσωμάτωση λέξεων, αμφίδρομο LSTM για κωδικοποίηση ακολουθίας και μηχανισμό αυτοπροσοχής για τη λήψη της πιο ενημερωτικής ακολουθίας. Οι Sawhney et al. [59] χρησιμοποίησαν LSTM, CNN και RNN για ανίχνευση αυτοκτονικού ιδεασμού. Οι Ji et al. [60] προτεινόμενο προσεκτικό δίκτυο σχέσεων με LSTM και μοντελοποίηση θεμάτων για κωδικοποίηση κειμένου και δείκτες κινδύνου.

Στο κοινόχρηστο έργο 2019 CLPsych Shared Task [61], εφαρμόστηκαν πολλές δημοφιλείς αρχιτεκτονικές DNN. Οι Hevia et al. [62] αξιολόγησαν το αποτέλεσμα της προεκπαίδευσης χρησιμοποιώντας διαφορετικά μοντέλα, συμπεριλαμβανομένου του RNN που βασίζεται σε GRU. Οι Morales et al. [63] μελέτησαν αρκετά δημοφιλή μοντέλα βαθιάς μάθησης, όπως CNN, LSTM, και Neural Network Synthesis (NeuNetS). Οι Matero et al. [64] πρότειναν μοντέλο διπλού πλαισίου χρησιμοποιώντας ιεραρχικά προσεκτικό RNN και αμφίδρομες αναπαραστάσεις κωδικοποιητών από μετασχηματιστές (BERT).

Μια άλλη υπο-κατεύθυνση είναι η λεγόμενη υβριδική μέθοδος που συνεργάζεται δευτερεύουσα μηχανική χαρακτηριστικών με τεχνικές μάθησης αναπαράστασης. Οι Chen et αϊ. [65] πρότειναν ένα υβριδικό μοντέλο ταξινόμησης μοντέλου συμπεριφοράς και μοντέλου γλώσσας αυτοκτονίας. Οι Zhao et al. [66] πρότειναν ένα μοντέλο D-CNN, το οποίο ενσωματώνει λέξεις και εξωτερικά χαρακτηριστικά πίνακα ως εισόδους για την ταξινόμηση των απόπειρων αυτοκτονίας πασχόντων από κατάθλιψη.

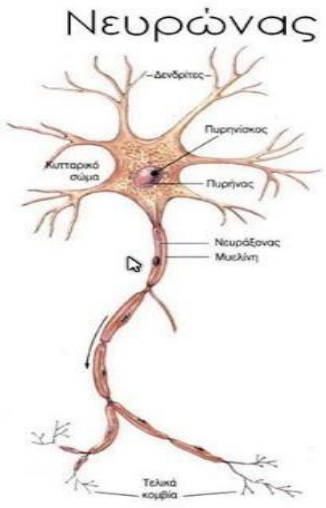


Σχήμα 4:Βαθιά νευρωνικά δίκτυα για ανίχνευση αυτοκτονικών ιδεών

2.1.6. Τεχνητά Νευρωνικά Δίκτυα

2.1.6.1. Νευρώνας

Αναφέρθηκε προηγουμένως ότι η Τεχνητή Νοημοσύνη και κατ' επέκταση η Μηχανική Μάθηση προσπαθεί να προσομοιώσει τη λειτουργία της ανθρώπινης μαθησιακής διεργασίας.



Σχήμα 5: Βιολογικός νευρώνας

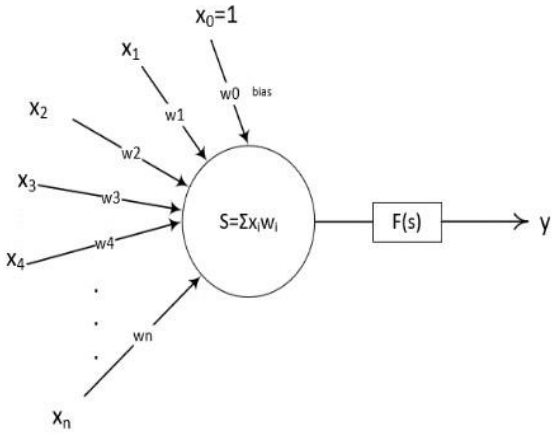
Γνωρίζουμε ότι η διεργασία αυτή εκτελείται στον ανθρώπινο εγκέφαλο με τη βοήθεια των βασικών δομικών στοιχείων του νευρικού συστήματος, των **νευρώνων**. Ιστολογικά ο νευρώνας αποτελείται από τρία μέρη: το κυτταρικό σώμα και τις αποφύσεις του, έναν ή περισσότερους δενδρίτες και τον νευράξονα. Κάθε νευρώνας ενώνεται με άλλους μέσω των **συνάψεων** και μεταδίδει ερεθίσματα σ' αυτούς ή σε άλλα κύτταρα απελευθερώνοντας νευροδιαβιβαστές στις συνάψεις. Τα ερεθίσματα (σήματα) που εισέρχονται μέσω των δενδριτών στο σώμα, συνδυάζονται και, αν το αποτέλεσμα ξεπερνά κάποιο όριο (κατώφλι), διαδίδεται μέσω του άξονα σε άλλους νευρώνες.

Ο εγκέφαλος έχει περίπου 100 δισεκατομμύρια νευρώνες, ο καθένας από τους οποίους ενώνεται κατά μέσο όρο με άλλους χίλιους με αποτέλεσμα να υπάρχουν περίπου 100 τρισεκατομμύρια συνάψεις. Ο χρόνος απόκρισης των βιολογικών νευρώνων είναι της τάξης των msec, αλλά λαμβάνονται πολύπλοκες αποφάσεις εκπληκτικά γρήγορα. Η επεξεργαστική ισχύς και η πληροφορία που περιέχει ο εγκέφαλος είναι διαμοιρασμένη σε όλο τον όγκο του

και επομένως αυτός μπορεί να προσομοιωθεί με ένα παράλληλο και κατανεμημένο υπολογιστικό σύστημα.

2.1.6.2. Τεχνητός Νευρώνας-Μετάδοση σήματος

Κατ' αντιστοιχία με τον εγκέφαλο, ένα τεχνητό νευρωνικό δίκτυο (ΤΝΔ – Artificial Neural Network) αποτελείται από ένα αριθμό στοιχείων που ονομάζονται **τεχνητοί νευρώνες**. Σε κάθε έναν από αυτούς καταφθάνει ένας αριθμός n από σήματα τα οποία υπόκεινται σε επεξεργασία. Καθένα από τα σήματα αυτά συνδέεται με μία τιμή, το επονομαζόμενο **βάρος** (το οποίο παίρνει τιμές ανάμεσα στο 0 και το 1) που υποδηλώνει το πόσο στενά συνδέονται οι νευρώνες που



Σχήμα 6: Τεχνητός Νευρώνας

μεταφέρουν αυτό το σήμα. Υποδηλώνει δηλαδή τη σημαντικότητα του συγκεκριμένου σήματος στο τελικό αποτέλεσμα: όσο μεγαλύτερη η τιμή του βάρους τόσο πιο σημαντικό είναι το συγκεκριμένο σήμα. Υπάρχει και ένα επιπλέον σήμα το οποίο ονομάζεται **μεροληψία (bias)** του νευρώνα (Σχήμα 6). Το σήμα αυτό έχει πάντα την τιμή 1 και η μεροληψία εξαρτάται από την τιμή του βάρους. Ο κάθε νευρώνας έχει μία μόνο έξοδο y (αντίστοιχη της σύναψης στον εγκέφαλο), μέσω της οποίας μεταβιβάζει το επεξεργασμένο σήμα στον επόμενο νευρώνα.

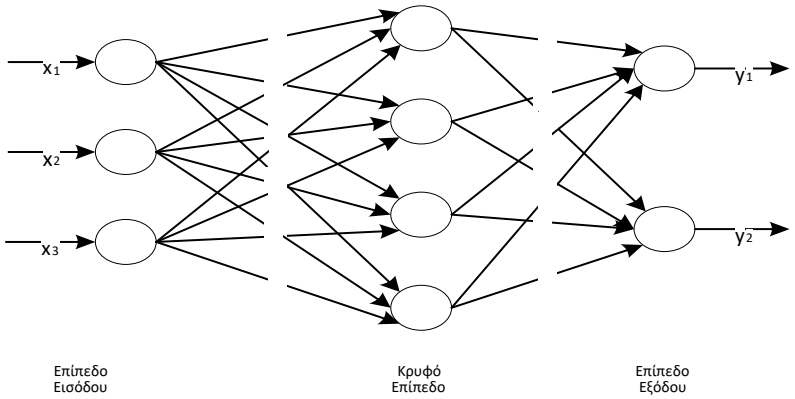
Όλα τα σήματα που φθάνουν στις εισόδους του νευρώνα αθροίζονται λαμβάνοντας υπόψη, όπως είπαμε, το βάρος του καθενός και υπολογίζεται η σταθμισμένη ποσότητα $\mathbf{S} = \sum_{\mathbf{i=0}}^{\mathbf{n}} \mathbf{x_i w_i}$.

Η έξοδος του νευρώνα προκύπτει από την εφαρμογή στην ποσότητα S της **συνάρτησης ενεργοποίησης ή μεταφοράς F** (activation or squashing function). Η συνάρτηση αυτή χρησιμοποιείται για να περιορίσει την τιμή εξόδου σε ένα εύρος τιμών στα διαστήματα [0,1] ή [-1, 1]. Η μεροληψία του νευρώνα, που αναφέρθηκε παραπάνω χρησιμοποιείται για αυξήσει ή να μειώσει την τιμή εξόδου. Το αν ο νευρώνας θα μεταφέρει την ποσότητα εξόδου στον επόμενο ή όχι εξαρτάται, συνήθως, από μια προκαθορισμένη τιμή που ονομάζεται **κατώφλι** (threshold). Αν η τιμή της εξόδου (σήμα) ξεπερνά το κατώφλι, τότε μεταφέρεται μέσω της σύναψης στον επόμενο νευρώνα και λέμε ότι ο νευρώνας **ενεργοποιήθηκε,** ενώ σε διαφορετική περίπτωση δεν γίνεται μεταφορά.

Ας δώσουμε ένα παράδειγμα νευρώνα με δύο εισόδους $x_0=1$, $x_1=0.7$ και $x_2=1.0$, βάρη $w_0=0.2$ $w_1=0.5$ και $w_2=0.5$ και κατώφλι $t=1.0$ Τότε $F(s) = x_1 \cdot w_1 + x_2 \cdot w_2 = 1.0 \cdot 0.2 + 0.7 \cdot 0.5 + 1.0 \cdot 0.5 = 1.05 > t$ και επομένως ο νευρώνας θα ενεργοποιηθεί. Αν όμως το βάρος του νευρώνα μεροληψίας ήταν $w_0= 0.1$, τότε $F(s)= 0.95 <t$ και ο νευρώνας δεν θα είχε ενεργοποιηθεί.

2.1.6.3. Τεχνητά νευρωνικά δίκτυα

Τα **τεχνητά νευρωνικά δίκτυα** (ΤΝΔ – Artificial Neural Networks) είναι συστήματα που επεξεργάζονται τα δεδομένα της εισόδου και αποτελούνται από ένα πλήθος τεχνητών νευρώνων, οργανωμένων σε διάφορα επίπεδα (layers). Όταν υπάρχουν περισσότερα του ενός επίπεδα με πολλούς νευρώνες, τότε μιλούμε για **βαθιά νευρωνικά δίκτυα** (deep neural



Σχήμα 7: Τεχνητό Νευρωνικό Δίκτυο

networks). Τα συγκεκριμένα δίκτυα χρησιμοποιούνται όταν έχουμε προβλήματα που δεν είναι γραμμικά διαχωρίσιμα.

Σε ένα τέτοιου είδους δίκτυο υπάρχουν διαφορετικά επίπεδα (Σχήμα 7). Ειδικότερα έχουμε:

- το **επίπεδο εισόδου** το οποίο δέχεται τα σήματα και τα διοχετεύει στο επόμενο επίπεδο. Στην ουσία δεν έχει νευρώνες και συνήθως δεν το μετρούμε σαν επίπεδο.
- τα **κρυφά επίπεδα** (κανένα, ένα ή περισσότερα) που αποτελούνται από μη-γραμμικούς νευρώνες. Τα κρυφά επίπεδα μπορούν να έχουν οποιονδήποτε αριθμό νευρώνων. Για σύνθετα προβλήματα (μη γραμμικά επιλύσιμα) υπάρχουν, συνήθως, περισσότερα του ενός κρυφά επίπεδα.
- το **επίπεδο εξόδου** όπου βρίσκονται ένας ή περισσότεροι νευρώνες γραμμικοί ή όχι ή συνδυασμός αυτών. Μη γραμμικούς νευρώνες (πχ σιγμοειδείς) έχουμε όταν θέλουμε στις εξόδους τους να έχουν διακριτές τιμές και γραμμικούς όταν στις εξόδους τους πρέπει να έχουν συνεχείς τιμές

  Ένα πολυεπίπεδο ΤΝΔ αναπαρίσταται συντομογραφικά ως εξής: p-m_1-m_2-…-$m_κ$-n, όπου p είναι ο αριθμός των εισόδων, κ ο αριθμός των κρυφών επιπέδων με m_1 νευρώνες στο 1ο, m_2 στο 2ο κοκ και n ο αριθμός των νευρώνων εξόδου. Γενικά δεν υπάρχουν κανόνες ούτε για τον αριθμό των νευρώνων κάθε επιπέδου ούτε για τον αριθμό των κρυφών επιπέδων του δικτύου. Αναλόγως των προβλημάτων που επιλύουμε, επιλέγουμε την αρχιτεκτονική του, συνήθως με τη μέθοδο των δοκιμών και λάθους (trial and error). Πάντως υπάρχουν κάποιοι εμπειρικοί κανόνες που βάζουν όρια στην αρχιτεκτονική του δικτύου όπως πχ ότι έχει αποδειχθεί πως ένα δίκτυο δεν μπορεί να μάθει περισσότερα παραδείγματα από το διπλάσιο του αριθμού των βαρών του .

  Ανάλογα με τις συνάψεις που υπάρχουν μεταξύ των νευρώνων στο δίκτυο τα ξεχωρίζουμε σε:

  - **πλήρως συνδεδεμένα** (fully connected) ΤΝΔ, όπου ο κάθε νευρώνας συνδέεται με όλους τους νευρώνες του επόμενου επιπέδου
  - **Μερικώς** συνδεδεμένα (partially connected) ΤΝΔ, όπου υπάρχουν νευρώνες που δεν συνδέονται με κάποιον ή κάποιους νευρώνες του επόμενου επιπέδου

- ΤΝΔ **με πρόσθια τροφοδότηση** (feedforward) όπου οι νευρώνες κάποιου επιπέδου ενώνονται μόνο με νευρώνες επόμενου επιπέδου
- ΤΝΔ **με ανατροφοδότηση** (feedback ή recurrent) όπου οι νευρώνες κάποιου επιπέδου μπορεί να ενώνονται με νευρώνες προηγούμενου επιπέδου

Στις περισσότερες των περιπτώσεων χρησιμοποιούμε δίκτυα με πρόσθια τροφοδότηση.

2.1.7. Επεξεργασία φυσικής γλώσσας (NLP)

Η επεξεργασία φυσικής γλώσσας (Native Language Processing) αποτελεί έναν κλάδο της τεχνητής νοημοσύνης (Artificial Intelligence) που βοηθά τους υπολογιστές να κατανοούν, να ερμηνεύουν και να χειρίζονται την ανθρώπινη γλώσσα. Η επεξεργασία της φυσικής γλώσσας χρησιμοποιείται από πολλούς επιστημονικούς κλάδους, συμπεριλαμβανομένης της επιστήμης της πληροφορικής και της υπολογιστικής γλωσσολογίας, προσπαθώντας να καλύψει το κενό που υπάρχει μεταξύ της ανθρώπινης επικοινωνίας και της κατανόησης των υπολογιστών. Ενώ η επεξεργασία της φυσικής γλώσσας δεν αποτελεί μια νέα επιστήμη, η τεχνολογία προχωράει γρήγορα χάρη στο αυξημένο ενδιαφέρον για τις επικοινωνίες από άνθρωπο σε μηχανή, καθώς και τη διαθεσιμότητα μεγάλων δεδομένων, ισχυρών υπολογιστών και ενισχυμένων αλγορίθμων.

Η επεξεργασία φυσικής γλώσσας βοηθά τους υπολογιστές να επικοινωνούν με τους ανθρώπους στη γλώσσα τους και να κλιμακώνουν άλλες εργασίες που σχετίζονται με τη γλώσσα. Για παράδειγμα, μας παρέχει τη δυνατότητα να μπορούν οι υπολογιστές να διαβάζουν κείμενο, να ακούν και να κατανοούν τον λόγο μας, να ερμηνεύουν και να μετρούν το συναίσθημα μας, όπως και να καθορίζουν ποια μέρη του λόγου ή του κειμένου είναι τα σημαντικά.

Οι σημερινές μηχανές μπορούν να αναλύσουν περισσότερα δεδομένα βασισμένα στη γλώσσα από ό,τι οι άνθρωποι, με συνεπή και αμερόληπτο τρόπο. Λαμβάνοντας υπόψη την εκπληκτική ποσότητα των αδόμητων δεδομένων που παράγονται καθημερινά, από ιατρικά αρχεία μέχρι δημοσιεύσεις σε μέσα κοινωνικής δικτύωσης, οι τεχνικές αυτοματισμού αποτελούν κρίσιμο εργαλείο για την αποτελεσματική ανάλυση δεδομένων κειμένου και λόγου. Η ανθρώπινη γλώσσα είναι εκπληκτικά πολύπλοκη και ποικίλει. Υπάρχουν άπειροι τρόποι, που μπορούμε να εκφραστούμε τόσο σε προφορικό όσο και σε γραπτό επίπεδο. Υπάρχουν

εκατοντάδες γλώσσες, που κάθε μία διαθέτει και μια πληθώρα από διαλέκτους. Επιπλέον σε κάθε γλώσσα υπάρχει ένα μοναδικό σύνολο κανόνων γραμματικής, συντακτικού και ορολογίας.

Στον γραπτό λόγο, συχνά κάνουμε λάθη, συντομεύουμε λέξεις και παραλείπουμε τα σημεία στίξης. Στον προφορικό λόγο, πολλοί έχουν τοπικές προφορές και χρησιμοποιούν τοπικές ή ξενόφερτες λέξεις. Επίσης στην ομιλία πρέπει να αντιμετωπιστούν προβλήματα προφοράς, βραδυγλωσσίας, παράλληλης ομιλίας ή ήχων-θορύβου. Οπότε για τη μοντελοποίηση της ανθρώπινης γλώσσας, υπάρχει ανάγκη για συντακτική και σημασιολογική κατανόηση και άριστη γνώση του συγκεκριμένου πεδίου.

Το NLP βοηθά στην επίλυση της ασάφειας στη γλώσσα και προσθέτει χρήσιμη αριθμητική δομή στα δεδομένα. Χρησιμοποιείται κυρίως για αναγνώριση ομιλίας και ανάλυση κειμένου. Η επεξεργασία φυσικής γλώσσας περιλαμβάνει πολλές διαφορετικές τεχνικές για την ερμηνεία της ανθρώπινης γλώσσας, από μεθόδους στατιστικής και μηχανικής μάθησης έως προσεγγίσεις βασισμένες σε κανόνες και αλγορίθμους. Χρειάζεται μια ευρεία σειρά προσεγγίσεων, διότι τα δεδομένα που είναι φωνητικά ή είναι κείμενο ποικίλλουν, όπως και οι πρακτικές εφαρμογές τους.

Οι βασικές εργασίες NLP περιλαμβάνουν τις τεχνικές tokenization (χωρισμός σε λεκτικά - tokens), parsing, lemmatization / stemming και tagging σε μέρη του λόγου, την αναγνώριση γλώσσας και τον προσδιορισμό σημασιολογικών σχέσεων. Σε γενικές γραμμές, οι τεχνικές του NLP κομματιάζουν τη γλώσσα σε μικρότερα στοιχειώδη κομμάτια, προσπαθώντας να κατανοήσουν τις σχέσεις μεταξύ των κομματιών και να διερευνήσουν πώς τα κομμάτια αυτά, συνεργάζονται ώστε να δημιουργούν νόημα.

Αυτά τα βασικά καθήκοντα χρησιμοποιούνται συχνά για επεξεργασία NLP υψηλότερου επιπέδου, όπως:

- Κατηγοριοποίηση περιεχομένου: Μια συλλογή δεδομένων βασισμένη σε γλωσσολογικές τεχνικές, όπως η αναζήτηση, η ευρετηρίαση και η ανίχνευση διπλοεγγραφών.
- Ανακάλυψη και μοντελοποίηση θεμάτων: Εύρεση με ακρίβεια του νοήματος και του θέματος σε συλλογές κειμένων και εφαρμογή προηγμένων αναλύσεων, όπως τεχνικές βελτιστοποίησης και πρόβλεψης.

- Συναφής εξαγωγή: Εξαγωγή δομημένων πληροφοριών από πηγές που βασίζονται σε κείμενο.
- Ανάλυση συναισθημάτων: Προσδιορισμός της διάθεσης και των υποκειμενικών απόψεων από μεγάλες ποσότητες κειμένου, συμπεριλαμβανομένου του μέσου συναισθήματος και της εξόρυξης γνώσης.
- Μετατροπή ομιλίας σε κείμενο και μετατροπή κειμένου σε ομιλία: Μετατροπή φωνητικών εντολών σε γραπτό κείμενο και αντίστροφα.
- Συνοπτική παρουσίαση εγγράφου: Δημιουργία αυτόματων συνόψεων μεγάλων κειμένων κειμένου.
- Μηχανική μετάφραση: Αυτόματη μετάφραση κειμένου ή ομιλίας από μια γλώσσα σε μία άλλη.

Σε όλες αυτές τις περιπτώσεις, ο πρωταρχικός στόχος είναι να ληφθούν ακατέργαστα δεδομένα μέσω της εισαγωγής κειμένου ή ήχου και να χρησιμοποιηθούν αλγόριθμοι ή τεχνικές γλωσσολογίας ώστε να μετατραπεί ή να εμπλουτιστεί ένα κείμενο με τέτοιο τρόπο ώστε να αποδίδει μεγαλύτερη αξία. Η επεξεργασία της φυσικής γλώσσας συμβαδίζει με την ανάλυση των κειμένων, η οποία μετράει, ομαδοποιεί και κατηγοριοποιεί τις λέξεις για να εξάγει τη δομή και τη σημασία τους από μεγάλες ποσότητες περιεχομένου. Υπάρχουν πολλές κοινές και πρακτικές εφαρμογές του NLP στην καθημερινότητά μας. Μία εφαρμογή του είναι στο ηλεκτρονικό ταχυδρομείο μας, στο email, στο φάκελο ανεπιθύμητης αλληλογραφίας, όπου ίσως έχετε παρατηρήσει ομοιότητες στο τίτλο του θέματος του email. Το φιλτράρισμα των ανεπιθύμητων μηνυμάτων γίνεται με την Bayesian τεχνική, μια στατιστική τεχνική NLP που συγκρίνει τις λέξεις σε spam με έγκυρα μηνύματα ηλεκτρονικού ταχυδρομείου για τον εντοπισμό της ανεπιθύμητης αλληλογραφίας. Ένα υποπεδίο του NLP είναι το NLU, που ονομάζεται φυσική γλώσσα κατανόησης, έχει αρχίσει να αυξάνεται σε δημοτικότητα λόγω των δυνατοτήτων του σε γνωστικές εφαρμογές και εφαρμογές τεχνητής νοημοσύνης. Το NLU υπερβαίνει τη δομική κατανόηση της γλώσσας, δημιουργεί λεξιλόγιο, και δημιουργεί μια καλά διαμορφωμένη ανθρώπινη γλώσσα από μόνο του. Οι αλγόριθμοι του NLU πρέπει να αντιμετωπίσουν το εξαιρετικά σύνθετο πρόβλημα της σημασιολογικής ερμηνείας, δηλαδή, να κατανοήσουν την προτεινόμενη έννοια

του λεκτικού ή γραπτού λόγου, με όλα τα λεπτά στοιχεία και τα συμπεράσματα που μπορούμε να καταλάβουμε εμείς.

Η εξέλιξη του NLP προς την NLU έχει πολλές σημαντικές επιπτώσεις τόσο για τις επιχειρήσεις όσο και για τους καταναλωτές. Φανταστείτε τη δύναμη ενός αλγορίθμου που μπορεί να κατανοήσει τη σημασία και την απόχρωση της ανθρώπινης γλώσσας σε πολλά πλαίσια, από την ιατρική και τη διδασκαλία μέχρι την αναγνώριση τάσεων αυτοκτονίας.

Η προ-επεξεργασία του κειμένου και των δεδομένων αποτελεί ουσιαστικό μέρος κάθε συστήματος NLP, καθώς οι χαρακτήρες, οι λέξεις και οι φράσεις που προσδιορίζονται σε αυτό το στάδιο είναι οι θεμελιώδεις μονάδες που μεταφέρονται σε όλα τα υπόλοιπα στάδια επεξεργασίας που ακολουθούν.

Δυστυχώς, οι λέξεις που εμφανίζονται σε έγγραφα και κείμενα έχουν πολλές δομικές παραλλαγές. Έτσι, πριν από την ανάκτηση πληροφοριών από τα κείμενα, οι τεχνικές προ-επεξεργασίας δεδομένων εφαρμόζονται στο στοχευόμενο σύνολο δεδομένων για να μειωθεί το μέγεθος του συνόλου δεδομένων κι έτσι να αυξηθεί η αποτελεσματικότητα του συστήματος.

Η προ-επεξεργασία περιλαμβάνει ένα σύνολο ενεργειών, οι οποίες προετοιμάζουν το κείμενο. Επειδή τα κείμενα περιέχουν συχνά μερικές συγκεκριμένες ειδικές μορφές, όπως αριθμούς, ημερομηνίες και τις συχνά χρησιμοποιούμενες λέξεις (όπως προθέσεις, άρθρα κτλ), που δεν βοηθούν στην εξόρυξη γνώσης από το κείμενο, πρέπει να εξαλειφθούν από αυτό.

Το πρώτο βήμα στην ανάλυση και εξόρυξη γνώσης από κείμενα, είναι να οριστεί σωστά το σώμα κειμένων που θα αναλυθεί. Εάν η φιλοδοξία είναι τα αποτελέσματα να γενικευθούν σε ένα μεγαλύτερο πληθυσμό εγγράφων, τότε εφαρμόζονται οι τυποποιημένοι κανόνες δειγματοληψίας. Δηλαδή τα έγγραφα θα πρέπει να επιλέγονται χρησιμοποιώντας κάποια τυχαία ή κάποια άλλη στρατηγική δειγματοληψίας και να είναι αντιπροσωπευτικά πάνω στο πεδίο που ασχολούμαστε. Μια πρόκληση που αντιμετωπίζεται πολλές φορές σε αυτό το στάδιο είναι η επανάληψη εγγραφών (διπλοεγγραφές), δηλαδή ότι μπορεί να υπάρχουν περισσότερες από μία περιπτώσεις του ίδιου εγγράφου σε ένα σώμα. Για παράδειγμα, στις βάσεις δεδομένων μεγάλων παγκόσμιων εφημερίδων, το ίδιο άρθρο μπορεί να υπάρχει περισσότερες από μία φορές, ίσως επειδή υπάρχουν διαφορετικές εκδόσεις μιας εφημερίδας σε διάφορες χώρες. Αυτή η αλληλοεπικάλυψη μπορεί να προκαλέσει στρέβλωση του συμπεράσματος με την υπερεκπροσώπηση ορισμένων εγγράφων. Η εύρεση και διαγραφή των διπλοεγγραφών μπορεί

να πραγματοποιηθεί με διάφορους αυτοματοποιημένους (χρησιμοποιώντας κάποιον αλγόριθμο) ή μη τρόπους, ελέγχοντας το σώμα του κειμένου, ώστε να διασφαλιστεί ότι κάθε έγγραφο αποτελεί μια μοναδική εγγραφή.

Μόλις καθοριστεί το σώμα, το επόμενο βήμα είναι να μετασχηματιστεί σε μορφή που να μπορεί να υποβληθεί σε ανάλυση. Η διαδικασία αυτή είναι συχνά δυσκίνητη και χρονοβόρα. Για παράδειγμα, τα έγγραφα που αποθηκεύονται σε μορφή .pdf, ίσως χρειαστεί να μετατραπούν και να αποθηκευτούν ως αρχεία κειμένου (.txt). Επίσης, αν τα κείμενα υπάρχουν μόνο σε χαρτί, όπως πολλά αρχειακά αντικείμενα, τότε πρέπει να σαρωθούν και να μετατραπούν σε αρχεία .txt, χρησιμοποιώντας λογισμικό οπτικής αναγνώρισης χαρακτήρων.

Στην ουσία παίρνουμε το σώμα ενός κειμένου και εκτελούμε σε αυτό κάποιους βασικούς μετασχηματισμούς και αναλύσεις, ώστε να μείνει το σώμα του κειμένου με αντικείμενα που θα είναι πολύ πιο χρήσιμα για την εκτέλεση κάποιου περαιτέρω, πιο ουσιαστικού αναλυτικού έργου.

Υπάρχουν 3 κύριες συνιστώσες της προ-επεξεργασίας κειμένων:

- Tokenization – αναγνώριση λέξεων
- normalization – στελέχωση
- subsitution – αντικατάσταση

2.1.8. Βασικές Μετρικές

Τα παραπάνω μοντέλα που παρουσιάστηκαν δεν είναι πάντοτε το ίδιο αποτελεσματικά ως προς την επίλυση των προβλημάτων. Για παράδειγμα, δεν ταξινομούν και δεν παρέχουν προβλέψεις με την ίδια ακρίβεια. Για την παρατήρηση και αξιολόγηση της απόδοσής τους χρησιμοποιούνται κάποιες μετρικές οδηγούν στην εξαγωγή χρήσιμων συμπερασμάτων. Για τα μοντέλα αυτές οι μετρικές βασίζονται σε ένα πίνακα γνωστό και ως πίνακα σύγχυσης(confusion matrix), ο οποίος χρησιμοποιείται για να περιγράψει την απόδοση ενός μοντέλου δοκιμών. Μια οπτική απεικόνιση αυτού του πίνακα είναι η ακόλουθη:

	Predicted class		
Actual Class		Class = Yes	Class = No
	Class = Yes	True Positive	False Negative
	Class = No	False Positive	True Negative

Πίνακας 2: Πίνακας Σύγχυσης (Confussion Table)

Επιπλέον, οι μετρικές βασίζονται και στις ακόλουθες παραμέτρους:

1) **True Positives (TP)** – πρόκειται για τις σωστά προβλεπόμενες θετικές τιμές, που σημαίνει ότι η τιμή της πραγματικής κλάσης είναι ναι και η τιμή της προβλεπόμενης κλάσης είναι ναι. Για παράδειγμα εάν η πραγματική τιμή κλάσης υποδεικνύει ότι ένα μήνυμα περιέχει τάσεις αυτοκτονίας, τότε και ισχύει το ίδιο και για την προβλεπόμενη κλάση.
2) **True Negatives (TN)** – πρόκειται για τις σωστά προβλεπόμενες αρνητικές τιμές που σημαίνει ότι η τιμή της πραγματικής κλάσης δεν είναι και η αξία της προβλεπόμενης κλάσης δεν είναι. Για παράδειγμα εάν η πραγματική κλάση υποδεικνύει ότι ένα μήνυμα δεν περιέχει τάσεις αυτοκτονίας, τότε και για την προβλεπόμενη κλάση ισχύει το ίδιο.
3) **False Positives (FP)** – πρόκειται για την περίπτωση που η πραγματική τάξη δεν είναι και η προβλεπόμενη τάξη είναι ναι. Για παράδειγμα, η πραγματική κλάση δείχνει το μήνυμα δεν περιέχει τάσεις αυτοκτονίας, και η προβλεπόμενη κλάση υπονοεί ότι το μήνυμα πιθανόν να περιέχει τάσεις αυτοκτονίας.
4) **False Negatives (FN)** – πρόκειται για την περίπτωση που η πραγματική τάξη είναι ναι, αλλά η προβλεπόμενη κατηγορία δεν είναι. Για παράδειγμα η πραγματική τιμή κλάσης υποδεικνύει ότι το μήνυμα περιέχει τάσεις αυτοκτονίας και η προβλεπόμενη κατηγορία σας λέει ότι το μήνυμα πιθανόν να περιέχει τάσεις αυτοκτονίας.

Παρακάτω παρατίθενται οι βασικότερες από τις μετρικές αξιολόγησης συστημάτων:

1) **Ορθότητα ή Πιστότητα (Accuracy)** είναι η αναλογία μιας σωστά προβλεπόμενης παρατήρησης προς τις συνολικές παρατηρήσεις.

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

Αυτό, όμως, από μόνο του μπορεί να οδηγήσει σε λανθασμένα συμπεράσματα, αν δεν λάβουμε υπόψη μας και άλλους παράγοντες. Για παράδειγμα, έστω ότι έχουμε ένα σύνολο 10.000 παραδειγμάτων από τα οποία τα 9.990 ανήκουν στην κλάση 0 και τα υπόλοιπα 10 στην κλάση 1. Αν το δίκτυό μας, από λανθασμένη κατασκευή ή εσφαλμένη εκπαίδευση κατηγοριοποιεί όλα τα δείγματα στην κλάση 0 τότε η πιστότητά του θα είναι 99,90%. Σαν αριθμός φαίνεται εξαιρετικός, όμως στην πραγματικότητα το μοντέλο δεν αναγνωρίζει καθόλου τη μία από τις δύο κλάσεις.

2) Ακρίβεια (Precision) είναι ο λόγος των σωστά προβλεπόμενων θετικών παρατηρήσεων προς τις συνολικές προβλεπόμενες θετικές παρατηρήσεις.

Η ακρίβεια ή τιμή θετικών προβλέψεων (precision or Positive Predicted Value – PPV) μας δείχνει πόσα από τα ταξινομημένα θετικά αποτελέσματα είναι πραγματικά θετικά.

$$\mathrm{PPV} = \frac{TP}{TP + FP}$$

Αντίστοιχα η τιμή αρνητικών προβλέψεων (Negative Predicted Value – NPV) μας δείχνει πόσα από τα ταξινομημένα αρνητικά αποτελέσματα είναι πραγματικά αρνητικά.

$$\mathrm{NPV} = \frac{TN}{TN + FN}$$

Όσο μικραίνει ο αριθμός των λανθασμένων προβλέψεων, τόσο μεγαλύτερες είναι οι συγκεκριμένες τιμές.

3) Ανάκληση (Recall/True Positive Rate): Ανάκληση είναι η αναλογία των σωστά προβλεπόμενων θετικών παρατηρήσεων σε όλες τις παρατηρήσεις στην πραγματική κατηγορία – ναι.

$$\mathrm{TPR} = \frac{TP}{TP + FN}$$

Επιλεκτικότητα (selectivity/True Negative rate): Επιλεκτικότητα είναι η αναλογία των σωστά προβλεπόμενων αρνητικών παρατηρήσεων σε όλες τις παρατηρήσεις στην πραγματική κατηγορία – όχι.

$$\text{TNR} = \frac{TN}{TN + FP}$$

Όσο πιο μεγάλες είναι οι τιμές, τόσο λιγότερα παραδείγματα έχουν ταξινομηθεί λάθος.

4) **F1-Score:** Η βαθμολογία F1 είναι ο σταθμισμένος μέσος όρος της ακρίβειας και ανάκλησης.

Οι δύο τελευταίες κατηγορίες τιμών που αναφέρθηκαν (ακρίβεια και ευαισθησία), εξετάζουν το πόσα από τα συνολικά δεδομένα ταξινομήθηκαν σωστά και το πόσα από τα ταξινομημένα δεδομένα έχουν ταξινομηθεί σωστά. Συχνά αυτές οι τιμές δεν μπορούν να μας δώσουν μια εικόνα της αποτελεσματικότητας του δικτύου γιατί η μια βγαίνει καλή και η άλλη όχι. Ο τρόπος να συνδυάσουμε τις δύο αυτές τιμές είναι μέσω του αρμονικού τους μέσου F1 score ή αλλιώς Fscore & F-measure.

$$\text{F1} = 2 \cdot \frac{Precision \cdot Recall}{Precision \cdot + Recall}$$

Η τιμή του F1 τείνει να είναι κοντά στη μικρότερη από τις δύο τιμές και μια υψηλή τιμή του σημαίνει ότι η ακρίβεια και ευαισθησία είναι ταυτόχρονα ικανοποιητικά μεγάλες.

