ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Η ΑΝΑΛΥΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΚΑΙ ΟΙ ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΣΤΗΝ ΕΚΠΑΙΔΕΥΣΗ

ΑΡΓΥΡΟΠΟΥΛΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΕΥΘΥΝΗ

……..Ραγάζου Βασιλική ……..
…….Ακαδημαϊκός Υπότροφος……

Λαμία 19 Φεβρουαρίου έτος 2025

ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Η ΑΝΑΛΥΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΚΑΙ ΟΙ ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΣΤΗΝ ΕΚΠΑΙΔΕΥΣΗ

ΑΡΓΥΡΟΠΟΥΛΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΕΥΘΥΝΗ

……..Ραγάζου Βασιλική ……..
…….Ακαδημαϊκός Υπότροφος…….

Λαμία 19 Φεβρουαρίου έτος 2025

UNIVERSITY OF THESSALY

SCHOOL OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE & TELECOMMUNICATIONS

SENTIMENT ANALYSIS AND ITS APPLICATION IN EDUCATION

ARGYROPOULOS KONSTANTINOS

FINAL THESIS

ADVISOR

……..Vasiliki ragazou……..
…….Adjunct Lecturer…….

Lamia February 19 year 2025

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

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

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

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

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

Ημερομηνία: 20/12/2025

Ο – Η Δηλ.

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

ΕΥΧΑΡΙΣΤΙΕΣ

Ξεκινώντας, να ευχαριστήσω την *επιβλέπουσα* κ. Ραγάζου, που ανέλαβε την επίβλεψη της πτυχιακής μου εργασίας και με καθοδήγησε αρκετά και μεθοδικά καθ' όλη την συγγραφή της παρούσας πτυχιακής εργασίας.

Τους *γονείς* μου,

Τις δύο *αδερφές* μου,

Τις/Τους *φίλες/φίλους* μου μέσα από τη σχολή, που μου έδωσαν την δική τους ανατροφοδότηση πάνω στην παρούσα πτυχιακή εργασία, εκ των οποίων θέλω ιδιαίτερα να ευχαριστήσω

Την *Αναστασία*,

Την *Ραλού*, και

Την *Μαριάννα*

Τέλος, τον τετράποδο φίλο μου, τον *Ρονάκο*, που πέρασε αρκετές από τις ώρες που έγραφα την πτυχιακή εργασία να κουρνιάζει δίπλα μου.

ΠΕΡΙΛΗΨΗ

Η παρούσα πτυχιακή ασχολείται με την ανάλυση συναισθημάτων (sentiment analysis) στην τριτοβάθμια εκπαίδευση. Ξεκινώντας από τις απαρχές της συναισθηματικής ανάλυσης, γύρω στη δεκαετία του ‘40, με την πρώιμη μορφή της να ασχολείται με αξιολογήσεις πάνω σε ταινίες που παρακολούθησε το κοινό. Με την πάροδο του καιρού, των συνεχών βελτιώσεων και των πειραματισμών στον τρόπο διεξαγωγής της, έφτασε σήμερα σε ανάλυση τόσο κριτικών και αξιολογήσεων σε σεμινάρια, όσο και σε ανάλυση συναισθήματος σε πραγματικό χρόνο, αναλύοντας φράσεις και κινήσεις ενός ατόμου μέσα σε ένα στιγμιότυπο. Έτσι, δημιουργούνται ερωτήματα όπως: “Ποια μέσα χρησιμοποιούνται για να επιτευχθεί η ανάλυση αυτή;”, “Τι προσφέρει η γνώση της συναισθηματικής κατάστασης των εκπαιδευόμενων, σε μια διαδικτυακή πλατφόρμα ή σε ένα εκπαιδευτικό ίδρυμα;” Η ανάλυση συναισθημάτων αποτελεί μια υποστηρικτική μεθοδολογία σε πολλούς τομείς της ανθρώπινης δραστηριότητας, και πολλές φορές πιο ισχυρή όταν πρόκειται για εμπορικούς τομείς. Αυτό, μαζί με τη δομή της και τον τρόπο λειτουργίας της αποτελούν το πρώτο μέρος της παρούσας εργασίας. Το δεύτερο μέρος απαρτίζεται από κώδικα ανάλυσης συναισθημάτων πάνω σε ένα σύνολο δεδομένων ειδικά διαμορφωμένο και επεξεργασμένο για τους σκοπούς της εργασίας.

ABSTRACT

This thesis examines sentiment analysis within the realm of higher education. Sentiment analysis originated in the 1940s, initially focusing on public appraisals of films. Over time, continuous enhancements and experimentation in its execution have culminated in the analysis of reviews and ratings during seminars, as well as real-time emotional assessment, scrutinizing an individual's expressions and gestures within a momentary observation. This prompts inquiries such as, "What instruments are employed to conduct this analysis?" and "What advantages does understanding learners' emotional states provide in an online platform or educational institution?" Sentiment analysis serves as a supplementary tool across several domains of human activity, particularly demonstrating greater efficacy in commercial sectors. This, along with its structure and functionality, constitutes the initial section of this study. The second section comprises sentiment analysis code applied to a dataset meticulously curated and processed for the objectives of the research.

Table of Contents

ΚΕΦΑΛΑΙΟ 1 Εισαγωγή

Τι είναι η ανάλυση συναισθημάτων

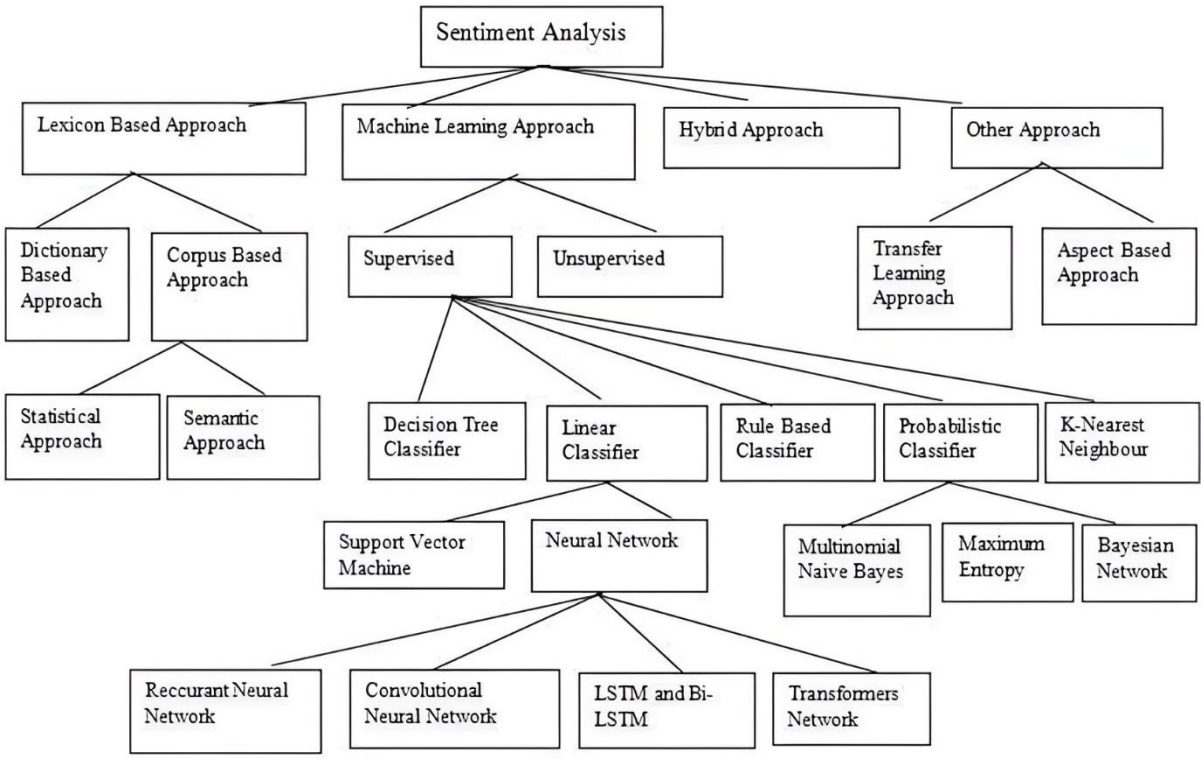
Η ανάλυση συναισθήματος, γνωστή κι ως εξόρυξη γνώμης (opinion mining), είναι μια τεχνική επεξεργασίας φυσικής γλώσσας (Natural Language Processing - NLP) που χρησιμοποιείται για να προσδιορίσει εάν τα δεδομένα έχουν θετικό, αρνητικό ή ουδέτερο χαρακτήρα. Η ανάλυση συναισθήματος εκτελείται συχνά σε δεδομένα κειμένου για να βοηθήσει τις επιχειρήσεις να παρακολουθούν το συναίσθημα της μάρκας και του προϊόντος στα σχόλια των πελατών και να κατανοούν τις ανάγκες των πελατών.

Πιθανότατα, η πρώτη δημοσίευση σχετικά με την γνώμη του κοινού ή ειδικών, ανήκει στον Stagner (*Bordoloi, et al. 2023)*. Μία άλλη πιθανή αρχική δημοσίευση για την ανάλυση συναισθημάτων ανήκει σε άρθρα που εξέδωσε ο/η Pang (*Pang et al. 2002)*. Η εν λόγω δημοσίευση είχε θέμα τις αξιολογήσεις ταινιών, όπου μέσα από κριτικές του κοινού έλαβε χώρα η ανάλυση συναισθήματος που βασίστηκε στη μηχανική μάθηση, μια μορφή τεχνητής νοημοσύνης στην οποία θεωρείται ότι τα συστήματα έχουν τη δυνατότητα της μάθησης μέσω δεδομένων, του εντοπισμού μοτίβων, καθώς και της δυνατότητας λήψης αποφάσεων με μικρή ανθρώπινη παρέμβαση. Η ταξινόμηση της πρώτης δημοσίευσης σχετικά με τη γνώμη κοινού και ειδικών έγινε πάνω σε ένα αρχείο, αναφορικά με το συναίσθημα, δηλαδή εξετάστηκε το πρόσημο των κριτικών(αρνητικό ή θετικό), παρά το θέμα αυτό καθαυτό.

Η ανάλυση συναισθήματος ξεκίνησε ως μια μεθοδολογία, η οποία αξιοποιήθηκε σαν εργαλείο για να συμβάλει στον υπολογισμό ικανοποίησης του κοινού, δηλαδή για σκοπούς Marketing. Στις μέρες μας, έχοντας πλέον αποκτήσει ακόμα περισσότερη δημοτικότητα και σπουδαιότητα, η χρήση της έχει επεκταθεί σε ακόμα περισσότερους τομείς της ανθρώπινης δραστηριότητας, όπως Recommender system (με την πιο «οικεία» σε εμάς μορφή του, τα Cookies των ιστοτόπων), λογοτεχνική φήμη καλλιτεχνών οποιουδήποτε τομέα, πολιτική (όπως παρακολούθηση πορείας και επιτυχίας προεκλογικών εκστρατειών) και στον συνοψισμό των απόψεων του κοινού.

Η ανάλυση συναισθήματος επικεντρώνεται στην πολικότητα του συναισθήματος *(θετικό, αρνητικό*) που εμφανίζεται σε ένα κείμενο, αλλά συνήθως την υπερβαίνει για να ανιχνεύσει συγκεκριμένες ψυχικές καταστάσεις -όπως θυμός, χαρά, λύπη-, τον επείγοντα χαρακτήρα *(επείγον, μη επείγον*), ακόμη και επιπρόσθετες πληροφορίες *(πχ. ενδιαφέρον ή μη ενδιαφέρον*). Πλέον, έχοντας εμβαθύνει περισσότερο στο φάσμα της πολικότητας του συναισθήματος, έχουν οριστεί και οι ενδιάμεσες κλιμακώσεις των προαναφερθέντων ταξινομήσεων, όπως παρουσιάζεται παρακάτω.

Μέσω της ανάλυσης συναισθημάτων στοχευμένα στην εκπαίδευση, καταγράφεται και επεξεργάζεται η γνώμη των εκπαιδευόμενων σχετικά με ένα θέμα, όπως ένα μάθημα, το εκπαιδευτικό προσωπικό, το πανεπιστήμιο κτλ. Η γνώμη τους αντλείται από σχόλια που γίνονται είτε δημόσια, είτε μέσω αξιολογήσεων (επί παραδείγματι πλατφόρμες και ιστοσελίδες όπως *Rate my Professor* ή ακόμα και μέσω της αξιολόγησης καθηγητών/τριών που διεξάγει το εκάστοτε ίδρυμα κατά τη διάρκεια των εξαμήνων). Παρατηρώντας και αξιοποιώντας τα παραπάνω, εξάγονται συμπεράσματα, αναφορικά με το αίσθημα ικανοποίησης/ θυμού/ δυσαρέσκειας ή ακόμα και ουδέτερη στάση αναφορικά με το εκάστοτε ερώτημα.



Εικ. 1. *Τρόποι Συναισθηματικής Ανάλυσης*

Το παραπάνω σχήμα επεξηγεί τους τρόπους και τις μορφές με τις οποίες μπορεί να αναπτυχθεί η συναισθηματική ανάλυση, καθώς και τα μοντέλα τα οποία μπορούν να αξιοποιηθούν για τον εκάστοτε ερευνητικό σκοπό. Συχνά, στις παλαιότερες κυρίως δημοσιεύσεις σχετικά με τη συναισθηματική ανάλυση, το συναίσθημα λαμβάνονταν υπόψιν ως δίπολο, δηλαδή το αρνητικό και το θετικό. Πλέον, λαμβάνεται υπόψιν και το ουδέτερο (όταν δηλαδή ένα σχόλιο δεν έχει ούτε θετικό, ούτε αρνητικό πρόσημο) και υπάρχει κλίμακα διαβάθμισης του συναισθήματος για κάθε σημείο συναισθηματικού ενδιαφέροντος (έγγραφο, πρόταση ή λέξη). Η κλίμακα μπορεί να είναι από το 1 έως το 5, με το 1 να απεικονίζει το εντελώς αρνητικό συναίσθημα, 2 το λίγο αρνητικό, 3 το ουδέτερο, 4 το ελαφρώς θετικό και 5 το αρκετά θετικό συναίσθημα. Ακόμη και με την ύπαρξη ουδέτερου συναισθήματος, κατά την ταξινόμηση και αξιολόγηση των ευρημάτων, δεν λαμβάνεται υπόψιν, μιας και δεν επιφέρει σημαντικές αλλαγές στο αποτέλεσμα των αξιολογήσεων, αλλά ούτε και τα επηρεάζει με κάποιον άλλον τρόπο.

Τρόπος λειτουργίας της ανάλυσης συναισθήματος

Συλλογή Δεδομένων

Η συλλογή δεδομένων για το μοντέλο μηχανικής μάθησης γίνεται με δύο τρόπους, είτε μέσω API (Application Programming Interface), διεπαφής με τη βοήθεια της οποίας τα δεδομένα που εισάγονται σε πραγματικό χρόνο, μεταδίδονται σε έναν server κατά την εκτέλεση μιας εφαρμογής και επιστρέφονται στο χρήστη με τις ανάλογες αποκρίσεις. Παραδείγματα αυτών αποτελούν ένα βίντεο στην εφαρμογή Microsoft Teams ή μια ζωντανή ροή μετάδοσης στο Instagram, είτε ακόμη και η εισαγωγή των δεδομένων χειροκίνητα από χρήστες. Πολλές οι περιπτώσεις όπου γίνεται προσπάθεια συναισθηματικής ανάλυσης σε

γλώσσα όπου δεν έχει προηγηθεί κάποια παρόμοια έρευνα, οπότε η ερευνητική ομάδα καλείται να δημιουργήσει βάσεις και σετ δεδομένων, καθώς και γλωσσικούς κανόνες χειροκίνητα (*Nikolic et al., 2019; Dervenis et al., 2024*), ώστε να καταχωρηθούν στο μοντέλο μηχανικής μάθησης για εκπαίδευση και δοκιμή.

