

Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Μεταπτυχιακό Πρόγραμμα Σπουδών *Κοινωνικά Πληροφοριακά
Συστήματα*

Μεταπτυχιακή Διατριβή



Ανάλυση των Συναισθημάτων της Κοινής Γνώμης στα Μέσα Κοινωνικής
Δικτύωσης

Μαρία Κωνσταντίνου

Επιβλέπων Καθηγητής
Κλήμης Νταλιάνης

Μάιος 2021

Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Μεταπτυχιακό Πρόγραμμα Σπουδών *Κοινωνικά Πληροφοριακά
Συστήματα*

Μεταπτυχιακή Διατριβή

Ανάλυση των Συναισθημάτων της Κοινής Γνώμης στα Μέσα Κοινωνικής
Δικτύωσης

Μαρία Κωνσταντίνου

Επιβλέπων Καθηγητής
Κλήμης Νταλιάνης

Η παρούσα μεταπτυχιακή διατριβή υποβλήθηκε προς μερική εκπλήρωση των
απαιτήσεων για απόκτηση μεταπτυχιακού τίτλου σπουδών
Στα Κοινωνικά Πληροφοριακά Συστήματα
από τη Σχολή Θετικών και Εφαρμοσμένων Επιστημών
του Ανοικτού Πανεπιστημίου Κύπρου.

Μάιος 2021

Περίληψη

Αντικείμενο της παρούσας μεταπτυχιακής διατριβής είναι η ανάλυση των συναισθημάτων της κοινής γνώμης στα μέσα κοινωνικής δικτύωσης. Η συλλογή και ανάλυση γραπτών μηνυμάτων τα οποία ανταλλάζουν οι χρήστες, μπορεί να προσφέρει σημαντικά στοιχεία τόσο στην επιστήμη της πληροφορικής όσο και σε πολλούς άλλους τομείς.

Κύριο θέμα της διατριβής είναι η παρουσίαση των πιο διαδεδομένων μεθόδων εξαγωγής συναισθημάτων από μη δομημένα κείμενα και η χρήση κάποιων εργαλείων για τον σκοπό αυτό. Πραγματοποιήθηκαν πειράματα εξαγωγής συναισθημάτων μέσα από γραπτά μηνύματα χρηστών στο Twitter με τη χρήση κατάλληλου λογισμικού και τα αποτελέσματα παρουσιάστηκαν και αξιολογήθηκαν.

Από τα αποτελέσματα της ανάλυσης, δεν έχουν εξαχθεί γενικευμένα συμπεράσματα όσον αφορά το είδος και την πολικότητα των συναισθημάτων. Επίσης το είδος του συναισθήματος, θετικό ή αρνητικό, και η επιτυχία της κατηγοριοποίησης του, διαφάνηκε να κρίνεται πιο πολύ από την ποιότητα και την εκφραστικότητα κάθε μηνύματος.

Προεκτάσεις της τρέχουσας μελέτης μπορεί να είναι η χρήση περισσότερων συνόλων δεδομένων για την εξαγωγή συναισθημάτων και η χρήση περισσότερων εργαλείων από τα ίδια σύνολα. Επιπλέον η συγγραφή κατάλληλου λογισμικού το οποίο θα υλοποιεί μία ή περισσότερες μεθόδους εξαγωγής συναισθημάτων, θα ήταν βοηθητικό για την εξαγωγή περισσότερων συμπερασμάτων.

Summary

The purpose of this dissertation is to analyse the feelings of public opinion on social media. The collection and analysis of text messages exchanged by users can provide important data to the computer science and many other fields.

The main topic of the dissertation is the presentation of the most common methods of extracting emotions from unstructured texts and the use of some tools for this purpose. Using appropriate software emotion experiments were performed through text messages of users on Tweeter the results of the experiments were presented and evaluated.

From the results of the analysis, no general conclusions have been drawn regarding the type and polarity of emotions. Also, the type of emotion, positive or negative, and the success of the categorisation, as it became clear that it was judged by the quality and expressiveness of each message.

Extensions of the current study may be the use of more data sets to extract emotions and the use of more tools from the same sets. In addition, writing appropriate software that implements one or more methods of extracting emotions would be helpful in drawing more conclusions.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα της διπλωματικής μου εργασίας κ. Κλήμη Νταλιάνη, για την βοήθεια και την κατανόηση του.

Επίσης, θα ήθελα να ευχαριστήσω την οικογένεια μου για την στήριξη της καθ' όλη τη διάρκεια των σπουδών μου.

Περιεχόμενα

Εισαγωγή	1
1.1 Σκοπός και κίνητρο της τρέχουσας Διατριβής	2
1.2 Δομή της Διατριβής	3
Βασικές έννοιες της εξαγωγής συναισθημάτων μέσω της Επεξεργασίας Φυσικής Γλώσσας	2
2.1 Η επεξεργασία Φυσικής Γλώσσας.....	2
2.2 Ανάλυση συναισθημάτων	3
2.3 Μεθοδολογίες	5
2.4 Παρουσίαση παρόμοιων μελετών	6
Μηχανική Μάθηση	8
3.1 Εισαγωγή	8
3.2 Τεχνητά Νευρωνικά Δίκτυα	10
3.3 Εξελικτικοί Αλγόριθμοι	14
3.4 Στατιστικές Μέθοδοι.....	16
3.5 Σημασιολογικά Δίκτυα	17
3.6 Σύγκριση Μεθόδων	17
Μέθοδοι εξαγωγής συναισθημάτων μέσω ανάλυσης φυσικής γλώσσας	20
4.1 Εισαγωγή.....	20
4.2 Ανάλυση των emoticons	20
4.3 LIWC	22
4.4 SentiStrength	23
4.5 SentiWordNet	23
4.6 SenticNet.....	23
4.7 SASA	24
4.8 HappinessIndex.....	24
4.9 PANAS-t.....	25
4.10 EmoTxt	26
4.11 Σύνοψη των μεθόδων	26
Παρουσίαση της μεθοδολογίας	29
5.1 Εισαγωγή	29
5.2 Το λογισμικό Dcipher Analytics	29
Παρουσίαση των δεδομένων προς ανάλυση	36
6.1 Εισαγωγή	36

6.2 Το σύνολο δεδομένων WASSA-2017	36
6.3 Το σύνολο δεδομένων Pre-processed Twitter tweets.....	39
Πειραματική Μελέτη των δεδομένων προς ανάλυση.....	43
7.1 Εισαγωγή	43
7.2 Εκτέλεση των πειραμάτων.....	43
7.3. Ανάλυση των πειραματικών αποτελεσμάτων	52
7.4. Παρουσίαση των πειραματικών αποτελεσμάτων	52
Συμπεράσματα	56
8.1 Εξαγωγή των συμπερασμάτων με βάση τα πειραματικά αποτελέσματα.....	56
8.1. Σύγκριση των αποτελεσμάτων με παρόμοιες διατριβές.....	57
8.3. Επίλογος και πιθανές προεκτάσεις.....	57
Παραρτήματα Α.....	59
Το σύνολο δεδομένων WASSA-2017	59
A.1 Απόσπασμα των πειραματικών αποτελεσμάτων για το συναίσθημα “θυμός”	59
A.2 Απόσπασμα των πειραματικών αποτελεσμάτων για το συναίσθημα “φόβος”	63
A.3 Απόσπασμα των πειραματικών αποτελεσμάτων για το συναίσθημα “χαρά”	66
A.4 Απόσπασμα των πειραματικών αποτελεσμάτων για το συναίσθημα “θλίψη”	70
Βιβλιογραφία	74

Κεφάλαιο 1

Εισαγωγή

Τα μέσα κοινωνικής δικτύωσης δίνουν την δυνατότητα στους χρήστες τους να συζητήσουν διάφορα θέματα, μέσω μιας πλατφόρμας η οποία είναι προσβάσιμη στο ευρύ κοινό, και να εκφράσουν την άποψή τους σε αυτά. Η συλλογή και η ανάλυση των συζητήσεων των χρηστών δίνουν σημαντικές πληροφορίες σε φορείς όπως κυβερνητικούς οργανισμούς, επιχειρήσεις αλλά και μη κοινωφελείς οργανώσεις (C. Paris, D. Georgeakopoulos 2015: 3-24) (Wang, R., D. Zhou, M. Jiang, J. Si and Y. Yang 2019: 41101-41124). Οι πληροφορίες αυτές μπορούν να χρησιμοποιηθούν στην συνέχεια σε μία πληθώρα εφαρμογών, όπως για παράδειγμα την πιο εύστοχη σχεδίαση ενημερωτικών εκστρατειών, την καλύτερη κατανόηση των καθημερινών προβλημάτων των πολιτών, τη δημιουργία νέων υπηρεσιών και προϊόντων, τη βελτίωση της επικοινωνίας με τους πελάτες κατά την προσφορά υπηρεσιών, κ.α.

Μια βασική πτυχή έρευνας των μηνυμάτων των κοινωνικών μέσων, είναι η ικανότητα ταξινόμησης τους για τον προσδιορισμό των υποκείμενων συναισθημάτων και γνώμης τα οποία εκφράζουν.

Αξιοσημείωτο βέβαια είναι, το γεγονός ότι υπάρχουν λεπτές αλλά σημαντικές διαφορές μεταξύ μίας γνώμης και ενός συναισθήματος. Από ψυχολογικής άποψης, τα συναισθήματα αποτελούν έκφραση μιας σύνθετης προσωπικής κατάστασης η οποία επηρεάζεται από διάφορους παράγοντες, όπως οι εμπειρίες του παρελθόντος, οι υποκειμενικές απόψεις, οι νόρμες και τα πιστεύω. Από την άλλη, η γνώμη αφορά συνήθως μια υποκειμενική άποψη για ένα θέμα.

Μπορούμε επίσης να πούμε ότι η γνώμη ενός ανθρώπου πάνω σε ένα θέμα είναι μία έκφραση της συναισθηματικής του κατάστασης, μέσα από ώριμη σκέψη και επεξεργασία. Όπως είναι φυσικό, τόσο η έκφραση των συναισθημάτων όσο και η γνώμη μιας ομάδας του πληθυσμού, έχουν ιδιαίτερο ενδιαφέρον από ερευνητικής άποψης με τη γνώμη να κατέχει κάποια αυξημένη βαρύτητα.

Στην συντριπτική πλειοψηφία των περιπτώσεων, η ανάλυση των συναισθημάτων των χρηστών ενός κοινωνικού δικτύου γίνεται μέσω των γραπτών μηνυμάτων τα οποία εισάγονται από τους ίδιους. Φυσικά σε ορισμένα κοινωνικά δίκτυα (π.χ. YouTube) θα μπορούσε να γίνει ανάλυση περαιτέρω δεδομένων (π.χ. video) αν και ο τομέας αυτός είναι εξαιρετικά περίπλοκος συγκριτικά με την ανάλυση των δεδομένων γραπτών μηνυμάτων. Ωστόσο, η ανάλυση των συναισθημάτων των χρηστών θέτει επιπλέον προκλήσεις, πέραν των τεχνικών της κατηγοριοποίησης κειμένου και της ομαδοποίησης στα σχόλια τους. Η σωστή εξαγωγή των συναισθημάτων από τα σχόλια των χρηστών εξαρτάται από το ειδικό κοινωνικό πλαίσιο του κάθε χρήστη, από την ιδιοσυγκρασία του, από το λεξιλόγιο το οποίο χρησιμοποιεί, τη σύνταξη καθώς και από το προσωπικό στυλ του ατόμου. Οι μελέτες που έγιναν σε αρχικό στάδιο στον τομέα της ανάλυσης συναισθήματος (Turney, P. D. 2002: 417-424) κατέδειξαν ότι η κατηγοριοποίηση συναισθήματος (sentiment classification) διαφέρει από την απλή κατηγοριοποίηση κειμένου. Επιπλέον προκλήσεις είναι η χρήση άτυπης γλώσσας, η χρήση διαφορετικών ορολογιών, οι συντομογραφίες και η μη τυπική γραμματική.

1.1 Σκοπός και κίνητρο της τρέχουσας Διατριβής

Η τρέχουσα διατριβή έχει ως σκοπό να προσθέσει στην ελληνική και κυπριακή βιβλιογραφία μία μελέτη για την εξαγωγή των συναισθημάτων από μη δομημένο κείμενο και πιο συγκεκριμένα από μηνύματα τα οποία έχουν δημοσιευθεί σε κοινωνικής δικτύωσης. Η διατριβή σκοπεύει να καλύψει το κενό το οποίο υπάρχει, στην ελληνική βιβλιογραφία, όσον αφορά την σύγκριση των μεθόδων εξαγωγής συναισθημάτων από μηνύματα κοινωνικής δικτύωσης. Πιο συγκεκριμένα, το βασικό ερώτημα το οποίο η διατριβή προσπαθεί να απαντήσει είναι το εξής: “αποτελούν οι τεχνικές Μηχανικής Μάθησης την μέθοδο η οποία ενδείκνυται περισσότερο για την εξαγωγή συναισθημάτων από μηνύματα τα οποία έχουν δημοσιευθεί σε μέσα κοινωνικής δικτύωσης”;

Για την επίτευξη του παραπάνω στόχου, έχει γίνει χρήση δύο κοινώς αποδεκτών συνόλων δεδομένων, τα οποία έχουν χρησιμοποιηθεί σε παρόμοιες έρευνες. Σε αντίθεση, οι άλλες έρευνες της ελληνικής βιβλιογραφίας (Ανεστάκης Α., 2015, Μηταρά Ε., 2017, Μώρος Ν., 2020), χρησιμοποιούν δεδομένα τα οποία έχουν ανακτηθεί από το Διαδίκτυο, κυρίως με την διαδικασία του web scraping, χωρίς αυτά να έχουν χρησιμοποιηθεί σε άλλες έρευνες. Τα σύνολα δεδομένα τα οποία κάνει χρήση η εν λόγω διατριβή, έχουν αναλυθεί με την χρήση του εργαλείου Dcipher Analytics, το οποίο παρέχει τόσο

παραδοσιακές μεθόδους όσο και μεθόδους Μηχανικής Μάθησης για την εξαγωγή συναισθημάτων από μη δομημένα κείμενα.

1.2 Δομή της Διατριβής

Η δομή της εν λόγω διατριβής είναι η εξής:

Στο 2^ο κεφάλαιο , γίνεται μια περιγραφή της ανάλυσης συναισθημάτων μέσω της ανάλυσης της φυσικής γλώσσας. Παρουσιάζονται οι μεθοδολογίες που προτείνονται για ανάλυση συναισθημάτων και γίνεται αναφορά σε άλλες μελέτες/ έρευνες σχετικές με το υπό μελέτη αντικείμενο.

Το 3^ο κεφάλαιο, παρουσιάζεται η μηχανική μάθηση ως εργαλείο για την εξαγωγή συναισθημάτων και την εξόρυξη γνώσης από κείμενα. Γίνεται αναφορά στις μεθόδους Τεχνικά Νευρωνικά Δίκτυα, τους Εξελικτικούς Αλγόριθμους, τις Στατιστικές Μεθόδους και τα Σηματολογικά Δίκτυα.

Στο 4^ο κεφάλαιο δίδεται η περιγραφή οκτώ μεθόδων ανάλυσης συναισθημάτων οι οποίες θεωρούνται ως οι πιο διαδομένες, καθώς και μια μεταξύ τους σύγκριση ως προς την καταλληλότητα τους αναλόγως της περίπτωσης χρήσης και του επιθυμητού αποτελέσματος.

Στο 5^ο κεφάλαιο γίνεται παρουσίαση της μεθοδολογίας που ακολουθήθηκε για τη συλλογή και ανάλυση των πειραμάτων και την εξαγωγή συναισθημάτων από μηνύματα του κοινωνικού δικτύου Twitter. Συγκεκριμένα παρουσιάζεται βήμα- βήμα η χρήση της δωρεάν έκδοσης του λογισμικού Dcipher Analytics.

Στο κεφάλαιο 6 παρουσιάζονται τα 2 σύνολα δεδομένων τα οποία χρησιμοποιήθηκαν προς ανάλυση και αφορούν αμφότερα συλλογή μηνυμάτων τα οποία έχουν δημοσιευθεί στο Twitter. Γίνεται παρουσίαση του δημοφιλούς συνόλου WASSA-2017 καθώς και ενός συνόλου μηνυμάτων το οποίο είναι διαθέσιμο στη πλατφόρμα ανάλυσης δεδομένων Kaggle.

Στο 7^ο κεφάλαιο παρουσιάζεται η εκτέλεση των πειραμάτων στα προαναφερθέντα σύνολα με την χρήση του Dcipher Analytics. Η εξαγωγή συναισθήματος έγινε με δυο διαφορετικούς τρόπους, με την χρήση κατάλληλου λεξικού λέξεων και με την χρήση μηχανικής μάθησης. Στη συνέχεια παρουσιάζονται τα πειραματικά αποτελέσματα αριθμητικά και διαγραμματικά.

Στο 8^ο κεφάλαιο παρουσιάζονται τα συμπεράσματα τα οποία προέκυψαν και οι πιθανές προεκτάσεις της παρούσας μεταπτυχιακής διατριβής.

Κεφάλαιο 2

Βασικές έννοιες της εξαγωγής συναισθημάτων μέσω της Επεξεργασίας Φυσικής Γλώσσας

2.1 Η επεξεργασία Φυσικής Γλώσσας

Η Επεξεργασία Φυσικής Γλώσσας (NLP – Natural Language Processing) αποτελεί έναν διεπιστημονικό κλάδο ο οποίος εφάπτεται της κλασικής επιστήμης της Πληροφορικής, κατέχοντας στοιχεία τα οποία πηγάζουν από την Τεχνητή Νοημοσύνη, όπως επίσης και από τον επιστημονικό τομέα της Υπολογιστικής Γλωσσολογίας. Η επεξεργασία φυσικής γλώσσας εστιάζει στις αλληλεπιδράσεις μεταξύ των υπολογιστών και των φυσικών (ανθρώπινων) γλωσσών. Κατά συνέπεια, τα προβλήματα τα οποία προσπαθεί να επιλύσει άπτονται μεταξύ άλλων της κατανόησης της, της εξαγωγής νοήματος από ανθρώπινα κείμενα, της κατηγοριοποίησης κειμένου ανάλογα με το περιεχόμενό του, της εξαγωγή συναισθημάτων από ανθρώπινα μηνύματα, κ.α.

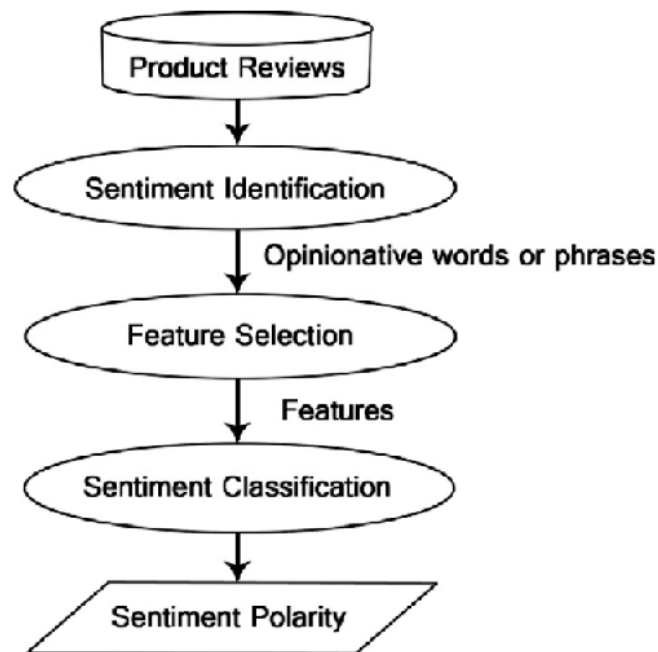
Στην λίστα που ακολουθεί παρουσιάζονται τα συχνότερα πεδία έρευνας της επεξεργασίας φυσικής γλώσσας. Κάποια από τα πεδία έρευνας έχουν εφαρμογές στην καθημερινή ζωή, ενώ άλλα υφίστανται ως υποβοηθητικά αντικείμενα έρευνας για την επίλυση μεγαλύτερων ζητημάτων.

- **Επεξεργασία και Ανάλυση κειμένου:** αποτελεί την πλέον συνηθισμένη εφαρμογή της επεξεργασίας φυσικής γλώσσας. Ο τομέας εστιάζει στην αναγνώριση της δομής του λόγου ενός κειμένου. Μερικά παραδείγματα είναι η φύση των σχέσεων του λόγου μεταξύ δύο προτάσεων (π.χ. επεξήγηση, αντίθεση κτλ.) καθώς και η αναγνώριση και η κατηγοριοποίηση των γλωσσικών πράξεων σε ένα κομμάτι κειμένου.
- **Αναγνώριση ομιλίας:** Η αυτόματη μετατροπή του ανθρώπινου λόγου όπως σε κείμενο με την χρήση ηλεκτρονικού υπολογιστή.

- **Αυτόματη παραγωγή λεκτικών μορφημάτων (stemming):** Η κατάτμηση των λέξεων στα μορφήματά τους καθώς και η αναγνώριση και κατηγοριοποίηση αυτών. Η δυσκολία του συγκεκριμένου πεδίου εξαρτάται σε μεγάλο βαθμό από την περιπλοκότητα της μορφολογίας της υπό εξέταση φυσικής γλώσσας.
- **Αυτόματη περίληψη κειμένου:** Η δημιουργία μίας αναγνώσιμης περίληψης ενός κειμένου.
- **Εξαγωγή πληροφοριών:** Η ανάκτηση πληροφοριών από μη δομημένα ή ημιδομημένα δεδομένα.
- **Κατηγοριοποίηση λέξεων σε μέρη του λόγου:** Η αυτόματη κατηγοριοποίηση της κάθε λέξης ενός κειμένου στο σωστό μέρος του λόγου στο οποίο αυτή ανήκει.
- **Αυτοματοποιημένη μετάφραση:** Η αυτόματη μετάφραση ενός κειμένου από μία ανθρώπινη γλώσσα σε μία άλλη.
- **Οπτική αναγνώριση χαρακτήρων:** Μετατροπή των προτάσεων ενός τυπωμένου κειμένου σε ψηφιακή μορφή.
- **Παραγωγή φυσικής γλώσσας:** Η μετατροπή πληροφοριών οι οποίες είναι αποθηκευμένες σε ψηφιακή μορφή σε αναγνώσιμο φυσικό λόγο.
- **Σύνθεση και Παραγωγή ομιλίας:** Η αυτόματη, τεχνητή παραγωγή του ανθρώπινου λόγου από τους υπολογιστές.

2.2 Ανάλυση συναισθημάτων

Η Ανάλυση Συναισθημάτων (Sentiment Analysis) ή Εξόρυξη Γνώμης (Opinion Mining) είναι η υπολογιστική μελέτη των απόψεων, των στάσεων και των συναισθημάτων των ανθρώπων με σημείο αναφοράς μια συγκεκριμένη οντότητα. Η οντότητα μπορεί να είναι ο άνθρωπος, το σύνολο ανθρώπων, ένα γεγονός, ένα κοινωνικό – πολιτικό – οικονομικό ζήτημα, ένα αντικείμενο. Πολύ συχνά οι οντότητες αυτές είναι αναθεωρήσιμες και οι διαδικασίες αναθεώρησης τους βασίζονται στην ανάλυση συναισθημάτων. Οι δύο εκφράσεις – Ανάλυση Συναισθημάτων ή Εξόρυξη Γνώμης είναι σχεδόν ταυτόσημες. Η διαφορά τους έγκειται στο ότι η εξόρυξη γνώμης αναζητά την στάση του ανθρώπου απέναντι στο αντικείμενο που εξετάζεται ενώ η ανάλυση συναισθήματος επεκτείνεται και στα συναισθήματα που αναδεικνύονται από τις πηγές προς μελέτη. Η γενική διαδικασία της ανάλυσης συναισθήματος παρουσιάζεται στην παρακάτω εικόνα.



Εικόνα 1. Γενική διαδικασία της ανάλυσης συναισθήματος.

Η Ανάλυση Συναισθήματος ακολουθεί μια διαδικασία ταξινόμησης η οποία πραγματοποιείται σε τρία στάδια. Στο πρώτο στάδιο γίνεται η κατηγοριοποίηση σε επίπεδο εγγράφου, το δεύτερο στάδιο γίνεται σε επίπεδο χαρακτηριστικών του λόγου και σε τρίτο στάδιο γίνεται σε επίπεδο προσδιορισμού συναισθήματος. Στο πρώτο στάδιο ένα έγγραφο κατατάσσεται ανάλογα με το περιεχόμενο του αν αυτό εκφράζει θετική ή αρνητική γνώμη ή συναίσθημα. Στο στάδιο αυτό ολόκληρο το έγγραφο αντιμετωπίζεται ως οντότητα. Σε δεύτερο στάδιο η ανάλυση γίνεται σε επίπεδο προτάσεων και γίνεται προσπάθεια να ανιχνευθεί το συναίσθημα που εκφράζει. Αρχικά εντοπίζεται εάν είναι υποκειμενική ή αντικειμενική έκφραση. Στην πρώτη περίπτωση αναζητείται το αν έχει θετικό ή αρνητικό ύφος. Επειδή συχνά δεν υπάρχει διαφοροποίηση στην ταξινόμηση σε επίπεδο εγγράφων ή προτάσεων – καθώς στην πραγματικότητα οι προτάσεις είναι σύντομα έγγραφα - είναι απαραίτητο ένα ακόμη στάδιο. Στο τρίτο στάδιο της ανάλυσης αναζητείται η ταξινόμηση σε επίπεδο έκφρασης συναισθήματος. Σε αυτό το στάδιο λαμβάνονται υπ' όψη τα χαρακτηριστικά της εξεταζόμενης οντότητας (το αντικείμενο της γνώμης των συμμετεχόντων) αλλά και – ενδεχομένως – τα χαρακτηριστικά των πηγών δεδομένων. Οι πηγές αυτές είναι κυρίως σχόλια σε κοινωνικά δίκτυα ή ιστολόγια, αξιολογήσεις προϊόντων, καταστάσεων ή λοιπών οντοτήτων (Medhat, Hassan, Korashy, 2014: 1096-1097).