(Cluster_analysis, 2019;Accuracy-precision-recall-f1-score-interpretation-ofperformance-measures, 2019)

2.1.9. Κατηγοριοποίηση μεθόδων ανίχνευσης αυτοκτονικού ιδεασμού

Η διάδοση της μηχανικής μάθησης διευκόλυνε την έρευνα σχετικά με την ανίχνευση αυτοκτονικών ιδεών από πολυτροπικά δεδομένα και παρείχε έναν πολλά υποσχόμενο τρόπο για αποτελεσματική έγκαιρη προειδοποίηση αυτοκτονικών τάσεων. Η τρέχουσα έρευνα επικεντρώνεται σε μεθόδους που βασίζονται σε κείμενο εξάγοντας χαρακτηριστικά και βαθιά μάθηση για αυτόματη μάθηση χαρακτηριστικών.

Πολλά κανονικά χαρακτηριστικά NLP όπως TF-IDF, θέματα, συντακτικά, συναισθηματικά χαρακτηριστικά και ρεαλιστικότητα , και μοντέλα βαθιάς μάθησης όπως το CNN και το LSTM χρησιμοποιούνται ευρέως από τους ερευνητές. Αυτές οι μέθοδοι απέκτησαν προκαταρκτική επιτυχία στην ανίχνευση αυτοκτονικού ιδεασμού, αλλά ορισμένες μέθοδοι μπορεί να μάθουν

μόνο στατιστικά στοιχεία και έλλειψη λογικής. Η πρόσφατη εργασία [57] ενσωμάτωσε την εξωτερική γνώση χρησιμοποιώντας βάσεις γνώσεων και οντολογία αυτοκτονίας για την εκτίμηση κινδύνου αυτοκτονίας με γνώση. Πήρε ένα αξιοσημείωτο βήμα προς την ανίχνευση γνώσης.

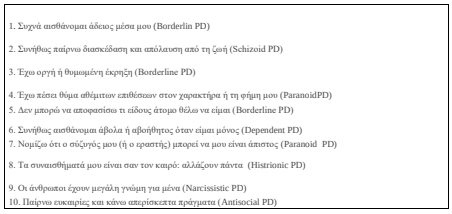
Κατηγορία	Δημοσιεύσεις	Μέθοδοι	Είσοδοι
Μηχανική χαρακτηριστικών	Ji et al. [17]	Word counts, POS, LIWC, TF-IDF + classifiers	Κείμενο
	Masuda et al. [40]	Multivariate/univariate logistic regression	Χαρακτηριστικά μεταβλητών
	Delgado-Gomez et al. [42]	International personal disorder examination screening questionnaire	Απαντήσεις ερωτηματολογίων
	Mulholland et al. [47]	Vocabulary features, syntactic features, semantic class features, N-gram	Στίχοι
	Okhapkina et al. [46]	Dictionary, TF-IDF + SVD	Κείμενο
	Huang et al. [48]	Lexicon, syntactic features, POS, tense	Κείμενο
	Pestian et al. [52]	Word counts, POS, concepts and readability score	Κείμενο
	Tai et al. [50]	self-measurement scale + ANN	Φόρμες αυτομέτρησης
	Shing et al. [13]	BoWs, empath, readability, syntactic, topic, LIWC, emotion, lexicon	Κείμενο
Βαθιά Μάθηση	Zhao et al. [66]	Word embedding, tabular features, D-CNN	Κείμενο +εξωτερικές πληροφορίες
	Shing et al. [13]	Word embedding, CNN, max pooling	Κείμενο
	Ji et al. [17]	Word embedding, LSTM, max pooling	Κείμενο
	Bento et al. [16]	Multi-task learning, neural networks	Κείμενο
	Hevia et al. [62]	Pretrained GRU, word embedding, document embedding	Κείμενο
	Morales et al. [63]	CNN, LSTM, NeuNetS, word embedding	Κείμενο
	Matero et al. [64]	Dual-context, BERT, GRU, attention, user-factor adaptation	Κείμενο
	Gaur et al. [57]	CNN, knowledge base, ConceptNet embedding	Κείμενο
	Coppersmith et al. [58]	GloVe, BiLSTM, self attention	Κείμενο
	Ji et al. [60]	Relation network, LSTM, attention, lexicon	Κείμενο

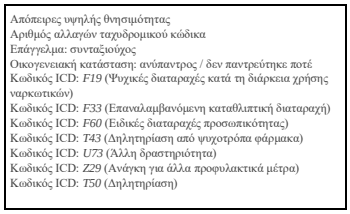
Πίνακας 3: Κατηγοριοποίηση των μεθόδων για την ανίχνευση αυτοκτονικού ιδεασμού

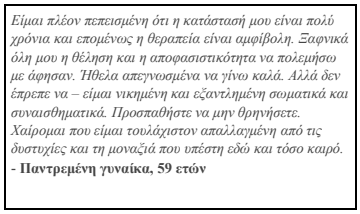
2.2. Εφαρμογές σε Τομείς

Έχουν εισαχθεί πολλές τεχνικές μηχανικής μάθησης για την ανίχνευση αυτοκτονικών ιδεών. Η σχετική υπάρχουσα έρευνα μπορεί επίσης να προβληθεί σύμφωνα με την πηγή δεδομένων. Η συγκεκριμένη εφαρμογή καλύπτει ένα ευρύ φάσμα τομέων, συμπεριλαμβανομένων ερωτηματολογίων, ηλεκτρονικών εγγραφών (EHR), σημειώσεων αυτοκτονίας και διαδικτυακού περιεχομένου χρήστη. Το Σχήμα 8 δείχνει μερικά παραδείγματα πηγής δεδομένων για ανίχνευση αυτοκτονικού ιδεασμού, όπου το Σχήμα 8α παραθέτει επιλεγμένες ερωτήσεις για την «Διεθνή εξέταση προσωπικών διαταραχών Ερωτηματολόγιο «(IPDE-SQ)

προσαρμοσμένο από [42], το Σχήμα 8β είναι επιλεγμένα αρχεία ασθενών από [67], Το Σχήμα 8γ είναι σημείωμα αυτοκτονίας από ιστότοπο[8], και το Σχήμα 8δ είναι ένα tweet και τα αντίστοιχα σχόλιά του από το Twitter.com . Ορισμένοι ερευνητές ανέπτυξαν επίσης λογισμικό για την πρόληψη αυτοκτονιών. Οι Berrouiguet et al. [68] ανέπτυξε μια εφαρμογή για κινητές συσκευές για την κατάσταση υγείας. Οι Meyer et al. [69] ανέπτυξε ένα εργαλείο ανίχνευσης αυτοκτονίας e-PASS (eSID) για ιατρούς.


1. Συχνά αισθάνομαι άδειος μέσα μου (Borderlin PD)
2. Συνήθως παίρνω διασκέδαση και απόλαυση από τη ζωή (Schizoid PD)
3. Έχω οργή ή θυμομένη έκρηξη (Borderline PD)
4. Έχω πέπαι θύμα αθέμιτων επιθέσεων στον χαρακτήρα ή τη φήμη μου (ParanoidPD)
5. Δεν μπορώ να αποφασίσω τι είδους άτομο θέλω να είμαι (Borderline PD)
6. Συνήθως αισθάνομαι άβολα ή αβοήθητος όταν είμαι μόνος (Dependent PD)
7. Νομίζω ότι ο σύζυγός μου (ή ο εραστής) μπορεί να μου είναι άπιστος (Paranoid PD)
8. Τα συναισθήματά μου είναι σαν τον καιρό: αλλάζουν πάντα (Histrionic PD)
9. Οι άνθρωποι έχουν μεγάλη γνώμη για μένα (Narcissistic PD)
10. Παίρνω ευκαιρίες και κάνω απερίσκεπτα πράγματα (Antisocial PD)


Απόπειρες υψηλής θνησιμότητας
Αριθμός αλλαγών ταχυδρομικού κώδικα
Επάγγελμα: συνταξιούχος
Οικογενειακή κατάσταση: ανύπαντρος / δεν παντρεύτηκε ποτέ
Κωδικός ICD: *F19* (Ψυχικές διαταραχές κατά τη διάρκεια χρήσης ναρκωτικών)
Κωδικός ICD: *F33* (Επαναλαμβανόμενη καταθλιπτική διαταραχή)
Κωδικός ICD: *F60* (Ειδικές διαταραχές προσωπικότητας)
Κωδικός ICD: *T43* (Δηλητηρίαση από ψυχοτρόπα φάρμακα)
Κωδικός ICD: *U73* (Άλλη δραστηριότητα)
Κωδικός ICD: *Z29* (Ανάγκη για άλλα προφυλακτικά μέτρα)
Κωδικός ICD: *T50* (Δηλητηρίαση)


Είμαι πλέον πεπεισμένη ότι η κατάστασή μου είναι πολύ χρόνια και επομένως η θεραπεία είναι αμφίβολη. Ξαφνικά όλη μου η θέληση και η αποφασιστικότητα να πολεμήσω με άφησαν. Ήθελα απεγνωσμένα να γίνω καλά. Αλλά δεν έπρεπε να – είμαι νικημένη και εξαντλημένη σωματικά και συναισθηματικά. Προσπαθήστε να μην θρηνήσετε. Χαίρομαι που είμαι τουλάχιστον απαλλαγμένη από τις δυστυχίες και τη μοναξιά που υπέστη εδώ και τόσο καιρό.
- Παντρεμένη γυναίκα, 59 ετών

(α) Ερωτηματολόγιο **(β)** EHR **(γ)** Σημειώσεις αυτοκτονίας **(δ)** Tweets

Σχήμα 8: Παραδείγματα περιεχομένου για ανίχνευση αυτοκτονικού ιδεασμού

2.2.1. Ερωτηματολόγια

Κριτήρια κλίμακας ψυχικής διαταραχής όπως το DSM-IV[9] και ICD-10[10] , και το IPDE-SQ παρέχουν ένα καλό εργαλείο για την αξιολόγηση της ψυχικής κατάστασης ενός ατόμου και της δυνητικότητάς του για αυτοκτονία. Αυτά τα κριτήρια και οι μετρήσεις εξέτασης μπορούν να χρησιμοποιηθούν για το σχεδιασμό ερωτηματολογίων για αυτο-μέτρηση ή πρόσωπο-με-πρόσωπο συνέντευξη ιατρού-ασθενούς.

Για να μελετήσουν την αξιολόγηση της αυτοκτονικής συμπεριφοράς, οι Delgado - Gomez et al. [10] εφάρμοσαν και συνέκριναν το IPDE-SQ και το «Barrat's Impulsiveness Scale» (έκδοση 11, BIS-11) για τον εντοπισμό ατόμων που ενδέχεται να επιχειρήσουν αυτοκτονία. Οι συγγραφείς διεξήγαγαν επίσης μια μελέτη για μεμονωμένα αντικείμενα από αυτές τις δύο κλίμακες. Η κλίμακα BIS-11 έχει 30 στοιχεία με βαθμολογίες 4 πόντων, ενώ το IPDE-SQ στο DSM-IV έχει 77 ερωτήσεις πραγματικού ελέγχου. Περαιτέρω, οι Delgado-Gomez et al. [42]

[8] https://paranorms.com/suicide-notes
[9] https://psychiatry.org/psychiatrists/practice/dsm
[10] https://apps.who.int/classifications/icd10/browse/2016/en

εισήγαγαν την «Κλίμακα Αξιολόγησης Κοινωνικής Αναπροσαρμογής Holmes-Rahe» (SRRS) και το IPDE-SQ, καθώς και σε δύο συγκριτικές ομάδες αυτών που διέπραξαν αυτοκτονία και αυτών που δε διέπραξαν αυτοκτονία. Το SRRS αποτελείται από 43 διαβαθμισμένα συμβάντα ζωής με διαφορετικά επίπεδα σοβαρότητας. Οι Harris et aï. [70] διεξήγαγαν μια έρευνα σχετικά με την κατανόηση των διαδικτυακών συμπεριφορών ατόμων που επιχείρησαν απόπειρα ή τετελεσμένη αυτοκτονία, με σκοπό να βοηθήσουν στην πρόληψη αυτοκτονιών. Ο Sueki [71] διενήργησε μια διαδικτυακή έρευνα πάνελ μεταξύ των χρηστών του Διαδικτύου για να μελετήσει τη σχέση μεταξύ της χρήσης Twitter που σχετίζεται με αυτοκτονία και της αυτοκτονικής συμπεριφοράς. Με βάση τα αποτελέσματα του ερωτηματολογίου, εφάρμοσαν διάφορες εποπτευόμενες μεθόδους μάθησης, όπως γραμμική παλινδρόμηση , σταδιακή γραμμική παλινδρόμηση, δέντρα αποφάσεων, Lars-en και SVMs για την ταξινόμηση αυτοκτονικών συμπεριφορών.

2.2.2. Ηλεκτρονικά Αρχεία Υγείας

Ο αυξανόμενος όγκος ηλεκτρονικών αρχείων υγείας (EHRs) άνοιξε το δρόμο για τεχνικές μηχανικής μάθησης για πρόβλεψη αυτοκτονίας. Τα αρχεία των ασθενών περιλαμβάνουν δημογραφικές πληροφορίες και ιστορικό που σχετίζεται με τη διάγνωση, τις εισαγωγές και επείγουσες επισκέψεις . Ωστόσο, λόγω των χαρακτηριστικών δεδομένων όπως η αραιότητα, το μεταβλητό μήκος των κλινικών σειρών και η ετερογένεια των αρχείων των ασθενών, πολλές προκλήσεις παραμένουν στη μοντελοποίηση ιατρικών δεδομένων για την πρόβλεψη αυτοκτονίας. Επιπλέον, οι διαδικασίες καταγραφής αλλάζουν λόγω της αλλαγής των πολιτικών υγειονομικής περίθαλψης και της ενημέρωσης των κωδικών διάγνωσης.

Υπάρχουν πολλά έργα πρόβλεψης του κινδύνου αυτοκτονίας βάσει των EHRs [72,73]. Οι Tran et al. [67] πρότειναν ένα ολοκληρωμένο πλαίσιο πρόβλεψης κινδύνου αυτοκτονίας με σχήμα εξαγωγής χαρακτηριστικών, ταξινομητές κινδύνου και διαδικασία βαθμονόμησης κινδύνου. Συγκεκριμένα, το κλινικό ιστορικό κάθε ασθενούς αντιπροσωπεύεται ως χρονική εικόνα. Οι Iliou et al. [74] πρότειναν μια μέθοδο προεπεξεργασίας δεδομένων για την ενίσχυση τεχνικών μηχανικής μάθησης για πρόβλεψη τάσης αυτοκτονίας ασθενών που πάσχουν από ψυχικές διαταραχές. Οι Nguyen et al. [75] διερεύνησαν πραγματικά διοικητικά δεδομένα ασθενών ψυχικής υγείας από νοσοκομείο για βραχυπρόθεσμες και μεσοπρόθεσμες αξιολογήσεις

κινδύνου αυτοκτονίας. Με την εισαγωγή τυχαίων δασών, μηχανών ενίσχυσης κλίσης και DNN, οι συγγραφείς κατάφεραν να ασχοληθούν με θέματα υψηλής διάστασης και πλεονασμού δεδομένων. Αν και η προηγούμενη μέθοδος απέκτησε προκαταρκτική επιτυχία, οι Iliou et al. [74] και Nguyen et al. [75] έχουν έναν περιορισμό στην πηγή δεδομένων που επικεντρώνεται σε ασθενείς με ψυχικές διαταραχές στα ιστορικά τους αρχεία. Οι Bhat και Goldman-Mellor [76] χρησιμοποίησαν ένα ανώνυμο γενικό σύνολο δεδομένων EHR για να χαλαρώσουν τον περιορισμό του ιστορικού που σχετίζεται με τη διάγνωση του ασθενούς, και εφάρμοσαν τα νευρωνικά δίκτυα ως μοντέλο ταξινόμησης για να προβλέψουν απόπειρες αυτοκτονίας.

2.2.3. Σημειώσεις Αυτοκτονίας

Οι σημειώσεις αυτοκτονίας είναι οι γραπτές σημειώσεις που άφησαν οι άνθρωποι πριν αυτοκτονήσουν. Συνήθως γράφονται σε γράμματα και σε απευθείας σύνδεση ιστολόγια και καταγράφονται σε ήχο ή βίντεο. Οι σημειώσεις αυτοκτονίας παρέχουν υλικό για την έρευνα NLP. Προηγούμενες προσεγγίσεις έχουν εξετάσει τις σημειώσεις αυτοκτονίας με τη χρήση ανάλυσης περιεχομένου [52], ανάλυσης συναισθήματος [44,77], και ανίχνευσης συγκίνησης [51]. Οι Pestian et al. [52] χρησιμοποίησαν μεταγραμμένες σημειώσεις αυτοκτονίας με δύο ομάδες συμπληρωτών και εκκινητών από άτομα που έχουν διαταραχή της προσωπικότητας ή πιθανές νοσηρές σκέψεις. Οι White και Mazlack [78] ανέλυσαν τις συχνότητες λέξεων σε σημειώσεις αυτοκτονίας χρησιμοποιώντας έναν ασαφή γνωστικό χάρτη για να διακρίνουν την αιτιότητα. Οι Liakata et al. [51] χρησιμοποίησαν ταξινομητές μηχανικής μάθησης σε 600 μηνύματα αυτοκτονίας με ποικίλο μήκος, διαφορετική ποιότητα αναγνωσιμότητας και σχολιασμούς πολλαπλών τάξεων.

Το συναίσθημα στο κείμενο παρέχει συναισθηματικές ενδείξεις κατανόησης αυτοκτονίας. Οι Desmet et al. [79] πραγματοποίησαν μια λεπτομερής ανίχνευση συναισθημάτων σε σημειώσεις αυτοκτονίας του έργου i2b2 του 2011. Ο Wicentowski and Sydes [80] χρησιμοποίησαν ένα σύνολο μέγιστης ταξινόμησης εντροπίας. Οι Wang et al. [44] και Kovaceviĉ et´ al. [81] πρότειναν υβριδική μηχανική μάθηση και μέθοδο βάσει κανόνων για την εργασία ταξινόμησης συναισθημάτων i2b2 σε σημειώσεις αυτοκτονίας.

Στην εποχή του κυβερνοχώρου, οι σημειώσεις αυτοκτονίας γράφονται πλέον σε διαδικτυακά ιστολόγια και μπορούν να αναγνωριστούν ότι ενέχουν τον πιθανό κίνδυνο αυτοκτονίας. Οι

Huang et al. [29] παρακολούθησαν διαδικτυακά ιστολόγια από το MySpace.com για τον εντοπισμό bloggers σε κίνδυνο. Οι Schoene και Dethlefs [82] εξήγαγαν γλωσσικά και συναισθηματικά χαρακτηριστικά για την ταυτότητα γνήσιων σημειώσεων αυτοκτονίας και σύγκρισης.

2.2.4. Διαδικτυακό Περιεχόμενο χρήστη

Η διάδοση των υπηρεσιών κινητής τηλεφωνίας στο Διαδίκτυο και της κοινωνικής δικτύωσης διευκολύνει τους ανθρώπους να εκφράζουν ελεύθερα τα δικά τους γεγονότα και συναισθήματα ζωής. Καθώς οι κοινωνικοί ιστότοποι παρέχουν έναν ανώνυμο χώρο για διαδικτυακή συζήτηση, ένας αυξανόμενος αριθμός ατόμων που πάσχουν από ψυχικές διαταραχές στρέφονται για να ζητήσουν βοήθεια. Υπάρχει μια ανησυχητική τάση ότι τα πιθανά θύματα αυτοκτονίας δημοσιεύουν τις αυτοκτονικές τους σκέψεις σε κοινωνικούς ιστότοπους όπως το Facebook, το Twitter, το Reddit και το MySpace. Οι πλατφόρμες κοινωνικών μέσων αποτελούν μια πολλά υποσχόμενη σήραγγα για την παρακολούθηση αυτοκτονικών σκέψεων και την πρόληψη των προσπαθειών αυτοκτονίας [83]. Τα μαζικά δεδομένα που δημιουργούνται από τους χρήστες παρέχουν μια καλή πηγή για τη μελέτη της γλώσσας των χρηστών του Διαδικτύου . Χρήση τεχνικών εξόρυξης δεδομένων στα κοινωνικά δίκτυα και η εφαρμογή τεχνικών μηχανικής μάθησης παρέχουν έναν δρόμο για την κατανόηση της πρόθεσης εντός των διαδικτυακών δημοσιεύσεων, την παροχή έγκαιρων προειδοποιήσεων και ακόμη και την ανακούφιση από τις αυτοκτονικές προθέσεις ενός ατόμου.

Το Twitter παρέχει μια καλή πηγή έρευνας για αυτοκτονία. Οι O'Dea et al. [12] συνέλεξαν tweets χρησιμοποιώντας το δημόσιο API και ανέπτυξαν την αυτόματη ανίχνευση αυτοκτονίας εφαρμόζοντας λογιστική παλινδρόμηση και SVM σε χαρακτηριστικά TF-IDF. Οι Wang et al. [84] βελτίωσαν περαιτέρω την απόδοση με την αποτελεσματική μηχανική χαρακτηριστικών. Οι Shepherd et al. [85] διεξήγαγαν ανάλυση δεδομένων με βάση την ψυχολογία για περιεχόμενο που υποδηλώνει τάσεις αυτοκτονίας στα κοινωνικά δίκτυα του Twitter. Οι συγγραφείς χρησιμοποίησαν τα δεδομένα από μια διαδικτυακή συνομιλία που ονομάζεται #dearmentalhealthprofessionals.

Μια άλλη διάσημη πλατφόρμα είναι το Reddit, το οποίο είναι ένα διαδικτυακό φόρουμ με συζητήσεις για συγκεκριμένα θέματα, έχει επίσης προσελκύσει μεγάλο ερευνητικό ενδιαφέρον

για τη μελέτη θεμάτων ψυχικής υγείας [86] και τον αυτοκτονικό ιδεασμό [38]. Μια κοινότητα στο Reddit που ονομάζεται SuicideWatch χρησιμοποιείται εντατικά για τη μελέτη της πρόθεσης αυτοκτονίας [17,87]. Οι De Choudhury et al. [87] εφάρμοσαν μια στατιστική μεθοδολογία για να ανακαλύψουν τη μετάβαση από ζητήματα ψυχικής υγείας στην αυτοκτονία. Οι Kumar et al. [88] εξέτασαν τη δραστηριότητα δημοσίευσης μετά τις αυτοκτονίες διασημοτήτων, μελέτησαν την επίδραση των αυτοκτονιών διασημοτήτων στο περιεχόμενο που σχετίζεται με αυτοκτονίες και πρότειναν μια μέθοδο για την πρόληψη των αυτοκτονιών υψηλού προφίλ.

Πολλές έρευνες [48,49] εργάζονται για τον εντοπισμό αυτοκτονικού ιδεασμού σε κινεζικά microblogs. Οι Guan et al. [89] μελέτησαν το προφίλ χρήστη και τις γλωσσικές δυνατότητες για την εκτίμηση της πιθανότητας αυτοκτονίας στα κινέζικα microblogs. Παραμένει επίσης κάποια εργασία χρησιμοποιώντας άλλες πλατφόρμες για ανίχνευση αυτοκτονικών ιδεών. Για παράδειγμα, οι Cash et al. [90] διεξήγαγαν μια μελέτη σχετικά με τα σχόλια των εφήβων και την ανάλυση περιεχομένου στο MySpace. Τα Steaming δεδομένα παρέχουν μια καλή πηγή ανάλυσης προτύπων χρήστη. Οι Vioules et al. [5] Διενήργησαν user-centric και post-centric ανάλυση συμπεριφοράς και εφάρμοσαν ένα πλαίσιο για τον εντοπισμό ξαφνικών συναισθηματικών αλλαγών στη ροή δεδομένων του Twitter για την παρακολούθηση ειδικών προειδοποιητικών σημείων. Η ροή ιστολογίου που συλλέχθηκε από δημόσια άρθρα ιστολογίων που γράφτηκαν από θύματα αυτοκτονίας χρησιμοποιήθηκε από τους Ren et al. [14] για να μελετήσουν τις συσσωρευμένες συναισθηματικές πληροφορίες.

2.3. Σύνοψη

Η πρόληψη των αυτοκτονιών παραμένει σημαντικό καθήκον στη σύγχρονη κοινωνία μας. Η έγκαιρη ανίχνευση αυτοκτονικού ιδεασμού είναι ένας σημαντικός και αποτελεσματικός τρόπος για την πρόληψη της αυτοκτονίας. Αυτή η διπλωματική εργασία διερευνά υπάρχουσες μεθόδους για την ανίχνευση αυτοκτονικού ιδεασμού από μια ευρεία προοπτική που καλύπτει κλινικές μεθόδους, όπως αλληλεπίδραση ασθενών με κλινικό ιατρό και ανίχνευση ιατρικού σήματος, ανάλυση περιεχομένου κειμένου, όπως φιλτράρισμα με βάση λεξικά και οπτικοποίηση νεφών λέξεων, μηχανική χαρακτηριστικών, συμπεριλαμβανομένων των πινάκων, των κειμένων και των συναισθηματικών χαρακτηριστικών και εκμάθηση

αναπαραγωγής με βάση τη βαθιά μάθηση, όπως κωδικοποιητές κειμένου που βασίζονται σε CNN και LSTM.

Οι εφαρμογές ανίχνευσης αυτοκτονικού ιδεασμού αποτελούνται κυρίως από τέσσερις τομείς , δηλαδή ερωτηματολόγια, ηλεκτρονικά αρχεία υγείας, σημειώσεις αυτοκτονίας και διαδικτυακό περιεχόμενο χρήστη. Ο Πίνακας 4 παρέχει μια σύντομη περίληψη των κατηγοριών, των πηγών δεδομένων και των μεθόδων. Μεταξύ αυτών των τεσσάρων κύριων τομέων, τα ερωτηματολόγια και τα EHR απαιτούν μέτρηση αυτοαναφοράς ή αλληλεπιδράσεις ασθενών-κλινικών και βασίζονται σε μεγάλο βαθμό σε κοινωνικούς λειτουργούς ή στα επαγγέλματα ψυχικής υγείας. Οι σημειώσεις αυτοκτονίας έχουν περιορισμό στην άμεση πρόληψη, καθώς πολλοί δράστες απόπειρας αυτοκτονίας αυτοκτονούν σε σύντομο χρονικό διάστημα μετά τη σύνταξη σημειώσεων αυτοκτονίας. Ωστόσο, παρέχουν μια καλή πηγή για ανάλυση περιεχομένου και τη μελέτη παραγόντων αυτοκτονίας. Ο τελευταίος τομέας διαδικτυακού περιεχομένου χρήστη είναι ένας από τους πιο ελπιδοφόρους τρόπους έγκαιρης προειδοποίησης και πρόληψης αυτοκτονιών, όταν είναι εξουσιοδοτημένοι με τεχνικές μηχανικής μάθησης. Με την ταχεία ανάπτυξη της ψηφιακής τεχνολογίας, το περιεχόμενο που δημιουργείται από τον χρήστη θα διαδραματίσει σημαντικότερο ρόλο στην ανίχνευση αυτοκτονικών ιδεών και άλλες μορφές δεδομένων, όπως τα δεδομένα υγείας που παράγονται από φορητές συσκευές, είναι πολύ πιθανό να βοηθήσουν στην παρακολούθηση του κινδύνου αυτοκτονίας στο κοντινό μέλλον.

Κατηγορίες	εξέταση αυτό-αναφορών πρόσωπο-με-πρόσωπο πρόληψη αυτοκτονίας αυτόματη ανίχνευση αυτοκτονικού ιδεασμού
δεδομένα	ερωτηματολόγια, σημειώσεις αυτοκτονίας, ιστολόγια αυτοκτονίας, ηλεκτρονικά αρχεία υγείας, online κοινωνικά κείμενα
μέθοδοι	Κλινικές μέθοδοι, ανάλυση περιεχομένου, μηχανική χαρακτηριστικών, βαθιά μάθηση

Πίνακας 4: Σύνοψη μελετών για την ανίχνευση αυτοκτονικών ιδεών

Παρουσιάζονται τέσσερις κύριες εφαρμογές για συγκεκριμένους τομείς σε ερωτηματολόγια, EHR, σημειώσεις αυτοκτονίας και διαδικτυακό περιεχόμενο χρήστη.

Οι περισσότερες εργασίες σε αυτόν τον τομέα έχουν διεξαχθεί από ψυχολόγους εμπειρογνώμονες με στατιστική ανάλυση, και επιστήμονες υπολογιστών με μηχανική μάθηση βασισμένη στη μηχανική χαρακτηριστικών γνώσεων και μάθηση εκπροσώπησης με βάση τη βαθιά μάθηση. Με βάση την τρέχουσα έρευνα, συνοψίσαμε τις υπάρχουσες εργασίες και προτείνουμε περαιτέρω νέες πιθανές εργασίες . Τέλος, συζητάμε ορισμένους περιορισμούς της τρέχουσας έρευνας και προτείνουμε μια σειρά μελλοντικών κατευθύνσεων, συμπεριλαμβανομένης της χρήσης αναδυόμενων τεχνικών μάθησης, της ερμηνεύσιμης κατανόησης προθέσεων, της χρονικής ανίχνευσης και της προληπτικής συνομιλίας.

Κεφάλαιο 3

3. Συγκριτική αξιολόγηση για τον εντοπισμό αυτοκτονικών ιδεών

3.1. Εισαγωγή

Η αυτοκτονία μπορεί να θεωρηθεί ως ένα από τα σοβαρότερα προβλήματα κοινωνικής υγείας στη σύγχρονη κοινωνία. Πολλοί παράγοντες μπορούν να οδηγήσουν σε αυτοκτονία, π.χ. προσωπικά ζητήματα, όπως απελπισία, έντονο άγχος, σχιζοφρένεια, αλκοολισμός ή παρορμητικότητα. Κοινωνικοί παράγοντες, όπως η κοινωνική απομόνωση, η υπερβολική έκθεση σε θανάτους. Ή αρνητικά γεγονότα ζωής, συμπεριλαμβανομένων τραυματικών γεγονότων, σωματικών ασθενειών, συναισθηματικών διαταραχών και προηγούμενων απόπειρών αυτοκτονίας. Χιλιάδες άνθρωποι σε όλο τον κόσμο πέφτουν θύματα αυτοκτονίας κάθε χρόνο, κάνοντας την πρόληψη της αυτοκτονίας να γίνει μια κρίσιμη παγκόσμια αποστολή δημόσιας υγείας.

Ο αυτοκτονικός ιδεασμός ή οι σκέψεις αυτοκτονίας είναι οι σκέψεις των ανθρώπων να αυτοκτονήσουν. Μπορεί να θεωρηθεί ως δείκτης κινδύνου αυτοκτονίας. Οι αυτοκτονικές

σκέψεις περιλαμβάνουν φευγαλέες σκέψεις, εκτεταμένες σκέψεις, λεπτομερή προγραμματισμό, παιχνίδι ρόλων, ελλιπείς προσπάθειες και ούτω καθεξής. Σύμφωνα με έκθεση του ΠΟΥ [91], 788.000 άνθρωποι εκτιμάται ότι αυτοκτόνησαν παγκοσμίως το 2015. Και ένας μεγάλος αριθμός ανθρώπων, ειδικά εφήβων, αναφέρθηκε ότι είχε αυτοκτονικό ιδεασμό. Έτσι, μια πιθανή προσέγγιση για την αποτελεσματική πρόληψη της αυτοκτονίας είναι η έγκαιρη ανίχνευση αυτοκτονικού ιδεασμού.

Με την ευρεία εμφάνιση τεχνολογιών κινητής τηλεφωνίας στο Διαδίκτυο και διαδικτυακών κοινωνικών δικτύων, υπάρχει μια αυξανόμενη τάση οι άνθρωποι να μιλούν για τις προθέσεις αυτοκτονίας τους σε διαδικτυακές κοινότητες. Αυτό το διαδικτυακό περιεχόμενο θα μπορούσε να είναι χρήσιμο για τον εντοπισμό των προθέσεων των ατόμων και την αυτοκτονική σκέψη τους. Μερικοί άνθρωποι, ειδικά οι έφηβοι, επιλέγουν να δημοσιεύουν τις αυτοκτονικές σκέψεις τους στα κοινωνικά δίκτυα, να ρωτούν πώς να αυτοκτονήσουν σε διαδικτυακές κοινότητες και να συνάπτουν διαδικτυακές συμφωνίες αυτοκτονίας. Η ανωνυμία της διαδικτυακής επικοινωνίας επιτρέπει επίσης στους ανθρώπους να εκφράζουν ελεύθερα τις πιέσεις και το άγχος που υφίστανται στον πραγματικό κόσμο. Αυτό το διαδικτυακό περιεχόμενο που δημιουργείται από τους χρήστες παρέχει μια άλλη πιθανή γωνία για έγκαιρη ανίχνευση και πρόληψη αυτοκτονιών.

Προηγούμενη έρευνα για την κατανόηση και την πρόληψη της αυτοκτονίας επικεντρώνεται κυρίως στις ψυχολογικές και κλινικές πτυχές της [9]. Πρόσφατα, πολλές μελέτες έχουν στραφεί σε μεθόδους επεξεργασίας φυσικής γλώσσας και ταξινόμηση αποτελεσμάτων ερωτηματολογίου μέσω εποπτευόμενης μάθησης, η οποία μαθαίνει μια λειτουργία χαρτογράφησης από 15 επισημασμένα δεδομένα κατάρτισης [92]. Ορισμένες από αυτές τις έρευνες έχουν χρησιμοποιήσει τη «Διεθνή Προσωπική Εξέταση Ερωτηματολόγιο Προβολής(Screening Questionnaire) », και ανέλυσαν ιστολόγια αυτοκτονίας και δημοσιεύσεις από ιστότοπους κοινωνικής δικτύωσης. Ωστόσο, αυτές οι μελέτες έχουν τους περιορισμούς τους. (1) τόσο από ψυχολογική όσο και από κλινική άποψη, η συλλογή δεδομένων και/ή των ασθενών είναι συνήθως δαπανηρή και ορισμένα διαδικτυακά δεδομένα μπορεί να βοηθήσουν στην κατανόηση σκέψεων και συμπεριφορών (2) τα απλά σύνολα χαρακτηριστικών και τα μοντέλα ταξινόμησης δεν είναι αρκετά προγνωστικά για να ανιχνεύσουν Τάσεις αυτοκτονίας.

Σε αυτήν την ενότητα, ερευνούμε το πρόβλημα της ανίχνευσης αυτοκτονικών ιδεών σε διαδικτυακούς κοινωνικούς ιστότοπους, με έμφαση στην κατανόηση και τον εντοπισμό των αυτοκτονικών σκέψεων στο περιεχόμενο των χρηστών στο διαδίκτυο. Πραγματοποιούμε μια ενδελεχή ανάλυση του περιεχομένου, των γλωσσικών προτιμήσεων και των περιγραφών του θέματος για να κατανοήσουμε τις σκέψεις αυτοκτονίας από την άποψη της εξόρυξης δεδομένων. Έξι διαφορετικά σύνολα πληροφοριακών χαρακτηριστικών εξήχθησαν και έξι εποπτευόμενοι αλγόριθμοι μάθησης συγκρίθηκαν για να ανιχνεύσουν τον αυτοκτονικό ιδεασμό στα δεδομένα. Είναι μια νέα εφαρμογή αυτόματης ανίχνευσης πρόθεσης αυτοκτονίας σε κοινωνικό περιεχόμενο με το συνδυασμό των προτεινόμενων αποτελεσματικών μοντέλων μηχανικής και ταξινόμησης χαρακτηριστικών.