Επεξεργασία Δεδομένων

Η επεξεργασία δεδομένων αποτελεί ένα από τα πιο σημαντικά σημεία της συναισθηματικής ανάλυσης, μιας και οι διαδικασίες που ακολουθούν βασίζονται στην ορθή επεξεργασία των δεδομένων. Πριν τη λειτουργία της συναισθηματικής ανάλυσης, εδραιώνεται η λειτουργία της προ-επεξεργασίας του κειμένου. Μέσω αυτής της διαδικασίας, η είσοδος κειμένου περνάει μέσα στο μοντέλο μάθησης, όπου με μια πολύ συγκεκριμένη διαδικασία, μπορεί και αναγνωρίζει λέξεις μεμονωμένα ώστε να εξάγει το συναίσθημα που αναγράφεται στο κείμενο. Έστω ότι μια φοιτήτρια σε μια αξιολόγηση της γράφει : «Το μάθημα είναι ΠΟΛΥ ενδιαφέρον 😊». Το μοντέλο μάθησης, θα λάβει την παραπάνω είσοδο, και θα χωρίσει τις λέξεις σε πεδία, με τον ισχύοντα κανόνα : όσες λέξεις τόσα και τα πεδία· άρα, η παραπάνω είσοδος θα έχει 5 πεδία, ως εξής : ‘το’, ‘μάθημα’, ‘είναι’, ‘πολύ’, ‘ενδιαφέρον’. Παρατηρείται πως, για να δημιουργηθούν τα πεδία, λαμβάνουν χώρα οι παρακάτω διαδικασίες :

1. Αφαίρεση του κενού χαρακτήρα μεταξύ λέξεων
2. Μετατροπή όλων των χαρακτήρων από κεφαλαία σε πεζά γράμματα και
3. Αφαίρεση όλων των περιττών χαρακτήρων και συμβόλων

Στην τρίτη διαδικασία, αφαιρούνται όλοι οι χαρακτήρες, καθώς και τα περιττά σύμβολα (όπως emoji, «~!@#» κ.ο.κ), μιας και δεν προσφέρουν τίποτα στη διαδικασία της συναισθηματικής ανάλυσης – τις περισσότερες φορές-. Δεν είναι λίγες οι περιπτώσεις όπου οι αξιολογήσεις εμπεριέχουν και emoji, όπου τις περισσότερες φορές «κουβαλούν» συναίσθημα, το οποίο και προσδίδει άλλη βαρύτητα στη συναισθηματική πολικότητα του κειμένου, αλλά τα μοντέλα τα αφαιρούν και έτσι χάνεται η πληροφορία αυτή. Έπειτα από την παραπάνω διαδικασία, όλες οι λέξεις/πεδία έχουν ένα σύμβολο/token. Αυτά τα σύμβολα είναι χαρακτήρες, λέξεις ή ακόμη και τμήματα λέξεων. Έτσι, μέσω αυτών τον συμβόλων επιτυγχάνεται η προ-επεξεργασία κειμένου, κι επιπλέον συμβάλλουν σε διάφορες *διεργασίες επεξεργασίας φυσικής γλώσσας (Natural Language Processing / NLP)*.

Εικ. 2. Πίνακας από emoticons και η συναισθηματική τους κατάταξη

Representation	Value
Emoji Name	FACE WITH TEARS OF JOY
UTF-8 Unicode Character(s)	😂
UTF-8 Character Count	1
Decimal HTML Entity	😂
Hexadecimal HTML Entity	😂
Hexadecimal Code Point(s)	1f602
Unicode Notation	U+1F602
Decimal Code Point(s)	128514
UTF-8 Hexadecimal (C Syntax)	0xF0 0x9F 0x98 0x82
UTF-8 Hexadecimal Bytes	F0 9F 98 82
UTF-8 Octal Bytes	360 237 230 202
UTF-16 Hexadecimal (C Syntax)	0xD83D 0xDE02
UTF-16 Hexadecimal	d83dde02
UTF-16 Decimal	55357 56834
UTF-32 Hexadecimal (C Syntax)	0x0001F602
UTF-32 Hexadecimal	01F602
UTF-32 Decimal	128514
Python	u"\U0001F602"
PHP	"\xf0\x9f\x98\ x82"
C / C++ / Java	"\uD83D\uDE02"
R-Encoding	<ed><a0><bd><ed><b8><82>
Emojitracker.com rank	1
Sentiment Score	0.805100583

Εικ. 3. Αναλυτική περιγραφή του emoticon “Face with tears of joy”

```python
from nltk.tokenize import word_tokenize

sentence = "What a beautiful day. I hope tomorrow will be better."

words = word_tokenize(sentence)

for i, word in enumerate(words, start=1):
    print(f"Word {i}: {word}")
```

word_tokenize.py hosted with ♥ by GitHub view raw

```
Word 1: What
Word 2: a
Word 3: beautiful
Word 4: day
Word 5: .
Word 6: I
Word 7: hope
Word 8: tomorrow
Word 9: will
Word 10: be
Word 11: better
Word 12: .
```

Word Tokenization

Εικ. 4. *Παράδειγμα δημιουργίας ετικετών και το tokenization εν δράσει.*

```python
sentence = "I can't believe."

characters = list("".join(nltk.word_tokenize(sentence)))

def printResult(data, dataType):
    for i, item in enumerate(data, start=1):
        print(f"{dataType} {i}: {item}")

printResult(characters, "Character")
```

character_tokenize.py hosted with ♥ by GitHub view raw

```
Character 1: I
Character 2: c
Character 3: a
Character 4: n
Character 5: '
Character 6: t
Character 7: b
Character 8: e
Character 9: l
Character 10: i
Character 11: e
Character 12: v
Character 13: e
Character 14: .
```

Character Tokenization

Εικ. 5. *Παράδειγμα tokenization σε χαρακτήρες.*

Παραπάνω, βλέπουμε δύο τρόπους με τους οποίους μπορεί να γίνει το tokenization, η ανίχνευση δηλαδή και ταξινόμηση των λέξεων/χαρακτήρων στο αντίστοιχο πεδίο, όπως αναφέρεται αργότερα. Δεν είναι λίγες οι φορές που η εντολή μέσω της οποίας γίνεται ο διαχωρισμός λέξεων, αντιμετωπίζει κάθε χαρακτήρα ως μια ξεχωριστή λέξη. Αυτό το είδος συμβολοποίησης είναι ένας χρήσιμος τρόπος διαχωρισμού και επεξεργασίας λέξεων σε πολύ συγκεκριμένες περιπτώσεις. Σε ένα μεγάλο ποσοστό περιπτώσεων, όμως, είναι πολύ

δύσχρηστος και χρειάζεται παραπάνω κανόνες και διεργασίες -και συνεπώς περισσότερους υπολογιστικούς πόρους- για να γίνει χρήσιμος και λειτουργικός.

Έπειτα από τη διαδικασία της συμβολοποίησης, τα δεδομένα είναι πλέον σε μορφή έτοιμα προς επεξεργασία, ώστε να αρχίσει η επεξεργασία του κειμένου. Αυτή, επίσης, χωρίζεται σε διάφορα στάδια με την εξής σειρά : συλλογή, επεξεργασία, ανάλυση και απεικόνιση δεδομένων. Παρακάτω, βλέπουμε τη διαδικασία συμβολοποίησης *Part-of-speech (POS-tagging)*. Ένα παράδειγμα πάνω σε αυτό είναι η εξής πρόταση : [(‘That’, ‘DT’), (‘movie’, ‘NN’), (‘was’, ‘VBD’), (‘a’, ‘DT’), (‘colossal’, ‘JJ’), (‘disaster’, ‘NN’), (‘I’, ‘PRP’), (‘absolutely’, ‘RB’), (‘hated’, ‘VBD’), (‘it’, ‘PRP’), (‘Waste’, ‘NN’) , (‘of’, ‘IN’), (‘time’, ‘NN’), (‘and’, ‘CC’), (‘money’, ‘NN’)]*https://careerfoundry.com/en/blog/data-analytics/sentiment-analysis/*, καθώς και οι παρακάτω ετικέτες αναλυτικότερα.

Tag	Description
ADJ	Adjective
ADV	Adposition
ADP	Adverb
AUX	Auxiliary
CCONJ	Coordinating Conjunction
DET	Determiner
INTJ	Interjection
NOUN	Noun
NUM	Numeral
PART	Particle
PRON	Pronoun
PROPN	Proper Noun
PUNCT	Punctuation
SCONJ	Subordinating Conjunction
SYM	Symbol
VERB	Verb
X	Other

Εικ. 6. *Ετικέτες στη συμβολοποίηση POS.*

Η συμβολοποίηση POS είναι μια διαδικασία μέσα από την οποία οι λέξεις δέχονται μια «ετικέτα», ανάλογα τον τύπο της λέξης (ρήμα, επίθετο κτλ). Έτσι, οι λέξεις παίρνουν την μορφή που είδαμε προηγουμένως, όπως η (‘colossal’, ‘JJ’)· είναι δηλαδή της μορφής (‘word’, ‘type’).

Στη συνέχεια, έπεται η επεξεργασία δεδομένων, και βασίζεται στο είδος των δεδομένων τα οποία καλείται να αναλύσει, μιας και αυτά ποικίλουν (ήχος, εικόνα, βίντεο ή απλό κείμενο). Με την επεξεργασία κειμένου να είναι η πιο απλή από τις μορφές και του βίντεο να είναι η πιο σύνθετη, λόγω του φόρτου και της πολυπλοκότητας της επεξεργασίας αυτού, πόσο μάλλον βίντεο σε ζωντανή μετάδοση.

Στην περίπτωση που τα δεδομένα μας είναι σε ηχητική μορφή, αυτά πρέπει να απομαγνητοφωνηθούν και να μετατραπούν σε γραπτή μορφή, ώστε να μπορούν να εισαχθούν στο μοντέλο μηχανικής μάθησης. Εάν πάλι τα δεδομένα είναι σε μορφή εικόνας, τότε κρίνεται απαραίτητη η χρήση εργαλείων που αναλύουν και μεταφράζουν την εικόνα σε λέξεις. Η μετάφραση της εικόνας σε λέξεις ονομάζεται *image capturing* και μπορεί να υλοποιηθεί με δύο τρόπους : την top-down και την bottom-up μέθοδο· η πρώτη, ξεκινάει από την άντληση πληροφοριών μέσω εικόνων, μετατρέποντας τες σε διάφορες λέξεις και η δεύτερη

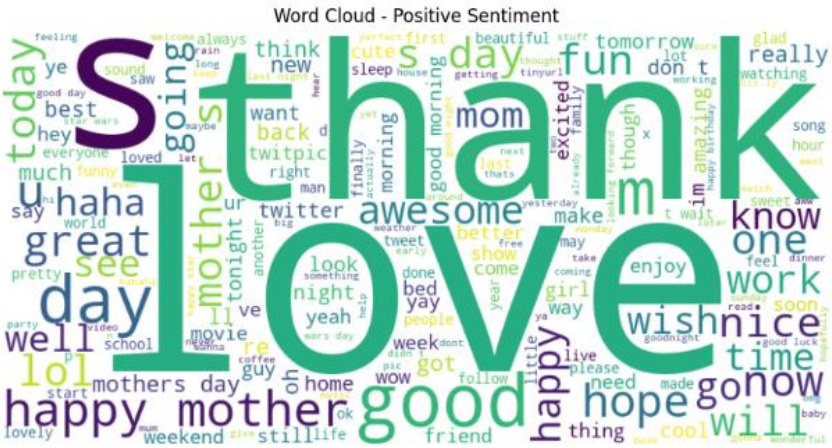
μεθοδολογία εκτιμά τις λέξεις που περιγράφουν καλύτερα μια εικόνα, όπου μέσω μιας διαδικασίας συνένωσης, παράγει προτάσεις.

Η επόμενη φάση συναισθηματικής ανάλυσης είναι η ανάλυση των αποκτηθέντων δεδομένων. Ξεκινάμε με την εκπαίδευση του μοντέλου: σε αυτή τη διαδικασία χωρίζουμε τα δεδομένα μας σε δύο άνισα σύνολα, ένα μεγάλο σύνολο για την εκπαίδευση του μοντέλου και ένα μικρότερο σύνολο για την δοκιμή του μοντέλου. Η κλίμακα με την οποία θα χωρίζονται τα σύνολα διαφέρει από έρευνα σε έρευνα, αλλά οι δύο άρχουσες κλίμακες είναι αυτές με λόγο 70:30 και κυριότερη αυτή της τάξης του 80:20, δηλαδή το 80% των δεδομένων θα αξιοποιηθεί στην εκπαίδευση του μοντέλου, και το 20% στη δοκιμή του μοντέλου.

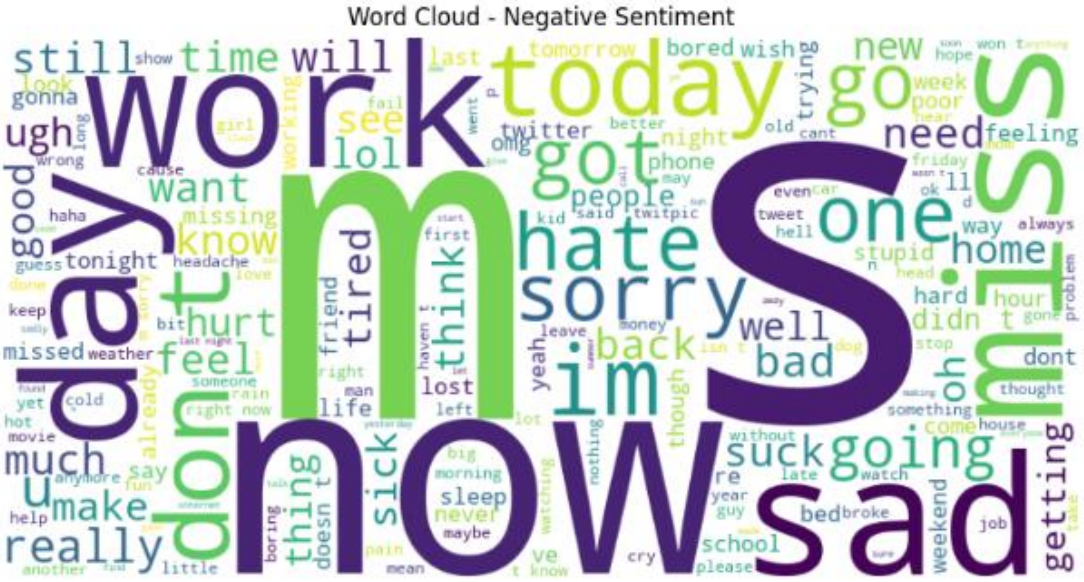
Επόμενη διεργασία είναι η γλωσσική επεξεργασία και ερμηνεία. Στο μεγαλύτερο μέρος τους, οι έρευνες έγιναν είτε σε αγγλικά ιδρύματα είτε στην αγγλική γλώσσα. Έτσι, χώρες που δεν έχουν την αγγλική γλώσσα ως κύρια γλώσσα, υστερούν στη βιβλιογραφία πάνω σε αυτό το θέμα, αλλά και στο ότι για να κάνουν μια έρευνα πάνω σε αυτό το κομμάτι πρέπει να δημιουργήσουν και να αναπτύξουν δικά τους μοντέλα για τη δική τους ομιλούμενη γλώσσα, καθώς και σετ δεδομένων στη συγκεκριμένη γλώσσα, ώστε να μπορεί να γίνει λειτουργική η χρήση του μοντέλου στη χώρα αυτή. Παραδείγματος χάριν, για να δημιουργήσουμε ένα μοντέλο μηχανικής μάθησης που υλοποιεί συναισθηματική ανάλυση στα ελληνικά, πρέπει να ψάξουμε εάν υφίσταται ήδη ένα σετ μοντέλων που να υλοποιεί αυτή τη διαδικασία. Εάν όχι, τότε θα πρέπει να δημιουργηθούν εκ νέου μοντέλα μηχανικής μάθησης για συναισθηματική ανάλυση που μπορούν να αναγνωρίζουν και να υποστηρίζουν το ελληνικό αλφάβητο, να εκπαιδευτεί το μοντέλο με σετ δεδομένων της ελληνικής γλώσσας, ώστε να μπορεί να δοκιμαστεί η απόδοσή του πάνω σε συστάδες κειμένων γραμμένες στα ελληνικά.

Συνεχίζοντας έχουμε τις εξής δύο ιδιαίτερες διαδικασίες, προσαρμοσμένα σύμβολα στο εκάστοτε θέμα μαζί με σετ δεδομένων, και την κατάταξη του θέματος που θα αναλυθεί. Δηλαδή, υπάρχουν συγκεκριμένες λέξεις, οι οποίες είτε λόγω της συναισθηματικής βαρύτητάς τους είτε λόγω της ιδιαιτερότητάς τους, για την λεκτική ανάλυση, λαμβάνουν κάποιες συγκεκριμένες ετικέτες, ώστε να διευκολυνθεί η ανάλυσή τους.