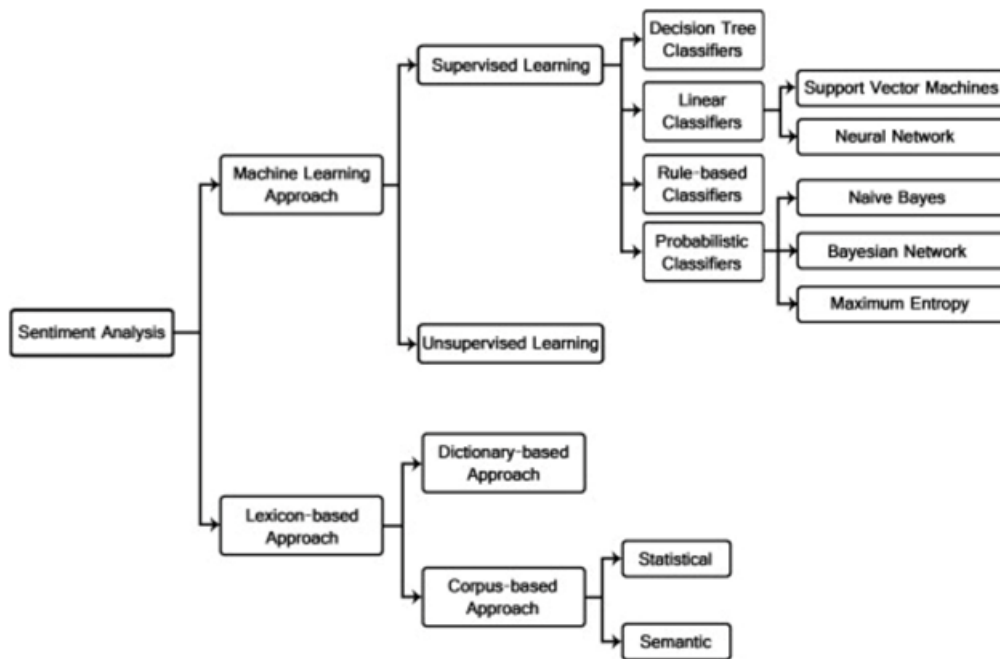
2.3 Μεθοδολογίες

Για την ανάλυση συναισθήματος έχουν προταθεί πολλές μέθοδοι. Επιχειρώντας την κατηγοριοποίηση τους μπορεί να διαμορφωθεί η ακόλουθη ταξινόμηση:

Διαδικασίες μηχανικής μάθησης: Πρόκειται για τεχνικές που μέσα από την επεξεργασία φυσικής γλώσσας αναζητείται να εντοπιστούν τα συναισθήματα που εκφράζουν τα κείμενα. Με τις διαδικασίες αυτές οι μηχανές είναι σε θέση να επεξεργάζονται ακολουθίες συμβολοσειρών και να επιστρέφουν τα συναισθήματα που αντιπροσωπεύουν. Οι τεχνικές αυτές χρησιμοποιούν ισχυρούς αλγορίθμους προκειμένου να εκπαιδεύονται οι μηχανές στην εκτίμηση συναισθημάτων όταν η είσοδος που δέχονται είναι κείμενα σε φυσική γλώσσα. Οι τεχνικές αυτές συχνά έχουν υψηλές απαιτήσεις σε προπαρασκευαστικές ενέργειες και υπολογιστικούς πόρους. Οι κατηγορίες των διαδικασιών που περιλαμβάνονται στην κατηγορία αυτή ακολουθούν την διάκριση της μηχανικής μάθησης σε εποπτευόμενη και μη εποπτευόμενη ενώ η εποπτευόμενη μηχανική μάθηση κατηγοριοποιείται περαιτέρω με βάση το είδος των αλγορίθμων που χρησιμοποιούνται (Llombart, 2017:1).

Διαδικασίες που βασίζονται σε λεξικά: Πρόκειται για διαδικασίες που βασίζονται στην ανάπτυξη πεπερασμένης έκτασης λεξικών αντιστοίχισης λέξεων και εκφράσεων με συγκεκριμένα συναισθήματα. Η αποδοτικότητα των τεχνικών αυτών βασίζεται σε μεγάλο βαθμό στην πληρότητα και την ευστοχία των λεξικών αυτών. Η περαιτέρω διάκριση τους περιλαμβάνει δύο κατηγορίες, τις τεχνικές που εξετάζουν λέξεις και εκείνες που εξετάζουν φράσεις (Musto, Semeraro, & Polignano, 2015).

Στην επόμενη εικόνα φαίνεται σχηματικά και συνοπτικά η διάκριση των τεχνικών ανάλυσης συναισθήματος.



Εικόνα 2. Η διάκριση των τεχνικών ανάλυσης συναισθήματος.

2.4 Παρουσίαση παρόμοιων μελετών

Στην ενότητα αυτή θα επικεντρωθούμε σε παρόμοιες με την παρούσα μελέτη έρευνες, κυρίως από συγγραφείς σε Ελλάδα και Κύπρο, και θα παρουσιάσουμε τις μεταξύ τους διαφορές.

Η έρευνα με τίτλο “Ανάλυση συναισθήματος από κείμενα, για οντότητες και χαρακτηριστικά αυτών” (Ανεστασάκης, 2015), επικεντρώνεται στην εξαγωγή συναισθημάτων από κείμενα και πιο συγκεκριμένα από σχόλια και αξιολογήσεις χρηστών από δύο ηλεκτρονικά καταστήματα. Η εξαγωγή των συναισθημάτων έγινε με συνδυασμό των λογισμικών DeixTo, NiosTo και Palo Pro. Τα συναισθήματα εξετάστηκαν κυρίως από την σκοπιά της πολικότητας των συναισθημάτων και πιο συγκεκριμένα κατά πόσο αυτά εκφράζουν θετικά ή αρνητικά συναισθήματα.

Σε έρευνα με τίτλο “Τεχνικές εξαγωγής και ανάλυσης συναισθημάτων σε μηνύματα του Twitter και υλοποίηση της ανάλυσης σε περιβάλλον cloud” (Μηταρά, 2017) έγινε εξαγωγή συναισθημάτων από μηνύματα τα οποία έχουν δημοσιευθεί σε κοινωνικά δίκτυα και πιο συγκεκριμένα στην υπηρεσία Twitter. Η εν λόγω έρευνα επικεντρώνεται περισσότερο στην επιστημονική μέθοδο η οποία ακολουθείται για την εξαγωγή συναισθημάτων και λιγότερο στην παρουσίαση των πειραματικών αποτελεσμάτων. Επιπρόσθετα, και αυτή η έρευνα εξετάζει κυρίως τα συναισθήματα από την πλευρά της πολικότητας (θετικά/αρνητικά).

Σε έρευνα με τίτλο “Ανάλυση συναισθήματος σε διαδικτυακές κριτικές υπηρεσιών φιλοξενίας” (Μώρος Ν., 2020) η εξαγωγή συναισθημάτων έγινε από σχόλια και γενικότερες κριτικές χρηστών, οι οποίες αφορούν υπηρεσίες φιλοξενίας (π.χ. ξενοδοχειακές μονάδες). Η συλλογή των δεδομένων έγινε με την χρήση τεχνικών Web Scraping, με την χρήση του λογισμικού Selenium, από την ιστοσελίδα booking.com. Στην εν λόγω έρευνα δίνεται περισσότερη έμφαση στην προσπάθεια εξαγωγής των συναισθημάτων με την χρήση τεχνικών κατηγοριοποίησης (classification).

Η τρέχουσα έρευνα διαφέρει από την βιβλιογραφία η οποία παρουσιάστηκε παραπάνω στα εξής σημεία:

- Επικεντρώνεται στην εξαγωγή συναισθημάτων από κοινωνικά δίκτυα με στόχο όχι μόνο την εξαγωγή της πολικότητας (θετικό ή αρνητικό συναίσθημα) αλλά όπου είναι δυνατό και του τύπου του συναισθήματος (π.χ. θυμός, χαρά, θλίψη, κτλ.).
- Δίνει ιδιαίτερη σημασία για τα αποτελέσματα τα οποία εξάγονται από τεχνικές Μηχανικής Μάθησης και τα συγκρίνει με τα αποτελέσματα τα οποία εξάγονται από περισσότερο παραδοσιακές τεχνικές (π.χ. Λεξικό Δεδομένων).
- Κάνει χρήση συνόλων δεδομένων τα οποία έχουν εξεταστεί και από άλλες μελέτες, (π.χ. το καλά δομημένο σύνολο δεδομένων WASSA-2017).

Κεφάλαιο 3

Μηχανική Μάθηση

3.1 Εισαγωγή

Ο όρος μάθηση, καθώς και όρος ευφυΐα, καλύπτουν ένα μεγάλο εύρος ανθρώπινων και τεχνητών διαδικασιών, με συνέπεια να είναι πολύ δύσκολο να τις ορίσουμε με ακρίβεια. Ένας ορισμός που θα μπορούσε να παρατίθεται σε κάποιο λεξικό, θα ήταν: “είναι η διαδικασία απόκτησης γνώσης μέσω συστηματικής μελέτης κι εφαρμογής” ή “αλλαγή του τρόπου σκέψης και των ενεργειών ενός ατόμου μέσω της εμπειρίας”.

Η ζωολογία και η ψυχολογία είναι δύο επιστήμες όπου ασχολούνται με την μάθηση σε ζώα και ανθρώπους αντίστοιχα. Η μελέτη αυτή επικεντρώνεται στη μάθηση που αφορά μηχανές ή αλλιώς την Μηχανική Μάθηση (Machine Learning). Πολλές από τις θεωρίες που έχουν αναπτυχθεί στους τομείς της ζωολογίας και της ψυχολογίας χρησιμοποιούνται και στη μηχανική μάθηση με κάποιες παραλλαγές. Όσον αφορά τις μηχανές, θα μπορούσαμε να πούμε ότι μια μηχανή “μαθαίνει” όταν αλλάζει την δομή της με τρόπο τέτοιο, ώστε να λειτουργεί πιο αποδοτικά στο μέλλον. Για παράδειγμα, όταν η απόδοση ενός συστήματος αναγνώρισης φωνής βελτιώνεται μετά από την επεξεργασία πολλών διαφορετικών δειγμάτων φωνής, τότε μπορούμε να πούμε με σχετική ασφάλεια, ότι το συγκεκριμένο σύστημα “έμαθε”.

Η Μηχανική Μάθηση αποτελεί ένα από τα σημαντικότερα εργαλεία για την εξαγωγή συναισθημάτων από κείμενα καθώς και την εξόρυξη γνώσης γενικότερα. Η μηχανική μάθηση δεν αποτελεί ξεχωριστή επιστήμη ούτε και μπορεί να θεωρηθεί ότι αποτελεί επιστημονικό πεδίο ενός και μόνο επιστημονικού τομέα, όπως για παράδειγμα η Πληροφορική ή η Ρομποτική. Αποτελεί γεγονός ότι στην εξάπλωση και την εξέλιξη της μηχανικής μάθησης διαδραμάτισαν σπουδαίο ρόλο πολλοί επιστημονικοί τομείς.

Στις πλείστες περιπτώσεις, τα συστήματα που σχετίζονται με την μηχανική μάθηση σχετίζονται και με το αντικείμενο της Τεχνητής Νοημοσύνης (Artificial Intelligence – AI). Συνήθως το αντικείμενο των συστημάτων αυτών είναι η αναγνώριση αντικειμένων, η διάγνωση με βάση κάποια δεδομένα, η κατασκευή σχεδίων, η πρόβλεψη, κτλ. Οι αλλαγές

οι οποίες πραγματοποιούνται σε ένα σύστημα με σκοπό να “μάθει” μπορεί να αφορούν είτε την αλλαγή των τιμών κάποιων παραμέτρων, σημαντικών για το σύστημα, είτε την δυναμική επανα-σχεδίαση της δομής τους. Έξαρτάται κάθε φορά από την φύση και το αντικείμενο του εν λόγω συστήματος για το ποιο από τα δύο μοντέλα θα ακολουθεί.

Τα είδη της εκπαίδευσης που μπορούν να εφαρμοστούν σε ένα σύστημα μηχανικής μάθησης είναι:

- μάθηση με επίβλεψη ή εποπτευόμενη μάθηση (supervised learning)
- μάθηση χωρίς επίβλεψη ή μη εποπτευόμενη μάθηση (unsupervised learning)
- μάθηση με ενίσχυση ή ενισχυτική μάθηση (reinforcement learning)

Στη μάθηση με επίβλεψη (supervised learning) το σύστημα εκπαιδεύεται με κάποια ήδη υπάρχοντα παραδείγματα, τα οποία είναι ζεύγη της μορφής (είσοδος, επιθυμητή έξοδος). Ποιοτικά θα μπορούσαμε να τα θεωρήσουμε και ως ζεύγη της μορφής: (ερώτηση, σωστή απάντηση). Η μάθηση με επίβλεψη είναι μια επαναληπτική διαδικασία κατά την οποία εμφανίζονται στο σύστημα εισόδοι (ερωτήσεις), παρέχονται οι αντίστοιχες έξοδοι (απαντήσεις), και συγκρίνοντας τις με τις επιθυμητές εξόδους, αυτό προσαρμόζει τις παραμέτρους του και τις εσωτερικές του δομές κατάλληλα. Στην ουσία το σύστημα προσπαθεί να ανακαλύψει την συσχέτιση μεταξύ των δεδομένων εισόδου με τα δεδομένα εξόδου. Συνεπώς η εκπαίδευση του συστήματος έγκειται στην ικανότητά του να απαντά σωστά σε ανάλογες εισόδους. Ο όρος μάθηση με επίβλεψη προκύπτει από το ανθρωπομορφικό ανάλογο του “επιβλέπωντος”, ο οποίος εποπτεύει το σύστημα μηχανικής μάθησης, θέτοντας ερωτήσεις και παρέχοντας ταυτόχρονα και τις αντίστοιχες σωστές απαντήσεις. Η εξέταση του εν λόγω συστήματος για το εάν εκπαιδεύτηκε σωστά, μπορεί να γίνει μετά την ολοκλήρωση της διαδικασίας, παρέχοντάς του παρόμοια δεδομένα, μη γνωστά από πριν και μετρώντας το ποσοστό λάθους στις απαντήσεις του. Στην μάθηση χωρίς επίβλεψη τα δεδομένα εκπαίδευσης δεν περιλαμβάνουν την επιθυμητή έξοδο αλλά μόνο τα δεδομένα εισόδου. Σε αυτή την περίπτωση στόχος της εκπαίδευσης δεν είναι η εύρεση της απεικόνισης εισόδου – εξόδου, αλλά η εξαγωγή από το σύστημα μηχανικής μάθησης χαρακτηριστικών ή συσχετίσεων μεταξύ των δεδομένων.

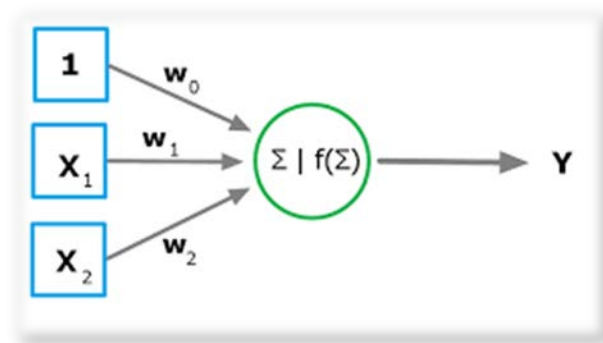
Η μάθηση με ενίσχυση είναι μια παραλλαγή της μάθησης με επίβλεψη, κατά την οποία παρέχεται λιγότερη πληροφορία στο σύστημα μηχανικής μάθησης από τον εξωτερικό επιβλέποντα. Η διαφορά έγκειται στο ότι στο σύστημα μάθησης δεν παρέχεται η επιθυμητή έξοδος για κάθε είσοδο, αλλά μόνο η τιμή μιας ποσότητας, που ονομάζεται σημείο ενίσχυσης (reinforcement signal) και δηλώνει εάν το σύστημα παρείχε σωστή ή

λάθος απόκριση χωρίς όμως να παρέχει λεπτομέρειες για το ποια είναι η σωστή απόκριση (μπορεί επίσης σε κάποιες χρονικές στιγμές κατά τη διάρκεια της εκπαίδευσης το σήμα ενίσχυσης να μην περιέχει πληροφορία σχετικά με το σωστό ή λάθος). Έτσι στην περίπτωση που το σύστημα μηχανικής μάθησης παρέχει λάθος έξοδο, ενημερώνεται ότι απάντησε λανθασμένα, αλλά δεν πληροφορείται σχετικά με το ποια είναι η σωστή έξοδος. Η χρήση τέτοιων συστημάτων μπορεί να γίνει σε περιβάλλοντα, όπου μία ακριβής απάντηση δεν έχει μεγάλη βαρύτητα κι ο χρήστης μπορεί να δεχθεί απαντήσεις προσεγγιστικές.

Τα Τεχνητά Νευρωνικά Δίκτυα αποτελούν ίσως τον πιο διαδεδομένο τομέα της μηχανικής μάθησης. Ακολουθεί ανάλυση της έννοιας των Τεχνητών Νευρωνικών Δικτύων πιο κάτω.

3.2 Τεχνητά Νευρωνικά Δίκτυα

Ένα Τεχνητό Νευρωνικό Δίκτυο στην ουσία εκτελεί σε αφαιρετικό επίπεδο, μια προσπάθεια προσομοίωσης, της λειτουργίας του ανθρώπινου εγκεφάλου. Ένα αρκετά διαδεδομένο μοντέλο κατηγοριοποίησης είναι τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ). Η πιο απλή μορφή ενός ΤΝΔ είναι ο αισθητήρας (Perceptron). Ο αισθητήρας είναι ένα μαθηματικό μοντέλο (εύκολα υλοποιήσιμο τόσο σε επίπεδο υλικού όσο και σε επίπεδο λογισμικού) το οποίο χρησιμοποιείται για τον διαχωρισμό γραμμικά διαχωρίσιμων παραδειγμάτων. Παρουσιάζεται η γραφική αναπαράσταση ενός αισθητήρα:



Διάγραμμα 1. Ο απλός αισθητήρας (Perceptron)

Ο απλός αισθητήρας είναι ένα σύστημα επεξεργασίας δεδομένων, το οποίο διαθέτει μία ή παραπάνω εισόδους και ακριβώς μία έξοδο. Οι εισοδοι ενός αισθητήρα λαμβάνουν αριθμητικές (πραγματικές) τιμές, οι οποίες συνήθως αντιστοιχούν στις τιμές των χαρακτηριστικών ενός στοιχείου. Η έξοδος του αισθητήρα είναι η κατηγορία στην οποία

ανήκει το στοιχείο αυτό. Επειδή η έξοδος του αισθητήρα είναι συνήθως 0 ή 1 (αν και θα μπορούσε να είναι διαφορετική), ο αισθητήρας χρησιμοποιείται συνηθέστερα για την κατηγοριοποίηση παραδειγμάτων που ανήκουν σε δύο μόνο δυνατές κατηγορίες. Μαθηματικά ο αισθητήρας ορίζεται ως εξής:

Έστω μία υπολογιστική μονάδα, η οποία ονομάζεται αισθητήρας και η οποία διαθέτει ένα σύνολο από n στον αριθμό εισόδους $\{x_1, x_2, \dots, x_n\}$. Σε κάθε είσοδο αντιστοιχεί ένα πραγματικό βάρος (*weight*). Τα βάρη συμβολίζονται ως εξής: $\{w_1, w_2, \dots, w_n\}$. Επιπλέον, ο αισθητήρας διαθέτει μία σταθερή είσοδο, η οποία είναι πάντα ίση με 1. Το βάρος της σταθερής αυτής εισόδου συμβολίζεται με Y και ονομάζεται *κατώφλι* (*threshold*).

Με την απόδοση τιμών στις εισόδους ενός αισθητήρα, αυτός υπολογίζει το εξής μέγεθος: $\sum_{i=1}^n w_i \cdot x_i$

Αν το παραπάνω μέγεθος είναι μεγαλύτερο ή ίσο της τιμής του κατωφλίου, τότε η έξοδος του αισθητήρα θα είναι 1 ενώ σε διαφορετική περίπτωση θα είναι 0. Πιο συγκεκριμένα, αν συμβολίσουμε με y την έξοδο του αισθητήρα, αυτή υπολογίζεται ως εξής:

$$y = f\left(\sum_{i=1}^n w_i \cdot x_i\right) = \begin{cases} 1, & \sum_{i=1}^n w_i \cdot x_i \geq \theta \\ 0, & \sum_{i=1}^n w_i \cdot x_i < \theta \end{cases}$$

Η παραπάνω συνάρτηση μπορεί επίσης να αναπαρασταθεί και ως εξής:

$$y = f(u) = \begin{cases} 1, & u \geq 0 \\ 0, & u < 0 \end{cases}$$

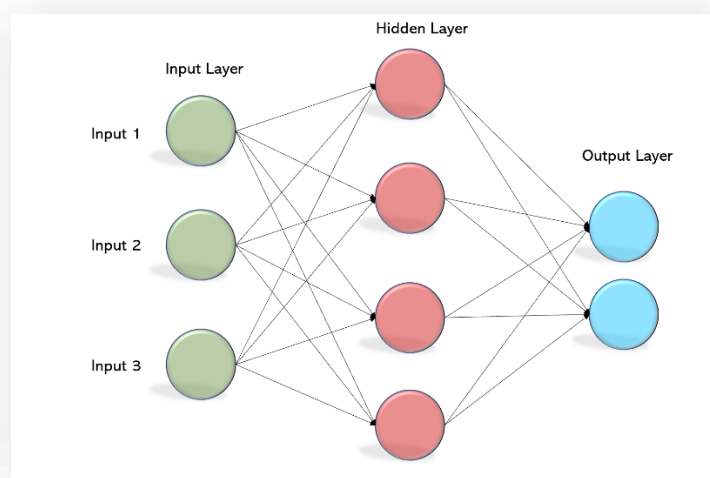
όπου $u = \sum_{i=1}^n w_i \cdot x_i - \theta$.

Η συνάρτηση f είναι χαρακτηριστική για κάθε αισθητήρα. Ένας αισθητήρας ο οποίος χρησιμοποιεί την παραπάνω συνάρτηση εξόδου έχει την δυνατότητα να διαχωρίσει, με 100% επιτυχία, ένα σύνολο από στοιχεία n χαρακτηριστικών, τα οποία ανήκουν σε ακριβώς δύο κατηγορίες. Πιο συγκεκριμένα, αν τα στοιχεία τα οποία αναπαρίστανται στον χώρο και είναι σε θέση να διαχωριστούν στις δύο κατηγορίες μέσω ενός υπερ-επιπέδου (ευθεία αν $n = 2$) τότε η μία από τις δύο κατηγορίες αναπαρίσταται με έξοδο 0 ενώ η άλλη με έξοδο 1. Τα στοιχεία τα οποία έχουν την παραπάνω ιδιότητα ονομάζονται γραμμικώς διαχωρίσιμα. Εκτός από την παραπάνω (περισσότερο διαδεδομένη μορφή της) συνάρτηση εξόδου, μπορεί να χρησιμοποιηθεί και οποιαδήποτε άλλη, γραμμική συνάρτηση. Μία άλλη πολύ συχνά χρησιμοποιούμενη συνάρτηση είναι η σιγμοειδής συνάρτηση (*sigmoid function*), η οποία ορίζεται ως εξής:

$$S(u) = \frac{1}{1 + e^{-cu}}$$

όπου $u = \sum_{i=1}^n w_i \cdot x_i - \theta$ και c μία πραγματική σταθερά.

Η σύνδεση των εξόδων κάποιων νευρώνων με τις εισόδους κάποιων άλλων, οδηγεί στην δημιουργία μίας δομής που ονομάζεται Τεχνητό Νευρωνικό Δίκτυο (ΤΝΔ). Υπάρχουν πολλές τοπολογίες Τεχνητών Νευρωνικών Δικτύων, με περισσότερο διαδεδομένη αυτή του πολύ-επίπεδου δικτύου εμπρόσθιας τροφοδότησης (feed-forward neural network ή Multilayer Perceptron). Μία γραφική αναπαράσταση ενός τέτοιου δικτύου παρουσιάζεται πιο κάτω:



Διάγραμμα 2. Γραφική αναπαράσταση ενός Τεχνητού Νευρωνικού Δικτύου εμπρόσθιας τροφοδότησης

Ένα ΤΝΔ εμπρόσθιας τροφοδότησης αποτελείται από ένα ή περισσότερα συνεχόμενα επίπεδα από αισθητήρες. Το πρώτο επίπεδο ονομάζεται επίπεδο εισόδου (input layer) ενώ το τελευταίο επίπεδο ονομάζεται επίπεδο εξόδου (output layer). Το ενδιάμεσο επίπεδο ονομάζεται κρυφό επίπεδο (hidden layer). Στην πραγματικότητα ένα ΤΝΔ μπορεί να αποτελείται από ένα ή περισσότερα κρυφά επίπεδα. Συνήθως ακολουθείται η σύμβαση ότι οι έξοδοι των νευρώνων του επιπέδου k τροφοδοτούν τις εισόδους των νευρώνων του επιπέδου $k + 1$.

Ένα ΤΝΔ μπορεί να χρησιμοποιηθεί ως ένας πολύ αποδοτικός κατηγοριοποιητής. Η συνήθης διαδικασία είναι η μεταβολή των βαρών των νευρώνων του ΤΝΔ, έτσι ώστε σε περίπτωση εισαγωγής των χαρακτηριστικών ενός στοιχείου στις εισόδους του, η έξοδος του ΤΝΔ να είναι η κατηγορία στην οποία το στοιχείο αυτό ανήκει. Η διαδικασία

μεταβολής των βαρών ενός ΤΝΔ ονομάζεται εκπαίδευση (training) και μάλιστα υπό επίβλεψη (supervised training), αφού η διαδικασία συνεχίζεται μέχρι το ΤΝΔ να δίνει ικανοποιητικές απαντήσεις για κάθε στοιχείο του συνόλου εκπαίδευσης (training set). Όλοι οι αλγόριθμοι εκπαίδευσης ενός ΤΝΔ στηρίζονται στην εξής υπόθεση: αν το ΤΝΔ είναι σε θέση να κατηγοριοποιεί ικανοποιητικά τα στοιχεία του συνόλου εκπαίδευσης, τότε κάθε νέο στοιχείο το οποίο θα εμφανιστεί στην είσοδο του είναι πολύ πιθανό να παρουσιάζει ομοιότητα με κάποιο στοιχείο από το σύνολο εκπαίδευσης και συνεπώς υπάρχει μεγάλη πιθανότητα η κατηγοριοποίηση του νέου αυτού στοιχείου να είναι σωστή.