Αυτό το τμήμα κάνει αξιοσημείωτες συνεισφορές και καινοτομίες στη βιβλιογραφία στα παρακάτω ακόλουθα σημεία:

1. **<u>Ανακάλυψη γνώσης:</u>** Αυτή είναι μια νέα εφαρμογή ανακάλυψης γνώσης και εξόρυξης δεδομένων για τον εντοπισμό αυτοκτονικών ιδεών στο περιεχόμενο των διαδικτυακών χρηστών. Προηγούμενες εργασίες σε αυτόν τον τομέα έχουν διεξαχθεί από ψυχολόγους ειδικούς στη στατιστική ανάλυση. Αυτή η προσέγγιση αποκαλύπτει γνώσεις σχετικά με τον αυτοκτονικό ιδεασμό από την άποψη της ανάλυσης δεδομένων. Πληροφορίες από την ανάλυση μας αποκαλύπτουν ότι τα αυτοκτονικά άτομα συχνά χρησιμοποιούν προσωπικές αντωνυμίες για να δείξουν το εγώ τους. Είναι πιο πιθανό να χρησιμοποιούν λέξεις που εκφράζουν αρνητικότητα, άγχος και θλίψη στο διάλογό τους. Είναι επίσης πιο πιθανό να επιλέξουν τον ενεστώτα για να περιγράψουν τα βάσανά τους και τον μελλοντικό χρόνο για να περιγράψουν την απελπισία και τα σχέδιά τους για αυτοκτονία.

2. **<u>Σύνολο δεδομένων και πλατφόρμα:</u>** Αυτή η ενότητα παρουσιάζει την πλατφόρμα Reddit και συλλέγει ένα νέο σύνολο δεδομένων για τον εντοπισμό αυτοκτονικών ιδεών. Το Reddit’s SuicideWatch BBS είναι ένα νέο διαδικτυακό κανάλι για άτομα με αυτοκτονικό ιδεασμό για να εκφράσουν το άγχος και τις πιέσεις τους. Οι κοινωνικοί εθελοντές ανταποκρίνονται με θετικούς, υποστηρικτικούς τρόπους για να ανακουφίσουν την κατάθλιψη και ελπίζουμε να αποτρέψουν πιθανές αυτοκτονίες. Αυτή η πηγή δεδομένων δεν είναι χρήσιμη μόνο για τον εντοπισμό αυτοκτονίας αλλά και για τη μελέτη του τρόπου

αποτελεσματικής πρόληψης της αυτοκτονίας μέσω αποτελεσματικής διαδικτυακής επικοινωνίας.

3. **<u>Χαρακτηριστικά, Μοντέλα και Συγκριτική Αξιολόγηση:</u>** Αντί να χρησιμοποιείτε βασικά μοντέλα με απλά χαρακτηριστικά για τον εντοπισμό αυτοκτονικών ιδεών, αυτή η προσέγγιση (1) προσδιορίζει πληροφοριακά χαρακτηριστικά από διάφορες προοπτικές, συμπεριλαμβανομένων στατιστικών, συντακτικών, γλωσσικών, χαρακτηριστικών ενσωμάτωσης λέξεων , και χαρακτηριστικά θέματος. (2) συγκρίνει με διαφορετικούς ταξινομητές τόσο από την παραδοσιακή όσο και από τη σκοπιά της βαθιάς μάθησης, όπως μηχανή διανύσματος υποστήριξης [93], τυχαίο δάσος [94], gradient boost δέντρο ταξινόμησης (GBDT) [95], XGBoost [96], MLFFNN [43] και Μακροπρόθεσμη Μνήμη (LSTM) [97]. Και (3) παρέχει σημεία αναφοράς για την ανίχνευση αυτοκτονικών ιδεών στο SuicideWatch στο Reddit, ένα ενεργό διαδικτυακό φόρουμ για επικοινωνία σχετικά με την αυτοκτονία.

Αυτή η ενότητα οργανώνεται ως εξής: Εισάγουμε τα σύνολα δεδομένων στην Ενότητα 3.2 μαζί με την εξερεύνηση δεδομένων και την ανακάλυψη γνώσης. Η ενότητα 3.3 περιγράφει την ταξινόμηση και τη μέθοδο εξαγωγής χαρακτηριστικών. Το τμήμα 3.4 είναι η πειραματική μελέτη. Ολοκληρώνουμε αυτήν την ενότητα στο 3.5.

3.2. Δεδομένα και Γνώση

Συλλέγουμε τα κείμενα αυτοκτονικού ιδεασμού από το Reddit και το Twitter και ελέγχουμε χειροκίνητα όλες τις αναρτήσεις για να βεβαιωθούμε ότι έχουν επισημανθεί σωστά. Οι κανόνες σχολιασμού και παραδείγματα αναρτήσεων εμφανίζονται στον Πίνακα 5:

Κατηγορίες	Κανόνες	Παραδείγματα
Κείμενο αυτοκτονίας	• Έκφραση αυτοκτονικών σκέψεων • Συμπεριλαμβανομένων πιθανών αυτοκτονικών ενεργειών	*Θέλω να τελειώσω τη ζωή μου απόψε.* *Χθες, προσπάθησα να κόψω τον καρπό μου, αλλά δεν τα κατάφερα.*

Κείμενο μη-αυτοκτονίας	• Επίσημη συζήτηση για αυτοκτονία • Αναφορά στην αυτοκτονία άλλων • Δεν έχει σχέση με την αυτοκτονία	*Το παγκόσμιο ποσοστό αυτοκτονιών αυξάνεται.* *Είμαι τόσο λυπημένος που ακούω ότι ο Robin Williams έβαλε τέλος στη ζωή του.* *Λατρεύω αυτήν την τηλεοπτική εκπομπή και παρακολουθώ κάθε εβδομάδα.*

Πίνακας 5: Κανόνες σχολιασμού και παραδείγματα κοινωνικών κειμένων

3.2.1. Σύνολο Δεδομένων Reddit

Το Reddit είναι μια εγγεγραμμένη διαδικτυακή κοινότητα που συγκεντρώνει κοινωνικές ειδήσεις και διαδικτυακές συζητήσεις. Αποτελείται από πολλές κατηγορίες θεμάτων και κάθε τομέας ενδιαφέροντος μέσα σε ένα θέμα ονομάζεται subreddit.

Σε αυτό το σύνολο δεδομένων, το περιεχόμενο των χρηστών στο διαδίκτυο περιλαμβάνει έναν τίτλο και ένα μέρος κειμένου. Για να διατηρήσουμε το απόρρητο, αντικαθιστούμε τις προσωπικές πληροφορίες με ένα μοναδικό αναγνωριστικό για την ταυτοποίηση του κάθε χρήστη. Συλλέξαμε δημοσιεύσεις με πιθανές προθέσεις αυτοκτονίας από ένα subreddit που ονομάζεται «Suicide Watch» (SW) [11] . Αναρτήσεις χωρίς αυτοκτονικό περιεχόμενο προέρχονταν από άλλα δημοφιλείς subreddits[12] [13]. Η συλλογή μη αυτοκτονικών δεδομένων είναι περιεχόμενο που δημιουργείται εξ ολοκλήρου από χρήστες και αποκλείονται οι δημοσιεύσεις συγκεντρώσεων ειδήσεων και διαχειριστή. Για να διευκολύνουμε τη μελέτη και την επίδειξη, θα μελετήσουμε το ισορροπημένο σύνολο δεδομένων στο Reddit και θα μελετήσουμε μη ισορροπημένα σύνολα δεδομένων στο Twitter όπως φαίνεται στην παρακάτω υποενότητα.

Το σύνολο δεδομένων Reddit περιλαμβάνει 3.549 δείγματα αυτοκτονικού ιδεασμού και πολλά κείμενα μη αυτοκτονίας. Συγκεκριμένα, κατασκευάζουμε δύο σύνολα δεδομένων για το Reddit που φαίνονται στον Πίνακα 6 Το πρώτο σύνολο δεδομένων περιλαμβάνει δύο subreddits, στα οποία το ένα προέρχεται από το suicideWatch και ένα άλλο από δημοφιλείς αναρτήσεις στο

[11] https: // w ww.reddit.com/r/SuicideWatch/
[12] https://www.reddit.com/r/all/
[13] https://www.reddit.com/r/popular/

Reddit. Το δεύτερο σύνολο δεδομένων αποτελείται από έξι δευτερεύοντα στοιχεία που περιλαμβάνουν το SuicideWatch και άλλα πέντε καυτά θέματα: Gaming[14], Jokes [15], Books [16], Movies [17] και AskReddit[18]. Στο δεύτερο σύνολο δεδομένων, ο συνδυασμός του SuicideWatch με οποιοδήποτε άλλο subreddit θα είναι ένα νέο ισοσκελισμένο υποσύνολο δεδομένων, για παράδειγμα, Suicide vs. Gaming και Suicide vs. Jokes. Αυτά τα δύο σύνολα δεδομένων θα μελετηθούν χωριστά στις υποενότητες 3.4.1 και 3.4.2 .

Dataset	Subreddits
1	SuicideWatch vs. Others (Non-suicide)
2	SuicideWatch vs. Gaming SuicideWatch vs. Jokes SuicideWatch vs. Books SuicideWatch vs. Movies SuicideWatch vs. AskReddit

Πίνακας 6: Δύο ισορροπημένα σύνολα δεδομένων Reddit

3.2.2. Σύνολο Δεδομένων Twitter

Πολλοί διαδικτυακοί χρήστες θέλουν επίσης να μιλήσουν για τον αυτοκτονικό ιδεασμό στα κοινωνικά δίκτυα. Ωστόσο, το Twitter είναι αρκετά διαφορετικό από το Reddit καθώς 1) το μήκος κάθε tweet περιορίζεται σε 140 χαρακτήρες[19] , 2) Οι χρήστες του tweet μπορεί να έχουν φίλους από τα κοινωνικά δίκτυα από τον πραγματικό κόσμο, ενώ οι χρήστες του Reddit είναι πλήρως ανώνυμοι, 3) ο τύπος επικοινωνίας και αλληλεπίδρασης είναι εντελώς διαφορετικός μεταξύ ιστότοπων κοινωνικής δικτύωσης και διαδικτυακών φόρουμ.

Το σύνολο δεδομένων Twitter συλλέγεται χρησιμοποιώντας μια τεχνική φιλτραρίσματος λέξεων-κλειδιών. Οι αυτοκτονικές λέξεις και φράσεις περιλαμβάνουν «αυτοκτονία», «πεθαίνω», «τερματίζω τη ζωή μου» και ούτω καθεξής. Πολλά από τα Tweets που συλλέγονται

[14] https://www.reddit.com/r/gaming/
[15] https://www.reddit.com/r/Jokes/
[16] https: //www.reddit.com/r/books/
[17] https://www.reddit.com/r/movies/
[18] https://www.reddit.com/r/AskReddit/
[19] Αυτό το όριο είναι τώρα 280 χαρακτήρες.

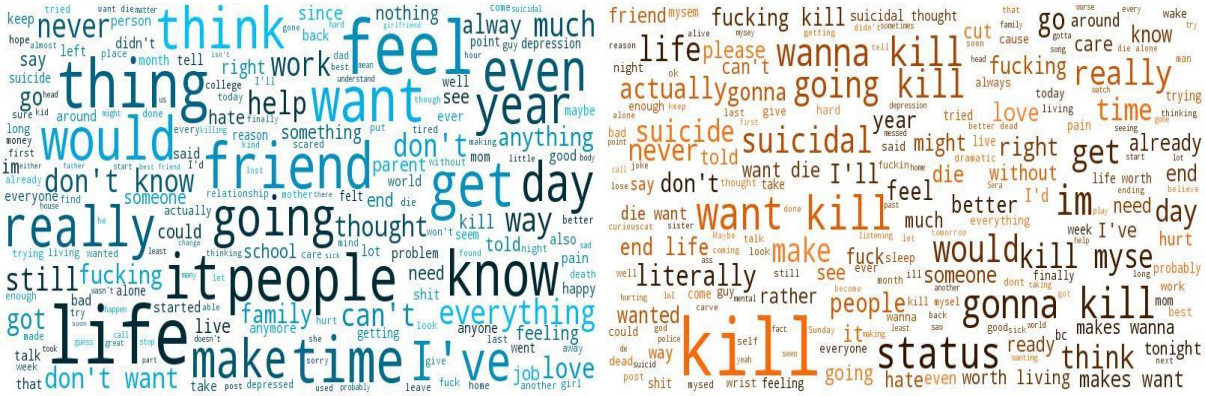
έχουν τις σχετικές με την αυτοκτονία λέξεις, αλλά πιθανότατα μιλούν για ταινία αυτοκτονίας ή διαφήμιση που δεν περιέχει αυτοκτονικό ιδεασμό. Επομένως, ελέγξαμε χειροκίνητα και επισημάναμε τα συλλεγμένα Tweets σύμφωνα με τους κανόνες σχολιασμού στον Πίνακα 5. Τέλος, το σύνολο δεδομένων Twitter έχει συνολικά 10.288 tweets με 594 tweets (περίπου 6%) με αυτοκτονικό ιδεασμό. Αυτό το σύνολο δεδομένων είναι ένα μη ισορροπημένο σύνολο δεδομένων και θα μελετηθεί στην Ενότητα 3.4.3.

3.2.3. Εξερεύνηση δεδομένων και ανακάλυψη γνώσης

Για να κατανοήσουμε τα άτομα που αυτοκτόνησαν, αναλύσαμε τις λέξεις, τις γλώσσες και τα θέματα στο περιεχόμενο των διαδικτυακών αναρτήσεων των χρηστών του διαδικτύου.

Word-cloud (σύννεφα λέξεων). Τα σύννεφα λέξεων χρησιμοποιήθηκαν για να παρέχουν μια οπτική κατανόηση των δεδομένων. Οι αναρτήσεις των χρηστών στο Reddit και τα tweets στο Twitter με πιθανό κίνδυνο αυτοκτονίας εμφανίζονται χωριστά στα Σχήματα. 9α και 9β. Όπως μπορούμε να δούμε, οι αυτοκτονικές αναρτήσεις χρησιμοποιούν συχνά λέξεις όπως «ζωή», «αυτοκτονία» και «σκοτώνω», παρέχοντας μια άμεση ένδειξη των αυτοκτονικών σκέψεων των χρηστών. Συχνά χρησιμοποιούνται επίσης λέξεις που εκφράζουν συναισθήματα ή προθέσεις, όπως «αισθάνομαι», «θέλω» και «γνωρίζω». Για παράδειγμα, ορισμένες αυτοκτονικές αναρτήσεις έγραφαν: «Νιώθω ότι δεν μου έχει μείνει κανένας και θέλω να το τελειώσω», «Θέλω να τελειώσω τη ζωή μου» και «Δεν ξέρω πόσο από αυτό ήταν ψυχολογικό τραύμα ».

Επιπλέον, οι κυρίαρχες λέξεις σε αυτές τις δύο κοινωνικές πλατφόρμες έχουν διαφορετικά στυλ λόγω των διατάξεων δημοσίευσης των πλατφορμών. Οι χρήστες του Reddit είναι πρόθυμοι να συνθέσουν τις αναρτήσεις τους με συγκεκριμένο τρόπο. Για παράδειγμα, περιγράφουν τα γεγονότα της ζωής τους και τις ιστορίες τους σχετικά με τους φίλους τους. Ενώ το περιεχόμενο στο Twitter είναι πολύ πιο απλό με εκφράσεις όπως «θέλω να σκοτώσω» και «θα σκοτώσω». Οι λεπτομέρειες συνήθως δεν περιλαμβάνονται στα tweets τους.



(α) σύννεφο λέξεων Reddit **(β)** σύννεφο λέξεων Twitter

Σχήμα 9: Οπτικοποίηση σύννεφων λέξεων αυτοκτονικών κειμένων στο Reddit και στο Twitter

Προτιμήσεις γλώσσας. Οι προτιμήσεις γλώσσας παρέχουν μια επισκόπηση των στατιστικών γλωσσικών πληροφοριών των δεδομένων. Οι παρατιθέμενες μεταβλητές που εμφανίζονται στον Πίνακα 7 εξήχθησαν χρησιμοποιώντας το LIWC 2015 [98]. Όλα αυτές οι κατηγορίες είναι χαρακτηριστικά που βασίζονται στον αριθμό των λέξεων. Υπολογίσαμε τη μέση τιμή κάθε μεταβλητής τόσο σε κείμενα που σχετίζονται με αυτοκτονία όσο και σε δημοσιεύσεις χωρίς αυτοκτονία. Όπως φαίνεται στον πίνακα, το περιεχόμενο με ή χωρίς αυτοκτονία διαφέρει αρκετά σε πολλά στοιχεία.

- Οι χρήστες με αυτοκτονικό ιδεασμό χρησιμοποιούν πολλές προσωπικές αντωνυμίες για να δείξουν το εγώ τους. Για παράδειγμα, «Θέλω να τελειώσω τη ζωή μου».
- Εκφράζουν περισσότερα αρνητικά συναισθήματα, όπως άγχος και θλίψη. Για παράδειγμα, «πνιγόμουν από ενοχές και κατάθλιψη για αρκετά χρόνια μετά».
- Όσον αφορά τον χρόνο, τα κείμενα με ιδέες αυτοκτονίας τείνουν να χρησιμοποιούν τον παρόντα και τον μελλοντικό χρόνο. Τείνουν να χρησιμοποιούν τον ενεστώτα για να περιγράψουν τον πόνο και την κατάθλιψή τους. Για παράδειγμα, «Νιώθω τόσο άσχημα». Ο μελλοντικός χρόνος χρησιμοποιείται για να περιγράψει τα απελπιστικά συναισθήματά τους για το μέλλον και τις προθέσεις αυτοκτονίας τους. Για παράδειγμα, «τελικά θα αυτοκτονήσω».
- Και οι δύο τύποι αναρτήσεων συζητούν για την οικογένεια, τους φίλους και κάνουν αναφορές σε γυναίκες ή άνδρες.

- Δεν αποτελεί έκπληξη το γεγονός ότι περισσότερες λέξεις που σχετίζονται με τον θάνατο εμφανίζονται σε κείμενα για αυτοκτονία. Για παράδειγμα, «σκοτώνω», «πεθαίνω», «τερματίζω τη ζωή» και «αυτοκτονία».
- Και οι δύο τύποι αναρτήσεων περιέχουν παρόμοιο αριθμό βρισιών.

Ένα από τα ευρήματα του Πίνακα 7 και του Σχήματος 9 είναι ότι τα άτομα με αυτοκτονικές σκέψεις τείνουν να δείχνουν άμεσα τις προθέσεις τους σε ανώνυμες διαδικτυακές κοινότητες που αντιμετωπίζουν κάποιου είδους προβλήματα στον πραγματικό κόσμο. Οι αναρτήσεις τους συχνά δείχνουν αρνητικά συναισθήματα με έντονο εγώ και πρόθεση.

Μέση καταμέτρηση λέξεων	Αυτοκτονία	Μη αυτοκτονία
προσωπικά ουσιαστικά	30.01	14.6
ποσοτικοποιητές	3.78	3.37
θετικό συναίσθημα	5.61	7.84
αρνητικό συναίσθημα	11.12	4.89
ανησυχία	1.46	0.55
θλίψη	3.86	0.63
παρελθούσα εστίαση	6.78	6.27
παρούσα εστίαση	34.81	17.86
μελλοντική εστίαση	4.06	1.76
οικογένεια	1.07	0.82
φίλοι	1.02	0.78
γυναικείες αναφορές	0.95	1.35
ανδρικές αναφορές	1.03	2.40
δουλειά	2.50	3.92
χρήματα	0.60	1.38
θάνατος	4.81	0.61
βρισιές	1.47	1.62

Πίνακας 7: Γλωσσικές στατιστικές πληροφορίες που εξήχθησαν από το LIWC

No.	10 κορυφαίες λέξεις για κάθε θέμα που σχετίζεται με την αυτοκτονία στο SuicideWatch
1.	money, working, suicide, gun, fucked, come, yet, failed, erase, don't
2.	said, got, went, started, friend, back, father, told, mother, girl
3.	im, school, go, year, time, know, one, ive, day, got
4.	im, don't, its, ive, cant, get, know, around, time, pain
5.	im, feel, like, want, know, friend, would, life, get, time
6.	imagine, cellophane, abandoned, anyone, medical, cheated, mr, surgery, yelling, letter
7.	im, want, life, like, get, feel, ive, know, year, even
8.	fucking, very, tomorrow, bottom, accept, sharp, n't, went, wife, attacked
9.	condition, suicide, also, hope, tx, california, chronic, jumping, crisis, age
10.	please, find, mother, car, social, live, need, accident, debt, month

Πίνακας 8: Θέματα λέξεις που εξάγονται από δημοσιεύσεις που περιέχουν αυτοκτονικές σκέψεις

Περιγραφή θέματος. Εξαγάγαμε 10 θέματα από δημοσιεύσεις που περιείχαν ιδέες αυτοκτονίας χρησιμοποιώντας τη μέθοδο μοντελοποίησης θεμάτων Latent Dirichlet (LDA) [99], όπως φαίνεται στον Πίνακα 8. Υπάρχουν ορισμένες αργκό του Διαδικτύου όπως «tx» (ευχαριστώ) και συντομογραφίες όπως «im» (είμαι) και «n't» («αρνητικός»). Στον τομέα της τυπικής επεξεργασίας φυσικής γλώσσας, οι προσωπικές λέξεις όπως «εγώ», «εμένα» και «εσύ» είναι λέξεις στάσης και πρέπει να αφαιρεθούν, αλλά τις διατηρήσαμε σε αυτήν την εξερεύνηση επειδή περιέχουν σημαντικές πληροφορίες. Έτσι, υπάρχουν πολλές προσωπικές αντωνυμίες που περιλαμβάνονται σε αυτά τα θέματα λέξεις, οι οποίες είναι πανομοιότυπες με τα αποτελέσματα στον Πίνακα 7.

Με ενδιαφέρον, παρατηρήσαμε ότι οι δημοσιεύσεις που περιέχουν αυτοκτονικά θέματα θα μπορούσαν να συνοψιστούν σε τρεις κατηγορίες: εσωτερικούς παράγοντες, εξωτερικούς κοινωνικούς παράγοντες και μικτούς εσωτερικούς/εξωτερικούς παράγοντες. Συγκεκριμένα, εσωτερικοί παράγοντες, συμπεριλαμβανομένων λέξεων όπως «ξέρω» (Θέματα 3, 4, 5 και 7), «θέλω», «αισθάνομαι» και «μου αρέσει» (Θέματα 5 και 7) και «ελπίζω» (Θέμα 9) εκφράζουν τα συναισθήματα, τις προθέσεις και τις επιθυμίες των ανθρώπων. Ενώ άλλες λέξεις όπως «χρήματα» και «εργασία» (Θέμα 1), «φίλος» (Θέματα 2 και 5), «σχολείο» (Θέμα 3), «χειρουργείο» (Θέμα 6), «κρίση» (Θέμα 9) και το «ατύχημα» (Θέμα 10) υποδηλώνουν ότι οι δημοσιεύσεις συνδέονται με κοινωνικούς παράγοντες. Στο Θέμα 3, 5, 9 και 10 παριστάνονται και οι δύο παράγοντες.

3.3. Μέθοδοι και Τεχνικές Λύσεις

3.3.1. Επεξεργασία χαρακτηριστικών

Με την προ-επεξεργασία και τον καθαρισμό των δεδομένων εκ των προτέρων, εξήγαμε διάφορα χαρακτηριστικά, συμπεριλαμβανομένων στατιστικών, χαρακτηριστικών που βασίζονται σε λέξεις (π.χ. αυτοκτονικές λέξεις, αντωνυμίες κ.λπ.), TF-IDF, σημασιολογικά και συντακτικά. Επιπλέον, χρησιμοποιήσαμε κατανεμημένα χαρακτηριστικά εκπαιδεύοντας νευρωνικά δίκτυα για να ενσωματώσουμε λέξεις σε διανυσματικές αναπαραστάσεις, μαζί με τα χαρακτηριστικά θεμάτων που εξάγονται από το LDA [99] ως χαρακτηριστικά χωρίς επίβλεψη.

Στατιστικά Χαρακτηριστικά: Οι αναρτήσεις που δημιουργούνται από χρήστες ποικίλλουν σε μήκος και ορισμένες στατιστικές δυνατότητες μπορούν να εξαχθούν από κείμενα. Ορισμένες αναρτήσεις χρησιμοποιούν σύντομες και απλές προτάσεις, ενώ άλλες χρησιμοποιούν σύνθετες προτάσεις και μεγάλες παραγράφους.

Μετά την τμηματοποίηση και την λεκτική ανάλυση, καταγράψαμε στατιστικά χαρακτηριστικά ως εξής:

- τον αριθμό των λέξεων, των λεκτικών μονάδων και των χαρακτήρων στον τίτλο
- τον αριθμό λέξεων, λεκτικών μονάδων, χαρακτήρων, προτάσεων και παραγράφων στο σώμα του κειμένου

Συντακτικά Χαρακτηριστικά: POS. Τα συντακτικά χαρακτηριστικά είναι χρήσιμες πληροφορίες σε εργασίες επεξεργασίας φυσικής γλώσσας. Εξάγουμε μέρη του λόγου (POS) [100], ως χαρακτηριστικά, για το μοντέλο μας, ανίχνευσης αυτοκτονικού ιδεασμού προκειμένου αυτό να συλλάβει τις παρόμοιες γραμματικές ιδιότητες στις αναρτήσεις των χρηστών.

Οι κοινές ετικέτες POS περιλαμβάνουν ουσιαστικά, ρήματα, μετοχές, άρθρα, αντωνυμίες, επιρρήματα και συνδέσμους. Προσδιορίστηκαν επίσης οι υποομάδες POS για να παρέχουν περισσότερες λεπτομέρειες σχετικά με τις γραμματικές ιδιότητες των αναρτήσεων. Κάθε ανάρτηση αναλύθηκε και τοποθετήθηκε με ετικέτα και ο αριθμός κάθε κατηγορίας στον τίτλο και στο κείμενο απλά μετρήθηκαν.

Γλωσσικά χαρακτηριστικά: LIWC. Οι αναρτήσεις των χρηστών στο διαδίκτυο περιέχουν συνήθως συναισθήματα, λέξεις σχετικότητας και παρενόχλησης. Τα λεξικά εφαρμόζονται ευρέως για την εξαγωγή αυτών των χαρακτηριστικών. Για να αναλύσουμε τα γλωσσικά και συναισθηματικά χαρακτηριστικά στα δεδομένα, χρησιμοποιήσαμε τη Γλωσσική έρευνα και τον αριθμό των λέξεων [98] (LIWC 2015[20]) που προτάθηκε και αναπτύχθηκε από το Πανεπιστήμιο του Τέξας στο Austin. Αυτή η προσέγγιση χρησιμοποιήθηκε σε προηγούμενη μελέτη [38]. Το εργαλείο περιέχει ένα ισχυρό εσωτερικά κατασκευασμένο λεξικό για αντιστοίχιση των λέξεων-στόχων σε αναρτήσεις κατά την ανάλυση δεδομένων. Εξήχθησαν περίπου 90 μεταβλητές . Εκτός από τα χαρακτηριστικά που βασίζονται στον αριθμό των λέξεων, θα μπορούσαμε να εξαγάγουμε χαρακτηριστικά που βασίζονται σε συναισθηματικό τόνο, γνωστικές διαδικασίες, αντιληπτικές διαδικασίες και πολλούς τύπους υβριστικών λέξεων. Οι συγκεκριμένες κατηγορίες περιλαμβάνουν τον αριθμό λέξεων, τη συνοπτική γλώσσα, τους γενικούς περιγραφείς, τις γλωσσικές διαστάσεις, τις ψυχολογικές κατασκευές, την προσωπική ανησυχία, τους άτυπους δείκτες γλώσσας και τα σημεία στίξης.

Χαρακτηριστικά Συχνότητας Λέξεων: TF-IDF. Πολλά είδη έκφρασης σχετίζονται με την αυτοκτονία. Χρησιμοποιήσαμε το TF-IDF για να εξαγάγουμε αυτά τα χαρακτηριστικά και να μετρήσουμε τη σημασία διαφόρων λέξεων τόσο από αυτοκτονικές αναρτήσεις όσο και από μη αυτοκτονικές αναρτήσεις. Το TF-IDF μετρά τον αριθμό των φορών που κάθε λέξη εμφανίζεται στα έγγραφα και προσθέτει ποινή ανάλογα με τη συχνότητα της λέξης σε ολόκληρο το σώμα. Η διαδικασία για τον υπολογισμό του TF-IDF είναι η εξής: Δεδομένου όρου και εγγράφου που σημειώνονται ως *t* και *d* αντίστοιχα, Αρχικά, υπολογίζουμε τον όρο συχνότητα με:

$$TF(t,d) = \frac{n_{t,d}}{\sum_{t' \in d} n_{t',d}}$$

όπου $n_{t,d}$ δηλώνει τον αριθμό ενός όρου σε ένα έγγραφο. Δεύτερον, υπολογίζουμε το αντίστροφο έγγραφο

[20] http://liwc.wpengine.com/

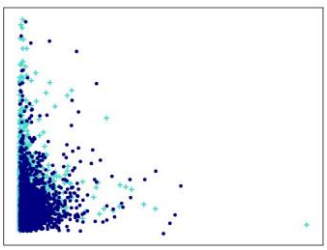
$$IDF(t,D) = \log \frac{N}{f_t}$$

όπου D είναι το σώμα, N είναι ο συνολικός αριθμός εγγράφων και f_t είναι ο αριθμός των εγγράφων που Τρίτον, παίρνουμε το TF-IDF ως:

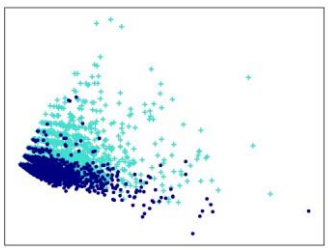
$$TF\text{-}IDF(t,d,D) = TF(t,d) \times IDF(t,D)$$

Λειτουργίες ενσωμάτωσης λέξεων: Η κατανεμημένη αναπαράσταση, η οποία είναι σε θέση να διατηρήσει τις σημασιολογικές πληροφορίες στα κείμενα, είναι δημοφιλής και χρήσιμη για πολλές εργασίες επεξεργασίας φυσικής γλώσσας. Ενσωματώνει λέξεις σε διανυσματικό χώρο. Υπάρχουν διάφορες τεχνικές για την ενσωμάτωση λέξεων. Χρησιμοποιήσαμε τη *word2vec* [101][21] για την εξαγωγή μιας κατανεμημένης σημασιολογικής αναπαράστασης των λέξεων.

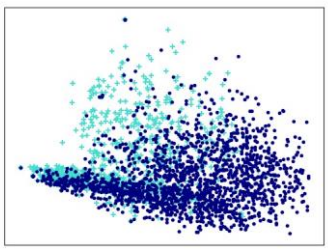
Υπάρχουν δύο αρχιτεκτονικές για τη *word2vec* ενσωμάτωση λέξεων, δηλαδή CBOW και Skip-gram. Το CBOW προβλέπει την παρούσα λέξη με βάση το πλαίσιο, το Skip-gram προβλέπει τις πλησιέστερες λέξεις στην τρέχουσα παρεχόμενη λέξη.



(α) Στατιστική

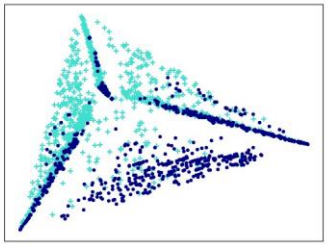


(β) POS

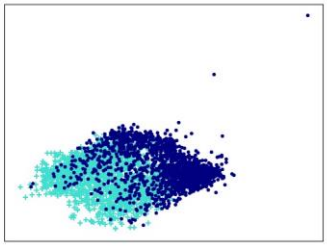


(γ) TF-IDF

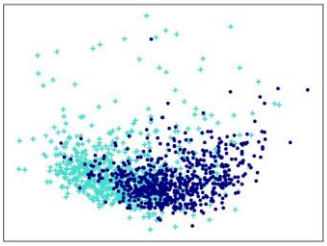
[21] https://code.google.com/archive/p/word2vec/



(δ) Θέματα



(ε) LIWC



(στ) Ενσωμάτωση λέξεων

Σχήμα 10: Οπτικοποίηση των εξαγόμενων χαρακτηριστικών χρησιμοποιώντας PCA

Χαρακτηριστικά θεμάτων: Οι αναρτήσεις αυτοκτονίας και οι μη αυτοκτονικές αναρτήσεις μιλούν για διαφορετικά θέματα που μπορούν να παρέχουν καλή κατανόηση για δύο κατηγορίες. Εφαρμόσαμε το Latent Dirichlet Allocation (LDA) [99] για να αποκαλύψουμε λανθάνοντα θέματα σε αναρτήσεις χρηστών. Κάθε θέμα είναι μια πιθανότητα ανάμειξης λέξεων στο θέμα και κάθε δημοσίευση είναι μια πιθανότητα μείγματος θεμάτων.

Δεδομένου του συνόλου των εγγράφων και του αριθμού των θεμάτων, χρησιμοποιήσαμε το LDA για να εξαγάγουμε τα θέματα από κάθε δημοσίευση και στη συνέχεια υπολογίζουμε την πιθανότητα ότι κάθε ανάρτηση ανήκε σε όλα τα γενικά θέματα. Ως εκ τούτου, οι αναρτήσεις αναπαρίστανται από τις θεματικές τους ιδιότητες ως διανύσματα πιθανότητας στο μήκος του αριθμού των θεμάτων.

Οπτικοποίηση χαρακτηριστικών: Για να κατανοήσουμε την πληροφόρηση αυτών των συνόλων χαρακτηριστικών, απεικονίζουμε τις δυνατότητες του συνόλου δεδομένων Reddit σε χώρο 2-διαστάσεων χρησιμοποιώντας την Ανάλυση βασικών στοιχείων (PCA) [102] στο Σχήμα 10. Τα αποτελέσματα καταδεικνύουν ότι πράγματι εξάγουμε χαρακτηριστικά που χωρίζουν σε μεγάλο βαθμό τα σημεία σε διαφορετικές κλάσεις..

3.3.2. Μοντέλα ταξινόμησης

Η ανίχνευση αυτοκτονίας στο κοινωνικό περιεχόμενο είναι ένα τυπικό πρόβλημα ταξινόμησης της εποπτευόμενης μάθησης όπως στον ορισμό 2.

Ορισμός 2 (Ανίχνευση αυτοκτονικού ιδεασμού στο Κείμενο) . Δεδομένου ενός εκπαιδευτικού συνόλου m εγγράφων με συγκεκριμένη τάξη $C = \{c_1, c_2, \dots, c_n\}$, η ανίχνευση αυτοκτονικού ιδεασμού σε κείμενο εκπαιδεύει έναν ταξινομητή χρησιμοποιώντας το σύνολο εκπαίδευσης και μαθαίνει έναν εποπτευόμενο ταξινομητή για πρόβλεψη στο σύνολο δοκιμών.

Δεδομένου ενός συνόλου δεδομένων $\{x_i, y_i\}_i^n$ *που* περιλαμβάνει ένα σύνολο κειμένων $\{x_i\}_i^n$ με ετικέτες$\{y_i\}_i^n$, εκπαιδεύσαμε ένα μοντέλο ταξινόμησης με επίβλεψη για να μάθουμε τη λειτουργία από τα ζεύγη δεδομένων εκπαίδευσης αντικειμένων εισόδου και εποπτικών σημάτων :

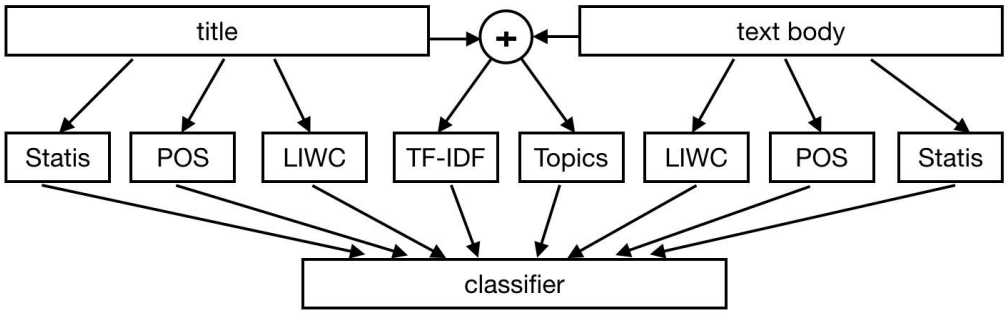
$$y_i = F(x_i)$$

όπου $y_i = 1$ σημαίνει ότι η έκφραση x_i είναι "κείμενο αυτοκτονίας (ST)", διαφορετικά $y_i = 0$ σημαίνει "όχι κείμενο αυτοκτονίας (non-ST)". Η εκπαίδευση ή η εκμάθηση του μοντέλου ταξινόμησης είναι να ελαχιστοποιήσει το σφάλμα πρόβλεψης στα δοσμένα δεδομένα εκπαίδευσης. Το σφάλμα πρόβλεψης πρέπει να παρουσιάζεται ως συνάρτηση απώλειας $L(y, F(x))$ όπου y είναι η πραγματική ετικέτα και $F(x)$ είναι η προβλεπόμενη ετικέτα χρησιμοποιώντας μοντέλο ταξινόμησης. Συνοπτικά, ο στόχος του αλγορίθμου εκπαίδευσης είναι να αποκτήσει ένα βέλτιστο μοντέλο πρόβλεψης $F(x)$ με την επίλυση της παρακάτω εργασίας βελτιστοποίησης:

$$\hat{F} = \arg\min_F \mathbb{E}_{x,y}[L(y, F(x))]$$

Διαφορετική μέθοδος ταξινόμησης μπορεί να έχει διαφορετικό ορισμό της συνάρτησης απώλειας και προκαθορισμένης δομής του μοντέλου. Χρησιμοποιήσαμε τόσο κλασικές μεθόδους ταξινόμησης με επίβλεψη μάθησης όσο και μεθόδους βαθιάς μάθησης για να λύσουμε την εργασία ταξινόμησης των αυτοκτονικών ιδεών .