Τελευταίο κομμάτι της ανάλυσης των δεδομένων είναι η απεικόνιση των αποτελεσμάτων, δηλαδή η οπτικοποίηση του αποτελέσματος της επεξεργασίας των δεδομένων ως προς διάφορους παράγοντες που επιλέγονται στην εκάστοτε έρευνα. Αυτό, περιλαμβάνει μια εικόνα που έχει ένα σύνολο λέξεων και με διάφορα μεγέθη απεικονίζεται η συχνότητα που η λέξη αυτή εμφανίστηκε-όσο πιο μεγάλη η λέξη τόσο πιο συχνή η εμφάνισή της-.



Εικ. 7. *Απεικόνιση των θετικών συναισθημάτων μιας έρευνας.*



Εικ. 8. *Απεικόνιση των αρνητικών συναισθημάτων μιας έρευνας.*

Οι παραπάνω εικόνες αποτελούν παράδειγμα της οπτικοποίησης των δεδομένων από μια έρευνα, όπου παρουσιάζονται τα αποτελέσματα συναισθηματικής ανάλυσης μιας έρευνας σχετικά με την εκπαιδευτική διαδικασία διαδικτυακά. Παρατηρείται ότι έχουν εμφανιστεί πολλές λέξεις πάνω στον πίνακα , οι οποίες όλες έλαβαν μέρος στην ανάλυση των δεδομένων σχετικά με τη συναισθηματική βαθμολογία, και συνέβαλαν με βαθμό ανάλογο του μεγέθους τους.

ΚΕΦΑΛΑΙΟ 2 Βιβλιογραφική Επισκόπηση

Σε αυτό το κεφάλαιο, παρουσιάζεται η βιβλιογραφική επισκόπηση που έγινε, σε έρευνες και βιβλιοθήκες από διάφορα επιστημονικά πεδία και ποικίλες ερευνητικές ομάδες, με σκοπό την απόκτηση σφαιρικής και ετερογενούς άποψης σχετικά με το θέμα. Οι κύριοι άξονες όπου κι έγινε η έρευνα ήταν οι εξής : Εκπαίδευση (διαδικτυακά), όπου αποτελείται από 16 έρευνες, ενδεικτικά: *Spatiotis et al., 2016; Nikolić et al., 2020; Kastrati et al., 2021; Senadhira et al., 2022; Us et al., 2022; Wen et al., 2023; Lasri et al., 2023; Dervenis et al., 2024*) , Κοινωνικά Μέσα Δικτύωσης, όπου αποτελείται από 8 έρευνες, ενδεικτικά: *Rokkou et al., 2021; Saemani & Setiyawati, 2022; Wen et al., 2023; Lasri et al., 2023;*), και η πανδημία του κορωναϊού που αποτελείται από τις εξής 3 έρευνες *Senadhira et al., 2022; Remali et al., 2022; Lasri et al., 2023* (Πίνακας 1).

Οι Kumar και Teeja (2014) εστιάζουν σε μια διεξοδική επισκόπηση της ανάλυσης συναισθήματος, περιγράφοντας την ανάπτυξή της, τη σημερινή της κατάσταση και την πιθανή μελλοντική της δυνότητα. Η εξόρυξη γνώμης, μια άλλη ονομασία για την ανάλυση συναισθήματος, ορίζεται ως η υπολογιστική εξέταση των σκέψεων, των συναισθημάτων και των συναισθημάτων των ανθρώπων, όπως αυτά εκφράζονται σε γραπτή γλώσσα. Παρουσιάζοντας τις απαρχές της και τις πρώιμες εργασίες, αποδεικνύεται πως οι πρώτες εργασίες στους τομείς της ανάλυσης κειμένου και της επεξεργασίας φυσικής γλώσσας (NLP) είναι το σημείο όπου πρωτοεμφανίστηκε η ανάλυση συναισθήματος. Αρχικά, δόθηκε έμφαση στην ταξινόμηση του συναισθήματος σε επίπεδο εγγράφου. Οι κύριες πηγές πληροφοριών για τις πρώτες μεθόδους ήταν τα λεξικά και οι αλγόριθμοι μηχανικής μάθησης. Δημοφιλείς μέθοδοι που χρησιμοποιήθηκαν ήταν η μέγιστη εντροπία, οι μηχανές διανυσμάτων υποστήριξης (SVM) και οι naive bayes. Προβλήματα που παρουσιάστηκαν κατά την καταγραφή της μελέτης απέδειξαν πως η ανθρώπινη γλώσσα είναι τόσο ευρεία, μπορεί να είναι δύσκολο να διακρίνει κανείς τον σαρκασμό και την ειρωνεία στο γραπτό περιεχόμενο. Η ακριβής κατανόηση του πλαισίου εξαρτάται από τη σαφή τοποθέτηση στο πλαίσιο όπου μεταφέρεται ένα συναίσθημα. Σημαντικό πρόβλημα είναι η πολυγλωσσική ανάλυση συναισθήματος καθώς η ανάπτυξη πολύγλωσσων και διαγλωσσικών μοντέλων είναι απαραίτητη και απαιτητική επειδή η ανάλυση συναισθημάτων σε πολλές γλώσσες προσθέτει ένα άλλο επίπεδο πολυπλοκότητας.

Στην παρακάτω έρευνα, διερευνάται η ανάλυση συναισθήματος που είναι ευαίσθητη ως προς το φύλο στα πλαίσια της διαδικτυακής εκπαίδευσης. Η κατανόηση του συναισθηματικού κλίματος σε αυτά τα πλαίσια είναι ζωτικής σημασίας καθώς οι διαδικτυακές πλατφόρμες μάθησης πολλαπλασιάζονται με ταχύτατους ρυθμούς. Η σημασία της συνεκτίμησης της γλώσσας και των συναισθημάτων που σχετίζονται με το φύλο στην ανάλυση συναισθήματος αναγνωρίζεται από τις ερευνητικές ομάδες, ιδίως σε εκπαιδευτικά περιβάλλοντα όπου η δυναμική του φύλου μπορεί να έχει σημαντικό αντίκτυπο. Στόχος της ομάδας αυτής είναι η δημιουργία ενός αλγορίθμου ανάλυσης συναισθήματος που να λαμβάνει υπόψη τη γλώσσα και τις συναισθηματικές εκφράσεις που είναι μοναδικές για κάθε φύλο. Αυτό υλοποιείται μέσω μιας προσπάθειας να πλαστεί μια πιο ολοκληρωμένη κατανόηση του συναισθηματικού περιβάλλοντος στην διαδικτυακή εκπαίδευση των εκπαιδευτικών. Η μελέτη αυτή, επιπλέον, εξετάζει τον τρόπο δημιουργίας του μοντέλου αυτού, συμπεριλαμβανομένου του τρόπου εύρεσης γλωσσικών δεικτών που είναι χαρακτηριστικοί για ένα συγκεκριμένο φύλο και του τρόπου ανάλυσης των συναισθηματικών εκφράσεων στον εκπαιδευτικό λόγο. Τελικά, η έρευνά αυτή προάγει γνώσεις για το πώς η ανάλυση συναισθήματος μπορεί να αναπαραστήσει με ακρίβεια το φάσμα των συναισθηματικών εμπειριών που έχουν οι

εκπαιδευτικοί όταν εργάζονται σε εικονικά περιβάλλοντα, με έμφαση στις λεπτές αποχρώσεις που αφορούν το φύλο.

Η ανάλυση συναισθήματος στα μέσα κοινωνικής δικτύωσης

Στην έρευνα της Lasri παρουσιάζεται μια μοναδική μέθοδος για την ανάλυση συναισθήματος σε tweets (δημοσιεύσεις δηλαδή που αναρτώνται στον ιστότοπο "Twitter"/ πλέον "X") σχετικά με την εξ αποστάσεως μάθηση στην τριτοβάθμια εκπαίδευση. Προτείνεται ένα μοντέλο που συνδυάζει μια αρχιτεκτονική αμφίδρομης μακράς βραχυπρόθεσμης μνήμης (*Bidirectional Long Short Term Memory* / *Bi-LSTM*) με μηχανισμούς αυτοπροσοχής/ self-attention based (είναι μια πρακτική που αξιοποιείται στη μηχανική μάθηση που ανιχνεύει και αξιοποιεί τις εισόδους ως προς το ποσοστό διασύνδεσης κι αλληλεπίδρασής τους). Στόχος αυτού του μοντέλου είναι να αναπαραστήσει τόσο τη διαδοχική δομή των δεδομένων κειμένου όσο και τη σημασία μεμονωμένων λέξεων ή φράσεων στα tweets. Ο αλγόριθμος μπορεί να σταθμίζει δυναμικά τη σημασία των διαφόρων όρων χρησιμοποιώντας την αυτοπροσοχή, γεγονός που βελτιώνει την κατανόηση του συναισθήματος που μεταφέρεται στα tweets. Μέσω δοκιμών σε ένα σύνολο δεδομένων που αντλήθηκαν από tweets σχετικά με την εξ αποστάσεως εκπαίδευση, η έρευνα αυτή δείχνει την αποτελεσματικότητα αυτής της στρατηγικής και παρουσιάζει αυξημένες επιδόσεις ανάλυσης συναισθήματος σε σύγκριση με παλαιότερες μεθόδους. *(Lasri et al., 2023).*

Στην επόμενη έρευνα της θεματικής αυτής, παρουσιάζεται η ενασχόληση με ένα σύστημα ανάλυσης συναισθήματος σε πραγματικό χρόνο στο Twitter, προσαρμοσμένο για πανεπιστήμια του Μαρόκο. Η εμφάνιση των μέσων κοινωνικής δικτύωσης και η αφθονία του περιεχομένου που δημιουργείται από τους χρήστες, ιδίως σε ιστότοπους όπως το Twitter, προσφέρουν την ευκαιρία να εξεταστούν οι απόψεις για τα ιδρύματα σε πραγματικό χρόνο. Η ερευνητική ομάδα προτείνει ένα σύστημα για τη διενέργεια ανάλυσης συναισθήματος σε tweets σχετικά με μαροκινά ιδρύματα με τη χρήση τεχνολογίας μεγάλων δεδομένων και μηχανικής μάθησης. Το σύστημα είναι σε θέση να κατηγοριοποιεί αυτόματα τα tweets σε θετικές, αρνητικές και ουδέτερες κριτικές χρησιμοποιώντας τεχνικές μηχανικής μάθησης. Οι τεχνολογίες μεγάλων δεδομένων χρησιμοποιούνται επίσης για τη διαχείριση του τεράστιου όγκου των tweets σε πραγματικό χρόνο, επιτρέποντας την άμεση ανάλυση αυτών. Η αρχιτεκτονική συστήματος που προτείνεται στη συγκεκριμένη έρευνα, περιλαμβάνει τα μοντέλα μηχανικής μάθησης που χρησιμοποιούνται, το μέσο επεξεργασίας δεδομένων και των στοιχείων της ανάλυσης συναισθήματος σε πραγματικό χρόνο. Ο στόχος είναι να παρέχει στους/στις ενδιαφερόμενους/ες των πανεπιστημίων διορατικές πληροφορίες σχετικά με το πώς αισθάνεται το κοινό για τα εκάστοτε ιδρύματά, στο Twitter. Μέσω δοκιμών, το άρθρο αναδεικνύει την αποτελεσματικότητα του συστήματος και υπογραμμίζει τις δυνατότητές του για την παρακολούθηση και τη βελτίωση της δημόσιας αντίληψης και φήμης των μαροκινών πανεπιστημίων. Η παρούσα έρευνα τονίζει την εστίασή της στην αξιοποίηση των μεγάλων δεδομένων/*big data* και της μηχανικής μάθησης για τη δημιουργία ενός συστήματος ανάλυσης συναισθήματος για την πλατφόρμα Twitter πάνω σε δεδομένα πραγματικού χρόνου.

Η συγκεκριμένη έρευνα αφορά την τριτοβάθμια εκπαίδευση στην Ινδονησία και πιο συγκεκριμένα το νόμο του Υπουργείου Παιδείας και Πολιτισμού της χώρας για την πρόληψη σεξουαλικής κακοποίησης στην τριτοβάθμια εκπαίδευση. Οι ερευνητές ανέλυσαν τις απόψεις που εκφράστηκαν σχετικά με αυτόν τον κανόνα χρησιμοποιώντας μια τεχνική διανυσματικής

μηχανής υποστήριξης (SVM). Το Permendikbud επιδιώκει να αντιμετωπίσει και να μειώσει τα προβλήματα που σχετίζονται με τη σεξουαλική βία στην τριτοβάθμια εκπαίδευση. Στόχος της μελέτης είναι η αυτόματη ταξινόμηση των συναισθημάτων από δεδομένα κειμένου σε θετικές, αρνητικές και ουδέτερες κατηγορίες με την εφαρμογή της SVM, μιας τεχνικής μηχανικής μάθησης για προβλήματα ταξινόμησης. Το σύνολο δεδομένων που χρησιμοποιήθηκε για την ανάλυση συναισθήματος περιελάμβανε σχόλια, απόψεις ή απαντήσεις που σχετίζονται με το Permendikbud που έγιναν σε διάφορους ιστότοπους, συμπεριλαμβανομένων φόρουμ, μέσων κοινωνικής δικτύωσης και επίσημων δηλώσεων. Τα ευρήματα της ανάλυσης συναισθήματος ρίχνουν φως στην κοινή γνώμη σε γενικό επίπεδο, καθώς και στις προοπτικές σχετικά με την αποτελεσματικότητα, την εφαρμογή και την επιρροή του κανονισμού στη σεξουαλική κακοποίηση σε περιβάλλοντα τριτοβάθμιας εκπαίδευσης. Τα αποτελέσματα της έρευνας έχουν σκοπό να παράσχουν στους νομοθέτες και τα ακαδημαϊκά ιδρύματα διορατικά σχόλια σχετικά με τον τρόπο βελτίωσης των εκστρατειών και των τακτικών τους για την πρόληψη των σεξουαλικών επιθέσεων στις πανεπιστημιουπόλεις.

Συνεχίζοντας στο πρίσμα των διεθνών εκπαιδευτικών συστημάτων γνωστοποιείται ένα αμφιλεγόμενο θέμα στην Κίνα που αφορά τις βαθμολογίες στην παγκόσμια κατάταξη πανεπιστημίων (*World University Rankings/ WUR*) που λαμβάνουν τα πανεπιστήμια της Κίνας. Στη μελέτη αυτή, προσφέρεται μια βιώσιμη προοπτική της τριτοβάθμιας εκπαίδευσης στο θέμα αυτό. Το παραπάνω επιτυγχάνεται μέσω της μοντελοποίησης θεμάτων με Latent Dirichlet Allocation (LDA) για να αυτό το περίπλοκο πρόβλημα. Παρά τη σημασία της για το παγκόσμιο τοπίο της τριτοβάθμιας εκπαίδευσης, η Παγκόσμια Κατάταξη Πανεπιστημίων μπορεί να είναι προβληματική και αντιφατική, ιδίως όταν πρόκειται για κινεζικά πανεπιστήμια. Τα δεδομένα σχετικά με την WUR και τα κινεζικά πανεπιστήμια συγκεντρώθηκαν από την ερευνητική ομάδα από διάφορες πηγές, συμπεριλαμβανομένων επιστημονικών δημοσιεύσεων, ειδήσεων και συνομιλιών στα μέσα κοινωνικής δικτύωσης. Η μελέτη χρησιμοποίησε την ανάλυση συναισθήματος για να διαπιστώσει ποιες ήταν οι σκέψεις των ανθρώπων -είτε θετικές, είτε αρνητικές, είτε ουδέτερες- σχετικά με το WUR και τα κινεζικά πανεπιστήμια. Ο εντοπισμός σημαντικών θεμάτων γύρω από το WUR και την κινεζική τριτοβάθμια εκπαίδευση βοηθήθηκε από τη μοντελοποίηση θεμάτων LDA. Τα αποτελέσματα της ανάλυσης συναισθήματος και της προαναφερθείσας μοντελοποίησης καλύπτονται στην εργασία, προσφέροντας πληροφορίες για την κοινή γνώμη, τις δυσκολίες και τις βιώσιμες λύσεις για τα κινεζικά κολέγια που επιθυμούν να αυξήσουν την κατάταξή τους χωρίς να θυσιάσουν τη βιωσιμότητα. Η έρευνα αυτή σπεύδει να δώσει μια ευρύτερη γνώση των περιπλοκών γύρω από τη WUR στην Κίνα και να παράσχει διορατικές πληροφορίες για τους/τις ενδιαφερόμενους/ες στον κλάδο της τριτοβάθμιας εκπαίδευσης, τους/τις εκπαιδευτικούς και τους/τις υπεύθυνους/ες χάραξης πολιτικής μαζί με την ενσωμάτωση της ανάλυσης συναισθήματος μέσω LDA.