Ο περισσότερο διαδεδομένος αλγόριθμος εκπαίδευσης ενός ΤΝΔ είναι ο αλγόριθμος οπισθοδιάδοσης του λάθους (Back Propagation – BP). Ο αλγόριθμος αυτός προσπαθεί να ελαχιστοποιήσει το τετραγωνικό σφάλμα της τρέχουσας εξόδου του ΤΝΔ σε σχέση με την επιθυμητή έξοδο. Πιο συγκεκριμένα, αν συμβολίσουμε ως P το σύνολο με τα στοιχεία εκπαίδευσης, με T_i την επιθυμητή έξοδο του δικτύου για το στοιχείο $p_i \in P$ και με O_i την πραγματική έξοδο του δικτύου για το στοιχείο αυτό, τότε ο αλγόριθμος BP προσπαθεί να ελαχιστοποιήσει το παρακάτω μέγεθος:

$$P \sum (T_i - O_i)^2$$

$$Total_{Error} = \sum_{i=1}^n$$

Αν το επίπεδο εξόδου περιλαμβάνει περισσότερους από ένα νευρώνα εξόδου, τότε η διαφορά της επιθυμητής με την πραγματική έξοδο, είναι στην ουσία μία αφαίρεση διανυσμάτων, η οποία ορίζεται ως εξής (υποθέτουμε ότι τα διανύσματα είναι διάστασης n):

$$T_i - O_i = \sum_{j=1}^n (T_{ij} - O_{ij})^2$$

όπου T_{ij} είναι το στοιχείο j του διανύσματος T_i (αντίστοιχα και για το δεύτερο διάνυσμα).

Η διαδικασία εκπαίδευσης του ΤΝΔ πραγματοποιείται σε κύκλους εκπαίδευσης, όπου σε κάθε κύκλο γίνονται ακριβώς οι ίδιες ενέργειες. Η εκπαίδευση σταματά όταν ικανοποιηθεί κάποιο κριτήριο τερματισμού, το οποίο συνήθως είναι είτε η ολοκλήρωση ενός αριθμού κύκλων είτε όταν το συνολικό τετραγωνικό σφάλμα των εξόδων μειωθεί κάτω από ένα κατώφλι ϵ ($Total_{Error} < \epsilon$). Οι ενέργειες οι οποίες πραγματοποιούνται σε έναν κύκλο εκπαίδευσης πραγματοποιούνται διαδοχικά για κάθε στοιχείο εκπαίδευσης

του συνόλου P. Ειδικότερα, οι υπολογισμοί για ένα στοιχείο $p_i \in P$ σε κάποιον κύκλο εκπαίδευσης είναι οι παρακάτω:

Εμπρόσθιο πέρασμα: στο στάδιο αυτό υπολογίζεται η έξοδος του κάθε νευρώνα σε κάθε επίπεδο του ΤΝΔ, με απώτερο στόχο τον υπολογισμό των εξόδων του τελευταίου επιπέδου. Για κάθε νευρώνα j , με n εισόδους, η έξοδος υπολογίζεται ως εξής:

$$u_j = \sum_{i=0}^n w_i \cdot x_i - \theta_j \quad \text{και} \quad y_j = f(u_j)$$

Στην περίπτωση που πρόκειται για νευρώνα κρυφού επιπέδου, η κάθε είσοδος x_i αποτελεί έξοδο νευρώνα προηγούμενου επιπέδου ενώ αν πρόκειται για νευρώνα του πρώτου επιπέδου (επιπέδου εισόδου) οι εισοδοί είναι οι τιμές των χαρακτηριστικών του τρέχοντος σημείου.

Υπολογισμός σφαλμάτων: στο στάδιο αυτό υπολογίζεται το σφάλμα για κάθε νευρώνα εξόδου j . Το σφάλμα υπολογίζεται ως εξής:

$$e_j = T_j - O_j$$

Πέρασμα προς τα πίσω: στο στάδιο αυτό το σφάλμα των νευρώνων εξόδου διαδίδεται προς τα πίσω, στους νευρώνες των προηγούμενων επιπέδων. Πιο συγκεκριμένα, για κάθε νευρώνα του ΤΝΔ υπολογίζεται το παρακάτω μέγεθος, το οποίο στην ουσία αναπαριστά τον «βαθμό ευθύνης» που έχει ο νευρώνας:

$$\delta_j = f'(u_j) \cdot e_j, \text{ αν ο } j \text{ είναι νευρώνας εξόδου}$$

$$\delta_j = f'(u_j) \cdot \sum_k w_{kj} \cdot \delta_k, \text{ αν ο } j \text{ είναι κρυφός νευρώνας και } k \text{ ο κάθε νευρώνας στον οποίον ο } j \text{ δίνει έξοδο.}$$

Υπολογισμός των μεταβολών των βαρών: για κάθε βάρος w_{ji} του νευρώνα j , το οποίο λαμβάνει ως είσοδο την έξοδο του νευρώνα i , η μεταβολή του υπολογίζεται ως εξής:

$$\Delta w_{ji} = \eta \cdot \delta_j \cdot O_i$$

Ανανέωση των βαρών: κάθε βάρος w_{ji} , το οποίο επισυνάπτεται στην i -οστή είσοδος του νευρώνα j ανανεώνει την τιμή του ως εξής:

$$w'_{ji} = w_{ji} + \Delta w_{ji}$$

Με την πάροδο ενός αριθμού κύκλων εκπαίδευσης, το συνολικό τετραγωνικό σφάλμα μειώνεται και έτσι με τον τρόπο αυτό εκπαιδεύεται το ΤΝΔ.

3.3 Εξελικτικοί Αλγόριθμοι

Ένας άλλος τομέας ο οποίος μπορεί να χρησιμοποιηθεί για την εξαγωγή συναισθημάτων από κείμενα, είναι οι Εξελικτικοί Αλγόριθμοι. Όσο αφορά τον τομέα της Τεχνητής

Νοημοσύνης, ένας εξελικτικός αλγόριθμος (Evolutionary Algorithm) αποτελεί τη προσομοίωση μίας βιολογικής διαδικασίας με σκοπό την επίλυση ενός προβλήματος. Ένας εξελικτικός αλγόριθμος χρησιμοποιεί μηχανισμούς που συναντώνται στην θεωρία εξέλιξης των ειδών (όπως η αναπαραγωγή, η μετάλλαξη και η φυσική επιλογή) αλλά και μηχανισμούς που προσομοιώνουν την συμπεριφορά των μελών ενός ζωικού πληθυσμού (π.χ. την συμπεριφορά των μυρμηγκιών). Συνήθως εφαρμόζονται σε προβλήματα βελτιστοποίησης (optimization problems) όπου κάθε πιθανή λύση παίζει τον ρόλο ενός ατόμου κάποιου πληθυσμού. Κάθε ένα από τα άτομα αξιολογείται με μία συνάρτηση καταλληλότητας ή συνάρτηση αξιολόγησης (fitness function).

Στην πράξη, οι Εξελικτικοί Αλγόριθμοι δεν εντοπίζουν την βέλτιστη λύση ενός προβλήματος αλλά μία καλή προσέγγιση αυτής. Επίσης, η φύση τους είναι γενική, κάτι το οποίο σημαίνει ότι μπορούν να εφαρμοστούν πάνω σε οποιοδήποτε είδος προβλήματος.

Δύο μεγάλες κατηγορίες Εξελικτικών Αλγορίθμων είναι οι Γενετικοί Αλγόριθμοι (Genetic Algorithms) και οι Αλγόριθμοι Νοημοσύνης Σμήνους (Swarm Intelligence). Οι Γενετικοί Αλγόριθμοι βασίζονται στην θεωρία της εξέλιξης των ειδών, η οποία προτάθηκε από τον Κάρολο Δαρβίνο. Ένας Γενετικός Αλγόριθμος εκτελεί μία «έξυπνη» αναζήτηση σε έναν μεγάλο χώρο πιθανών λύσεων ενός προβλήματος. Αυτό το οποίο πρακτικά κάνει είναι να ξεκινά με ένα μικρό αρχικό αριθμό πιθανών λύσεων (πληθυσμός) και στην συνέχεια να αξιολογεί και να βελτιστοποιεί τις πιθανές αυτές λύσεις.

Τα συστήματα νοημοσύνης σμήνους αποτελούνται από έναν πληθυσμό απλών και αυτόνομων πρακτόρων (agents) οι οποίοι επικοινωνούν μεταξύ τους, καθώς και με το περιβάλλον στο οποίο βρίσκονται σε τοπικό επίπεδο. Τα συστήματα αυτά έχουν εμπνευστεί από διάφορα βιολογικά συστήματα (M. Milonas, 1994). Οι πράκτορες εφαρμόζουν πολύ απλούς κανόνες συμπεριφοράς και δεν υπάρχει καμία κεντρική δομή ελέγχου που να υπαγορεύει τον τρόπο συμπεριφοράς τους. Οι τοπικές (και πολλές φορές τυχαίες) αλληλεπιδράσεις μεταξύ των πρακτόρων οδηγούν στην εμφάνιση μιας ευφυούς, καθολικής συμπεριφοράς, η οποία όμως είναι άγνωστη σε κάθε αυτόνομο πράκτορα.

Όσον αφορά την αναγνώριση συναισθημάτων από κείμενα, ένας Εξελικτικός Αλγόριθμος χρησιμοποιείται, όταν ένα πρόβλημα μπορεί να διατυπωθεί ως πρόβλημα βελτιστοποίησης. Πιο συγκεκριμένα, κάθε διαφορετικό συναίσθημα μπορεί να αναπαρασταθεί με μία συνάρτηση της μορφής $f(w_1, w_2, \dots, w_n)$, όπου κάθε w_i αποτελεί μία λέξη του κειμένου. Η συνάρτηση επιστρέφει μία πραγματική τιμή στο διάστημα $[0, 1]$, η οποία αντιστοιχεί στην βεβαιότητα του κατά πόσο το εν λόγω συναίσθημα

εκφράζεται στο κείμενο αυτό. Ανάλογα με τον εξελικτικό αλγόριθμο ο οποίος θα επιλεγεί, οι τελεστές (δηλαδή οι ενέργειες) επί του κειμένου είναι διαφορετικές. Συνήθως, οι Εξελικτικοί Αλγόριθμοι απαιτούν περισσότερο χρόνο εκτέλεσης σε σχέση με τις υπόλοιπες μεθόδους εξαγωγής συναισθημάτων.

3.4 Στατιστικές Μέθοδοι

Οι στατιστικές μέθοδοι που χρησιμοποιούνται για την εξόρυξη πληροφορίας δεν διαφέρουν σημαντικά από αυτές που χρησιμοποιούνται για την εξόρυξη δεδομένων γενικότερα. Η μόνη διαφορά είναι στα στοιχεία τα οποία επεξεργάζονται. Οι περισσότερες μέθοδοι συνήθως, χωρίζουν το κείμενο σε σύμβολα (tokens) ή φράσεις και προσπαθούν να αναθέσουν μια ταμπέλα (κατηγοριοποίηση) η οποία αντιπροσωπεύει την κατηγορία στην οποία ανήκει η φράση ή το σύμβολο. Η πιο γνωστή στατιστική μέθοδος κατηγοριοποίησης είναι η κατηγοριοποίηση Bayesian.

Η κατηγοριοποίηση Bayesian βασίζεται στη στατιστική θεωρία κατηγοριοποίησης του Bayes. Ο στόχος είναι να κατηγοριοποιηθεί ένα δείγμα X σε μια από τις δεδομένες κατηγορίες (C_1, C_2, \dots, C_n) χρησιμοποιώντας ένα μοντέλο πιθανότητας που ορίζεται σύμφωνα με την θεωρία Bayes. Κάθε κατηγορία χαρακτηρίζεται από μια εκ των προτέρων πιθανότητα (a priori probability) παρατήρησης της κλάσης (κατηγορίας) C_i . Επίσης, υποθέτουμε ότι το δείγμα δεδομένων X ανήκει σε μια κλάση C_i , με την υπό συνθήκη συνάρτηση πυκνότητας πιθανότητας: $p(X/C_i) \in [0,1]$. Κατόπιν, χρησιμοποιώντας τους ανωτέρω ορισμούς και βασιζόμενοι στη θεωρία Bayes, καθορίζουμε την εκ των υστέρων (posterior) δεσμευμένη πιθανότητα $p(C_i|X)$ ως εξής:

$$p(C_i|X) = \frac{p(C_i|X)p(C_i)}{p(X)}$$

Ο απλούστερος Bayesian κατηγοριοποιητής είναι ο γνωστός Naive Bayesian. Αυτός υποθέτει ότι σε μια δεδομένη κατηγορία, η επίδραση ενός γνωρίσματος (attribute) είναι ανεξάρτητη από τις τιμές των άλλων γνωρισμάτων. Η υπόθεση γίνεται για απλοποίηση των υπολογισμών που εμπλέκονται και καλείται υπό συνθήκη ανεξαρτησία (conditional independence). Ένας άλλος Bayesian κατηγοριοποιητής είναι τα Bayesian Belief Networks. Πρόκειται για γραφικά μοντέλα τα οποία αντίθετα με τους Naive Bayesian κατηγοριοποιητές, επιτρέπουν την παρουσίαση των εξαρτήσεων μεταξύ των υποσυνόλων των γνωρισμάτων. Η πλήρης θεωρητική και μαθηματική παρουσίαση των παραπάνω κατηγοριοποιητών ξεφεύγει από τα όρια της συγκεκριμένης μελέτης και για τον λόγο αυτό αναφέρονται απλά συνοπτικά.

3.5 Σημασιολογικά Δίκτυα

Ένα σημασιολογικό δίκτυο (semantic network), είναι ένα γράφος, ο οποίος αποτελείται από κόμβους και κατευθυνόμενες ακμές. Οι κόμβοι απεικονίζουν έννοιες, αντικείμενα και συμβάντα, ενώ οι ακμές δυαδικές σχέσεις. Τέτοια σημασιολογικά δίκτυα μπορούν να χρησιμοποιηθούν και στην εξαγωγή πληροφορίας από κείμενα, προσομοιώνοντας την λειτουργία της συνειρμικής μνήμης του ανθρώπινου εγκεφάλου (associative memory). Η βασική ιδέα για την λειτουργία της συνειρμικής μνήμης του ανθρώπινου εγκεφάλου είναι αυτή της ενεργοποίησης κάποιων γεγονότων από κάποια άλλα. Για παράδειγμα, εάν ένα πρόσωπο βιώσει ένα γεγονός Α, μπορεί να θυμηθεί ένα παρόμοιο γεγονός Β, το οποίο θα τον οδηγήσει στην ανάκτηση από την μνήμη του ενός παρόμοιου γεγονότος Γ. Η παραπάνω διαδικασία, μπορεί να προσομοιωθεί σε ένα σημασιολογικό δίκτυο, όπου τα αποθηκευμένα γεγονότα αναπαρίστανται με τους κόμβους, ενώ η “ενεργοποίηση” αυτών πραγματοποιείται μέσω των ακμών που τους ενώνουν.

3.6 Σύγκριση Μεθόδων

Κάθε μία από τις μεθόδους που παρουσιάστηκαν πιο πάνω, προσεγγίζουν το θέμα της εξόρυξης συναισθημάτων από μη δομημένα κείμενα με διαφορετική λογική και σκοπιμότητα. Για παράδειγμα οι περισσότερες μέθοδοι μηχανικής μάθησης υποθέτουν ότι πριν την εισαγωγή κειμένων προς εξέταση, οι εσωτερικές δομές του συστήματος στο οποίο εφαρμόζονται θα έχουν εκπαιδευτεί σύμφωνα με κάποιο σύνολο δεδομένων εκπαίδευσης (training set). Με βάση τις παραπάνω παραδοχές, δεν είναι δυνατό να πούμε γενικότερα ότι μια μέθοδος είναι καλύτερη από μια άλλη, αλλά θα πρέπει να εξετάσουμε το ζήτημα ειδικά για κάθε περίπτωση εφαρμογής.

Μπορούμε να πούμε ότι οι τομείς εφαρμογής χωρίζονται σε δύο μεγάλες κατηγορίες: κείμενα της ίδιας φύσης, αποθηκευμένα κυρίως σε μία βάση ή αποθήκη δεδομένων και κείμενα εντελώς διαφορετικής φύσης, όπως για παράδειγμα διάφορες ιστοσελίδες στο Διαδίκτυο. Τυπικά, στα κείμενα της πρώτης κατηγορίας, ο χρήστης του συστήματος εξόρυξης ενδιαφέρεται κυρίως για τις εμφανίσεις των οντοτήτων και συσχετίσεων μεταξύ αυτών συγκεκριμένου είδους (π.χ. ιατρικοί όροι), ενώ στα κείμενα της δεύτερης κατηγορίας ο χρήστης ενδιαφέρεται κυρίως για την κατηγοριοποίησή τους με βάση το θεματικό τους αντικείμενο. Τέλος ένα σημαντικό σημείο, το οποίο θα πρέπει να προσέξει κάποιος για την εφαρμογή μιας μεθόδου, είναι το υπολογιστικό κόστος. Για παράδειγμα,

η εκτέλεση των πράξεων σε ένα Τεχνητό Νευρωνικό Δίκτυο δεν έχει μεγάλη πολυπλοκότητα, αλλά ο έλεγχος της ομοιότητας μιας συμβολοσειράς με μια άλλη (ομοιότητας κι όχι ισότητας) έχει ενδεχομένως μεγαλύτερο υπολογιστικό κόστος. Μία συγκριτική παρουσίαση των διάφορων τεχνικών προσεγγίσεων οι οποίες παρουσιάστηκαν στο τρέχον κεφάλαιο, παρουσιάζεται παρακάτω, στον Πίνακα 1

Προσέγγιση	Σύντομη περιγραφή
Λεκτική Ανάλυση (Tokenization)	Η λεκτική ανάλυση (tokenization) είναι η διαδικασία του τεμαχισμού ενός κειμένου σε μικρότερα υπό-τμήματα, τα οποία ονομάζονται "σύμβολα" (tokens).
Κανονικοποίηση Μορφής (Normalization)	Αφορά την αντιστοίχιση του κάθε τμήματος/συμβόλου στην ομάδα συμβόλων στην οποία ανήκει.
Δεικτοδότηση (Indexing)	Η διαδικασία της δεικτοδότησης (indexing) χρησιμοποιείται από ένα σύστημα εξαγωγής πληροφοριών από φυσικό κείμενο για την γρηγορότερη πρόσβαση στα σύμβολα (tokens) τα οποία βρέθηκαν κατά τις παραπάνω διαδικασίες.
Μηχανική Μάθηση	Αφορά την διαδικασία "εκπαίδευσης" ενός συστήματος με την χρήση έτοιμων παραδειγμάτων (της μορφής εισόδου-εξόδου) με την ελπίδα ότι το σύστημα θα απαντήσει με την σωστή έξοδο στην περίπτωση παρόμοιων εισόδων.
Τεχνητά Νευρωνικά Δίκτυα	Ένα Τεχνητό Νευρωνικό Δίκτυο είναι στην ουσία μια προσπάθεια προσομοίωσης, σε αφαιρετικό επίπεδο, της λειτουργίας του ανθρώπινου εγκεφάλου.
Εξελικτικοί Αλγόριθμοι	Στον τομέα της Τεχνητής Νοημοσύνης, ένας Εξελικτικός Αλγόριθμός (Evolutionary Algorithm) είναι μία προσομοίωση μιας βιολογικής διαδικασίας με σκοπό την

	<p>επίλυση ενός προβλήματος. Ένας Εξελικτικός Αλγόριθμος χρησιμοποιεί μηχανισμούς που συναντώνται στην θεωρία εξέλιξης των ειδών (όπως η αναπαραγωγή, η μετάλλαξη και η φυσική επιλογή) αλλά και μηχανισμούς που προσομοιώνουν την συμπεριφορά μεταξύ των μελών ενός ζωικού πληθυσμού (π.χ. την συμπεριφορά των μυρμηγκιών).</p>
<p>Σημασιολογικά Δίκτυα</p>	<p>Ένα σημασιολογικό δίκτυο (semantic network), είναι ένας γράφος, ο οποίος αποτελείται από κόμβους και κατευθυνόμενες ακμές. Οι κόμβοι απεικονίζουν έννοιες, αντικείμενα και συμβάντα, ενώ οι ακμές δυαδικές σχέσεις.</p>

Πίνακας 1. Συγκριτική παρουσίαση των διάφορων τεχνικών προσεγγίσεων

Κεφάλαιο 4

Μέθοδοι εξαγωγής συναισθημάτων μέσω ανάλυσης φυσικής γλώσσας

4.1 Εισαγωγή

Η εν λόγω ενότητα παρέχει μια σύντομη περιγραφή οκτώ μεθόδων ανάλυσης συναισθημάτων που διερευνήθηκαν σε αυτή την μελέτη. Παρουσιάζονται οι πιο δημοφιλείς στη βιβλιογραφία (δηλαδή οι πιο γνωστές και ευρέως χρησιμοποιούμενες) οι οποίες καλύπτουν ποικίλες τεχνικές όπως η χρήση της Επεξεργασίας Φυσικής Γλώσσας (NLP), η χρήση του Amazon Mechanical Turk (AMT) για τη δημιουργία καθορισμένων συνόλων δεδομένων, η χρήση ψυχομετρικών κλιμάκων για τον προσδιορισμό συναισθημάτων που βασίζονται στη διάθεση, η χρήση τεχνικών μηχανικής μάθησης με επίβλεψη και χωρίς επίβλεψη και άλλες. Η επικύρωση αυτών των μεθόδων ποικίλλει πολύ, από τη χρήση κατασκευασμένων παραδειγμάτων έως μια μεγάλη συλλογή δεδομένων με ετικέτα.

4.2 Ανάλυση των emoticons

Ο απλούστερος τρόπος ανίχνευσης της πολικότητας (δηλαδή θετικού ή αρνητικού συναισθήματος) ενός μηνύματος βασίζεται στα emoticons τα οποία περιέχει. Τα emoticons έχουν γίνει δημοφιλή τα τελευταία χρόνια, στο βαθμό που ορισμένα (π.χ. <3) περιλαμβάνονται πλέον στο Αγγλικό Λεξικό της Οξφόρδης (Oxford English Dictionary 2011). Τα emoticons βασίζονται κυρίως στις εκφράσεις προσώπου και αντιπροσωπεύουν συναισθήματα χαρούμενου ή λυπημένου ανθρώπου, αν και υπάρχει ένα ευρύ φάσμα παραλλαγών, όπως για παράδειγμα το emoticon<3 το οποίο αντιπροσωπεύει την καρδιά και εκφράζει αγάπη ή στοργή. Για την εξαγωγή της πολικότητας από τα emoticons, χρησιμοποιείται ένα σύνολο κοινών emoticons όπως αυτά που εμφανίζονται στον Πίνακα 4.1. Ο πίνακας περιλαμβάνει επίσης τις δημοφιλείς

παραλλαγές που εκφράζουν τις πρωταρχικές πολικότητες των θετικών, αρνητικών και ουδέτερων συναισθημάτων. Τα μηνύματα που περιέχουν περισσότερα από ένα emoticon συσχετίζονται με την πολικότητα του πρώτου emoticon που εμφανίζεται στο κείμενο.

Συναίσθημα	Emoticons
Θετικό	:) :] :} :o) :o] :o} :-] :-) :-} =) =] =} =^] =^) =^} :B :-D :-B :^D :^B =B =^B =^D :') :'] :'} ='') ='] ='}' <3 ^^ ^_^ ^_^ ^^ :* ==* :-* ;) ;] ;} :-p :-P :-b :^p :^P :^b =p \o\ /o/ :P :p :b =b =^p =^P =^b \o/
Αρνητικό	:([: { :o(:o[:o{ :-[:-(:-{ =(=[={ =^(=^{ >=(>=[>={ >=(>:-{ >:-[>:-(>=^[>:-(

	:-[:- (=(=[={ =^ [>:-=(>=[>=^(:'(:'[:'{ ='{ ='(='[=\ :\ =/ :/ =\$ o.OO_oOo :\$-{: >:-{: >=^{ :o{
Ουδέτερο	: = :- >.<><>_< :o :O =0 -.- -_- :x -.-' -_- -_-,' :x =X :# =# :-x :-@ :-# :^x :^#

Πίνακας 2: παραδείγματα emoticons και των αντίστοιχων συναισθημάτων που αυτά εκφράζουν

4.3 LIWC

Το LIWC (Linguistic Enquiry and Word Count) (Y. R. Tausczik and J. W. Pennebaker 2010: 24-54) είναι εργαλείο ανάλυσης κειμένου το οποίο αξιολογεί δομικά, γνωστικά και συναισθηματικά στοιχεία ενός δεδομένου κειμένου με βάση τη χρήση ενός λεξικού-ευρετηρίου που περιέχει λέξεις, καθώς και τις αντίστοιχες κατηγορίες στις οποίες ανήκουν. Ειδικά για τον εντοπισμό θετικών και αρνητικών εκφράσεων σε ένα δεδομένο κείμενο, το LIWC παρέχει περαιτέρω σύνολα κατηγοριών συναισθημάτων και όχι μόνο θετικών και αρνητικών. Για παράδειγμα, η λέξη “συμφωνώ” ανήκει στις ακόλουθες κατηγορίες (σύνολα): σύμφωνη γνώμη, θετικό συναίσθημα και γνωστική διαδικασία.

Το λογισμικό LIWC είναι εμπορικό και παρέχει διάφορες επιλογές, όπως επιτρέποντας στους χρήστες να συμπεριλαμβάνουν δικά τους λεξικά αντί για τα ήδη υπάρχοντα. Το LIWC είναι διαθέσιμο στη διεύθυνση <http://www.liwc.net/>.

4.4 SentiStrength

Οι μέθοδοι που βασίζονται στη μηχανική μάθηση είναι κατάλληλες για εφαρμογές οι οποίες χρειάζονται μοντέλα αναγνώρισης συναισθημάτων ή προσαρμοσμένων συναισθημάτων. Στη βιβλιογραφία έχουν προταθεί διάφορες βασικές αρχιτεκτονικές για τον προσδιορισμό του συναισθήματος με βάση τις τεχνικές μηχανικής μάθησης.