Η δομή της μεθόδου εξαγωγής χαρακτηριστικών μας φαίνεται στο Σχήμα 11. Όπως αναφέρθηκε στην Ενότητα 3.3.1, τα χαρακτηριστικά περιλαμβάνουν στατιστικά στοιχεία, μετρήσεις POS, χαρακτηριστικά LIWC, διανύσματα TF-IDF και χαρακτηριστικά πιθανότητας θεμάτων. Μεταξύ αυτών των χαρακτηριστικών, εφαρμόσαμε χαρακτηριστικά POS και χαρακτηριστικά LIWC τόσο στον τίτλο όσο και στο κείμενο των αναρτήσεων χρηστών. Συνδυάσαμε τον τίτλο και το σώμα σε ένα κομμάτι κειμένου για να εξαγάγουμε διανύσματα πιθανότητας θεμάτων και διανύσματα TF-IDF. Όλα τα εξαγόμενα χαρακτηριστικά εισήχθησαν στους ταξινομητές.



Σχήμα 11: Η δομή του μοντέλου για το σύνολο δεδομένων Reddit

3.4. Εμπειρική Αξιολόγηση

3.4.1. Σύγκριση και ανάλυση αυτοκτονίας έναντι μη αυτοκτονίας

Μέθοδος	Χαρακτηριστικά	Acc.	Prec.	Recall	F1-score	AUC
SVM	Statis	0.8064	0.8045	0.8189	0.8116	0.8061
	Statis+Topic	0.8609	0.881	0.8406	0.8603	0.8613
	Statis+Topic+TF-IDF	0.8571	0.8414	0.8865	0.8634	0.8565
	Statis+Topic+TF-IDF+POS	0.8674	0.8545	0.8916	0.8727	0.8670
	Statis+Topic+TF-IDF+POS+LIWC	0.9123	0.9144	0.9133	0.9138	0.9123
Random Forest	Statis	0.7732	0.8094	0.7258	0.7653	0.7741
	Statis+Topic	0.8973	0.8922	0.9082	0.9001	0.8971
	Statis+Topic+TF-IDF	0.8915	0.8795	0.912	0.8954	0.8911
	Statis+Topic+TF-IDF+POS	0.8986	0.8801	0.9273	0.9031	0.8981
	Statis+Topic+TF-IDF+POS+LIWC	0.9357	0.9213	0.9554	0.938	0.9353
GBDT	Statis	0.7505	0.7632	0.7398	0.7513	0.7507
	Statis+Topic	0.898	0.8856	0.9184	0.9017	0.8976
	Statis+Topic+TF-IDF	0.896	0.89	0.9082	0.899	0.8958
	Statis+Topic+TF-IDF+POS	0.8928	0.8893	0.9018	0.8955	0.8926
	Statis+Topic+TF-IDF+POS+LIWC	0.9461	0.9354	0.9605	0.9478	0.9458
XGBoost	Statis	0.7667	0.7822	0.7513	0.7664	0.7670
	Statis+Topic	0.8999	0.8938	0.912	0.9028	0.8997
	Statis+Topic+TF-IDF	0.9019	0.8941	0.9158	0.9049	0.9016
	Statis+Topic+TF-IDF+POS	0.9103	0.8998	0.9273	0.9133	0.9100
	Statis+Topic+TF-IDF+POS+LIWC	**0.9571**	0.9499	**0.9668**	**0.9583**	**0.9569**

MLFFNN	Statis	0.7647	0.7742	0.7742	0.7742	0.7731
	Statis+Topic	0.8821	0.8740	0.8525	0.8631	0.8961
	Statis+Topic+TF-IDF	0.8606	0.8369	0.8401	0.8385	0.8855
	Statis+Topic+TF-IDF+POS	0.9068	0.9038	0.8868	0.8952	0.9369
	Statis+Topic+TF-IDF+POS+LIWC	0.9283	0.9391	0.9205	0.9295	0.9403
LSTM	*word2vec* word embedding	0.9266	**0.9786**	0.8750	0.9239	0.9276

Πίνακας 9: Σύγκριση διαφορετικών μεθόδων χρησιμοποιώντας διαφορετικά χαρακτηριστικά

Αυτή η ενότητα συγκρίνει διάφορες μεθόδους ταξινόμησης χρησιμοποιώντας διαφορετικούς συνδυασμούς χαρακτηριστικών με διασταυρούμενη επικύρωση 10 φορές. Τα συγκεκριμένα μοντέλα ταξινόμησης περιλαμβάνουν Support Vector Machine [93], Random Forest [94], Gradient Boost Classification Tree (GBDT) [95], XGBoost [96] και Multilayer Feed Forward Neural Net (MLFFNN) [43]. Το SVM είναι σε θέση να επιλύσει προβλήματα που δεν είναι γραμμικά διαχωρίσιμα στον χαμηλότερο χώρο με την κατασκευή ενός υπερπλάνου σε χώρο με μεγάλες διαστάσεις. Μπορεί να προσαρμοστεί σε πολλά είδη ταξινόμησης. Τα Random Forest, GBDT και XGBoost είναι μέθοδοι δέντρων που χρησιμοποιούν δέντρα αποφάσεων ως βασικούς ταξινομητές και παράγουν μια μορφή επιτροπής για να αποκτήσουν καλύτερη απόδοση από οποιονδήποτε ταξινομητή βάσης. Το MLFFNN λαμβάνει τις διαφορετικές δυνατότητες ως είσοδο και μαθαίνει τον συνδυασμό τους με μη γραμμικότητα.

Για σύγκριση και επίλυση του προβλήματος της μη κατανόησης της σημασιολογικής σημασίας και της συντακτικής δομής των προτάσεων, η βαθιά μάθηση παρέχει ισχυρές επιδόσεις. Χρησιμοποιήσαμε δίκτυο Long Short Term Memory (LSTM) [97], ένα υπερσύγχρονο βαθύ νευρωνικό δίκτυο. Το LSTM λαμβάνει ως είσοδο τον τίτλο και το κείμενο των αναρτήσεων χρηστών με ενσωμάτωση λέξης και χρησιμοποιεί το κελί μνήμης για να διατηρήσει την κατάσταση για μεγάλες περιόδους, καταγράφοντας τις μακροπρόθεσμες εξαρτήσεις στην ανίχνευση μακρών συνομιλιών.

Όπως φαίνεται στον Πίνακα 9, η απόδοση όλων των μεθόδων αυξάνεται συνδυάζοντας περισσότερα χαρακτηριστικά. Αυτή η παρατήρηση επικυρώνει την αποτελεσματικότητα και την πληροφόρηση των εξαγόμενων χαρακτηριστικών μας. Ωστόσο, η συμβολή κάθε χαρακτηριστικού ποικίλλει, γεγονός που οδηγεί σε διακυμάνσεις στα αποτελέσματα μεμονωμένων μεθόδων. Το XGBoost είχε την καλύτερη απόδοση από αυτές τις έξι μεθόδους

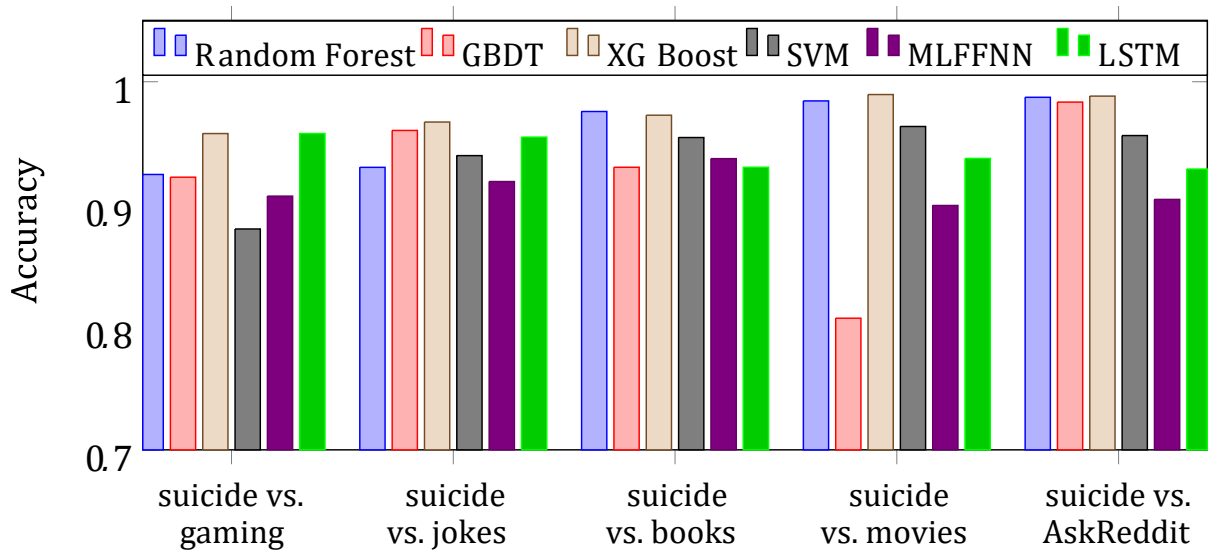
όταν έλαβε όλες τις ομάδες χαρακτηριστικών ως εισόδους. Παρόλο που το LSTM δεν απαιτεί επεξεργασία χαρακτηριστικών και φημίζεται για την υπερσύγχρονη απόδοσή του σε πολλές άλλες εργασίες επεξεργασίας φυσικής γλώσσας, δεν απέδωσε τόσο καλά όσο ορισμένες από τις άλλες μεθόδους μάθησης συνόλου με επαρκή χαρακτηριστικά σε αυτήν την περίπτωση. Το Random Forest, το GBDT, το XGBoost και το MLFFNN με τα κατάλληλα χαρακτηριστικά παρήγαγαν καλύτερη ακρίβεια και βαθμολογία F1 από το LSTM στο σύνολο δεδομένων Reddit. Ομολογουμένως, η βαθιά μάθηση με ενσωμάτωση λέξεων είναι μάλλον βολική και τυπικά επιτυγχάνει επαρκή αποτελέσματα, ακόμη και χωρίς περίπλοκα μηχανικά χαρακτηριστικά.

Η μέτρηση απόδοσης AUC σε κάθε ταξινόμηση είναι η περιοχή κάτω από την καμπύλη χαρακτηριστικών λειτουργίας του δέκτη με όλα τα εξαγόμενα χαρακτηριστικά. Στην τελευταία στήλη του Πίνακα 9, η AUC(Area Under the Curve) έχει μια αυξανόμενη τάση με περισσότερα συνδυασμένα χαρακτηριστικά. Η μέθοδος XGBoost αποκτά την υψηλότερη AUC από 0,9569 ενώ άλλες μέθοδοι έχουν πολύ παρόμοια τιμή AUC πάνω από 0,9.

3.4.2. Αυτοκτονία εναντίον Ενιαίων θεμάτων Subreddits

Για να αξιολογήσουμε την ταξινόμηση της αυτοκτονίας με άλλες συγκεκριμένες διαδικτυακές κοινότητες, επεκτείναμε τα σύνολα δεδομένων και τα πειράματά μας σε άλλα συγκεκριμένα subreddits, όπως “gaming”, “Jokes”, “books”, “movies” and “AskReddit”.

Τα αποτελέσματα φαίνονται στο Σχήμα 12 . Η χρήση των δυνατοτήτων που εξήχθησαν με την προσέγγισή μας ήταν πολύ αποτελεσματικός τρόπος για την ταξινόμηση των αναρτήσεων αυτοκτονικού ιδεασμού από άλλους τομείς subreddits. Στην πραγματικότητα, τα αποτελέσματα ταξινόμησης για το σύνολο αυτοκτονικών δεδομένων έναντι του συνόλου δεδομένων subreddit ήταν καλύτερα από το σύνολο αυτοκτονικών έναντι μη αυτοκτονικών δεδομένων, όπου τα μη αυτοκτονικά δείγματα αποτελούνται από πολλούς δημοφιλείς τομείς subreddit. Σε αυτά τα πειράματα, το XGBoost παρήγαγε τα καλύτερα αποτελέσματα σε “ movies ” και “AskReddit” όσον αφορά την ακρίβεια και τις βαθμολογίες F1. Το LSTM και το Random Forest υπερ-απόδωσαν τα άλλα μοντέλα σε “ gaming ” και “ books ” αντίστοιχα.



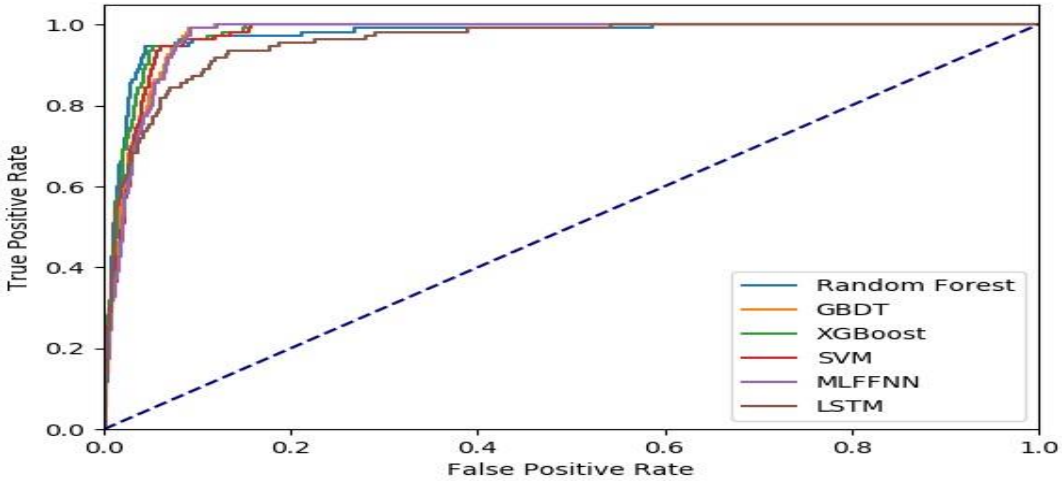
Σχήμα 12: Ταξινόμηση για αυτοκτονικό ιδεασμό του SuicideWatch έναντι άλλων έξι υποδιαιρέσεων

3.4.3. Πειράματα στο Σύνολο Δεδομένων του Twitter

Για να αξιολογήσουμε την απόδοση των λειτουργιών μας και των μοντέλων ταξινόμησης, κάνουμε ένα άλλο πείραμα στο σύνολο δεδομένων του Twitter. Το κείμενο των tweets χωρίς σώμα μεγάλου κειμένου διαφέρει με το κείμενο Reddit. Έτσι, για το πειραματικό περιβάλλον, υπάρχει μια μικρή διαφορά μεταξύ τους. Εξαιρούμε τον αριθμό των παραγράφων στα στατιστικά χαρακτηριστικά, τις λειτουργίες POS και LIWC των κειμένων. Οι υπόλοιπες ρυθμίσεις είναι παρόμοιες με το προηγούμενο πείραμά μας. Λαμβάνοντας υπόψη την ανισορροπία της τάξης στα δεδομένα του Twitter, υιοθετούμε τεχνικές υπο δειγματοληψίας. Τα αποτελέσματα είναι οι μέσες μετρήσεις κάθε υπο δειγματοληψίας δεδομένων που παρουσιάζονται στον Πίνακα 10. Οι χαρακτηριστικές καμπύλες λειτουργίας του δέκτη αυτών των μεθόδων φαίνονται στο Σχήμα 13. Σε αυτά τα σύνολα δεδομένων, το τυχαίο δάσος αποκτά καλύτερη απόδοση από τα περισσότερα μοντέλα, εκτός από τη μέτρηση ακρίβειας κατά την οποία το MLFFNN κερδίζει ελαφρώς καλύτερο αποτέλεσμα.

Model	Acc.	Prec.	Recall	F1	AUC
Random Forest	**0.9638**	0.9638	**0.9917**	**0.9646**	**0.9862**
GBDT	0.9500	0.9413	0.9603	0.9503	0.9825
XGBoost	0.9591	0.9425	0.9782	0.9597	0.9843
SVM	0.9485	0.9261	0.9755	0.9497	0.9813
MLFFNN	0.9412	**0.9661**	0.9194	0.9421	0.9823
LSTM	0.9108	0.9399	0.8802	0.9059	0.9747

Πίνακας 10: Σύγκριση διαφορετικών μοντέλων που χρησιμοποιούν όλα τα επεξεργασμένα χαρακτηριστικά στα δεδομένα του Twitter



Σχήμα 13: Η καμπύλη λειτουργίας του δέκτη με έξι μεθόδους με όλα τα επεξεργασμένα χαρακτηριστικά

Κεφάλαιο 4

4. Δίκτυο Προσεκτικών Σχέσεων

4.1. Εισαγωγή

Η ψυχική υγεία είναι ένα παγκόσμιο ζήτημα, ιδιαίτερα σοβαρό στις περισσότερες ανεπτυγμένες χώρες και πολλές αναδυόμενες αγορές. Σύμφωνα με έκθεση του ΠΟΥ[22], 1 στους 4 ανθρώπους παγκοσμίως υποφέρουν από ψυχική διαταραχή σε κάποιο βαθμό. Και 3 στα 4 άτομα με σοβαρές ψυχικές διαταραχές δεν λαμβάνουν θεραπεία, γεγονός που επιδεινώνει το πρόβλημα. Το σχήμα 14 δείχνει τον επιπολασμό διαταραχών ψυχικής υγείας και χρήσης ουσιών το 2016[23]. Εν μέρει λόγω σοβαρών ψυχικών διαταραχών, 900.000 άτομα αυτοκτονούν κάθε χρόνο σε όλο τον κόσμο, καθιστώντας την αυτοκτονία τη συχνότερη αιτία θανάτου

[22] Σχέδιο δράσης για την ψυχική υγεία 2013 - 2020, διαθέσιμο στη διεύθυνση http://www.who.int/mental_health/action_plan_2013/ mhap_brochure.pdf? Ua = 1

[23] Έκδοση από το Σιάτλ, Ηνωμένες Πολιτείες: Institute for Health Metrics and Evaluation (IHME), 2017. διατίθεται στη διεύθυνση http://hdx.healthdata.org/gbd-results-tool. Λήψη από https://ourworldindata.org/mental-health

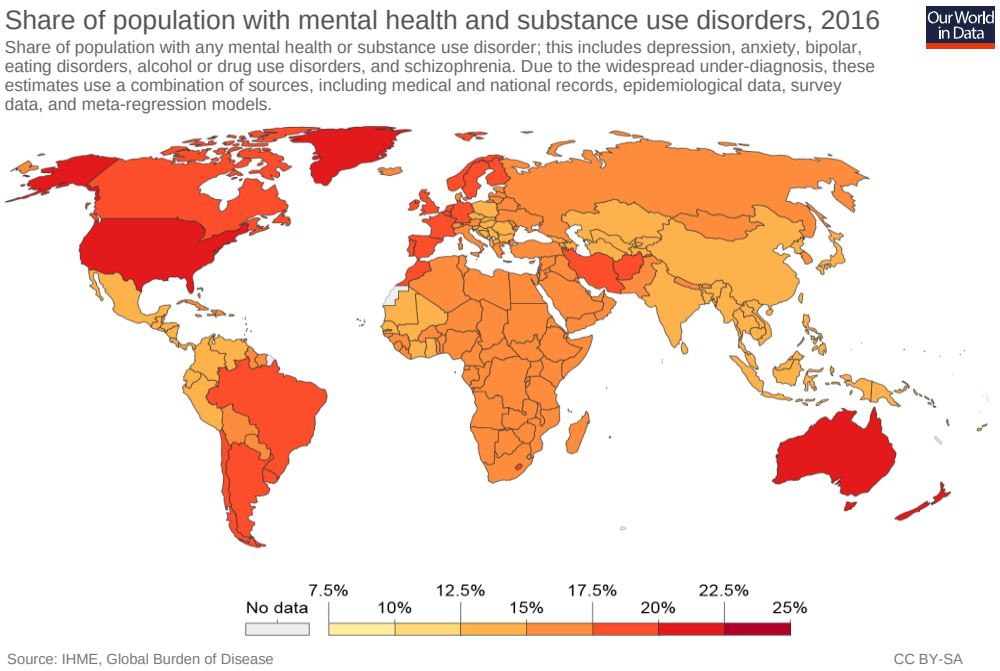
μεταξύ των νέων. Άτομα τα οποία διαπράττουν απόπειρες αυτοκτονίας αναφέρονται επίσης ότι πάσχουν από ψυχικές διαταραχές. Η Εθνική Συμμαχία των ΗΠΑ για τις ψυχικές Ασθένειες ανέφερε ότι το 46% των θυμάτων αυτοκτονίας έχουν βιώσει παθήσεις ψυχικής υγείας[24].

Με την πρόοδο των υπηρεσιών κοινωνικών δικτύων, οι άνθρωποι αρχίζουν να εκφράζουν τα συναισθήματά τους στα φόρουμ και αναζητούν ηλεκτρονική υποστήριξη. Οι τακτικοί τρόποι πρόληψης περιλαμβάνουν προφορική διαβούλευση και ψυχολογική παρέμβαση. Ωστόσο, λόγω της σπανιότητας και της ανισότητας των δημόσιων πόρων στις υπηρεσίες υγείας [103], πολλά θύματα δεν μπόρεσαν να λάβουν αποτελεσματικές θεραπείες ακόμη και εάν μερικά από αυτά πάσχουν από σοβαρές ψυχικές διαταραχές. Η μετάβαση από τον αυτοκτονικό ιδεασμό στη δράση είναι μια μακροπρόθεσμη διαδικασία. Οι Gilat et al. [104] κλιμάκωσαν τους κινδύνους αυτοκτονίας σε τέσσερα επίπεδα, δηλαδή, μη αυτοκτονικές, αυτοκτονικές σκέψεις ή επιθυμίες, προθέσεις αυτοκτονίας και αυτοκτονική πράξη ή σχέδιο. Πριν από τον αυτοκτονικό ιδεασμό, τα θύματα μπορεί να υποφέρουν από διαφορετικά είδη άλλων ψυχικών διαταραχών. Σύμφωνα με τις μετα-αναλύσεις, υποκείμενες ψυχικές διαταραχές μπορούν να οδηγήσουν σε αυτοκτονία, ειδικά σε χώρες υψηλού εισοδήματος με ποσοστό 90%[25]. Η υπηρεσία κοινωνικής δικτύωσης έχει γίνει ένα από τα πιο χρήσιμα εργαλεία για την παροχή υποστήριξης και ανατροφοδότησης για άτομα με προβλήματα ψυχικής υγείας [85]. Για την παροχή αποτελεσματικής έγκαιρης πρόληψης αυτοκτονίας λόγω περιορισμένων πόρων υποστήριξης, είναι απαραίτητο να υπολογίσουμε αυτόματα τα επίπεδα κινδύνου και να παρέχουμε υποστήριξη συνομιλίας ανάλογα των θεμάτων των θυμάτων για την ανακούφισή τους. Το κίνητρό μας είναι να χρησιμοποιήσουμε τεχνικές βαθιάς μάθησης για να μπορέσουμε να εντοπίσουμε έγκαιρα και να προσδιορίσουμε τα επίπεδα κινδύνου των ανθρώπων, τα οποία μπορούν να βοηθήσουν τους κοινωνικούς λειτουργούς ή τους ειδικούς να κατανοήσουν εκ των προτέρων την κατάσταση των ανθρώπων όταν προσπαθούν να ανακουφίσουν τα προβλήματα ψυχικής υγείας τους. Η τεχνική αυτόματης ανίχνευσης μπορεί να εφαρμοστεί στην

[24] NAMI έκθεση σε κίνδυνο της αυτοκτονίας, διαθέσιμη σε https://www.nami.org/Learn-More/Mental-Health-Conditions/Related-Conditions/Suicide

[25] Έκθεση των Χάνα Ρίτσι και Μαξ Ρόζερ δημοσιευμένη στο διαδίκτυο στο OurWorldInData.org. Ανακτήθηκε από https: // ourworldindata.org/mental-health

παρακολούθηση της ψυχικής υγείας και να βοηθήσει στη διευκόλυνση της διαδικτυακής υποστήριξης.



Σχήμα 14: Παγκόσμιο Δίκτυο Συνεργασίας για ψυχικές ασθένειες. Αποτελέσματα Global Burden of Disease Study 2016 (GBD 2016)

Ο εντοπισμός των επιπέδων κινδύνου αυτοκτονικού ιδεασμού και των τύπων ψυχικών διαταραχών από τους οποίους οι διαδικτυακοί χρήστες ή οι σχέσεις τους μπορεί να υποφέρουν αποτελεί ένα τυπικό πρόβλημα ταξινόμησης κειμένου. Υπάρχουν πολλοί τύποι ψυχικών διαταραχών σύμφωνα με δύο κύρια διαγνωστικά σχήματα για τον εντοπισμό ψυχικών διαταραχών, δηλαδή, Διαγνωστικό και Στατιστικό Εγχειρίδιο Ψυχικών Διαταραχών (DSM-5)[26] και Κεφάλαιο V Ψυχικές και Συμπεριφορικές διαταραχές της διεθνούς στατιστικής ταξινόμησης ασθενειών και σχετικών προβλημάτων υγείας 10η αναθεώρηση (ICD-10)[27]. Ο αυτοκτονικός ιδεασμός και οι ψυχικές διαταραχές όπως η κατάθλιψη, το άγχος και η διπολική διαταραχή στο διαδικτυακό κοινωνικό περιεχόμενο μοιράζονται αρκετά παρόμοια θέματα, όπως η χρήση της γλώσσας, η διανομή θεμάτων και η συναισθηματική πολικότητα. Τα περισσότερα από αυτά περιέχουν πολλές αρνητικές εκφράσεις. Το δημοφιλές θέμα μεταξύ

[26] https://www.psychiatry.org/psychiatrists/practice/dsm

[27] http://apps.who.int/classifications/icd10/browse/2016/en#/V

αυτών των δημοσιεύσεων είναι αρκετά κοινό, συμπεριλαμβανομένου του εργασιακού άγχους, των οικογενειακών ζητημάτων και της προσωπικής κρίσης. Έτσι, η ταξινόμηση των αυτοκτονικών ιδεών και άλλων θεμάτων ψυχικής υγείας απαιτεί προσοχή για να κατανοήσουμε καλά τις λεπτές διαφορές μεταξύ αυτών των χαρακτηριστικών.

Τα θέματα αυτοκτονίας και ψυχικής υγείας θα μπορούσαν να κατηγοριοποιηθούν ως διαφορετικά επίπεδα τα οποία θα μπορούσαν να θεωρηθούν ως πρόβλημα ταξινόμησης πολλαπλών κατηγοριών. Υπάρχει μια έκρηξη πρόσφατων έργων σχετικά με την ταξινόμηση κειμένου χρησιμοποιώντας βαθιά νευρωνικά δίκτυα. Αλλά η ταξινόμηση της ψυχικής υγείας και των ιδεών αυτοκτονίας είναι μια πιο συγκεκριμένη εργασία που απαιτεί να επικεντρωθεί στη χρήση της γλώσσας των πιθανών θυμάτων. Παρατηρώντας ότι οι δημοσιεύσεις των ατόμων εμφανίζουν συναισθήματα ή η έκφραση των πόνων περιέχει το συναίσθημά τους ως ένα βαθμό, προτείνουμε να συλλάβουμε αυτές τις χρήσιμες και σημαντικές πληροφορίες για να μάθουμε πλουσιότερη αναπαράσταση των προτάσεων και του λόγου για τις πράξεις των ανθρώπων και την ψυχική ή κοινωνική κατάσταση των ανθρώπων. Σε αυτήν την ενότητα, αναπτύσσουμε αρκετά δημοφιλή μοντέλα βαθιάς μάθησης για ταξινόμηση κειμένου σε ορισμένα υπάρχοντα σύνολα δεδομένων και αυτο-συλλεγόμενα σύνολα δεδομένων σε πραγματικούς ιστότοπους κοινωνικής δικτύωσης και προτείνουμε ένα νέο μοντέλο αιτιολογίας σχέσεων ενισχυμένης σχέσης για την παροχή ακριβέστερης ταξινόμησης των επιπέδων κινδύνου αυτοκτονίας και του αυτοκτονικού ιδεασμού έναντι άλλων ψυχικών διαταραχών όπως κατάθλιψη και άγχος.

Οι συνεισφορές μας θα μπορούσαν να συνοψιστούν ως εξής:

- Αυτή η ενότητα επικεντρώνεται στον εντοπισμό της ιδέας της αυτοκτονίας και των διαφόρων ειδών ψυχικών διαταραχών για έγκαιρη προειδοποίηση. Συγκεκριμένα, λαμβάνουμε υπόψη την ανίχνευση σε επίπεδο χρήστη και μετά το επίπεδο.

- Για τη βελτίωση της απόδοσης του εντοπισμού κινδύνου, προτείνουμε το μοντέλο σχέσης δικτύου με προσοχή για συλλογισμό πάνω από την αναπαράσταση κειμένου και δύο σύνολα δεικτών κινδύνου, δηλαδή, συναισθηματική κατάσταση με βάση το λεξικό και λανθάνοντα θέματα εντός των αναρτήσεων.

- Πειράματα σε δημόσια σύνολα δεδομένων και τα δικά μας συλλεγμένα σύνολα δεδομένων δείχνουν ότι η προτεινόμενη μέθοδος μας μπορεί να βελτιώσει την προγνωστική απόδοση.

Αυτή η ενότητα οργανώνεται ως εξής. Σχετικές εργασίες για ψυχικές διαταραχές, αυτοκτονικός ιδεασμός, ταξινόμηση κειμένου και σχετικούς συλλογισμούς αναλύονται στην Ενότητα 4.2. Στην Ενότητα 4.3, εισάγουμε την προτεινόμενη μέθοδο που εισάγει το συναισθηματικό λεξικό και το μοντέλο θεμάτων στη συλλογιστική με δίκτυο σχέσεων με βάση την προσοχή. Τα σύνολα δεδομένων εισάγονται στην Ενότητα 4.4, μαζί με μια απλή και διερευνητική ανάλυση. Οι πειραματικές ρυθμίσεις και τα αποτελέσματα παρουσιάζονται στην Ενότητα 4.5. Στην Ενότητα 4.6, κάνουμε ένα συμπέρασμα και έχουμε μια σύντομη προοπτική για μελλοντικές εργασίες.

4.2. Σχετική Εργασία

4.2.1. Ταξινόμηση Κειμένου

Η ταξινόμηση κειμένου γνώρισε ραγδαία ανάπτυξη με την ανάπτυξη βαθιών νευρωνικών δικτύων. Οι κατανεμημένες τεχνικές αναπαράστασης λέξεων όπως το word2vec [54] και το GloVe [55] παρέχουν ισχυρά εργαλεία για την αναπαράσταση κειμένου. Ο Kim [105] πρότεινε συνελικτικά νευρωνικά δίκτυα για ταξινόμηση προτάσεων . Για να συλλάβουμε τη μακροπρόθεσμη εξάρτηση στις προτάσεις, εφαρμόστηκε η μακροπρόθεσμη μνήμη (LSTM) [106]. Οι Lai et. al [107] πρότειναν επαναλαμβανόμενα συνελικτικά νευρωνικά δίκτυα που συνδυάζουν δύο δημοφιλείς αρχιτεκτονικές νευρωνικών δικτύων για ταξινόμηση κειμένου. Μηχανισμός προσοχής [108] χρησιμοποιείται επίσης ευρέως στην κατηγοριοποίηση κειμένου. Οι Yang et al. [109] πρότειναν ιεραρχικά δίκτυα προσοχής χρησιμοποιώντας μηχανισμό προσοχής σε επίπεδο λέξης και πρότασης. Οι Lin et al. [110] πρότειναν την αυτό-προσοχή για να μάθουν δομημένη ενσωμάτωση πρότασης.

4.2.2. Σχετικός συλλογισμός

Ο σχετικός συλλογισμός με δίκτυα σχέσεων (relation networks - RN) χρησιμοποιείται αρχικά για την ανακάλυψη αντικειμένου σκηνής εκμεταλλευόμενος τις σχέσεις μεταξύ αντικειμένων [111]. Τα δίκτυα RN εισάγονται περαιτέρω στο σχετικό συλλογισμό για οπτική απάντηση

ερωτήσεων υπολογίζοντας τη βαθμολογία σχέσης των χαρτών χαρακτηριστικών των ζευγών αντικειμένων και την αναπαράσταση ερωτήσεων [112]. Στην κοινότητα της μάθησης αναπαράστασης της βάσης γνώσης, μελετάται επίσης ο σχετικός συλλογισμός μεταξύ θεμάτων και αντικειμένων στις βάσεις γνώσης [113]. Όσον αφορά το σενάριο εφαρμογής μας για την ανίχνευση αυτοκτονικού ιδεασμού, είναι κρίσιμο να κατανοήσουμε τη σχέση μεταξύ αυτοκτονίας και δεικτών κινδύνου όπως το συναίσθημα του ατόμου και τα γεγονότα της ζωής.

4.3. Μέθοδοι

4.3.1. Ορισμός Προβλήματος

Η ανίχνευση ιδεών αυτοκτονίας και ψυχικών διαταραχών στο κοινωνικό περιεχόμενο είναι τεχνικά μια συγκεκριμένη εργασία ταξινόμησης κειμένου. Σε αυτήν την ενότητα, διεξάγουμε λεπτομερή εκτίμηση κινδύνου αυτοκτονίας και ταξινόμηση πολλαπλών θεμάτων ψυχικής υγείας, τα οποία φυσικά θεωρούνται ως ταξινόμηση πολλαπλών κατηγοριών. Για τον εκλεπτυσμένο κίνδυνο αυτοκτονίας, τα επίπεδα κινδύνου περιλαμβάνουν κανένα, χαμηλό, μέτριο και σοβαρό κίνδυνο, ενώ για την ταξινόμηση της ψυχικής υγείας, συγκεκριμένες ψυχικές διαταραχές είναι η κατάθλιψη, το άγχος, η διπολική διαταραχή κ.ο.κ. Και υπάρχουν δύο δευτερεύουσες εργασίες για συγκεκριμένες ρυθμίσεις δεδομένων σε κοινωνικό περιεχόμενο, δηλαδή ταξινόμηση μετά το επίπεδο(post-level) και ταξινόμηση σε επίπεδο χρήστη(user-level). Η πρώτη λαμβάνει ως είσοδο τη μοναδική ανάρτηση p , ενώ η δεύτερη ανιχνεύει την απόπειρα αυτοκτονίας με πολλαπλές αναρτήσεις $P = \{ p_1 , p_2 , \ldots , p_n \}$.