Η ερευνητική ομάδα των *Pooja και Bhalla (2022)* εξετάζεται η ανάλυση συναισθήματος στο πλαίσιο της εκπαίδευσης υψηλής ποιότητας. Η ανάλυση συναισθήματος μπορεί να χρησιμοποιηθεί για να μάθουμε περισσότερα των απόψεων και των συναισθημάτων των ενδιαφερομένων στον τομέα της εκπαίδευσης, συμπεριλαμβανομένων των εκπαιδευτικών, των μαθητών και των γονέων. Η μελέτη καλύπτει μεγάλο εύρος χρήσεων της ανάλυσης συναισθήματος, συμπεριλαμβανομένης της εξέτασης των σχολίων των μαθητών, της αξιολόγησης της αποτελεσματικότητας των καθηγητών, του υπολογισμού της συμμετοχής των μαθητών και του προσδιορισμού του τρόπου με τον οποίο το ευρύ κοινό αισθάνεται για τα πανεπιστήμια. Προκειμένου να συγκεντρωθούν μελέτες περιπτώσεων και περιπτώσεις χρήσης της ανάλυσης συναισθήματος σε εκπαιδευτικά πλαίσια, η ερευνητική ομάδα εξέτασε την ήδη υπάρχουσα βιβλιογραφία. Συνολικά, η έρευνα παρέχει μια ολοκληρωμένη επισκόπηση του

τρόπου με τον οποίο η ανάλυση συναισθήματος μπορεί να συμβάλει στη βελτίωση της ποιότητας της εκπαίδευσης με την αξιοποίηση των πληροφοριών από τα δεδομένα συναισθήματος.

Η μελέτη της ερευνητικής ομάδας του Nicolic επικεντρώνεται στην ανάλυση συναισθήματος με βάση τις πτυχές των αξιολογήσεων της τριτοβάθμιας εκπαίδευσης στη σερβική γλώσσα. Η απαρχή αυτής προήλθε από: την αξιολόγηση του συναισθήματος που εντοπίζεται σε κριτικές για την τριτοβάθμια εκπαίδευση, με ιδιαίτερη έμφαση στις εγκαταστάσεις, την υποστήριξη των φοιτητών/τριών, την ποιότητα της διδασκαλίας και το περιεχόμενο των μαθημάτων. Η ερευνητική ομάδα χρησιμοποιεί προσεγγίσεις ανάλυσης συναισθήματος με βάση τις αξιολογήσεις, οι οποίες συνεπάγονται την κατάτμηση αυτών σε διάφορες ομάδες και στη συνέχεια την εξακρίβωση του συναισθήματος (θετικό, αρνητικό ή ουδέτερο) που συνδέεται με το εκάστοτε στοιχείο. Οι αξιολογήσεις συλλέχθηκαν και επεξεργάστηκαν για ανάλυση από διάφορες πηγές που σχετίζονται με την τριτοβάθμια εκπαίδευση στη Σερβία και κύρια πηγή το "http://oceniprofesora.com", ο αντίστοιχος σερβικός ιστότοπος του *Rate My Professor*. Οι αξιολογήσεις αυτές αφορούν ποικίλα θέματα και προσφέρουν μια εμπεριστατωμένη εικόνα του τοπίου των συναισθημάτων. Στο πλαίσιο της τριτοβάθμιας εκπαίδευσης, όπου οι αξιολογήσεις διαφέρουν και καλύπτουν ένα εύρος θεμάτων, εξετάζονται οι δυσκολίες στον εντοπισμό συγκεκριμένων χαρακτηριστικών μέσα στις αξιολογήσεις. Η συγγραφική ομάδα κατηγοριοποιεί το συναίσθημα που συνδέεται με κάθε ανακαλυφθείσα πτυχή εφαρμόζοντας τεχνικές ανάλυσης συναισθήματος. Η κατηγοριοποίηση αυτή βοηθά στον προσδιορισμό των χαρακτηριστικών που κρίνονται ελκυστικά ή μη ελκυστικά από μια εξεταστική επιτροπή. Η παραπάνω έρευνα φτάνει στο συμπέρασμα πως εκμαιεύονται συμπεράσματα σχετικά με τα γενικά μοτίβα συναισθήματος στις κριτικές της τριτοβάθμιας εκπαίδευσης. Με βάση τις εκάστοτε απόψεις της κριτικής επιτροπής, που είναι σε θέση να εντοπίσει τους δυνατούς τομείς των ιδρυμάτων τριτοβάθμιας εκπαίδευσης, αλλά και τους τομείς όπου το κάθε ίδρυμα υστερεί, ώστε να τεθεί προς ανάπτυξη. Τα ιδρύματα τριτοβάθμιας εκπαίδευσης μπορούν να βρουν αξία στα ευρήματα αυτής της έρευνας.

Συνεχίζοντας την βιβλιογραφική επισκόπηση και περνώντας στο κομμάτι των τεχνικών επεξεργασίας φυσικής γλώσσας (NLP), η συγκεκριμένη έρευνα τις χρησιμοποιεί για την ανάλυση των αξιολογήσεων των φοιτητών, προκειμένου να διερευνήσει τα βασικά στοιχεία που καθιστούν τα μαζικά ανοικτά διαδικτυακά μαθήματα/ MOOC (όπως Udemy, Coursera, Domestika κτλ) επιτυχημένα. Η έρευνα αυτή αποκαλύπτει μια σειρά από κρίσιμα στοιχεία που είναι απαραίτητα, όπως δέσμευση, καθώς τα MOOC που κρατούν με επιτυχία την προσοχή των συμμετεχόντων και τους παρακινούν καθ' όλη τη διάρκεια του μαθήματος θεωρούνται επιτυχημένα. Αλληλεπιδραστικότητα επειδή οι μαθητές/τριες είναι πιο πιθανό να προσελκύονται σε μαθήματα που περιλαμβάνουν διαδραστικά στοιχεία όπως συζητήσεις, εργασίες και κουίζ. Οι θετικές κριτικές και η ικανοποίηση των εκπαιδευομένων συσχετίζονται άμεσα με την ικανότητα των εκπαιδευτών/τριών να επικοινωνούν με σαφή και συνοπτικό τρόπο. Δέσμευση του/της εκπαιδευτικού: Οι καθηγητές/τριες που συμμετέχουν ενεργά στις συζητήσεις της τάξης, παρέχουν έγκαιρη ανατροφοδότηση και απαντούν σε ερωτήματα των φοιτητών/τριων χαίρουν μεγάλης εκτίμησης. Το μέγεθος των πόρων καθιστά ένα μαθησιακό περιβάλλον αποτελεσματικό και εξαρτάται από την ύπαρξη αναγνωσμάτων, βίντεο και πρόσθετων πόρων του υψηλότερου διαμετρήματος, ώστε να εγείρει και να επιστήσει το ενδιαφέρον της μαθητικής του κοινότητας. Κοινότητα αφορά την αίσθηση του «ανήκειν» στην τάξη, όπου οι μαθητές/τριες μπορούν να συμμετέχουν και να συνεργάζονται μεταξύ τους, κάνοντας την εμπειρία πιο ευχάριστη. Ευελιξία, μιας και οι εκπαιδευόμενοι/ες δίνουν στα MOOC που παρέχουν ευέλικτο χρονοδιάγραμμα και επιλογές ρυθμού ευνοϊκές κριτικές.

Εφαρμογές στον πραγματικό κόσμο, όπου η μαθητική κοινότητα εκτιμά τα μαθήματα που κάνουν τη σύνδεση μεταξύ αφηρημένων ιδεών και πραγματικών, εργασιακών καταστάσεων. Αποδεικνύεται, συμπερασματικά, πως τα ακαδημαϊκά προγράμματα εκτιμώνται ιδιαίτερα όταν καλλιεργούν ένα φιλόξενο και ενθαρρυντικό περιβάλλον. Οι συχνές ενημερώσεις σχετικά με νέο περιεχόμενο επιφέρει διατήρηση αλλά ακόμα και αύξηση στα επίπεδα ενδιαφέροντος και συνδιαλλαγής με το εκπαιδευτικό προσωπικό.

Ο συνδυασμός της μηχανικής μάθησης με την ανάλυση συναισθήματος στα εκπαιδευτικά συστήματα εξετάζεται στη μελέτη της Deepa. Με την εξέταση της ανατροφοδότησης που λήφθηκε για τον εντοπισμό συναισθηματικών αντιδράσεων και περιοχών που χρήζουν ανάπτυξης, επιδιώκεται η βελτίωση των εκπαιδευτικών εμπειριών των μαθητών. Το σύστημα επεξεργάζεται και αναλύει την ανατροφοδότηση όπου λήφθηκε από ανατροφοδότηση των μαθητών μέσω γραπτών αξιολογήσεων, χρησιμοποιώντας τεχνικές επεξεργασίας φυσικής γλώσσας (NLP) και βαθιάς μάθησης, παρέχοντας πληροφορίες σχετικά με το συναίσθημα των μαθητών που θα μπορούσαν να ενημερώσουν για εκπαιδευτικές μεταρρυθμίσεις και προσαρμοσμένες στρατηγικές μάθησης.

Μια διεξοδική ανάλυση της χρήσης της ανάλυσης συναισθήματος σε εκπαιδευτικά περιβάλλοντα αναλύεται στη μελέτη του Dalipi, με στόχο τη διεξοδική κατηγοριοποίηση της βιβλιογραφίας και των ευρημάτων σχετικά με την ανάλυση της ανατροφοδότησης των φοιτητών με τη χρήση της βαθιάς μάθησης (*Deep Learning/ DL*), της μηχανικής μάθησης (*Machine Learning/ MC*) και της επεξεργασίας φυσικής γλώσσας (*NLP*). Τα σημαντικά ευρήματα της έρευνας περιλαμβάνουν το πεδίο εφαρμογής και τη μεθοδολογία αξιοποιώντας το πλαίσιο PRISMA, μέσω του οποίου οι συγγραφείς πραγματοποίησαν συστηματική ανάλυση χαρτογράφησης με έμφαση στις δημοσιεύσεις που δημοσιεύθηκαν μεταξύ 2015 και 2020. Από την αρχική συλλογή από 612 μελετών, εξέτασαν 92 σχετικές μελέτες. Τα ευρήματα αποκαλύπτουν ότι η χρήση των προσεγγίσεων Deep Learning στην ανάλυση συναισθήματος έχει αυξηθεί κατακόρυφα, γεγονός που υποδηλώνει ότι πρόκειται για μια πρόσφατη τάση. Ακόμα και με αυτή την επέκταση, ο χειρισμός του τεράστιου όγκου και του ποικίλου φάσματος των δεδομένων ανατροφοδότησης των φοιτητών εξακολουθεί να παρουσιάζει δυσκολίες. Δυσκολίες που αντιμετωπίζουν και προτάσεις που στοχεύουν στη βελτίωση του συγκεκριμένου πεδίου, μιας και μελέτη υποδεικνύει ορισμένες σημαντικές κατευθύνσεις για περαιτέρω έρευνα, συμπεριλαμβανομένης της απαίτησης για τυποποιημένο λογισμικό ανάλυσης συναισθήματος, οργανωμένα σύνολα δεδομένων και βελτιωμένες τεχνικές για την έκφραση και τον εντοπισμό συναισθημάτων.

Η μελέτη της Us εξετάζει τον τρόπο με τον οποίο τα κολέγια μπορούν να χρησιμοποιήσουν τα μέσα κοινωνικής δικτύωσης για να προωθήσουν την περιβαλλοντικά συνειδητοποιημένη εικόνα τους. Τα πανεπιστήμια δίνουν όλο και μεγαλύτερη έμφαση στην περιβαλλοντικής βιωσιμότητα λόγω των αυξανόμενων περιβαλλοντικών ανησυχιών και της σημασίας της βιωσιμότητας. Η μελέτη αξιολογεί πόσο καλά λειτουργούν τα μέσα κοινωνικής δικτύωσης για την προώθηση των "πράσινων" εμπορικών σημάτων των πανεπιστημίων χρησιμοποιώντας εξόρυξη κειμένου και ανάλυση συναισθήματος. Στόχος της παρούσας μελέτης είναι να εξετάσει το περιεχόμενο των μέσων κοινωνικής δικτύωσης που αφορά την "πράσινη" επωνυμία των πανεπιστημίων, να αξιολογήσει την κοινή γνώμη σχετικά με αυτές τις πρωτοβουλίες και να προσφέρει προτάσεις για τη βελτίωση του μάρκετινγκ της πράσινης επωνυμίας των πανεπιστημίων μέσω των μέσων κοινωνικής δικτύωσης. Για να επιτύχουν τους στόχους τους, συγγραφική ομάδα χρησιμοποίησε τεχνικές ανάλυσης συναισθήματος και εξόρυξης κειμένου χρησιμοποιώντας δεδομένα από τα μέσα κοινωνικής δικτύωσης. Για να προσδιοριστεί η αντίληψη και η στάση του κοινού απέναντι στις δραστηριότητες πράσινης

επωνυμίας, έγινε ανάλυση συναισθήματος. Σύμφωνα με την ερευνητική ομάδα, η πράσινη επωνυμία των πανεπιστημίων προβάλλεται συχνά στα μέσα κοινωνικής δικτύωσης μέσω αναρτήσεων που εστιάζουν σε ορισμένα θέματα, συμπεριλαμβανομένων των προγραμμάτων περιβαλλοντικής εκπαίδευσης, των φιλικών προς το περιβάλλον πρακτικών και των έργων βιωσιμότητας. Στο ευρύτερο σύνολο των ευρυμάτων διαπιστώθηκε ότι οι άνθρωποι αισθάνονται θετικά για τις δραστηριότητες πράσινης επωνυμίας των πανεπιστημίων. Παρ' όλα αυτά, υπήρξαν και περιπτώσεις ουδέτερης ή αρνητικής στάσης, οι οποίες συχνά συνδέονταν με τις θεωρούμενες ως ανεπαρκείς οικολογικές προσπάθειες.

Η ανάλυση συναισθήματος στα μέσα κοινωνικής δικτύωσης κατά την πανδημία του κορωναϊού

Υπό το πρίσμα της πανδημίας του Covid-19, η έρευνα της Senadhira αναφέρει μια μελέτη ανάλυσης συναισθήματος που διεξήχθη σε δεδομένα του Twitter σχετικά με την ηλεκτρονική μάθηση. Η επιδημία προκάλεσε μια δραματική αλλαγή στην εκπαίδευση, με αποτέλεσμα να παρατηρηθεί μια έξαρση των συζητήσεων και των απόψεων στο Twitter σχετικά με την εξ αποστάσεως και τη διαδικτυακή εκπαίδευση. Κατά τη διάρκεια της επιδημίας, η ερευνητική ομάδα συγκέντρωσε ένα σημαντικό σύνολο δεδομένων από tweets που συζητούσαν την ηλεκτρονική εκπαίδευση. Οι ερευνητές χρησιμοποίησαν τεχνικές ανάλυσης συναισθήματος για να εξετάσουν τα συναισθήματα που αποδίδονταν σε αυτά τα tweets, προκειμένου να κατανοήσουν καλύτερα τις πεποιθήσεις, τις στάσεις και τα συναισθήματα του κοινού σχετικά με τη διαδικτυακή μάθηση σε αυτή τη δύσκολη περίοδο. Τα δεδομένα αντικατοπτρίζουν τις θετικές, αρνητικές και ουδέτερες απόψεις που εξέφρασαν οι χρήστες του Twitter σχετικά με τη διαδικτυακή μάθηση εν μέσω της πανδημίας Covid-19, παρέχοντας εικόνα για τις ευρύτερες τάσεις των συναισθημάτων.