Το SentiStrength εκτιμά το θετικό ή το αρνητικό συναίσθημα σε σύντομα κείμενα, ακόμη και στην περίπτωση που αυτά περιέχουν άτυπη γλώσσα. Για σύντομα κοινωνικά κείμενα γραμμένα στην Αγγλική γλώσσα, έχει ακρίβεια σε ανθρώπινο επίπεδο και για κάθε ένα κείμενο προσδίδει μία τιμή στα παρακάτω δύο διαστήματα τιμών:

- -1 (όχι αρνητικό) έως -5 (εξαιρετικά αρνητικό)
- 1 (όχι θετικό) έως 5 (εξαιρετικά θετικό)

4.5 SentiWordNet

Το SentiWordNet (Esuli and Sebastiani, 2006: 417-422) είναι ένα εργαλείο που χρησιμοποιείται ευρέως στην εξόρυξη συναισθημάτων από κείμενο, και βασίζεται σε ένα λεξικό της Αγγλικής γλώσσας το οποίο ονομάζεται WordNet (G. A. Miller, 1995: 30-41). Αυτό το λεξικό ομαδοποιεί λέξεις, ουσιαστικά, ρήματα και άλλες γραμματικές ομάδες σε σύνολα συνωνύμων τα οποία ονομάζονται synsets. Το SentiWordNet προσδίδει μία από τρεις βαθμολογίες σε κάθε synset για την ένδειξη του συναισθήματος στις λέξεις που αυτό περιέχει: θετικό, αρνητικό και αντικειμενικό (ουδέτερο). Οι βαθμολογίες, οι οποίες παίρνουν τιμές από το σύνολο τιμών $[0, 1]$ προσδίδονται με την χρήση ενός ημι-εποπτευόμενου μοντέλου μηχανικής μάθησης. Για παράδειγμα, ας υποθέσουμε ότι ένα synset "s", το οποίο έχει εξαχθεί από ένα tweet περιέχει τα ακόλουθα δεδομένα: $s = [\text{κακό, άσχημα, τρομερό}]$. Το SentiWordNet θα δώσει σκορ 0.0 για θετικό, 0.850 για αρνητικό και 0.150 για αντικειμενικό-ουδέτερο συναίσθημα, αντίστοιχα.

4.6 SenticNet

Το SenticNet (E. Cambria, R. Speer, C. Havasi, A. Hussain, 2010) είναι μία μέθοδος εξόρυξης συναισθημάτων η οποία βασίζεται στη Τεχνητή Νοημοσύνη και τα σημασιολογικά δίκτυα. Ο σκοπός του SenticNet είναι η εύρεση της πολικότητας η οποία προκύπτει από τις φράσεις ενός κειμένου βασισμένο στην σημασιολογική ανάλυση αυτού και όχι στην συντακτική ανάλυση. Η μέθοδος χρησιμοποιεί τεχνικές Επεξεργασίας Φυσικής Γλώσσας

για την εύρεση της πολικότητας για περίπου 14000 έννοιες. Για παράδειγμα, για την ανάλυση της φράσης “Βαριέμαι, είναι Δευτέρα πρωί”, το SenticNet θα προσπαθήσει αρχικά να αναλύσει έννοιες, όπως το “Βαριέμαι” και “Δευτέρα πρωί”. Στην συνέχεια αναθέτει την τιμή πολικότητας -0.383 για την φράση “βαριέμαι” και την τιμή 0.228 για την φράση “Δευτέρα πρωί”. Το τελικό αποτέλεσμα για το παράδειγμα είναι -0.077, το οποίο προκύπτει από την μέση τιμή των δύο παραπάνω τιμών.

4.7 SASA

Το SASA αποτελεί ένα εργαλείο Μηχανικής Μάθησης, η ονομασία του οποίου προκύπτει από την φράση Sail Sentiment Analyzer (SASA) (H. Wang, D. Can, A. Kazemzadeh, F. Bar, S. Narayanan, 2012: 115-120). Το SASA χρησιμοποιεί Τεχνητά Νευρωνικά Δίκτυα για την λειτουργία του και έχει εκπαιδευθεί μέσω 17000 μηνυμάτων από την πλατφόρμα Twitter, τα οποία δημοσιεύθηκαν κατά την εκλογική περίοδο του έτους 2012 στις ΗΠΑ. Το εργαλείο έχει επίσης αξιολογηθεί από την υπηρεσία Amazon Mechanical Turk (AMT) της εταιρείας Amazon. Κατά την εν λόγω αξιολόγηση, εργαζόμενοι της Amazon πραγματοποίησαν εισαγωγή δεδομένων, προσδίδοντας ετικέτες σε ένα σύνολο μηνυμάτων. Οι ετικέτες έπαιρναν τις τιμές θετικό, αρνητικό, ουδέτερο ή μη αναγνωρίσιμο. Στην συνέχεια, το ίδιο σύνολο μηνυμάτων δόθηκε σαν είσοδο στο εργαλείο SASA και τα αποτελέσματα ήταν πολύ κοντά με αυτά που προέκυψαν από την μη αυτόματη απόδοση ετικετών από την υπηρεσία AMT.

4.8 HappinessIndex

Το Happiness Index (P. S. Dodds, C. M. Danforth, 2009: 441–456) αποτελεί ένα εργαλείο εξαγωγής συναισθημάτων το οποίο χρησιμοποιεί μία συλλογή 1034 λέξεων, οι οποίες έχουν τις αντίστοιχες ετικέτες πολικότητας. Το σύνολο των λέξεων ονομάζεται Affective Norms for English Words (ANEW). Το Happiness Index προσδίδει μία τιμή πολικότητας σε ένα κείμενο, η οποία παίρνει τιμές στο σύνολο [1,9], με το 1 να σημαίνει εξαιρετικά αρνητικό και το 9 εξαιρετικά θετικό. Πιο συγκεκριμένα, για κάθε λέξη του κειμένου η οποία υπάρχει μέσα στο σύνολο ANEW, υπολογίζεται ένας μέσος όρος των τιμών αυτών. Όπως είναι φυσικό, οι λέξεις στο σύνολο ANEW έχουν ήδη τις τιμές 1 έως 9.

4.9 PANAS-t

Το PANAS-t (P. Gonçalves, F. Benevenuto, M. Cha, 2013) αποτελεί μία ψυχομετρική κλίμακα (ή κλίμακα συναισθημάτων) η οποία έχει ως σκοπό την αναγνώριση συναισθηματικών διακυμάνσεων των χρηστών από τα μηνύματα τα οποία δημοσιεύουν στην υπηρεσία Twitter. Η μέθοδος βασίζεται πάνω σε μία άλλη μέθοδο, η οποία ονομάζεται Positive Affect Negative Affect Scale (PANAS) (D. Watson, L. Clark, 1985), και αποτελεί μία καλά ορισμένη μέθοδο της επιστήμης της Ψυχολογίας. Η PANAS-t βασίζεται σε ένα μεγάλο σύνολο λέξεων, οι οποίες είναι συνδυασμένες με 11 συναισθήματα. Τα συναισθήματα αυτά είναι τα εξής:

- ευγένεια
- διαβεβαίωση
- ηρεμία
- έκπληξη
- φόβος
- θλίψη
- ενοχή
- εχθρότητα
- συστολή
- κόπωση
- προσοχή

Σημειώνεται ότι τα παραπάνω συναισθήματα αντιστοιχούν τόσο σε θετικές όσο και σε αρνητικές πολικότητες. Η μέθοδος έχει σχεδιαστεί με τρόπο τέτοιο έτσι ώστε να είναι σε θέση να ανιχνεύει αύξηση ή μείωση των εν λόγω συναισθημάτων με την πάροδο του χρόνου. Για να μπορεί να συσχετίσει ένα κείμενο σε ένα συγκεκριμένο συναίσθημα, η PANAS-t χρησιμοποιεί αρχικά μία βάση με τις κανονικοποιημένες τιμές για κάθε συναίσθημα, βασισμένη στα καθολικά δεδομένα. Έπειτα η μέθοδος υπολογίζει την τιμή $P(s)$ για κάθε συναίσθημα "s". Το εν λόγω μέγεθος παίρνει τιμές στο διάστημα [-1.0, 1.0]. Η κάθε τιμή αντιστοιχεί σε αύξηση ή αντίστοιχη μείωση του εν λόγω συναισθήματος. Για παράδειγμα, αν για ένα δεδομένο σύνολο μηνυμάτων (tweets) η τιμή για το συναίσθημα $s = \text{"έκπληξη"}$ είναι $P(\text{"έκπληξη"}) = 0.250$, αυτό σημαίνει ότι το συναίσθημα αυτό είχε μία αύξηση της τάξης του 25% για μία συγκεκριμένη ημέρα. Αντίστοιχα, η τιμή $P(s) = -0.015$ σημαίνει ότι το συναίσθημα "s" μειώθηκε κατά 1.5% σε μία συγκεκριμένη ημέρα.

4.10 EmoTxt

Το σύστημα EmoTxt (Fabio Calefato, Filippo Lanubile, Nicole Novielli, 2017) αποτελεί ένα σύστημα εξαγωγής/αναγνώρισης συναισθημάτων από κείμενο, το οποίο έχει εκπαιδευθεί και ελεγχθεί με την χρήση 9000 περίπου μηνυμάτων, τα οποία περιλαμβάνουν ερωτήσεις, απαντήσεις και σχόλια από διάφορες πηγές στο Διαδίκτυο. Για την εκπαίδευση του μοντέλου, έγινε χρήση Μηχανών Διανυσμάτων Υποστήριξης (Support Vector Machines – SVM) καθώς και του στατιστικού εργαλείου R. Οι Μηχανές Διανυσμάτων Υποστήριξης αποτελούν ένα πολύ χρήσιμο μοντέλο στην περίπτωση δεδομένων μεγάλων διαστάσεων, όπου πολλές τιμές μπορούν να λείπουν από τις διαστάσεις κάθε παραδείγματος εκπαίδευσης. Τα συναισθήματα τα οποία προσπαθεί να αναγνωρίσει το EmoTxt είναι τα ακόλουθα:

- Χαρά
- Αγάπη
- Λύπη
- Θυμός
- Έκπληξη
- Φόβος

Για κάθε ένα από τα παραπάνω συναισθήματα, το μοντέλο προσδίδει μία πιθανότητα εμφάνισής του, στο υπό εξέταση κείμενο.

4.11 Σύνοψη των μεθόδων

Στον Πίνακα 3 παρουσιάζονται με συγκεντρωτικό τρόπο οι μέθοδοι εξόρυξης συναισθημάτων από φυσική γλώσσα, οι οποίοι παρουσιάστηκαν στο τρέχον κεφάλαιο. Για κάθε μέθοδο παρουσιάζεται μία σύντομη περιγραφή καθώς και οι συνήθεις περιπτώσεις χρήσης της, όπου η μέθοδος κρίνεται περισσότερο κατάλληλη να χρησιμοποιηθεί.

Μέθοδος	Σύντομη Περιγραφή	Περιπτώσεις Χρήσης
Ανάλυση των emoticons	Προσπάθεια εύρεσης της πολικότητας ενός μηνύματος με βάση τα emoticons τα οποία	Εύρεση πολικότητας σε μηνύματα κοινωνικών δικτύων. Χρήσιμο για τις περιπτώσεις τις οποίες δεν υπάρχει καθόλου

	παρουσιάζονται σε αυτό.	ενδιαφέρον για τη ποσόστωση του συναισθήματος, παρά μόνο την ανίχνευσή του.
LIWC	Χρήση λεξικού-ευρετηρίου λέξεων όπου κάθε λέξη αντιστοιχεί σε κάποια συναισθήματα.	Χρήσιμο στην περίπτωση που είναι επιθυμητή η ανίχνευση περισσότερων συναισθημάτων και όχι μόνο θετικών/αρνητικών.
SentiStrength	Ανίχνευση θετικών και αρνητικών συναισθημάτων, σε σύντομα μηνύματα, στις παρακάτω κλίμακες: Θετικά: [1, 5] Αρνητικά: [-1, -5]	Τα διαθέσιμα μηνύματα είναι μικρά σε μέγεθος. Χρήσιμο όταν επιδιώκεται μόνο η πολικότητα των συναισθημάτων. Είναι επιθυμητή η έκφραση της πολικότητας σε ένα εύρος τιμών, και όχι μόνο την ανίχνευσή της.
SentiWordNet	Ομαδοποιεί τις λέξεις ενός κειμένου σε synsets και υπολογίζει το συναίσθημα για κάθε ένα από αυτά.	Χρήσιμο για την ανίχνευση της πολικότητας μηνυμάτων στο διάστημα [0, 1].
SenticNet	Εξάγει την πολικότητα ενός μηνύματος στο διάστημα [-1, 1], δίνοντας όμως βάση στα σημασιολογικά χαρακτηριστικά του και όχι μόνο στις λέξεις που αυτό περιέχει.	Χρήσιμο για την εξαγωγή πολικότητας στο εύρος τιμών [-1, 1]. Όταν επιδιώκεται η σημασιολογική ανάλυση των μηνυμάτων, ανεξαρτήτως του αν περιέχουν αρνητικές ή θετικές λέξεις στο περιεχόμενό τους.
SASA	Μοντέλο εξόρυξης συναισθημάτων το οποίο έχει εκπαιδευθεί μέσω μηνυμάτων της εταιρείας Twitter. Για	Ιδανικό στην περίπτωση εξόρυξης συναισθημάτων από μηνύματα της υπηρεσίας Twitter.

	κάθε μήνυμα προσδίδει μία εκ των τιμών: θετικό, αρνητικό, ουδέτερο και μη αναγνωρίσιμο.	Δεν επιδιώκεται καθόλου η ποσόστωση του συναισθήματος, παρά μόνο η ανίχνευσή του.
Happiness Index	Ανίχνευση πολικότητας σε ένα κείμενο με τιμές στο διάστημα [1, 9].	Εύρεση πολικότητας στο εύρος τιμών [1, 9].
PANAS-t	Αναγνώριση 11 διαφορετικών συναισθημάτων σε ένα κείμενο. Το κάθε συναίσθημα το οποίο αναγνωρίζεται παίρνει τιμές στο εύρος [-1, 1]	Επιδιώκεται η αναγνώριση των εν λόγω 11 συναισθημάτων σε ένα κείμενο. Επιπλέον, το πόσο δυνατό είναι ένα συναίσθημα και συνεπώς είναι σημαντική η απόδοση μίας εκ των τιμών στο διάστημα [-1, 1].
EmoTxt	Αναγνώριση 6 διαφορετικών συναισθημάτων σε ένα κείμενο: Χαρά Αγάπη Λύπη Θυμός Έκπληξη Φόβος	Όταν υπάρχει ενδιαφέρον για αναγνώριση των εν λόγω 6 συναισθημάτων σε ένα κείμενο. Δεν υπάρχει μόνο ενδιαφέρον στο να ανιχνευθεί το κυρίαρχο συναίσθημα αλλά και η πιθανότητα/βεβαιότητα εμφάνισης του καθενός για τα εν λόγω συναισθήματα.

Πίνακας 3: Σύνοψη των μεθόδων εξόρυξης συναισθημάτων από φυσική γλώσσα.

Κεφάλαιο 5

Παρουσίαση της μεθοδολογίας

5.1 Εισαγωγή

Στο εν λόγω κεφάλαιο θα παρουσιάσουμε την μεθοδολογία η οποία ακολουθήθηκε για τη συλλογή, ανάλυση και παρουσίαση των πειραματικών αποτελεσμάτων στη παρούσα μεταπτυχιακή διατριβή. Σκοπός είναι να δοθεί μία εποπτική εικόνα στον αναγνώστη πριν την λεπτομερή εκτέλεση των πειραμάτων.

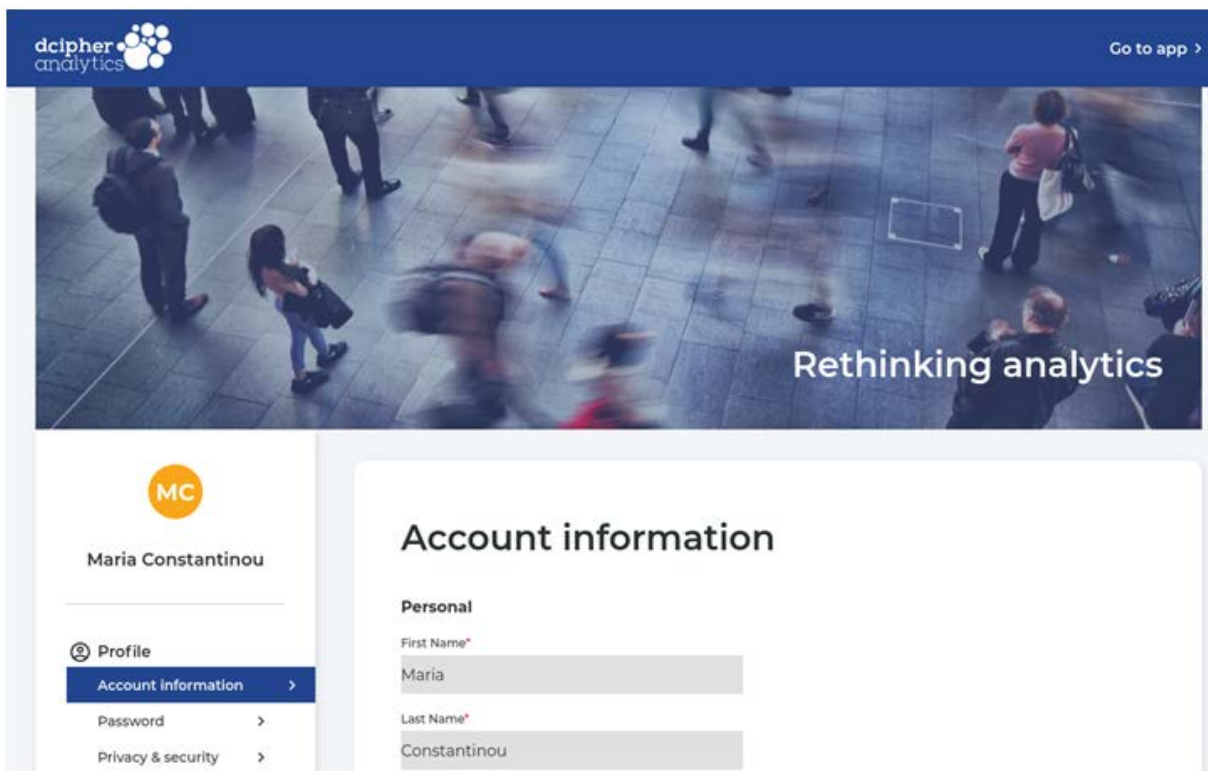
Στόχος των πειραμάτων τα οποία θα εκτελεστούν είναι να εξαχθούν τα συναισθήματα από μηνύματα του κοινωνικού δικτύου Twitter, με την χρήση κατάλληλου λογισμικού και στην συνέχεια να ελεγχθεί κατά πόσο η εξαγωγή των εν λόγω συναισθημάτων ήταν επιτυχής. Για να μπορεί να γίνει ο εν λόγω έλεγχος, θα χρησιμοποιηθούν σύνολα δεδομένων για τα οποία γνωρίζουμε ήδη το συναίσθημα το οποίο εκφράζεται για κάθε ένα από τα επιμέρους μηνύματα. Για την εξαγωγή των συναισθημάτων επιλέξαμε να χρησιμοποιήσουμε το λογισμικό Dcipher Analytics (21), το οποίο διαθέτει λειτουργίες ανάλυσης κειμένου καθώς και εργαλεία εξαγωγής συναισθημάτων από αυτό.

5.2 Το λογισμικό Dcipher Analytics

Το λογισμικό Dcipher Analytics (21) (Εικόνα 1) αποτελεί ένα πολύ δημοφιλές λογισμικό ανάλυσης μη δομημένου και ημι-δομημένου κειμένου, το οποίο χρησιμοποιείται κατά κόρον για την εξαγωγή συναισθημάτων από κείμενα.



Εικόνα 1.: Το λογότυπο του Dcipher Analytics



Εικόνα 2: δημιουργία λογαριασμού για την χρήση της πλατφόρμας Dcipher

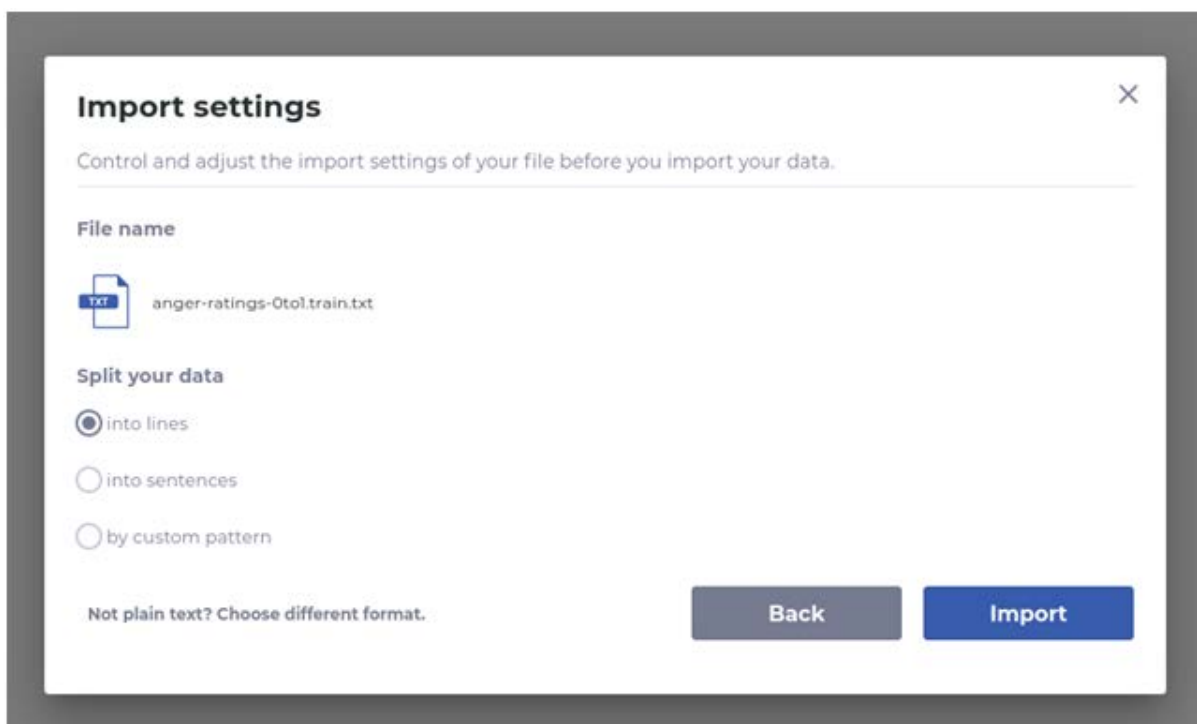
Το εν λόγω λογισμικό είναι διαθέσιμο μέσω του Διαδικτύου, ως μια διαδικτυακή εφαρμογή. Παρόλο που η χρήση του γίνεται επί πληρωμή, προσφέρεται και σε δωρεάν δοκιμαστικές εκδόσεις. Για την χρήση του θα πρέπει ο χρήστης να προβεί αρχικά σε εγγραφή χρήστη (Εικόνα 2) όπως έχει γίνει για σκοπούς της εν λόγω μελέτης.

Η χρήση της πλατφόρμας Dcipher είναι εύκολη και γρήγορη. Συγκεκριμένα, ο χρήστης θα πρέπει αρχικά να δημιουργήσει μία μελέτη (project) για να μπορεί να εκτελέσει τα κατάλληλα πειράματα. Μετά την δημιουργία της μελέτης, ο χρήστης θα πρέπει να ανεβάσει στην πλατφόρμα Dcipher τα κατάλληλα δεδομένα, τα οποία θα επεξεργαστεί. Τα δεδομένα αυτά δίνονται συνήθως σε μορφή αρχείου κειμένου, σε κάποιο από τα δημοφιλή πρότυπα, όπως για παράδειγμα CSV ή JSON. Ένα παράδειγμα εισαγωγής παρουσιάζεται στην Εικόνα 3.



Εικόνα 3: εισαγωγή δεδομένων σε μία μελέτη του Dcipher

Επίσης κατά την διαδικασία εισαγωγής, μπορούμε να επιλέξουμε διάφορες παραμέτρους, όπως για παράδειγμα το αν η κάθε εγγραφή του αρχείου αφορά μία γραμμή αυτού ή αν η τμηματοποίηση γίνεται με διαφορετικό τρόπο (Εικόνα 4).



Εικόνα 4: διάφορες παράμετροι κατά την εισαγωγή των δεδομένων

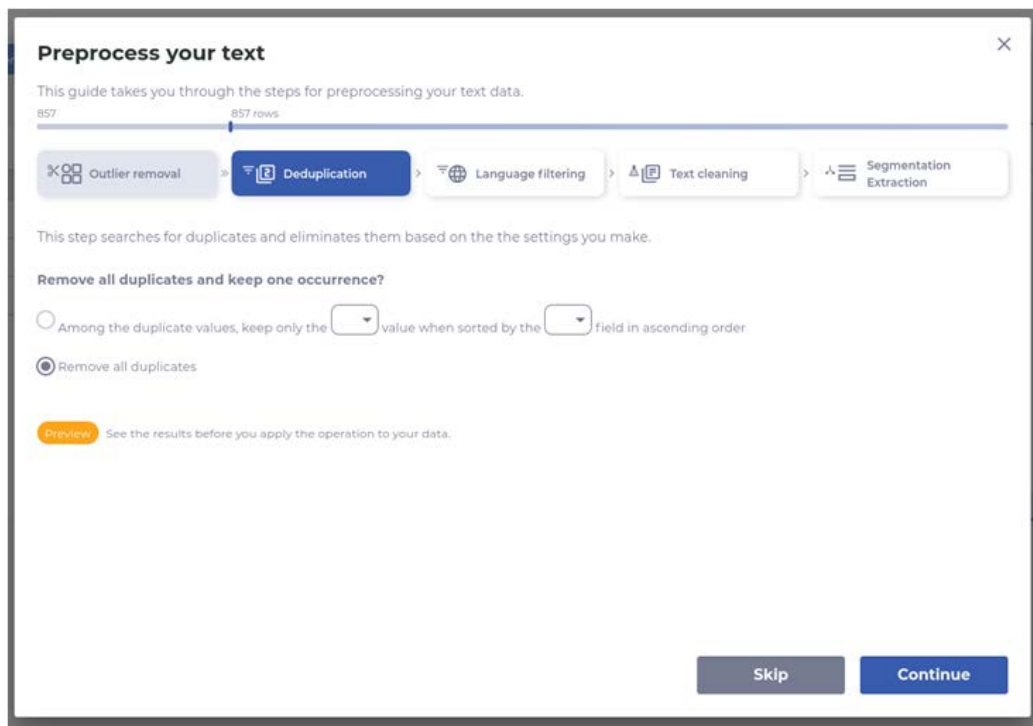
Στην συνέχεια, μετά την εισαγωγή των δεδομένων, ο χρήστης μπορεί να δει και να τροποποιήσει το σύνολο δεδομένων το οποίο ανέβασε, χρησιμοποιώντας τα διαθέσιμα εργαλεία της πλατφόρμας, όπως αυτό παρουσιάζεται στην Εικόνα 5.

The screenshot shows a data pipeline interface. On the left, a 'Schema' panel lists fields: 'anger-test' (entity collection), 'id' (id), and 'value' (long text). On the right, a 'Table View' panel shows a table with 857 rows. The table has columns 'id' and 'value'. The 'value' column contains various text snippets, some of which are duplicates of the same text.

id	value
10764	"@realDonaldTrump support thanks for saving my wife and I were getting our iPhones today and then being ba...
10765	"@realDonaldTrump not turning up? Why am I not surprised. Late for work again! #turninganger/0.688"
10766	"@realDonaldTrump Officers... should NOT have the right to just 'shoot' human beings without provocation. It's wro...
10767	"@realDonaldTrump already/anger/0.688"
10768	"@realDonaldTrump... fat party and I instantly get belabored about it. #turninganger/0.688"
10769	"@realDonaldTrump safe magic and tracking but resent them/anger/0.688"
10770	"@realDonaldTrump @SheriffClem @FarkensHeader This man is an effort to our justice system/anger/0.6...
10771	"@realDonaldTrump LET SHAKES IN MY HOUSE, I BET IT @his_Bus_Truck I KILL THAT BUGGED WHEN I GET MY ...
10772	"@realDonaldTrump unless the big boys 'with Simon State' taking 70%. It's expect an investigation forthw...
10773	"@realDonaldTrump black armed thug with a record carrying gun illegally gets shot by black cop. #outrage...
10774	"@realDonaldTrump... unless Lee and says 'have you tried' (changing her name)" @W5014... #wack/anger/0.688"

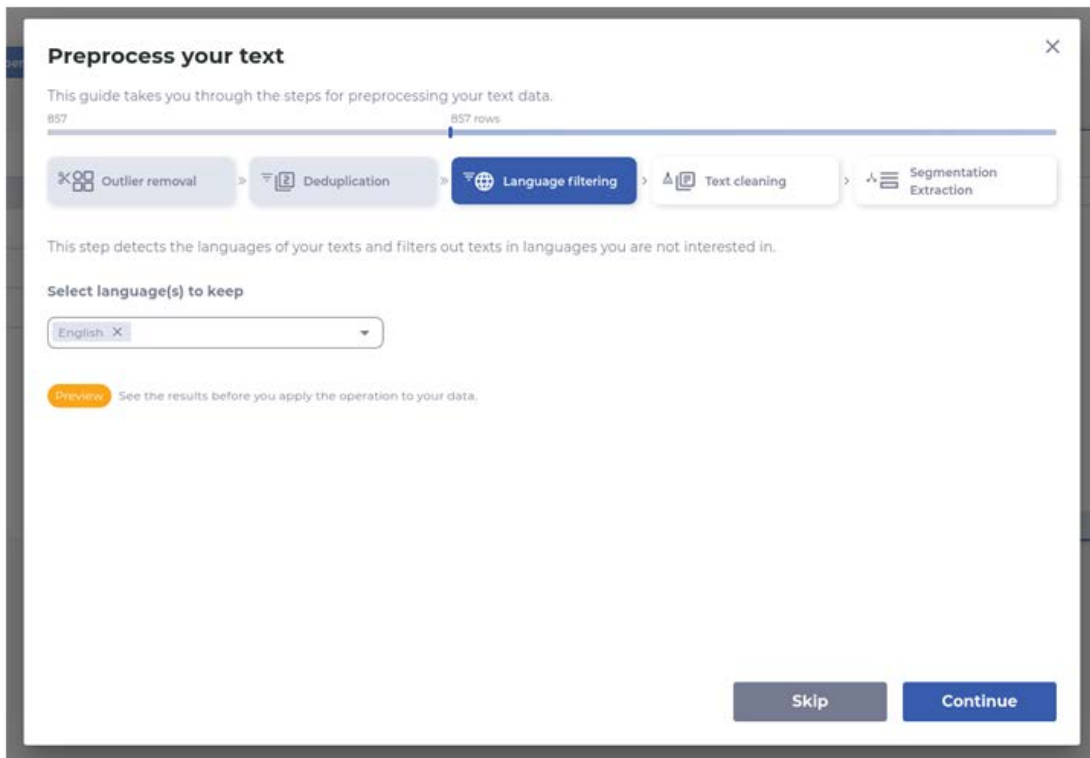
Εικόνα 5: παράδειγμα θέασης και τροποποίησης ενός συνόλου δεδομένων

Το σύνολο δεδομένων μπορεί να τροποποιηθεί με την χρήση των εργαλείων τα οποία παρέχονται κάτω από την επιλογή Preprocess. Πιο συγκεκριμένα, ο χρήστης μπορεί να επιλέξει την αφαίρεση των διπλότυπων εγγραφών όπως παρουσιάζεται στην Εικόνα 6.



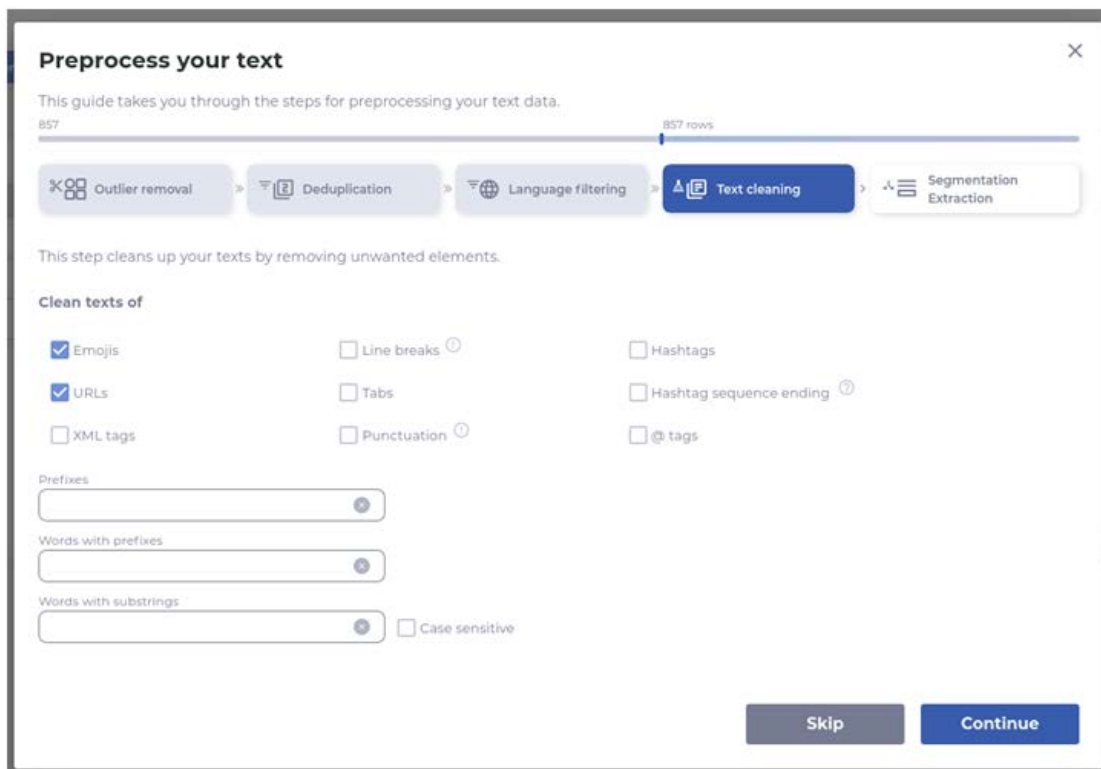
Εικόνα 6.: εργαλείο αφαίρεσης διπλότυπων εγγραφών

Ο χρήστης στην συνέχεια μπορεί να επιλέξει να κρατήσει μόνο τις εγγραφές που αφορούν μία συγκεκριμένη γλώσσα (ή περισσότερες), όπως παρουσιάζεται στην Εικόνα 7.



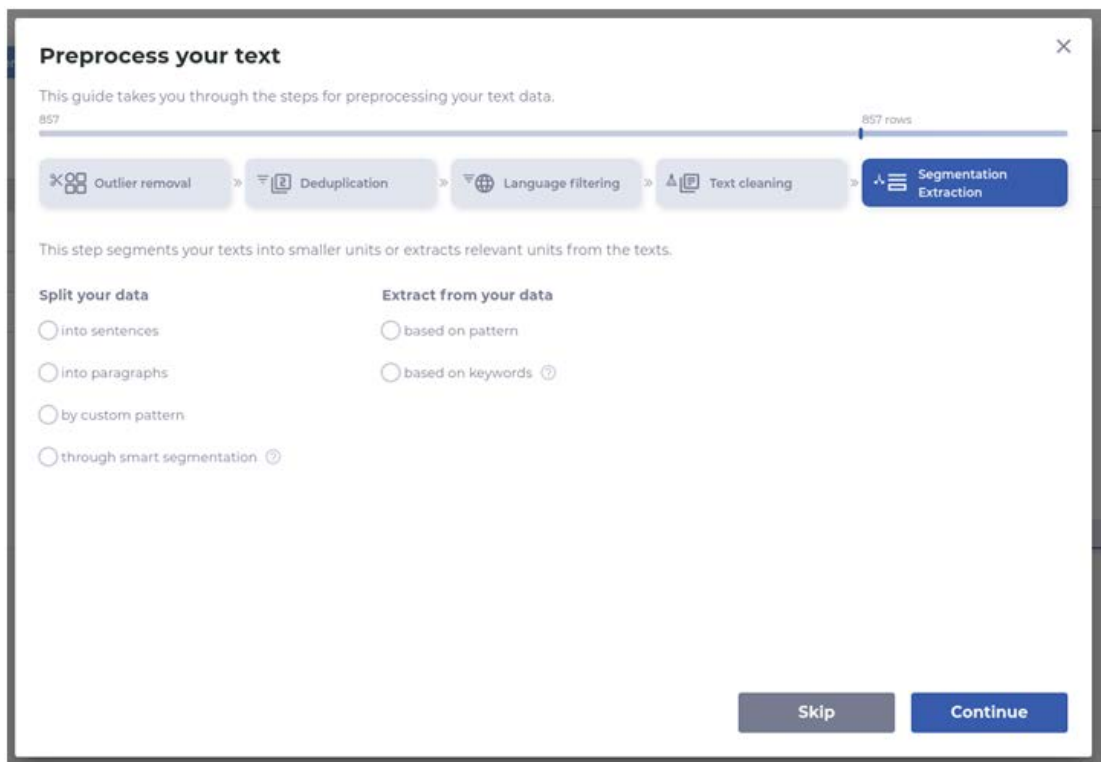
Εικόνα 7. : επιλογή των εγγραφών που αφορούν μία ή περισσότερες γλώσσες.

Το επόμενο βήμα (Εικόνα 8) αφορά τον καθαρισμό του κειμένου από περιεχόμενο το οποίο αφορά μη ωφέλιμη πληροφορία, όπως είναι εικονίδια (emojis) ή και URLs. Φυσικά, στην περίπτωση εξαγωγής συναισθημάτων από κείμενα, τα εικονίδια μπορούν σε κάποιες περιπτώσεις να συνεισφέρουν στον προσδιορισμό του συναισθήματος το οποίο παρουσιάζεται στο κείμενο.



Εικόνα 8.: αφαίρεση στοιχείων από ένα κείμενο, έτσι ώστε αυτό να μην εμπεριέχει επιπρόσθετη, μη ωφέλιμη πληροφορία.

Στο τελευταίο βήμα της διαδικασίας της προ επεξεργασίας το εργαλείο Dcipher Analytics προσφέρει το σπάσιμο των μεγάλων κειμένων σε μικρότερα τμήματα πληροφορίας, όπως για παράδειγμα σε προτάσεις ή σε παραγράφους (Εικόνα 9). Ειδικά στις περιπτώσεις όπως αυτές της εξαγωγής συναισθημάτων, θεωρείται ευκολότερο να ανιχνευθεί το κυρίαρχο συναίσθημα σε μία πρόταση (ή εν τέλει σε λίγες προτάσεις) από ότι να ανιχνευθεί σε ένα μεγαλύτερο κείμενο.



Εικόνα 9.: τμηματοποίηση των εγγραφών του κειμένου σε μικρότερα τμήματα.

Κεφάλαιο 6

Παρουσίαση των δεδομένων προς ανάλυση

6.1 Εισαγωγή

Στο κεφάλαιο αυτό, παρουσιάζονται τα σύνολα δεδομένων τα οποία χρησιμοποιήθηκαν τόσο για την εκπαίδευση όσο και στο μετέπειτα έλεγχο του μοντέλου μηχανικής μάθησης. Τα σύνολα δεδομένων τα οποία χρησιμοποιήθηκαν αναφέρονται πιο κάτω:

6.2 Το σύνολο δεδομένων WASSA-2017

Το σύνολο δεδομένων WASSA-2017 (14) αποτελεί ένα από τα πλέον δημοφιλή σύνολα δεδομένων για την αναγνώριση συναισθημάτων μέσα από κείμενα. Πιο συγκεκριμένα, αποτελεί μία συλλογή μηνυμάτων τα οποία έχουν δημοσιευθεί στην υπηρεσία Twitter και για τα οποία είναι ήδη γνωστό τόσο το συναίσθημα το οποίο εκφράζεται σε κάθε μήνυμα όσο και η ένταση αυτού. Κάθε μήνυμα έχει την παρακάτω μορφή:

[Κωδικός μηνύματος] [Κείμενο μηνύματος] [Συναίσθημα] [Ένταση]

Η ένταση του συναισθήματος είναι μία πραγματική τιμή, οι οποία παίρνει τιμές στο διάστημα $[0,1]$. Στον Πίνακα 4 που ακολουθεί παρουσιάζεται ένα δείγμα από κάποια μηνύματα από το εν λόγω σύνολο δεδομένων.

Κωδικός	Μήνυμα	Συναίσθημα	Ένταση
30000	Just got back from seeing @GaryDelaney in Burslem. AMAZING!! Face still hurts from laughing so much #hilarious	joy	0.98
30001	Oh dear an evening of absolute hilarity I don't think I have laughed so much in a long time! 😊	joy	0.96
30002	Been waiting all week for this game ♡♡♡#cheer #friday ♡	joy	0.94
30003	@gardiner_love : Thank you so much, Gloria! You're so sweet, and thoughtful! You just made my day more joyful! I love you too! 😊❤️	joy	0.94
30004	I feel so blessed to work with the family that I nanny for ♡ nothing but love & appreciation, makes me smile.	joy	0.94
30005	Today I reached 1000 subscribers on YT!! , #goodday, #thankful	joy	0.93

30006	@Singaholic121 Good morning, love! Happy first day of fall. Let's make some awesome #autumnmemories #annabailey #laughter #smile	joy	0.92
30007	#BridgetJonesBaby is the best thing I've seen in ages! So funny, I've missed Bridget! #love #TeamMark	joy	0.92
30008	Just got back from seeing @GaryDelaney in Burslem. AMAZING!! Face still hurts from laughing so much	joy	0.92
30009	@IndyMN I thought the holidays could not get any more cheerful, and then I met you. #TheNiceBot	joy	0.92
30010	I'm just still . So happy .\nA blast	joy	0.92
30011	It's meant to be!! #happy #happy	joy	0.92
30012	🌟👏 Yeah👏 PAUL👏👏🌟 #glorious #BB18	joy	0.92
30013	My morning started off amazing!! Hopefully	joy	0.92

	the whole day is going as i want it to go!\n #GreatDay		
30014	😁 @cailamarsai you've had me 😁 😁 time watching @black_ishABC after you've lost your #glasses! It was #hilarious! @mrbabyboogaloo	joy	0.92
30015	@iamTinaDatta love you so much #smile 😊😊	joy	0.9

Πίνακας 4: Παραδείγματα εγγραφών για το σύνολο δεδομένων WASSA-2017

6.3 Το σύνολο δεδομένων Pre-processed Twitter tweets

Το δεύτερο σύνολο δεδομένων το οποίο χρησιμοποιήθηκε στην μελέτη είναι διαθέσιμο στην δημοφιλή πλατφόρμα ανάλυσης δεδομένων Kaggle (23). Το σύνολο αποτελείται από τρία επιμέρους υποσύνολα μηνυμάτων της υπηρεσίας Twitter: ένα για θετικά, αρνητικά και ουδέτερα μηνύματα αντίστοιχα. Για κάθε μήνυμα παρουσιάζεται μόνο το αντίστοιχο συναίσθημα, αλλά δεν φαίνεται ο βαθμός του συναισθήματος. Η ταξινόμηση όλων των μηνυμάτων έχει ήδη πραγματοποιηθεί ενώ τα δεδομένα δεν χωρίζονται σε σύνολα εκπαίδευσης και ελέγχου. Ένα παράδειγμα των δεδομένων του εν λόγω συνόλου δεδομένων παρουσιάζεται παρακάτω, στον Πίνακα 5.

Μήνυμα	Συναίσθημα
beauty and personality. :)	Θετικό
KISSES The Fashion Icon	Θετικό
Please Visit Or Likes Share :)	Θετικό
Beautiful album from the greatest unsung guitar genius of our time - and I've met the great backstage	Θετικό
Good luck to Rich riding for great project in this Sunday. Can you donate?	Θετικό
Omg he... kissed... him crying with joy happy annivming and papi!!!!	Θετικό
love happy thanks happy	Θετικό
Join vote for the singer! Do spread the word. :D	Θετικό
Thanks for the great review! smilenYay another art raffle! Everything you need to know is in the picture :D	Θετικό
Hello I hope you visit Luxor its amazing city in Egypt pleas check	Θετικό
We got a Vive tracker in the office and our intern went to work.	Θετικό
Take a look at favourites.io You can do this and more happy	Θετικό
Go back to school for music! I think I will in time happy	Θετικό
Sixth spot not applicable Team! Higher pa! :) KISSES TheFashionIcon	Θετικό
thanks for being top influencers in my community last week happy	Θετικό
Share the love: thanks for being top new followers this week happy	Θετικό
Want this? retweet this tweet if you want your to be included in the nhc scrapbook!! i'll be making something out of all of the users	Θετικό
Happy 420 happy pass it to your followers!	Θετικό
Who wants some good vibes? Watch this on YouTube or TV100 application now for FREE happy	Θετικό
Great to work with you too such a lovely place!	Θετικό
What looks bad today will be a blessing tomorrow happy Keep your faith stirred up each day!	Θετικό
You deserve a vacation happy	Θετικό
Never known how hard it is until last to days how to resist temptation to enter politics debates on social media happy stood firm so far	Θετικό
Some summer salad inspiration for the BBQ smile	Θετικό

Hey thanks for being top new followers this week! Much appreciated happy	Θετικό
Gouna 20km away is also quite nice - honestly anywhere away from crazy cairo is Sixth spot not applicable Team!	Θετικό
Higher pa! :)KISSES TheFashionIcon	Θετικό
Hey thanks for being top new followers this week!	Θετικό
Much appreciated happy thanks for the recent follow. Much appreciated happy	Θετικό
Thanks Lala.. Much needed happy	Θετικό
How unhappy some dogs like it though	Αρνητικό
talking to my over driver about where I'm going he said he'd love to go to New York too but since Trump it's probably not	Αρνητικό
Does anybody know if the Rand's likely to fall against the dollar?	Αρνητικό
I got some money I need to change into R but it keeps getting stronger unhappy	Αρνητικό
I miss going to gigs in Liverpool unhappy	Αρνητικό
There isnt a new Riverdale tonight ? unhappy	Αρνητικό
Don't like how jittery caffeine makes me sad	Αρνητικό
My area's not on the list unhappy think I'll go LibDems anywayI want fun plans this weekend unhappy	Αρνητικό
When can you notice me.unhappy what?	Αρνητικό
Ahhhhh! You recognized LOGAN!!!	Αρνητικό
Cinemax shows have a BAD track record for getting cancelled unhappy	Αρνητικό
Errr dude.... They're gone unhappy Asked other league memeber to check the guys are go	Αρνητικό
Not you again sad	Αρνητικό
Why would Harvey be going to prison? unhappy	Αρνητικό
Missing in crying Seaside area.	Αρνητικό
Becoz if we will depend on your promoting its waste of hard work to all team who	Αρνητικό
I thought you'll save me crying major waffle cravings right now sad	Αρνητικό
cant speak japanese :(Αρνητικό

How can people do stuff like this unhappy	Αρνητικό
please just stop confining animals in zoos unhappy	Αρνητικό

Πίνακας 5: παράδειγμα θετικών και αρνητικών μηνυμάτων από το σύνολο δεδομένων

Κεφάλαιο 7

Πειραματική Μελέτη των δεδομένων προς ανάλυση

7.1 Εισαγωγή

Στο κεφάλαιο αυτό, παρουσιάζονται τα πειράματα τα οποία εκτελέστηκαν στα πιο πάνω σύνολα δεδομένων, μέσω της πλατφόρμας Dcipher Analytics. Πιο συγκεκριμένα, για κάθε σύνολο δεδομένων, έγινε προσπάθεια να εξαχθεί το δηλωθέν συναίσθημα με δύο διαφορετικούς τρόπους:

- Με την χρήση κατάλληλου λεξικού λέξεων.
- Με την χρήση Μηχανικής Μάθησης και πιο συγκεκριμένα με τη μέθοδο Deep Learning.

Η εκτέλεση των πειραμάτων παρουσιάζεται πιο κάτω.

7.2 Εκτέλεση των πειραμάτων


Έχουν εκτελεστεί τα πειράματα τόσο για το σύνολο δεδομένων WASSA-2017 όσο και για το σύνολο δεδομένων Pre-processed Twitter tweets της πλατφόρμας Kaggle. Παρουσιάζεται αναλυτικά η εκτέλεση των πειραμάτων για το πρώτο σύνολο δεδομένων. Οι ενέργειες οι οποίες εκτελέστηκαν για το δεύτερο σύνολο δεδομένο είναι ακριβώς οι ίδιες. Όσον αφορά το σύνολο δεδομένων WASSA-2017, χρησιμοποιήθηκαν τα δεδομένα από το σύνολο εκπαίδευσης, αφού περιείχαν μεγαλύτερο αριθμό εγγράφων από τα υπόλοιπα σύνολα δεδομένων. Τα εργαλεία της πλατφόρμας Dcipher Analytics είναι ήδη εκπαιδευμένα για εξαγωγή συναισθημάτων από κείμενα της αγγλικής γλώσσας και συνεπώς δεν προέκυψε η ανάγκη περαιτέρω εκπαίδευσης.

Βήμα 1ο: δημιουργήθηκε μία νέα pipeline και φορτώθηκαν σε αυτή τα σύνολα δεδομένων για κάθε ένα από τα συναισθήματα. Εικόνα 10.


File upload ✕

Upload your data files to Dcipher Cloud and import data from them into your projects.

Upload a file from your device



Choose file(s) or drag them here

Files	Uploaded at	Size	# files
 TXT files	-	404 KB	4 ^
<input type="checkbox"/> anger-ratings-0to1.train.txt	3 minutes ago	93 KB	
<input type="checkbox"/> fear-ratings-0to1.train.txt	3 minutes ago	130 KB	
<input type="checkbox"/> joy-ratings-0to1.train.txt	3 minutes ago	90 KB	
<input type="checkbox"/> sadness-ratings-0to1.train.txt	3 minutes ago	91 KB	

[Select file\(s\)](#)


Εικόνα 10.: τα σύνολα δεδομένα τα οποία χρησιμοποιήθηκαν.

Βήμα 2ο: κάθε εγγραφή έχει οριστεί να αποτελεί μία διαφορετική γραμμή και όχι μια διαφορετική λέξη ή πρόταση. Εικόνα 11.

Import settings ✕

Control and adjust the import settings of your file before you import your data.

File name

 anger-ratings-0to1.train.txt

Split your data

into lines

into sentences

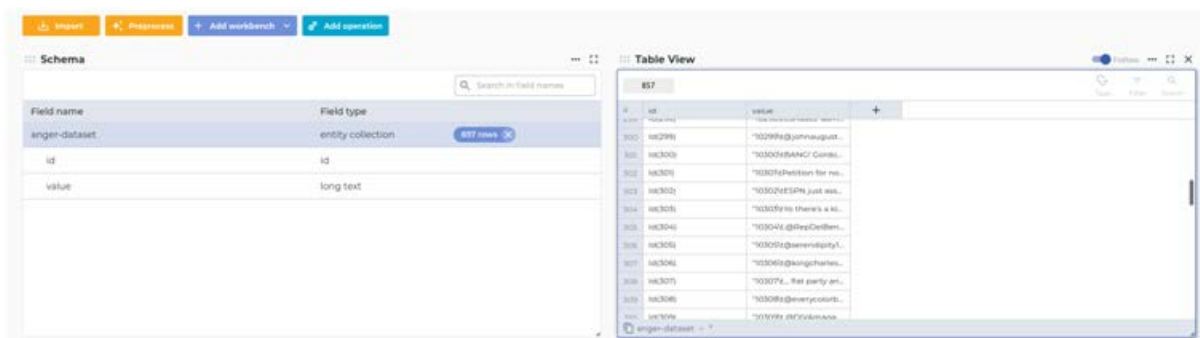
by custom pattern

Not plain text? Choose different format.

Back
Import

Εικόνα 11.: ορισμός της έννοιας της εγγραφής.

Βήμα 3ο: το κάθε αρχείο φορτώθηκε με την σειρά για την εξαγωγή συναισθημάτων. Στην πιο κάτω εικόνα παρουσιάζεται το αρχείο για το συναίσθημα “θυμός” (anger).



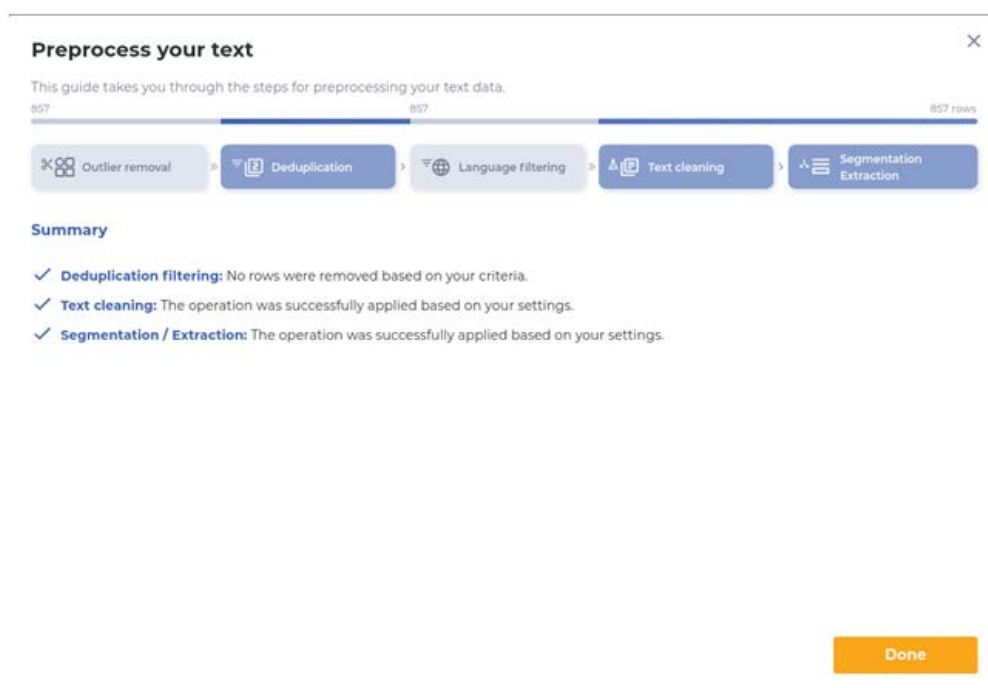
The screenshot shows a data processing interface with two main panels. On the left is the 'Schema' panel, which lists the fields and their types for the 'anger-dataset':

Field name	Field type
anger-dataset	entity collection
id	id
value	long text

On the right is the 'Table View' panel, which displays a table with 857 rows. The columns are 'id' and 'value'. The 'value' column contains various text entries, such as '10296@jornhaugott...', '10300@bainc' Games...', '10301@hudson for me...', '10302@ESPM just aw...', '10303@ there's a ki...', '10304@ @theCeltic...', '10305@ @senneplyf...', '10306@ @kongcharco...', '10307@... Red party am...', '10308@ @everycourt...', and '10309@ @'ndubious'.

Εικόνα 12.: το αρχείο το οποίο αντιστοιχεί στο συναίσθημα “θυμός”.

Βήμα 4ο: αφαιρέθηκαν τυχόν διπλότυπες εγγραφές. Στο συγκεκριμένο βήμα δεν κρίθηκε αναγκαία κάποια άλλη προ-επεξεργασία των δεδομένων.



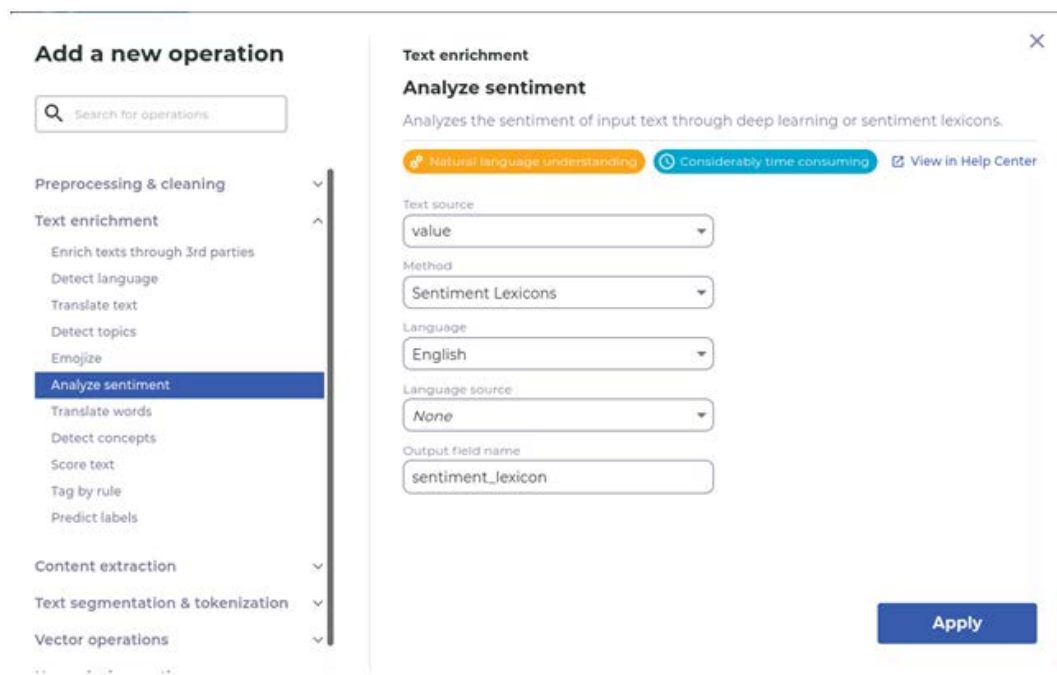
The screenshot shows a 'Preprocess your text' workflow interface. It includes a progress bar with 857 rows and a sequence of steps: Outlier removal, Deduplication, Language filtering, Text cleaning, and Segmentation/Extraction. A 'Summary' section below the steps provides the following information:

- ✓ **Deduplication filtering:** No rows were removed based on your criteria.
- ✓ **Text cleaning:** The operation was successfully applied based on your settings.
- ✓ **Segmentation / Extraction:** The operation was successfully applied based on your settings.

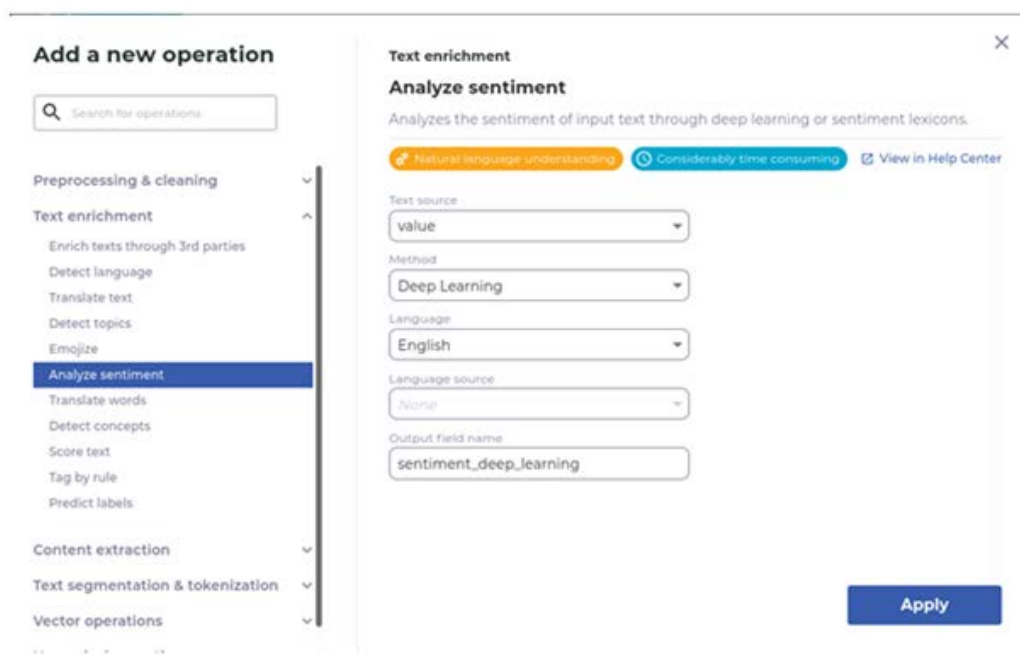
A 'Done' button is visible at the bottom right of the interface.

Εικόνα 13.: αφαίρεση των διπλότυπων εγγραφών.

Βήμα 5ο: εκτελέστηκαν οι ενέργειες εξαγωγής συναισθημάτων με την χρήση λεξικού λέξεων και με την χρήση μηχανικής μάθησης. Για κάθε μία από τις εγγραφές του αρχείου δημιουργήθηκαν δύο περαιτέρω στήλες με τα αποτελέσματα της εξαγωγής συναισθημάτων.



Εικόνα 14.: εξαγωγή συναισθημάτων με βάση λεξικό λέξεων.



Εικόνα 15.: εξαγωγή συναισθημάτων με την χρήση μηχανικής μάθησης.

#	id	value	sentiment_lexicon	sentiment_deep_learning
17	Id(16)	"10016\t@ArizonaCov...	{"score": -1, "label": "NEG...	{"score": 0.9980657696...
18	Id(17)	"10017\tI hate my law...	{"score": -0.5, "label": "N...	{"score": -0.999334990...
19	Id(18)	"10018\twhy are peopl...	{"score": -0.3333333333...	{"score": -0.999735295...
20	Id(19)	"10019\tI'm about to b...	{"score": -1, "label": "NEG...	{"score": -0.9045416712...
21	Id(20)	"10020\tMaking my b...	{"score": -1, "label": "NEG...	{"score": 0.9879708886...
22	Id(21)	"10021\tBes! You don't...	{"score": -1, "label": "NEG...	{"score": -0.999346673...
23	Id(22)	"10022\tI got a NO RE...	{"score": -1, "label": "NEG...	{"score": -0.998886764...
24	Id(23)	"10023\tTasers immo...	{"score": -1, "label": "NEG...	{"score": -0.9109618067...
25	Id(24)	"10024\t@LeKingCoq ...	{"score": -1, "label": "NEG...	{"score": 0.9730685353...
26	Id(25)	"10025\tActually fumi...	{"score": -1, "label": "NEG...	{"score": -0.993725836...
27	Id(26)	"10026\t.@DIVAmaga...	{"score": -0.5, "label": "N...	{"score": -0.9971417704...

Εικόνα 16.: το αποτέλεσμα της εκτέλεσης των πειραμάτων για το συναίσθημα "θυμός".

Ένα τμήμα των αποτελεσμάτων παρουσιάζεται παρακάτω (το αποτέλεσμα είναι σε μορφή JSON):

...

```
{"id": {"_type": "Id", "value": [134]}, "value": "10134\tHow was Natalie one of the top three favorites?! #toofaced and #bitter ḡŸ™,\tanger\t0.667", "sentiment_lexicon": {"score": -0.3333333333333333, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9958796501159668, "label": "NEGATIVE"}},
```

```
{"id": {"_type": "Id", "value": [135]}, "value": "10135\t@dceaminer Democrats and their voters have zero tolerance for honesty. They associate honesty with anger and hate.\tanger\t0.667", "sentiment_lexicon": {"score": -0.2, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.847712516784668, "label": "NEGATIVE"}},
```

```
{"id": {"_type": "Id", "value": [136]}, "value": "10136\t@gerfingerpoken #rage?? The #CrookedCourt said #rage MANY times to explain away the brutal killings of #Petits by #Hayes & #Komisarjevsky\tanger\t0.667", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.96907639503479, "label": "POSITIVE"}},
```

```
{"id": {"_type": "Id", "value": [137]}, "value": "10137\tTo'Why can't we brothers protect one another? No one's serious, and it makes me furious. Don't be misled, just think of Fred.' #mayfield\tanger\t0.578", "sentiment_lexicon": {"score": -0.3333333333333333,
```

"label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.8526448011398315, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [138]}, "value": "10138\t.... wakes up and says 'have you tried changing her nappy?' ġÿĩġÿŠġÿ¼ !!!\tanger\t0.667", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9996054768562317, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [139]}, "value": "10139\t@ManUtd you have had from me over the years is irrelevant. Its an absolute joke. #manutd #ticketing #noloyalty #joke #notimpressed\tanger\t0.667", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9986734390258789, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [140]}, "value": "10140\tAll hell is breaking loose in Charlotte. #CharlotteProtest #anger #looting\tanger\t0.667", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9973357319831848, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [141]}, "value": "10141\tThere's this Bpharm4 guy Eish that guy brings anger into my life. When I see him nje like darkness fills me @RamatsobaneM will know\tanger\t0.667", "sentiment_lexicon": {"score": -0.5, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9309929609298706, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [142]}, "value": "10142\ti had an hour of football practice under the boiling sun and now i have 2hr volleyball practice under the BOILING SUN AGAIN\tanger\t0.667", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9768871665000916, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [143]}, "value": "10143\ttesting #angry\tanger\t0.667", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.7475862503051758, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [144]}, "value": "10144\t@WatcherDev @WarcraftDevs put so much artefact points into fury and cant raid... literally wasted 1 month of my life! Thanks blizzard\tanger\t0.667", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}},

"NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9949068427085876, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [145]}, "value": "10145\tHow is it suppose to work if you do that? Wtf dude? Thanks for pissing me off. #furious\tanger\t0.833", "sentiment_lexicon": {"score": -0.3333333333333333, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.7136101126670837, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [146]}, "value": "10146\ti was angry.\tanger\t0.667", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9989290833473206, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [147]}, "value": "10147\tThat feel when you travel 700 miles to pick up a form that arrives in the post two days after you leave. #fume\tanger\t0.667", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.509036123752594, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [148]}, "value": "10148\t@cineworld 'Congratulations your Free 1 month has been activated' Then charges Â£34.80 the same month. Absolutely furious ğŸ˜¡\tanger\t0.667", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.5932085514068604, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [149]}, "value": "10149\t@tmz @HarveyLevinTMZ Hell hath no fury like a women scorned. It's the affair. Not the parenting\tanger\t0.667", "sentiment_lexicon": {"score": -0.3333333333333333, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9486053586006165, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [150]}, "value": "10150\tstraight people are canoodling on the quad and I'm #offended\tanger\t0.648", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9789780974388123, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [151]}, "value": "10151\tI can definitely feel my irritation today, lord give me strength\tanger\t0.667", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9999302625656128, "label": "NEGATIVE"}},

```
{
  "id": {"_type": "Id", "value": [152]},
  "value": "10152\t@RobertTaitWHU against Chelsea anything is possible haha. I'm fuming about that bet",
  "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"},
  "sentiment_deep_learning": {"score": -0.8629835247993469, "label": "NEGATIVE"}
}
```

```
{
  "id": {"_type": "Id", "value": [153]},
  "value": "10153\tWhen you break a record in #madden I wish it didn't say the same shit after every rush like you just broke the record. #shutupphilsimms",
  "sentiment_lexicon": {"score": -0.6, "label": "NEGATIVE"},
  "sentiment_deep_learning": {"score": -0.9605789184570312, "label": "NEGATIVE"}
}
```

...

Ακολουθήθηκε η ίδια ακριβώς διαδικασία και για τα σύνολα δεδομένων τα οποία αντιστοιχούν στα συναισθήματα “φόβου” (fear dataset), “χαράς” (joy dataset) και “θλίψης” (sadness dataset). Το τελικό αποτέλεσμα της εκτέλεσης των αντίστοιχων πειραμάτων παρουσιάζεται στις Εικόνες 17 – 19.

#	id	value	sentiment_lexicon	sentiment_deep_le...
1	Id(0)	"20000\tI feel like I a...	{"score":-0.6666666666...	{"score":-0.956540763...
2	Id(1)	"20001\tI get so nervo...	{"score":-1,"label":"NEG...	{"score":-0.938857555...
3	Id(2)	"20002\tI lost my blin...	{"score":-1,"label":"NEG...	{"score":-0.999934792...
4	Id(3)	"20003\tI feel like I a...	{"score":-0.6,"label":"N...	{"score":-0.81152433156...
5	Id(4)	"20004\tThis is the sc...	{"score":-1,"label":"NEG...	{"score":0.9992350339...
6	Id(5)	"20005\t@mgcsartwo...	{"score":-1,"label":"NEG...	{"score":-0.831703007...
7	Id(6)	"20006\tI have to fina...	{"score":-1,"label":"NEG...	{"score":-0.5385147333...
8	Id(7)	"20007\t@laura221b I ...	{"score":-0.5,"label":"N...	{"score":-0.8331692814...
9	Id(8)	"20008\tMy bus was i...	{"score":-1,"label":"NEG...	{"score":0.9995350837...
10	Id(9)	"20009\tMy bus was i...	{"score":-1,"label":"NEG...	{"score":0.9963985681...
11	Id(10)	"20010\tprocrasting is...	{"score":-0.5,"label":"N...	{"score":-0.984234273...

Εικόνα 17.: το αποτέλεσμα της εκτέλεσης των πειραμάτων για το συναίσθημα “φόβος”.

#	id	value	sentiment_lexicon	sentiment_deep_je...
1	Id(0)	"30000\tJust got bac...	["score":0.333333333333...	["score":-0.957634389...
2	Id(1)	"30001\tOh dear an ev...	["score":1,"label":"POSI...	["score":-0.845896303...
3	Id(2)	"30002\tBeen waiting...	["score":1,"label":"POSI...	["score":0.98561513423...
4	Id(3)	"30003\t@gardiner_lo...	["score":1,"label":"POSI...	["score":0.9999992847...
5	Id(4)	"30004\tI feel so bless...	["score":1,"label":"POSI...	["score":0.6479146480...
6	Id(5)	"30005\tToday I reach...	["score":1,"label":"POSI...	["score":-0.8193522095...
7	Id(6)	"30006\t@Singaholicl...	["score":0.6,"label":"PO...	["score":0.9546520709...
8	Id(7)	"30007\t#BridgetJon...	["score":0.2,"label":"PO...	["score":0.99583154916...
9	Id(8)	"30008\tJust got bac...	["score":0,"label":"NEU...	["score":-0.766564309...
10	Id(9)	"30009\t@IndyMN I t...	["score":1,"label":"POSI...	["score":0.5765566825...
11	Id(10)	"30010\tI'm just still ...	["score":1,"label":"POSI...	["score":0.9980286955...

Εικόνα 18.: το αποτέλεσμα της εκτέλεσης των πειραμάτων για το συναίσθημα “χαρά”.

#	id	value	sentiment_lexicon	sentiment_deep_je...
1	Id(0)	"40000\tDepression s...	["score":-1,"label":"NEG...	["score":0.99411976337...
2	Id(1)	"40001\tFeeling wort...	["score":-1,"label":"NEG...	["score":-0.898248612...
3	Id(2)	"40002\tFeeling wort...	["score":-1,"label":"NEG...	["score":-0.7649041414...
4	Id(3)	"40003\tMy #Fibromy...	["score":-0.6666666666...	["score":-0.999420046...
5	Id(4)	"40004\tI'm think ima...	["score":-1,"label":"NEG...	["score":-0.9945613741...
6	Id(5)	"40005\tSo when I try...	["score":-1,"label":"NEG...	["score":-0.733900785...
7	Id(6)	"40006\tExtreme sad...	["score":-1,"label":"NEG...	["score":0.9944534897...
8	Id(7)	"40007\tmy life in on...	["score":-1,"label":"NEG...	["score":0.9718669056...
9	Id(8)	"40008\tPanic attacks...	["score":-1,"label":"NEG...	["score":0.6416663527...
10	Id(9)	"40009\tFeel so grim ...	["score":-1,"label":"NEG...	["score":0.9803246855...
11	Id(10)	"40010\tHonestly dep...	["score":-1,"label":"NEG...	["score":0.9416096210...

Εικόνα 19.: το αποτέλεσμα της εκτέλεσης των πειραμάτων για το συναίσθημα “θλίψη”.

Τα πειράματα αφορούσαν τα 4 υποσύνολα δεδομένων (για τα συναισθήματα “θυμός”, “φόβος”, “χαρά” και “θλίψη”) τα οποία ανήκουν στο γενικότερο σύνολο δεδομένων WASSA-2017. Το σύνολο δεδομένων *Pre-processed Twitter tweets* της πλατφόρμας Kaggle περιέχει τρία σύνολα δεδομένων, τα οποία αφορούν θετικά, ουδέτερα και αρνητικά συναισθήματα αντίστοιχα. Και για τα τρία αυτά σύνολα δεδομένων εκτελέστηκε η ίδια ακριβώς διαδικασία. Κρίθηκε σκόπιμο να μην εμφανιστούν αναλυτικά τα αντίστοιχα βήματα της πειραματικής μελέτης, καθώς δεν διαφέρουν από αυτά που παρουσιάστηκαν παραπάνω για το πρώτο σύνολο δεδομένων. Τα συγκεντρωτικά πειραματικά αποτελέσματα τόσο για το πρώτο σύνολο δεδομένων όσο και για το δεύτερο παρουσιάζονται αναλυτικά πιο κάτω. Για μία περισσότερη αναλυτική παρουσίαση των εν λόγω πειραματικών αποτελεσμάτων, ο αναγνώστης μπορεί να διατρέξει στο Παράρτημα Α.

7.3. Ανάλυση των πειραματικών αποτελεσμάτων

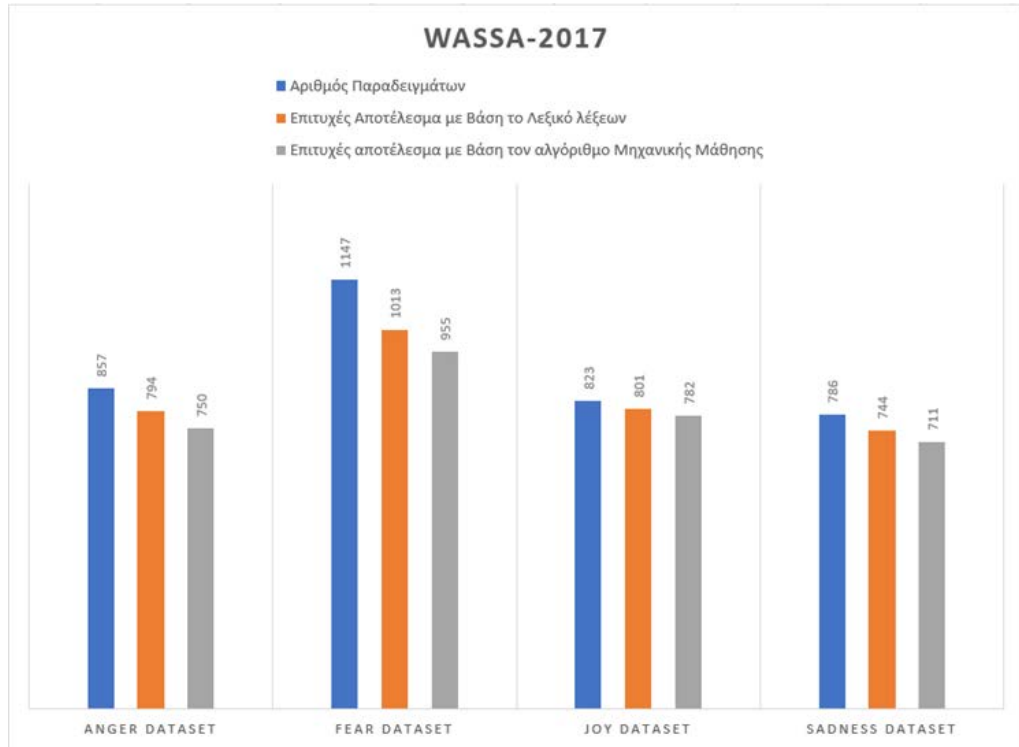
Στην εν λόγω παράγραφο γίνεται παρουσίαση και σύγκριση των πειραματικών αποτελεσμάτων τα οποία εκτελέστηκαν πιο πάνω. Επειδή κάθε ένα από τα δύο σύνολα δεδομένων WASSA-2017 και *Pre-processed Twitter tweets* αποτελούνται από περαιτέρω υπό-σύνολα δεδομένων, η εν λόγω σύγκριση αφορά τόσο τα ευρύτερα σύνολα δεδομένων μεταξύ τους όσο και μεταξύ των υποσυνόλων μέσα στο ευρύτερο σύνολο δεδομένων.

7.4. Παρουσίαση των πειραματικών αποτελεσμάτων

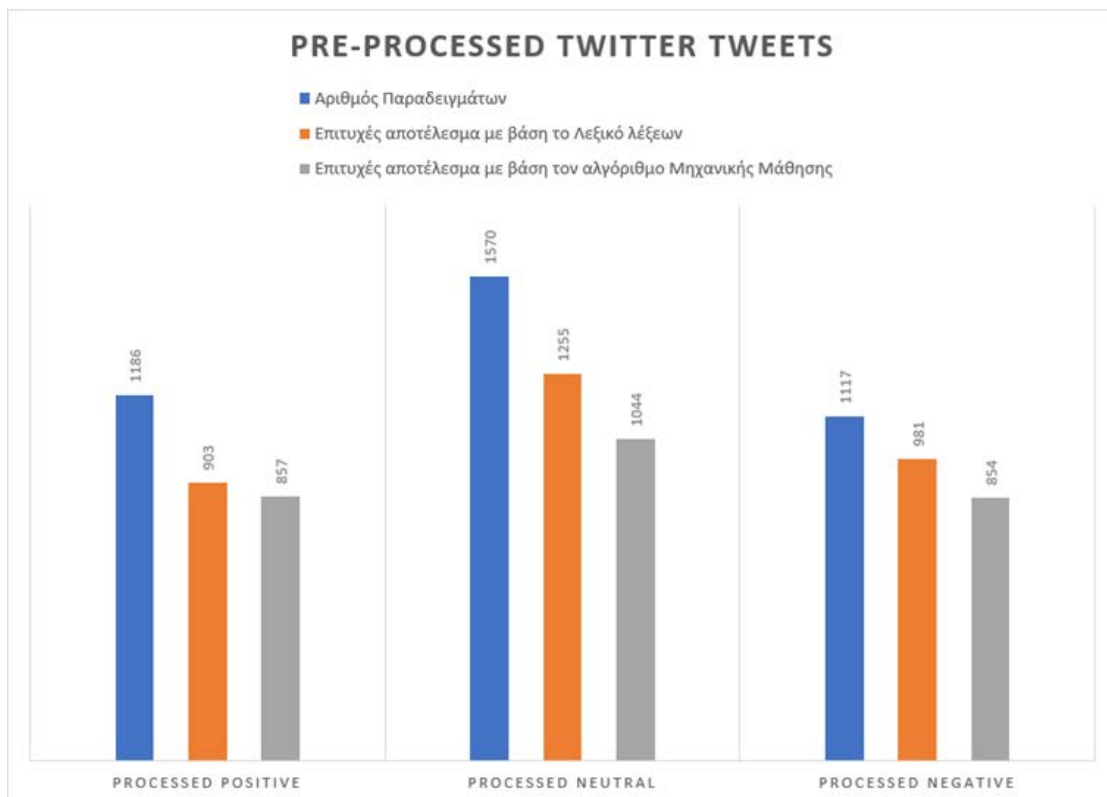
Σύνολο Δεδομένων	Υποσύνολο Δεδομένων	Συναίσθημα	Αριθμός Παραδειγμάτων	Επιτυχές αποτέλεσμα με βάση το Λεξικό λέξεων	Επιτυχές αποτέλεσμα με βάση τον αλγόριθμο Μηχανικής Μάθησης
WASSA-2017	Anger Dataset	Θυμός (αρνητικό)	857	794	750
WASSA-2017	Fear Dataset	Φόβος (αρνητικό)	1147	1013	955
WASSA-2017	Joy Dataset	Χαρά (θετικό)	823	801	782
WASSA-2017	Sadness Dataset	Θλίψη (αρνητικό)	786	744	711
Pre-processed Twitter tweets	Processed Positive	Θετικά συναισθήματα	1186	903	857
Pre-processed Twitter tweets	Processed Neutral	Ουδέτερα συναισθήματα	1570	1255	1044
Pre-processed Twitter tweets	Processed Negative	Αρνητικά συναισθήματα	1117	981	854

Πίνακας 6: τα αποτελέσματα των πειραμάτων με βάση τους αλγορίθμους του εργαλείου Dcipher Analytics.

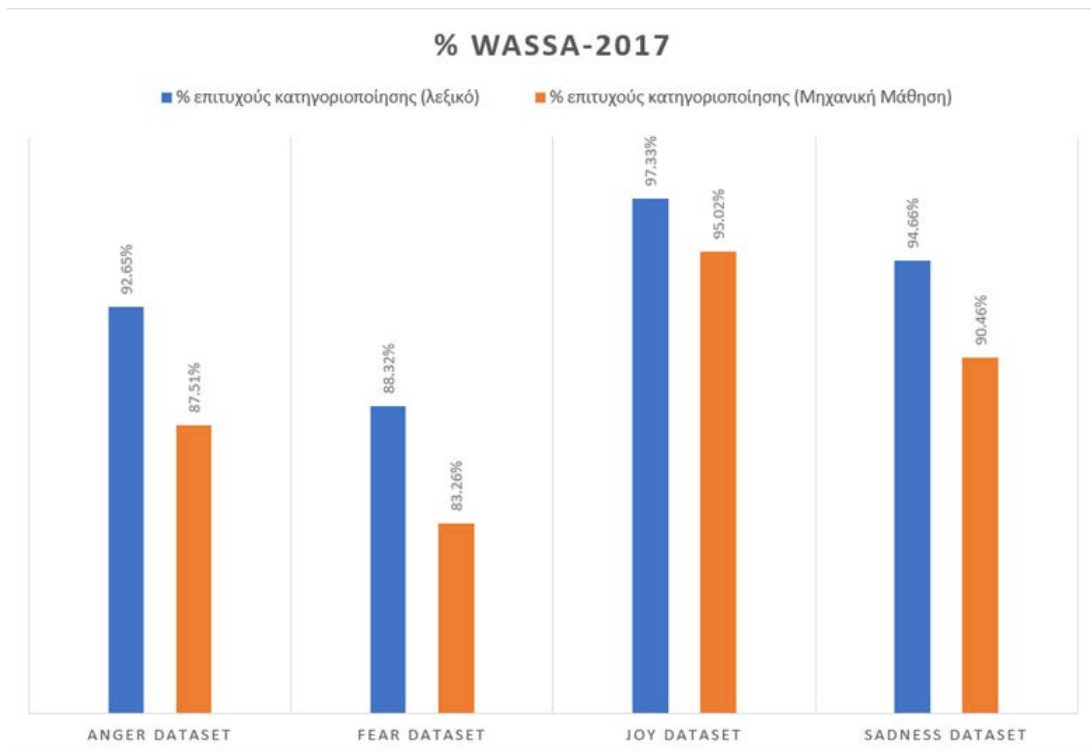
Ακολουθούν τα αποτελέσματα διαγραμματικά



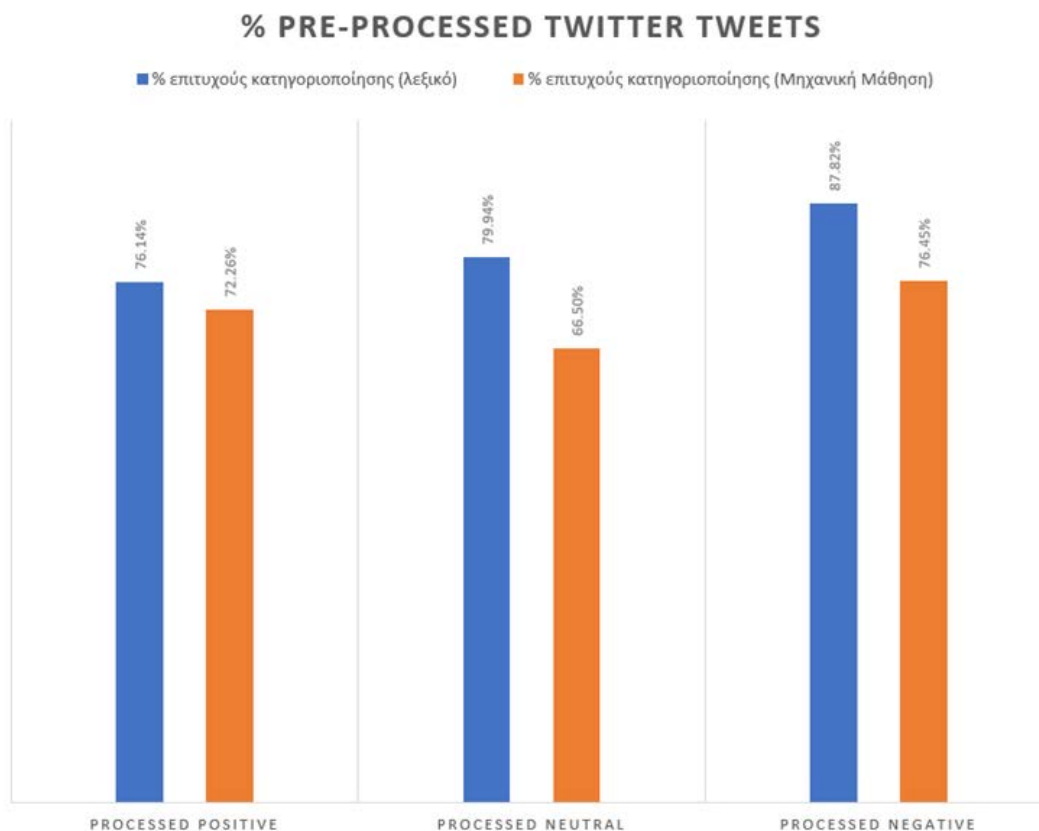
Διάγραμμα 3: αριθμητικά αποτελέσματα του συνόλου δεδομένων *WASSA-2017* για κάθε ένα από τα συναισθήματα, τόσο με βάση το Λεξικό δεδομένων όσο και με βάση τον αλγόριθμο Μηχανικής Μάθησης.



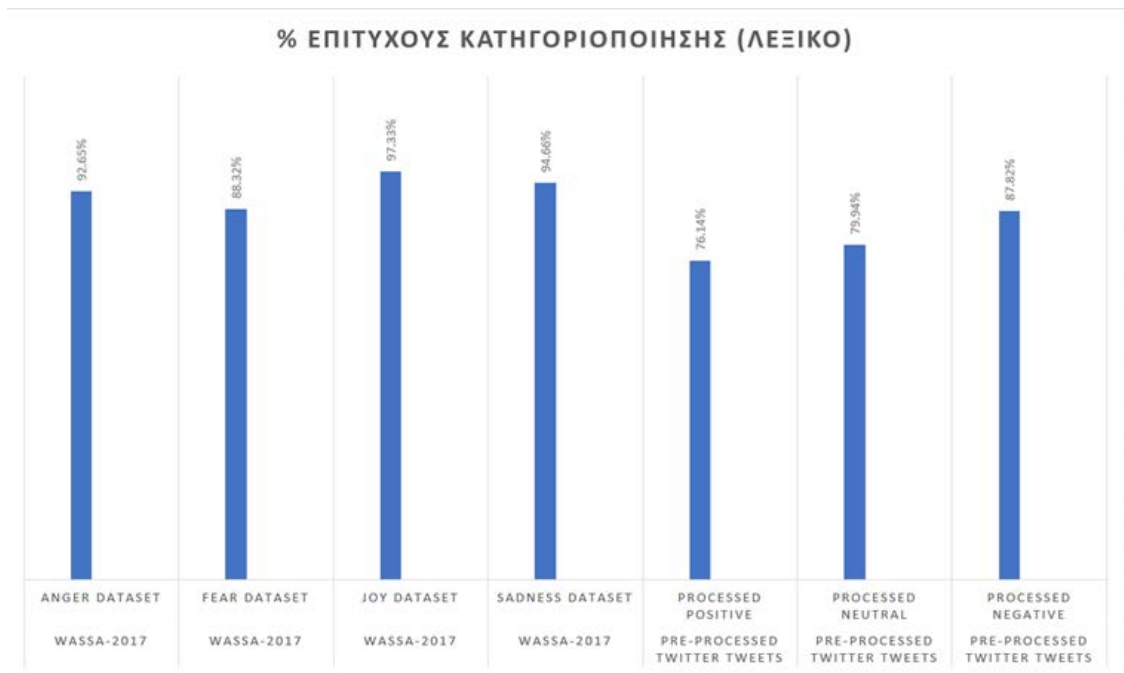
Διάγραμμα 4.: τα αριθμητικά αποτελέσματα του συνόλου δεδομένων *Pre-processed Twitter Tweets* για κάθε ένα από τα συναισθήματα, τόσο με βάση το Λεξικό δεδομένων όσο και με βάση τον αλγόριθμο Μηχανικής Μάθησης.



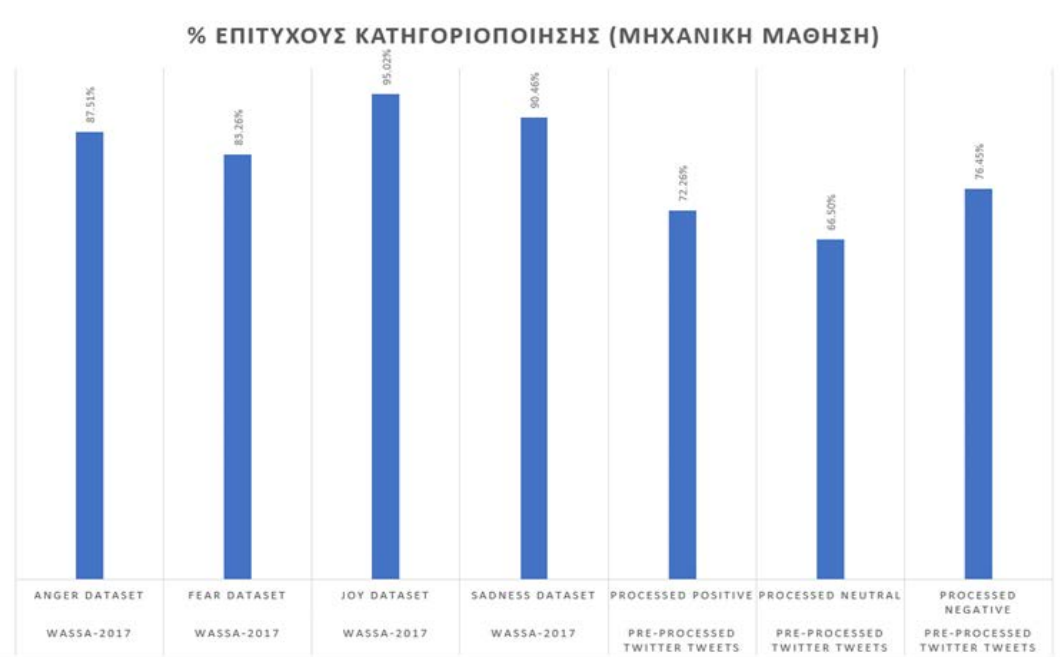
Διάγραμμα 5.: τα ποσοστιαία αποτελέσματα του συνόλου δεδομένων *WASSA-2017* για κάθε ένα από τα συναισθήματα, τόσο με βάση το Λεξικό δεδομένων όσο και με βάση τον αλγόριθμο Μηχανικής Μάθησης.



Διάγραμμα 6.: τα ποσοστιαία αποτελέσματα του συνόλου δεδομένων Pre-processed Twitter Tweets για κάθε ένα από τα συναισθήματα, τόσο με βάση το Λεξικό δεδομένων όσο και με βάση τον αλγόριθμο Μηχανικής Μάθησης.



Διάγραμμα 7.: ποσοστιαία αποτελέσματα για κατηγοριοποίηση των συναισθημάτων για το σύνολο δεδομένων *WASSA-2017* και για το σύνολο δεδομένων *Pre-processed Twitter Tweets* με βάση το Λεξικό δεδομένων.



Διάγραμμα 8.: τα ποσοστιαία αποτελέσματα για κατηγοριοποίησης των συναισθημάτων τόσο για το σύνολο δεδομένων *WASSA-2017* όσο και για το σύνολο δεδομένων *Pre-processed Twitter Tweets* με βάση τον αλγόριθμο Μηχανικής Μάθησης.

Κεφάλαιο 8

Συμπεράσματα

8.1 Εξαγωγή των συμπερασμάτων με βάση τα πειραματικά αποτελέσματα

Παρατηρήθηκε ότι η εξαγωγή συναισθημάτων με βάση το Λεξικό δεδομένων δίνει ελαφρώς καλύτερα αποτελέσματα από την εξαγωγή συναισθημάτων με βάση τον αλγόριθμο Μηχανικής Μάθησης. Δεδομένου του γεγονότος ότι όλα τα μηνύματα ήταν γραμμένα στην αγγλική γλώσσα και επίσης το ότι τόσο το Λεξικό Δεδομένων όσο και ο αλγόριθμος Μηχανικής Μάθησης της υπηρεσίας Dcipher Analytics, προορίζονται κυρίως για την εν λόγω γλώσσα, μπορούμε να κάνουμε την εν λόγω γενίκευση.

Για το σύνολο δεδομένων WASSA-2017 παρατηρήθηκε ότι, το ποσοστό επιτυχίας ήταν μεγαλύτερο στο συναίσθημα (“χαράς”) το οποίο ήταν θετικό, εν συγκρίσει με τα υπόλοιπα υποσύνολα δεδομένων τα οποία εκφράζουν αρνητικά συναισθήματα (“θυμό”, “φόβο”, “λύπη”).

Όσον αφορά το σύνολο δεδομένων *Pre-processed Twitter Tweets*, παρατηρήθηκε ότι το ποσοστό επιτυχίας για τα αρνητικά συναισθήματα ήταν μεγαλύτερο σε σχέση με τα υποσύνολα δεδομένων των θετικών και των ουδέτερων συναισθημάτων αντίστοιχα.

Με βάση τα πιο πάνω, δεν προέκυψε γενικευμένο συμπέρασμα που να αφορά το είδος και την πολικότητα των συναισθημάτων. Αυτό που μπορούμε να πούμε με ασφάλεια είναι ότι το ποσοστό επιτυχίας ήταν ελαφρώς μεγαλύτερο στο σύνολο δεδομένων WASSA-2017 από ότι στο σύνολο δεδομένων *Pre-processed Twitter Tweets*. Αυτό μπορεί να προκύπτει από το γεγονός του ότι τα μηνύματα του συνόλου δεδομένων WASSA-2017 είναι περισσότερο εκφραστικά και επίσης εμπεριέχουν emoticons. Όπως φαίνεται λοιπόν από τα πειραματικά αποτελέσματα, όσον αφορά την εξαγωγή συναισθημάτων από αδόμητο κείμενο, και ειδικότερα σε μικρού μεγέθους μηνύματα, τα emoticons παίζουν καθοριστικό ρόλο.

8.1. Σύγκριση των αποτελεσμάτων με παρόμοιες διατριβές

Στην εν λόγω διατριβή πραγματοποιήσαμε πειραματική έρευνα για την εξαγωγή συναισθημάτων πάνω σε δύο καλά ορισμένα σύνολα δεδομένων, κάνοντας χρήση κατάλληλου λογισμικού. Η ειδοποιός διαφορά της τρέχουσας έρευνας με παρόμοιες, για την ελληνική βιβλιογραφία, είναι η παρουσίαση των αριθμητικών πειραματικών αποτελεσμάτων και την εξαγωγή συμπερασμάτων από αυτά. Πιο συγκεκριμένα, σε σχέση με την έρευνα “Ανάλυση συναισθήματος από κείμενα, για οντότητες και χαρακτηριστικά αυτών” (Ανεστάκης Α., 2015), στην τρέχουσα έρευνα παρουσιάζονται τόσο διαγραμματικά όσο και σε μορφή πίνακα τα αποτελέσματα της ανάλυσης του λογισμικού που χρησιμοποιήθηκε (στην έρευνα που αναφέρθηκε παραπάνω δεν δίνεται κάτι παρόμοιο). Επίσης, όσον αφορά την έρευνα “Τεχνικές εξαγωγής και ανάλυσης συναισθημάτων σε μηνύματα του Twitter και υλοποίηση της ανάλυσης σε περιβάλλον cloud” (Μηταρά Ε., 2017), στο πειραματικό κομμάτι γίνεται μία περιγραφή της διαδικασίας της οποίας ακολουθήθηκε αλλά επίσης δεν δίνονται στο τέλος τα τελικά ευρήματα. Συνεπώς, η τρέχουσα διατριβή έρχεται να καλύψει το κενό στην ελληνική και κυπριακή βιβλιογραφία, όσον αφορά την παρουσίαση σαφών πειραματικών ευρημάτων.

8.3. Επίλογος και πιθανές προεκτάσεις

Στην τρέχουσα μελέτη επιτευχθήκαν οι πιο κάτω στόχοι:

Παρουσιάστηκαν οι βασικές έννοιες της Επεξεργασίας Φυσικής Γλώσσας και ειδικότερα της εξαγωγής συναισθημάτων από αδόμητο κείμενο. Οι εν λόγω μέθοδοι καλύπτουν ένα ευρύ φάσμα διάφορων τομέων της επιστήμης της Πληροφορικής αλλά και του κλάδου της Στατιστικής.

Παρουσιάστηκαν οι πιο διαδεδομένες μέθοδοι και εργαλεία εξαγωγής συναισθημάτων από αδόμητο κείμενο και η καταλληλότητα χρήσης τους για κάθε περίπτωση. Η σύγκριση παρουσιάστηκε σε κατάλληλο πίνακα.

Παρουσιάστηκε η μεθοδολογία της εκτέλεσης των πειραμάτων και έγινε η εκτέλεση για δύο σύνολα δεδομένων: το σύνολο δεδομένων WASSA-2017 και το σύνολο δεδομένων Pre-processed Twitter Tweets. Το κάθε ένα από τα σύνολα αυτά δεδομένων εμπεριείχε 4 και 3 σύνολα δεδομένων αντίστοιχα. Για την εκτέλεση των πειραμάτων έγινε χρήση του λογισμικού Dcipher Analytics. Πιο συγκεκριμένα, χρησιμοποιήθηκαν δύο διαφορετικές

μέθοδοι εξαγωγής συναισθημάτων από αδόμητο κείμενο: με χρήση κατάλληλου Λεξικού δεδομένων για την αγγλική γλώσσα και με χρήση αλγορίθμων Μηχανικής Μάθησης.

Εκτελέστηκαν τα πειράματα για κάθε ένα από τα σύνολα δεδομένων και παρουσιάσαμε αναλυτικά τον τρόπο με τον οποίο αυτά εκτελέστηκαν.

Τα αποτελέσματα των πειραμάτων παρουσιάστηκαν σε μορφή πινάκων και σε μορφή διαγραμμάτων αντίστοιχα.

Οι Πιθανές προεκτάσεις που θα μπορούσαν να ακολουθήσουν για την αποκόμιση περισσότερων στοιχείων είναι:

Η Χρήση περισσότερων συνόλων δεδομένων για την εξαγωγή συναισθημάτων και η μεταξύ τους σύγκριση. Η ανάλυση δεδομένων στοχευμένων θεμάτων/συζητήσεων τα οποία μπορεί να κινήσουν συγκεκριμένα συναισθήματα, ίσως να ήταν βοηθητικά ως προς την αποκόμιση περισσότερων συμπερασμάτων.

Η χρήση περισσότερων εργαλείων για την εξαγωγή συναισθημάτων από τα σύνολα δεδομένων και κυρίως εργαλεία και λεξικά που αφορούν τα emoticons.

Η συγγραφή κατάλληλου λογισμικού (μικρής έκτασης) το οποίο θα υλοποιεί μία ή περισσότερες μεθόδους εξαγωγής συναισθημάτων.

Παραρτήματα Α

Το σύνολο δεδομένων WASSA- 2017

Στο παράρτημα παρουσιάζεται ένα τμήμα από το σύνολο δεδομένων WASSA-2017. Πιο συγκεκριμένα, για κάθε ένα από τα υποσύνολα συναισθημάτων που το απαρτίζουν (θυμός, φόβος, χαρά και θλίψη) παρουσιάζονται μερικές από τις εγγραφές του, καθώς και το αποτέλεσμα της πειραματικής ανάλυσης η οποία έλαβε μέρος.

A.1 Απόσπασμα των πειραματικών αποτελεσμάτων για το συναίσθημα “θυμός”

...

```
{"id": {"_type": "Id", "value": [2]}, "value": "10002\t@DPD_UK I asked for my parcel to be delivered to a pick up store not my address #fuming #poorcustomerservice\tanger\t0.896", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.941780149936676, "label": "NEGATIVE"}},
```

```
{"id": {"_type": "Id", "value": [3]}, "value": "10003\tso ef whichever .... wipe pulled the fire alarm in davis bc I was sound asleep #pissed #angry #upset #tired #sad #tired #hangry #####\tanger\t0.896", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9999969005584717, "label": "NEGATIVE"}},
```

```
{"id": {"_type": "Id", "value": [4]}, "value": "10004\tDon't join @BTCare they put the phone down on you, talk over you and are rude. Taking money out of my acc willynilly! #fuming\tanger\t0.896", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.6253727674484253, "label": "NEGATIVE"}},
```

```
{"id": {"_type": "Id", "value": [5]}, "value": "10005\tMy blood is boiling\tanger\t0.875", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9989370703697205, "label": "POSITIVE"}},
```

{\"id\": {\"_type\": \"Id\", \"value\": [6]}, \"value\": \"10006\\tWhen you've still got a whole season of Wentworth to watch and a stupid cunt in work ruins it for us ġŸġŸ @_KirstyGA #raging #oldcunt\\tanger\\t0.875\", \"sentiment_lexicon\": {\"score\": -0.6666666666666666, \"label\": \"NEGATIVE\"}, \"sentiment_deep_learning\": {\"score\": -0.9969385862350464, \"label\": \"NEGATIVE\"}},

{\"id\": {\"_type\": \"Id\", \"value\": [7]}, \"value\": \"10007\\t@bt_uk why does tracking show my equipment delivered, when it wasn't? Why is my service suddenly delayed? We've already 3 weeks. #fuming\\tanger\\t0.875\", \"sentiment_lexicon\": {\"score\": -1.0, \"label\": \"NEGATIVE\"}, \"sentiment_deep_learning\": {\"score\": -0.7685023546218872, \"label\": \"NEGATIVE\"}},

{\"id\": {\"_type\": \"Id\", \"value\": [8]}, \"value\": \"10008\\t@TeamShanny legit why i am so furious with him, people are such ...ing idiots.\\tanger\\t0.875\", \"sentiment_lexicon\": {\"score\": -1.0, \"label\": \"NEGATIVE\"}, \"sentiment_deep_learning\": {\"score\": 0.9556384682655334, \"label\": \"POSITIVE\"}},

{\"id\": {\"_type\": \"Id\", \"value\": [9]}, \"value\": \"10009\\tHow is it suppose to work if you do that? Wtf dude? Thanks for pissing me off. \\tanger\\t0.875\", \"sentiment_lexicon\": {\"score\": 0.0, \"label\": \"NEUTRAL\"}, \"sentiment_deep_learning\": {\"score\": -0.7320806384086609, \"label\": \"NEGATIVE\"}},

{\"id\": {\"_type\": \"Id\", \"value\": [10]}, \"value\": \"10010\\tim so mad about power rangers. im incensed. im furious.\\tanger\\t0.667\", \"sentiment_lexicon\": {\"score\": -1.0, \"label\": \"NEGATIVE\"}, \"sentiment_deep_learning\": {\"score\": 0.7334132194519043, \"label\": \"POSITIVE\"}},

{\"id\": {\"_type\": \"Id\", \"value\": [11]}, \"value\": \"10011\\tWont use using @mothercareuk @Mothercarehelp again!! These guys cant get nothing right!! #fuming\\tanger\\t0.854\", \"sentiment_lexicon\": {\"score\": 0.0, \"label\": \"NEUTRAL\"}, \"sentiment_deep_learning\": {\"score\": -0.9838253855705261, \"label\": \"NEGATIVE\"}},

{\"id\": {\"_type\": \"Id\", \"value\": [12]}, \"value\": \"10012\\tBitches aggravate like what inspires you to be the biggest cunt known to man kind?\\tanger\\t0.854\", \"sentiment_lexicon\": {\"score\": -0.5, \"label\": \"NEGATIVE\"}, \"sentiment_deep_learning\": {\"score\": -0.9810991883277893, \"label\": \"NEGATIVE\"}},

{\"id\": {\"_type\": \"Id\", \"value\": [13]}, \"value\": \"10013\\tWhy does @dapperlaughs have to come to Glasgow on a night I am working. I am ...ing gutted, been waiting for an appearance for ages #raging\\tanger\\t0.938\", \"sentiment_lexicon\": {\"score\": -1.0, \"label\":

"NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9985195994377136, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [14]}, "value": "10014\t... fuming ġŸ~¤\tanger\t0.854", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.5084742307662964, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [15]}, "value": "10015\tZero help from @ups customer service. Just pushing the buck back and forth and promising callbacks that donâ€™t happen. #anger #loathing\tanger\t0.854", "sentiment_lexicon": {"score": -0.5, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9668256044387817, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [16]}, "value": "10016\t@ArizonaCoyotes not to mention the GRA guy stops me but let's the 2 ppl in front of me go. WTF. My blood is boiling.\tanger\t0.854", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9980657696723938, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [17]}, "value": "10017\tI hate my lawn mower. If it had a soul, I'd condemn it to the fiery pits of Hell.\tanger\t0.833", "sentiment_lexicon": {"score": -0.5, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.999334990978241, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [18]}, "value": "10018\twhy are people so offended by kendall he ends photo shoot like seriously shut the ... up\tanger\t0.833", "sentiment_lexicon": {"score": -0.3333333333333333, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9997352957725525, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [19]}, "value": "10019\tI'm about to block everyone everywhere posting about the storm. I think everyone is aware of the damn rain and what not so quit. #damn \tanger\t0.812", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9045416712760925, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [20]}, "value": "10020\tMaking my buddy cry !! Bitch wait for revenge ġŸ¤—ġŸ‘ŒġŸ»\tanger\t0.812", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9879708886146545, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [21]}, "value": "10021\tBes! You don't just tell a true blooded hoopjunkie to switch a f*c@n' team that juz destroyed your own team. You juz don't!

#insult\tanger\t0.891", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"},
"sentiment_deep_learning": {"score": -0.9993466734886169, "label": "NEGATIVE"}},
{"id": {"_type": "Id", "value": [22]}, "value": "10022\tI got a NO RESPONSE again from an
#irate customer of #amazon ğŸ˜ğŸ˜\nwhy can't they understand that it's not always
amazon fault.\n#carrierfault\tanger\t0.812", "sentiment_lexicon": {"score": -1.0,
"label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.99888676404953, "label":
"NEGATIVE"}},
{"id": {"_type": "Id", "value": [23]}, "value": "10023\tTasers immobilize, if you taser
someone why the ... do you need to shoot them one second later?! This is really sick! #rage
#wtf #murder\tanger\t0.812", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"},
"sentiment_deep_learning": {"score": -0.9109618067741394, "label": "NEGATIVE"}},
{"id": {"_type": "Id", "value": [24]}, "value": "10024\t@LeKingCoq @MistaAggy
@GiovanniiDC The way my blood is boiling, little bastards!\tanger\t0.792",
"sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning":
{"score": 0.9730685353279114, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [25]}, "value": "10025\tActually fuming I have nothing to
wear Saturday\tanger\t0.792", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"},
"sentiment_deep_learning": {"score": -0.9937258362770081, "label": "NEGATIVE"}},
{"id": {"_type": "Id", "value": [26]}, "value": "10026\t.@DIVAmagazine than straight
people. Even the arse straight guys who think that means a threesome is fine. #angry
#fuming\tanger\t0.792", "sentiment_lexicon": {"score": -0.5, "label": "NEGATIVE"},
"sentiment_deep_learning": {"score": -0.9971417784690857, "label": "NEGATIVE"}},
{"id": {"_type": "Id", "value": [27]}, "value": "10027\tBloody parking ticket ğŸ˜ğŸ˜,
#fuming\tanger\t0.792", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"},
"sentiment_deep_learning": {"score": -0.973110556602478, "label": "NEGATIVE"}},
{"id": {"_type": "Id", "value": [28]}, "value": "10028\tYou ever just find that the people
around you really irritate you sometimes? That's me right now ğŸ˜\tanger\t0.792",
"sentiment_lexicon": {"score": -0.3333333333333333, "label": "NEGATIVE"},
"sentiment_deep_learning": {"score": -0.9986163377761841, "label": "NEGATIVE"}},
{"id": {"_type": "Id", "value": [29]}, "value": "10029\tI get so angry at people that don't
know that you don't have a stop sign on Francis and you do at Foster #road
\tanger\t0.792", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"},
"sentiment_deep_learning": {"score": -0.9977673292160034, "label": "NEGATIVE"}},
...

A.2 Απόσπασμα των πειραματικών αποτελεσμάτων για το συναίσθημα “φόβος”

...

```
{ "id": { "_type": "Id", "value": [0] }, "value": "20000\tI feel like I am drowning. #depression #anxiety #failure #worthless\tfear\t0.979", "sentiment_lexicon": { "score": -0.6666666666666666, "label": "NEGATIVE" }, "sentiment_deep_learning": { "score": -0.9565407633781433, "label": "NEGATIVE" } },  
{ "id": { "_type": "Id", "value": [1] }, "value": "20001\tI get so nervous even thinking about talking to ***** I wanna die\tfear\t0.979", "sentiment_lexicon": { "score": -1.0, "label": "NEGATIVE" }, "sentiment_deep_learning": { "score": -0.9388575553894043, "label": "NEGATIVE" } },  
{ "id": { "_type": "Id", "value": [2] }, "value": "20002\tI lost my blinders ... #panic\tfear\t0.975", "sentiment_lexicon": { "score": -1.0, "label": "NEGATIVE" }, "sentiment_deep_learning": { "score": -0.9999347925186157, "label": "NEGATIVE" } },  
{ "id": { "_type": "Id", "value": [3] }, "value": "20003\tI feel like I am drowning. #depression #failure #worthless\tfear\t0.938", "sentiment_lexicon": { "score": -0.6, "label": "NEGATIVE" }, "sentiment_deep_learning": { "score": -0.8115243315696716, "label": "NEGATIVE" } },  
{ "id": { "_type": "Id", "value": [4] }, "value": "20004\tThis is the scariest American Horror Story out of all of them... I'm gonna have to watch in the daytime. #frightened\tfear\t0.938", "sentiment_lexicon": { "score": -1.0, "label": "NEGATIVE" }, "sentiment_deep_learning": { "score": 0.9992350339889526, "label": "POSITIVE" } },  
{ "id": { "_type": "Id", "value": [5] }, "value": "20005\t@mgcsartwork I nearly started crying and having a full on panic attack after tatinof bc of the crowds so I feel him\tfear\t0.938", "sentiment_lexicon": { "score": -1.0, "label": "NEGATIVE" }, "sentiment_deep_learning": { "score": -0.8317030072212219, "label": "NEGATIVE" } },  
{ "id": { "_type": "Id", "value": [6] }, "value": "20006\tI have to finally tell my therapist about my sexuality ... last frontier ... not sure I can do it in the AM #fear #SingleGirlProblems\tfear\t0.938", "sentiment_lexicon": { "score": -1.0, "label": "NEGATIVE" }, "sentiment_deep_learning": { "score": -0.5385147333145142, "label": "NEGATIVE" } },
```

{"id": {"_type": "Id", "value": [7]}, "value": "20007\t@laura221b I don't think I've ever moved so fast in a panic in all my life ğŸ˜, Gave me such a fright ğŸ˜~\tfear\t0.938", "sentiment_lexicon": {"score": -0.5, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.8331692814826965, "label": "NEGATIVE"}},
{"id": {"_type": "Id", "value": [8]}, "value": "20008\tMy bus was in a car crash... I'm still shaking a bit... This week was an absolute horror and this was the icing on the cake... #terrible\tfear\t0.938", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.999535083770752, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [9]}, "value": "20009\tMy bus was in a car crash... I'm still shaking a bit... This week was an absolute horror and this was the icing on the cake... \tfear\t0.920", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9963985681533813, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [10]}, "value": "20010\tprocrastinating is fun until im an hour away from the time its due and i still havent finished so i have a panic attack\tfear\t0.920", "sentiment_lexicon": {"score": -0.5, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9842342734336853, "label": "NEGATIVE"}},
{"id": {"_type": "Id", "value": [11]}, "value": "20011\tMy #anxiety is rising tonight and I'm not sure why. Sometimes I wonder if I'm a magnet for any free-floating anxiety in the universe.\tfear\t0.917", "sentiment_lexicon": {"score": -0.2, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.8403345942497253, "label": "NEGATIVE"}},
{"id": {"_type": "Id", "value": [12]}, "value": "20012\tIm not a #nervouswreck, Im a #nervous #pileup. #GoingCrazy #HelpMe #Insane #Antisocial\tfear\t0.917", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9980965256690979, "label": "NEGATIVE"}},
{"id": {"_type": "Id", "value": [13]}, "value": "20013\t#panic Panic attack from fear of starting new medication\tfear\t0.917", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9978972673416138, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [14]}, "value": "20014\tJob interview in the afternoon #nervous #ek\tfear\t0.917", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9126363396644592, "label": "NEGATIVE"}},
{"id": {"_type": "Id", "value": [15]}, "value": "20015\tI am beyond mad that I lost track of a brown spider in my brown carpet. Where did you go? ğŸ• will be sneaking up on me

\tfear\t0.917", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"},
 "sentiment_deep_learning": {"score": -0.9994009733200073, "label": "NEGATIVE"}},
 {"id": {"_type": "Id", "value": [16]}, "value": "20016\tanxiety level- 100\tfear\t0.905",
 "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning":
 {"score": 0.9275128245353699, "label": "POSITIVE"}},
 {"id": {"_type": "Id", "value": [17]}, "value": "20017\tWah just woke up frm a ...
 nightmare\tfear\t0.898", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"},
 "sentiment_deep_learning": {"score": 0.994804322719574, "label": "POSITIVE"}},
 {"id": {"_type": "Id", "value": [18]}, "value": "20018\tI get so much PUSSY\nP - panic
 attacks \nU - uncontrollable anxiety \nS - suicidal fantasies \nS - sadness \nY -
 yearning for death\tfear\t0.896", "sentiment_lexicon": {"score": -1.0, "label":
 "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.6638769507408142, "label":
 "NEGATIVE"}},
 {"id": {"_type": "Id", "value": [19]}, "value": "20019\tCosplaying properly for the first time
 on Saturday! Pretty nervous..\tfear\t0.896", "sentiment_lexicon": {"score": -
 0.3333333333333333, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -
 0.7983490824699402, "label": "NEGATIVE"}},
 {"id": {"_type": "Id", "value": [20]}, "value": "20020\t@Markgatiss I'm surrounded by
 those Trump voters. You're right, it is ... terrifying. #redstate #despair\tfear\t0.896",
 "sentiment_lexicon": {"score": -0.5, "label": "NEGATIVE"}, "sentiment_deep_learning":
 {"score": 0.9847176671028137, "label": "POSITIVE"}},
 {"id": {"_type": "Id", "value": [21]}, "value": "20021\tI ignored broken tooth for so long, now
 have abscess. Need dentist but #fear makes it hard for me to go..45 and still can't go to
 dentist\tfear\t0.896", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"},
 "sentiment_deep_learning": {"score": -0.9682455062866211, "label": "NEGATIVE"}},
 {"id": {"_type": "Id", "value": [22]}, "value": "20022\t@RogueCoder250 We are in so much
 trouble!! I don't think the Rev will see the funny side of our project.
 #nervous\tfear\t0.896", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"},
 "sentiment_deep_learning": {"score": 0.5703338980674744, "label": "POSITIVE"}},
 {"id": {"_type": "Id", "value": [23]}, "value": "20023\tim crying katherine is the only one
 whos like talking to me during my anxiety attack im gonna faint\tfear\t0.896",
 "sentiment_lexicon": {"score": -0.6, "label": "NEGATIVE"}, "sentiment_deep_learning":
 {"score": 0.8445157408714294, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [24]}, "value": "20024\tEver been really lonely and your phone keeps blowing up, but you just canâ€™t pick it up and respond to people? #anxiety #recluse #issues\tfear\t0.893", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.8450671434402466, "label": "NEGATIVE"}},
 {"id": {"_type": "Id", "value": [25]}, "value": "20025\tSuper shitting it about this tattoo #nervous\tfear\t0.880", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9699088931083679, "label": "POSITIVE"}},
 {"id": {"_type": "Id", "value": [26]}, "value": "20026\tBreaking out in hives for the first time since college finals. #anxiety sucks\tfear\t0.880", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9760040044784546, "label": "POSITIVE"}},
 {"id": {"_type": "Id", "value": [27]}, "value": "20027\t@tchop_StL @YouTube \nI get sick to my stomach everytime I see this video & for the helicopter crew to make such comments is revolting.\tfear\t0.875", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9805352091789246, "label": "POSITIVE"}},
 {"id": {"_type": "Id", "value": [28]}, "value": "20028\tI am beyond mad that I lost track of a brown spider in my brown carpet. Where did you go? ğŸ• will be sneaking up on me #frightened\tfear\t0.875", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9997333884239197, "label": "NEGATIVE"}},
 {"id": {"_type": "Id", "value": [29]}, "value": "20029\t6. 5 minutes to perform and I'm #nervous I wish these noises would stop #handicapped #annoyed\tfear\t0.875", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9992594122886658, "label": "NEGATIVE"}},
 ...

A.3 Απόσπασμα των πειραματικών αποτελεσμάτων για το συναίσθημα “χαρά”

[{"id": {"_type": "Id", "value": [0]}, "value": "30000\tJust got back from seeing @GaryDelaney in Burslem. AMAZING!! Face still hurts from laughing so much #hilarious\tjoy\t0.980", "sentiment_lexicon": {"score": 0.3333333333333333, "label":

"POSITIVE"}, "sentiment_deep_learning": {"score": -0.9576343894004822, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [1]}, "value": "30001\tOh dear an evening of absolute hilarity I don't think I have laughed so much in a long time! ġŸ~\tjoy\t0.958", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": -0.845896303653717, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [2]}, "value": "30002\tBeen waiting all week for this game âœŸ,âœŸ,âœŸ, #cheer #friday âœŸ,\tjoy\t0.940", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9856151342391968, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [3]}, "value": "30003\t@gardiner_love : Thank you so much, Gloria! You're so sweet, and thoughtful! You just made my day more joyful! I love you too! ġŸ~ŠġŸ•\tjoy\t0.938", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9999992847442627, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [4]}, "value": "30004\tI feel so blessed to work with the family that I nanny for âœŸ, nothing but love & appreciation, makes me smile.\tjoy\t0.938", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.6479146480560303, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [5]}, "value": "30005\tToday I reached 1000 subscribers on YT!! , #goodday, #thankful\tjoy\t0.926", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": -0.8193522095680237, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [6]}, "value": "30006\t@Singaholic121 Good morning, love! Happy first day of fall. Let's make some awesome #autumnmemories #annabailey #laughter #smile\tjoy\t0.924", "sentiment_lexicon": {"score": 0.6, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9546520709991455, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [7]}, "value": "30007\t#BridgetJonesBaby is the best thing I've seen in ages! So funny, I've missed Bridget! #love #TeamMark\tjoy\t0.922", "sentiment_lexicon": {"score": 0.2, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9958315491676331, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [8]}, "value": "30008\tJust got back from seeing @GaryDelaney in Burslem. AMAZING!! Face still hurts from laughing so much \tjoy\t0.920", "sentiment_lexicon": {"score": 0.0, "label": "NEUTRAL"}, "sentiment_deep_learning": {"score": -0.7665643095970154, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [9]}, "value": "30009\t@IndyMN I thought the holidays could not get any more cheerful, and then I met you. #TheNiceBot\tjoy\t0.917", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.5765566825866699, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [10]}, "value": "30010\tI'm just still . So happy .\nA blast\tjoy\t0.917", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9980286955833435, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [11]}, "value": "30011\tIt's meant to be!! #happy #happy\tjoy\t0.917", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.7039275765419006, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [12]}, "value": "30012\tğŸŸ¥âš-ï, Yeahâ€¼ï, PAULâ€¼ï,âš-ïğŸŸ¥ #glorious #BB18\tjoy\t0.917", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.879400372505188, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [13]}, "value": "30013\tMy morning started off amazing!! Hopefully the whole day is going as i want it to go!\n #GreatDay \tjoy\t0.917", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.7977091670036316, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [14]}, "value": "30014\tğŸŸ± @cailamarsai you've had me ðŸŸ~, ðŸŸ~, the whole time watching @black_ishABC after you've lost your #glasses! It was #hilarious! @mrbabyboogaloo\tjoy\t0.900", "sentiment_lexicon": {"score": 0.3333333333333333, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": -0.9875819683074951, "label": "NEGATIVE"}},
{"id": {"_type": "Id", "value": [15]}, "value": "30015\t@iamTinaDatta love you so much #smile ðŸŸŠğŸŸŠ\tjoy\t0.896", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9999886751174927, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [16]}, "value": "30016\t@WyoWiseGuy @LivingVertical however, REI did offer me the job today as well. Can't believe how exponentially freaking joyous I feel...!!!\tjoy\t0.896", "sentiment_lexicon": {"score": 0.5, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.998536229133606, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [17]}, "value": "30017\t2 days until #GoPackGo and 23 days until #GoGipeGo..... I'm so excited! #smiling\tjoy\t0.880", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": -0.7795673608779907, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [18]}, "value": "30018\t@TheMandyMoore You are beyond wonderful. Your singing prowess is phenomenal but damn... I'm just elated to watch you act again. #ThisIsUs ĩā—\tjoy\t0.879", "sentiment_lexicon": {"score": 0.6666666666666666, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.999903678894043, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [19]}, "value": "30019\t@luckiiCHARM_ Luckii, I'm changing in so many ways bc of Him!! It's a scary but joyful feeling, making me so strong.\tjoy\t0.877", "sentiment_lexicon": {"score": 0.5, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9892100691795349, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [20]}, "value": "30020\t@JoshNoneYaBiz I love parody accounts! Well done. Vote for #Trump. #lol #hilarious\tjoy\t0.875", "sentiment_lexicon": {"score": 0.5, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9630861878395081, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [21]}, "value": "30021\tWhen you wake up from a dream laughing at something stupid, and that makes you laugh more #hilarious\tjoy\t0.875", "sentiment_lexicon": {"score": 0.3333333333333333, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9999508857727051, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [22]}, "value": "30022\tknow that I have my future planned out, I feel so much happier #goals #life #happy #igotthis #yay\tjoy\t0.875", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": -0.9147905111312866, "label": "NEGATIVE"}},
{"id": {"_type": "Id", "value": [23]}, "value": "30023\tOnline now !!!:) all day come play with me !! I'm happy horny playful sweet sour;) \tjoy\t0.875", "sentiment_lexicon": {"score": 0.6666666666666666, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9998621940612793, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [24]}, "value": "30024\t@grahnort wonderful experience watching you yesterday at. @BBCLetItShine thankyou for the #laughter\tjoy\t0.872", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9999996423721313, "label": "POSITIVE"}},
{"id": {"_type": "Id", "value": [25]}, "value": "30025\t@itsyourgirl_Z happy birthday :) have a blessed day love from Toronto :) #bday \tjoy\t0.868", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.999920129776001, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [26]}, "value": "30026\t@WSJNordics You make the world a more joyful place. #TheNiceBot\tjoy\t0.864", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9987419247627258, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [27]}, "value": "30027\tMorning all! Of course it is sunny on this Monday morning to cheerfully welcome us back to work.)\tjoy\t0.863", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9517779350280762, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [28]}, "value": "30028\t@Singaholic121 Good morning, love! Happy first day of fall. Let's make some awesome #autumnmemories #annabailey\tjoy\t0.860", "sentiment_lexicon": {"score": 0.5, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9550067186355591, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [29]}, "value": "30029\tIt's the #FirstDayofFall and I'm so happy. Sipping my #PumpkinSpice flavored coffee and #smiling! Happy Fall everyone! #amwriting\tjoy\t0.860", "sentiment_lexicon": {"score": 1.0, "label": "POSITIVE"}, "sentiment_deep_learning": {"score": 0.9998301267623901, "label": "POSITIVE"}},

...

A.4 Απόσπασμα των πειραματικών αποτελεσμάτων για το συναίσθημα “θλίψη”

...

[{"id": {"_type": "Id", "value": [0]}, "value": "40000\tDepression sucks! #depression\tsadness\t0.958", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9941197633743286, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [1]}, "value": "40001\tFeeling worthless as always #depression\tsadness\t0.958", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.8982486128807068, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [2]}, "value": "40002\tFeeling worthless as always\tsadness\t0.958", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.7649041414260864, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [3]}, "value": "40003\tMy #Fibromyalgia has been really bad lately which is not good for my mental state. I feel very overwhelmed #anxiety #bipolar #depression\tsadness\t0.946", "sentiment_lexicon": {"score": -0.6666666666666666,

"label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9994200468063354, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [4]}, "value": "40004\tIm think ima lay in bed all day and sulk. Life is hitting me to hard rn\tsadness\t0.934", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9945613741874695, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [5]}, "value": "40005\tSo when I try I fail... and when I don't try... I still fail... #confused and #lost\tsadness\t0.917", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.733900785446167, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [6]}, "value": "40006\tExtreme sadness\tsadness\t0.917", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.994453489780426, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [7]}, "value": "40007\tmy life in one word is depressing\tsadness\t0.917", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9718669056892395, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [8]}, "value": "40008\tPanic attacks are the worst. Feeling really sick and still shaking. I should be a sleep. #anxiety #depression\tsadness\t0.917", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.6416663527488708, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [9]}, "value": "40009\tFeel so grim + ugly atm\tsadness\t0.896", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9803246855735779, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [10]}, "value": "40010\tHonestly depression is kicking my ... lately.. ğŸ˜™\tsadness\t0.896", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9416096210479736, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [11]}, "value": "40011\tAt my age all I see is gray. Is it gray because of my bad eyes or my perspective #depression #healingjustice\tsadness\t0.896", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.7945331335067749, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [12]}, "value": "40012\tAfter 3 idk why I start feeling so depress, sad and lonely.\tsadness\t0.896", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.7945331335067749, "label": "NEGATIVE"}},

"NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9998663663864136, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [13]}, "value": "40013\tI sulk too much for my own good.\tsadness\t0.896", "sentiment_lexicon": {"score": 0.0, "label": "NEUTRAL"}, "sentiment_deep_learning": {"score": -0.9973055124282837, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [14]}, "value": "40014\tA night where depression is winning... #depression #fml #help\tsadness\t0.896", "sentiment_lexicon": {"score": -0.5, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9729069471359253, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [15]}, "value": "40015\tGoing home is depressing\tsadness\t0.896", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9997145533561707, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [16]}, "value": "40016\tWe can't even grieve over one black body before another one pops up due to pig brutality\tsadness\t0.892", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9955358505249023, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [17]}, "value": "40017\tIt's a gloomy ... day\tsadness\t0.888", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9976094961166382, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [18]}, "value": "40018\t@Fra93_bruno wow I'm just really sadden by that. Terrible\tsadness\t0.875", "sentiment_lexicon": {"score": 0.0, "label": "NEUTRAL"}, "sentiment_deep_learning": {"score": 0.892791211605072, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [19]}, "value": "40019\tMy heads still in Ibiza but my body is sat at me desk at work #depressing\tsadness\t0.875", "sentiment_lexicon": {"score": -0.3333333333333333, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9964646100997925, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [20]}, "value": "40020\t@Ashleyavitia_ don't leave me #sad\tsadness\t0.875", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.8903270363807678, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [21]}, "value": "40021\tDepression sucks! \tsadness\t0.875", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9984307885169983, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [22]}, "value": "40022\tIt's been a week of awful connectivity with @TMobile no service or only 4G is not what Im paying for. #unhappy #poorservice\tsadness\t0.870", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9994074106216431, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [23]}, "value": "40023\tThis the most depressing shit ever\tsadness\t0.861", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.9654092788696289, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [24]}, "value": "40024\tSometimes The Worst Place You Can Be Is In Your Own Head.\n\n#quotes #worstenemy #depression #thinktoomuch\tsadness\t0.854", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9997512698173523, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [25]}, "value": "40025\t@TMobileHelp my service is down in Cleveland, OH. What's going on? #unhappy\tsadness\t0.854", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9968351721763611, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [26]}, "value": "40026\tWhat a sad evening - clearing out all of Harvey's cage and belongings. Now so final. Goodbye my little man... #depressing\tsadness\t0.854", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": 0.8977293372154236, "label": "POSITIVE"}},

{"id": {"_type": "Id", "value": [27]}, "value": "40027\tWhen you have 15 doe run the opposite side of you ğŸ™ #depression\tsadness\t0.854", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.5929955840110779, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [28]}, "value": "40028\tThe moment of the day when you have to start to plaster a smile in your face. #depression\tsadness\t0.854", "sentiment_lexicon": {"score": -0.3333333333333333, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9565337896347046, "label": "NEGATIVE"}},

{"id": {"_type": "Id", "value": [29]}, "value": "40029\tI hate when it's gloomy outside because it always gets me in a depressing mood\tsadness\t0.839", "sentiment_lexicon": {"score": -1.0, "label": "NEGATIVE"}, "sentiment_deep_learning": {"score": -0.9978570342063904, "label": "NEGATIVE"}},

...

Βιβλιογραφία

- [1] Nepal, S., C. Paris, D. Georgeakopoulos, Social media for government services Springer, 2015.
- [2] Wang, R., D. Zhou, M. Jiang, J. Si and Y. Yang, “A survey on opinion mining: From stance to product aspect” IEEE Access vol 7, pp. 41101-41124, 2019.
- [3] Turney, P. D., Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th annual meeting on association for computational linguistics , pp. 417-424, 2002.
- [4] Ανεστάκης Α., Ανάλυση συναισθήματος από κείμενα, για οντότητες και χαρακτηριστικά αυτών, 2015.
- [5] Μηταρά Ε., Τεχνικές εξαγωγής και ανάλυσης συναισθημάτων σε μηνύματα του Twitter και υλοποίηση της ανάλυσης σε περιβάλλον cloud, 2017.
- [6] Μώρος Ν., Ανάλυση συναισθήματος σε διαδικτυακές κριτικές υπηρεσιών φιλοξενίας, 2020.
- [7] Medhat, W., Hassan, A. & Korashy, H., Sentiment analysis algorithms and applications: A survey, pp. 1096-1097, 2014.
- [8] Llobert, O. R., Using Machine Learning Techniques for Sentiment Analysis, 1, 2017.
- [9] Musto, C., Semeraro, G. & Polignano, M., A comparison of Lexicon-based approaches for Sentiment Analysis of microblog posts, 2015.
- [10] M. Milonas, Swarms, phase transitions, and collective intelligence, In Langton CG Ed., Artificial Life III, Addison Wesley, Reading, MA, 1994.
- [11] OMG! Oxford English Dictionary grows a heart: Graphic symbol for love (and that exclamation) are added as words.tinyurl.com/klv36p.
- [12] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. Journal of Language and Social Psychology, 29(1):24–54, 2010.
- [13] Esuli and Sebastiani. Sentwordnet: A publicly available lexical resource for opinion mining. In International Conference on Language Resources and Evaluation (LREC), pages 417–422, 2006.
- [14] G. A. Miller. Wordnet: a lexical database for english. Communications of the ACM, 38(11):39–41, 1995.

- [15] E. Cambria, R. Speer, C. Havasi, and A. Hussain. Senticnet: A publicly available semantic resource for opinion mining. In AAI Fall Symposium Series, 2010.
- [16] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan. A system for real-time twitter sentiment analysis of 2012 u.s. presidential election cycle. In ACL System Demonstrations, pages 115–120, 2012.
- [17] P. S. Dodds and C. M. Danforth. Measuring the happiness of large-scale written expression: songs, blogs, and presidents. *Journal of Happiness Studies*, 11(4):441–456, 2009.
- [18] P. Gonçalves, F. Benevenuto, and M. Cha. PANAS-t: A Psychometric Scale for Measuring Sentiments on Twitter. abs/1308.1857v1, 2013.
- [19] D. Watson and L. Clark. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of Personality and Social Psychology*, 54(1):1063–1070, 1985.
- [20] Fabio Calefato, Filippo Lanubile, Nicole Novielli. EmoTxt: A Toolkit for Emotion Recognition from Text, 2017, <https://arxiv.org/pdf/1708.03892.pdf>
- [21] <https://www.dcipheranalytics.com/>
- [22] <http://saifmohammad.com/WebPages/EmotionIntensity-SharedTask.html>
- [23] <https://www.kaggle.com/shashank1558/preprocessed-twitter-tweets>