4.3.2. Μοντέλο Αρχιτεκτονικής

Το προτεινόμενο μοντέλο αποτελείται από δύο βήματα, δηλαδή, μονάδα αναπαράστασης μετά και συσχέτισης σχέσεων, όπως απεικονίζεται στο Σχήμα 15. Η αναπαράσταση μετά περιλαμβάνει δύο μέρη εξαγωγής δεικτών κατάστασης που σχετίζονται με τον κίνδυνο και κωδικοποιητή κειμένου LSTM. Η ενότητα σχέσης όπως φαίνεται στο διακεκομμένο πλαίσιο του Σχήματος 15 χρησιμοποιεί το δίκτυο σχέσεων vanilla για να αιτιολογήσει τη σύνδεση μεταξύ δεικτών κατάστασης και αναρτήσεων χρήστη, και μηχανισμό προσοχής για την ιεράρχηση προτεραιότητας πιο σημαντικών βαθμολογιών σχέσης των σχετικών συλλογισμών.

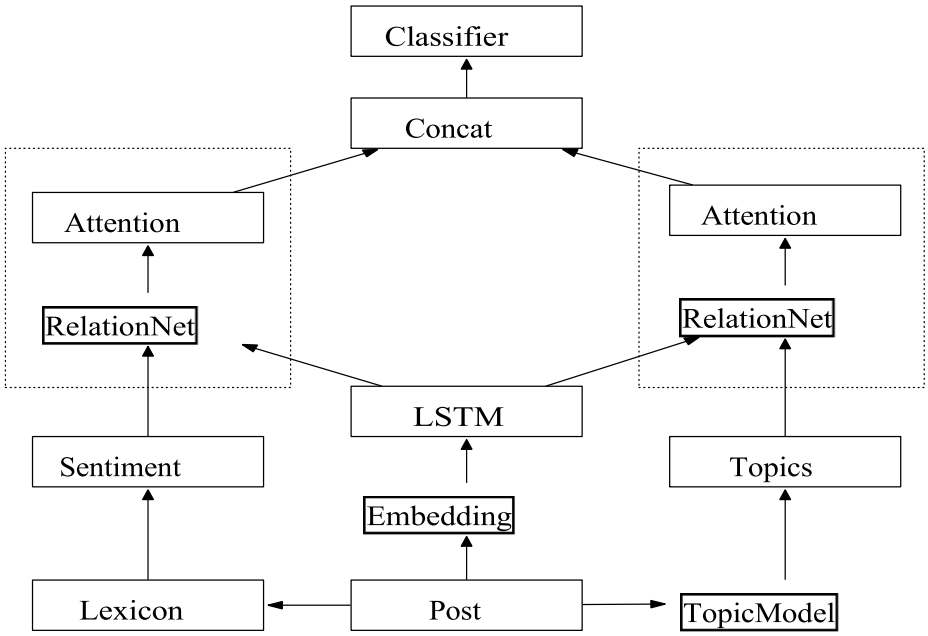
4.3.3. Κωδικοποίηση κειμένου και δείκτες κινδύνου

Η ακολουθία αναρτήσεων χρήστη είναι ενσωματωμένη σε διανύσματα λέξεων $p = \{w_1, w_2, \ldots, w_n\} \in \mathbb{R}^{l \times d}$. Εφαρμόζουμε αμφίδρομο LSTM στην εξίσωση. 4.1 για κωδικοποίηση κειμένου, για την καταγραφή της παρακείμενης εξάρτησης των λέξεων.

$$\overrightarrow{h_t} = \overrightarrow{LSTMcell}(w_t, \overrightarrow{h_{t-1}})$$
$$\overleftarrow{h_t} = \overleftarrow{LSTMcell}(w_t, \overleftarrow{h_{t+1}}) \quad (4.1)$$

Η κρυφή κατάσταση επιτυγχάνεται ενώνοντας κάθε κατεύθυνση ως $h_t = \text{concat}(\overrightarrow{h_t}, \overleftarrow{h_t})$, όπου $h_t \in \mathbb{R}^{l \times 2n}$ δίνεται n ως ο αριθμός κρυφών μονάδων.

Οι συναισθηματικές πληροφορίες παίζουν σημαντικό ρόλο όταν οι άνθρωποι εκφράζουν τα βάσανα και τα συναισθήματά τους σε διαδικτυακά κοινωνικά δίκτυα. Για τη μέτρηση του συναισθήματος, λαμβάνουμε λεξικά συναισθημάτων ως πρόσθετες πληροφορίες, συγκεκριμένα, χρησιμοποιούνται λεξικά συναισθημάτων για συγκεκριμένους τομείς [114] από κοινότητες στο Reddit. Τα λεξικά συναισθήματος προκαλούνται από λέξεις σπόρου με ενσωμάτωση λέξεων για συγκεκριμένο τομέα και πλαίσιο διάδοσης ετικέτας. Οι εξαγόμενες πληροφορίες συναισθημάτων της δημοσίευσης που σημειώνονται ως $s \in \mathbb{R}^l$ λειτουργούν ως δείκτες κατάστασης που αντιπροσωπεύουν την εσωτερική συναισθηματική κατάσταση των συγγραφέων των δημοσιεύσεων. Αντίστοιχα, εξωτερικοί δείκτες όπως τα γεγονότα της ζωής των ανθρώπων αποκαλύπτουν μια άλλη διάσταση ως δείκτη κινδύνου. Για να συλλάβουμε εξωτερικούς παράγοντες αυτοκτονικού ιδεασμού ή διαταραχών ψυχικής διαταραχής, εισάγουμε το μοντέλο θεμάτων για να μάθουμε τοπικά χαρακτηριστικά χωρίς επίβλεψη. Συγκεκριμένα, το Latent Dirichlet Allocation (LDA) [115] εφαρμόζεται για την εξαγωγή λανθάνουσων θεμάτων σε κοινωνικές αναρτήσεις για να αντιπροσωπεύει τα βάσανα των ανθρώπων, όπως γεγονότα της ζωής, κοινωνική έκθεση και άλλη εμπειρία στον πραγματικό κόσμο. Τα διανύσματα βαθμολογίας πιθανότητας των αναρτήσεων που ανήκουν σε όλα τα εξαγόμενα θέματα αναπαρίστανται ως $v \in \mathbb{R}^m$, όπου m είναι ο αριθμός των θεμάτων.



Σχήμα 15: Η αρχιτεκτονική του προτεινόμενου μοντέλου

4.3.4. Δίκτυο Σχέσεων με Προσοχή

Το δίκτυο σχέσεων [112] είναι μια νευρωνική ενότητα για σχεσιακό συλλογισμό. Αρχικά προτάθηκε να αποτυπωθεί η σχέση μεταξύ αντικειμένων. Δεδομένα αντικείμενα $O = \{o_1, o_2, \ldots, o_n\}$ και συναρτήσεις των f_φ και g_θ, ένα δίκτυο σχέσεων ορίζεται στην Εξίσωση 4.2. Η έξοδος του g_θ ονομάζεται μαθημένη «σχέση», ενώ η συνάρτηση f_φ λειτουργεί ως ταξινομητής.

$$RN(O) = f_\varphi\left(\sum_{i,j} g_\theta\,(o_i, o_j)\right) \quad (4.1)$$

Στόχος μας είναι να αιτιολογήσουμε παράγοντες κινδύνου αυτοκτονικού ιδεασμού και ψυχικών διαταραχών. Έτσι, παίρνουμε την κωδικοποίηση κειμένου και τους δείκτες κατάστασης ως την είσοδο των δικτύων σχέσεων για να υπολογίσουμε τις βαθμολογίες σχέσης μεταξύ κάθε διακριτικού σε αναρτήσεις και δείκτες κατάστασης που διαμορφώνονται βάσει του χρόνου και των χαρακτηριστικών του θέματος. Για να ενισχυθεί ο σχεσιακός συλλογισμός, ο μηχανισμός προσοχής ενσωματώνεται με την ενότητα σχέσεων, αποδίδοντας βάρη προσοχής στις σχέσεις που έχουν μάθει. Η ιδέα του προσεκτικού δικτύου σχέσεων φαίνεται στο Σχήμα 16. Το κείμενο που αντιπροσωπεύει κωδικοποιείται από ένα δίκτυο LSTM το οποίο

καταγράφει τη διαδοχική ανεξαρτησία και στη συνέχεια συνδυάζεται με τους δείκτες διευρυμένης κατάστασης. Εδώ, εξετάζουμε δύο δείκτες συναισθημάτων και χαρακτηριστικών θεμάτων, με τις διευρυμένες αναπαραστάσεις να συμβολίζονται ως $S = [s, s, \ldots, s] \in \mathbb{R}^{l \times l}$ και $V = [v, v, \ldots, v] \in \mathbb{R}^{l \times m}$αντίστοιχα. Στη συνέχεια, εισάγονται σε δίκτυα σχέσεων για τον υπολογισμό του διανύσματος σχέσης $r_i \in \mathbb{R}^k$ με multiple layer perception (MLP) όπως στην Εξίσωση 4.3 για τον δείκτη συναισθήματος.

$$r_i = \text{MLP}(h_i, s_i) \tag{4.3}$$

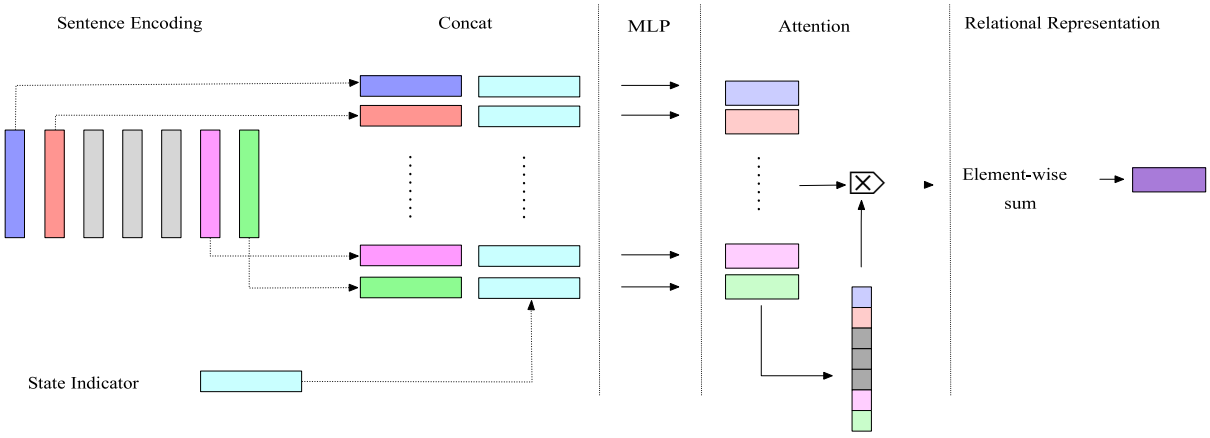
Η προσοχή υπολογίζεται ως εξής:

$$\alpha = \text{softmax}([r_1, r_2, \ldots, r_l]W^T + b) \tag{4.4}$$

όπου $W \in \mathbb{R}^{1 \times k}$, $b \in \mathbb{R}^l$ και $\alpha \in \mathbb{R}^l$. Με βάση το προϊόν από άποψη στοιχείων, η προσεκτική αναπαράσταση των μαθημένων σχέσεων μπορεί να υπολογιστεί ως :

$$\tilde{r} = \alpha \otimes [r_1, r_2, \ldots, r_i] \tag{4.5}$$

όπου $\tilde{r} \in \mathbb{R}^{l \times k}$. Εφαρμόζοντας το άθροισμα των στοιχείων στο $\tilde{r}$, παίρνουμε την τελική προσεκτική σχεσιακή αναπαράσταση.



Σχήμα 16: Δίκτυο σχέσεων με προσοχή

4.3.5. Ταξινόμηση

Το τελευταίο βήμα είναι να χρησιμοποιηθεί η ταξινομημένη αναπαράσταση, η οποία περιέχει διαδοχικές πληροφορίες και σχέση δεικτών κινδύνου, για ταξινόμηση. Συγκεκριμένα,

συνδυάζουμε σχεσιακές αναπαραστάσεις του $e = [\tilde{r}_s, \tilde{r}_v]$ από δύο κανάλια όπως φαίνεται στο σχήμα 15 και χρησιμοποιούμε το πλήρως συνδεδεμένο στρώμα με μη γραμμική συνάρτηση ενεργοποίησης του $f(\cdot)$ για να παράγουμε τα logits για πρόβλεψη ως εξής.

$$l = \mathrm{f}(W_l e + b_l)$$

$$\mathcal{P} = \mathrm{softmax}(W_o l + b_o) \tag{4.6}$$

όπου $\mathrm{W}_l \in \mathbb{R}^{d_l \times d_e}$, $\mathrm{b}_l \in \mathbb{R}^{d_l}$, $\mathrm{W}_o \in \mathbb{R}^{c \times d_l}$, $\mathrm{b}_o \in \mathbb{R}^{c}$, $P \in \mathbb{R}^{c}$ Για ταξινόμηση πολλαπλών κλάσεων, η προβλεπόμενη ετικέτα παράγεται από

$$\hat{y} = arg\max_i(\mathcal{P}_i) \tag{4.7}$$

4.3.6. Εκπαίδευση

Υπάρχουν δύο φάσεις κατάρτισης του προτεινόμενου μοντέλου μας, δηλαδή η εκπαίδευση του μοντέλου θεμάτων LDA και του μοντέλου ταξινόμησης. Για το μοντέλο θεμάτων, το LDA αναλαμβάνει μια γενεσιουργή διαδικασία εγγράφων ως τυχαία μίγματα σε λανθάνοντα θέματα και ένα θέμα μπορεί να συναχθεί ως κατανομή πάνω στις λέξεις, όπου το Bayesian συμπέρασμα χρησιμοποιείται για την εκμάθηση διαφόρων κατανομών. Στην πράξη, χρησιμοποιούμε τη βιβλιοθήκη Gensim[28] για τη δημιουργία του μοντέλου θεμάτων κατά την εφαρμογή.

Για τον τελικό στόχο της ανίχνευσης του αυτοκτονικού ιδεασμού και των ασθενειών ψυχικής υγείας , χρησιμοποιούμε τη διασταυρούμενη απώλεια εντροπίας με κανονικοποίηση L2 ως εξής.

$$L = -\frac{1}{\sum_{s=1}^{N} c(s)} \sum_{i=1}^{N} \sum_{j=1}^{c(i)} \log \mathcal{P}_{i,j}\left[\mathrm{y}_{i,j}\right] + \lambda \|\theta\|_2 \tag{4.8}$$

όπου $c(s)$ είναι το σύνολο ετικετών, το θ αντιπροσωπεύει όλες τις προπονούμενες παραμέτρους και λ είναι ο συντελεστής κανονικοποίησης ή ο λεγόμενος ρυθμός διάσπασης βάρους.

[28] https://radimrehurek.com/gensim/

Εφαρμόζουμε τον αλγόριθμο Adam [116] για τη βελτιστοποίηση της αντικειμενικής συνάρτησης.

4.4. Δεδομένα

Χρησιμοποιούμε δεδομένα από δύο δημοφιλείς ιστότοπους κοινωνικής δικτύωσης, δηλαδή, το Reddit και το Twitter, με συνολικά τρία σύνολα δεδομένων που προέρχονται από τους ανωτέρω ιστότοπους. Δύο από αυτά προέρχονται από το Reddit με ένα δημόσιο σύνολο δεδομένων και ένα πρώτο που συλλέχθηκε σε αυτήν την ενότητα. Δημοσιεύσεις ατόμων από ένα ενεργό subreddit για online υποστήριξη στο Reddit, που ονομάζεται "SuicideWatch" (SW) [29], χρησιμοποιούνται εντατικά σε αυτά τα δύο σύνολα δεδομένων. Το τελευταίο συλλέγεται από το Twitter συνδυάζοντας αρκετές υπάρχουσες πηγές δεδομένων. Αυτά τα σύνολα δεδομένων καλύπτουν την αυτοκτονία και άλλα θέματα ψυχικής υγείας, με συγκεκριμένες κατηγορίες που αναφέρονται στο ICD-10 όπως παρατίθενται στον Πίνακα 11.

Κατηγορία	Περιγραφές κατηγορίας που αναφέρονται στο ICD-10
Αυτοκτονία	Εσκεμμένος αυτοτραυματισμός, αυτοκτονικός ιδεασμός (τάσεις)
Κατάθλιψη	Σε τυπικά ήπια, μέτρια ή σοβαρά καταθλιπτικά επεισόδια, ο ασθενής υποφέρει από διαταραχή της διάθεσης, μείωση της ενέργειας και μείωση της δραστηριότητας.
Άγχος	Φοβικό άγχος και άλλοι παράγοντες αποτροπής άγχους
Διπολική διαταραχή	Μια δυσλειτουργία που χαρακτηρίζεται από δύο ή περισσότερα επεισόδια στα οποία διαταράσσεται σημαντικά η διάθεση και τα επίπεδα δραστηριότητας του ασθενούς
PTSD	Προκύπτει ως μια καθυστερημένη και παρατεταμένη αντίδραση σε ένα αγχωτικό γεγονός ή κατάσταση (είτε σύντομης είτε μακράς διάρκειας) ενός εξαιρετικά απειλητικού ή καταστροφικού χαρακτήρα

Πίνακας 11: Περιγραφή ψυχικών διαταραχών στο ICD-10

4.4.1. UMD Reddit Σύνολο Δεδομένων Αυτοκτονίας

Το UMD Reddit Σύνολο Δεδομένων Αυτοκτονίας [13] συλλέχθηκε από ανώνυμα φόρουμ συζήτησης στο Reddit.com . Περιέχει αναρτήσεις 620 χρηστών στο σετ εκπαίδευσης και 245

[29] https://www.reddit.com/r/SuicideWatch

χρηστών στο σετ δοκιμών, από 11,128 χρήστες στο subreddit “SuicideWatch” και 11,129 χρήστες σε άλλες δευτερεύουσες ενότητες. Σημειώνεται από εργαζόμενους στο πλήθος και ειδικούς σε ανθρώπους μέσω πλατφόρμας πλήθους. Ο κίνδυνος αυτοκτονίας κλιμακώνεται σε τέσσερα επίπεδα, δηλαδή, χωρίς κίνδυνο (α), χαμηλό (β), μέτριο (γ) και σοβαρό κίνδυνο (δ). Παρέχει επίσης χονδροειδείς ετικέτες όπου σε κανέναν κίνδυνο και χαμηλό κίνδυνο δίνεται η ετικέτα 0, ο μέτριος και ο σοβαρός κίνδυνος επισημαίνονται ως 1, μαζί με την ομάδα ελέγχου ως η ετικέτα -1.

Αυτό το σύνολο δεδομένων κυκλοφόρησε ως κοινόχρηστη εργασία CLPsych 2018 [13] και στη συνέχεια μια νέα έκδοση του λειτούργησε ως κοινόχρηστη εργασία CLPsych 2019 [61]. Σε αυτήν την ενότητα, χρησιμοποιούμε σύνολα δεδομένων που προέρχονται από το σύνολο δεδομένων UMD σε τέσσερις κατηγορίες κινδύνου αυτοκτονίας σε επίπεδο χρήστη. Οι στατιστικές πληροφορίες του συνόλου δεδομένων απεικονίζονται στον πίνακα 12. Επιπλέον, συμπεριλαμβάνουμε χρήστες ελέγχου (με ετικέτα "Κανένας") σε αυτό το σύνολο σχολιασμών.

Σχόλιο	Αριθμοί	% των επιπέδων α/β/γ/δ/
πλήθος	621	26%/10%/24%/40%
ειδικός	245	29%/9%/25%/37%

Πίνακας 12: Στατιστικές πληροφορίες του UMD Reddit Συνόλου Δεδομένων Αυτοκτονίας

Χρησιμοποιούμε τις μετασχηματισμένες ετικέτες από την ακατέργαστη ετικέτα σύμφωνα με την αρχική περιγραφή αυτού του συνόλου δεδομένων. Συγκεκριμένα, οι ακατέργαστες ετικέτες "γ" ή "δ" μετατρέπονται σε 1, οι ακατέργαστες ετικέτες "α" ή "β" μετατρέπονται σε 0 και η ετικέτα ενός χρήστη ελέγχου είναι -1 εξ ορισμού. Χωρίσαμε ολόκληρο το σύνολο δεδομένων σε σύνολα εκπαίδευσης, επικύρωσης και δοκιμών, όπως παρατίθενται στον Πίνακα 13.

Ετικέτα	#/% of train	#/% of valid.	#/% of test
-1	495/49.8489%	126/50.6024%	245/50.000%
0	188/31.2185%	89/35.7430%	86/17.551%
1	310/18.9325%	34/13.6546%	159/32.449%

Πίνακας 13: Στατιστικές πληροφορίες του συνόλου δεδομένων UMD με διαίρεση εκπαίδευσης/επικύρωσης/δοκιμής

4.4.2. SWMH Reddit Σύνολο Δεδομένων

Καθώς τα σοβαρά ζητήματα ψυχικής υγείας είναι πολύ πιθανό να οδηγήσουν σε ιδέες αυτοκτονίας, συλλέγουμε επίσης ένα άλλο σύνολο δεδομένων από ορισμένες υποδιαιρέσεις που σχετίζονται με την ψυχική υγεία στο Reddit.com για να προωθήσουμε τη μελέτη των ψυχικών διαταραχών και των αυτοκτονικών ιδεών. Ονομάζουμε αυτό το σύνολο δεδομένων ως Reddit SuicideWatch και Mental Health Collection, ή για συντομία SWMH, όπου οι συζητήσεις περιλαμβάνουν πρόθεση που σχετίζεται με την αυτοκτονία και ψυχικές διαταραχές όπως κατάθλιψη, άγχος και διπολική διαταραχή. Χρησιμοποιούμε το επίσημο API του Reddit[30] και αναπτύσουμε έναν ιστό αράχνης(web spider) για να συλλέξουμε τα στοχευμένα φόρουμ. Αυτή η συλλογή περιέχει συνολικά 54.412 δημοσιεύσεις. Συγκεκριμένες υποδιαιρέσεις παρατίθενται στον Πίνακα 14, καθώς και ο αριθμός και το ποσοστό των θέσεων που συλλέχθηκαν στη διαίρεση train-val-test.

Σε αυτές τις κοινότητες ή τα λεγόμενα subreddits, οι άνθρωποι συζητούν για τις ψυχικές διαταραχές των δικών τους ή των συγγενών τους και ζητούν συμβουλές ή βοήθεια. Πραγματοποιούμε πειραματική ανάλυση σε αυτό το σύνολο δεδομένων για να εντοπίσουμε συζητήσεις σχετικά με την αυτοκτονία και τις ψυχικές διαταραχές.

Subreddit	#/% of train	#/% of valid.	#/% of test
Κατάθλιψη	11,940/34.29	3,032/34.83	3,774/34.68
SuicideWatch	6,550/18.81	1,614/18.54	2,018/18.54
Άγχος	6,136/17.62	1,508/17.32	1,911/17.56
Από το στήθος μου	5,265/15.12	1,332/15.30	1,687/15.50
Διπολική Διαταραχή	4,932/14.16	1,220/14.01	1,493/13.72

Πίνακας 14: Στατιστικές πληροφορίες για SuicideWatch και subreddits που σχετίζονται με την ψυχική υγεία, δηλαδή, σύνολο δεδομένων SWMH

4.4.3. Συλλογή Συνόλου Δεδομένων Twitter

Το τρίτο σύνολο δεδομένων είναι μια συλλογή διαφορετικών υποσυνόλων από το Twitter. Συλλογές δειγμάτων από δύο σύνολα δεδομένων αποτελούνται από τα περισσότερα δείγματα

[30] https://www.reddit.com/dev/api/

αυτού του συνόλου δεδομένων. Πρώτον, 594 περιπτώσεις tweets που περιέχουν ιδέες αυτοκτονίας προέρχονται από τους Ji et al. [17], με επιπλέον 606 tweets που συλλέχθηκαν χειροκίνητα από αυτήν την εργασία. Δεύτερον, ο ίδιος αριθμός καταχωρίσεων κατάθλιψης και μετατραυματικής διαταραχής (PTSD) λαμβάνεται από το κοινό σύνολο εργασιών CLPsych 2015 [117]. Αυτό το σύνολο δεδομένων είναι διαθέσιμο κατόπιν αιτήματος[31]. Τέλος, η ομάδα ελέγχου όπου οι χρήστες του Twitter δεν έχουν προσδιοριστεί ότι έχουν ψυχική κατάσταση ή αυτοκτονικό ιδεασμό αποτελείται από δειγματοληψία κανονικών tweets από προηγούμενα σύνολα δεδομένων [17,117]. Τέλος, αυτή η συλλογή δεδομένων Twitter περιλαμβάνει συνολικά 4.800 tweets με τέσσερις κατηγορίες αυτοκτονίας, κατάθλιψης, PTSD και ελέγχου.

4.4.4. Γλωσσικές ενδείξεις και πολικότητα συναισθημάτων

Έχουμε μια σύντομη αναλυτική ανάλυση των δεδομένων. Ορισμένες επιλεγμένες γλωσσικές στατιστικές πληροφορίες του Σύνολο δεδομένων UMD που εξήχθη από το λογισμικό Linguistic Inquiry and Word Count (LIWC) [32] φαίνονται στον Πίνακα 15. Ο κίνδυνος αυτοκτονίας αυξάνεται μεταξύ των ετικετών -1, 0 και 1. Τα αποτελέσματα της γλωσσικής έρευνας δείχνουν ότι το αρνητικό συναίσθημα, το άγχος και η θλίψη εκφράζονται περισσότερο σε δημοσιεύσεις με υψηλό κίνδυνο αυτοκτονίας. Οι ίδιες τάσεις υπάρχουν σε οικογενειακά θέματα, αναφορές που σχετίζονται με τον θάνατο και προσβλητικές λέξεις. Φυσικά, τα θετικά συναισθήματα παρουσιάζονται λιγότερο σε περιπτώσεις με υψηλό κίνδυνο αυτοκτονίας.

Γλωσσικές ενδείξεις	ετικέτα -1	ετικέτα 0	ετικέτα 1
Θετικό συναίσθημα	3.30	3.12	2.96
Αρνητικό συναίσθημα	1.56	2.30	2.74
άγχος	0.17	0.33	0.41
θλίψη	0.28	0.50	0.68
οικογένεια	0.29	0.39	0.47
φίλος	0.43	0.56	0.54
εργασία	2.54	1.92	1.80
χρήματα	1.13	0.71	0.61
θάνατος	0.22	0.29	0.36
προσβλητικές λέξεις	0.23	0.33	0.40

[31] Request for data access via http://www.cs.jhu.edu/~mdredze/datasets/clpsych_shared_task_2015/
[32] http://liwc.wpengine.com

Πίνακας 15: Επιλεγμένες γλωσσικές στατιστικές πληροφορίες του συνόλου δεδομένων UMD που εξήχθησαν από το LIWC

4.5. Πειράματα

Για να αξιολογήσουμε την απόδοση του προτεινόμενου μοντέλου μας, το συγκρίνουμε με αρκετά μοντέλα ταξινόμησης κειμένου σε τρία σύνολα δεδομένων πραγματικού κόσμου. Οι βασικές γραμμές και οι εμπειρικές ρυθμίσεις εισάγονται και τα αποτελέσματα αναφέρονται και συζητούνται σε αυτήν την ενότητα.

4.5.1. Βασική γραμμή και ρυθμίσεις

Συγκρίναμε πέντε δημοφιλή μοντέλα ταξινόμησης με την προτεινόμενη μέθοδο. Αυτά τα βασικά μοντέλα περιγράφονται ως εξής:

- **fastText** [118]: ένα αποτελεσματικό μοντέλο ταξινόμησης κειμένου με αναπαράσταση προτάσεων λέξεων και γραμμικό ταξινομητή.
- **CNN** [105]: εφαρμόζει συνελικτικά νευρωνικά δίκτυα πάνω από την ενσωμάτωση λέξης της πρότασης για την παραγωγή χαρτών χαρακτηριστικών και στη συνέχεια χρησιμοποιεί τη μέγιστη συγκέντρωση των χαρακτηριστικών.
- **LSTM** [106]: παίρνει διαδοχικά διανύσματα λέξεων ως είσοδο στα επαναλαμβανόμενα κελιά LSTM και εφαρμόζει συγκέντρωση πάνω στην έξοδο για την τελική αναπαράσταση. Συνδυάζοντας την κατεύθυνση προς τα εμπρός και προς τα πίσω, γίνεται αμφίδρομος LSTM (BiLSTM).
- **RCNN** [107]: αυτό το μοντέλο εφαρμόζει αρχικά το μοντέλο LSTM [106] για τη λήψη διαδοχικών πληροφοριών και στη συνέχεια εφαρμόζει το CNN [105] για περαιτέρω εξαγωγή χαρακτηριστικών. Έχει αμφίδρομη έκδοση χρησιμοποιώντας το BiLSTM.

- **SSA** [110]: πρότεινε έναν δομημένο μηχανισμό αυτόματης προσοχής με πολλαπλούς λυκίσκους εισάγοντας μια 2D μήτρα για ενσωμάτωση αναπαράστασης. Η αυτοεκτίμηση εφαρμόζεται στις διαδοχικές κρυφές καταστάσεις του δικτύου LSTM.

Όλα τα βασικά μοντέλα και η προτεινόμενη μέθοδος μας υλοποιούνται από την PyTorch[33] και τρέχουν σε ένα απλό GPU (Nvidia GeForce GTX 1080 Ti). Εκπαιδεύουμε τα μοντέλα για 50 εποχές από προεπιλογή, ορίζοντας το μέγεθος της παρτίδας να είναι 128 και 16 ανάλογα με το μέγεθος των συνόλων δεδομένων. Συγκεκριμένα, το μέγεθος παρτίδας του συνόλου δεδομένων UMD είναι 16, και για τη συλλογή δεδομένων SWMH και Twitter, το μέγεθος παρτίδας είναι 128. Για τη ενσωμάτωση λέξης, χρησιμοποιούμε προ-εκπαιδευμένη αναπαράσταση λέξης GloVe [55], με είτε στατική είτε δυναμική ενσωμάτωση. Η προτεινόμενη μέθοδος μας απαριθμεί όλα τα 250 λεξικά δευτερεύουσας έκδοσης του Reddit και τον αριθμό των θεμάτων από 5 έως 20. Επιλέγουμε την καλύτερη απόδοση επικύρωσης σε πολλαπλές διαδρομές και αναφέρουμε την απόδοση δοκιμών ως πειραματικά αποτελέσματα.

Ο στόχος της αυτόματης ανίχνευσης είναι αφιερωμένος στην παραγωγή αποτελεσματικών διαγνώσεων (δηλαδή true positive) και στη μείωση των λανθασμένων διαγνώσεων (δηλαδή false positive) για την αποφυγή του άγχους και του άγχους των ασθενών που προκαλούνται από την ψευδή ανίχνευση. Έτσι, κατά τη διαδικασία αξιολόγησης, εστιάζουμε μόνο στην ακρίβεια πρόβλεψης, αλλά επίσης αναφέρουμε τη μέση σταθμισμένη μέτρηση F-score. Για μη ισορροπημένα σύνολα δεδομένων, εφαρμόζουμε ποινή βάρους στην αντικειμενική συνάρτηση και αναφέρουμε τα σταθμισμένα μέσα αποτελέσματα.

4.5.2. Αποτελέσματα

Αξιολογούμε την πειραματική απόδοση σε τρία σύνολα δεδομένων που συλλέχθηκαν από το Reddit και το Twitter. Για το σύνολο δεδομένων αυτοκτονίας UMD και το σύνολο δεδομένων SWMH, τα αναφερόμενα αποτελέσματα σταθμίζονται κατά μέσο όρο.

Αυτοκτονία UMD

[33] https://www.pytorch.org/

Πρώτα εφαρμόζουμε τη μέθοδο και τις βασικές γραμμές μας στο σύνολο δεδομένων αυτοκτονίας UMD για ταξινόμηση σε επίπεδο χρήστη. Για την επεξεργασία ενός συνόλου αναρτήσεων από χρήστες, όλες οι αναρτήσεις χρηστών συνδυάζονται ως αναπαράσταση σε επίπεδο χρήστη. Τα αποτελέσματα τεσσάρων μετρήσεων πιστότητας**(Accuracy)**, ακριβείας**(Precision)**, ανάκλησης(**Recall**) και βαθμολογίας F1**(F1-score)** είναι πολύ κοντά σε αυτό το σύνολο δεδομένων. Το μοντέλο BiLSTM αποκτά την υψηλότερη πιστότητα **(Accuracy)** 56,94% και το μοντέλο μας ακολουθεί στη δεύτερη θέση 56,73%. Αλλά το μοντέλο μας έχει υψηλότερη βαθμολογία F1 από όλες τις βασικές γραμμές. Παρατηρώντας αυτά τα πολύ στενά αποτελέσματα , στη συνέχεια προχωρούμε σε περαιτέρω ανάλυση των αποτελεσμάτων κάθε τάξης στην επόμενη ενότητα.

Model	Accuracy	Precision	Recall	F1
fastText	0.5327	0.5300	0.5327	0.5202
CNN	0.5531	0.4498	0.5531	0.4935
LSTM	0.5612	0.4625	0.5612	0.5071
BiLSTM	**0.5694**	0.5029	**0.5694**	0.5233
RCNN	0.5592	0.4953	0.5592	0.5111
SSA	0.5633	0.4711	0.5633	0.4839
RN	0.5673	**0.5405**	0.5673	**0.5453**

Πίνακας 16: Σύγκριση διαφορετικών μοντέλων σε σύνολα δεδομένων UMD για ταξινόμηση σε επίπεδο χρήστη, όπου η πιστότητα, η ακρίβεια, η ανάκληση και η βαθμολογία F1 σταθμίζονται κατά μέσο όρο

Reddit SWMH

Στη συνέχεια, πραγματοποιούμε πειράματα στο σύνολο δεδομένων Reddit SWMH, το οποίο περιέχει τόσο αυτοκτονικό ιδεασμό όσο και θέματα ψυχικής υγείας για τη μελέτη της προγνωστικής απόδοσης του μοντέλου μας. Είναι ένα μεγαλύτερο σύνολο δεδομένων με περισσότερες περιπτώσεις σε σύγκριση με το σύνολο δεδομένων UMD. Τα πειράματα σε αυτό το σύνολο δεδομένων μας βοηθούν να έχουμε μια εικόνα της συλλογιστικής δύναμης του δικτύου σχέσεων για κείμενο που σχετίζεται με την ψυχική υγεία με παρόμοια χαρακτηριστικά. Όπως φαίνεται στον Πίνακα 17, το μοντέλο μας ξεπερνά όλα τα βασικά μοντέλα όσον αφορά και τις τέσσερις μετρήσεις

Model	Accuracy	Precision	Recall	F1
fastText	0.5722	0.5760	0.5722	0.5721
CNN	0.5657	0.5925	0.5657	0.5556
LSTM	0.5934	0.6032	0.5934	0.5917
BiLSTM	0.6196	0.6204	0.6196	0.6190
RCNN	0.6096	0.6161	0.6096	0.6063
SSA	0.6214	0.6249	0.6214	0.6226
RN	**0.6474**	**0.6510**	**0.6474**	**0.6478**

Πίνακας 17: Σύγκριση διαφορετικών μοντέλων στη συλλογή Reddit SWMH, όπου η πιστότητα, η ακρίβεια, η ανάκληση και η βαθμολογία F1 σταθμίζονται κατά μέσο όρο

Συλλογή Twitter

Τέλος, πραγματοποιούμε πειράματα στο σύνολο δεδομένων του Twitter με παρόμοια αποτελέσματα με των προηγούμενων πειραμάτων. Σε αντίθεση με τις αναρτήσεις στο Reddit, τα tweets σε αυτό το σύνολο δεδομένων είναι σύντομες ακολουθίες λόγω του ορίου μήκους του tweet των 280 χαρακτήρων. Τα αποτελέσματα όλων των βασικών μεθόδων και της προτεινόμενης μεθόδου μας φαίνονται στον Πίνακα 18. Μεταξύ αυτών των ανταγωνιστικών μεθόδων, το μοντέλο μας κερδίζει την καλύτερη απόδοση σε αυτές τις τέσσερις μετρικές, με βελτίωση 1,77% και 1,82% από το δεύτερο καλύτερο BiLSTM μοντέλο από άποψη πιστότητας**(Accuracy)**, και F1-score αντίστοιχα. Η προτεινόμενη μέθοδος μας εισάγει βοηθητικές πληροφορίες για το συναίσθημα που βασίζεται στο λεξικό και τα χαρακτηριστικά γνωρίσματα που έχουν διδαχθεί από το σώμα και χρησιμοποιεί σχεσιακό συλλογισμό για τη μοντελοποίηση της αλληλεπίδρασης μεταξύ κωδικοποιήσεων κειμένου που βασίζονται σε LSTM και δεικτών κινδύνου. Η κωδικοποίηση πλουσιότερων πληροφοριών και η αποτελεσματική συλλογιστική βοηθούν το μοντέλο μας να αυξήσει την απόδοση σε σύντομη ταξινόμηση tweet.