Συνεχίζοντας στη θεματική του Covid-19, όπου κι επήλθαν ριζικές αλλαγές στον τρόπο που διεξάγεται η μάθηση και περάσαμε σε μια νέα και όχι τόσο οικεία σε εμάς εποχή διδασκαλίας, την εξ’ αποστάσεως εκπαίδευση. Τα ιδρύματα τριτοβάθμιας εκπαίδευσης σε όλον τον κόσμο έχουν επηρεαστεί σημαντικά από την πανδημία, η οποία προκάλεσε την ταχεία μετάβαση στη διαδικτυακή μάθηση. Στόχος της μελέτης αυτής είναι να κατανοήσει τις απόψεις που εκφράζει το κοινό σε συζητήσεις σχετικά με την ηλεκτρονική εκπαίδευση κατά τη διάρκεια αυτής της υγειονομικής (κι όχι μόνο) κρίσης. Η ερευνητική ομάδα συγκέντρωσε και εξέτασε ένα σύνολο δεδομένων από διαδικτυακές συζητήσεις από τα μέσα κοινωνικής δικτύωσης, φόρουμ ή ιστότοπους μάθησης. Επιδιώχθηκε ο προσδιορισμός των τις γενικές στάσεις -είτε ευνοϊκές, είτε δυσμενείς, είτε ουδέτερες- σχετικά με τη διαδικτυακή μάθηση στην τριτοβάθμια εκπαίδευση κατά τη διάρκεια της πανδημίας, χρησιμοποιώντας εργαλεία ανάλυσης συναισθήματος. Τα ευρήματά της ομάδας αυτής ρίχνουν φως στις στάσεις, τις πεποιθήσεις και τις δυσκολίες που αντιμετωπίζουν οι εκπαιδευτικοί, οι φοιτητές/τριες και τα ιδρύματα κατά την προσαρμογή τους στην ηλεκτρονική μάθηση. Τα ευρήματα της μελέτης μπορούν να παράσχουν στους εκπαιδευτικούς σημαντικές πληροφορίες για τη βελτίωση της διαδικτυακής εκπαίδευσης σε περιόδους κρίσης όπως αυτή της πανδημίας του Covid-19.

Η ανάλυση συναισθήματος στην ελληνική γλώσσα

Το τελευταίο κομμάτι του θεωρητικού μέρους της παρούσας πτυχιακής ασχολείται με τις τεχνικές ανάλυσης συναισθήματος σχετικά με την ελληνική γλώσσα. Ξεκινώντας με τη μελέτη της εκπαιδευτικής ομάδας της Rokkou, χρησιμοποιώντας μεθόδους αιχμής ειδικά σχεδιασμένες για τις ανάγκες και τις ιδιαιτερότητες της ελληνικής γλώσσας, δημιουργήθηκε ένα εξειδικευμένο μοντέλο και ο στόχος του είναι η αύξηση της ακρίβειας των μοντέλων της ταξινόμησης συναισθήματος στην ελληνική γλώσσα. Για τη συγκέντρωση δεδομένων των ελληνικών κειμένων συλλέχθηκε ένα ποικίλο σύνολο δεδομένων από διάφορες πηγές, όπως ιστολόγια, άρθρα ειδήσεων και μέσα κοινωνικής δικτύωσης. Αυτό εγγυάται ένα ευρύ φάσμα συναισθημάτων και καταστάσεων. Πριν την κύρια διαδικασία επεξεργασίας δεδομένων, τα δεδομένα υποβλήθηκαν σε διαδικασίες προεπεξεργασίας προσαρμοσμένες ειδικά για τα ελληνικά. Σε αυτές περιλαμβάνονται η συμβολοποίηση, η αφαίρεση των λέξεων παύσης (είναι, και κτλ), ώστε η διαχείριση των μοναδικών ελληνικών γραμμάτων και η κανονικοποίηση του κειμένου να ληφθούν υπόψιν οι μορφολογικές αποκλίσεις. Εφαρμόστηκαν τόσο εκσυγχρονισμένες όσο και συμβατικές τεχνικές αφαίρεσης χαρακτηριστικών. Μεταξύ αυτών είναι η δημιουργία μοντέλου, όπου δημιουργήθηκαν και αξιολογήθηκαν τα ακόλουθα μοντέλα μηχανικής μάθησης και βαθιάς μάθησης: Οι μηχανές διανυσμάτων υποστήριξης (SVM), αλγόριθμοι Naive Bayes και Random Forest είναι μερικά παραδείγματα παραδοσιακών μοντέλων μηχανικής μάθησης. Η αξιολόγηση του μοντέλου έγινε με τη χρήση μετρήσεων όπως η ακρίβεια, η ανάκληση και το F1-score. Η ευρωστία και η γενικευσιμότητα των μοντέλων διασφαλίστηκαν με τη χρήση της διασταυρούμενης επικύρωσης (n-fold cross validation), μιας τεχνικής που μοιράζει το σύνολο δεδομένων σε N ίσα τμήματα, κρατώντας ένα τμήμα του N για την δοκιμή και τα υπόλοιπα N-1 τμήματα για εκπαίδευση του μοντέλου. Σύμφωνα με τη μελέτη, τα μοντέλα βαθιάς μάθησης ξεπέρασαν σε απόδοση τα συμβατικά μοντέλα μηχανικής μάθησης, ιδίως εκείνα που βασίζονταν σε RNNs/ Recurrent Neural Networks, *επαναλαμβανόμενα νευρωνικά δίκτυα,* που αξιοποιούν διαδοχικές σειρές δεδομένων ώστε να διευθετηθούν κοινά προβλήματα που εμφανίζονται κατά την αναγνώριση ομιλίας και μετάφρασης αυτής. Το μοντέλο που παρουσίασε την υψηλότερη ακρίβεια ήταν το μοντέλο όπου και προτάθηκε στην παρούσα έρευνα, γεγονός που υποδηλώνει ότι η προτεινόμενη μέθοδος αποδίδει ικανοποιητικά στην καταγραφή του συναισθήματος στα ελληνικά κείμενα.

Η μελέτη του Spatiotis στοχεύει στη δημιουργία και την αξιολόγηση ενός συστήματος ανάλυσης συναισθήματος ειδικά σχεδιασμένου για τα ελληνικά. Λόγω των ιδιαιτεροτήτων της ελληνικής γλώσσας, οι συγγραφείς θέλουν να αναπτύξουν ένα ισχυρό μοντέλο με την ικανότητα να κατηγοριοποιεί αξιόπιστα τα κείμενα ανάλογα με το συναίσθημα. Οι συγγραφείς εξασφάλισαν ένα ευρύ φάσμα θεμάτων και συναισθημάτων συγκεντρώνοντας μια πλούσια συλλογή ελληνόγλωσσης βιβλιογραφίας από πολλαπλές πηγές. Έγινε προεπεξεργασία των κειμένων για να αντιμετωπιστούν γλωσσικά στοιχεία που είναι μοναδικά για την ελληνική γλώσσα, όπως λέξεις παύσης, η κλιτική μορφολογία και τα διακριτικά στοιχεία, μοναδικά στην ελληνική γλώσσα, όπως το ω, ο τόνος στα φωνήεντα κτλ. Για παράδειγμα, τα n-grams και οι ετικέτες στα μέρη του λόγου είναι παραδείγματα κλασικών λεξικών χαρακτηριστικών- οι ενσωματώσεις λέξεων είναι παράδειγμα ενός προηγμένου χαρακτηριστικού που χρησιμοποιήθηκε για την αναπαράσταση των κειμένων. Σημασιολογικά, τα συμφραζόμενα δεδομένα εξήχθησαν από τα κείμενα με τη χρήση των Word2Vec και Term Frequency - Inverse Document Frequency (TF-IDF, στατιστική μέθοδος NLP όπου υπολογίζει τη σημασία ενός έγγραφο). Το προτεινόμενο μοντέλο ενσωματώνει αλγορίθμους μηχανικής μάθησης, όπως μηχανές διανυσματικών μηχανών υποστήριξης (SVM), Naive Bayes και Random Forests. Η κατηγοριοποίηση του συναισθήματος ελληνικών κειμένων ήταν μια άλλη διεργασία που αξιοποιήθηκε για την αξιολόγηση της αποτελεσματικότητας των μοντέλων βαθιάς μάθησης, συμπεριλαμβανομένων των νευρωνικών δικτύων με συνελικτική σύνδεση (*Convolutional Neural Network/ CNN*)-τεχνική μηχανικής μάθησης που αξιοποιεί τα δεδομένα

που λαμβάνονται από εικόνες- και των νευρωνικών δικτύων με επαναλαμβανόμενη σύνδεση (*Recurrent Neural Network/ RNN*). Για την αξιολόγηση των μοντέλων χρησιμοποιήθηκαν τυποποιημένα μέτρα όπως το F1-score, η ανάκληση και η ακρίβεια. Για να διασφαλιστεί ότι τα μοντέλα ήταν αξιόπιστα και ευρέως εφαρμόσιμα, έγινε διασταυρούμενη επικύρωση. Όσον αφορά την ακρίβεια και τη συνολική απόδοση, τα μοντέλα βαθιάς μάθησης -ιδιαίτερα τα RNN- εμφανίζουν καλύτερες επιδόσεις από τους συμβατικούς αλγορίθμους μηχανικής μάθησης. Υψηλή ακρίβεια επιτεύχθηκε από το προτεινόμενο μοντέλο, αποδεικνύοντας τη χρησιμότητα των επιλεγμένων προσεγγίσεων και στρατηγικών εξαγωγής χαρακτηριστικών, τονίζοντας την αξία της προσαρμοσμένης εξαγωγής χαρακτηριστικών και των μεθόδων προεπεξεργασίας για μη αγγλικές γλώσσες. Τα αποτελέσματα υποδεικνύουν την ανώτερη καταλληλότητα των τεχνικών βαθιάς μάθησης για εργασίες ανάλυσης συναισθήματος στα ελληνικά, θέτοντας στέρεες βάσεις για περαιτέρω μελέτη και πρακτικές εφαρμογές στον τομέα αυτό.

Κλείνοντας, η έρευνα των Dervenis με τους συνεργάτες του *(Dervenis et al., 2024)* χρησιμοποιεί τεχνικές ανάλυσης συναισθήματος για να αναλύσει τα δεδομένα των φοιτητών προκειμένου να αξιολογήσει τις δεξιότητες των εκπαιδευτών/τριων στην τριτοβάθμια εκπαίδευση. Η κεντρική ιδέα είναι η παρουσίαση μιας αμερόληπτης αξιολόγησης του έργου των καθηγητών/τριών με βάση τις απόψεις που μοιράζονται οι φοιτητές/τριες. Συγκεντρώθηκαν απόψεις φοιτητών πανεπιστημίων και κολεγίων, με τα σχόλια να καλύπτουν ένα ευρύ φάσμα εκπαιδευτικών θεμάτων, συμπεριλαμβανομένων των σχεδίων μαθημάτων, των στρατηγικών διδασκαλίας και των αλληλεπιδράσεων μεταξύ καθηγητών και φοιτητών. Η προεπεξεργασία των μηνυμάτων ανατροφοδότησης βοήθησε στην τακτοποίηση και την ετοιμασία των δεδομένων για ανάλυση. Η συμβολοποίηση, η εξάλειψη των λέξεων παύσης και το stemming ή lemmatization (τεχνικές που αναγνωρίζουν ομόρριζες λέξεις και αποθηκεύουν τη ρίζα σαν σημείο επεξεργασίας συναισθήματος) ήταν μερικές από τις διαδικασίες που χρησιμοποιήθηκαν για την μορφοποίηση του κειμένου. Η ανατροφοδότηση κατηγοριοποιήθηκε σε καλή, αρνητική και ουδέτερη χρησιμοποιώντας ανάλυση συναισθήματος. Εφαρμόστηκαν τόσο τεχνικές μηχανικής μάθησης όσο και συμβατικές τεχνικές ανάλυσης συναισθήματος. Για να βρεθεί η καλύτερη στρατηγική για την ταξινόμηση συναισθήματος, διερευνήθηκαν τεχνικές όπως αλγόριθμοι μάθησης με επίβλεψη και προσεγγίσεις που βασίζονται σε λεξικά Αφαιρέθηκαν χαρακτηριστικά κειμένου για να συμβάλλουν στην ταξινόμηση συναισθήματος. Ανάμεσά τους ήταν οι ενσωματώσεις λέξεων, τα n-grams και η συχνότητα όρων-αντίστροφης συχνότητας εγγράφων (TF-IDF). Δημιουργήθηκε και αξιολογήθηκε ένας αριθμός μοντέλων για να διαπιστωθεί πόσο ακριβής ήταν η ταξινόμηση του συναισθήματος. Οι μηχανές διανυσμάτων υποστήριξης (SVM), τα Naive Bayes και τα τυχαία δάση είναι παραδείγματα μοντέλων μηχανικής μάθησης που συγκρίθηκαν. Για την αξιολόγηση των μοντέλων χρησιμοποιήθηκαν τυποποιημένα κριτήρια αξιολόγησης όπως η ακρίβεια, η ανάκληση και το F1-score. Η μελέτη προέβαλε ότι η ανάλυση συναισθήματος μπορεί να κατηγοριοποιήσει αποτελεσματικά τα σχόλια των μαθητών σε διάφορες ομάδες συναισθήματος, αναγνωρίζοντας αντιλήψεις σχετικά με την αποτελεσματικότητα των καθηγητών. Τα ευρήματα διευκόλυναν τον εντοπισμό τομέων για ανάπτυξη, επισημαίνοντας ορισμένες πτυχές της διδασκαλίας που προσέλκυσαν επαίνους ή αρνητική κριτική.

Παρακάτω φαίνεται συνοπτικά η βιβλιογραφική επισκόπηση που συνέβαλε στην υλοποίηση της παρούσας πτυχιακής με τη μορφή ενός πίνακα, που περιλαμβάνει τους/τις συγγραφείς των δημοσιεύσεων, το επίπεδο στο οποίο εφαρμόστηκε η συναισθηματική ανάλυση, το μέγεθος του δείγματος, το μοντέλο μηχανικής μάθησης που εφαρμόστηκε στο εκάστοτε άρθρο καθώς και κάποια ειδικά χαρακτηριστικά που εμφανιζόντουσαν στα άρθρα αυτά.

ΣΥΓΓΡΑΦΕΙΣ	ΕΠΙΠΕΔΟ ΔΙΑΧΩΡΙΣΜΟΥ	ΔΕΙΓΜΑ	ΜΟΝΤΕΛΟ	ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ
Imane Lasri Anouar Riadsolh, Mourad Elbelkacemi	Πρόταση	24642 αγγλικά tweets	Self-attention-based Bi-LSTM	GloVe word embedding (ενσωμάτωσης λέξεων)
K. I. Senadhira, R. A. H. M. Rupasingha, B. T. G. S. Kumara	Πρόταση	8976 tweets	Artificial Neural Network (ANN)	Term Frequency Inverse Document Frequency (TF-IDF) vectorizer.
Mireia Usart, Carme Grimalt-Álvaro, Adolf Maria Iglesias-Estradé	Πρόταση	N = 48 φοιτήτριες/τές	Support Vector Machine (SVM)	Linear Kernel TASS Corpus
Nor Ain Syafiqah Remali, Mohd Razif Shamsuddin, Shuzlina Abdul-Rahman	Πρόταση	38,602 tweets	SVM, Naïve Bayes (NB), Decision Tree (DCT), and Random Forest (RF)	Power BI data visualization
Imane Lasri, Anouar Riadsolh, Mourad Elbelkacemi	Πρόταση	1798 γαλλικά tweets	RF, multinomial NB classifier, logistic regression, DCT, linear SVC (Classifier), and XGBoost TF-IDF	count vectorizer feature extraction techniques Twint 10-fold cross validation Kafka Kibana (data visualization)

Reinhard Alfaries Saemani, Nina Setiyawati	Πρόταση	1252 Tweets	SVM	C, Gamma, Kernel = RBF
Pooja · Rajni Bhalla	Λέξεις	21 έρευνες	SVM, NB	Kernel techniques
Yating Wen , Xiaodong Zhao , Xingguo Li and Yuqi Zang	Φράσεις	18.466 κινεζικά σχόλια	Latent Dirichlet Allocation (LDA)	SnowNLP Visualization Diagram
Lingyao Li , John Johnson , William Aarhus , Dhawal Shah	Πρόταση	116 MOOCS (N= 183.574 κριτικές)	NLP	DNA charts Vader Sentiment tool Radar Chart
Nikola Nikolic Olivera Grljevic Aleksandar Kovacevic	Φράσεις	6.335 κριτικές	SVM,	
Hoar, Benjamin B., Ramachandran, Roshini, Levis-Fitzgerald, Marc, Sparck, Erin M., Wu, Ke, Liu, Chong	Πρόταση	1.603 κριτικές	NLP	multigrams
Mrtdaa Mohammed Almosawi , Dr. Salma A. Mahmood	Λέξεις	2.217 λέξεις	NB, K-Nearest Neighbors(K-NN) and SVM	Bigram Pie chart
Deepa S M , Revathi N , Sivakami K , Piyush Kumar Pareek	Πρόταση	33 μεταβλητές and 1044 παρατηρήσεις	CNN, DT, RF, SVM	Unigram Correlation matrix single and multi-variant regression techniques

Zenun Kastrati, Fisnik Dalipi, Ali Shariq Imran, Krenare Pireva Nuci, Mudasir Ahmad Wani	Αρχείο	92 Έρευνες	SVM, NB, DT, k-NN, Neural Networks (NN)	PRISMA framework, Systematic Mapping, PICO Framework Bigrams
Georgia Rokkou, Vasilis Triantafyllou, Nikolaos Spatiotis, Michael Paraskevas	Πρόταση	Tweets	SVM, NB, RF	TF-IDF N-gram
Nikolaos Spatiotis, Iosif Mporas, Michael Paraskevas, Isidoros Perikos	Αρχείο	11.156 σχόλια	SVM, NB, RF, CNN, RNN	Word2Vec TF-IDF Unigram
Charalampos Dervenis, Panos Fitsilis, Omiros Iatrellis, and Athanasios Koustelios	Φράσεις	700+ σχόλια	NLP	Visualization Diagram
Akshi Kumar, Teeja Mary Sebastian	Αρχείο	36 δημοσιεύσεις	NLP	Multigram

Πίνακας 1. Συγκεντρωτικός πίνακας ερευνών που συμμετέχουν στην παρούσα πτυχιακή

ΚΕΦΑΛΑΙΟ 3 Μεθοδολογία Έρευνας

Στο κεφάλαιο αυτό παρουσιάζεται και αναλύεται το σύνολο δεδομένων (dataset) το οποίο χρησιμοποιήθηκε ως είσοδος τροφοδότησης του μοντέλου ανάλυσης συναισθημάτων.