Model	Accuracy	Precision	Recall	F1
fastText	0.7927	0.7924	0.7927	0.7918
CNN	0.7885	0.7896	0.7885	0.7887
LSTM	0.8021	0.8094	0.8021	0.8039
BiLSTM	0.8208	0.8207	0.8208	0.8195
RCNN	0.8094	0.8089	0.8094	0.8090
SSA	0.8156	0.8149	0.8156	0.8152
RN	**0.8385**	**0.8381**	**0.8385**	**0.8377**

Πίνακας 18: Σύγκριση επιδόσεων στο σύνολο δεδομένων Twitter, όπου η πιστότητα, η ακρίβεια, η ανάκληση και η βαθμολογία F1 σταθμίζονται κατά μέσο όρο

4.5.3. Απόδοση σε κάθε τάξη

Αυτή η ενότητα μελετά την απόδοση σε κάθε κατηγορία συνόλου δεδομένων UMD. Επιλέγουμε δύο γραμμές βάσης με καλύτερη απόδοση για σύγκριση. Τα αποτελέσματα φαίνονται στον Πίνακα 19. Το προτεινόμενο μοντέλο που βασίζεται σε RN είναι κακό στην πρόβλεψη ανάρτησης χωρίς αυτοκτονία, αλλά καλό στην πρόβλεψη αναρτήσεων με υψηλό κίνδυνο αυτοκτονίας.

Δυστυχώς, και τα τρία αυτά μοντέλα έχουν πολύ χαμηλή ικανότητα πρόβλεψης αναρτήσεων με χαμηλό κίνδυνο αυτοκτονίας. Στο σύνολο δεδομένων UMD με μικρό όγκο εμφανίσεων, αυτά τα μοντέλα τείνουν να προβλέπουν δημοσιεύσεις ως κλάσεις με περισσότερες εμφανίσεις, παρόλο που εφαρμόζουμε ποινή στην αντικειμενική συνάρτηση.

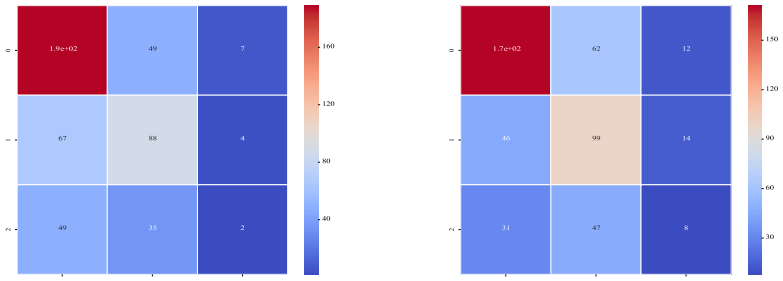
Ετικέτα	Μετρικές	BiLSTM	SSA	RN
-1	Precision	0.62	0.57	0.69
	Recall	0.77	**0.92**	0.70
	F1-score	0.69	0.70	0.69
1	Precision	0.51	0.57	0.48
	Recall	0.55	0.31	**0.62**
	F1-score	0.53	0.41	0.54
0	Precision	0.15	0.00	0.24
	Recall	0.02	0.00	0.09
	F1-score	0.04	0.00	0.13

Πίνακας 19: Απόδοση σε κάθε κατηγορία συνόλου δεδομένων αυτοκτονίας UMD

4.5.4. Ανάλυση σφαλμάτων

Αυτή η ενότητα πραγματοποιεί ανάλυση σφαλμάτων, λαμβάνοντας ως παράδειγμα το σύνολο δεδομένων UMD. Όπως αναφέρθηκε προηγουμένως στην τελευταία ενότητα, οι περισσότερες μέθοδοι υποφέρουν από κακή απόδοση στην πρόβλεψη αναρτήσεων χαμηλού κινδύνου. Το

Σχήμα 17 δείχνει τους χάρτες θερμότητας της μήτρας σύγχυσης του BiLSTM και του μοντέλου μας βασισμένου σε RN , όπου οι άξονες 0, 1 και 2 αντιπροσωπεύουν την ετικέτα των -1, 1 και 0. Αυτές οι δύο μέθοδοι τείνουν να προβλέπουν περισσότερες περιπτώσεις ως μηδενικές ή υψηλού κινδύνου. Και η προτεινόμενη μέθοδος μας έχει ένα ελαφρώς καλύτερο αποτέλεσμα από την αντίστοιχη της. Παρατηρούμε επίσης ότι το προτεινόμενο μοντέλο μας μπορεί να επιτύχει υψηλότερη πιστότητα **(Accuracy)**, 59,18% στο σύνολο δεδομένων UMD. Αποτυγχάνει όμως στην πρόβλεψη αυτοκτονικού ιδεασμού χαμηλού κινδύνου, με παρόμοια απόδοση άλλων γραμμών βάσης.



(α) BiLSTM **(β)** RN

Σχήμα 17: Πίνακας σύγχυσης σε σύνολο δεδομένων UMD

Χρησιμοποιούμε λεξικά συναισθημάτων και μοντελοποίηση θεμάτων για ακριβείς δείκτες κινδύνου που σχετίζονται με το συναίσθημα και το θέμα για σχετικούς συλλογισμούς με κωδικοποίηση κειμένου. Αυτή η διαδικασία προεπεξεργασίας μπορεί να προκαλέσει διάδοση σφαλμάτων. Το συναίσθημα ποικίλλει σε διαφορετικές κοινωνικές κοινότητες. Η χρήση υπαρχόντων λεξικών σε δημοφιλείς κοινότητες μπορεί να έχει περιορισμούς. Στη μελλοντική εργασία, θα εξετάσουμε τη δημιουργία λεξικών από κοινότητες που σχετίζονται με την ψυχική υγεία.

Κεφάλαιο 5

5. Συμπέρασμα

Θέματα ψυχικής υγείας, όπως το άγχος, η κατάθλιψη και οι αυτοκτονικός ιδεασμός, γίνονται όλο και πιο ανησυχητικά στη σύγχρονη κοινωνία. Σοβαρές ψυχικές διαταραχές, αλλά χωρίς αποτελεσματική θεραπεία είναι πολύ πιθανό να στραφούν σε αυτοκτονία. Οι λόγοι που αυτοκτονούν οι άνθρωποι είναι περίπλοκοι, συμπεριλαμβανομένων κοινωνικών παραγόντων όπως η κοινωνική απομόνωση· προσωπικά θέματα, όπως αλκοολισμός ή αποτυχία σταδιοδρομίας· ή επίδραση αρνητικών γεγονότων της ζωής. Συνεπώς, απαιτείται επειγόντως η ανάπτυξη αποτελεσματικών τεχνικών πρόληψης αυτοκτονίας. Η έγκαιρη ανίχνευση αυτοκτονικού ιδεασμού είναι μία από τις πιο αποτελεσματικές μεθόδους για την πρόληψη της αυτοκτονίας. Η έρευνα για την κατανόηση της πρόθεσης αυτοκτονίας και την πρόληψη της αυτοκτονίας επικεντρώνεται κυρίως στις ψυχολογικές και κλινικές πτυχές της και στην ταξινόμηση των αποτελεσμάτων του ερωτηματολογίου μέσω εποπτευόμενης μάθησης. Ωστόσο, η συλλογή δεδομένων και/ή ασθενών είναι συνήθως δαπανηρή, τόσο από ψυχολογική όσο και από κλινική άποψη.

Με την πρόοδο στα μέσα κοινωνικής δικτύωσης, όλο και περισσότερα άτομα εκφράζουν τα συναισθήματά τους και τα βάσανά τους στο διαδίκτυο. Ανώνυμοι διαδικτυακοί ιστότοποι παρέχουν ένα βολικό μέρος για να αλληλεπιδρούν οι άνθρωποι με άλλους χρησιμοποιώντας ασύγχρονη επικοινωνία. Ο όγκος του κειμένου αυξάνεται συνεχώς με τη δημοτικότητα των υπηρεσιών κοινωνικής δικτύωσης και η πρόληψη των αυτοκτονιών παραμένει ένα σημαντικό καθήκον στη σύγχρονη κοινωνία μας. Είναι επομένως ουσιαστικό να αναπτυχθούν νέες μέθοδοι για την ανίχνευση διαδικτυακών κειμένων που περιέχουν ιδέες αυτοκτονίας με την ελπίδα ότι μπορεί να προληφθεί η αυτοκτονία. Το κοινωνικό περιεχόμενο σε διαδικτυακές κοινότητες για την κατάθλιψη παρέχει χαρακτηριστικά θεμάτων και ψυχογλωσσικές ενδείξεις για την αυτόματη ανίχνευση και πρόβλεψη αυτοκτονικού ιδεασμού. Η χρήση τεχνικών εξόρυξης δεδομένων σε κοινωνικά δίκτυα και η εφαρμογή νευρωνικών δικτύων παρέχουν τη δυνατότητα να κατανοήσουμε την πρόθεση μέσα σε διαδικτυακές αναρτήσεις και ακόμη και να ανακουφίσουμε τις αυτοκτονικές προθέσεις ενός ατόμου .

Αυτή η πτυχιακή εργασία πραγματοποιεί πρώτα μια ολοκληρωμένη βιβλιογραφική ανασκόπηση σχετικά με τις τρέχουσες μεθόδους ανίχνευσης αυτοκτονικού ιδεασμού, συμπεριλαμβανομένων κλινικών μεθόδων που βασίζονται στην αλληλεπίδραση μεταξύ κοινωνικών λειτουργών ή ειδικών και στοχευμένων ατόμων, και τεχνική μηχανικής μάθησης με μηχανική χαρακτηριστικών ή βαθιά μάθηση για αυτόματη ανίχνευση με βάση διαδικτυακά

κοινωνικά περιεχόμενα. Οι εφαρμογές ανίχνευσης αυτοκτονικού ιδεασμού για συγκεκριμένους τομείς εξετάζονται επίσης σύμφωνα με τις πηγές δεδομένων τους, δηλαδή ερωτηματολόγια, ηλεκτρονικές καταστάσεις υγείας , σημειώσεις αυτοκτονίας και περιεχόμενο σε απευθείας σύνδεση χρήστη.

Εν συνεχεία, για να κατανοήσουμε τον αυτοκτονικό ιδεασμό μέσω διαδικτυακού περιεχομένου που δημιουργείται από χρήστες με στόχο την έγκαιρη ανίχνευση μέσω εποπτευόμενης μάθησης, αναλύουμε τις γλωσσικές προτιμήσεις των χρηστών και τις περιγραφές θεμάτων που αποκαλύπτουν σημαντικές γνώσεις για τον εντοπισμό τάσεων αυτοκτονίας. Τα άτομα που έχουν τάσεις αυτοκτονίας εκφράζουν έντονα αρνητικά συναισθήματα, άγχος και απελπισία. Οι σκέψεις αυτοκτονίας μπορεί να αφορούν οικογένεια και φίλους. Και θέματα που συζητούν αφορούν τόσο προσωπικά όσο και κοινωνικά ζητήματα. Για να ανιχνεύσουμε ιδέες αυτοκτονίας, εξάγουμε πολλά ενημερωτικά σύνολα χαρακτηριστικών, συμπεριλαμβανομένων στατιστικών, συντακτικών, γλωσσικών, ενσωμάτωσης λέξεων και χαρακτηριστικών θεμάτων και συγκρίνουμε έξι ταξινομητές, συμπεριλαμβανομένων τεσσάρων παραδοσιακών εποπτευόμενων ταξινομητών και δύο μοντέλων νευρωνικών δικτύων. Μια πειραματική μελέτη καταδεικνύει τη σκοπιμότητα και την πρακτικότητα της προσέγγισης και παρέχει σημεία αναφοράς για την ανίχνευση αυτοκτονικών ιδεών στις ενεργές διαδικτυακές πλατφόρμες: Reddit SuicideWatch και Twitter. Ενώ η εκμετάλλευση πιο αποτελεσματικών συνόλων χαρακτηριστικών, σύνθετα μοντέλα ή άλλοι παράγοντες όπως οι χρονικές πληροφορίες μπορεί να βελτιώσουν την ανίχνευση αυτοκτονικής ταυτότητας - αυτές θα είναι οι μελλοντικές μας κατευθύνσεις, η συμβολή και ο αντίκτυπος αυτού του τμήματος είναι τριπλός: (1) παροχή πλούσιας γνώσης στην κατανόηση αυτοκτονικού ιδεασμού, (2) εισαγωγή συνόλων δεδομένων για την ερευνητική κοινότητα για τη μελέτη αυτού του σημαντικού προβλήματος , και (3) πρόταση πληροφοριακών χαρακτηριστικών και αποτελεσματικών μοντέλων για την ανίχνευση αυτοκτονικών ιδεών.

Η ταξινόμηση του αυτοκτονικού ιδεασμού και άλλων ψυχικών διαταραχών είναι ένα δύσκολο έργο καθώς μοιράζονται αρκετά παρόμοια μοτίβα στη χρήση της γλώσσας και την συναισθηματική πολικότητα. Σε αυτή την πτυχιακή εργασία, ενισχύουμε την αναπαράσταση κειμένου με βαθμολογίες συναισθημάτων που βασίζονται σε λεξικό και λανθάνοντα θέματα και προτείνουμε τη χρήση δικτύων σχέσεων για τον εντοπισμό αυτοκτονικών ιδεών και ψυχικών διαταραχών με σχετικούς δείκτες κινδύνου. Η ενότητα σχέσεων είναι περαιτέρω

εξοπλισμένη με τον μηχανισμό προσοχής για να δώσει προτεραιότητα σε πιο σημαντικά χαρακτηριστικά σχέσης. Μέσω πειραμάτων σε τρία σύνολα δεδομένων πραγματικού κόσμου, το μοντέλο μας υπερτερεί των περισσότερων ομολόγων του.

Η πρόοδος των τεχνικών βαθιάς μάθησης έχει ενισχύσει την έρευνα για την ανίχνευση αυτοκτονικών ιδεών. Στο μελλοντικό έργο, θα μελετηθούν πιο αναδυόμενες τεχνικές μάθησης, όπως ο μηχανισμός προσοχής και τα γραφικά νευρωνικά δίκτυα για μάθηση αναπαράστασης κειμένου αυτοκτονίας. Μπορούν επίσης να χρησιμοποιηθούν και άλλα παραδείγματα μάθησης, όπως η μεταφορά μάθησης, η εχθρική εκπαίδευση και η ενισχυτική μάθηση. Για παράδειγμα, η γνώση σχετικά με τον τομέα ανίχνευσης ψυχικής υγείας μπορεί να μεταφερθεί για ανίχνευση αυτοκτονικού ιδεασμού και τα γενεσιουργά αντίπαλα δίκτυα μπορούν να χρησιμοποιηθούν για τη δημιουργία εχθρικών δειγμάτων για αύξηση δεδομένων. Στις υπηρεσίες κοινωνικής δικτύωσης, οι αναρτήσεις με αυτοκτονικό ιδεασμό βρίσκονται στη μεγάλη ουρά της διανομής διαφορετικών κατηγοριών αναρτήσεων. Προκειμένου να επιτευχθεί αποτελεσματική ανίχνευση στην κακή ισορροπημένη κατανομή του σεναρίου σε πραγματικό κόσμο, η μάθηση με ελάχιστες δυνατότητες μπορεί να χρησιμοποιηθεί για εκπαίδευση σε λίγες επισημασμένες θέσεις με αυτοκτονικό ιδεασμό μεταξύ του μεγάλου κοινωνικού σώματος.

--.--

Βιβλιογραφία-Αναφορές

---

[1] S. Hinduja, JW Patchin, Bullying, cyber bullying και αυτοκτονία, Αρχείο έρευνας αυτοκτονίας 14 (3) (2010) 206–221.

[2] AE Crosby, B. Han, LAG Ortega, SE Parks, J. Gfroerer, αυτοκτονικές σκέψεις και συμπεριφορές μεταξύ ενηλίκων ηλικίας ≥ 18 ετών - Ηνωμένες Πολιτείες, 2008-2009., Εβδομαδιαία έκθεση νοσηρότητας και θνησιμότητας 60 (13) (2011) 1 - 22.

[3] J. Joo, S. Hwang, JJ Gallo, Ιόντα θανάτου και αυτοκτονικός ιδεασμός σε δείγμα κοινότητας που δεν πληρούν τα κριτήρια για μείζονα κατάθλιψη, Κρίση (2016) 161–165.

[4] MK Nock, G. Borges, EJ Bromet, J. Alonso, M. Angermeyer, A. Beautrais, R. Bruffaerts, WT Chiu, G. De Girolamo, S. Gluzman, et αϊ. , Διακρατικός επιπολασμός και παράγοντες κινδύνου για αυτοκτονικό ιδεασμό, σχέδια και προσπάθειες, The British Journal of Psychiatry 192 (2) (2008) 98–105.

[5] MJ Vioules, B. Moulahi, J. Az` e, S. Bringay, Εντοπισμός αναρτήσεων που σχετίζονται με αυτοκτονία σε ροές δεδομένων twitter, I BM Journal of Research and Development 62 (1) (2018) 1–21. doi:10.1147/JRD. 2017.2768678

[6] AJ Ferrari, RE Norman, G. Freedman, AJ Baxter, JE Pirkis, MG Harris, A. Page, E. Carnahan, L. Degenhardt, T. Vos, et al., Το βάρος που οφείλεται στις διανοητικές και χρήσης ουσιών διαταραχές ως παράγοντες κινδύνου για αυτοκτονία: ευρήματα από την παγκόσμια μελέτη της ασθένειας 2010, PloS one 9 (4) (2014) e91936.

[7] RC O'Connor, MK Nock, Η ψυχολογία της αυτοκτονικής συμπεριφοράς, The Lancet Psychiatry 1 (1) (2014) 73-85.

[8] J. Lopez-Castroman, B. Moulahi, J. Aze, S. Bringay, J. Deninotti, S. Guillaum e, E. Baca-´ Garcia, εξόρυξη δεδομένων από κοινωνικά δίκτυα για τη βελτίωση της πρόληψης αυτοκτονιών: Μια κριτική, περιοδικό της έρευνας νευροεπιστήμης (2019) 1–10.

[9] V. Venek, S. Scherer, L.-P. Morency, J. Pestian, et αϊ., Αξιολόγηση αυτοκτονικού κινδύνου για εφήβους στην κλινική ενδοεπικοινωνία -ασθενούς , Συναλλαγές IEEE στο Affective Computing 8 (2) (2017) 204–215.

[10] D. Delgado-Gomez, H. Blasco-Fontecilla, AA Alegria, T. Legido-Gil, A. Artes-Rodriguez, E. Baca-Garcia, Βελτίωση της ακρίβειας της κατάταξης αυτοκτονίας, Τεχνητή νοημοσύνη στην ιατρική 52 (3) (2011) 165–168.

[11] G. Liu, C. Wang, K. Peng, H. Hu ang, Y. Li, W. Cheng, SocInf: Επιθέσεις συμπερασμάτων μελών σε δεδομένα υγείας των κοινωνικών μέσων με μηχανική μάθηση, Συναλλαγές IEEE σε Υπολογιστικά Κοινωνικά Συστήματα (2019) 907 - 921.

[12] B. O'Dea, S. Wan, PJ Batterham, AL Calear, C. Paris, H. Christensen, Εντοπίζοντας αυτοκτονία στο twitter, Παρεμβάσεις στο Διαδίκτυο 2 (2) (2015) 183–188.

[13] H.-C. Shing, S. Nair, A. Zirikly, M. Friedenberg, H. Daume III, P. Resnik, Ειδικοί, πλήθος πόρων και μηχανική εκτίμηση του κινδύνου αυτοκτονίας μέσω διαδικτυακών

δημοσιεύσεων, σε: Πρακτικά του Πέμπτου Εργαστηρίου Υπολογιστικής Γλωσσολογίας και Κλινικής Ψυχολογία: Από το πληκτρολόγιο στην κλινική, 2018, σελ. 25–36.

[14] F. Ren, X. Kang, C. Quan, Εξέταση συσσωρευμένων συναισθηματικών χαρακτηριστικών σε ιστολόγια αυτοκτονίας με ένα μοντέλο θεμάτων συναισθημάτων, IEEE Journal of Biomedical and He alth Informatics 20 (5) (2016) 1384–1396.

[15] L. Yue, W. Chen, X. Li, W. Zuo, M. Yin, Μια έρευνα ανάλυσης συναισθημάτων στα μέσα κοινωνικής δικτύωσης, τα συστήματα γνώσης και πληροφοριών (2018) 1–47.

[16] A. Benton, M. Mitchell, D. Hovy, Εκμάθηση πολλαπλών εργασιών για την ψυχική υγεία με χρήση τόσο κειμένου μέσων, arXiv preprint arXiv: 1712.03538 (2017).

[17] S. Ji, CP Yu, S.-f. Fung, S. Pan, G. Long, Εποπτευόμενη μάθηση για ανίχνευση αυτοκτονικών ιδεών σε περιεχόμενο χρήστη στο διαδίκτυο, Πολυπλοκότητα 2018 (2018) 1–11.

[18] S. Ji, G. Long, S. Pan, T. Zhu, J. Jiang, S. Wan g, Ανίχνευση αυτοκτονικού ιδεασμού με προστασία δεδομένων σε διαδικτυακές κοινότητες, σε: Διεθνές Συνέδριο για Συστήματα Βάσεων Δεδομένων για Προηγμένες Εφαρμογές, Springer , Cham, 2019, σελ. 225–229.

[19] J. Tighe, F. Shand, R. Ridani, A. Mackinnon, N. De La Mata, H. Christe nsen, Ibobbly κινητή παρέμβαση για την πρόληψη αυτοκτονιών σε αυστραλιανούς αυτόχθονες νέους: μια πιλοτική τυχαιοποιημένη ελεγχόμενη δοκιμή, ανοιχτή BMJ 7 (1) (2017) e013518.

[20] NNG de Andrade, D. Pawson, D. Muriello, L. Donahue, J. Guadagno, Ethics and τεχνητή νοημοσύνη: πρόληψη αυτοκτονιών στο facebook, Philosophy & Technology 31 (4) (2018) 669–684.

[21] LC McKernan, EW Clayton, CG Walsh, Προστασία της ζωής διατηρώντας την ελευθερία: Ηθικές συστάσεις για την πρόληψη αυτοκτονιών με τεχνητή νοημοσύνη, Σύνορα στην ψυχιατρική 9 (2018) 650.

[22] KP Linthicum, KM Schafer, JD Ribeiro, Μηχανική μάθηση στην επιστήμη αυτοκτονίας: Εφαρμογές και ηθική, Επιστημονικές συμπεριφορές και νόμος 37 (3) (2019) 214–222.

[23] S. Scherer, J. Pestian, L.-P. Morency, Διερεύνηση των χαρακτηριστικών της ομιλίας των αυτοκτονικών εφήβων, το: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2013, pp. 709-713.

[24] D. Sikander, M. Arvaneh, F. Amico, G. Healy, T. Ward, D. Kearney, E. Mohedano, J. Fagan, J. Yek, AF Smeaton, et al., Προβλέποντας τον κίνδυνο αυτοκτονίας χρησιμοποιώντας καρδιακό ρυθμό ηρεμίας, σε: Signal and Annual Processing Association and Summit and Conference (APSIPA), 2016 Asia-Pacific, IEEE, 2016, σελ. 1-4.

[25] Οι MA Just, L. Pan, VL Cherkassky, DL McMakin, C. Cha, MK Nock, D. Brent, H μηχανική μάθηση νευρικών αναπαραστάσεων αυτοκτονιών και συναισθημάτων εννοεί τη νεανική αυτοκτονία, την ανθρώπινη συμπεριφορά της φύσης 1 (12) (2017 911.

[26] N. Jiang, Y. Wang, L. Sun, Y. Song, H. Sun, Μια μελέτη erp για τη σιωπηρή επεξεργασία συναισθημάτων σε καταθλιπτικά άτομα που αποπειράθηκαν αυτοκτονία, σε: Πληροφορική στην Ιατρική και την Εκπαίδευση (ITME), 2015 7η International Conference on, IEEE, 2015, σελ. 37-40.

[27] M. Lotito, E. Cook, Μια ανασκόπηση των οργάνων και προσεγγίσεων εκτίμησης κινδύνου αυτοκτονίας, Κλινικός ψυχικής υγείας 5 (5) (2015) 216–223.

[28] Z. Tan, X. Liu, X. Liu, Q. Cheng, T. Zhu, Σχεδιασμός άμεσων μηνυμάτων microblog για να προσελκύσουν χρήστες κοινωνικών μέσων με ιδέες αυτοκτονίας: συνέντευξη και μελέτη έρευνας στο weibo, Journal of medical Internet research 19 (12) (2017) e381.

[29] Y.Π. Huang, T. Goh, CL Liew, Κυνηγώντας σημειώσεις αυτοκτονίας στο web 2.0-προκαταρκτικά ευρήματα , σε: Εργαστήρια πολυμέσων, 2007. ISMW'07. Ένατο Διεθνές Συμπόσιο IEEE, IEEE, 2007, σελ. 517–521.

[30] KD Varathan, N. Talib, Σύστημα ανίχνευσης αυτοκτονιών που βασίζεται στο twitter, στο: Science and Information Conference (SAI), 2014, IEEE, 2014, σελ. 785-788.

[31] J. Jashinsky, S. Burton, C. Hanson, J. West, C. Giraud-Carrier, M. Barnes, T. Argyle, Παρακολούθηση παραγόντων κινδύνου αυτοκτονίας μέσω twitter στις ΗΠΑ, Crisis: The Journal of Crisis Intervention και πρόληψη αυτοκτονιών 35 (1) (2014) 51-59.

[32] JF Gunn, D. Lester, Twiterter postings και αυτοκτονία: Μια ανάλυση των αναρτήσεων μιας θανατηφόρας αυτοκτονίας στις 24 ώρες πριν από το θάνατο, Suicidologi 17 (3) (2015) 28–30.

[33] G. Coppersmith, R. Leary, E. Whyne, T. Wood, Ποσοτικός αυτοκτονικός ιδεασμός μέσω της χρήσης γλώσσας στα μέσα κοινωνικής δικτύωσης, σε: Jo int Statistics Meetings Proceedings, Statistic Computing Section, JSM, 2015, pp. 1–15 .

[34] GB Colombo, P. Burnap, A. Hodorog, J. Scourfield, Ανάλυση της συνδεσιμότητας και της επικοινωνίας των αυτοκτονικών χρηστών στο twitter, Επικοινωνίες υπολογιστών 73 (2016) 291–300.

[35] G. Coppersmith, K. Ngo, R. Leary, A. Wood, Διερευνητική ανάλυση των κοινωνικών μέσων πριν από μια απόπειρα αυτοκτονίας, σε: Πρακτικά του Τρίτου Εργαστηρίου Υπολογιστικής Γλωσσολογίας και Κλινικής Ψυχολογίας, 2016, σελ. 106– 117.

[36] P. Solano, M. Ustulin, E. Pizzorn o, M. Vichi, M. Pompili, G. Serafini, M. Amore, Μια προσέγγιση βάσει Google για την παρακολούθηση του κινδύνου αυτοκτονίας, Ψυχιατρική έρευνα 246 (2016) 581–586.

[37] ME Larsen, N. Cummins, TW Boonstra, B. O'Dea, J. Tighe, J. Nicholas, F. Shand, J. Epps, H. Christensen, Η χρήση της τεχνολογίας στην πρόληψη αυτοκτονιών, σε: Μηχανική στην Ιατρική και Βιολογία Εταιρεία (EMBC), 2015 37ο Ετήσιο Διεθνές Συνέδριο του IEEE, IEEE, 2015, σελ. 7316–7319.

[38] HY Huang, M. Bashir, διαδικτυακή κοινότητα και πρόληψη αυτοκτονιών: Διερεύνηση των γλωσσικών στοιχείων και μεροληψία απάντησης, σε: CHI'16, 2016.

[39] M. De Choudhury, E. Kiciman, Η γλώσσα της κοινωνικής υποστήριξης στα μέσα κοινωνικής δικτύωσης και η επίδρασή της στον κίνδυνο αυτοκτονικού ιδεασμού, στο: Ενδέκατο Διεθνές Συνέδριο AAAI για τα Διαδικτυακά και Κοινωνικά Μέσα, AAAI, 2017, σελ. 1-10 .

[40] N. Masuda, I. Kurahashi, H. Onari, Αυτοματισμός ιδεών ατόμων σε διαδικτυακά κοινωνικά δίκτυα, PloS one 8 (4) (2013).

[41] S. Chattopadhyay, Μια μελέτη για την ανάλυση κινδύνου αυτοκτονίας, σε: Ηλεκτρονική δικτύωση, εφαρμογή και υπηρεσίες, 9ο Διεθνές Συνέδριο 2007, IEEE, 2007, σελ. 74–78.

[42] D. Delgado-Gomez, H. Blasco-Fontecilla, F. Sukno, MS Ramos-Plasencia, E. Baca-Garcia, Ταξινόμηση απόπειρων αυτοκτονίας : Προς προγνωστικά μοντέλα συμπεριφοράς αυτοκτονίας, Neurocomputing 92 (2012) 3–8.

[43] S. Chattopadhyay, Ένα μαθηματικό μοντέλο εκτίμησης αυτοκτονικών προθέσεων σε ενήλικες, American Journal of Biomedical Engineering 2 (6) (2012) 251–262.

[44] W. Wang, L. Chen, M. Tan, S. Wang, AP Sheth, Ανακαλύπτοντας ένα λεπτό συναίσθημα σε σημειώσεις αυτοκτονίας, Βιοϊατρικές πληροφορίες πληροφορικής 5 (Συμπλήρωμα 1) (2012) 137.

[45] A. Abboute, Y. Boudjeriou, G. Entringer, J. Aze, S. Bringay, P. Poncelet, Mining twitter για «πρόληψη αυτοκτονιών, σε: Διεθνές Συνέδριο για τις Εφαρμογές της Φυσικής Γλώσσας σε Βάσεις Δεδομένων / Πληροφοριακά Συστήματα, Springer, 2014, σελ. 250–253.

[46] E. Okhapkina, V. Okhapkin, O. Kazarin, Προσαρμογή μεθόδων ανάκτησης πληροφοριών για τον εντοπισμό καταστροφικών πληροφοριακών πληροφοριών σε κοινωνικά δίκτυα, σε: Προηγμένα εργαστήρια δικτύωσης πληροφοριών και εφαρμογών (WAINA), 31ο διεθνές συνέδριο 2017 στο , 2017, σελ. 87–92.

[47] M. Mulholland, J. Quinn, τάσεις αυτοκτονίας: Η αυτόματη ταξινόμηση των αυτοκτονικών και μη- αυτοκτονικών στιχουργών χρησιμοποιώντας nlp., Στο: IJCNLP, 2013, σελ. 680–684.

[48] X. Huang, L. Zhang, D. Chiu, T. Liu, X. Li, T. Zhu, Εντοπισμός αυτοκτονικού ιδεασμού σε κινεζικά microblogs με ψυχολογικά λεξικά, σε: Ubiquitous Intelligence and Computing, 2014 IEEE 11th Intl Conf on και IEEE 11th Intl Conf on and Autonomic and Trusted Computing, και IEEE 14th Intl Conf on Scalable Computing and Communications and the Associated Workshops (UTC-ATC-ScalCom), IEEE, 2014, σελ. 844-84

[49] X. Huang, X. Li, T. Liu, D. Chiu, T. Zhu, L. Zhang, Τοπικό μοντέλο για τον εντοπισμό αυτοκτονικού ιδεασμού στο κινεζικό microblog, σε: Πρακτικά του 29ου Συνεδρίου Ειρηνικού Ασίας για τη Γλώσσα, τις Πληροφορίες and Computation, 2015, σελ. 553–562.

[50] Y.-M. Tai, H.-W. Chiu, Τεχνητή ανάλυση νευρικού δικτύου σχετικά με την αυτοκτονία και το ιστορικό αυτοτραυματισμών των ταϊβανέζων στρατιωτών, σε: Καινοτόμος

Υπολογισμός, Πληροφορίες και Έλεγχος, 2007. ICICIC'07. Δεύτερο Διεθνές Συνέδριο, IEEE, 2007, σελ. 363–363.

[51] M Liakata, J . H. Kim, S. Saha, J. Hastings, D. Rebholzschuhmann, Τρεις υβριδικοί ταξινομητές για την ανίχνευση συναισθημάτων σε σημειώσεις αυτοκτονίας, Biomedical Informatics Insights, 2012, Suppl. 1 (2012-01-30) 2012 ((Συμπ. 1)) (2012) 175–184.

[52] J. Pestian, H. Nasrallah, P. Matykiewi cz, A. Bennett, A. Leenaars, Ταξινόμηση σημειώσεων αυτοκτονίας με χρήση επεξεργασίας φυσικής γλώσσας: Ανάλυση περιεχομένου, Βιοϊατρικές πληροφορίες πληροφορικής 2010 (3) (2010) 19.

[53] SR Braithwaite, C. Giraud-Carrier, J. West, MD Barnes, CL Hanson, Επικύρωση αλγορίθμων μηχανικής μάθησης για δεδομένα twitter ενάντια σε καθιερωμένα μέτρα αυτοκτονίας, ψυχική υγεία JMIR 3 (2) (2016) e21.

[54] T. Mikolov, K. Chen, G. Corrado, J. Dean, Αποτελεσματική εκτίμηση των αναπαραστάσεων λέξεων σε διανυσματικό χώρο, arXiv preprint arXiv: 1301.3781 (2013).

[55] J. Pennington, R. Socher, C. Manning, Glove: Παγκόσμιοι φορείς για την αναπαράσταση λέξεων, σε: Πρακτικά του συνεδρίου του 2014 για εμπειρικές μεθόδους στην επεξεργασία φυσικής γλώσσας (EMNLP), 2014, σελ. 1532–1543.

[56] S. Ji, G. Long, S. Pan, T. Zhu, J. Jiang, S. Wang, X. Li, μεταφορά γνώσης μέσω συλλογής μοντέλων για διαδικτυακή κοινωνική φροντίδα, arXiv preprint arXiv: 1905.07665 (2019).

[57] M. Gaur, A. Alambo, JP Sain, U. Kursuncu, K. Thirunarayan, R. Kavuluru, A. Sheth, R. Welton, J. Pathak, Εκτίμηση γνώσης της σοβαρότητας του κινδύνου αυτοκτονίας για έγκαιρη επέμβαση , στο: The World Wide Web Conference, ACM, 2019, σελ. 514–525.

[58] G. Coppersm ith, R. Leary, P. Crutchley, A. Fine, Επεξεργασία φυσικών γλωσσών των κοινωνικών μέσων ως έλεγχος για τον κίνδυνο αυτοκτονίας, Biomedical Informatics Insights 10 (2018) 1–11.

[59] R. Sawhney, P. Manchanda, P. Mathur, R. Shah, R. Singh, Εξερεύνηση και εκμάθηση αυτοκτονικών ιδεοτήτων σε κοινωνικά μέσα με βαθιά μάθηση, σε: Πρακτικά του 9ου Εργαστηρίου σχετικά με τις υπολογιστικές προσεγγίσεις στην υποκειμενικότητα, το συναίσθημα και την ανάλυση κοινωνικών μέσων, 2018, σελ. 167–175.

[60] S. Ji, X. Li, Z. Huang, E. Cambria, αυτοκτονικός ιδεασμός και διανοητική διαταραχή ανίχνευσης με προσεκτικά δίκτυα σχέσεων, arXiv preprint arXiv: 2004.07601 (2020).

[61] A. Zirikly, P. Resnik, O. Uzuner, K. Hollingshead, Clpsych 2019 κοινό έργο: Πρόβλεψη του βαθμού κινδύνου αυτοκτονίας σε θέσεις reddit, σε: Πρακτικά του έκτου

εργαστηρίου για την Υπολογιστική Γλωσσολογία και Κλινική Ψυχολογία, 2019, σελ. 24–33.