Το σύνολο δεδομένων δημιουργήθηκε από αξιολογήσεις φοιτητριών/τών ενός πανεπιστημίου της Βόρειας Ινδίας και το οποίο σχημάτισε τη βάση της ανάλυσης της ανατροφοδότησης εκπαιδευτικών ιδρυμάτων για το συγκεκριμένο πανεπιστήμιο.

Το συγκεκριμένο dataset αποτελείται από τις εξής στήλες: CourseContent, Examination, Labwork, Library_Facilities, Sentiment και Text. Όλες οι στήλες -πλην των δύο τελευταίων- εμφανίζονται από δύο φορές η κάθε μια, μία φορά για να παρουσιάσουν την γραπτή αξιολόγηση και μία για να αποτυπώσουν την συναισθηματική πολικότητα στην κριτική αυτή. Στη συγκεκριμένη πτυχιακή εργασία αναλύθηκαν και επεξεργάστηκαν οι δύο τελευταίες στήλες, δηλαδή Sentiment και Text, οι οποίες αποτελούν διπλότυπο των ήδη υπαρχουσών στηλών και μετονομάστηκαν έτσι για διευκόλυνση κατά τον προγραμματισμό του κώδικα. Οι στήλες αυτές περιλαμβάνουν την συναισθηματική πολικότητα της εκάστοτε αξιολόγησης (δηλαδή -1,0,1) -στήλη Sentiment- και το κείμενο της εκάστοτε κριτικής -στήλη Text- αντίστοιχα.

Στην επεξεργασία και εκτέλεση του κώδικα έλαβαν μέρος διάφορες συναρτήσεις και συγκεκριμένοι αλγόριθμοι, εφαρμοσμένα πάντα στην ανάλυση συναισθήματος. Ο κώδικας είναι γραμμένος αποκλειστικά σε Python, όπου και χρησιμοποιήθηκαν βιβλιοθήκες και πακέτα ήδη εγκατεστημένα στις πηγές της γλώσσας, όπως για παράδειγμα numpy, pandas και seaborn.

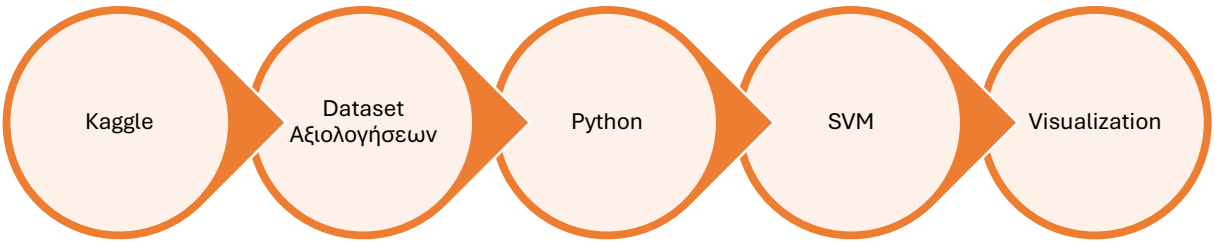
Ξεκινώντας με τον αλγόριθμο που χρησιμοποιήθηκε για την ανάλυση του κώδικα, είναι ο Support Vector Machine (SVM). Η χρήση του συμβάλει στην ορθή λειτουργία σε datasets μικρού ή μεσαίου μεγέθους (δηλαδή με μικρό πλήθος εγγραφών και πεδίων, ακριβώς όπως το dataset που χρησιμοποιήθηκε σε αυτήν την εργασία). Επιπλέον, ο SVM έχει καλή απόδοση σε πολυδιάστατα δεδομένα, μέσω της χρήσης του TF-Idf που χρησιμοποιήθηκε επίσης στον κώδικα. Τέλος, μέσα από τον SVM, ενισχύεται η αξιοπιστία του μοντέλου ως προς την αποφυγή overfitting αυτού, δηλαδή το να λειτουργεί με τον ίδιο τρόπο το μοντέλο που δημιουργήθηκε, ανεξάρτητα από την είσοδο που του παρέχουμε, καθιστώντας το μοντέλο αναξιόπιστο και δύσχρηστο σε περαιτέρω datasets και δοκιμές.

Μέσω του SVM αξιοποιήθηκε και η συνάρτηση GridSearch(), η οποία συμβάλει στον συντονισμό υπερπαραμέτρων. Αυτό σημαίνει πως η διαδικασία εύρεσης των βέλτιστων συνδυασμών υπερπαραμέτρων αυτοματοποιείται για το εκάστοτε μοντέλο μηχανικής μάθησης. Σκοπός της παραπάνω διαδικασίας είναι η εύρεση του βέλτιστου συνδυασμού δεδομένων εισόδου για το καλύτερο δυνατό αποτέλεσμα απόδοσης του μοντέλου που παράχθηκε. Στο τελευταίο κομμάτι της επεξεργασίας του dataset, το training, ο αλγόριθμος SVM είναι ο καταλυτικός παράγοντας που ελέγχει, δοκιμάζει και αποδίδει τις τελικές τιμές των παραμέτρων στις αντίστοιχες μεταβλητές, ελέγχοντας την ορθότερη και αποδοτικότερη λειτουργία των μεταβλητών μέσω των υπερπαραμέτρων, όπου αναλύονται εκτενέστερα στο Κεφάλαιο 4.

Μέσω του SVM έγινε και χρήση της συνάρτησης SMOTE, η οποία εφαρμόζεται για την απαλοιφή ανισορροπιών στο σετ δεδομένων, ώστε να παραχθεί ένα πιο ομαλό (και συνεπώς πιο αξιόπιστο) μοντέλο μηχανικής μάθησης. Μέσω συγκεκριμένων παραμέτρων η συνάρτηση SMOTE μπορεί να διαχειριστεί και σύνθετα δεδομένα, κάνοντας την να λειτουργεί εύρυθμα και σε επεξεργασία και ανάλυση περίπλοκων και σύνθετων δεδομένων. Παρακάτω δίνεται ένα παράδειγμα του σετ δεδομένων που αναλύθηκε για την εξαγωγή αποτελεσμάτων για την παρούσα εργασία.

teaching	teaching	coursecontent	coursecontent
0	teacher are punctual but they should also give us the some practical knowledge other than theortical	0	content of courses are average
1	Good	-1	Not good
1	Excellent lectures are delivered by teachers and all teachers are very punctual.	1	All courses material provide very good knowledge in depth .
1	Good	-1	Content of course is perfectly in line with the teaching philosophy that is being perpetuated here i.e cramming,rote learning.
1	teachers give us all the information required to improve the performance.	1	content of courses improves my knowledge

Πίνακας 2. Στιγμιότυπο πίνακα από το dataset με στήλες συναισθημάτων και αξιολογήσεων.



Εικ. 9. Το παραπάνω σχήμα απεικονίζει συνοπτικά την διαδικασία από την συλλογή δεδομένων μέχρι και την εκτέλεση του κώδικα.

Στο παραπάνω σχήμα παρουσιάζονται τα βήματα τα οποία έγιναν για την υλοποίηση της παρούσας πτυχιακής όσον αφορά το κομμάτι του κώδικα. Αρχικά, έγινε η αναζήτηση της βάσης δεδομένων στο Kaggle, όπου μέσω της γλώσσας προγραμματισμού python ξεκίνησε η συγγραφή του κώδικα για την ανάλυση και επεξεργασία της βάσης δεδομένων. Όπως έγινε ήδη γνωστό, ο κώδικας βασίστηκε στο μοντέλο μηχανικής μάθησης Support Vector Machine όπου και δημιουργήθηκε και εκπαιδεύτηκε ο αναλυτής. Τέλος, με την εκτέλεση του προγράμματος, εκτυπώνονται στην οθόνη τα αποτελέσματα της εκπαίδευσης του αναλυτή σε οπτικοποιημένη μορφή, την οποία θα δούμε στο επόμενο κεφάλαιο.

ΚΕΦΑΛΑΙΟ 4 Υλοποίηση

Στο παρόν κεφάλαιο θα γίνει μια αναλυτική παρουσίαση του κώδικα που χρησιμοποιήθηκε καθώς και των αποτελεσμάτων που παράχθηκαν από την εκτέλεση αυτού. Πιο συγκεκριμένα, ο κώδικας στον οποίο βασίστηκε η παρούσα εργασία είναι ο παρακάτω :

```
import numpy as np
import pandas as pd
import nltk
import matplotlib.pyplot as plt
from wordcloud import WordCloud
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split, GridSearchCV, cross_val_score
from sklearn.svm import SVC
from sklearn.metrics import classification_report, accuracy_score, confusion_matrix
from imblearn.over_sampling import SMOTE
import logging
from sklearn.utils import parallel_backend
from joblib import parallel_backend
from sklearn import metrics
import seaborn as sns


def process_data():
    with parallel_backend('threading'):
        verbose=0

# Configure logging, deixnei info level minimata  opou kai ta kanei assign sto GridSearchCV

logging.basicConfig(level=logging.INFO)
logger = logging.getLogger('GridSearchCV')

# Load feedback data
df = pd.read_excel('C:/Users/Koe/Desktop/finalDataset0_2.xlsx')

# Convert text to lowercase and ensure it is string type
df['text'] = df['text'].astype(str).str.lower()

# Tokenization
df['tokens'] = df['text'].apply(nltk.word_tokenize)

# Remove stopwords
stopwords = nltk.corpus.stopwords.words('english')
df['tokens'] = df['tokens'].apply(lambda x: [word for word in x if word not in stopwords])

X = df['text']
y = df['sentiment']
```

```
# Check class balance
print("\nClass distribution:\n")
print(y.value_counts())

# Word Cloud for Ratings | theyre only used to show how many iterations of each category were found between [-1,1]
feedbackratings = df['sentiment']
counts = feedbackratings.value_counts()
counts.index = counts.index.map(str)

# Word Cloud for Comments
feedbackcomments = df['text']
textonly2 = feedbackcomments.to_string(index=False)
print("\n")
print(feedbackcomments.head())

# Plotting word clouds
wc1 = WordCloud(width=800, height=400, background_color='white').generate_from_frequencies(counts)
plt.figure(figsize=(10, 6))
plt.imshow(wc1, interpolation='bilinear')
plt.axis('off')
plt.title('Word Cloud - Sentiment Ratings Frequency')

wc2 = WordCloud(width=800, height=400, background_color='white').generate(textonly2)
plt.figure(figsize=(10, 6))
plt.imshow(wc2, interpolation='bilinear')
plt.axis('off')
plt.title('Word Cloud - Comments')

plt.show()

# Train-test splitting

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Vectorization

vectorizer = TfidfVectorizer()
X_train_vectors = vectorizer.fit_transform(X_train)
X_test_vectors = vectorizer.transform(X_test)

# Xrisi SMOTE wste na apaloifthoun anisoropies
smote = SMOTE(random_state=42)
X_train_vectors, y_train = smote.fit_resample(X_train_vectors, y_train)

# Hyperparameter tuning mesw tou GridSearchCV diladi gia ta beltisa dinata apotelesmata
```

```
        param_grid = {'C': [0.1, 1, 10, 100], 'gamma': [1, 0.1, 0.01, 0.001], 'kernel': ['linear', 'rbf']}

        # Afto edw to xrisimopoiisa giati mou ebgaze 150 minimata san : [CV] END...C=100, gamma=0.01, kernel=rbf; totaltime=0.0s
        with parallel_backend('threading'):
            grid = GridSearchCV(SVC(), param_grid, refit=True, verbose=0)  # Set verbose=0 to suppress detailed output
            grid.fit(X_train_vectors, y_train)

        print(f"\nBest Parameters: {grid.best_params_}\n")

        # Ekpaidefsi montelou me tis kaliteres parametrous mesw GridSearch
        model = SVC(C=grid.best_params_['C'], gamma=grid.best_params_['gamma'], kernel=grid.best_params_['kernel'], class_weight='balanced')
        model.fit(X_train_vectors, y_train)

        # 5-fold Cross validation scores, wste na exw 80:20 ratio giati exw mikro sample
        cv_scores = cross_val_score(model, X_train_vectors, y_train, cv=5)
        print(f"Cross-validation scores: {cv_scores}\n")
        print(f"Mean cross-validation score: {np.mean(cv_scores)}\n")


        # Problepsi
        y_pred = model.predict(X_test_vectors)

        # Emfanisi metrikwn

        print("Classification Report:")
        print("\n")
        print(classification_report(y_test, y_pred, zero_division=0))

        print("\nAccuracy Score:", accuracy_score(y_test, y_pred))

        # Print confusion matrix

        cf_matrix = confusion_matrix(y_test, y_pred)
        sns.heatmap(cf_matrix/np.sum(cf_matrix), annot=True,  cmap='Greens')

        print("\n")

        plt.show()
```

Έπειτα, όπως αναφέρθηκε και προηγουμένως, η εκτέλεση του προγράμματος έχει ως έξοδο τα παρακάτω μηνύματα με τα ακόλουθα αποτελέσματα:

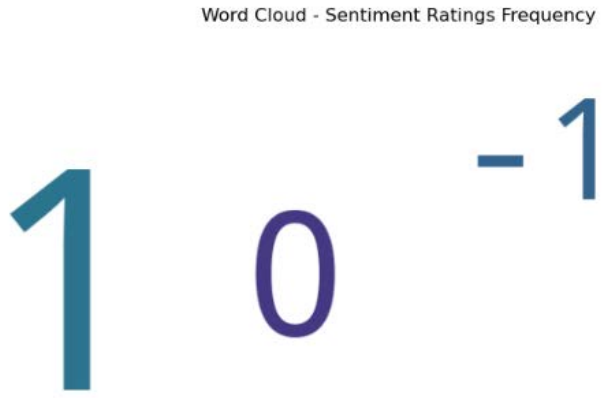
SENTIMENT	
1	154
0	19
-1	12

Πίνακας 3. Αριστερά εμφανίζεται η πολικότητα και το εύρος της συναισθηματικής τιμής και δεξιά το πόσες φορές εμφανίστηκε στο dataset

	Precision	Recall	F1-Score	Support
-1	0.00	0.00	0.00	1
0	1.00	0.40	0.57	5
1	0.89	1.00	0.94	31
Accuracy			0.89	37
Macro avg	0.63	0.47	0.50	37
Weighted avg	0.88	0.89	0.86	37
Accuracy Score	0.89189189189			

Πίνακας 4. Ο παραπάνω πίνακας περιέχει συγκεντρωτικά την έξοδο του προγράμματος από την εκτέλεση του κώδικα.

και παράγονται τα εξής WordClouds για την απεικόνιση των αποτελεσμάτων της ανάλυσης συναισθήματος:



Εικ. 10. Η παραπάνω εικόνα είναι αποτέλεσμα της εκτέλεσης του κώδικα που αναπτύχθηκε για ερευνητικούς σκοπούς της παρούσας πτυχιακής εργασίας, όπου παρουσιάζεται το εύρος μέσα στο οποίο κινείται το αποτέλεσμα συναισθήματος

Εικ. 11. Η παραπάνω εικόνα είναι αποτέλεσμα της εκτέλεσης του κώδικα που αναπτύχθηκε για ερευνητικούς σκοπούς της παρούσας πτυχιακής εργασίας, όπου και φαίνονται οι πιο συχνά χρησιμοποιημένες λέξεις στις αξιολογήσεις των συμμετεχόντων στην εκπαιδευτική διαδικασία.

Η παρατήρηση της πρώτης εικόνας, δηλαδή η εικόνα με τα -1, 0 και 1 ξεχωρίζει τις τρεις τιμές που παίρνουν τα συναισθήματα κατά την ανάλυση συναισθήματος, όπως έχει γνωστοποιηθεί και στο *Κεφάλαιο 2* σε αντίστοιχες εργασίες. Δηλαδή, τα συναισθήματα μπορούν να πάρουν την τιμή -1 για να χαρακτηρίσουν ένα αρνητικό συναίσθημα, την τιμή 0 ώστε να χαρακτηρίσουν ένα ουδέτερο συναίσθημα και την τιμή 1 με σκοπό τον χαρακτηρισμό ενός συναισθήματος ως θετικό. Κάτι που αξίζει να σημειωθεί είναι πως η πλειοψηφία των φοιτητών και φοιτητριών κρατούν θετική στάση και αισθάνονται ικανοποίηση από την διδασκαλία του ιδρύματός τους.

Εξετάζοντας τη δεύτερη εικόνα, φαίνεται πως την άρχουσα θέση στις βαθμολογήσεις των εκπαιδευομένων εξέχει η λέξη «good», ακολουθούμενη από τις λέξεις «excellent» και «activities» αντίστοιχα. Αυτό είναι θετικό, μιας και στο συγκεκριμένο WordCloud λήφθηκαν υπόψιν μόνο οι λέξεις με θετικό πρόσημο, γεγονός που υποδηλώνει πως τα άτομα που παίρνουν μέρος στην εκπαιδευτική διαδικασία είναι ικανοποιημένα από το εκπαιδευτικό τους ίδρυμα και από τις δραστηριότητες που λαμβάνουν χώρα γύρω από αυτήν.

Συνεχίζοντας, ας ρίξουμε μια ματιά στις μετρικές και στα μέρη αυτών, επεξηγώντας τη σημασία και τη χρήση τους.

- print(f"\nBest Parameters: {grid.best_params_}\n") το οποίο έχει ως έξοδο στην οθόνη:
  Best Parameters: {'C': 1, 'gamma': 1, 'kernel': 'rbf'}

  C: 1 → Χαρακτηρίζει το κόστος της λάθος κατάταξης των δεδομένων.

  Gamma:1 → Η παρούσα παράμετρος επηρεάζει την ευαισθησία του μοντέλου. Εάν η τιμή της παραμέτρου αυτής είναι πολύ μεγάλη, τότε το εύρος στο οποίο επηρεάζει τα διανύσματα υποστήριξης, χωρίς ρύθμιση με την παράμετρο «C» ώστε να προλάβουν την υπερβολική προσαρμογή του μοντέλου / overfitting-Το overfitting είναι συμπεριφορά του μοντέλου όπου αυτό παρέχει ακριβή αποτελέσματα για τα δεδομένα που χρησιμοποιούνται στην εκπαίδευση αυτού, μα αδυνατεί να προσφέρει ακριβή αποτελέσματα κατά την

εισαγωγή νέων δεδομένων. Εάν όμως το Gamma είναι μικρό, δεν θα είναι σε θέση να αποτυπώσει την πολυπλοκότητα των δεδομένων.

Kernel: rbf (Radial Basis Function)→ Η συγκεκριμένη παράμετρος είναι αρκετά περίπλοκη και συνάμα αποδοτική, αφού συνδυάζει διαφορετικούς πολυνωμικούς πυρήνες-kernels, αλγόριθμοι ειδικοί για την ανάλυση μοτίβων- για την απεικόνιση διαφορετικών και μη γραμμικών δεδομένων, βάσει διαφορετικών σκοπιών.

Έπειτα, έχουμε τα:

- print(f"Cross-validation scores: {cv_scores}\n")
- print(f"Mean cross-validation score: {np.mean(cv_scores)}\n"), με έξοδο το παρακάτω

Cross-validation scores: [1. 1. 1. 1. 1.]

Mean cross-validation score: 1.0

Cross-validation (διασταυρούμενη επικύρωση): είναι η διαδικασία μέσω τις οποίας χωρίζεται το δείγμα των δεδομένων σε δύο (άνισα) μέρη, ένα για την εκπαίδευση και ένα για τη δοκιμή του μοντέλου μηχανικής μάθησης, και όπως αναφέραμε παραπάνω, η πιο διαδεδομένη μορφή cross-validation είναι η 80/20.

Τέλος, με το κλείσιμο της συναισθηματικής ανάλυσης και της παρουσίασης των αποτελεσμάτων αυτής έχουμε τις εξής γραμμές κώδικα:

- ```
  # Problepsi
      y_pred = model.predict(X_test_vectors)
  # Emfanisi metrikwn
  print("Classification Report:")
  print("\n")
      print(classification_report(y_test, y_pred, zero_division=0))
      print("\nAccuracy Score:", accuracy_score(y_test, y_pred))
      # Print confusion matrix
  print("\nConfusion Matrix:\n")
      print(confusion_matrix(y_test, y_pred))
  ```

Αναλύοντας ξεχωριστά τις παραπάνω μετρικές διακρίνουμε τις εξής:

- Precision →Η ακρίβεια, σε συνδυασμό με την ανάκληση, φανερώνουν το ποσοστό στο οποίο μπορεί το μοντέλο να ταυτοποιήσει με σωστό τρόπο τα συναισθήματα
- Recall → Η ανάκληση συμβάλλει, όπως είδαμε παραπάνω, στην ταυτοποίηση και ταξινόμηση συναισθημάτων στην εκάστοτε κατηγορία
- F1-score → Η μετρική που εξισορροπεί τα αποτελέσματα, με τον συνδυασμό της ακρίβειας και της ανάκλησης, όπου λαμβάνεις υπόψιν και τα ψευδώς αρνητικά, μαζί με τα ψευδώς αρνητικά αποτελέσματα.
- Support → Η υποστήριξη φανερώνει το πραγματικό πλήθος τον επαναλήψεων της εκάστοτε κλάσης μέσα στο σετ δεδομένων.
- Accuracy → Η ακρίβεια παρουσιάζει τα ορθώς ταξινομημένα στοιχεία συναισθήματος
- Macro avg(average) → Όλες οι κλάσεις συμμετέχουν ισότιμα για τον υπολογισμό της μέσης τελικής μετρικής
- Weighted avg → Η συνεισφορά της εκάστοτε κλάσης στον υπολογισμό της μέσης τελικής μετρικής είναι ανάλογη του μεγέθους της, από όπου και «ζυγίζεται».
  ```

- Confusion Matrix(μήτρα σύγχυσης) → Είναι ένας συγκεντρωτικός πίνακας όπου αναπαρίσταται η ανάλυση των προβλέψεων που έλαβαν χώρα σε ένα μοντέλο, λεπτομερώς.

Από τα αποτελέσματα της εκτέλεσης κώδικα παρατηρούμε πως οι τιμές F1-score και Precision για την πολικότητα συναισθήματος με τιμή «1» δίνει τα καλύτερα δυνατά αποτελέσματα -0.94 και 0.89 αντίστοιχα- σε σύγκριση με τις υπόλοιπες κατηγορίες για τις άλλες δύο μετρικές (0,-1). Οι τιμές των F1-score και Precision πλησιάζουν αρκετά τη μονάδα, γεγονός άκρως θετικό, διότι δηλώνεται με αυτόν τον τρόπο πως οι κατατάξεις αυτές επιτυγχάνουν ορθή πρόβλεψη και κατάρτιση των δεδομένων στην εκάστοτε κατηγορία.

Επιπλέον, από την ανάλυση που έγινε στα τρία μετρικά [-1, 0, 1] αναδεικνύεται πως το μοντέλο ανταπεξέρχεται σωστά στην ταξινόμηση τόσο στα ψευδώς θετικά αλλά και στα ψευδώς αρνητικά αποτελέσματα (όπως Precision & Recall), γεγονός που υποστηρίζει την ορθή επιλογή του Support Vector Machine ως αλγόριθμο υλοποίησης της ανάλυσης συναισθήματος. Ο αλγόριθμος βοηθάει σε datasets με μικρό ή μεσαίο πλήθος πεδίων, κάτι που ισχύει στην παρούσα υλοποίηση. Επιπροσθέτως, ο SVM επιτρέπει την αυτοματοποιημένη ανάλυση και κατάρτιση των δεδομένων σε κατηγορίες, ελαχιστοποιώντας την ανθρώπινη επέμβαση στην διεξαγωγή αποτελεσμάτων.

```
sentiment
 1    154
 0     19
-1     12
Name: count, dtype: int64


0    extracurricular activities are excellent and p...
1                                                 good
2    extra curricular activities also help students...
3    complete wastage of time. again this opinion i...
4    extracurricular activities increases mental an...
Name: text, dtype: object

Best Parameters: {'C': 1, 'gamma': 1, 'kernel': 'rbf'}

Cross-validation scores: [1. 1. 1. 1. 1.]

Mean cross-validation score: 1.0

Classification Report:


              precision    recall  f1-score   support

          -1       0.00      0.00      0.00         1
           0       1.00      0.40      0.57         5
           1       0.89      1.00      0.94        31

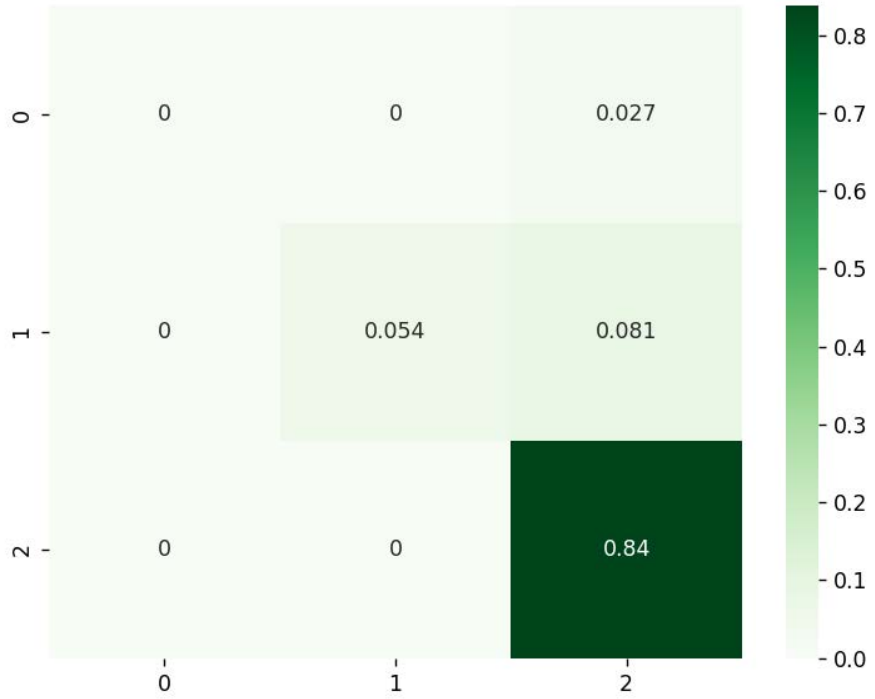
    accuracy                           0.89        37
   macro avg       0.63      0.47      0.50        37
weighted avg       0.88      0.89      0.86        37


Accuracy Score: 0.8918918918918919

Confusion Matrix:

[[ 0  0  1]
 [ 0  2  3]
 [ 0  0 31]]
```

Εικ. 12. Το παραπάνω στιγμιότυπο οθόνης εμπεριέχει συγκεντρωτικά τα αποτελέσματα που αναφέρθηκαν προηγουμένως.



Εικ. 13. Η παραπάνω εικόνα είναι η μήτρα σύγχυσης (confusion matrix) που εμφανίζεται στην οθόνη μετά την εκτέλεση του προγράμματος, απεικονίζοντας τις ορθές προβλέψεις του αναλυτή μας.

ΚΕΦΑΛΑΙΟ 5 Συμπεράσματα

Η παρούσα πτυχιακή εργασία πραγματεύεται τη συναισθηματική ανάλυση στην τριτοβάθμια εκπαίδευση, μια μεθοδολογία που έχει πλέον εξελιχθεί σε ισχυρό εργαλείο για την εξαγωγή σημαντικών πληροφοριών, σχετικά με το πως αισθάνονται οι μετέχοντες/μετέχουσες για την εκπαιδευτική διαδικασία. Στην εργασία αυτή, παρουσιάζεται πως οι τεχνικές μηχανικής μάθησης (ML) και επεξεργασίας φυσικής γλώσσας (NLP) μπορούν να αξιοποιηθούν για να υπολογίσουν τον βαθμό ικανοποίησης των φοιτητών/τριών αλλά και πως η συναισθηματική ανάλυση μπορεί να λειτουργήσει ως δείκτης ο οποίος φανερώνει την ποιότητα των εκπαιδευτικών προγραμμάτων.

Αρχικά, η ανάλυση συναισθημάτων, γνωστή και ως εξόρυξη γνώμης (opinion mining) ξεκίνησε ως ένα εργαλείο για εμπορικούς σκοπούς, όπως η παρακολούθηση της επιτυχίας μιας ταινίας ή ενός προϊόντος. Με την πάροδο του χρόνου, η χρήση της ανάλυσης συναισθημάτων επεκτάθηκε και εδραιώθηκε σε πολλούς τομείς, όπως πολιτική, marketing αλλά και εκπαίδευση. Μέσω της διαρκούς εξέλιξης και βελτίωσης της συναισθηματικής ανάλυσης, αναπτύχθηκε και εδραιώθηκε η δυνατότητα ανάλυσης των εκπαιδευομένων σε πραγματικό χρόνο ή μέσω σχολίων και αξιολογήσεών τους, γεγονός που συμβάλλει ώστε τα εκπαιδευτικά ιδρύματα να βελτιώσουν και να διευρύνουν τις υπηρεσίες τους με σκοπό τη βέλτιστη ανταπόκριση στις ανάγκες των φοιτητών/τριών.

Ευρήματα και Συνεισφορά

Η συναισθηματική ανάλυση στην εκπαίδευση προσφέρει πολύτιμα στοιχεία για την ισχυροποίηση και ενίσχυση της μαθησιακής εμπειρίας. Μέσω της παρούσας πτυχιακής εργασίας αναδείχθηκε η σημασία της συλλογής δεδομένων, της επεξεργασίας και της ανάλυσής τους, με σκοπό την καταγραφή και κατανόηση των συναισθημάτων των εκπαιδευομένων.

Υψίστης σημασίας ήταν η ανάλυσή του πώς οι φοιτητές/τριες αξιολογούν και αντιμετωπίζουν τα μαθήματα, το διδακτικό προσωπικό και γενικά την εκπαιδευτική εμπειρία μέσα στο εκάστοτε διδακτικό πλαίσιο (ενδεικτικά διαλέξεις, περιεχόμενα και στοχοθεσία αυτών, κατανόηση και ποιότητα εκπαιδευτικού ιδρύματος). Τα ευρήματα αυτών, έπειτα από εκτενές φιλτράρισμα, δείχνουν πως οι εκπαιδευόμενοι/ες εκφράζουν έντονα τις πεποιθήσεις τους όσον αφορά τις διδακτικές μεθόδους, τα μέσα που χρησιμοποιούνται στην διεξαγωγή της εκπαιδευτικής διαδικασίας, αλλά και την αλληλεπίδραση και τις διαπροσωπικές σχέσεις που αναπτύσσουν με το εκπαιδευτικό προσωπικό, στοιχεία που στοχεύει να βελτιώσει η ανάλυση συναισθήματος.

Επιπρόσθετα, τα μοντέλα μηχανικής μάθησης, σε συνδυασμό με την εφαρμογή των μεθόδων NLP κατέστησαν εφικτή τη δημιουργία ισχυρών και αποτελεσματικών εργαλείων που είναι σε θέση να προσαρμοστούν σε διαφορετικές γλώσσες και ποικίλα εκπαιδευτικά πλαίσια. Συγκεκριμένα για την ελληνική γλώσσα, παρουσιάστηκε πως πρέπει να επιλυθούν αρκετές προκλήσεις λόγω της πολυπλοκότητας και μοναδικότητας του ελληνικού αλφαβήτου, αλλά και της δομής της γραμματικής, καθιστώντας πιο χρονοβόρα και απαιτητική την ανάπτυξη μοντέλων μηχανικής μάθησης και ανάλυσης συναισθημάτων για την ελληνική γλώσσα. Παρά τις αντιξοότητες αυτές, όμως, οι τεχνικές αυτές ευδοκίμησαν και αποδείχθηκαν άκρως αποτελεσματικές στην ορθή αποτύπωση της συναισθηματικής πολικότητας (θετικό, αρνητικό και ουδέτερο)σε αξιολογήσεις επάνω στην εκπαίδευση.

Η ανάλυση συναισθημάτων διατηρεί αλλά και ενισχύει την ικανότητα των εκπαιδευτικών ιδρυμάτων να αξιολογούν τις αδυναμίες και τις ευστοχίες τους. Τα ευρήματα της παρούσας εργασίας δείχνουν ότι οι φοιτητές/τριες έχουν σε ιδιαίτερη εκτίμηση την αλληλεπίδραση με

τους/τις εκπαιδευτικούς τους και την ποιότητα των μερών και υλικών που συμμετέχουν στη διεξαγωγή της διδακτικής διαδικασίας, εν αντιθέσει με τις αρνητικές αξιολογήσεις, που αφοσιώνονται κυρίως στην έλλειψη των προαναφερθέντων και της ευελιξίας της διδακτικής διαδικασίας. Μέσω της ανάλυσης και επεξεργασίας αυτής, τα εκπαιδευτικά ιδρύματα μπορούν να προσαρμόσουν τις εκπαιδευτικές στρατηγικές και την στοχοθεσία τους ώστε να προσφέρουν μια πιο ολοκληρωμένη και ικανοποιητική διδακτική εμπειρία.

Η παρούσα πτυχιακή εργασία μπορεί να θεωρηθεί ως ένα μικρό θεμέλιο για πιο αναλυτικές και πρακτικές εφαρμογές της συναισθηματικής ανάλυσης στον εκπαιδευτικό τομέα της Ελλάδας. Κυριότερο πεδίο εφαρμογής είναι η αξιολόγηση των εκπαιδευτικών λειτουργιών σε πραγματικό χρόνο όπου οι φοιτητές και οι φοιτήτριες είναι σε θέση να παρέχουν άμεση ανατροφοδότηση και σε πραγματικό χρόνο, αξιοποιώντας ψηφιακά μέσα και πλατφόρμες όπου η ανατροφοδότηση αυτή αναλύεται και επεξεργάζεται άμεσα. Αυτό καθιστά αμεσότερη και ταχύτερη την ανταπόκριση των εκπαιδευτικών ιδρυμάτων στις ανάγκες και τους προβληματισμούς των φοιτητών/τριών.

Η ανάλυση συναισθήματος μπορεί να αξιοποιηθεί και για την αξιολόγηση της απόδοσης των εκπαιδευτικών, αφού προσφέρει μια αντικειμενική μέτρηση και εικόνα του βαθμού ικανοποίησης των εκπαιδευομένων μέσω των μεθόδων διδασκαλίας και της γενικής συμπεριφοράς τους. Σε ένα μελλοντικό πλαίσιο, η χρήση τέτοιων τεχνικών μπορεί ακόμα και να οδηγήσει σε πιο προσαρμοσμένα εκπαιδευτικά προγράμματα που να καλύπτουν, πιο εξατομικευμένα, τις ανάγκες των εκπαιδευομένων.

Η μελλοντική έρευνα μπορεί να επικεντρωθεί στην ανάπτυξη πιο εξειδικευμένων μοντέλων ανάλυσης συναισθήματος για την ελληνική γλώσσα, με στόχο να ξεπεραστούν οι τρέχουσες προκλήσεις. Η δημιουργία νέων σετ δεδομένων και η βελτίωση των τεχνικών ανάλυσης και επεξεργασίας κειμένου (ανάλυση, προεπεξεργασία, tokenization, decluttering) μπορεί να αυξήσει την ακρίβεια και την αποτελεσματικότητα των μοντέλων και να διευρύνει την χρήση της ανάλυσης συναισθήματος σε διάφορα και πιθανώς πιο εξειδικευμένα πεδία της εκπαίδευσης. Επιπλέον, η ενσωμάτωση διαφόρων ερεθισμάτων, όπως οπτικοακουστικό υλικό, εν παραδείγματι από τις ζωντανές διαλέξεις, θα μπορούσε να ανοίξει το δρόμο για μια νέα κατεύθυνση που θα βελτιώσει την ποιότητα των αναλύσεων

Συμπερασματικά, η ανάλυση συναισθήματος είναι μια ισχυρή μεθοδολογία που δύναται να προσφέρει ανεκτίμητες πληροφορίες σχετικά με τις συναισθηματικές αντιδράσεις των φοιτητών στην εκπαιδευτική τους εμπειρία. Η παρούσα εργασία συμβάλλεις την καλύτερη δυνατή κατανόηση αυτής της μεθοδολογίας και δείχνει πως μπορεί να αξιοποιηθεί για την βελτίωση και αναβάθμιση της ποιότητας της εκπαίδευση στην τριτοβάθμια εκπαίδευση. Με την κατάλληλη προσαρμογή και εκσυγχρόνιση των μεθόδων αυτών, αλλά και την εξατομικευμένη προσαρμογή τους στις ανάγκες και τους σκοπούς του εκάστοτε εκπαιδευτικού ιδρύματος, τα εκπαιδευτικά ιδρύματα μπορούν να εξασφαλίσουν μια πιο θετική, ευχάριστη και υποστηρικτική μαθησιακή εμπειρία για όλα τα μετέχοντα μέλη στη διαδικασία αυτή.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Abhishek. (2024, August 12). How part-of-speech tag, dependency and constituency parsing aid in understanding text data?. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/07/part-of-speechpos-tagging-dependency-parsing-and-constituency-parsing-in-nlp/.

Arvinthsss.(2023, January 27). Python School Students Report project. https://www.kaggle.com/code/arvinthsss/python-school-students-report-project/notebook/ Retrieved on July 15th, 2024.

Bordoloi, M., & Biswas, S. K. (2023). Sentiment analysis: A survey on design framework, applications and future scopes. *Artificial Intelligence Review*, *56*(11), 12505–12560. https://doi.org/10.1007/s10462-023-10442-2.

Dervenis, C. (2024). Assessing Teacher Competencies in Higher Education: A Sentiment Analysis of Student Feedback. *International Journal of Information and Education Technology*, *14*(4), 533–541. https://doi.org/10.18178/ijiet.2024.14.4.2074.

Ezgi Gökdemir. (2023, December 24). What is Tokenization? - SabancıDx - Medium. Medium; SabancıDx. https://medium.com/sabancidx/what-is-tokenization-a60c77ea6424.

Geeksforgeeks. (2019, January 28). NLP | Part of Speech - Default Tagging. GeeksforGeeks. https://www.geeksforgeeks.org/nlp-part-of-speech-default-tagging/ Retrieved on January 3rd, 2024.

Hoar, B. B., Ramachandran, R., Levis-Fitzgerald, M., Sparck, E. M., Wu, K., & Liu, C. (2023). Enhancing the Value of Large-Enrollment Course Evaluation Data Using Sentiment Analysis. *Journal of Chemical Education*, *100*(10), 4085–4091. https://doi.org/10.1021/acs.jchemed.3c00258.

Karabiber, F. (n.d.). TF-IDF — Term Frequency-Inverse Document Frequency – LearnDataSci. Www.learndatasci.com. https://www.learndatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/

Kastrati, Z., Dalipi, F., Imran, A. S., Pireva Nuci, K., & Wani, M. A. (2021). Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study. *Applied Sciences*, *11*(9), 3986. https://doi.org/10.3390/app11093986.

Kumar, A., & Sebastian, T. M. (2012). Sentiment Analysis: A Perspective on its Past, Present and Future. *International Journal of Intelligent Systems and Applications*, *4*(10), 1–14. https://doi.org/10.5815/ijisa.2012.10.01.

Lasri, I., Riadsolh, A., & Elbelkacemi, M. (2023). Real-time Twitter Sentiment Analysis for Moroccan Universities using Machine Learning and Big Data Technologies. *International Journal of Emerging Technologies in Learning (IJET)*, *18*(05), 42–61. https://doi.org/10.3991/ijet.v18i05.35959.

Lasri, I., Riadsolh, A., & ElBelkacemi, M. (2023). Self-Attention-Based Bi-LSTM Model for Sentiment Analysis on Tweets about Distance Learning in Higher Education. *International Journal of Emerging Technologies in Learning (IJET)*, *18*(12), 119–141. https://doi.org/10.3991/ijet.v18i12.38071.

Li, L., Johnson, J., Aarhus, W., & Shah, D. (2022). Key factors in MOOC pedagogy based on NLP sentiment analysis of learner reviews: What makes a hit. *Computers & Education*, *176*, 104354. https://doi.org/10.1016/j.compedu.2021.104354.

Medallia. (n.d.). *Real time text analytics software*. https://monkeylearn.com/sentiment-analysis/ Retrieved on December 20th, 2023.

Nikolić, N., Grljević, O., & Kovačević, A. (2020). Aspect-based sentiment analysis of reviews in the domain of higher education. *The Electronic Library*, *38*(1), 44–64. https://doi.org/10.1108/EL-06-2019-0140.

Pooja, & Bhalla, R. (2022). A Review Paper on the Role of Sentiment Analysis in Quality Education. *SN Computer Science*, *3*(6), 469. https://doi.org/10.1007/s42979-022-01366-9.

Remali, N. A. S., Shamsuddin, M. R., & Abdul-Rahman, S. (2022). Sentiment Analysis on Online Learning for Higher Education During Covid-19. *2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, 142–147. https://doi.org/10.1109/AiDAS56890.2022.9918788.

Rinaldi, A. M., Russo, C., & Tommasino, C. (2023). Automatic image captioning combining natural language processing and deep neural networks. Results in Engineering, 18, 101107. https://doi.org/10.1016/j.rineng.2023.101107.

Rokkou, G., Spatiotis, N., Triantafyllou, V., & Paraskevas, M. (2021). A new approach applying Sentiment Analysis in Greek Language. *25th Pan-Hellenic Conference on Informatics*, 235–241. https://doi.org/10.1145/3503823.3503867.

S M, D., N, R., K, S., & Pareek, P. K. (2022). Machine Learning based Education System with Sentiment Analysis for Students. *2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon)*, 1–6. https://doi.org/10.1109/MysuruCon55714.2022.9972555.

Saemani, R. A., & Setiyawati, N. (2022). Analisis Sentimen Permendikbud Pencegahan Dan Penanganan Kekerasan Seksual Di Lingkungan Perguruan Tinggi Menggunakan Support Vector Machine(SVM). *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *8*(1), 65–71. https://doi.org/10.33480/jitk.v8i1.2807.

Senadhira, K. I., Rupasingha, R. A. H. M., & Kumara, B. T. G. S. (2022). Sentiment Analysis on Twitter Data Related to Online Learning During the Covid-19 Pandemic. *2022 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, 131–136. https://doi.org/10.1109/SCSE56529.2022.9905190.

Sentiment Analysis 101. (n.d.). KDnuggets. https://www.kdnuggets.com/2015/12/sentiment-analysis-101.html. Retrieved on February 18th, 2024.

Spatiotis, N., Mporas, I., Paraskevas, M., & Perikos, I. (2016). Sentiment Analysis for the Greek Language. *Proceedings of the 20th Pan-Hellenic Conference on Informatics*, 1–4. https://doi.org/10.1145/3003733.3003769.

Us, Y., Pimonenko, T., Lyulyov, O., Chen, Y., & Tambovceva, T. (2022). Promoting Green Brand of University in Social Media: Text Mining and Sentiment Analysis. *Virtual Economics*, *5*(1), 24–42. https://doi.org/10.34021/ve.2022.05.01(2).

Usart, M., Grimalt-Álvaro, C., & Iglesias-Estradé, A. M. (2023). Gender-sensitive sentiment analysis for estimating the emotional climate in online teacher education. *Learning Environments Research*, *26*(1), 77–96. https://doi.org/10.1007/s10984-022-09405-.

Wen, Y., Zhao, X., Li, X., & Zang, Y. (2023). Explaining the Paradox of World University Rankings in China: Higher Education Sustainability Analysis with Sentiment Analysis and LDA Topic Modeling. *Sustainability*, *15*(6), 5003. https://doi.org/10.3390/su15065003.

What is overfitting?. IBM. (2021, October 15). https://www.ibm.com/topics/overfitting. Retrieved on January 26th , 2024.

ΠΑΡΑΡΤΗΜΑ Α

```
import numpy as np

import pandas as pd

import nltk

import matplotlib.pyplot as plt

from wordcloud import WordCloud

from sklearn.feature_extraction.text import TfidfVectorizer

from sklearn.model_selection import train_test_split, GridSearchCV, cross_val_score

from sklearn.svm import SVC

from sklearn.metrics import classification_report, accuracy_score, confusion_matrix

from imblearn.over_sampling import SMOTE

import logging

from sklearn.utils import parallel_backend

from joblib import parallel_backend

from sklearn import metrics

import seaborn as sns



def process_data():

    with parallel_backend('threading'):

        verbose=0



# Configure logging, deixnei info level minimata  opou kai ta kanei assign sto GridSearchCV



logging.basicConfig(level=logging.INFO)
```

```
logger = logging.getLogger('GridSearchCV')


# Load feedback data

df = pd.read_excel('C:/Users/Koe/Desktop/finalDataset0_2.xlsx')


# Convert text to lowercase and ensure it is string type

df['text'] = df['text'].astype(str).str.lower()


# Tokenization

df['tokens'] = df['text'].apply(nltk.word_tokenize)


# Remove stopwords

stopwords = nltk.corpus.stopwords.words('english')

df['tokens'] = df['tokens'].apply(lambda x: [word for word in x if word not in stopwords])


X = df['text']

y = df['sentiment']


# Check class balance

print("\nClass distribution:\n")

print(y.value_counts())


# Word Cloud for Ratings | theyre only used to show how many iterations of each category
were found between [-1,1]

feedbackratings = df['sentiment']

counts = feedbackratings.value_counts()
```

```
counts.index = counts.index.map(str)


# Word Cloud for Comments

feedbackcomments = df['text']

textonly2 = feedbackcomments.to_string(index=False)

print("\n")

print(feedbackcomments.head())


# Plotting word clouds

wc1 = WordCloud(width=800, height=400,
background_color='white').generate_from_frequencies(counts)

plt.figure(figsize=(10, 6))

plt.imshow(wc1, interpolation='bilinear')

plt.axis('off')

plt.title('Word Cloud - Sentiment Ratings Frequency')


wc2 = WordCloud(width=800, height=400, background_color='white').generate(textonly2)

plt.figure(figsize=(10, 6))

plt.imshow(wc2, interpolation='bilinear')

plt.axis('off')

plt.title('Word Cloud - Comments')


plt.show()


# Train-test splitting
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)


# Vectorization


vectorizer = TfidfVectorizer()

X_train_vectors = vectorizer.fit_transform(X_train)

X_test_vectors = vectorizer.transform(X_test)


# Xrisi SMOTE wste na apaloifthoun anisoropies

smote = SMOTE(random_state=42)

X_train_vectors, y_train = smote.fit_resample(X_train_vectors, y_train)


# Hyperparameter tuning mesw tou GridSearchCV diladi gia ta beltisa dinata apotelesmata


param_grid = {'C': [0.1, 1, 10, 100], 'gamma': [1, 0.1, 0.01, 0.001], 'kernel': ['linear', 'rbf']}


# Afto edw to xrisimopoiisa giati mou ebgaze 150 minimata san : [CV] END...C=100, gamma=0.01, kernel=rbf; totaltime=0.0s

with parallel_backend('threading'):

    grid = GridSearchCV(SVC(), param_grid, refit=True, verbose=0)  # Set verbose=0 to suppress detailed output

    grid.fit(X_train_vectors, y_train)


print(f"\nBest Parameters: {grid.best_params_}\n")


# Ekpaidefsi montelou me tis kaliteres parametrous mesw GridSearch
```

```
model = SVC(C=grid.best_params_['C'], gamma=grid.best_params_['gamma'],
kernel=grid.best_params_['kernel'], class_weight='balanced')

model.fit(X_train_vectors, y_train)


# 5-fold Cross validation scores, wste na exw 80:20 ratio giati exw mikro sample

cv_scores = cross_val_score(model, X_train_vectors, y_train, cv=5)

print(f"Cross-validation scores: {cv_scores}\n")

print(f"Mean cross-validation score: {np.mean(cv_scores)}\n")



# Problepsi

y_pred = model.predict(X_test_vectors)


# Emfanisi metrikwn


print("Classification Report:")

print("\n")

print(classification_report(y_test, y_pred, zero_division=0))


print("\nAccuracy Score:", accuracy_score(y_test, y_pred))


# Print confusion matrix


cf_matrix = confusion_matrix(y_test, y_pred)

sns.heatmap(cf_matrix/np.sum(cf_matrix), annot=True,  cmap='Greens')
```

```
print("\n")


plt.show()
```