[62] AG Hevia, R. C. Menendez, D. Gayo-Avello, Ανάλυση της χρήσης των υπαρχόντων συστημάτων για την κοινή εργασία «clpsych 2019», σε: Πρακτικά του 6ου Εργαστηρίου Υπολογιστικής Γλωσσολογίας και Κλινικής Ψυχολογίας, 2019, σελ. 148–151.

[63] M. Morales, P. Dey, T. Theisen, D. Belitz, N. Chernova, Μια διερεύνηση συστημάτων βαθιάς μάθησης για την εκτίμηση κινδύνου αυτοκτονίας, σε: Πρακτικά του 6ου Εργαστηρίου Υπολογιστικής Γλωσσολογίας και Κλινικής Ψυχολογίας, 2019, σελ. 177–181.

[64] M. Matero, A. Idnani, Y. Son, S. Giorgi, H. Vu, M. Zamani, P . Limbachiya, SC Guntuku, HA Schwartz, Εκτίμηση κινδύνου αυτοκτονίας με γλώσσα διπλού πλαισίου πολλαπλών επιπέδων και bert, σε: Πρακτικά του Έκτου Εργαστηρίου Υπολογιστικής Γλωσσολογίας και Κλινικής Ψυχολογίας, 2019, σελ. 39–44.

[65] L. Chen, A. Aldayel, N. Bogoy chev, T. Gong, Παρόμοια μυαλά αναρτώνται: Εκτίμηση του κινδύνου αυτοκτονίας χρησιμοποιώντας ένα υβριδικό μοντέλο, σε: Πρακτικά του 6ου εργαστηρίου για την Υπολογιστική Γλωσσολογία και την Κλινική Ψυχολογία, 2019, σελ. 152–157.

[66] X. Zhao, S. Lin, Z. Huang, Ταξινόμηση κειμένου της τρύπας δέντρων micro -blog με βάση το συνελικτικό νευρωνικό δίκτυο, σε: Πρακτικά του Διεθνούς Συνεδρίου για τους Αλγόριθμους, την Πληροφορική και την Τεχνητή Νοημοσύνη του 2018, ACM, 2018, σ. . 61.

[67] T. Tran, D. Phung, W. Luo, R. Harvey, M. Berk, S. Venkatesh, Ένα ολοκληρωμένο πλαίσιο για την πρόβλεψη κινδύνου αυτοκτονίας, σε: Πρακτικά του 19ου διεθνούς συνεδρίου ACM SIGKDD για την ανακάλυψη γνώσεων και τα δεδομένα εξόρυξη, ACM, 2013, σελ. 1410–1418.

[68] S. Berrouiguet, R. Billot, P. Lenca, P. Tanguy, E. Baca-Garcia, M. Simonnet, B. Gourve nnec, Προς εφαρμογές ηλεκτρονικής υγείας για την πρόληψη αυτοκτονιών, το 2016 IEEE First International Conference on Connected Health: Εφαρμογές, Συστήματα και Τεχνολογίες Μηχανικών (CHASE), IEEE, 2016, σελ. 346–347.

[69] D. Meyer, J.-A. Abbott, I. Rehm, S. Bhar, A. B arak, G. Deng, K. Wallace, E. Ogden, B. Klein, Ανάπτυξη εργαλείου ανίχνευσης αυτοκτονικών ιδεών για ρυθμίσεις πρωτοβάθμιας υγειονομικής περίθαλψης: χρήση διαδικτυακών ψυχοκοινωνικών δεδομένων ανοιχτής πρόσβασης, Τηλεϊατρική και ηλεκτρονική υγεία 23 (4) (2017) 273–281.

[70] KM Harris, JP McLean, J. Sh effield, Suicidal and online: Πώς μας ενημερώνουν οι διαδικτυακές συμπεριφορές για αυτόν τον πληθυσμό υψηλού κινδύνου ; Μελέτες θανάτου 38 (6) (2014) 387–394.

[71] H. Sueki, Ο συσχετισμός της χρήσης twitter που σχετίζεται με την αυτοκτονία με αυτοκτονική συμπεριφορά: μια εγκάρσια μελέτη νεαρών χρηστών του Διαδικτύου στην Ιαπωνία, Journal of συναισθηματικές διαταραχές 170 (2015) 155-160.

[72] KW Hammond, RJ Laundry, TM OLeary, WP Jones, Χρήση αναζήτησης κειμένου για τον αποτελεσματικό προσδιορισμό του επιπολασμού των προσπαθειών αυτοκτονίας κατά τη διάρκεια ζωής μεταξύ βετεράνων, το 2013 46ο Hawaii Internationa l Conference on System Sciences, IEEE, 2013, pp. 2676– 2683.

[73] CG Walsh, JD Ribeiro, JC Franklin, Πρόβλεψη κινδύνου απόπειρων αυτοκτονίας με την πάροδο του χρόνου μέσω της μηχανικής μάθησης, Clinical Psychological Science 5 (3) (2017) 457–469.

[74] T. Iliou, G. Konstantopoulou, M. Ntekouli, D. Lymberopoulos, K. Assimakopoulos, D. Galiatsatos, G. Anastassopoulos, Machine learning μέθοδος προεπεξεργασίας για την πρόβλεψη αυτοκτονίας, σε: L. Iliadis, I. Maglogiannis (Eds.), Artificial Intelligence Applications and Innovations, Springer In ternational Publishing, Cham, 2016, σελ. 53-60.

[75] T. Nguyen, T. Tran, S. Gopakumar, D. Phung, S. Venkatesh, Μια αξιολόγηση τυχαιοποιημένων μεθόδων μηχανικής μάθησης για περιττά δεδομένα: Πρόβλεψη βραχυπρόθεσμου και μεσοπρόθεσμου κινδύνου αυτοκτονίας από διοικητικά αρχεία και εκτιμήσεις κινδύνου, arXiv preprint arXiv: 1605.011 16 (2016).

[76] HS Bhat, SJ Goldman-Mellor, Πρόβλεψη απόπειρών αυτοκτονίας εφήβων με νετρωνικά δίκτυα, στο: Εργαστήριο NIPS 2017 για Μηχανική Εκμάθηση για την Υγεία, 2017, σελ. 1-8.

[77] JP Pestian, P Matykiewicz, M Linn-Ριπή, B South, O. Uzuner, J. Wiebe, K B . Cohen, J. Hurdle, C. Brew, Ανάλυση συναισθημάτων σημειώσεων αυτοκτονίας: Μια κοινή εργασία, Biomedical informatics insights 5 (Συμπλήρωμα 1) (2012) 3.

[78] E. White, LJ Mazlack, Discerning suicide notes causality using fuzzy cognitive maps, in: Fuzzy Systems: Fuzzy Systems (FUZZ), 2011 I EEE International Conference on, IEEE, 2011, pp. 2940–2947.

[79] B. Desmet, V. Hoste, Ανίχνευση συναισθημάτων σε σημειώσεις αυτοκτονίας, Expert Systems with Applications 40 (16) (2013) 6351–6358.

[80] R. Wicentowski, MR Sydes, Ανίχνευση συναισθημάτων σε σημειώσεις αυτοκτονίας χρησιμοποιώντας μέγιστη εντροπία ταξινόμησης, Biomedical informatics insights 5 (2012) BII – S8972.

[81] A. Konecˆný, A. Dehghan, JA Keane, G. Nenadic, Κατηγοριοποίηση θέματος των δηλώσεων σε σημειώσεις αυτοκτονίας με ολοκληρωμένους κανόνες και μηχανική μάθηση, Biomedical informatics insights 5 (2012) BII – S8978.

[82] AM Schoene, N. Dethlefs, Αυτόματη ταυτοποίηση σημειώσεων αυτοκτονίας από γλωσσικά και συναισθηματικά χαρακτηριστικά, στο: Πρακτικά του 10ου εργαστηρίου SIGHUM για τη γλωσσική τεχνολογία για την πολιτιστική κληρονομιά, τις κοινωνικές επιστήμες και τις ανθρωπιστικές επιστήμες, 201 6, σ. 128–133.

[83] J. Robinson, G. Cox, E. Bailey, S. Hetrick, M. Rodrigues, S. Fisher, H. Herrman, Μέσα κοινωνικής δικτύωσης και πρόληψη αυτοκτονιών: μια συστηματική ανασκόπηση, Έγκαιρη παρέμβαση στην ψυχιατρική 10 (2) (2016) 103–121.

[84] Y. Wang, S. Wan, C. Paris, Ο ρόλος των χαρακτηριστικών και του πλαισίου στην ανίχνευση ιδεών αυτοκτονίας, σε: Πρακτικά του εργαστηρίου της Australasian Language Technology Workshop 2016, 2016, σελ. 94–102.

[85] A. Shepherd, C. Sanders, M. Doyle, J. Shaw, Χρήση κοινωνικών μέσων για υποστήριξη και ανατροφοδότηση από χρήστες υπηρεσιών ψυχικής υγείας : θεματική ανάλυση μιας συνομιλίας στο twitter, BMC ψυχιατρική 15 (1) (2015) 29.

[86] M. De Choudhury, S. De, Ομιλία για την ψυχική υγεία στο reddit: Αυτο-αποκάλυψη, κοινωνική υποστήριξη και ανωνυμία., Στο: ICWSM, 2014, σελ. 1–10.

[87] M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, M. Kumar, Ανακαλύπτοντας μετατοπίσεις στον αυτοκτονικό ιδεασμό από το περιεχόμενο της ψυχικής υγείας στα κοινωνικά μέσα, στο: CHI, ACM, 2016, σελ. 2098–2110.

[88] M. Kumar, M. Dredze, G. Coppersmith, M. De Choudhury, Εντοπισμός αλλαγών στο περιεχόμενο αυτοκτονίας που εκδηλώθηκε στα μέσα κοινωνικής δικτύωσης μετά από αυτοκτονίες διασημοτήτων, σε: Πρακτικά του 26ου Συνεδρίου ACM για υπερκειμένα και κοινωνικά μέσα, ACM, 2015, σελ. 85–94.

[89] L. Guan, B. Hao, Q. Cheng, PS Yip, T. Zhu, Προσδιορισμός χρηστών Κινέζων microblogs με υψηλή πιθανότητα αυτοκτονίας χρησιμοποιώντας προφίλ που βασίζονται στο Διαδίκτυο και γλωσσικά χαρακτηριστικά: μοντέλο ταξινόμησης, ψυχική υγεία JMIR 2 (2) (2015) e17.

[90] SJ Cash, M. Thelwall, SN Peck, JZ Ferrell, JA Bridge, Εφηβικές δηλώσεις αυτοκτονίας στο myspace, Cyberpsychology, Behavior, and Social Networking 16 (3) (2013) 166-174.

[91] Ποσοστά αυτοκτονιών, δεδομένα του Παγκόσμιου Παρατηρητηρίου Υγείας (GHO), διαθέσιμα στη διεύθυνση: http://www.who.int/gho/mental_health/self_rates/el/ (2015).

[92] M. Mohri, A. Rostamizadeh, A. Talwalkar, Foundations of machine learning, MIT press, 2012.

[93] C. Cortes, V. Vapnik, Support vector machine, Machine learning 20 (3) (1995) 273–297.

[94] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.

[95] JH Friedman, Greedy function approximation: a gradient boosting machine, Annals of statistics (2 001) 1189–1232.

[96] T. Chen, C. Guestrin, Xgboost: Ένα κλιμακούμενο σύστημα ενίσχυσης δέντρων, στο: Πρακτικά του 22ου διεθνούς συνεδρίου για την ανακάλυψη της γνώσης και την εξόρυξη δεδομένων, ACM, 2016, σελ. 785–794.

[97] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (8) (1997) 1735–1780.

[98] J. W. Pennebaker, R. L. Boyd, K. Jordan, K. Blackburn, The development and psychometric properties of LIWC2015, Tech. rep. (2015).

[99] DM Blei, AY Ng, MI Jordan, Latent dirichlet alokation, Journal of machine Learning research 3 (Jan) (2003) 993–1022.

[100] A. Voutilainen, Part-of-speech tagging, The Oxford handbook of computational linguistics(2003) 219–232.

[101] T. Mikolov, K. Chen, G. Corrado, J. Dean, Αποτελεσματική εκτίμηση των αναπαραστάσεων λέξεων στο διανυσματικό χώρο, arXiv preprint arXiv: 1301.3781 (2013).

[102] I. T. Jolliffe, Principal component analysis and factor analysis, in: Principal component analysis, Springer, 1986, pp. 115–128.

[103] K. Jacob, V. Patel, Ταξινόμηση ψυχικών διαταραχών: μια παγκόσμια προοπτική ψυχικής υγείας, The Lancet 383 (9926) (2014) 1433–1435.

[104] I. Gilat, Y. Tobin, G. Shahar, Προσφέροντας υποστήριξη σε αυτοκτονικά άτομα σε μια online ομάδα υποστήριξης, Archives of Suicide Research 15 (3) (2011) 195–206.

[105] Y. Kim, Convolutional νευρωνικά δίκτυα για ταξινόμηση προτάσεων, στο: Πρακτικά του Συνεδρίου του 2014 ο Εμπειρικές Μέθοδοι στην Επεξεργασία Φυσικής Γλώσσας (EMNLP), 2014, σελ. 1746–1751.

[106] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (8) (1997) 1735–1780.

[107] S. Lai, L. Xu, K. Liu, J. Zhao, Επαναλαμβανόμενα συνελικτικά νευρωνικά δίκτυα για ταξινόμηση κειμένου., Στο: AAAI, Vol. 333, 2015, σελ. 2267–2273.

[108] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, arXiv preprint arXiv: 1409.0473 (2014).

[109] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, E. Hovy, Hierarchical attention networks for document classification, in: NAACL, 2016, pp. 1480–1489.

[110] Z. Lin, M. Feng, C. N. d. Santos, M. Yu, B. Xiang, B. Zhou, Y. Bengio, A structured selfattentive sentence embedding, arXiv preprint arXiv:1703.03130 (2017).

[111] D. Raposo, A. Santoro, D. Barrett, R. Pascanu, T. Lillicrap, P. Battaglia, Discovering objects and their relations from entangled scene representations, arXiv preprint arXiv:1702.05068 (2017).

[112] A. Santoro, D. Raposo, DG Barrett, M. Malinowski, R. Pascanu, P. Battaglia, T. Lillicrap, A simple neuronal network module for relational reasoning, στο: NIPS, 2017, σελ. 4967–4976.

[113] R. Socher, D. Chen, C. D. Manning, A. Ng, Reasoning with neural tensor networks for knowledge base completion, in: NIPS, 2013, pp. 926–934.

[114] W. L. Hamilton, K. Clark, J. Leskovec, D. Jurafsky, Inducing-domain-specific sentiment lexicons from unlabelled corpora, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Vol. 2016, 2016, σελ. 595.

[115] DM Blei, AY Ng, M. I . Jordan, Latent dirichlet alokation, Journal of Machine Learning Research 3 (Jan) (2003) 993–1022.

[116] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[117] G. Coppersmith, M. Dredze, C. Harman, K. Hollingshe ad, M. Mitchell, Clpsych 2015 shared task: Depression and ptsd on twitter, in: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 2015, σελ. 31–39.

[118] A. Joulin, E. Grave, P. Bojanowski, T. Mikolov, Bag of tricks for efficient text classification, arXiv preprint arXiv:1607.01759 (2016).

Παράρτημα Α: Ο κώδικας της εφαρμογής

Έξι μοντέλα εφαρμόστηκαν, συγκεκριμένα αυτά είναι τα εξής: λογιστική παλινδρόμηση(logistic regression), τυχαίο δάσος(random forest), δέντρο απόφασης βαθμιαίας κλίσης(gradient boosting decision tree), xgboost, μηχανή διανύσματος υποστήριξης(support vector machine) και δίκτυα LSTM.

Τα πέντε πρώτα παραπάνω μοντέλα στο Reddit και στο Twitter υλοποιήθηκαν από τα αρχεία python clf.py and python clf_reddit.py. Το μοντέλο LSTM για το Reddit και το Twitter από τα αρχεία python lstm.py και python lstm_reddit.py. python lstm_word2vec.py και python lstm_word2vec_reddit.py.

Τα σενάρια είναι γραμμένα σε Python 3.9(64-bit).

Αρχείο clf.py

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
# Python version: 3.9
import string
import nltk
import numpy as np
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer
from sklearn.model_selection import KFold
from gensim import corpora
from gensim.models.ldamodel import LdaModel
from nltk.stem.wordnet import WordNetLemmatizer
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
import xgboost as xgb
from tabulate import tabulate
from options import arg_clf
from helpers import load_df, evaluate_prediction


def f_basic(data):
    print("Processing basic features ...")
    num_title_words, num_title_token, num_title_char = [], [], []
    for title in data['tweets']:
        num_title_words.append(len(title.split()))
        tokens = nltk.word_tokenize(title)
        num_title_token.append(len(tokens))
        num_title_char.append(len(title))
    features = {'title_words': num_title_words, 'title_token': num_title_token, 'title_char': num_title_char,}
    return pd.DataFrame(features, columns=['title_words', 'title_token', 'title_char'])


def f_liwc(dataset_name):
    # extracted features using LIWC
    print("Processing LIWC features ...")
    liwc_twitter = pd.read_csv('./data/liwc_features/liwc_{}.csv'.format(dataset_name))
    liwc = liwc_twitter[liwc_twitter.columns[3:]]
    return liwc


def get_all_tags(data):
    print("Processing POS features ...")
    tags_all = []
    for title in data['tweets']:
        tagged_text = nltk.pos_tag(nltk.word_tokenize(title))
        for word, tag in tagged_text:
            if tag not in tags_all:
                tags_all.append(tag)
    return tags_all


def f_pos(data, tags_all):
    tag_dict, tag_count = {}, {}
    for tag in tags_all:
        tag_dict[tag] = 0
        tag_count[tag] = []
    for title in data['tweets']:
        tagged_text = nltk.pos_tag(nltk.word_tokenize(title))
        for word, tag in tagged_text:
            tag_dict[tag]+=1
        for count,tag in zip(tag_dict.values(), tag_dict.keys()):
            tag_count[tag].append(count)
```

```
    return pd.DataFrame(tag_count, index=None)


def f_tfidf(data):
    print("Processing TF-IDF features ...")
    X = data['tweets']
    count_vect = CountVectorizer(stop_words='english', ngram_range=(1, 1), max_features=50)
    X_counts = count_vect.fit_transform(X)
    tfidf_transformer = TfidfTransformer()
    X_train_tfidf = tfidf_transformer.fit_transform(X_counts)
    df_tfidf = pd.DataFrame(X_train_tfidf.todense())
    return df_tfidf


def f_topics(data, topic_num):
    def cleaning(article):
        punctuation = set(string.punctuation)
        lemmatize = WordNetLemmatizer()
        one = " ".join([i for i in article.lower().split() if i not in stopwords])
        two = "".join(i for i in one if i not in punctuation)
        three = " ".join(lemmatize.lemmatize(i) for i in two.lower().split())
        return three

    def pred_new(doc):
        one = cleaning(doc).split()
        two = dictionary.doc2bow(one)
        return two

    print("Processing Topics features ...")
    stopwords = set(nltk.corpus.stopwords.words('english'))
    text = data['tweets'].map(cleaning)
    text_list = []
    for t in text:
        temp = t.split()
        text_list.append([i for i in temp if i not in stopwords])
    dictionary = corpora.Dictionary(text_list)
    doc_term_matrix = [dictionary.doc2bow(doc) for doc in text_list]
    ldamodel = LdaModel(doc_term_matrix, num_topics=topic_num, id2word = dictionary, passes=50)
    probs = []  # list of probability vectors
    for t in text:
        prob = ldamodel[(pred_new(t))]
        d = dict(prob)
        for i in range(topic_num):
            if i not in d.keys():
                d[i] = 0
        temp = []
        for i in range(topic_num):
            temp.append(d[i])
        probs.append(temp)
    return pd.DataFrame(probs, index=None)


if __name__ == '__main__':
    args = arg_clf()
    df_data = load_df(args.dataset)
    df_data.columns = ['tweets', 'y']
    if args.num_features == 1:
        df_basic = f_basic(df_data)
        df_all = pd.concat([df_basic, df_data['y']], axis=1)
    elif args.num_features == 2:
        df_basic = f_basic(df_data)
        df_tfidf = f_tfidf(df_data)
        df_features = pd.concat([df_basic, df_tfidf], axis=1)
        df_all = pd.concat([df_features, df_data['y']], axis=1)
```

```
    elif args.num_features == 3:
        df_basic = f_basic(df_data)
        df_tfidf = f_tfidf(df_data)
        tags_all = get_all_tags(df_data)
        df_pos = f_pos(df_data, tags_all)
        df_features = pd.concat([df_basic, df_tfidf, df_pos], axis=1)
        df_all = pd.concat([df_features, df_data['y']], axis=1)
    elif args.num_features == 4:
        df_basic = f_basic(df_data)
        df_tfidf = f_tfidf(df_data)
        tags_all = get_all_tags(df_data)
        df_pos = f_pos(df_data, tags_all)
        df_topic = f_topics(df_data, args.num_topics)
        df_features = pd.concat([df_basic, df_tfidf, df_pos, df_topic], axis=1)
        df_all = pd.concat([df_features, df_data['y']], axis=1)
    elif args.num_features == 5:
        df_basic = f_basic(df_data)
        df_tfidf = f_tfidf(df_data)
        tags_all = get_all_tags(df_data)
        df_pos = f_pos(df_data, tags_all)
        df_topic = f_topics(df_data, args.num_topics)
        df_liwc = f_liwc(args.dataset)
        df_features = pd.concat([df_basic, df_tfidf, df_pos, df_topic, df_liwc], axis=1)
        df_all = pd.concat([df_features, df_data['y']], axis=1)
    else:
        raise ValueError("Error: number of features groups")

    result_average, h = [], ['Model', 'Acc.', 'Pre.', 'Rec.', 'F1', 'AUC']
    lr_acc, lr_pre, lr_rec, lr_f1, lr_auc = [], [], [], [], []
    rf_acc, rf_pre, rf_rec, rf_f1, rf_auc = [], [], [], [], []
    gbdt_acc, gbdt_pre, gbdt_rec, gbdt_f1, gbdt_auc = [], [], [], [], []
    xgb_acc, xgb_pre, xgb_rec, xgb_f1, xgb_auc = [], [], [], [], []
    if args.dataset == 'Twitter' or args.dataset == 'twitter':
        num_sampling = 5
    else:
        num_sampling = 1
    for i in range(num_sampling):
        if args.dataset == 'Twitter' or args.dataset == 'twitter':
            # under sampling
            df_pos = df_all.loc[df_all['y'] == 1]
            df_neg = df_all.loc[df_all['y'] == 0]
            df_sample = pd.concat([df_pos, df_neg.sample(len(df_pos['y']))])
            df_sample = df_sample.dropna()
            X = df_sample[df_sample.columns[:-1]].as_matrix()
            y = df_sample['y'].as_matrix()
        else:
            df_all = df_all.dropna()
            X = df_all[df_all.columns[:-1]].as_matrix()
            y = df_all['y'].as_matrix()

        # 10-fold cross validation
        num_fold = 10
        kf = KFold(n_splits=num_fold, shuffle=True, random_state=0)
        for train_index, test_index in kf.split(X):
            num_fold -= 1
            X_train, X_test = X[train_index], X[test_index]
            y_train, y_test = y[train_index], y[test_index]
            # Logisitc Regression
            clf = LogisticRegression(penalty='l2', tol=1e-6)
            clf.fit(X_train, y_train)
            y_pred = clf.predict_proba(X_test)[:,1]
            acc, pre, rec, f1, auc = evaluate_prediction(y_test, y_pred, k_th=num_fold,
                                                         model_name='Logistic Regression', dataset_name=args.dataset)
            lr_acc.append(acc)
```

```
        lr_pre.append(pre)
        lr_rec.append(rec)
        lr_f1.append(f1)
        lr_auc.append(auc)
        # Random Forest
        clf = RandomForestClassifier(n_estimators=20, max_depth=8, random_state=0)
        clf.fit(X_train, y_train)
        y_pred = clf.predict_proba(X_test)[:, 1]
        acc, pre, rec, f1, auc = evaluate_prediction(y_test, y_pred, k_th=num_fold,
                                        model_name='Random Forest', dataset_name=args.dataset)
        rf_acc.append(acc)
        rf_pre.append(pre)
        rf_rec.append(rec)
        rf_f1.append(f1)
        rf_auc.append(auc)
        # GBDT
        clf = GradientBoostingClassifier(max_depth=8, random_state=0)
        clf.fit(X_train, y_train)
        y_pred = clf.predict_proba(X_test)[:, 1]
        acc, pre, rec, f1, auc = evaluate_prediction(y_test, y_pred, k_th=num_fold,
                                        model_name='GBDT', dataset_name=args.dataset)
        gbdt_acc.append(acc)
        gbdt_pre.append(pre)
        gbdt_rec.append(rec)
        gbdt_f1.append(f1)
        gbdt_auc.append(auc)
        # XGBoost
        dtrain = xgb.DMatrix(X_train, label=y_train, missing=-999)
        dtest = xgb.DMatrix(X_test, label=y_test, missing=-999)
        params = {'max_depth': 10, 'eta': 0.1, 'silent': 1, 'objective': 'binary:logistic', 'nthread': -1}
        num_round = 10000
        watchlist = [(dtrain, 'train'), (dtest, 'test')]
        model = xgb.train(params, dtrain, num_round, watchlist, early_stopping_rounds=50, verbose_eval=10)
        y_pred = model.predict(dtest)
        acc, pre, rec, f1, auc = evaluate_prediction(y_test, y_pred, k_th=num_fold,
                                        model_name='XGBoost', dataset_name=args.dataset)
        xgb_acc.append(acc)
        xgb_pre.append(pre)
        xgb_rec.append(rec)
        xgb_f1.append(f1)
        xgb_auc.append(auc)
result_average.append(['Logistic Regression', np.mean(lr_acc), np.mean(lr_pre), np.mean(lr_rec), np.mean(lr_f1), np.mean(lr_auc)])
result_average.append(['Random Forest', np.mean(rf_acc), np.mean(rf_pre), np.mean(rf_rec), np.mean(rf_f1), np.mean(rf_auc)])
result_average.append(['GBDT', np.mean(gbdt_acc), np.mean(gbdt_pre), np.mean(gbdt_rec), np.mean(gbdt_f1), np.mean(gbdt_auc)])
result_average.append(['XGB', np.mean(xgb_acc), np.mean(xgb_pre), np.mean(xgb_rec), np.mean(xgb_f1), np.mean(xgb_auc)])
print(tabulate(result_average, headers=h))
```

Αρχείο clf_reddit.py

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
# Python version: 3.9

import string
import nltk
import numpy as np
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer
from sklearn.model_selection import KFold
from gensim import corpora
from gensim.models.ldamodel import LdaModel
from nltk.stem.wordnet import WordNetLemmatizer
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
import xgboost as xgb
from tqdm import tqdm
from tabulate import tabulate
from options import arg_clf
from helpers import load_df, evaluate_prediction


def f_basic(data):
    print("Processing basic featues ...")
    num_title_words, num_title_token, num_title_char, num_title_sent = [], [], [], []
    num_body_words, num_body_token, num_body_para, num_body_sent = [], [], [], []
    for title in data['title']:
        num_title_words.append(len(title.split()))
        tokens = nltk.word_tokenize(title)
        num_title_token.append(len(tokens))
        num_title_char.append(len(title))
        sentences = nltk.tokenize.sent_tokenize(title, language='english')
        num_title_sent.append(len(sentences))
    for body in data['usertext']:
        temp_words, temp_token, temp_sent = 0, 0, 0
        for para in body:
            temp_words += len(para.split())
            temp_token += len(nltk.word_tokenize(para))
            temp_sent += len(nltk.tokenize.sent_tokenize(para, language='english'))
        num_body_words.append(temp_words)
        num_body_token.append(temp_token)
        num_body_sent.append(temp_sent)
        num_body_para.append(len(body))
    features = {'title_words': num_title_words, 'title_token': num_title_token, 'title_char': num_title_char,
                'title_sent': num_title_sent, 'body_words': num_body_words, 'body_token': num_body_token,
                'body_sent': num_body_sent, 'body_para': num_body_para}
    return pd.DataFrame(features, columns=['title_words', 'title_token', 'title_char', 'title_sent',
                                           'body_words', 'body_token', 'body_sent', 'body_para'])


def f_liwc(subreddit):
    print("Processing LIWC features ...")
    liwc_title = pd.read_csv('./data/liwc_features/liwc_{}_title.csv'.format(subreddit))
    liwc_body = pd.read_csv('./data/liwc_features/liwc_{}_body.csv'.format(subreddit))
    liwc = pd.concat((liwc_title[liwc_title.columns[4:]], liwc_body[liwc_body.columns[4:]]), axis=1)
    return liwc


def get_all_tags(data):
    print("Processing POS features ...")
    tags_all = []
    for title in data['title']:
        tagged_text = nltk.pos_tag(nltk.word_tokenize(title))
```

```
        for word, tag in tagged_text:
            if tag not in tags_all:
                tags_all.append(tag)
    for body in data['usertext']:
        for para in body:
            tagged_text = nltk.pos_tag(nltk.word_tokenize(para))
            for word, tag in tagged_text:
                if tag not in tags_all:
                    tags_all.append(tag)
    return tags_all


def f_pos(data, tags_all):
    tag_dict, tag_count, tag_count_body = {}, {}, {}
    for tag in tqdm(tags_all):
        tag_dict[tag] = 0
        tag_count[tag] = []
        tag_count_body[tag] = []
    for title in tqdm(data['title']):
        tagged_text = nltk.pos_tag(nltk.word_tokenize(title))
        for word, tag in tagged_text:
            tag_dict[tag] += 1
        for count, tag in zip(tag_dict.values(), tag_dict.keys()):
            tag_count[tag].append(count)
    for tag in tags_all:
        tag_dict[tag] = 0
    for body in tqdm(data['usertext']):
        for para in body:
            tagged_text = nltk.pos_tag(nltk.word_tokenize(para))
            for word, tag in tagged_text:
                tag_dict[tag] += 1
        for count, tag in zip(tag_dict.values(), tag_dict.keys()):
            tag_count_body[tag].append(count)
    return pd.concat((pd.DataFrame(tag_count, index=None), pd.DataFrame(tag_count_body, index=None)), axis=1)


def f_tfidf(data):
    print("Processing TF-IDF features ...")
    X = []
    for t, b in zip(data['title'], data['usertext']):
        X.append(t + ' ' + b)
    count_vect = CountVectorizer(stop_words='english', ngram_range=(1, 1), max_features=50)
    X_counts = count_vect.fit_transform(X)
    tfidf_transformer = TfidfTransformer()
    X_tfidf = tfidf_transformer.fit_transform(X_counts)
    return pd.DataFrame(X_tfidf.todense())


def f_topics(data, topic_num):
    print("Processing Topics features ...")
    def cleaning(article):
        punctuation = set(string.punctuation)
        lemmatize = WordNetLemmatizer()
        one = " ".join([i for i in article.lower().split() if i not in stopwords])
        two = "".join(i for i in one if i not in punctuation)
        three = " ".join(lemmatize.lemmatize(i) for i in two.lower().split())
        return three

    def pred_new(doc):
        one = cleaning(doc).split()
        two = dictionary.doc2bow(one)
        return two

    def load_title_body(data):
```

```
        text =[]
        for i in range(len(data["y"])):
            temp = str(data["title"][i])[2:-2]
            for j in data["usertext"][i]:
                temp = temp + ' ' + str(j)[2:-2]
            text.append(temp)
        return text

    stopwords = set(nltk.corpus.stopwords.words('english'))
    text_all = load_title_body(data)
    df = pd.DataFrame({'text': text_all}, index=None)
    text = df.applymap(cleaning)['text']
    text_list = []
    for t in text:
        temp = t.split()
        text_list.append([i for i in temp if i not in stopwords])

    dictionary = corpora.Dictionary(text_list)
    doc_term_matrix = [dictionary.doc2bow(doc) for doc in text_list]
    ldamodel = LdaModel(doc_term_matrix, num_topics=topic_num, id2word = dictionary, passes=50)
    probs = []
    for text in text_all:
        prob = ldamodel[(pred_new(text))]
        d = dict(prob)
        for i in range(topic_num):
            if i not in d.keys():
                d[i] = 0
        temp = []
        for i in range(topic_num):
            temp.append(d[i])
        probs.append(temp)
    return pd.DataFrame(probs, index=None)


if __name__ == '__main__' :
    args = arg_clf()
    df_data = load_df(args.dataset)
    if args.num_features == 1:
        df_basic = f_basic(df_data)
        df_all = pd.concat([df_basic, df_data['y']], axis=1)
    elif args.num_features == 2:
        df_basic = f_basic(df_data)
        df_tfidf = f_tfidf(df_data)
        df_features = pd.concat([df_basic, df_tfidf], axis=1)
        df_all = pd.concat([df_features, df_data['y']], axis=1)
    elif args.num_features == 3:
        df_basic = f_basic(df_data)
        df_tfidf = f_tfidf(df_data)
        tags_all = get_all_tags(df_data)
        df_pos = f_pos(df_data, tags_all)
        df_features = pd.concat([df_basic, df_tfidf, df_pos], axis=1)
        df_all = pd.concat([df_features, df_data['y']], axis=1)
    elif args.num_features == 4:
        df_basic = f_basic(df_data)
        df_tfidf = f_tfidf(df_data)
        tags_all = get_all_tags(df_data)
        df_pos = f_pos(df_data, tags_all)
        df_topic = f_topics(df_data, args.num_topics)
        df_features = pd.concat([df_basic, df_tfidf, df_pos, df_topic], axis=1)
        df_all = pd.concat([df_features, df_data['y']], axis=1)
    elif args.num_features == 5:
        df_basic = f_basic(df_data)
        df_tfidf = f_tfidf(df_data)
        tags_all = get_all_tags(df_data)
```

```
    df_pos = f_pos(df_data, tags_all)
    df_topic = f_topics(df_data, args.num_topics)
    df_liwc = f_liwc(args.dataset)
    df_features = pd.concat([df_basic, df_tfidf, df_pos, df_topic, df_liwc], axis=1)
    df_all = pd.concat([df_features, df_data['y']], axis=1)
else:
    raise ValueError("Error: number of features groups")
result_average, h = [], ['Model', 'Acc.', 'Pre.', 'Rec.', 'F1', 'AUC']
lr_acc, lr_pre, lr_rec, lr_f1, lr_auc = [], [], [], [], []
rf_acc, rf_pre, rf_rec, rf_f1, rf_auc = [], [], [], [], []
gbdt_acc, gbdt_pre, gbdt_rec, gbdt_f1, gbdt_auc = [], [], [], [], []
xgb_acc, xgb_pre, xgb_rec, xgb_f1, xgb_auc = [], [], [], [], []
df_all = df_all.dropna()
X = df_all[df_all.columns[:-1]].as_matrix()
y = df_all['y'].as_matrix()

# 10-fold cross validation
num_fold = 10
kf = KFold(n_splits=num_fold, shuffle=True, random_state=0)
for train_index, test_index in kf.split(X):
    num_fold -= 1
    X_train, X_test = X[train_index], X[test_index]
    y_train, y_test = y[train_index], y[test_index]
    # Logisitc Regression
    clf = LogisticRegression(penalty='l2', tol=1e-6)
    clf.fit(X_train, y_train)
    y_pred = clf.predict_proba(X_test)[:,1]
    acc, pre, rec, f1, auc = evaluate_prediction(y_test, y_pred, k_th=num_fold,
                                    model_name='Logistic Regression', dataset_name=args.dataset)
    lr_acc.append(acc)
    lr_pre.append(pre)
    lr_rec.append(rec)
    lr_f1.append(f1)
    lr_auc.append(auc)
    # Random Forest
    clf = RandomForestClassifier(n_estimators=20, max_depth=8, random_state=0)
    clf.fit(X_train, y_train)
    y_pred = clf.predict_proba(X_test)[:, 1]
    acc, pre, rec, f1, auc = evaluate_prediction(y_test, y_pred, k_th=num_fold,
                                    model_name='Random Forest', dataset_name=args.dataset)
    rf_acc.append(acc)
    rf_pre.append(pre)
    rf_rec.append(rec)
    rf_f1.append(f1)
    rf_auc.append(auc)
    # GBDT
    clf = GradientBoostingClassifier(max_depth=8, random_state=0)
    clf.fit(X_train, y_train)
    y_pred = clf.predict_proba(X_test)[:, 1]
    acc, pre, rec, f1, auc = evaluate_prediction(y_test, y_pred, k_th=num_fold,
                                    model_name='GBDT', dataset_name=args.dataset)
    gbdt_acc.append(acc)
    gbdt_pre.append(pre)
    gbdt_rec.append(rec)
    gbdt_f1.append(f1)
    gbdt_auc.append(auc)
    # XGBoost
    dtrain = xgb.DMatrix(X_train, label=y_train, missing=-999)
    dtest = xgb.DMatrix(X_test, label=y_test, missing=-999)
    params = {'max_depth': 10, 'eta': 0.1, 'silent': 1, 'objective': 'binary:logistic', 'nthread': -1}
    num_round = 10000
    watchlist = [(dtrain, 'train'), (dtest, 'test')]
    model = xgb.train(params, dtrain, num_round, watchlist, early_stopping_rounds=50, verbose_eval=10)
    y_pred = model.predict(dtest)
```

```
    acc, pre, rec, f1, auc = evaluate_prediction(y_test, y_pred, k_th=num_fold,
                                        model_name='XGBoost', dataset_name=args.dataset)
    xgb_acc.append(acc)
    xgb_pre.append(pre)
    xgb_rec.append(rec)
    xgb_f1.append(f1)
    xgb_auc.append(auc)
result_average.append(['Logistic Regression', np.mean(lr_acc), np.mean(lr_pre), np.mean(lr_rec), np.mean(lr_f1), np.mean(lr_auc)])
result_average.append(['Random Forest', np.mean(rf_acc), np.mean(rf_pre), np.mean(rf_rec), np.mean(rf_f1), np.mean(rf_auc)])
result_average.append(['GBDT', np.mean(gbdt_acc), np.mean(gbdt_pre), np.mean(gbdt_rec), np.mean(gbdt_f1), np.mean(gbdt_auc)])
result_average.append(['XGB', np.mean(xgb_acc), np.mean(xgb_pre), np.mean(xgb_rec), np.mean(xgb_f1), np.mean(xgb_auc)])
print(tabulate(result_average, headers=h))
```

Αρχείο helpers.py

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
# Python version: 3.9

import csv
import numpy as np
import pandas as pd
import datetime
from sklearn import metrics


def load_df(dataset_name):
    """
    Load data as DataFrame for logistic regression and ensemble models, without sampling
    :param dataset_name: name of dataset
    :return: DataFrame
    """
    print('Loading dataset DataFrame')
    if dataset_name == 'EP' or dataset_name == 'ep':
        file_name = './data/experience_project/ep.csv'
        df = pd.read_csv(file_name)
        return df
    elif dataset_name == 'twitter' or dataset_name == 'Twitter':
        file_name = './data/twitter/twitter.xlsx'
        df = pd.read_excel(file_name)
        return df
    elif dataset_name in ['popular', 'all', 'AskReddit', 'books', 'gaming', 'movies', 'Jokes']:
        file_name = './data/reddit/{}.csv'.format(dataset_name)
        df = pd.read_csv(file_name)
        return df
    else:
        raise ValueError("Error: unrecognized dataset")


def load_data(dataset_name):
    """
    Load data for NN models, with under sampling from Twitter dataset
    :param dataset_name: name of dataset
    :return: X, y
    """
    print('Loading text dataset')
    if dataset_name == 'EP' or dataset_name == 'ep':
        file_name = './data/experience_project/ep.csv'
        df = pd.read_csv(file_name)
        X = df['body'].as_matrix()
        y = df['y'].as_matrix()
        return X, y
    elif dataset_name == 'twitter' or dataset_name == 'Twitter':
        file_name = './data/twitter/twitter.xlsx'
        df = pd.read_excel(file_name)
        df = df.dropna()
        df_pos = df.loc[df['y'] == 1]
        df_neg = df.loc[df['y'] == 0]
        df = pd.concat([df_pos, df_neg.sample(len(df_pos['y']))])
        df = df.sample(frac=1)
```

```
        X = df['tweets'].as_matrix()
        y = df['y'].as_matrix()
        return X, y
    elif dataset_name in ['popular', 'all', 'AskReddit', 'books', 'gaming', 'movies', 'Jokes']:
        file_name = './data/reddit/{}.csv'.format(dataset_name)
        df = pd.read_csv(file_name)
        X1 = df['title'].as_matrix()
        X2 = df['usertext'].as_matrix()
        y = df['y'].as_matrix()
        return X1, X2, y
    else:
        raise ValueError("Error: unrecognized dataset")


def find_threshold(fpr, tpr, threshold):
    rate = np.array(tpr) + np.array(fpr)
    return threshold[np.argmax(rate)]


def evaluate_prediction(y_test, y_pred, k_th, model_name, dataset_name):
    fpr, tpr, th = metrics.roc_curve(y_test, y_pred)
    roc_auc = metrics.auc(fpr, tpr)

    df_results = pd.DataFrame({'y_test': y_test, 'y_pred': y_pred}, index=None)
    df_results.to_csv('./save/prediction_{}_{}_{}.csv'.format(dataset_name, model_name, k_th), index=False)

    o_threshold = 0.5
    for i in range(len(y_pred)):
        if y_pred[i] >= o_threshold:
            y_pred[i] = 1
        else:
            y_pred[i] = 0

    acc = metrics.accuracy_score(y_test, y_pred)
    pre = metrics.precision_score(y_test, y_pred)
    rec = metrics.recall_score(y_test, y_pred)
    f1 = metrics.f1_score(y_test, y_pred)

    dict_eval = {'date': datetime.date.today(),
                 'model': model_name,
                 'accuracy': acc,
                 'precision': pre,
                 'recall': rec,
                 'f-score': f1,
                 'roc': roc_auc,
                 'note': '{}_th fold'.format(k_th),
                 'dataset': dataset_name
                 }
    with open('./output/{}.csv'.format(dataset_name), 'a') as f:
        field_names = ['date', 'model', 'accuracy', 'precision', 'recall', 'f-score', 'roc', 'note', 'dataset']
        writer = csv.DictWriter(f, fieldnames=field_names)
        writer.writerow(dict_eval)
    return acc, pre, rec, f1, roc_auc
```

Αρχείο lstm.py

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
# Python version: 3.9

import os
import numpy as np
from tqdm import tqdm
from tabulate import tabulate
from sklearn.model_selection import KFold
from keras.models import Model
from keras.layers import Dense, Embedding, Input
from keras.layers import LSTM, GlobalMaxPool1D, Dropout, BatchNormalization
from keras.preprocessing import text, sequence
from keras.callbacks import EarlyStopping

from options import arg_lstm
from helpers import load_data, evaluate_prediction


def build_LSTM(args):
    inp = Input(shape=(args.max_seq_len, ))
    x = Embedding(input_dim=args.max_num_words, output_dim=args.embedding_dim)(inp)
    x = LSTM(units=args.lstm_units, activation='tanh', dropout=args.dropout_rate,
return_sequences=True)(x)
    x = GlobalMaxPool1D()(x)
    x = Dropout(args.dropout_rate)(x)
    x = BatchNormalization()(x)
    x = Dense(args.dense_units, activation="relu")(x)
    x = Dropout(args.dropout_rate)(x)
    x = BatchNormalization()(x)
    x = Dense(1, activation="sigmoid")(x)
    model = Model(inputs=inp, outputs=x)
    model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
    return model


if __name__ == '__main__':
    # parse parameters
    args_lstm = arg_lstm()
    if args_lstm.dataset == 'twitter' or args_lstm.dataset == 'Twitter':
        n_sampling = 5
    else:
        n_sampling = 1
    h, result = ['Model', 'Acc.', 'Pre.', 'Rec.', 'F1', 'AUC'], []
    list_acc, list_pre, list_rec, list_f1, list_auc = [], [], [], [], []
    while n_sampling > 0:
        # load data
        X, y = load_data(dataset_name=args_lstm.dataset)

        # split train test
        kfold = KFold(n_splits=10, shuffle=True, random_state=1234)
        num_fold = 0
        for train_ix, test_ix in tqdm(kfold.split(y)):
            list_sentences_train = X[train_ix]
            y_train = y[train_ix]
            list_sentences_test = X[test_ix]
            y_test = y[test_ix]
```

```
        tokenizer = text.Tokenizer(num_words=args_lstm.max_num_words)
        tokenizer.fit_on_texts(list(list_sentences_train))
        list_tokenized_train = tokenizer.texts_to_sequences(list_sentences_train)
        list_tokenized_test = tokenizer.texts_to_sequences(list_sentences_test)
        X_train = sequence.pad_sequences(list_tokenized_train, maxlen=args_lstm.max_seq_len,
dtype='float64')
        X_test = sequence.pad_sequences(list_tokenized_test, maxlen=args_lstm.max_seq_len,
dtype='float64')

        # train model and predict
        model = build_LSTM(args=args_lstm)

        early = EarlyStopping(monitor="val_loss", mode="min", patience=20)
        model.fit(X_train, y_train, batch_size=args_lstm.batch_size, epochs=args_lstm.epochs,
              validation_split=args_lstm.valid_split, callbacks=[early])

        y_pred = model.predict(X_test).reshape(y_test.shape)

        acc, pre, rec, f1, auc = evaluate_prediction(y_test, y_pred, k_th=num_fold,
model_name='LSTM',
                                             dataset_name=args_lstm.dataset)

        list_acc.append(acc)
        list_pre.append(pre)
        list_rec.append(rec)
        list_f1.append(f1)
        list_auc.append(auc)
        result.append(['LSTM', acc, pre, rec, f1, auc])
        num_fold += 1
    n_sampling -= 1
  result.append(['average', np.mean(list_acc), np.mean(list_pre),
            np.mean(list_rec), np.mean(list_f1), np.mean(list_auc)])
  print(tabulate(result, headers=h))
```

Αρχείο lstm_reddit.py

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
# Python version: 3.9

import numpy as np
from tqdm import tqdm
from tabulate import tabulate
from sklearn.model_selection import KFold
from keras.models import Model
from keras.layers import Dense, Embedding, Input
from keras.layers import LSTM, GlobalMaxPool1D, Dropout, BatchNormalization
from keras.layers.merge import concatenate
from keras.preprocessing import text, sequence
from keras.callbacks import EarlyStopping

from options import arg_lstm
from helpers import load_data, evaluate_prediction


def build_LSTM(args):

    inp1 = Input(shape=(args.max_seq_len, ))
    inp2 = Input(shape=(args.max_seq_len, ))
    x1 = Embedding(input_dim=args.max_num_words, output_dim=args.embedding_dim)(inp1)
    x2 = Embedding(input_dim=args.max_num_words, output_dim=args.embedding_dim)(inp2)
    x1 = LSTM(units=args.lstm_units, activation='tanh', return_sequences=True)(x1)
    x2 = LSTM(units=args.lstm_units, activation='tanh', return_sequences=True)(x2)
    x = concatenate([x1, x2])
    x = GlobalMaxPool1D()(x)
    x = Dropout(args.dropout_rate)(x)
    x = BatchNormalization()(x)
    x = Dense(args.dense_units, activation="relu")(x)
    x = Dropout(args.dropout_rate)(x)
    x = BatchNormalization()(x)
    x = Dense(1, activation="sigmoid")(x)
    model = Model(inputs=[inp1, inp2], outputs=x)
    model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
    return model


if __name__ == '__main__':
    # parse parameters
    args_lstm = arg_lstm()
    if args_lstm.dataset == 'twitter' or args_lstm.dataset == 'Twitter':
        n_sampling = 5
    else:
        n_sampling = 1
    h, result = ['Model', 'Acc.', 'Pre.', 'Rec.', 'F1', 'AUC'], []
    list_acc, list_pre, list_rec, list_f1, list_auc = [], [], [], [], []
    while n_sampling > 0:
        # load data
        X1, X2, y = load_data(dataset_name=args_lstm.dataset)

        # split train test
        kfold = KFold(n_splits=10, shuffle=True, random_state=1234)
```

```
        num_fold = 0
        for train_ix, test_ix in tqdm(kfold.split(y)):
            title_train = X1[train_ix]
            title_test = X1[test_ix]
            usertext_train = X2[train_ix]
            usertext_test = X2[test_ix]
            y_train = y[train_ix]
            y_test = y[test_ix]

            tokenizer = text.Tokenizer(num_words=args_lstm.max_num_words)
            tokenizer.fit_on_texts(list(title_train)+list(usertext_train))
            tokenized_title_train = tokenizer.texts_to_sequences(title_train)
            tokenized_title_test = tokenizer.texts_to_sequences(title_test)
            tokenized_usertext_train = tokenizer.texts_to_sequences(usertext_train)
            tokenized_usertext_test = tokenizer.texts_to_sequences(usertext_test)
            X1_train = sequence.pad_sequences(tokenized_title_train, maxlen=args_lstm.max_seq_len,
dtype='float64')
            X1_test = sequence.pad_sequences(tokenized_title_test, maxlen=args_lstm.max_seq_len,
dtype='float64')
            X2_train = sequence.pad_sequences(tokenized_usertext_train, maxlen=args_lstm.max_seq_len,
dtype='float64')
            X2_test = sequence.pad_sequences(tokenized_usertext_test, maxlen=args_lstm.max_seq_len,
dtype='float64')

            # train model and predict
            model = build_LSTM(args=args_lstm)

            early = EarlyStopping(monitor="val_loss", mode="min", patience=20)
            model.fit([X1_train, X2_train], y_train, batch_size=args_lstm.batch_size, epochs=args_lstm.epochs,
                      validation_split=0.1, callbacks=[early])

            y_pred = model.predict([X1_test, X2_test]).reshape(y_test.shape)

            acc, pre, rec, f1, auc = evaluate_prediction(y_test, y_pred, k_th=num_fold, model_name='LSTM',
                                                         dataset_name=args_lstm.dataset)

            list_acc.append(acc)
            list_pre.append(pre)
            list_rec.append(rec)
            list_f1.append(f1)
            list_auc.append(auc)
            result.append(['LSTM', acc, pre, rec, f1, auc])
            num_fold += 1
        n_sampling -= 1
    result.append(['average', np.mean(list_acc), np.mean(list_pre),
                   np.mean(list_rec), np.mean(list_f1), np.mean(list_auc)])
    print(tabulate(result, headers=h))
```

Αρχείο lstm_word2vec.py

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
# Python version: 3.9

import numpy as np
from tabulate import tabulate
from gensim.models import KeyedVectors
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences
from keras.layers import Dense, Input, LSTM, Embedding, Dropout
from keras.models import Model
from keras.layers.normalization import BatchNormalization
from keras.callbacks import EarlyStopping
from sklearn.model_selection import KFold
from tqdm import tqdm

from options import arg_lstm
from helpers import load_data, evaluate_prediction


def build_LSTM(args):
    # define the model structure
    embedding_layer = Embedding(nb_words, args.embedding_dim, weights=[embedding_matrix],
                                input_length=args.max_seq_len, trainable=False)
    lstm_layer = LSTM(args.lstm_units, activation='tanh', dropout=args.dropout_rate)
    sequence_input = Input(shape=(args.max_seq_len,), dtype='int32')
    embedded_sequences = embedding_layer(sequence_input)
    x = lstm_layer(embedded_sequences)
    x = Dropout(args.dropout_rate)(x)
    x = BatchNormalization()(x)
    x = Dense(args.dense_units, activation='relu')(x)
    x = Dropout(args.dropout_rate)(x)
    x = BatchNormalization()(x)
    preds = Dense(1, activation='sigmoid')(x)
    model = Model(inputs=sequence_input, outputs=preds)
    model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['acc'])
    return model


if __name__ == "__main__":
    # parse parameters
    args_lstm = arg_lstm()

    print('Loading pertrained embedding file')
    word2vec = KeyedVectors.load_word2vec_format(args_lstm.embedding_file, binary=True)
    print('Found %s word vectors of word2vec' % len(word2vec.vocab))

    X, y = load_data(dataset_name=args_lstm.dataset)
    all_text = []
    for i in range(len(X)):
        all_text.append(X[i])
```

```
tokenizer = Tokenizer(num_words=args_lstm.max_num_words)
tokenizer.fit_on_texts(all_text)

word_index = tokenizer.word_index
print('Found %s unique tokens' % len(word_index))
print('Preparing embedding matrix')
nb_words = min(args_lstm.max_num_words, len(word_index)) + 1
embedding_matrix = np.zeros((nb_words, args_lstm.embedding_dim))
for word, i in word_index.items():
    if word in word2vec.vocab:
        embedding_matrix[i] = word2vec.word_vec(word)
print('Null word embeddings: %d' % np.sum(np.sum(embedding_matrix, axis=1) == 0))

if args_lstm.dataset == 'twitter' or args_lstm.dataset == 'Twitter':
    n_sampling = 5
else:
    n_sampling = 1
h, result = ['Model', 'Acc.', 'Pre.', 'Rec.', 'F1', 'AUC'], []
list_acc, list_pre, list_rec, list_f1, list_auc = [], [], [], [], []
while n_sampling > 0:
    X, y = load_data(dataset_name=args_lstm.dataset)

    # 10 fold
    num_fold = 10
    kf = KFold(n_splits=num_fold, shuffle=True, random_state=0)

    h, result = ['Model', 'Acc.', 'Pre.', 'Rec.', 'F1', 'AUC'], []
    list_acc, list_pre, list_rec, list_f1, list_auc = [], [], [], [], []
    for train_index, test_index in tqdm(kf.split(X)):
        num_fold -= 1
        text_train, text_test = X[train_index], X[test_index]
        y_train, y_test = y[train_index], y[test_index]

        train_text_seq = tokenizer.texts_to_sequences(text_train)
        test_text_seq = tokenizer.texts_to_sequences(text_test)

        train_tweet = pad_sequences(train_text_seq, maxlen=args_lstm.max_seq_len)
        test_tweet = pad_sequences(test_text_seq, maxlen=args_lstm.max_seq_len)
        train_labels = np.array(y_train)
        test_labels = np.array(y_test)

        # sample train/validation data
        np.random.seed(1234)
        perm = np.random.permutation(len(train_tweet))
        idx_train = perm[:int(len(train_tweet) * (1 - args_lstm.valid_split))]
        idx_val = perm[int(len(train_tweet) * (1 - args_lstm.valid_split)):]
        data_train = train_tweet[idx_train]
        labels_train = train_labels[idx_train]
        data_val = train_tweet[idx_val]
        labels_val = train_labels[idx_val]

        # train the model
        model = build_LSTM(args=args_lstm)
```

```
        early_stopping = EarlyStopping(monitor='val_loss', patience=10)
        hist = model.fit(data_train, labels_train, validation_data=(data_val, labels_val),
                         epochs=args_lstm.epochs, batch_size=args_lstm.batch_size, shuffle=True, callbacks=[early_stopping])
        # predict
        print('Testing')
        preds = model.predict(test_tweet, batch_size=32, verbose=1)
        y_pred = preds.ravel()
        acc, pre, rec, f1, auc = evaluate_prediction(y_test=y_test, y_pred=y_pred,
                                             k_th=num_fold, model_name='LSTM-word2vec', dataset_name=args_lstm.dataset)
        list_acc.append(acc)
        list_pre.append(pre)
        list_rec.append(rec)
        list_f1.append(f1)
        list_auc.append(auc)
        result.append(['LSTM-word2vec', acc, pre, rec, f1, auc])
    n_sampling -= 1
result.append(['average', np.mean(list_acc), np.mean(list_pre),
               np.mean(list_rec), np.mean(list_f1), np.mean(list_auc)])
print(tabulate(result, headers=h))
```

Αρχείο lstm_word2vec_reddit.py

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
# Python version: 3.9

import numpy as np
from sklearn.model_selection import KFold
from gensim.models import KeyedVectors
from keras.preprocessing.text import Tokenizer
from keras.preprocessing import sequence
from keras.layers import Dense, Input, LSTM, Embedding, Dropout
from keras.layers.merge import concatenate
from keras.models import Model
from keras.layers.normalization import BatchNormalization
from keras.callbacks import EarlyStopping
from tqdm import tqdm
from tabulate import tabulate
from options import arg_lstm
from helpers import load_data, evaluate_prediction


def build_LSTM(args, mat_embedding):
    embedding_layer = Embedding(nb_words, args.embedding_dim, weights=[mat_embedding],
                                input_length=args.max_seq_len, trainable=False)
    lstm_layer = LSTM(args.lstm_units, dropout=args.dropout_rate_lstm, recurrent_dropout=args.dropout_rate)
    sequence_1_input = Input(shape=(args.max_seq_len,), dtype='int32')
    embedded_sequences_1 = embedding_layer(sequence_1_input)
    x1 = lstm_layer(embedded_sequences_1)
    sequence_2_input = Input(shape=(args.max_seq_len,), dtype='int32')
    embedded_sequences_2 = embedding_layer(sequence_2_input)
    x2 = lstm_layer(embedded_sequences_2)
    merged = concatenate([x1, x2])
    merged = Dropout(args.dropout_rate)(merged)
    merged = BatchNormalization()(merged)
    merged = Dense(args.dense_units, activation='relu')(merged)
    merged = Dropout(args.dropout_rate)(merged)
    merged = BatchNormalization()(merged)
    preds = Dense(1, activation='sigmoid')(merged)
    model = Model(inputs=[sequence_1_input, sequence_2_input], outputs=preds)
    model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['acc'])
    return model


if __name__ == '__main__':
    args_lstm = arg_lstm()
    X1, X2, y = load_data(dataset_name=args_lstm.dataset)
    all_text = []
    for i in range(len(X1)):
        all_text.append(X1[i])
        all_text.append(X2[i])

    tokenizer = Tokenizer(num_words=args_lstm.max_num_words)
    tokenizer.fit_on_texts(all_text)
    word_index = tokenizer.word_index
    print('Found %s unique tokens' % len(word_index))
```

```
print('Loading pertrained embedding file')
word2vec = KeyedVectors.load_word2vec_format(args_lstm.embedding_file, binary=True)
print('Found %s word vectors of word2vec' % len(word2vec.vocab))

print('Preparing embedding matrix')
nb_words = min(args_lstm.max_num_words, len(word_index)) + 1
embedding_matrix = np.zeros((nb_words, args_lstm.embedding_dim))
for word, i in word_index.items():
    if word in word2vec.vocab:
        embedding_matrix[i] = word2vec.word_vec(word)
print('Null word embeddings: %d' % np.sum(np.sum(embedding_matrix, axis=1) == 0))

h, result = ['Model', 'Acc.', 'Pre.', 'Rec.', 'F1', 'AUC'], []
list_acc, list_pre, list_rec, list_f1, list_auc = [], [], [], [], []

# 10 fold
num_fold = 10
kf = KFold(n_splits=num_fold, shuffle=True, random_state=0)
for train_ix, test_ix in tqdm(kf.split(y)):
    title_train = X1[train_ix]
    title_test = X1[test_ix]
    usertext_train = X2[train_ix]
    usertext_test = X2[test_ix]
    y_train = y[train_ix]
    y_test = y[test_ix]

    tokenized_title_train = tokenizer.texts_to_sequences(title_train)
    tokenized_title_test = tokenizer.texts_to_sequences(title_test)
    tokenized_usertext_train = tokenizer.texts_to_sequences(usertext_train)
    tokenized_usertext_test = tokenizer.texts_to_sequences(usertext_test)
    X1_train = sequence.pad_sequences(tokenized_title_train, maxlen=args_lstm.max_seq_len, dtype='float64')
    X1_test = sequence.pad_sequences(tokenized_title_test, maxlen=args_lstm.max_seq_len, dtype='float64')
    X2_train = sequence.pad_sequences(tokenized_usertext_train, maxlen=args_lstm.max_seq_len, dtype='float64')
    X2_test = sequence.pad_sequences(tokenized_usertext_test, maxlen=args_lstm.max_seq_len, dtype='float64')

    # train model and predict
    model = build_LSTM(args=args_lstm, mat_embedding=embedding_matrix)

    early = EarlyStopping(monitor="val_loss", mode="min", patience=20)
    model.fit([X1_train, X2_train], y_train, batch_size=args_lstm.batch_size, epochs=args_lstm.epochs,
              validation_split=0.1, callbacks=[early])

    preds = model.predict([X1_test, X2_test])
    y_pred = preds.ravel()
    acc, pre, rec, f1, auc = evaluate_prediction(y_test, y_pred, k_th=num_fold, model_name='LSTM',
                                                 dataset_name=args_lstm.dataset)

    list_acc.append(acc)
    list_pre.append(pre)
    list_rec.append(rec)
    list_f1.append(f1)
    list_auc.append(auc)
    result.append(['LSTM-word2vec', acc, pre, rec, f1, auc])
    num_fold += 1
result.append(['average', np.mean(list_acc), np.mean(list_pre),
               np.mean(list_rec), np.mean(list_f1), np.mean(list_auc)])
print(tabulate(result, headers=h))
```

Αρχείο options.py

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
# Python version: 3.9

import argparse


def arg_clf():
    parser = argparse.ArgumentParser()
    parser.add_argument('--dataset', type=str, default='reddit', help="name of dataset")
    parser.add_argument('--num_features', type=int, default=1, help="number of features groups")
    parser.add_argument('--num_topics', type=int, default=10, help="number of topics")
    args = parser.parse_args()
    return args


def arg_rnn():
    parser = argparse.ArgumentParser()
    parser.add_argument('--max_seq_len', type=int, default=1000, help="max length of sequences")
    parser.add_argument('--max_num_words', type=int, default=20000, help="max number of words")
    parser.add_argument('-d', '--embedding_dim', type=int, default=300, help="embedding dimension")
    parser.add_argument('--rnn_units', type=int, default=128, help="units of RNN")
    parser.add_argument('--dropout_rate', type=float, default=0.1, help='dropout rate')
    parser.add_argument('--dense_units', type=int, default=32, help="units of Dense layer")

    parser.add_argument('--dataset', type=str, default='twitter', help="name of dataset")
    embedding_file = '/data/shji/datasets/word2vec/GoogleNews-vectors-negative300.bin'
    parser.add_argument('-f', '--embedding_file', type=str, default=embedding_file, help="embedding file")
    parser.add_argument('--embedding_type', type=str, default='word2vec', help="the type of word embedding")
    parser.add_argument('--valid_split', type=float, default=0.1, help="ratio of validation split")
    parser.add_argument('-act', '--activation', type=str, default='relu', help="type of activation function")
    parser.add_argument('--patience', type=int, default=10, help="number of epochs with no improvement after which training will be stopped")
    parser.add_argument('--batch_size', type=int, default=64, help="batch size")
    parser.add_argument('--epochs', type=int ,default=200, help="training epochs")
    args = parser.parse_args()
    return args


def arg_lstm():
    parser = argparse.ArgumentParser()
    parser.add_argument('--max_seq_len', type=int, default=1000, help="max length of sequences")
    parser.add_argument('--max_num_words', type=int, default=50000, help="max number of words")
    parser.add_argument('-d', '--embedding_dim', type=int, default=300, help="embedding dimension")
    parser.add_argument('--lstm_units', type=int, default=128, help="units of LSTM")
    parser.add_argument('--dropout_rate', type=float, default=0.1, help='dropout rate')
    parser.add_argument('--dropout_rate_lstm', type=float, default=0.2, help='dropout rate of lstm unit')
    parser.add_argument('--dense_units', type=int, default=32, help="units of Dense layer")

    parser.add_argument('--dataset', type=str, default='twitter', help="name of dataset")
    embedding_file = '/data/shji/datasets/word2vec/GoogleNews-vectors-negative300.bin'
    parser.add_argument('-f', '--embedding_file', type=str, default=embedding_file, help="embedding file")
    parser.add_argument('--embedding_type', type=str, default='word2vec', help="the type of word embedding")
    parser.add_argument('--valid_split', type=float, default=0.1, help="ratio of validation split")
    parser.add_argument('-act', '--activation', type=str, default='relu', help="type of activation function")
    parser.add_argument('--patience', type=int, default=10, help="number of epochs with no improvement after which training will be stopped")
    parser.add_argument('--batch_size', type=int, default=64, help="batch size")
    parser.add_argument('--epochs', type=int ,default=200, help="training epochs")
    args = parser.parse_args()
    return args

if __name__ == '__main__':
    a = arg_lstm()
```

Αρχείο rnn.py

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
# Python version: 3.6

import numpy as np
from tqdm import tqdm
from tabulate import tabulate
from sklearn.model_selection import KFold
from keras.models import Model
from keras.layers import Dense, Embedding, Input
from keras.layers import SimpleRNN, GlobalMaxPool1D, Dropout, BatchNormalization
from keras.preprocessing import text, sequence
from keras.callbacks import EarlyStopping

from options import arg_rnn
from helpers import load_data, evaluate_prediction


def build_RNN(args):
    inp = Input(shape=(args.max_seq_len, ))
    x = Embedding(args.max_num_words, args.embedding_dim)(inp)
    x = SimpleRNN(args.rnn_units, return_sequences=True)(x)
    x = GlobalMaxPool1D()(x)
    x = Dropout(rate=args.dropout_rate)(x)
    x = BatchNormalization()(x)
    x = Dense(units=args.dense_units, activation="relu")(x)
    x = Dropout(rate=args.dropout_rate)(x)
    x = BatchNormalization()(x)
    x = Dense(1, activation="sigmoid")(x)
    model = Model(inputs=inp, outputs=x)
    model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
    return model


if __name__ == '__main__':
    # parse parameters
    args_rnn = arg_rnn()
    if args_rnn.dataset == 'twitter' or args_rnn.dataset == 'Twitter':
        n_sampling = 5
    else:
        n_sampling = 1
    h, result = ['Model', 'Acc.', 'Pre.', 'Rec.', 'F1', 'AUC'], []
    list_acc, list_pre, list_rec, list_f1, list_auc = [], [], [], [], []
    while n_sampling > 0:
        # load data
        X, y = load_data(dataset_name=args_rnn.dataset)

        # split train test
        kfold = KFold(n_splits=10, shuffle=True, random_state=1234)
        num_fold = 0
        for train_ix, test_ix in tqdm(kfold.split(y)):
            list_sentences_train = X[train_ix]
            y_train = y[train_ix]
            list_sentences_test = X[test_ix]
            y_test = y[test_ix]
```

```
        tokenizer = text.Tokenizer(num_words=args_rnn.max_num_words)
        tokenizer.fit_on_texts(list(list_sentences_train))
        list_tokenized_train = tokenizer.texts_to_sequences(list_sentences_train)
        list_tokenized_test = tokenizer.texts_to_sequences(list_sentences_test)
        X_train = sequence.pad_sequences(list_tokenized_train, maxlen=args_rnn.max_seq_len, dtype='float64')
        X_test = sequence.pad_sequences(list_tokenized_test, maxlen=args_rnn.max_seq_len, dtype='float64')

        # train model and predict
        model = build_RNN(args=args_rnn)
        early = EarlyStopping(monitor="val_loss", mode="min", patience=20)
        model.fit(X_train, y_train, batch_size=args_rnn.batch_size, epochs=args_rnn.epochs,
                  validation_split=0.1, callbacks=[early])

        y_pred = model.predict(X_test).reshape(y_test.shape)

        acc, pre, rec, f1, auc = evaluate_prediction(y_test, y_pred, k_th=num_fold, model_name='RNN',
                                                     dataset_name=args_rnn.dataset)

        list_acc.append(acc)
        list_pre.append(pre)
        list_rec.append(rec)
        list_f1.append(f1)
        list_auc.append(auc)
        result.append(['RNN', acc, pre, rec, f1, auc])
        num_fold += 1
    n_sampling -= 1
result.append(['average', np.mean(list_acc), np.mean(list_pre),
               np.mean(list_rec), np.mean(list_f1), np.mean(list_auc)])
print(tabulate(result, headers=h))
```

Αρχείο rnn_reddit.py

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
# Python version: 3.9

import numpy as np
from tqdm import tqdm
from tabulate import tabulate
from sklearn.model_selection import KFold
from keras.models import Model
from keras.layers import Dense, Embedding, Input
from keras.layers import SimpleRNN, GlobalMaxPool1D, Dropout, BatchNormalization
from keras.layers.merge import concatenate
from keras.preprocessing import text, sequence
from keras.callbacks import EarlyStopping

from options import arg_rnn
from helpers import load_data, evaluate_prediction


def build_RNN(args):
    inp1 = Input(shape=(args.max_seq_len, ))
    inp2 = Input(shape=(args.max_seq_len, ))
    x1 = Embedding(args.max_num_words, args.embedding_dim)(inp1)
    x2 = Embedding(args.max_num_words, args.embedding_dim)(inp2)
    x1 = SimpleRNN(args.rnn_units, return_sequences=True)(x1)
    x2 = SimpleRNN(args.rnn_units, return_sequences=True)(x2)
    x = concatenate([x1, x2])
    x = GlobalMaxPool1D()(x)
    x = Dropout(args.dropout_rate)(x)
    x = BatchNormalization()(x)
    x = Dense(args.dense_units, activation="relu")(x)
    x = Dropout(args.dropout_rate)(x)
    x = BatchNormalization()(x)
    x = Dense(1, activation="sigmoid")(x)
    model = Model(inputs=[inp1, inp2], outputs=x)
    model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
    return model


if __name__ == '__main__':
    # parse parameters
    args_rnn = arg_rnn()
    if args_rnn.dataset == 'twitter' or args_rnn.dataset == 'Twitter':
        n_sampling = 5
    else:
        n_sampling = 1
    h, result = ['Model', 'Acc.', 'Pre.', 'Rec.', 'F1', 'AUC'], []
    list_acc, list_pre, list_rec, list_f1, list_auc = [], [], [], [], []
    while n_sampling > 0:
        # load data
        X1, X2, y = load_data(dataset_name=args_rnn.dataset)

        # split train test
        kfold = KFold(n_splits=10, shuffle=True, random_state=1234)
        num_fold = 0
```

```
    for train_ix, test_ix in tqdm(kfold.split(y)):
        title_train = X1[train_ix]
        title_test = X1[test_ix]
        usertext_train = X2[train_ix]
        usertext_test = X2[test_ix]
        y_train = y[train_ix]
        y_test = y[test_ix]

        tokenizer = text.Tokenizer(num_words=args_rnn.max_num_words)
        tokenizer.fit_on_texts(list(title_train)+list(usertext_train))
        tokenized_title_train = tokenizer.texts_to_sequences(title_train)
        tokenized_title_test = tokenizer.texts_to_sequences(title_test)
        tokenized_usertext_train = tokenizer.texts_to_sequences(usertext_train)
        tokenized_usertext_test = tokenizer.texts_to_sequences(usertext_test)
        X1_train = sequence.pad_sequences(tokenized_title_train, maxlen=args_rnn.max_seq_len, dtype='float64')
        X1_test = sequence.pad_sequences(tokenized_title_test, maxlen=args_rnn.max_seq_len, dtype='float64')
        X2_train = sequence.pad_sequences(tokenized_usertext_train, maxlen=args_rnn.max_seq_len, dtype='float64')
        X2_test = sequence.pad_sequences(tokenized_usertext_test, maxlen=args_rnn.max_seq_len, dtype='float64')

        # train model and predict
        model = build_RNN(args=args_rnn)
        early = EarlyStopping(monitor="val_loss", mode="min", patience=20)
        model.fit([X1_train, X2_train], y_train, batch_size=args_rnn.batch_size, epochs=args_rnn.epochs,
                  validation_split=0.1, callbacks=[early])

        y_pred = model.predict([X1_test, X2_test]).reshape(y_test.shape)
        acc, pre, rec, f1, auc = evaluate_prediction(y_test, y_pred, k_th=num_fold, model_name='RNN',
                                                     dataset_name=args_rnn.dataset)

        list_acc.append(acc)
        list_pre.append(pre)
        list_rec.append(rec)
        list_f1.append(f1)
        list_auc.append(auc)
        result.append(['RNN', acc, pre, rec, f1, auc])
        num_fold += 1
    n_sampling -= 1
    result.append(['average', np.mean(list_acc), np.mean(list_pre),
                   np.mean(list_rec), np.mean(list_f1), np.mean(list_auc)])
    print(tabulate(result, headers=h))
```

Υπεύθυνη Δήλωση Συγγραφέα:

Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν.1599/1986, η παρούσα εργασία αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης.