ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ
Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών
στα Πληροφοριακά Συστήματα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Ο Ρόλος της Επικοινωνίας στα Μέσα Κοινωνικής Δικτύωσης και Ανάλυση Συναισθήματος: Η Περίπτωση της Χρηματοοικονομικής Πληροφορίας – The Impact of Communication in Social Media and Sentiment Analysis: The Case of Financial Information»

Βουλγαράκη Μίνα

Επιβλέπουσα καθηγήτρια: Αικατερίνη Κύρτσου

Θεσσαλονίκη, 2018

Η παρούσα διπλωματική εργασία, η οποία εκπονήθηκε στο πλαίσιο απόκτησης μεταπτυχιακού διπλώματος στα Πληροφορικά Συστήματα (MIS) και τα λοιπά αποτελέσματα της, αποτελούν συνιδιοκτησία του Πανεπιστημίου Μακεδονία και του φοιτητή, ο καθένας από τους οποίους έχει το δικαίωμα ανεξάρτητης χρήσης και αναπαραγωγής τους (στο σύνολο ή τμηματικά) για διδακτικούς και ερευνητικούς σκοπούς, σε κάθε περίπτωση αναφέροντας τον τίτλο, το συγγραφέα και το Πανεπιστήμιο Μακεδονίας όπου εκπονήθηκε η ΔΕ, καθώς και τον επιβλέποντα και την επιτροπή κρίσης.

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω θερμά την επιβλέπουσα καθηγήτρια μου, κ. Κύρτσου Αικατερίνη για την πολύτιμη καθοδήγηση της και την εμπιστοσύνη που μου έδειξε καθ' όλη τη διάρκεια της εκπόνησης της διπλωματικής μου εργασίας.

Επίσης, θα ήθελα να ευχαριστήσω την οικογένεια μου για τη συμπαράσταση τους και την ηθική τους υποστήριξη σε όλη τη διάρκεια των σπουδών μου, καθώς και το σύντροφο μου Ανδρέα, για την αμέριστη βοήθεια του σε κάθε μου προσπάθεια.

Περίληψη

Στην παρούσα διπλωματική εργασία, θα ερευνήσουμε αν και κατά πόσο το κλίμα που δημιουργείται στα μέσα κοινωνικής δικτύωσης αναφορικά με τις τράπεζες και τις ασφαλιστικές / επενδυτικές εταιρείες, είτε αυτό είναι θετικό, είτε αρνητικό, αποτυπώνεται στη συμπεριφορά των πελατών τους, υφιστάμενων ή εν δυνάμει, προς αυτές.

Πάνω σε αυτό, θα προσπαθήσουμε να κατηγοριοποιήσουμε σχόλια σε μορφή κειμένου (tweets). Τα δεδομένα θα αντληθούν από το κοινωνικό δίκτυο Twitter και θα αφορούν σε σχόλια των οικονομικών μονάδων είτε αυτές είναι πολίτες, είτε επενδυτικοί ή πολιτικοί φορείς, σχετικά με τον τραπεζικό κλάδο. Έπειτα από επεξεργασία θα γίνει κατηγοριοποίηση τους σε θετικά, ουδέτερα και αρνητικά, με στόχο να ελέγξουμε την αποτελεσματικότητα των αλγόριθμων μηχανικής μάθησης Naïve Bayes και Deep Learning, προκειμένου να δούμε πως οι οικονομικοί οργανισμοί μπορούν να λάβουν την πληροφόρηση που χρειάζονται για το κλίμα που επικρατεί στην αγορά. Επιπλέον, για να αποτυπώσουμε την επίδραση των διογκωμένων πληροφοριών στη λιανική τραπεζική, θα προσπαθήσουμε να επιβεβαιώσουμε τις θεωρίες της ψυχολογίας των μαζών και της οικονομικής ετερογένειας. Για το λόγο αυτό θα πραγματοποιηθεί έρευνα με ερωτηματολόγιο μέσω των social media.

Λέξεις Κλειδιά: «Twitter, tweets, τράπεζες, ανάλυση κειμένου, μηχανική μάθηση, οικονομική ψυχολογία, συναισθηματική ανάλυση»

Abstract

In this dissertation, we will investigate whether the climate that is created in social media for banks and insurance / investment companies, whether it's positive or negative, is reflected in the behavior of their clients, existing or potential, to them.

On that, we will try to categorize text format comments (tweets). Our data will be taken from Twitter and they will be comments about the banking industry and by economic units, whether they are citizens or investment or political actors. After processing, they will be categorized into positive, neutral and negative, in order to examine the effectiveness of Naive Bayes and Deep Learning algorithms, in order to see how financial organizations can have the information they need for the prevailing climate of the markets. In addition, to capture the impact of inflated information on retail banking, we will try to confirm the theories of crowd psychology and economic heterogeneity. For this reason a questionnaire will take place through social media.

Keywords: «Twitter, tweets, banks, text analysis, mechanical learning, economic psychology, emotional analysis »

Περιεχόμενα

Εισαγωγή **7**
Κεφάλαιο 1: Ανάλυση Συναισθήματος **9**
1.1. Περιγραφή και Κατηγορίες Ανάλυσης Συναισθήματος 9
1.2. Εφαρμογές Ανάλυσης Συναισθήματος 9
1.3. Προσεγγίσεις Ανάλυσης Συναισθήματος 11
1.3.1. Ανάλυση Βασισμένη σε Λεξικό Προσέγγισης 11
1.3.2. Μηχανική Μάθηση 13
1.4. Εισαγωγή στην έννοια των Big Data 15
1.4.1. Χαρακτηριστικά των Big Data 16
1.4.2. Η σχέση μεταξύ Big Data και Internet of Thing (IoT) 17
1.4.3. Μέθοδοι επεξεργασίας Big Data 17
1.5. Ανάλυση κειμένου 18
1.5.1. Επεξεργασία φυσικής γλώσσας (NLP) 18
1.5.2. Εξόρυξη, Κατηγοριοποίηση, Ανάκτηση κειμένου – Εξαγωγή πληροφοριών 19
1.5.3. Ανάλυση συναισθήματος με μηχανική μάθηση 23
1.5.3.1. Ο αλγόριθμος Naive Bayes 23
1.5.3.2.Τεχνητά νευρωνικά δίκτυα 24
1.5.3.3. Deep Learning 25
1.5.4. Το Κοινωνικό Δίκτυο Twitter 26
1.5.5. Εφαρμογή sentiment analysis σε tweets με NLP tools 27
1.6. Εμπειρική Έρευνα – Συλλογή tweets και ανάλυση συναισθήματος με τη χρήση του RapidMiner και του AYLIEN 28
1.6.1. Συλλογή tweets 28
1.6.2. Ανάλυση συναισθήματος με τη χρήση των RapidMiner και AYLIEN 31
Κεφάλαιο 2: Συμπεριφορική Χρηματοοικονομική **51**
2.1. Γενικά περί Συμπεριφορικής Χρηματοοικονομικής 51
2.2. Ψυχολογία και κανόνες λήψης αποφάσεων 51
2.3 Οικονομική ετερογένεια και είδη επενδυτών 52
2.4 Φήμες και χρηματοοικονομική συμπεριφορά 54
2.4.1. Θεωρία της ψυχολογίας των μαζών 55
2.5. Εμπειρική Έρευνα – Ανάλυση ερωτηματολογίου 56
2.5.1. Μεθοδολογία έρευνας 56
2.5.2. Επεξεργασία δεδομένων και ανάλυση αποτελεσμάτων 58
Συμπεράσματα **73**
Βιβλιογραφία – Αναφορές **75**

Εισαγωγή

Η ανθρώπινη διάθεση ως προς τα διάφορα ζητήματα ενδιαφέροντος αποτελεί ιδιαίτερα σημαντική πληροφορία, τόσο ως προς τα κοινωνικά, όσο και ως προς τα οικονομικοπολιτικά ζητήματα.

Οι παραδοσιακές μέθοδοι εκτίμησης της άποψης και των συναισθημάτων, όπως π.χ. οι δημοσκοπήσεις, είναι αρκετά χρονοβόρες και χωρίς άμεσα αποτελέσματα. Έτσι, προκύπτει η ανάγκη για εύρεση πιο άμεσων λύσεων στη διαχείριση των πληροφοριών και στην εξαγωγή συμπερασμάτων από αυτές. Η λύση δίδεται από τα κοινωνικά δίκτυα, η αυξημένη χρήση των οποίων επιτρέπει την εξόρυξη και ανάκτηση πληροφοριών και καθιστά τον τομέα της αυτόματης εξαγωγής συναισθήματος έναν αναπτυσσόμενο και πολύ σημαντικό κλάδο.

Οι χρήστες του διαδικτύου έχουν πάψει πλέον να είναι παθητικοί και χρησιμοποιούν όλο και περισσότερο τα κοινωνικά δίκτυα όχι μόνο για να επικοινωνήσουν, αλλά και για να ανταλλάξουν απόψεις, να αξιολογήσουν προϊόντα και υπηρεσίες, ή απλά να πουν τη γνώμη τους για θέματα της επικαιρότητας ή προσωπικά τους ζητήματα. Ο όγκος των πληροφοριών που προκύπτει από τις αναρτήσεις στα κοινωνικά δίκτυα είναι τεράστιος και δίνει μεγάλες δυνατότητες άντλησης και επεξεργασίας.

Στη σύγχρονη κοινωνία η απόκτηση γνώσης μπορεί να δώσει σημαντικό πλεονέκτημα σε όποιον την αξιολογήσει και την αξιοποιήσει σωστά, κάνοντας τις πληροφορίες που προκύπτουν από την ανάλυση συναισθήματος όλο και πιο χρήσιμες.

Η συνεχής τεχνολογική εξέλιξη στον τομέα της πληροφορίας, δε θα μπορούσε να αφήσει ανεπηρέαστο τον κλάδο των τραπεζών. Η στροφή των καταναλωτών προς τις νέες τεχνολογίες και η αυξημένη χρήση των κοινωνικών δικτύων, ωθεί τις τράπεζες στην αναζήτηση τρόπων για την ευκολότερη άντληση πληροφοριών και την πιο άμεση αξιολόγηση τους, ώστε να μπορέσουν να ανταποκριθούν στις ανάγκες των πελατών τους και να γίνουν πιο ανταγωνιστικές. Οι τράπεζες στην προσπάθεια τους αυτή, αναζητούν καινοτόμες ιδέες και δίνουν ιδιαίτερη έμφαση στην άριστη εξυπηρέτηση του πελάτη, εκμεταλλευόμενες την πρόοδο της τεχνολογίας και δημιουργώντας νέες ευκαιρίες για την απόκτηση ανταγωνιστικού πλεονεκτήματος.

Στην περίοδο της οικονομικής κρίσης, όλοι οι οικονομικοί οργανισμοί και περισσότερο οι τράπεζες, καταβάλλουν προσπάθειες για την οικονομική τους

αναδιάρθρωση, λαμβάνοντας μέτρα όπως για παράδειγμα το πάγωμα των προσλήψεων, η μείωση του προσωπικού μέσω εθελούσιων αποχωρήσεων και η συρρίκνωση του δικτύου των καταστημάτων τους. Το γεγονός αυτό σε συνάρτηση με τις δυσκολίες που αντιμετωπίζουν οι καταναλωτές στην αποπληρωμή των υποχρεώσεων τους, τόσο προς τις τράπεζες, όσο και γενικότερα, δημιουργεί πολλές φορές ένα αρνητικό κλίμα προς αυτές.

Η ταύτιση των τραπεζών με την οικονομική κρίση των τελευταίων ετών, καθώς και η πολυπλοκότητα των σχέσεων που χαρακτηρίζουν την ανθρώπινη συμπεριφορά και τις ανθρώπινες αντιδράσεις απέναντι στα οικονομικά και πολιτικά ερεθίσματα που δέχονται οι οικονομικές μονάδες, δημιουργούν στις τράπεζες την ανάγκη να αποκρυπτογραφήσουν και να αξιολογήσουν τη διάθεση, τόσο των πελατών, όσο και των επενδυτών.

Στην παρούσα διπλωματική εργασία, θα ερευνήσουμε αν και κατά πόσο το κλίμα που δημιουργείται στα κοινωνικά δίκτυα αναφορικά με τις τράπεζες, είτε αυτό είναι θετικό, είτε αρνητικό, αποτυπώνεται στη συμπεριφορά των πελατών τους, υφιστάμενων ή εν δυνάμει, προς αυτές.

Στο πρώτο κεφάλαιο θα γίνει γενική αναφορά στην ανάλυση συναισθήματος, στα big data, στην ανάλυση κειμένου και το κοινωνικό δίκτυο twitter. Στο δεύτερο κεφάλαιο θα παρουσιάσουμε την έρευνα μας πάνω στις τυχόν μεταβολές στην επενδυτική συμπεριφορά των οικονομικών μονάδων και στο αν αυτή επηρεάζεται από τα σχόλια τρίτων στα κοινωνικά δίκτυα.

Κάθε κεφάλαιο θα περιέχει το κομμάτι της έρευνας που το αφορά. Αρχικά, στο πρώτο κεφάλαιο, για αντιληφθούμε τον τρόπο που λειτουργεί η ανάλυση συναισθήματος, θα γίνει συλλογή σχολίων σχετικών με τις τράπεζες που δραστηριοποιούνται στην Ελλάδα και θα κατηγοριοποιηθούν βάσει του συναισθήματος που εκφράζουν. Για το σκοπό αυτό θα χρησιμοποιηθεί το λογισμικό RapidMiner και τα εργαλεία AYLIEN API και Twitter Archiver. Έπειτα, στο δεύτερο κεφάλαιο, θα δημιουργηθεί ένα ερωτηματολόγιο και θα αναρτηθεί σε διάφορα group στο κοινωνικό δίκτυο Facebook ή θα σταλεί με email προς απάντηση. Στη συνέχεια, αφού συλλεχθεί ένας ικανός αριθμός συμπληρωμένων ερωτηματολογίων, θα αναλυθεί προκειμένου να διερευνήσουμε το ενδεχόμενο οι οικονομικές μονάδες να επηρεάζονται ως προς τις επενδυτικές τους αποφάσεις, από σχόλια τρίτων στο διαδίκτυο. Η εργασία θα ολοκληρωθεί με την παράθεση των συμπερασμάτων που θα προκύψουν.

Κεφάλαιο 1: Ανάλυση Συναισθήματος

1.1. Περιγραφή και Κατηγορίες Ανάλυσης Συναισθήματος

Η ανάλυση συναισθήματος είναι η διαδικασία με την οποία εξάγονται πληροφορίες σχετικά με τη συναισθηματική κατάσταση ενός χρήστη, μέσω του γραπτού λόγου του και αποτελεί υποκατηγορία της ταξινόμησης κειμένου. Για την ταξινόμηση του κειμένου σε συναισθηματικές κλάσεις χρησιμοποιούνται στατιστικές μέθοδοι, μέθοδοι μηχανικής μάθησης και τεχνικές επεξεργασίας φυσικού λόγου (NPL) (Liu, 2012).

Βασικός στόχος της εξόρυξης γνώμης είναι να κατηγοριοποιήσει την πολικότητα της πάνω στην έκφραση της για ένα θέμα, προσδιορίζοντας τις θετικές και αρνητικές εκφράσεις στα σχόλια των χρηστών του διαδικτύου, κάνοντας πιο αξιόπιστες τις αξιολογήσεις τους για προϊόντα και υπηρεσίες. Αρχικά, η ανάλυση συναισθήματος διαχωρίζεται βάσει της κατάστασης την οποία προσδιορίζει. Μπορεί δηλαδή να προσδιορίζει είτε τη συναισθηματική κατάσταση του συγγραφέα τη δεδομένη χρονική στιγμή, είτε αυτή που θέλει σκοπίμως να μεταδώσει στους αναγνώστες, είτε απλά την έκφραση της άποψης του πάνω σε ένα θέμα. Σε κάθε περίπτωση, η ταξινόμηση των συναισθημάτων που εκφράζονται στο κείμενο μπορεί να γίνει κατηγοριοποιώντας τα ως χαρά, λύπη, θυμό ή γενικότερα όπως θετικά, αρνητικά και ουδέτερα. Οι γενικές κατηγορίες μπορούν να γίνουν και πέντε, προσθέτοντας τις μάλλον θετικά και μάλλον αρνητικά. Η αναζήτηση του συναισθήματος σε ένα κείμενο μπορεί να αφορά σε ολόκληρο το κείμενο, κάποιες προτάσεις που περιέχει ή ακόμη και μερικές μόνο φράσεις, όταν αυτές αναφέρονται στο θέμα για το οποίο αναζητούμε το συναίσθημα.

1.2. Εφαρμογές Ανάλυσης Συναισθήματος

Πρόκειται για σχετικά καινούργιο τομέα έρευνας, ο οποίος αφορά στην ταξινόμηση κειμένων σε κατηγορίες ανάλογα με το συναίσθημα που εκφράζουν. Ενώ έως και τον προηγούμενο αιώνα είχε ερευνηθεί επαρκώς η ταξινόμηση των κειμένων με βάση το θέμα τους (topic classification), στις αρχές του 21^{ου} αιώνα οι έρευνες στράφηκαν στην ταξινόμηση τους ως προς το συναίσθημα. Έτσι, χρησιμοποιήθηκαν

στατιστικές μέθοδοι και αλγόριθμοι μηχανικής μάθησης προκειμένου να κατηγοριοποιηθούν κείμενα.

Στις μέρες μας, το ερευνητικό αυτό πεδίο προσελκύει πολύ μεγάλο ενδιαφέρον για δύο βασικούς λόγους. Ο πρώτος είναι η αύξηση της χρήσης του διαδικτύου και άρα η πρόσβαση σε άπειρα δεδομένα. Το ίντερνετ πλέον έχει γίνει μέρος της καθημερινότητας των ανθρώπων και χρησιμοποιείται για οποιαδήποτε πληροφορία ή επικοινωνία. Σε αυτό έχει συμβάλει κατά πολύ και η εξέλιξη των κινητών τηλεφώνων. Με την έλευση των smartphones, δισεκατομμύρια άνθρωποι ανά τον κόσμο απέκτησαν άμεση πρόσβαση στο διαδίκτυο και άρα σε μεγάλους όγκους δεδομένων (big data) όπως κείμενα, εικόνες, βίντεο κλπ. Τα κοινωνικά δίκτυα, δίνουν επίσης τη δυνατότητα στους χρήστες να επικοινωνήσουν με άλλους ή απλά να εκφραστούν. Έτσι, οι ερευνητές μπορούν να διαχειριστούν όλα αυτά τα δεδομένα για να εξάγουν συμπεράσματα σχετικά με τη συναισθηματική κατάσταση των χρηστών. Ο δεύτερος λόγος που το πεδίο της ανάλυσης συναισθήματος προσελκύει μεγάλο ενδιαφέρον, είναι η τεχνολογική πρόοδος και άρα η αύξηση των πόρων που χρειάζονται για τη μέτρηση του συναισθήματος. Η εξέλιξη της τεχνολογίας δίνει τις απαραίτητες δυνατότητες επεξεργαστικής ισχύος που χρειάζονται για την εφαρμογή της μηχανικής μάθησης και της ανάλυσης συναισθήματος, καθώς σήμερα όλοι έχουν τη δυνατότητα να τρέξουν τους απαιτητικούς αλγόριθμους, ακόμη και στους προσωπικούς τους υπολογιστές (Turney, 2002 & Pang, et al., 2002).

Ως αποτέλεσμα των όσων αναφέραμε παραπάνω, έχει αναζωπυρωθεί το ενδιαφέρον γύρω από τη μηχανική μάθηση και την τεχνητή νοημοσύνη, κάνοντας πιο αποτελεσματική την ανάλυση συναισθήματος. Έτσι, όλο και περισσότερες εταιρείες και οργανισμοί χρησιμοποιούν αντίστοιχους αλγόριθμους σε δεδομένα που εξάγουν από το διαδίκτυο αναφορικά με τα προϊόντα ή τις υπηρεσίες τους. Με βάση τα αποτελέσματα που προκύπτουν, μπορούν να βγουν συμπεράσματα για την αποδοχή ή όχι των προϊόντων τους από τους καταναλωτές. Ένα άλλο πεδίο εφαρμογής της συναισθηματικής ανάλυσης είναι ο εκδημοκρατισμός των κοινωνικών δικτύων, δηλαδή η εξόρυξη πληροφοριών σχετικά με τη θέση της κοινής γνώμης σε οικονομικοπολιτικά ή κοινωνικά ζητήματα (Pang & Lee, 2008).

1.3. Προσεγγίσεις Ανάλυσης Συναισθήματος

Η συναισθηματική ανάλυση μπορεί να προσεγγισθεί κυρίως με δύο τρόπους, τη βασισμένη σε λεξικό ανάλυση (lexicon–based) και τη μηχανική μάθηση (Esuliand & Sebastiani, 2006).

1.3.1. Ανάλυση Βασισμένη σε Λεξικό Προσέγγισης

Στον lexicon – based τρόπο χρησιμοποιούνται συναισθηματικά λεξικά για να βαθμολογηθούν κείμενα ως προς τα συναισθήματα που εκφράζουν. Για την υλοποίηση αυτής της μεθόδου προστίθεται συναισθηματική βαθμολογία στις επιμέρους λέξεις ή φράσεις του κειμένου, οι οποίες έχουν σημειωθεί ως προς το συναισθηματικό τους περιεχόμενο και έχουν κατοχυρωθεί σε ένα σχετικό λεξικό, το οποίο χαρακτηρίζεται ως word level. Σε αυτού του είδους τα λεξικά, οι λέξεις ή φράσεις συνοδεύονται από αντίστοιχες βαθμολογίες βάσει των οποίων συμπεραίνουμε κατά πόσο ταιριάζουν σε μία από τις δύο κατηγορίες συναισθήματος: θετική ή αρνητική. Κάθε βαθμολογία έχει το αντίστοιχο πρόσημο για να εκφράσει την κλίμακα της κατηγορίας. Υπάρχουν ωστόσο και ορισμένα λεξικά που χρησιμοποιούν πιο συγκεκριμένες κατηγορίες όπως π.χ. χαρά, λύπη, θυμός κλπ.

Στην περίπτωση που επιλεχθεί η μέθοδος του λεξικού ώστε να κατηγοριοποιηθεί ένα κείμενο, ακολουθείται η εξής διαδικασία: Χρησιμοποιώντας έναν αλγόριθμο (tokenization), αναλύουμε το κείμενο σε λέξεις ή φράσεις και έπειτα αναζητούμε κάθε token στο συναισθηματικό λεξικό και σημειώνουμε τη βαθμολογία του. Τέλος, αθροίζουμε τις επιμέρους βαθμολογίες για να προκύψει ο χαρακτηρισμός του συνολικού κειμένου, το οποίο ταξινομείται ως θετικό ή αρνητικό βάσει του πρόσημου του αθροίσματος.

Η μέθοδος αυτή είναι πολύ απλή, χρειάζεται χαμηλή υπολογιστική ισχύ για την εξαγωγή αποτελεσμάτων και δεν προαπαιτεί κανένα κείμενο εκπαίδευσης, γεγονός το οποίο γλυτώνει χρόνο και ανθρώπινη εργασία. Ωστόσο, η απόδοση της στην ταξινόμηση του κειμένου περιορίζεται από αδυναμίες. Οι κύριοι περιορισμοί που αντιμετωπίζει είναι οι εξής (Taboada et al., 2011):

Άρνηση. Η διαδικασία που περιγράψαμε αγνοεί τις λέξεις που εκφράζουν άρνηση όπως π.χ. δεν, όχι, ποτέ, κανένας, τίποτα κ.α., οι οποίες όταν βρεθούν σε μία

πρόταση μπορούν να αλλάξουν τελείως το νόημα της προηγούμενης ή επόμενη λέξης. Για παράδειγμα, ενώ η πρόταση «το φαγητό ήταν καλό» εκφράζει θετική αξιολόγηση, η πρόταση «το φαγητό δεν ήταν καλό», εκφράζει αρνητική αξιολόγηση. Η πιο συνηθισμένη επιλογή αντιμετώπισης είναι να αντιστρέφουμε την πόλωση των λέξεων που ακολουθούν την άρνηση μέχρι το επόμενο σημείο στίξης ή τον επόμενο αντιθετικό σύνδεσμο όπως π.χ. αλλά, παρόλα αυτά κ.λ.π. Και πάλι όμως προκύπτουν αδυναμίες. Αν για παράδειγμα, υποθέσουμε ότι η λέξη καλό βαθμολογείται με +3 και η φράση «πολύ καλό» με +5, τότε η άρνηση «δεν ήταν καλό» θα έχει μεγαλύτερη βαθμολογία από την άρνηση «δεν ήταν πολύ καλό» κάτι το οποίο δε συνάδει με την αντίληψη μας. Για ορθότερα αποτελέσματα οι Taboada et al, πρότειναν την ολίσθηση της βαθμολογίας της επόμενης από την άρνηση λέξης κατά μία σταθερά προς την αντίθετη κατεύθυνση για να προκύψει η βαθμολογία της άρνησης.

Μεταβολή έντασης. Εκτός από την άρνηση που αναλύσαμε παραπάνω, υπάρχουν και άλλες λέξεις οι οποίες μεταβάλλουν την ένταση των επόμενων ή προηγούμενων λέξεων. Υπάρχουν για παράδειγμα λέξεις όπως πραγματικά, περισσότερο, αληθινά, που αυξάνουν την ένταση της επόμενης λέξης και λέξεις όπως λιγότερο, χαμηλότερο που μειώνουν την ένταση. Για την αντιμετώπιση της αδυναμίας αυτής, ερευνητές όπως οι Kennedy & Inkpen (2006) και οι Polanyi & Zaenen (2006), πρότειναν μια απλή πρόσθεση ή αφαίρεση. Αν δηλαδή βρεθεί στο κείμενο μία λέξη μεταβολής έντασης, η βαθμολογία της επόμενης λέξης θα αυξηθεί ή θα μειωθεί ανάλογα κατά μία σταθερή ποσότητα. Παρόλα αυτά, το πρόβλημα δεν αντιμετωπίζεται πλήρως καθώς όλες οι λέξεις έντασης δεν έχουν την ίδια βαρύτητα. Παραδείγματος χάριν, η λέξη εξαιρετικά έχει πολύ μεγαλύτερη ένταση από τη λέξη περισσότερο. Σε αυτή την περίπτωση οι Taboada et al, πρότειναν να αντιστοιχίζεται ένα ποσοστό σε κάθε λέξη έντασης, το οποίο θα προστίθεται ή θα αφαιρείται ανάλογα από το 100% και το ποσοστό που θα προκύπτει θα πολλαπλασιάζεται με τη βαθμολογία της επόμενης λέξης.

Σειρά των λέξεων. Μία άλλη αδυναμία της lexicon – based προσέγγισης, είναι ότι δε λαμβάνει υπόψη τη σειρά των λέξεων μέσα στο κείμενο, η οποία μπορεί να αλλάξει τελείως το νόημα μίας πρότασης και το συναίσθημα που προκύπτει. Για παράδειγμα αν πούμε «είναι αλήθεια, δεν μου άρεσε το φαγητό» και «δεν είναι αλήθεια, μου άρεσε το φαγητό» προκύπτει τελείως αντίθετο συναίσθημα παρότι οι λέξεις και των δύο προτάσεων είναι ακριβώς οι ίδιες. Για να αντιμετωπιστεί αυτή η

αδυναμία, θα πρέπει όπως εξηγήσαμε παραπάνω, να επισημάνουμε εκτός από την άρνηση και την εμβέλεια της.

Αντιθετικοί σύνδεσμοι. Θα πρέπει να δίνεται ιδιαίτερη σημασία σε λέξεις όπως αλλά, παρότι, ωστόσο, παρόλα αυτά, διότι συνδέουν δύο προτάσεις με διαφορετικό συναίσθημα. Συνήθως το βασικό συναίσθημα βρίσκεται μετά τον αντιθετικό σύνδεσμο. Π.χ. η φράση «το φαγητό ήταν καλό αλλά πολύ ακριβό» δείχνει μάλλον αρνητικό συναίσθημα, ενώ η φράση «το φαγητό ήταν ακριβό αλλά πολύ καλό» δείχνει μάλλον θετικό συναίσθημα.

Ιδιωματισμοί και έννοιες. Καθώς η μέθοδος lexicon – based, εξετάζει την κάθε λέξη ξεχωριστά, δε λαμβάνει υπόψη φράσεις οι οποίες δρουν σαν λέξη και εκφράζουν συναισθήματα συνολικά. Όπως για παράδειγμα η φράση «μια φορά στα χίλια χρόνια» η οποία σημαίνει πολύ σπάνια. Αν τη χρησιμοποιήσουμε σε μία πρόταση όπως π.χ. «τέτοιο φαγητό τρως μία φορά στα χίλια χρόνια», θα της δώσει ένα ιδιαίτερα θετικό νόημα. Άρα χρειάζεται να χρησιμοποιηθεί ένα λεξικό ιδιωματισμών και να αναλυθεί το κείμενο σε επίπεδο φράσεων και όχι λέξεων. Επίσης, υπάρχουν λέξεις και φράσεις των οποίων το νόημα διαφέρει ανάλογα με την περίπτωση στην οποία χρησιμοποιούνται και είναι ίσως η μεγαλύτερη δυσκολία που καλείται να αντιμετωπίσει η συναισθηματική ανάλυση. Σύμφωνα με τους Poria et al. (2015), το κείμενο αναλύεται σε προτάσεις και κάθε πρόταση αναλύεται σε έννοιες, οι οποίες αναζητούνται σε ειδικά σχεδιασμένα λεξικά.

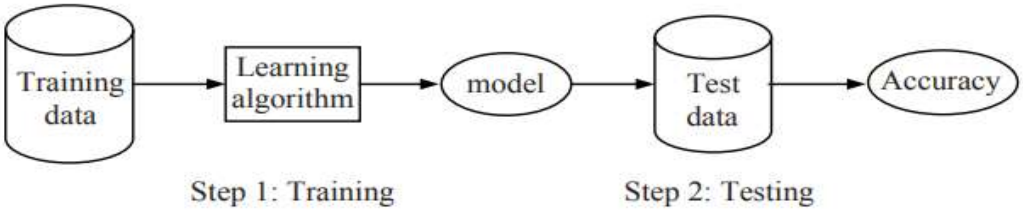
Ειρωνεία. Είναι πολύ δύσκολο ακόμη και για τον άνθρωπο να διακρίνει τα όρια μεταξύ χιούμορ και σαρκασμού, πόσο μάλλον για μια μηχανή. Με ειρωνική διάθεση συνήθως εκφράζονται αρνητικά σχόλια. Για παράδειγμα η πρόταση «σε αυτό το εστιατόριο σίγουρα θα ζήσετε μια ξεχωριστή εμπειρία» μπορεί να εκφράζει ένα πολύ θετικό συναίσθημα ή ένα αρνητικό που εκφράζεται ειρωνικά. Έχουν αναπτυχθεί ειδικές μέθοδοι για τη ταξινόμηση μιας πρότασης σε ειρωνική ή μη.

1.3.2. Μηχανική Μάθηση

Ως μηχανική μάθηση εννοούμε τον κλάδο της πληροφορικής που δημιουργεί και αναλύει αλγόριθμους και τεχνικές ώστε οι υπολογιστές να εξελίσσονται μέσα από την εμπειρία τους. Κατά τον Arthur Samuel (1959), ως μηχανική μάθηση ορίζεται «*το πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί για αυτό το σκοπό*» (Simon, 2013). Βασικό της χαρακτηριστικό

είναι ότι μοντελοποιεί δεδομένα ώστε να τα χρησιμοποιήσει για τη δημιουργία προβλέψεων πάνω σε αυτά και εφαρμόζεται σε εργασίες υπολογισμού όπου δεν υπάρχει δυνατότητα να προγραμματιστούν συγκεκριμένοι αλγόριθμοι. Ενώ αρχικά η έρευνα γύρω από τη μηχανική μάθηση επικεντρώθηκε στην επίλυση προβλημάτων και σε συμβολικές μεθόδους, στις μέρες μας έχει εξελιχθεί ώστε να μπορεί να χρησιμοποιηθεί σε κάθε κλάδο όπως της υγείας, του εμπορίου, της παραγωγής, του αθλητισμού κ.α.

Όπως αποτυπώνεται και στην εικόνα 1, για την «εκπαίδευση» ενός συστήματος, του δίνονται δεδομένα εκπαίδευσης και χρησιμοποιώντας έναν αλγόριθμο μηχανικής μάθησης, το σύστημα εκπαιδεύεται πάνω στα δεδομένα αυτά. Κατά τη διαδικασία αυτή, αποκτάει εμπειρία από την οποία μαθαίνει και δημιουργεί μοντέλα δεδομένων. Έπειτα του δίνονται δεδομένα δοκιμής, που δεν έχει συναντήσει ξανά και από τις προβλέψεις που θα κάνει πάνω σε αυτά, θα αξιολογηθεί η απόδοση του συστήματος (Liu, 2011).



Εικόνα 1: Η βασική διαδικασία μάθησης: εκπαίδευση και αξιολόγηση (Bing, Liu., "Web data mining – exploring hyperlinks, contents and usage data, 2011, p.66)

Ανάλογα με το αν τα δεδομένα που εισάγονται στο σύστημα κατά την εκπαίδευση του, συνοδεύονται με κάποια επιθυμητή απόκριση, αν δηλαδή το κατευθύνουν κατά κάποιο τρόπο, η μηχανική μάθηση χωρίζεται στις εξής κατηγορίες (Bishop, 2006):

Επιτηρούμενη μάθηση. Το υπολογιστικό πρόγραμμα δέχεται δεδομένα εισόδου, καθώς και τα επιθυμητά αποτελέσματα και στόχος του είναι να μάθει ένα γενικό κανόνα, προκειμένου να προβλέπει τα αποτελέσματα σε νέα δεδομένα εισόδου.

Μη επιτηρούμενη μάθηση. Το υπολογιστικό πρόγραμμα, χωρίς να του δίνονται κάποια επιθυμητά αποτελέσματα στον αλγόριθμο μάθησης, πρέπει να βρει τη δομή πίσω από τα δεδομένα εισόδου.

Ενισχυτική μάθηση. Το υπολογιστικό πρόγραμμα αλληλεπιδρά με ένα δυναμικό περιβάλλον στο οποίο πρέπει να επιτευχθεί ένας συγκεκριμένος στόχος, χωρίς να

ενημερώνεται αν έχει φτάσει κοντά στο στόχο του. Χαρακτηριστικά παραδείγματα ενισχυτικής μηχανικής μάθησης είναι τα αυτοκινούμενα οχήματα και τα παιχνίδια με αντιπάλους.

1.4. Εισαγωγή στην έννοια των Big Data

Τα big data είναι μία αφηρημένη έννοια, η οποία χαρακτηρίζει σύνολα δεδομένων τα οποία δεν είναι δυνατόν να αποθηκευτούν και να διαχειριστούν από τις παραδοσιακές βάσεις δεδομένων όπως π.χ. η SQL. Σύμφωνα με τους Mayer - Schonberger & Cukier (2013) ως big data ορίζονται «*μεγάλα σύνολα δεδομένων και πληροφοριών απ' όπου μπορούν να εξαχθούν χρήσιμα συμπεράσματα στη βάση χαρακτηριστικών όπως ο μεγάλος όγκος, η ποικιλία, η ταχύτητα και η αξία τους*».

Με την εξέλιξη της τεχνολογίας και την αποτελεσματικότερη αποθήκευση δεδομένων, οι βάσεις δεδομένων κατέληξαν να μη συμβαδίζουν με τις ανάγκες που ολοένα και μεγάλωναν. Το πρόβλημα έγινε ιδιαιτέρως ορατό με την εκλαΐκευση της ψηφιακής τεχνολογίας κατά τη δεκαετία του 1990. Την τελευταία δεκαετία δε, η διαχείριση των big data απασχολεί όλο και περισσότερο, σχεδόν όλες τις μεγάλες εταιρείες, καθώς ο όγκος των δεδομένων αναμένεται να αυξάνεται κατά περίπου 40% ανά έτος μέχρι το 2020 (Khan et al., 2014). Επιπλέον, έχει παρατηρηθεί ότι το 90% των δεδομένων που υπάρχουν σήμερα, έχει δημιουργηθεί από το 2014 και έπειτα. Σύμφωνα με την έρευνα «Predicts 2013: Information Innovation» της Gartner, το 42% των ερωτηθέντων είχαν ήδη επενδύσει σε τεχνολογίες διαχείρισης big data ή τουλάχιστον σκόπευαν να το πράξουν μέσα στο 2013[1].

Οι άνθρωποι, μέσω της τεχνολογίας που χρησιμοποιούν, παράγουν αχανείς ποσότητες ψηφιακών δεδομένων. Έτσι, οι ανάγκες για περισσότερο χώρο αποθήκευσης έχει αυξηθεί κατά πολύ τα τελευταία χρόνια, όπως επίσης και οι απαιτήσεις των χρηστών. Το γεγονός αυτό κάνει τις υπάρχουσες μορφές επεξεργασίας δεδομένων να μοιάζουν απαρχαιωμένες, αφού δε μπορούν να ανταπεξέλθουν στις ανάγκες των χρηστών και έτσι όλες οι επιλογές παρακολούθησης και επεξεργασίας δεδομένων εστιάζουν κυρίως στα big data.

Η αποτελεσματική διαχείριση όλου αυτού του όγκου των πληροφοριών που πλέον βρίσκεται στη διάθεση όλων, αποτελεί τη μεγάλη τεχνολογική πρόκληση της

[1] Business-analytics.gr, «Big data: Η επιτυχία περνά μέσα από τα analytics»

εποχής και είναι ένα δύσκολο υπολογιστικά έργο, το οποίο όμως μπορεί να δώσει πολύτιμα στοιχεία με την κατάλληλη επεξεργασία. Μερικά παραδείγματα από datasetsπου παράγονται από συσκευές πολιτών και εταιρειών είναι οι αναρτήσεις στα social media, τα βίντεο στο YouTube, οι αναζητήσεις στο Google και οι γεωγραφικές τοποθεσίες που καταγράφονται μέσω των κινητών τηλεφώνων. Η επεξεργασία τόσο μεγάλου όγκου δεδομένων συνήθως αφορά σε μεγάλες εταιρείες και οργανισμούς, που μπορούν να διαθέσουν τους απαραίτητους πόρους για την ανάπτυξη ή αγορά τεχνολογίας με σκοπό τη διαχείριση του. Εταιρείες όπως η Google, η Amazon και το eBay επενδύουν τεράστια ποσά σε τεχνολογίες επεξεργασίας των διαθέσιμων δεδομένων τους, γεγονός που τους διευκολύνει στη δημιουργία νέων προϊόντων και υπηρεσιών. Όμως και εταιρείες μικρότερου μεγέθους, που δε χρειάζονται την επεξεργασία μεγάλων όγκων δεδομένων για την ανάπτυξη των δραστηριοτήτων τους, μπορούν να εκμεταλλευτούν τα big data μέσω έτοιμων εργαλείων analytics ή μέσα από λύσεις digital marketing, ώστε να αναλύσουν τη συμπεριφορά των πελατών τους και να ικανοποιήσουν έτσι τις ανάγκες τους με τρόπους που στο παρελθόν ήταν ανέφικτοι.

1.4.1. Χαρακτηριστικά των Big Data

Τα κύρια χαρακτηριστικά των big data είναι ο όγκος, η ποικιλία και η ταχύτητα (Wang et al., 2016).

Όγκος. Αναφέρεται στο πλήθος των δεδομένων, ο οποίος στην περίπτωση των big data είναι τεράστιος. Ως χαρακτηριστικό, χρησιμεύει στην αποθήκευση και την επεξεργασία των δεδομένων. Ο όγκος από την άλλη πλευρά, είναι και το βασικό πρόβλημα των big data,καθώς καθημερινά δημιουργούνται εκατομμύρια δεδομένα.

Ποικιλία. Αναφέρεται στο ότι τα δεδομένα δεν έχουν σταθερή δομή και συνήθως δεν είναι σε πλήρως επεξεργάσιμη μορφή. Μπορεί να είναι δομημένα (π.χ. από σχεσιακές βάσεις δεδομένων), ημιδομημένα (π.χ. κοινωνικά δίκτυα) ή μη δομημένα (π.χ. πολυμέσα). Η ποικιλία τους επηρεάζει τη σημασία και τη μεταβλητότητα τους και έτσι αποτελεί πρόκληση για τις παραδοσιακές βάσεις δεδομένων.

Ταχύτητα. Αναφέρεται στο χρόνο που χρειάζεται για τη δημιουργία και την αποθήκευση των δεδομένων, καθώς και στην πρόσβαση σε αυτά και αποτελεί βασική ιδιότητα των big data που παράγονται καθημερινά.

1.4.2. Η σχέση μεταξύ Big Data και Internet of Thing (IoT)

Η έννοια του Internet of Things αφορά στη χρήση ενσωματωμένων αισθητήρων για τη συλλογή δεδομένων, σε αντικείμενα της καθημερινότητας από βιομηχανικά μηχανήματα μέχρι wearable συσκευές. Οι συσκευές συνδέονται σε μία εφαρμογή που έχει σχεδιαστεί για να αντιμετωπίζει συγκεκριμένες ανάγκες και έχουν σχεδιαστεί ώστε να αναγνωρίζουν τις πληροφορίες που χρειάζονται και να τις εκμεταλλευτούν κατάλληλα[2].

Παραδείγματα της χρήσης του διαδικτύου πραγμάτων είναι οι αισθητήρες για την αυτόματη ρύθμιση της θέρμανσης ή του φωτισμού, το αυτόματο πότισμα ανάλογα με τις καιρικές συνθήκες, οι μετρήσεις των καρδιακών παλμών κατά τη διάρκεια της άσκησης και πολλά άλλα. Κοινώς, το Internet of Things είναι η τεχνολογία του μέλλοντος που θα χρησιμοποιηθεί για τη διευκόλυνση της καθημερινότητας των ανθρώπων.

Η μεγαλύτερη ίσως πρόκληση που αντιμετωπίζουν όσοι παράγουν τεχνολογικά συστήματα IoT, είναι η επεξεργασία του τεράστιου όγκου πληροφοριών που λαμβάνουν μέσω των συσκευών με τους ενσωματωμένους αισθητήρες και οι οποίες προέρχονται από δεδομένα του πραγματικού κόσμου. Τα big data που παράγονται συγκεκριμένα από συσκευές IoT διαφέρουν από τα υπόλοιπα, ως προς την ετερογένεια και τον αδόμητο χαρακτήρα τους. Τα βασικά χαρακτηριστικά τους είναι ότι η πληθώρα των δεδομένων που δημιουργούνται από μεγάλο αριθμό τερματικών, είναι ημιδομημένα ή αδόμητα και χρησιμεύουν μόνο εφόσον επεξεργαστούν και αναλυθούν (Chenetal., 2014). Αν και η ανάπτυξη big data μέσω του IoT βρίσκεται σε πρώιμο στάδιο, έχει αναγνωριστεί ότι οι δύο τεχνολογίες είναι αλληλένδετες και πρέπει να αναπτυχθούν από κοινού.

1.4.3. Μέθοδοι επεξεργασίας Big Data

Όσον αφορά στον τρόπο εξαγωγής πληροφοριών από τα big data, θα αναφέρουμε τις βασικές μεθόδους επεξεργασίας τους (Yaqoob et al.,2016).

---

[2] ibm.com, «What is the Internet of Things, and how does it work?». *Internet of Things blog*. 2016-11-17»

Ευρετηρίαση. Χρησιμοποιείται για το γρήγορο εντοπισμό δεδομένων και στα big data μέσω ενός ευρετηρίου μπορούν να ομαδοποιηθούν οι πληροφορίες, δημιουργώντας αρχεία. Ωστόσο η συγκεκριμένη μέθοδος είναι αρκετά ακριβή.

Καταμερισμός. Αποτελεί μία πολύ αποτελεσματική μέθοδο ανάκτησης δεδομένων, η οποία δε χρειάζεται τη χρήση ευρετηρίου και αποδίδει καλύτερα όταν έχουμε διακριτά και τυχαία δεδομένα, οργανωμένα σε σειρά. Ενώ από τη μία με τον κατακερματισμό επιτυγχάνεται η ταχεία ανάγνωση των δεδομένων, από την άλλη έχει υψηλή πολυπλοκότητα και γραμμική ανίχνευση.

Φιλτράρισμα. Με τη χρήση του κατάλληλου φίλτρου αναζήτησης εξοικονομούμε χώρο αποθήκευσης δεδομένων και ελαχιστοποιούμε τις περιπτώσεις εξαγωγής λανθασμένων πληροφοριών. Παρότι η συγκεκριμένη μέθοδος περιορίζει σημαντικά τη δυνατότητα διαγραφής, εξυπηρετεί στην απόδοση του αποθηκευτικού χώρου και έχει υψηλή ταχύτητα στην εύρεση πληροφοριών.

1.5. Ανάλυση κειμένου

1.5.1. Επεξεργασία φυσικής γλώσσας (NLP)

Η επεξεργασία φυσικής γλώσσας αποτελεί ένα πεδίο της επιστήμης των ηλεκτρονικών υπολογιστών που αφορά στην αλληλεπίδραση μεταξύ υπολογιστή και ανθρώπινης (φυσικής) γλώσσας. Επίσης, σχετίζεται με το πεδίο της τεχνητής νοημοσύνης. Βασική ασχολία της είναι η κατανόηση της φυσικής γλώσσας, ώστε να δώσει στους υπολογιστές τη δυνατότητα να εξάγουν συμπεράσματα όταν δέχονται δεδομένα από αυτή[3].

Στις μέρες μας, οι NLP αλγόριθμοι βασίζονται στη μηχανική μάθηση με στατιστικές μεθόδους. Οι αλγόριθμοι αυτοί δέχονται ως είσοδο σύνολα χαρακτηριστικών, τα οποία προκύπτουν από την επεξεργασία των δεδομένων εισόδου. Αλγόριθμοι όπως τα δέντρα αποφάσεων, ήταν από τους πρώτους που χρησιμοποιήθηκαν και παρήγαγαν συστήματα από απόλυτους κανόνες «if–then». Οι τελευταίες έρευνες, οι οποίες εστίασαν σε στατιστικά μοντέλα για απλές αποφάσεις βάσει πιθανοτήτων, υπολόγιζαν πραγματικές τιμές στα χαρακτηριστικά εισόδου. Οι αλγόριθμοι αυτοί είναι ικανοί να εκφράσουν με μεγάλη βεβαιότητα όχι μόνο μία

[3] nlp.stanford.edu, « GloVe: Global Vectors for Word Representation »

αλλά πολλές πιθανές απαντήσεις και έτσι μπορούν να παράγουν πιο αξιόπιστα αποτελέσματα[4].

Παρακάτω θα αναφέρουμε μερικούς από τους πιο συνηθισμένους τομείς έρευνας στην επιστήμη της φυσικής γλώσσας, από τους οποίους κάποιοι είναι άμεσα εφαρμόσιμοι σε πραγματικά προβλήματα και κάποιοι αξιοποιούνται περισσότερο ως υπο-ενέργειες επίλυσης μεγαλύτερων προβλημάτων.

Παραγωγή φυσικής γλώσσας. Μετατρέπει τις πληροφορίες που προέρχονται από βάσεις δεδομένων σε φυσική γλώσσα που μπορεί να διαβαστεί από άνθρωπο.

Κατανόηση φυσικής γλώσσας. Μετατρέπει μέρος κειμένων σε λογικές δομές, εύκολα διαχειρίσιμες από υπολογιστές.

Αναγνώριση χαρακτήρων από εικόνες. Διαβάζει το κείμενο που εισάγεται από μία εικόνα που απεικονίζει χαρακτήρες.

Αναγνώριση ομιλίας. Αναγνωρίζει τη λεκτική αναπαράσταση της ομιλίας μέσω ενός ηχητικού ντοκουμέντου και μετατρέπει την ομιλία σε κείμενο.

Αυτόματη μετάφραση. Μεταφράζει κείμενο από μία φυσική γλώσσα σε μία άλλη. Ωστόσο αντιμετωπίζει προβλήματα κυρίως λόγω των διαφορών μεταξύ των ανθρώπινων γλωσσών και της πολυπλοκότητας των ανθρώπινων γνώσεων πάνω σε αυτές.

Απάντηση σε ερωτήσεις. Οι ερωτήσεις μπορεί να έχουν τόσο συγκεκριμένες απαντήσεις οι οποίες να είναι οι σωστές, μπορεί όμως και να είναι θεωρητικές, δηλαδή να μην έχουν συγκεκριμένη απάντηση που να θεωρείται σωστή.

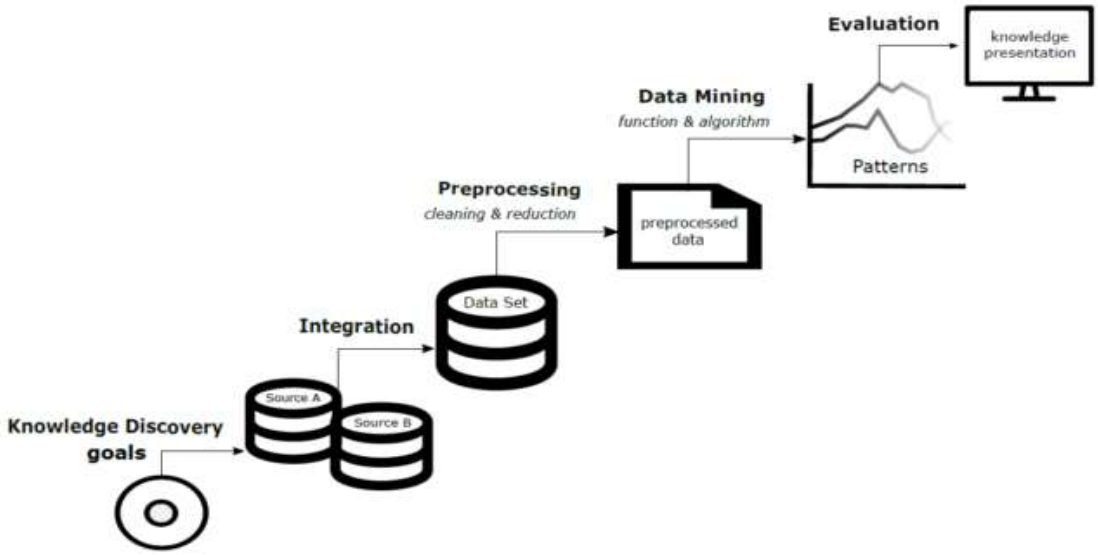
1.5.2. Εξόρυξη, Κατηγοριοποίηση, Ανάκτηση κειμένου – Εξαγωγή πληροφοριών

Ως εξόρυξη κειμένου εννοούμε την ανακάλυψη χρήσιμων πληροφοριών από δεδομένα φυσικής γλώσσας, αλλά και την ανάπτυξη εφαρμογών αυτόματης διαχείρισης κειμένων. Το εμπορικό ενδιαφέρον για τέτοιου είδους εφαρμογές άνθισε τα τελευταία χρόνια λόγω της έκρηξης του παγκόσμιου ιστού και της τεράστιας αύξησης των διαθέσιμων ψηφιακών κειμένων.

Τα δεδομένα κειμένου παρουσιάζουν ορισμένα προβλήματα. Αρχικά, τα κείμενα όταν είναι στην αρχική τους μορφή, δεν είναι άμεσα επεξεργάσιμα από τους αλγόριθμους μηχανικής μάθησης και χρειάζεται να προηγηθεί ένα στάδιο

[4] nlp.stanford.edu, « GloVe: Global Vectors for Word Representation »

κατάλληλου μετασχηματισμού. Σημαντικό πρόβλημα επίσης, αποτελεί το μέγεθος των κειμένων, αφού κάθε λέξη που περιλαμβάνουν ισοδυναμεί με ένα χαρακτηριστικό. Ακόμη, το γεγονός ότι κάποιες από τις λέξεις που περιλαμβάνονται στο κείμενο μπορεί να έχει περισσότερες από μία έννοιες (π.χ. συνώνυμες λέξεις), ενδεχομένως να αποτελέσει πρόβλημα. Δυσκολία υπάρχει περίπτωση να επιφέρει και το ότι στα κείμενα ηλεκτρονικού ταχυδρομείου χρησιμοποιείται ανεπίσημη γλώσσα και ίσως να περιέχουν συντομογραφίες κ.λ.π.



Εικόνα 2: Τα στάδια της εξόρυξης δεδομένων (Ανακτήθηκε από: mti.binus, «Knowledge Discovery Process»)

Όπως βλέπουμε και στην εικόνα 2, τα στάδια της εξόρυξης γνώσης από κείμενο είναι τα εξής (Bing, 2011):

Συλλογή δεδομένων. Χρειάζεται να γίνει συλλογή των κειμένων από το ίντερνετ.

Προεπεξεργασία. Η προεπεξεργασία είναι μια πολύ σημαντική διαδικασία, με την οποία αφού επιλεχθούν τα κείμενα, αφαιρείται οποιοδήποτε περιττό στοιχείο. Έτσι, επιτυγχάνεται ο «καθαρισμός» των κειμένων από «βρώμικα» δεδομένα όπως ελλιπή, ασυνεπή ή λανθασμένα.

Μετασχηματισμός. Σε αυτό το στάδιο τα κείμενα μετατρέπονται σε επεξεργάσιμη μορφή από τους αλγόριθμους της μηχανικής μάθησης.

Συρρίκνωση. Το στάδιο αυτό χρειάζεται λόγω του μεγάλου μεγέθους των αρχικών κειμένων.

Εξόρυξη. Στο στάδιο αυτό εφαρμόζονται οι αλγόριθμοι μηχανικής μάθησης και ως αποτέλεσμα είτε περιγράφονται τα δεδομένα που μπορεί να οδηγήσουν σε χρήσιμη γνώση, είτε προκύπτουν μοντέλα που μπορούν να αξιοποιηθούν για την αυτόματη διαχείριση των κειμένων.

Αξιολόγηση. Αξιολογούνται τα μοντέλα που προκύπτουν από την εφαρμογή των αλγόριθμων μηχανικής μάθησης.

Αξιοποίηση. Αφού αξιολογηθούν τα μοντέλα, γίνεται προσπάθεια να αξιοποιηθούν τα αποτελέσματα της εξόρυξης.

Οι βασικότερες μέθοδοι εξόρυξης κειμένων, τις οποίες θα αναλύσουμε παρακάτω, είναι η κατηγοριοποίηση κειμένου, η ανάκτηση και η εξαγωγή πληροφοριών.

Κατηγοριοποίηση κειμένου

Κατηγοριοποίηση ή αλλιώς ταξινόμηση κειμένου είναι η διαδικασία με την οποία προσδιορίζεται η κατηγορία στην οποία ανήκει ένα κείμενο, μία μέτρηση ή οτιδήποτε άλλο. Κατά τη διαδικασία αυτή, εισάγεται ένα σύνολο δεδομένων τα οποία περιγράφουν το αντικείμενο και μετά από επεξεργασία, εξάγεται η κατηγορία στην οποία ανήκει. Κύριο χαρακτηριστικό της είναι ότι υπάρχει ένα προκαθορισμένο σύνολο κατηγοριών σε κάποια από τις οποίες κατατάσσεται ένα αντικείμενο. Το εν λόγω αντικείμενο μπορεί να καταταγεί σε καμία, μία ή περισσότερες κατηγορίες [Βλαχάβας κ.α., 2002].

Εφαρμογές που κατηγοριοποιούν κείμενα ξεκίνησαν να εφαρμόζονται από τη δεκαετία του 1960, ωστόσο από το 2000, λόγω της μεγάλης αύξησης των ηλεκτρονικών κειμένων, γνωρίζουν μεγάλη άνθιση. Η κυρίαρχη προσέγγιση για τη δημιουργία των ταξινομητών βασίζεται σε τεχνικές μηχανικής μάθησης, όπου μέσω ενός συνόλου παραδειγμάτων που χρησιμοποιούνται για την εκπαίδευση του, δίνονται τα χαρακτηριστικά κάθε κατηγορίας στον ταξινομητή [Βλαχάβας κ.α., 2002].

Το διαδίκτυο είναι ένα πολύ δυναμικό περιβάλλον, με πληθώρα πηγών κειμένων, όπως οι ιστοσελίδες, τα email, τα forum συζητήσεων, τα blogs, οι ροές ειδήσεων κ.α. Στις μέρες μας, υπάρχουν πολύ ενδιαφέρουσες εφαρμογές της αυτόματης ταξινόμησης τέτοιου είδους κειμένων. Οι γνωστότερες είναι το φιλτράρισμα της

ανεπιθύμητης αλληλογραφίας (spam) και των πορνογραφικού περιεχομένου σελίδων (popup) και η δημιουργία εξατομικευμένων διαφημίσεων και ειδησεογραφικών ροών.

Πρόβλημα στην κατηγοριοποίηση των κειμένων μπορεί να προκύψει λόγω των πολλαπλών ετικετών με τις οποίες σχετίζονται. Για παράδειγμα ένα κείμενο στο οποίο οι χρήστες έχουν αναθέσει πολλά hashtags, μπορεί να επηρεάσει αρνητικά τον ταξινομητή ως προς την ποιότητα της πρόβλεψης και του χρόνου εκπαίδευσης και αξιολόγησης (Tsoumakas et al., 2008).

Ανάκτηση κειμένου

Η ανάκτηση κειμένου είναι η εύρεση των εγγράφων που περιέχουν τις πληροφορίες και όχι τις ίδιες τις πληροφορίες που εξετάζει. Προκειμένου να ανακτηθούν τα κείμενα, χρησιμοποιούνται στατιστικές μέθοδοι για την αυτόματη επεξεργασία των δεδομένων ως εξής: εισάγονται στο σύστημα έγγραφα κειμένου και ένα ερώτημα και το σύστημα εξάγει ένα σύνολο εγγράφων σχετικά με αυτό. Οι προκλήσεις που καλείται να αντιμετωπίσει ένα έξυπνο σύστημα ανάκτησης πληροφοριών, αφορούν στη σημασία των λέξεων (π.χ. συνώνυμα ή λέξεις με διφορούμενες έννοιες), στη σειρά των λέξεων μέσα στην πρόταση, καθώς μπορεί να αλλάξει εντελώς το νόημα της και η αξιοπιστία της πηγής (Hearst, 1999).

Εξαγωγή πληροφοριών

Οι πληροφορίες που περιέχονται στα κείμενα φυσικής γλώσσας δεν είναι άμεσα κατάλληλες να αναλυθούν από τον υπολογιστή. Παρόλα αυτά, οι υπολογιστές είναι σχεδιασμένοι να «σκανάρουν» μεγάλα σύνολα κειμένων και να εξάγουν πληροφορίες από μεμονωμένες λέξεις ή φράσεις.

Η διαδικασία εξαγωγής πληροφοριών περιλαμβάνει ορισμένα βήματα επεξεργασίας, όπως ο διαχωρισμός των συμβολοσειρών σε tokens, συνήθως λέξεις, η κατάτμηση της πρότασης σε μέρη του λόγου και η αναγνώριση των ονομασιών (πρόσωπα, τοποθεσίες, εταιρείες κλπ). Για να λειτουργήσει, εισάγεται ένα σύνολο κειμένων στο σύστημα εξαγωγής πληροφοριών και ένα ερώτημα. Το σύστημα βρίσκει τις προτάσεις με τη σχετική πληροφορία, απορρίπτοντας οτιδήποτε άσχετο και την εξάγει σε μία προκαθορισμένη μορφή.

1.5.3. Ανάλυση συναισθήματος με μηχανική μάθηση

Όπως έχουμε ήδη αναφέρει, η ανάλυση συναισθήματος, δηλαδή η τοποθέτηση της άποψης του συγγραφέα πάνω σε ένα θέμα, στη θετική ή αρνητική κατηγορία, μπορεί να γίνει με μεθόδους μηχανικής μάθησης. Ο Peter Turney το 2002, ανέπτυξε ένα σύστημα αυτόματης ταξινόμησης των κριτικών για ταινίες και προϊόντα, χρησιμοποιώντας μία μέθοδο μη επιτηρούμενης μάθησης, η οποία βασιζόταν στην αμοιβαία πληροφορία μεταξύ μίας φράσης και των λέξεων «τέλειο» και «φτωχό». Επίσης το ίδιο έτος, οι Pang, Lee και Vaithynathan, εφάρμοσαν αλγόριθμους επιβλεπόμενης μηχανικής μάθησης, τους οποίους θα αναπτύξουμε παρακάτω, προκειμένου να ταξινομήσουν συναισθηματικά, κριτικές ταινιών. Αν και η πιο συχνή προσέγγιση ταξινόμησης ενός κειμένου είναι η δυαδική, δηλαδή η ταξινόμηση σε μία από τις δύο κατηγορίες (θετική ή αρνητική), δεν είναι η μοναδική. Οι Pang & Lee (2005) αύξησαν τις κατηγορίες ταξινόμησης σε πέντε, όπου οι χρήστες καλούνταν να βαθμολογήσουν τις ταινίες επιλέγοντας από ένα έως πέντε αστέρια. Τα επόμενα χρόνια, άρχισαν να χρησιμοποιούνται ολοένα και πιο πολύπλοκα μοντέλα μηχανικής μάθησης, όπως αυτά που λαμβάνουν υπόψη τη σειρά των λέξεων μέσα στο κείμενο και τα τελευταία χρόνια, μοντέλα που χρησιμοποιούν νευρωνικά δίκτυα, τα οποία παρουσιάζουν πολύ αξιοσημείωτα αποτελέσματα.

Ωστόσο, η εφαρμογή τεχνικών μηχανικής αναγνώρισης συναισθημάτων από κείμενο εμφανίζει προβλήματα. Ένα από αυτά είναι η αναγνώριση του ουδέτερου συναισθήματος. Δεν περιέχουν όλες οι προτάσεις είτε θετικό, είτε αρνητικό συναίσθημα. Έτσι, όταν η τεχνική που χρησιμοποιείται εστιάζει σε αυτά τα δύο μόνο, χωρίς να λάβει υπόψη της την ουδέτερη κλάση, το αποτέλεσμα είναι να κατηγοριοποιούνται οι ουδέτερες λέξεις είτε στη θετική, είτε στην αρνητική κατηγορία και να δημιουργείται μια υπερπροσαρμογή του μοντέλου, δίνοντας μη ακριβή αποτελέσματα. Το πρόβλημα μπορεί να αντιμετωπιστεί με τους αλγόριθμους επιβλεπόμενης μηχανικής μάθησης SVM (Kopper & Schler, 2005).

1.5.3.1. Ο αλγόριθμος Naive Bayes

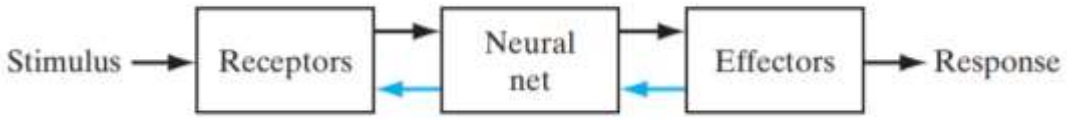
Ο αλγόριθμος Naïve Bayes είναι αλγόριθμος επιβλεπόμενης μηχανικής μάθησης και ένας απλός ταξινομητής βασισμένος σε πιθανότητες, ο οποίος όπως υποδηλώνει η ονομασία του, βασίζεται στο θεώρημα του Bayes και στην απλοϊκή (naive) θεωρία

της ανεξαρτησίας μεταξύ των χαρακτηριστικών της κάθε κλάσης. Είναι από τους βασικούς και πιο αποδοτικούς αλγόριθμους για την ταξινόμηση κειμένων. Στα θετικά της εφαρμογής του συγκεκριμένου αλγορίθμου, συγκαταλέγονται η απλότητα του στη χρήση, η καλή απόδοση, οι χαμηλές απαιτήσεις σε μνήμη και ο μικρός χρόνος εκπαίδευσης. Από την άλλη πλευρά όμως, ο αλγόριθμος Naïve Bayes συχνά υπερεκτιμά τις πιθανότητες απαντήσεων, γεγονός που τον κάνει ίσως ένα κακό εκτιμητή. Υπάρχουν διαφορετικές παραλλαγές, οι διαφορές των οποίων αφορούν στις υποθέσεις που κάνουν αναφορικά με την κατανομή των πιθανοτήτων (Liu, 2011).

1.5.3.2.Τεχνητά νευρωνικά δίκτυα

Τα νευρωνικά δίκτυα, η ονομασία των οποίων είναι εμπνευσμένη από τους νευρώνες του εγκεφάλου, χρησιμοποιούνται σε προβλήματα επιβλεπόμενης μηχανικής μάθησης με πολύ καλά αποτελέσματα. Το συγκεκριμένο μοντέλο μηχανικής μάθησης έχει σαν σκοπό να ρυθμίσει τις εσωτερικές παραμέτρους και για την εκπαίδευση του χρησιμοποιείται μία συνάρτηση κόστους.

Σε ένα νευρωνικό δίκτυο, οι μονάδες εισόδου είναι τα χαρακτηριστικά και οι μονάδες εξόδου οι τάξεις. Τα βάρη στις συνδέσεις μεταξύ εισόδων και εξόδων είναι οι μεταξύ τους συσχετίσεις. Προκειμένου να ταξινομηθεί ένα κείμενο, τα βάρη των χαρακτηριστικών του εισάγονται στις εισόδου του δικτύου, ενεργοποιούνται και η απόφαση καθορίζεται από την τελική τιμής της εξόδου (εικόνα 3).



Εικόνα3: Block diagram representation of nervous system (Simon, Haykin., "Neural Networks and Learning Machines", 2008, p.6)

Η πιο απλή μορφή νευρωνικού δικτύου είναι ο γραμμικός ταξινομητής perceptron, ο οποίος αποτελείται από ένα μόνο νευρώνα. Ο νευρώνας αυτός δέχεται σήματα εισόδου και μία σταθερή είσοδο που αντιστοιχεί στην πόλωση και χρησιμοποιώντας ένα γραμμικό μοντέλο, παράγει ένα σήμα εξόδου. Ωστόσο οι δυνατότητες του perceptron είναι περιορισμένες, καθώς μπορεί να επιλύει μόνο προβλήματα τα οποία είναι γραμμικά διαχωρίσιμα (Haykin, 2009).

Μια πιο σύνθετη μορφή νευρωνικού δικτύου είναι τα νευρωνικά δίκτυα πολλών επιπέδων. Τα νευρωνικά αυτά δίκτυα, μπορούν να επιλύσουν προβλήματα που δεν είναι διαχωρίσιμα γραμμικά και αποτελούνται από το επίπεδο εισόδου, το οποίο αποστέλλει σήματα εισόδου στους νευρώνες του κρυφού επιπέδου, ένα ή πολλά κρυφά επίπεδα και το επίπεδο εξόδου, το οποίο περιέχει γραμμικούς ή μη νευρώνες.

1.5.3.3. Deep Learning

Η μέθοδος Deep Learning λειτουργεί σε φάσεις (Souza et al., 2017). Πρώτη είναι η φάση της προεκμάθησης «layer wise training» και ακολουθεί η φάση της οριστικής ρύθμισης «fine tuning». Η μέθοδος αυτή λειτουργεί με παρόμοιο με τα νευρωνικά δίκτυα τρόπο, με τη διαφορά ότι επειδή η μάθηση γίνεται σε κάθε στρώμα, τα βάρη ρυθμίζονται μεταξύ όλων των κρυφών στρωμάτων και όχι μόνο μεταξύ των στρωμάτων εισόδου και εξόδου με το κρυφό στρώμα.

Για τη φάση της εκπαίδευσης, χρειάζεται να γίνει εξίσωση των τιμών εισόδου και εξόδου. Οι τιμές έρχονται μέσα από ένα κρυφό στρώμα, στο στρώμα εισόδου και το βάρος των νευρωνικών δικτύων προσαρμόζεται με σκοπό την ελαχιστοποίηση των θορύβων. Μόλις ολοκληρωθεί η εκπαίδευση ενός στρώματος, το δίκτυο προχωράει στην εκπαίδευση του επόμενου. Έτσι, κάθε κρυφό στρώμα γίνεται στρώμα εισόδου όταν η εκμάθηση προχωράει στο επόμενο στρώμα.

Στη φάση της προεκμάθησης η εξαγωγή των τιμών γίνεται με μη επιβλεπόμενη μάθηση, οπότε για την ορθή ταξινόμηση τους, χρειάζεται η φάση της οριστικής ρύθμισης. Σε αυτή τη φάση προστίθεται ένα στρώμα εξόδου στα στρώματα που ολοκλήρωσαν την προεκμάθηση και γίνονται οι τελικές ρυθμίσεις του δικτύου. Για την επιβλεπόμενη μάθηση στα στρώματα εξόδου, συνήθως χρησιμοποιούνται αλγόριθμοι μηχανικής μάθησης όπως οι Support Vector Machines (SVMs).

Οι Μηχανές Διανυσμάτων Υποστήριξης SVMs είναι μία δημοφιλής μέθοδος επιβλεπόμενης μηχανικής μάθησης, η οποία είναι ιδιαίτερα αποτελεσματική σε πολυδιάστατους χώρους, ακόμη και όταν τα χαρακτηριστικά είναι περισσότερα από τα δείγματα και δεν απαιτεί μεγάλη μνήμη. Επίσης, επιλύουν γραμμικά και μη γραμμικά προβλήματα ταξινόμησης χρησιμοποιώντας προκαθορισμένες συναρτήσεις. Το βασικό μειονέκτημα αυτής της μεθόδου είναι ότι δεν υπολογίζει τις πιθανότητες απευθείας, όπως ο αλγόριθμος Naïve Bayes (Liu, 2011).

1.5.4. Το Κοινωνικό Δίκτυο Twitter

Το Twitter είναι ένας από τους δημοφιλέστερους ιστότοπους κοινωνικής δικτύωσης και επιτρέπει στους χρήστες του να κοινοποιήσουν και να διαβάσουν tweets, δηλαδή μηνύματα μεγέθους το πολύ 280 χαρακτήρων. Η λειτουργία του ξεκίνησε το 2006 και σήμερα έχει πάνω από 300 εκατομμύρια ενεργούς χρήστες[5].

Τα βασικά χαρακτηριστικά των tweets, εκτός από το μικρό μέγεθος των μηνυμάτων, είναι ο ανεπίσημος λόγος, οι συντομεύσεις και τα ακρωνύμια που περιέχουν τα μηνύματα, η επιμήκυνση των λέξεων ώστε να δοθεί έμφαση και τα ειδικά tokens όπως usernames, hashtags, emoticons, urls και retweets. Τα usernames αναφέρονται σε ένα συγκεκριμένο χρήστη του Twitter και ξεκινούν με το σύμβολο «@». Τα hashtags ξεκινούν με το σύμβολο «#», δεν περιέχουν κενά ή σύμβολα και δηλώνουν το θέμα του tweet, τη συναισθηματική κατάσταση του χρήστη ή οτιδήποτε άλλο θέλει ο χρήστης να δηλώσει με αυτό. Τα emoticons (εικόνα 4) είναι εικονίδια που προσομοιάζουν ανθρώπινες εκφράσεις όπως χαρά, λύπη, θυμός, αγανάκτηση και χρησιμοποιούνται για να εκφράσουν τη συναισθηματική κατάσταση γενικά των χρηστών ή πάνω σε ένα συγκεκριμένο θέμα. Οι πληροφορίες που μπορούν να εξαχθούν από αυτά είναι πολύτιμες. Τα urls είναι ηλεκτρονικές διευθύνσεις που μπορεί να επισημάνει ένας χρήστης στο μήνυμα του. Τα retweets είναι κοινοποιήσεις του tweet ενός χρήστη από ένα άλλο χρήστη.

Εικόνα 4: Παράδειγμα των πιο συνηθισμένων emoticons

Το γεγονός ότι το Twitter είναι ένα κοινωνικό δίκτυο το οποίο έχει υιοθετηθεί ευρέως, θεωρείται ότι αποτελεί μία καλή αντανάκλαση του τι συμβαίνει στον κόσμο και αντικατοπτρίζει την τάση που επικρατεί πάνω σχεδόν στα πάντα. Για παράδειγμα δείχνει την τάση για πολιτικά και κοινωνικά θέματα, για τον κινηματογράφο, για προϊόντα και υπηρεσίες και πάρα πολλά άλλα. Οι τάσεις αυτές μπορούν να χρησιμοποιηθούν από εταιρείες για θέματα μάρκετινγκ, όπως π.χ. δημιουργία

[5] el.wikipedia.org, «Twitter»

διαφημίσεων ή λανσάρισμα νέων προϊόντων. Συνεπώς αντιλαμβανόμαστε όλοι ότι η ανάλυση των tweets μπορεί να γίνει χρυσορυχείο για εταιρείες και οργανισμούς προκειμένου να αποκτήσουν συγκριτικό πλεονέκτημα απέναντι στους ανταγωνιστές τους.

1.5.5. Εφαρμογή sentiment analysis σε tweets με NLP tools

Για να εφαρμοστεί η ανάλυση συναισθήματος στα tweets που συλλέγονται χρειάζεται να χρησιμοποιηθεί κάποιο NLP toolkit. Τα Natural Language Processing toolkits προσφέρουν επιλογές για κάθε NLP πεδίο και τα κατάλληλα εργαλεία για την ανάλυση συναισθήματος. Πολλά από αυτά τα εργαλεία συνδυάζουν την NLP με τη μηχανική μάθηση. Στην εργασία μας, για την ανάλυση των tweets, θα χρησιμοποιήσουμε ένα συνδυασμό του λογισμικού RapidMiner και των εργαλείων AYLIEN API και Twitter Archiver.

RapidMiner. Είναι μία data science πλατφόρμα λογισμικού η οποία παρέχει ένα ολοκληρωμένο περιβάλλον εξόρυξης δεδομένων, μηχανικής μάθησης και ανάλυσης προγνωστικών. Χρησιμοποιείται τόσο για επαγγελματικούς, όσο και για εκπαιδευτικούς σκοπούς και υποστηρίζει όλα τα στάδια της ανάπτυξης διαδικασιών ανάλυσης δεδομένων, καθώς διαθέτει όλα τα απαραίτητα εργαλεία για την εισαγωγή και την επεξεργασία δεδομένων, αλλά και την παρουσίαση των αποτελεσμάτων που προκύπτουν από αυτές.

AYLIEN API. Είναι ένα εργαλείο ανάλυσης κειμένου που χρησιμοποιείται για τη σύνοψη των δεδομένων, την ανεύρεση μεμονωμένων εννοιών μέσα στο κείμενο και τη συναισθηματική ανάλυση του. Λειτουργεί καλύτερα με τη χρήση άλλων προσθέτων για την εισαγωγή δεδομένων από το ιστό, όπως για παράδειγμα το Twitter Archiver. Είναι δωρεάν έως 1.000 ερωτήματα.

Twitter Archiver. Χρησιμοποιείται για την αναζήτηση συγκεκριμένων hashtags στο Twitter, τα οποία και μπορεί να αποθηκεύσει σε αρχείο excel. Δέχεται απλά και σύνθετα ερωτήματα αναζήτησης, όπως π.χ. για την τοποθεσία, τη γλώσσα κλπ. Είναι δωρεάν αν η ενημέρωση των δεδομένων θα γίνεται ανά μία ώρα.

1.6. Εμπειρική Έρευνα – Συλλογή tweets και ανάλυση συναισθήματος με τη χρήση του RapidMiner και του AYLIEN

1.6.1. Συλλογή tweets

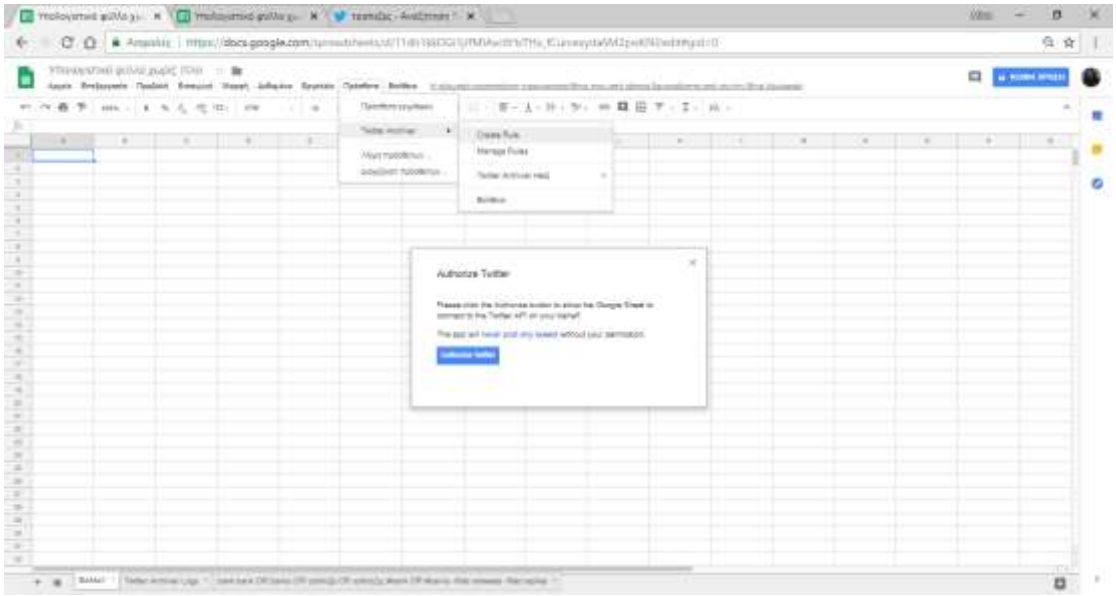
Στην ενότητα αυτή θα γίνει η συλλογή των tweets, τα οποία και θα χρησιμοποιήσουμε για την ανάλυση συναισθήματος.

Για το λόγο αυτό δημιουργήσαμε ένα προσωπικό λογαριασμό στο Twitter. Για να συλλέξουμε τα tweets χρησιμοποιήσαμε το Twitter Archiver του Google και δημιουργήσαμε ένα Google Spreadsheet στο οποίο αποθηκεύσαμε τα tweets που προέκυψαν από την αναζήτηση που κάναμε με λέξεις κλειδιά ή hashtags. Παρακάτω θα αναπτύξουμε αναλυτικά τα βήματα που ακολουθήσαμε για τα οποία εκπαιδευτήκαμε κυρίως από τη σελίδα https://www.labnol.org/internet/save-twitter-hashtag-tweets/6505/ και από διάφορα βίντεο στο Youtube όπως ενδεικτικά το «Save Tweets in a Google Spreadsheet» του χρήστη AMit Agarwal (https://www.youtube.com/watch?v=MGU7azCYFpw).

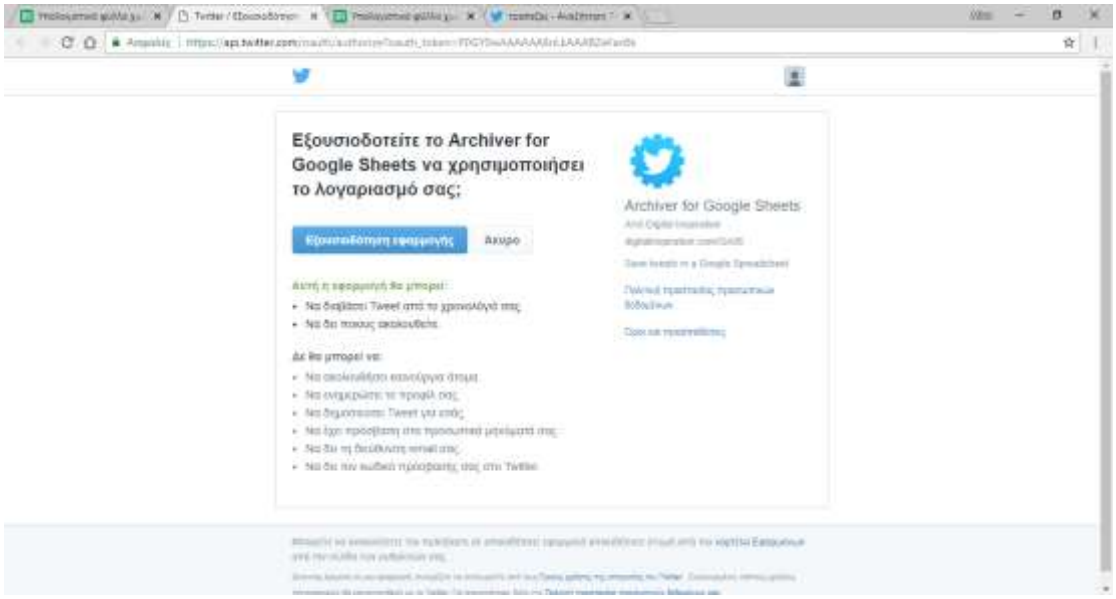
Αρχικά εγκαταστήσαμε το Twitter Archiver μέσω της διεύθυνσης:
https://chrome.google.com/webstore/detail/twitter-archiver/pkanpfekacaojdncfgbjadedbggbbphi

Εικόνα 5: Η εγκατάσταση του Twitter Archiver

Στη συνέχεια από το menu «πρόσθετα» επιλέξαμε «twitter archiver» και «authorize twitter», για να επιτρέψουμε την πρόσβαση στο λογαριασμό μας στο Twitter (εικόνες 4 & 5).



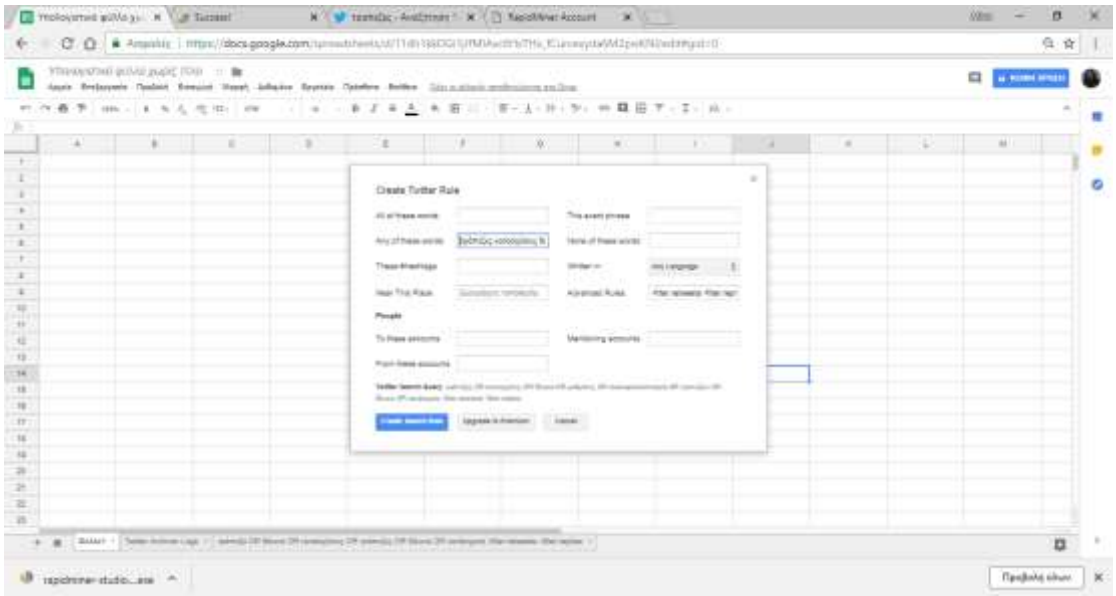
Εικόνα 6: Σύνδεση του Twitter Archiver με Twitter account



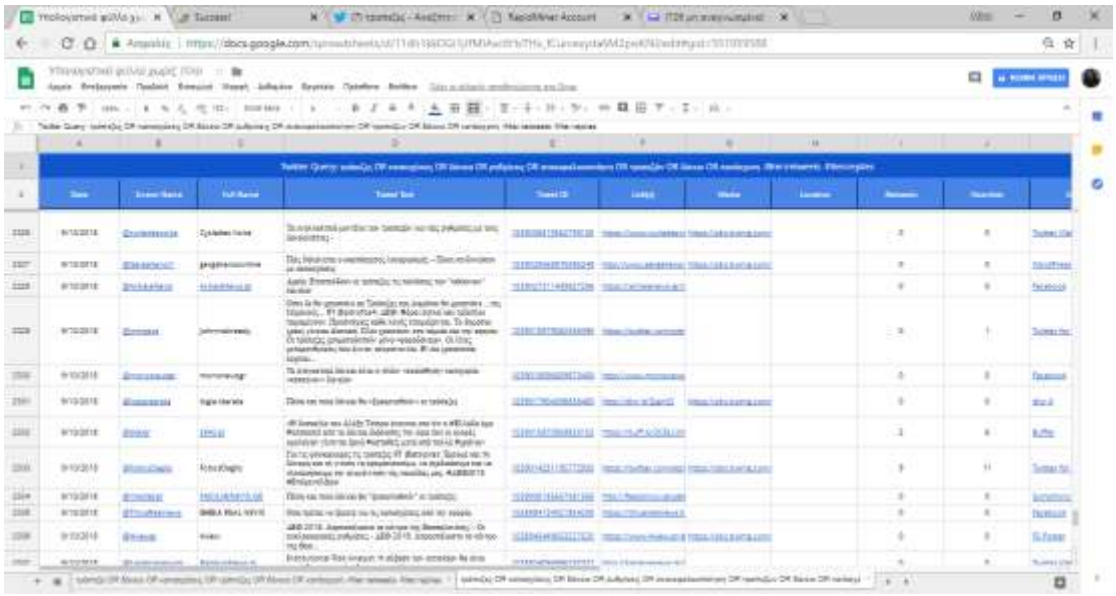
Εικόνα 7: Σύνδεση του Twitter Archiver με Twitter account (συνέχεια)

Έπειτα από το menu «πρόσθετα» επιλέξαμε «create search rule» και μετά «start tracking». Όπως βλέπουμε στις εικόνες 8, 9 & 10, αφού εισάγαμε τις λέξεις κλειδιά «τράπεζες, κατασχέσεις, δάνεια, ρυθμίσεις, ανακεφαλαιοποίηση, τραπεζών, δάνειο,

κατάσχεση», δημιουργήσαμε ένα Google Spreadsheet με τα tweets που ταιριάζουν στην αναζήτηση που κάναμε.



Εικόνα 8: Εισαγωγή λέξεων κλειδιά για τη δημιουργία Google Spreadsheet



Εικόνα 9: Αποτέλεσμα δημιουργίας Google Spreadsheet

9/10/2018	@AchaiaNews	AchaiaNews.gr	Αχαΐα: Επισπεύδουν οι τράπεζες τις πωλήσεις των "κόκκινων" δανείων
9/10/2018	@jnmazos	Johnmakready	Όσοι δε θα χρωστάνε ας Τράπεζες και Δημόσιο θα χρωστάνε ...της Μιχαλούς... RT @pitrofos4: ΔΕΘ: Φόροι διπλοί και τρίδιπλοι παραμένουν. Προσλήψεις κάθε λογής ετοιμάζονται. Το δημόσιο χρέος γίνεται ιδιωτικό. Όλοι χρωστούν στα ταμεία και την εφορία. Οι τράπεζες χρηματοδοτούν μόνο «φοροδάνεια». Οι λίγες μεταρρυθμίσεις που έγιναν ακυρώνονται. Η νέα χρεωκοπία έρχεται...
9/10/2018	@mononewsgr	mononewsgr	Τα στεγαστικά δάνεια είναι η πλέον «ευαίσθητη» κατηγορία «κόκκινων» δανείων
9/10/2018	@logiastarata	logia starata	Πόσα και ποια δάνεια θα «ξεφορτωθούν» οι τράπεζες
9/10/2018	@skaigr	SKAI.gr	«Η δυσκολία του Αλέξη Τσίπρα έγκειται στο ότι η #Ελλάδα έχει #αποκοπεί από τα δάνεια διάσωσης την ώρα που οι αγορές ομολόγων γίνονται ξανά #ασταθείς μετά από πολλά #χρόνια»
9/10/2018	@FotiosDaglis	FotiosDaglis	Για τις φουκαριαρες τις τράπεζες RT @atsipras: Έχουμε και τη δύναμη και τη γνώση να οραματιστούμε, να σχεδιάσουμε και να υλοποιήσουμε την αναγέννηση της πατρίδας μας. #ΔΕΘ2018 #ΕπόμενηΜέρα

Εικόνα 10: Αποτέλεσμα δημιουργίας Google Spreadsheet

Τα tweets που επιστράφηκαν δημοσιεύθηκαν την περίοδο 09-16/09/2018 κατά την οποία έλαβε χώρα η Διεθνής Έκθεση της Θεσσαλονίκης και κατά συνέπεια υπήρξε έντονη πολιτική κινητοποίηση στη χώρα. Το χρονικό διάστημα ήταν μόνο 7 μέρες διότι η εφαρμογή λόγω του free edition δεν επιτρέπει αναζήτηση για περισσότερες ημέρες. Το περιεχόμενο των tweets αφορούσε κυρίως σε σχόλια και σε αναμετάδοση σχολίων πολιτών και δημοσιογράφων σχετικά με τον τραπεζικό κλάδο, τις δηλώσεις πολιτικών προσώπων για τις τράπεζες και την αναμετάδοση σχετικών άρθρων.

1.6.2. Ανάλυση συναισθήματος με τη χρήση των RapidMiner και AYLIEN

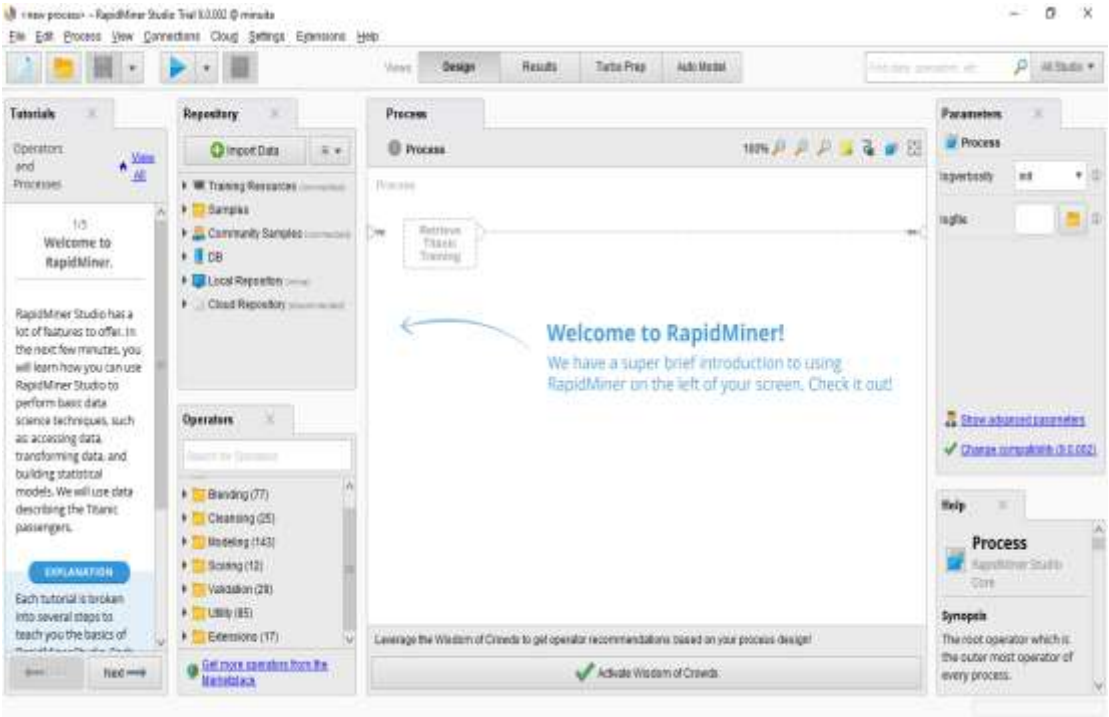
Για την παρούσα εργασία κατεβάσαμε από την επίσημη ιστοσελίδα της εταιρείας RapidMiner Inc. (http://www.rapidminer.com) την έκδοση RapidMiner Studio 9.0.002 trial edition, η οποία παρέχεται δωρεάν για εκπαιδευτικούς σκοπούς.

Για την εκμάθηση της εφαρμογής χρησιμοποιήθηκε ο οδηγός ReadExcel της εταιρείας RapidMiner για την έκδοση 9.0 (latest): https://docs.rapidminer.com/studio/operators/data_access/files/read/read_excel.html, καθώς και βίντεο αναρτημένα στο Youtube και πιο συγκεκριμένα τα βίντεο 01 έως 09 του επίσημου καναλιού της εταιρείας RapidMiner Inc.

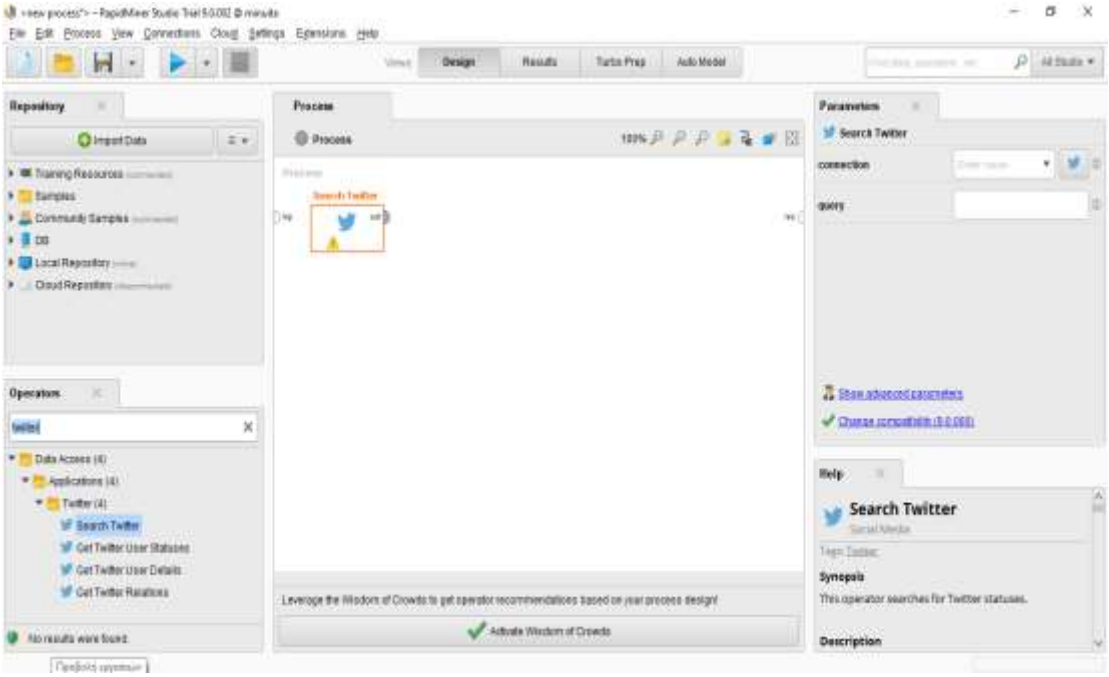
(https://www.youtube.com/user/RapidIVideos). Επίσης χρησιμοποιήσαμε τη δωρεάν έκδοση του Text Analysis API της εταιρείας AYLIEN (https://aylien.com/text-api/).

Παρακάτω θα αναλύσουμε τα βήματα που ακολουθήσαμε για την ανάλυση συναισθήματος στα tweets.

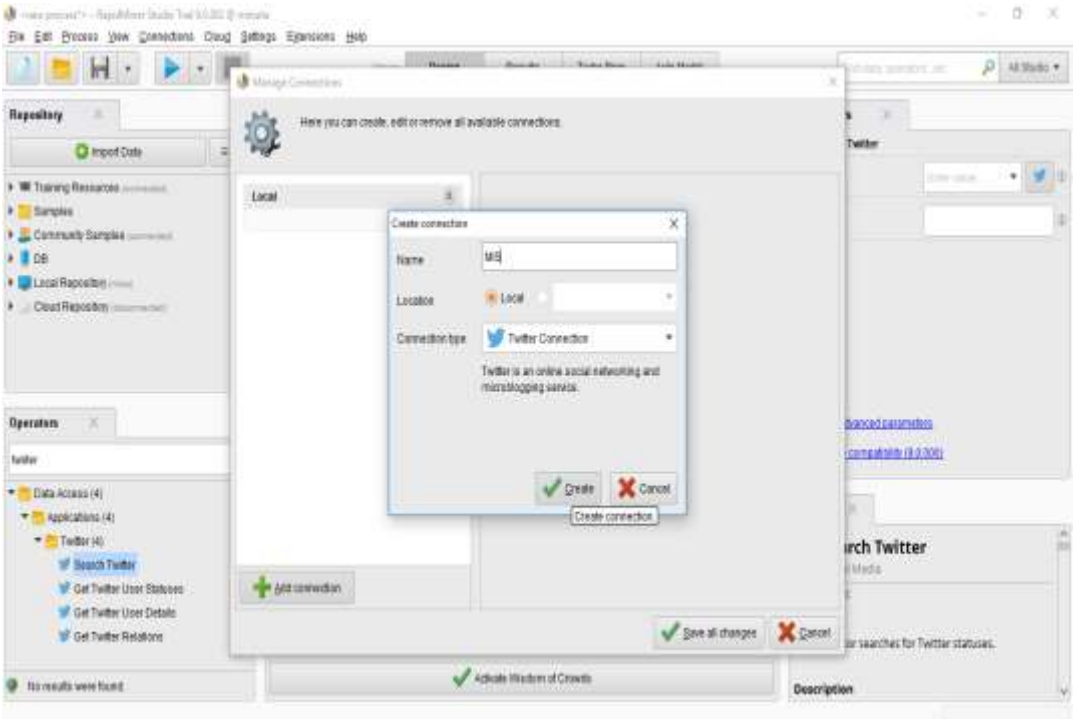
Για τη συλλογή των tweets εκτός από τα στοιχεία του Google Spreadsheet, χρησιμοποιήσαμε και μία αναζήτηση με σχετικό plug in μέσα από το RapidMiner. Αρχικά ανοίξαμε ένα κενό process και προσθέσαμε το «search Twitter» στους operators. Έπειτα από την παράμετρο connection, ακολουθήσαμε τα βήματα για να το συνδέσουμε με το λογαριασμό μας στο Twitter (εικόνες 11 έως 15).



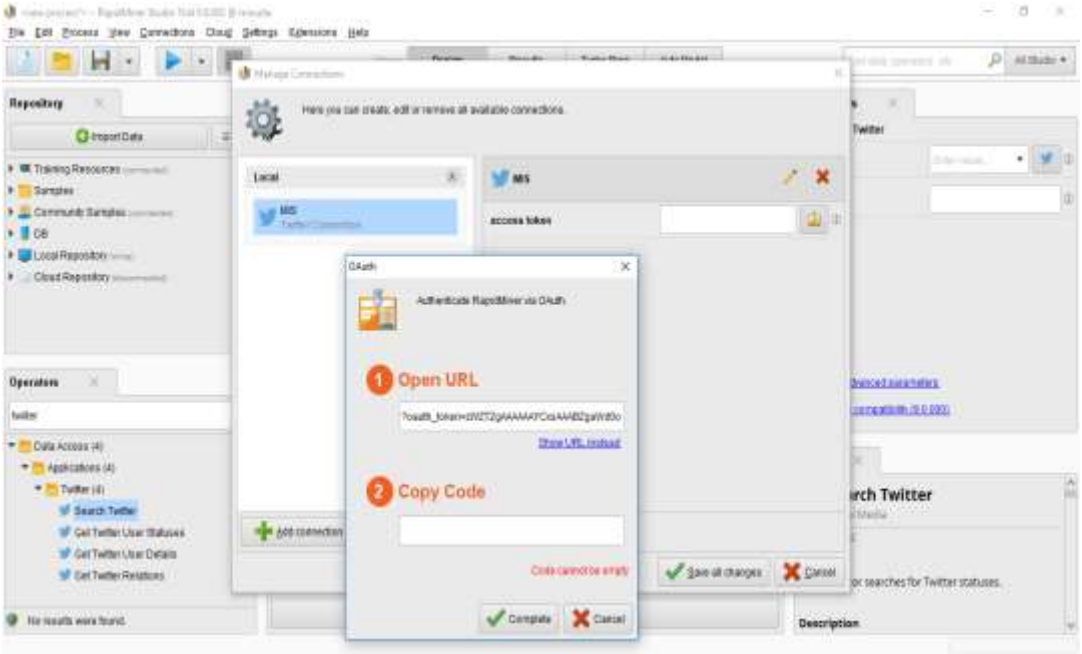
Εικόνα 11: Σύνδεση RapidMiner με Twitter account. Βήμα 1ο



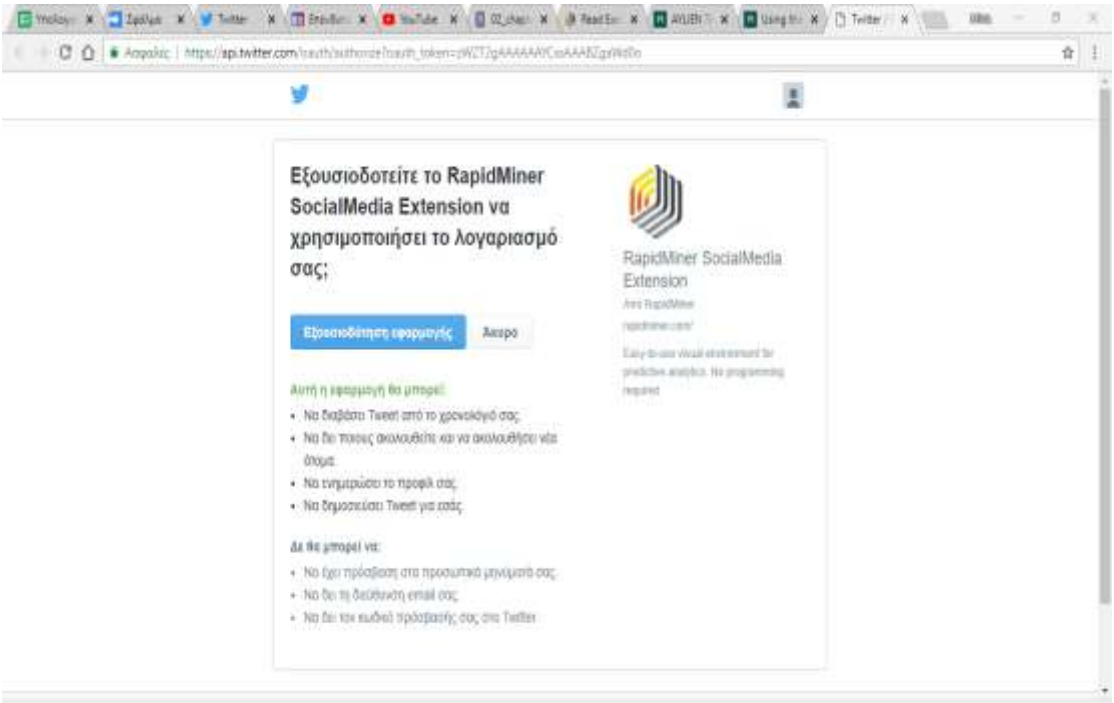
Εικόνα 12: Σύνδεση RapidMiner με Twitter account. Βήμα 2ο



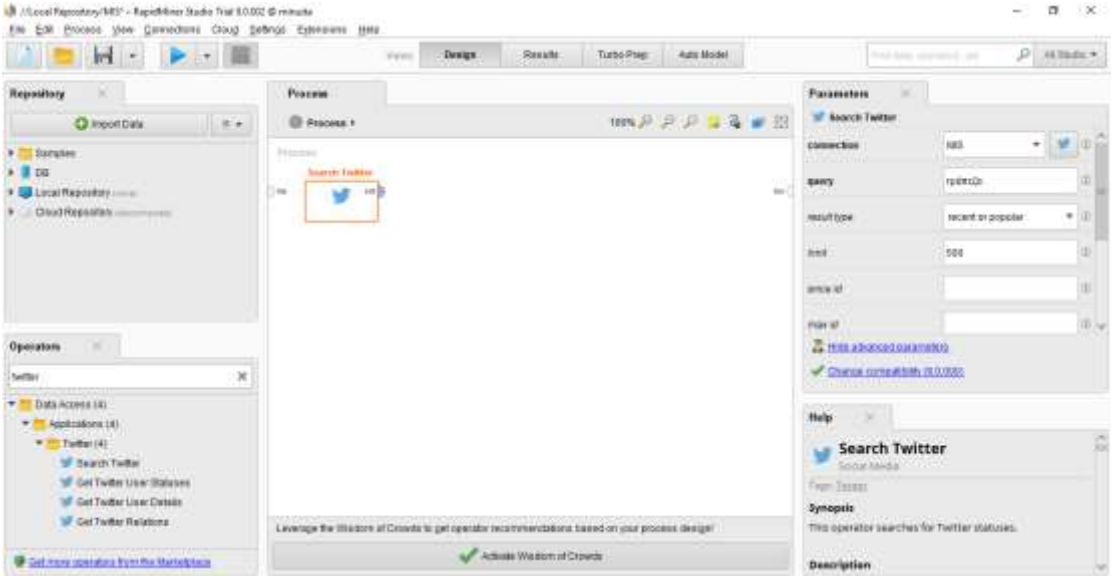
Εικόνα 13: Σύνδεση RapidMiner με Twitter account. Βήμα 3ο



Εικόνα 14: Σύνδεση RapidMiner με Twitter account. Βήμα 4ο



Εικόνα 15: Σύνδεση RapidMiner με Twitter account. Βήμα 5ο



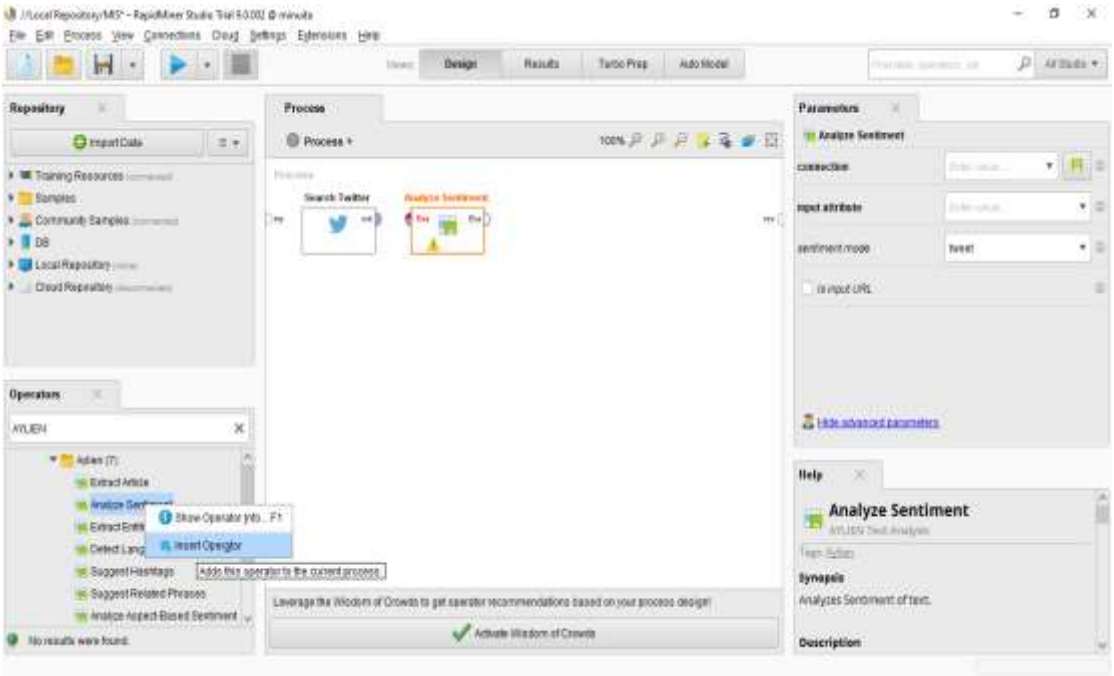
Εικόνα 16: Εισαγωγή λέξεων κλειδιά και ορίου αποτελεσμάτων

Για να προσθέσουμε tweets βάλαμε στο πεδίο query ως κριτήριο αναζήτησης «τράπεζα» και επιλέξαμε ως όριο το 500 για να μας δώσει όσο το δυνατόν μεγαλύτερο αριθμό αποτελεσμάτων, χωρίς καθυστερήσεις στην επεξεργασία (εικόνα 16). Στη συνέχεια κάναμε αναζήτηση στους operators για «text analysis» και προσθέσαμε το AYLIEN (εικόνες 17 & 18), στο οποίο όπως είπαμε νωρίτερα, έχουμε ήδη δημιουργήσει προσωπικό λογαριασμό.

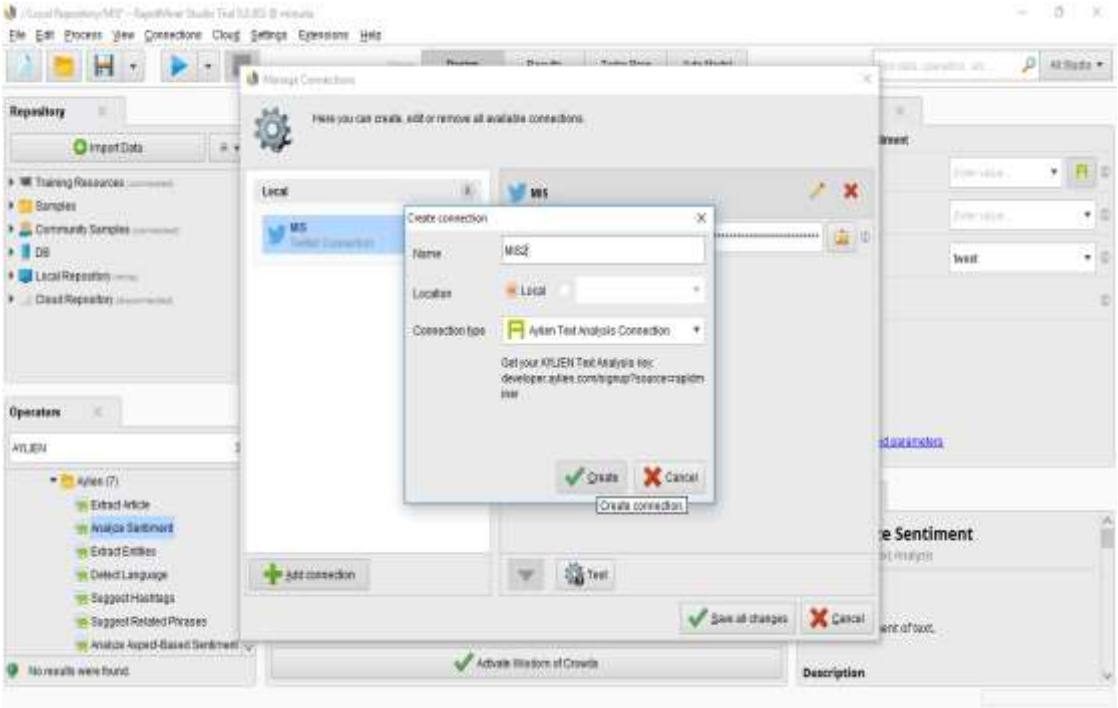
Εικόνα 17: Προσθήκη του operator AYLIEN στο RapidMiner. Βήμα 1ο

Εικόνα 18: Προσθήκη του operator AYLIEN στο RapidMiner. Βήμα 2ο

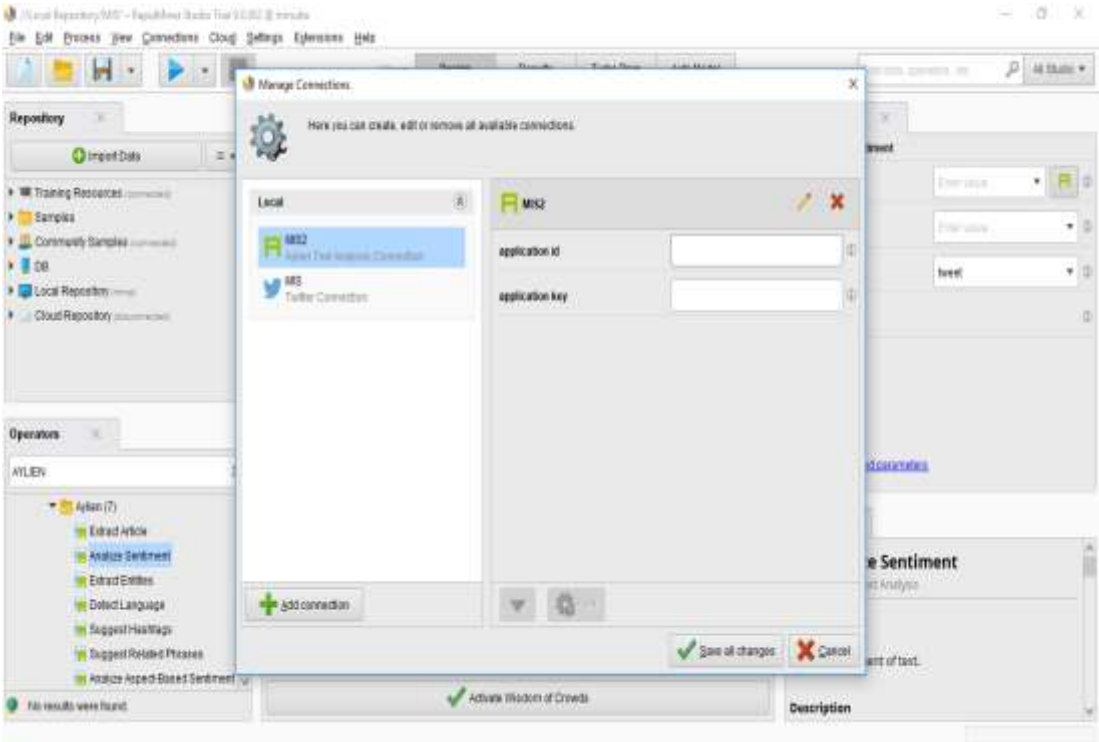
Έπειτα από την παράμετρο connection, ακολουθήσαμε τα βήματα για να το συνδέσουμε με το λογαριασμό μας στο AYLIEN (εικόνες 19 έως 22).



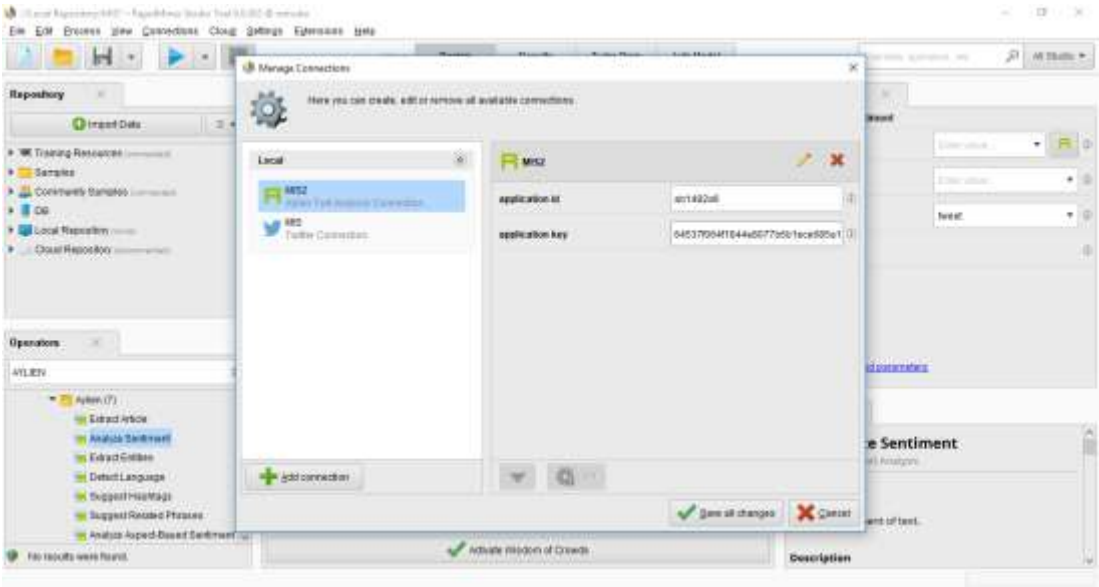
Εικόνα 19: Σύνδεση RapidMiner με AYLIEN account. Βήμα 1ο



Εικόνα 20: Σύνδεση RapidMiner με AYLIEN account. Βήμα 2ο

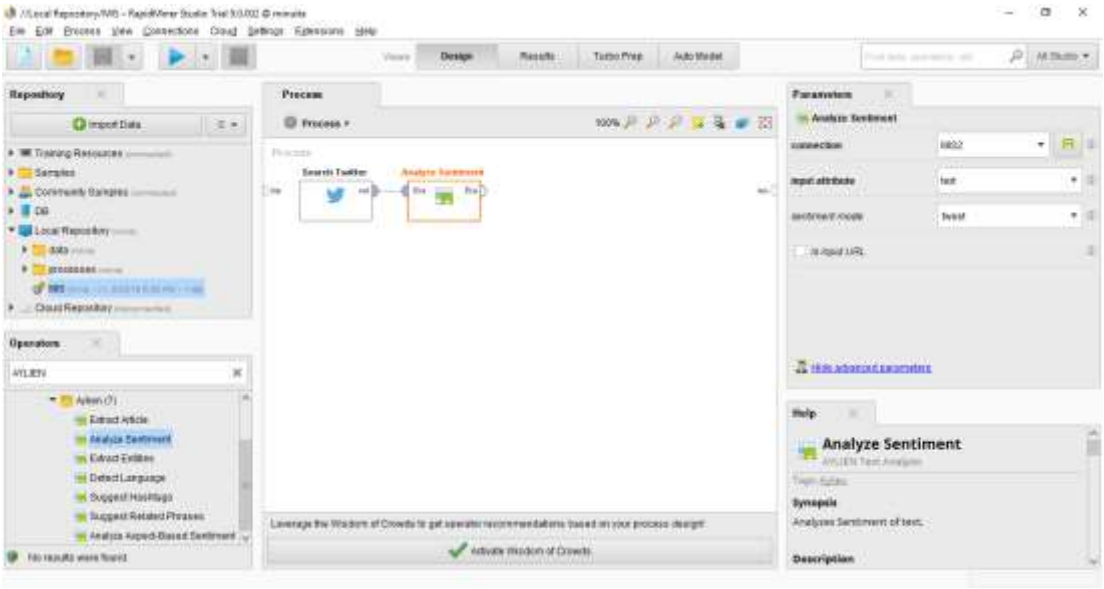


Εικόνα 21: Σύνδεση RapidMiner με AYLIEN account. Βήμα 3ο



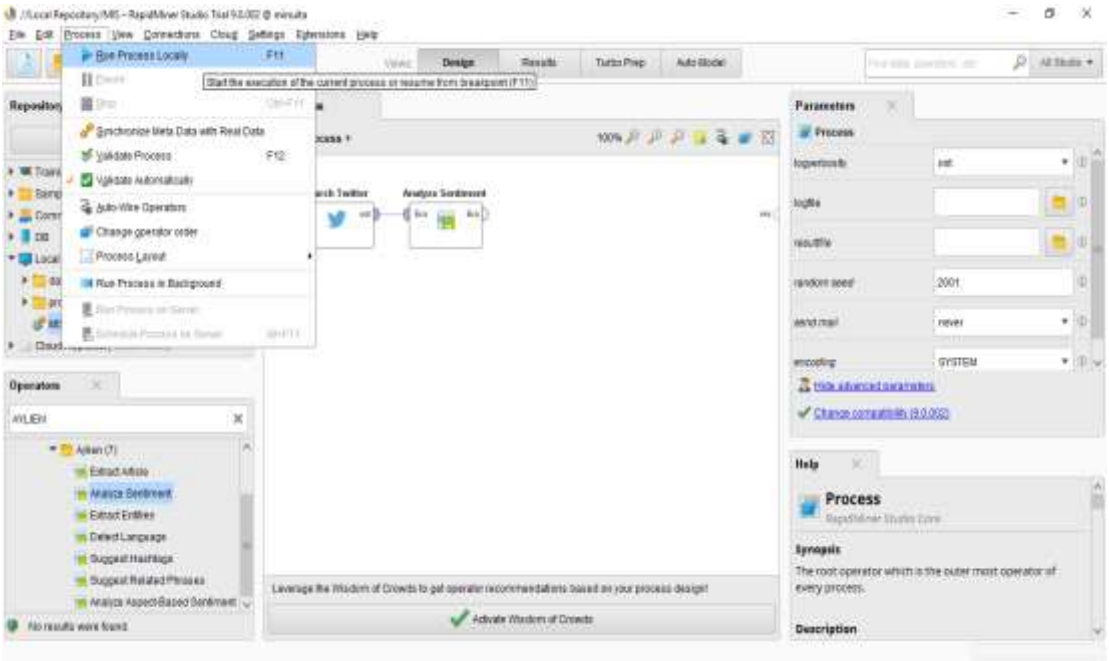
Εικόνα 22: Σύνδεση RapidMiner με AYLIEN account. Βήμα 4ο

Στις παραμέτρους του operator, στο πεδίο «input attribute» επιλέξαμε text και στο πεδίο «sentiment mode» επιλέξαμε tweet (εικόνα 23).

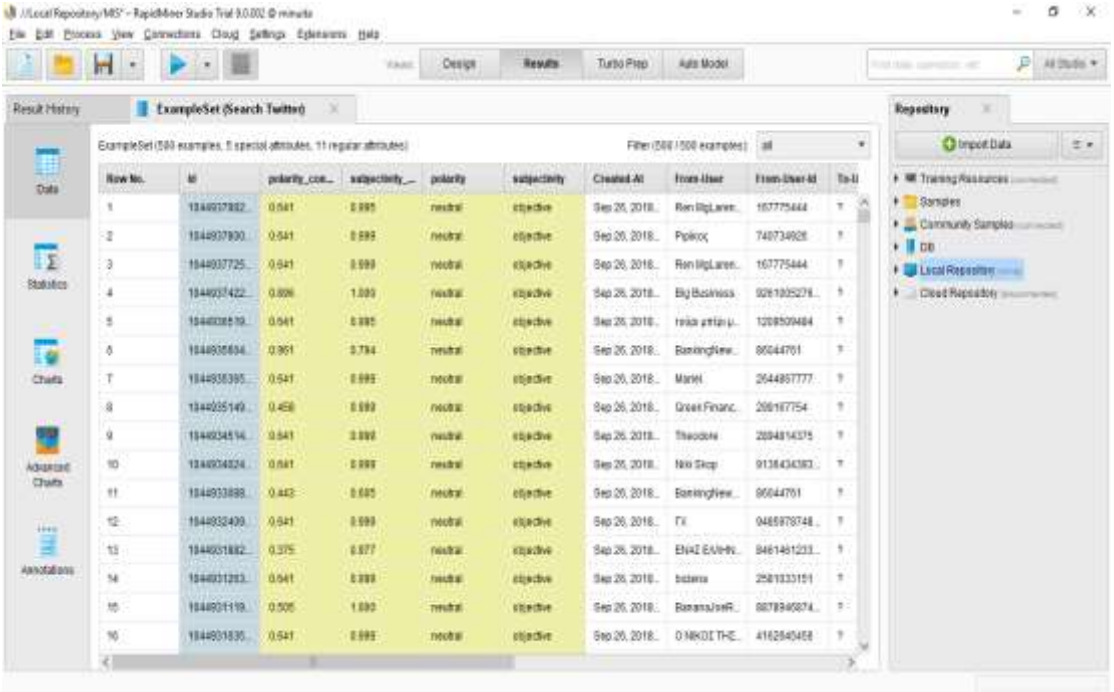


Εικόνα 23: Εισαγωγή παραμέτρων διαδικασίας σύνδεσης των operators Search Twitter & AYLIEN

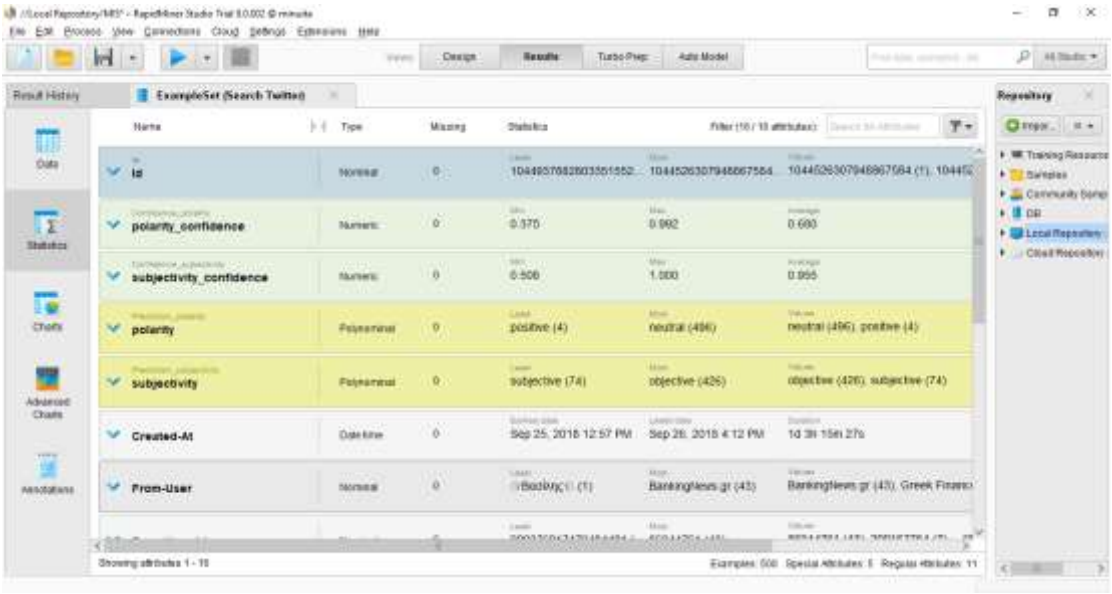
Τέλος, από το menu process επιλέξαμε «run process» (εικόνα 24) και το σύστημα μας εμφάνισε τα αποτελέσματα με τα δεδομένα μας, καθώς και σχετικά στατιστικά στοιχεία και γραφήματα (εικόνα 25).



Εικόνα 24: Έναρξη διαδικασίας εξαγωγής tweets και ανάλυσης του συναισθήματος

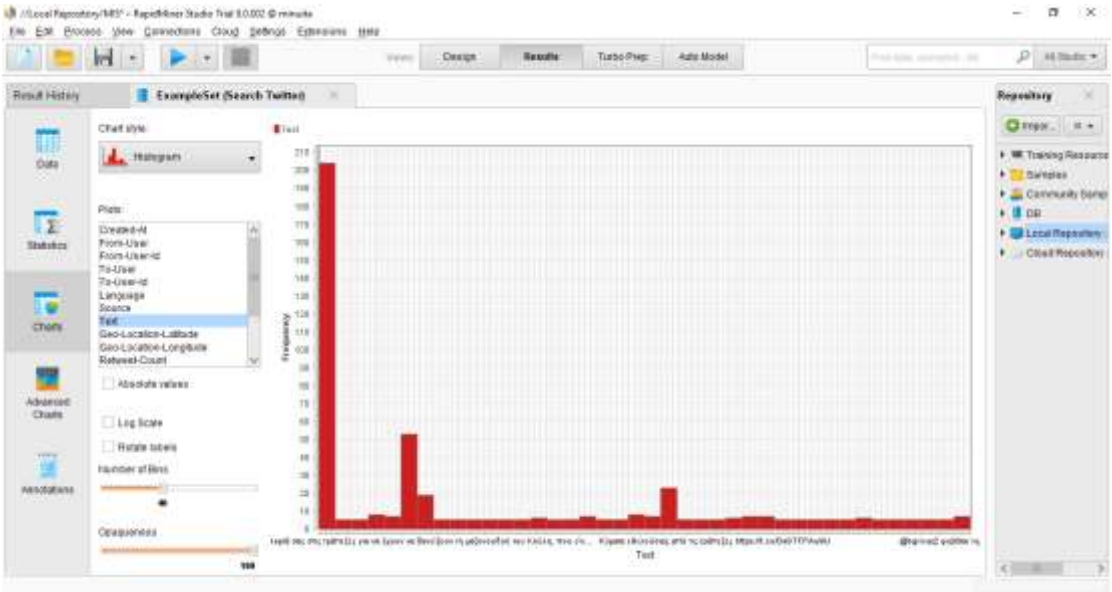


Εικόνα 25: Αποτέλεσμα εξαγωγής tweets και ανάλυσης του συναισθήματος



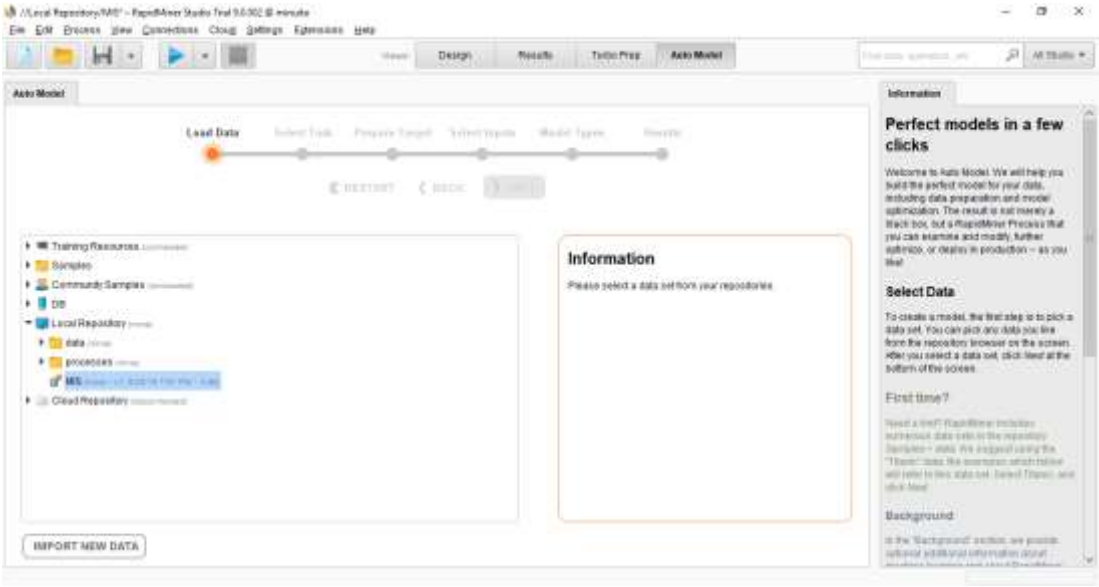
Εικόνα 26: Αποτέλεσμα εξαγωγής tweets και ανάλυσης του συναισθήματος. Κατηγοριοποίηση κειμένου.

Στα στατιστικά στοιχεία που προκύπτουν (εικόνα 26), παρατηρούμε ότι προκύπτει χαμηλή εμπιστοσύνη στην εκτίμηση της ανάλυσης συναισθήματος (polarity_confidence), με μέσο όρο 0.68, η οποία κατηγοριοποιεί τη συντριπτική πλειοψηφία των tweets ως ουδέτερα (496), μόλις 4 tweets ως θετικά και κανένα ως αρνητικό. Στην εικόνα 27 βλέπουμε ότι τα περισσότερα tweets αφορούν σε σχόλια σχετικά με το δανεισμό πολιτικού κόμματος στην Ελλάδα.



Εικόνα 27: Πλήθος tweets ανά περιεχόμενο κειμένου

Τώρα για να προσθέσουμε τα στοιχεία του Google Spreadsheet, πηγαίνουμε στην καρτέλα «Auto Model» και επιλέγουμε «import new data» (εικόνα 28).

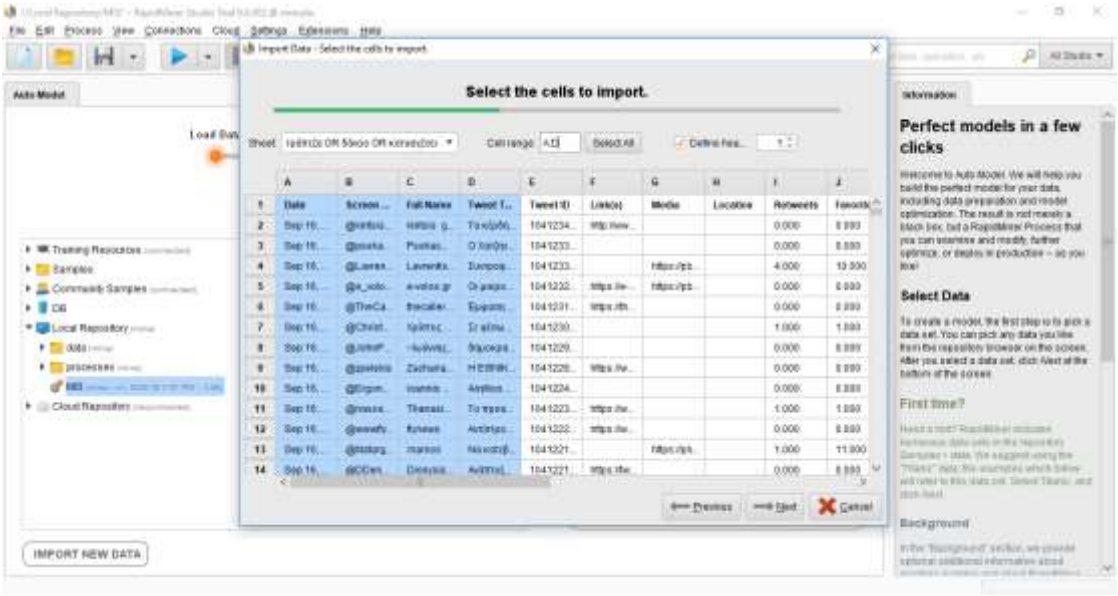


Εικόνα 28: Προσθήκη του Google Spreadsheet στο RapidMiner

Στην ενότητα Load Data ακολουθούμε τα βήματα που χρειάζονται ώστε να κάνουμε browse το αρχείο excel με τα tweets που δημιουργήσαμε χρησιμοποιώντας το Twitter Archiver (εικόνα 29).

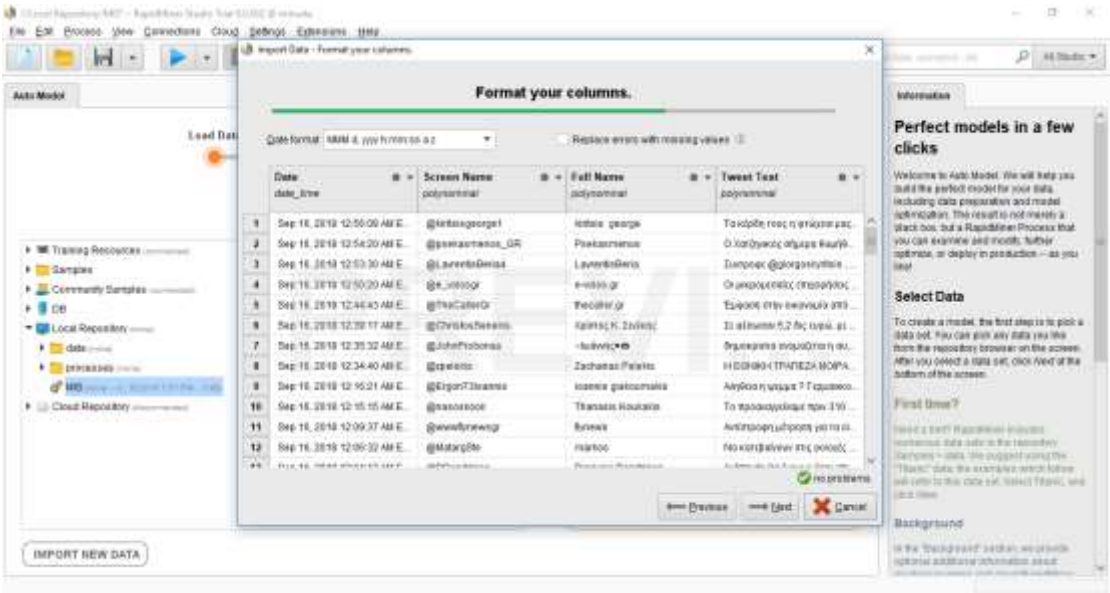
Εικόνα 29: Εισαγωγή του Google Spreadsheet

Επιλέγουμε τις στήλες που μας ενδιαφέρουν. Στην προκειμένη περίπτωση τις στήλες A έως D (εικόνα 30).



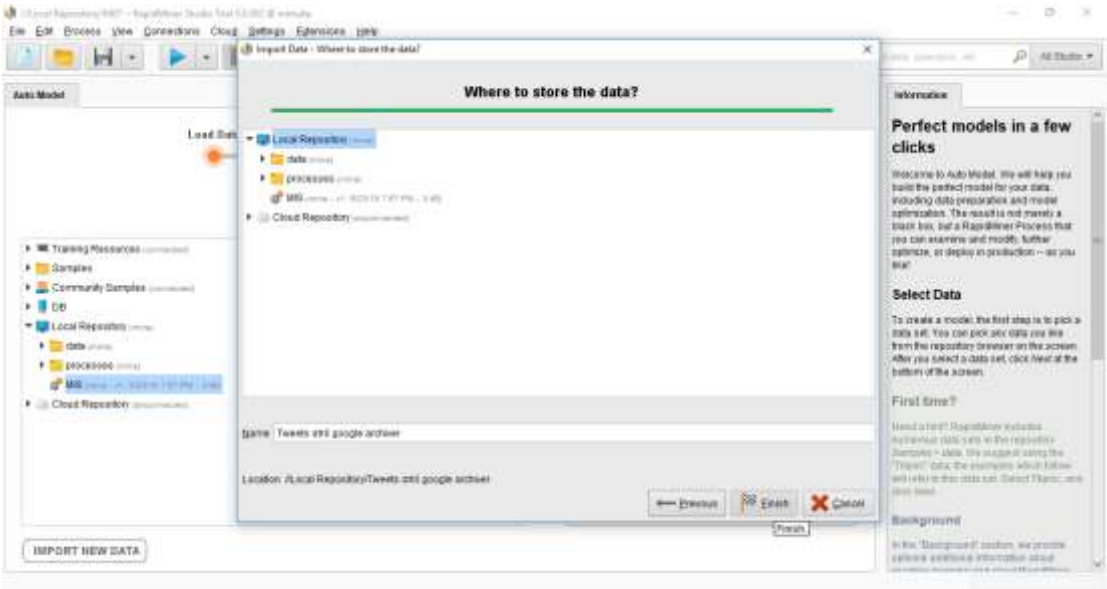
Εικόνα 30: Επιλογή στηλών

Αφού δεν προκύπτει κάποιο πρόβλημα από τον έλεγχο, συνεχίζουμε με next (εικόνα 31).



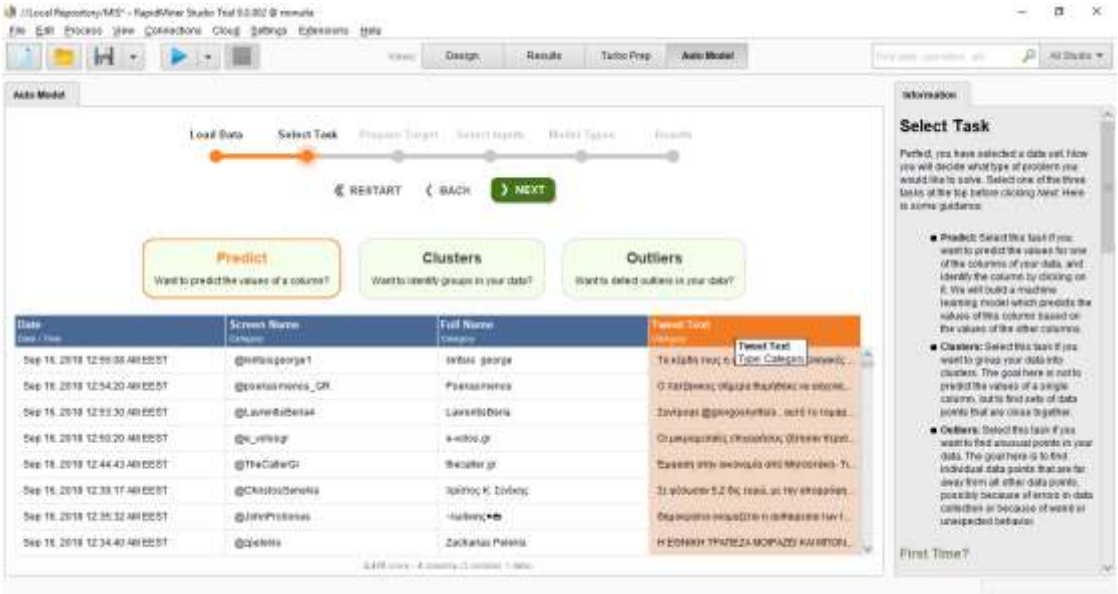
Εικόνα 31: Έλεγχος εισαγωγής δεδομένων

Και τέλος επιλέγουμε το φάκελο processes και πατάμε finish (εικόνα 32).



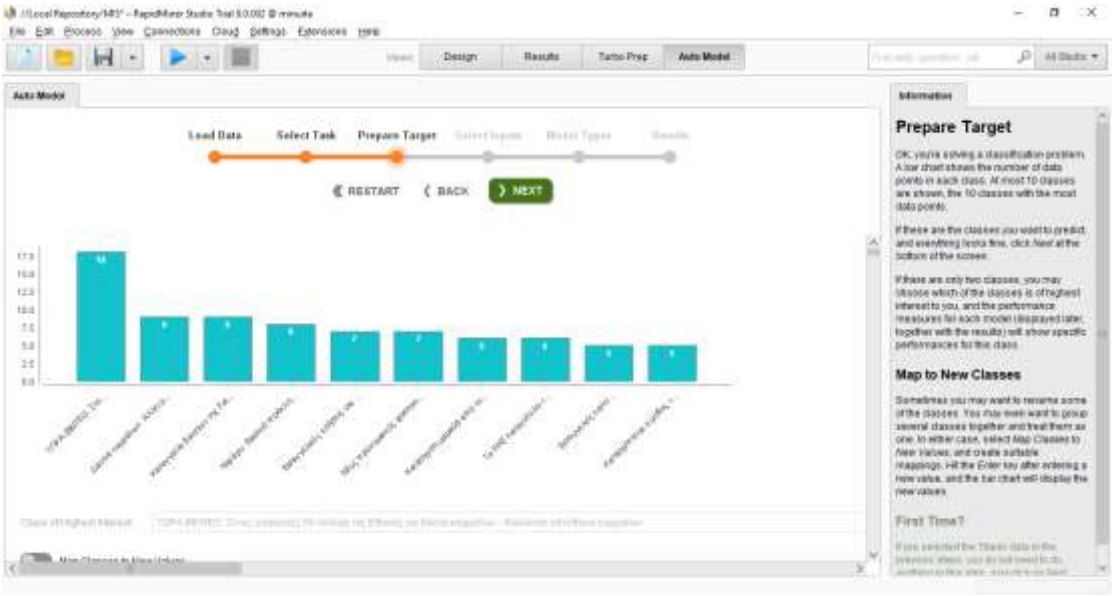
Εικόνα 32: Ολοκλήρωση εισαγωγής δεδομένων

Συνεχίζουμε στην ενότητα Select Task. Επιλέγουμε τη στήλη «tweet text» και πατάμε predict και έπειτα next (εικόνα 33).



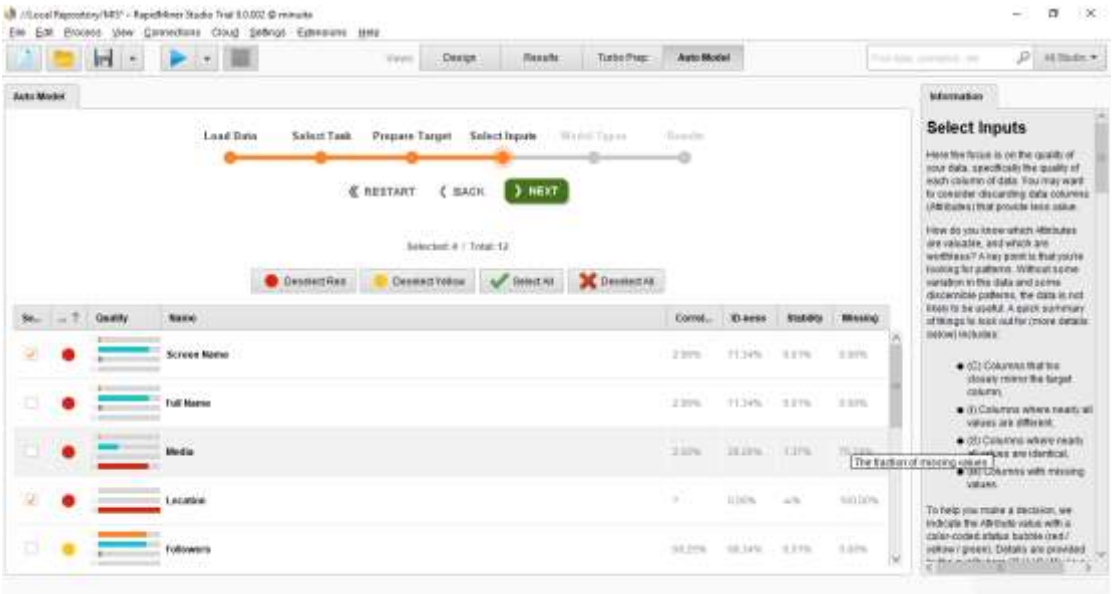
Εικόνα 33: Κατηγοριοποίηση κειμένου. Βήμα 1ο

Στην ενότητα Prepare Target συνεχίζουμε με next (εικόνα 34).



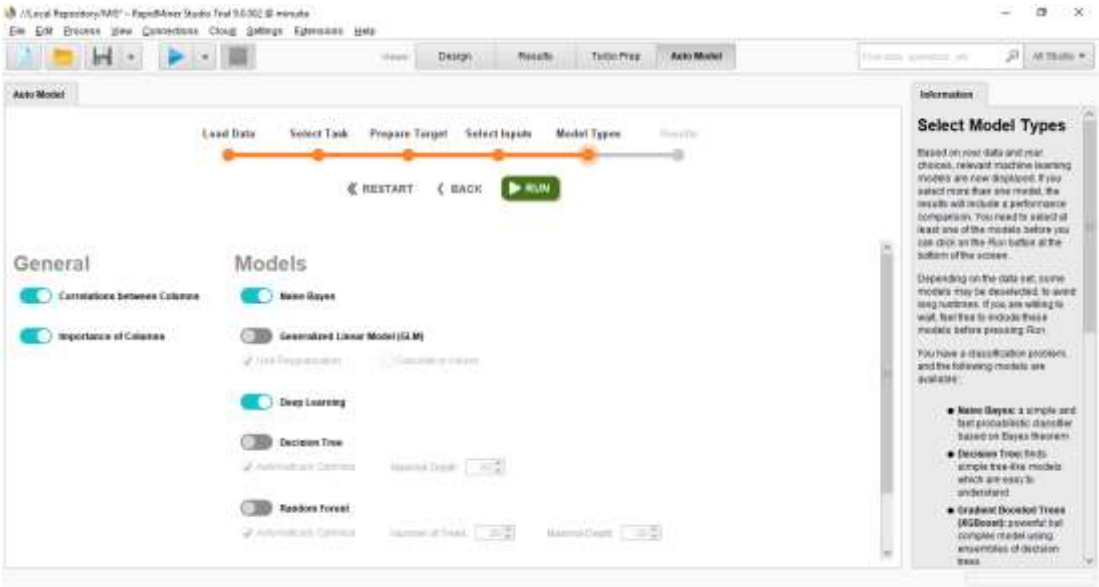
Εικόνα 34: Κατηγοριοποίηση κειμένου. Βήμα 2ο

Στην ενότητα Select Input επιλέγουμε τα quality: screen name, location, date και links. Συνεχίζουμε με next (εικόνα 35).



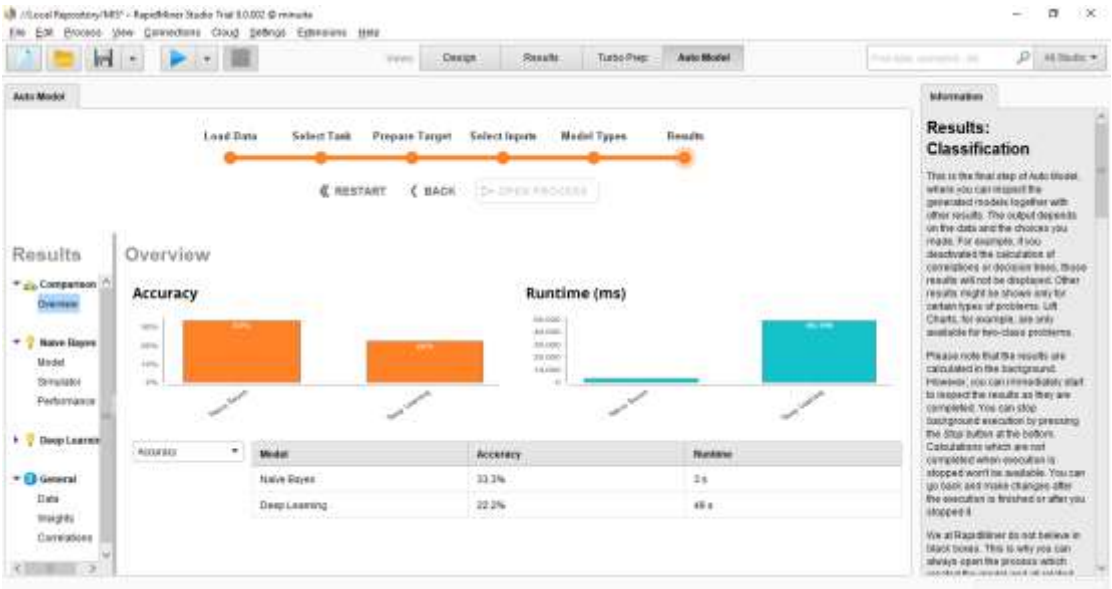
Εικόνα 35: Κατηγοριοποίηση κειμένου. Βήμα 3ο

Στην ενότητα Model Types, για την ταξινόμηση τριών κλάσεων γνώμης (θετική, αρνητική, ουδέτερη), επιλέγουμε τους αλγόριθμους Naïve Bayes και Deep Learning και πατάμε run (εικόνα 36).



Εικόνα 36: Κατηγοριοποίηση κειμένου. Βήμα 4ο

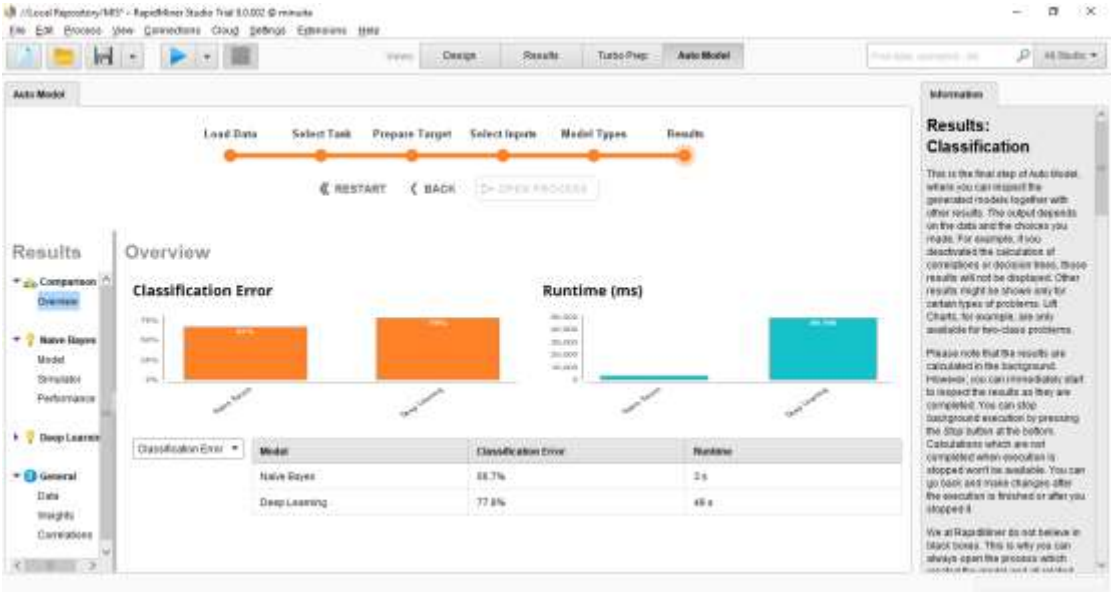
Από τα αποτελέσματα που εξήχθησαν, όπως βλέπουμε και στην εικόνα 37, όσον αφορά στον αλγόριθμο Naïve Bayes, η ακρίβεια του αποτελέσματος ισούται με 33,3%, ενώ όσον αφορά στον αλγόριθμο Deep Learning ισούται με 22,2%. Παρατηρούμε ότι και στις δύο περιπτώσεις έχουμε πολύ χαμηλή αποτελεσματικότητα.



Εικόνα 37: Αποτελεσματικότητα αλγόριθμων Naïve Bayes και Deep Learning

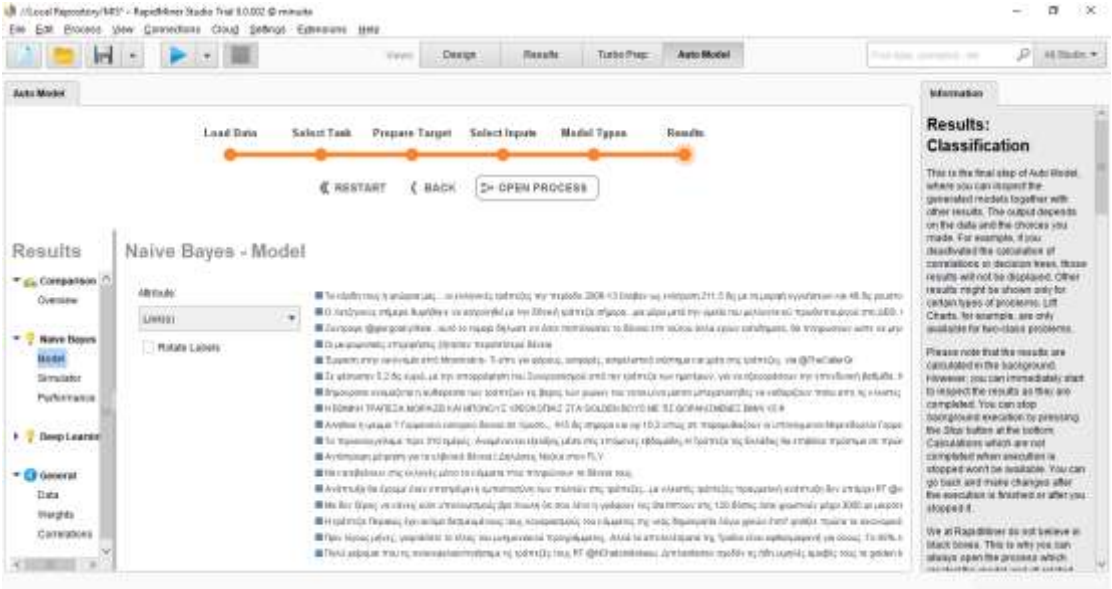
Βλέποντας τα αποτελέσματα από άλλη «οπτική γωνία», δηλαδή από την πλευρά του σφάλματος, βλέπουμε ότι για τον αλγόριθμο Naive Bayes το σφάλμα ανέρχεται

σε 66,7% και για τον αλγόριθμο Deep Learning σε 77,8% (εικόνα 38). Δηλαδή και στις δύο περιπτώσεις είναι πολύ υψηλό.

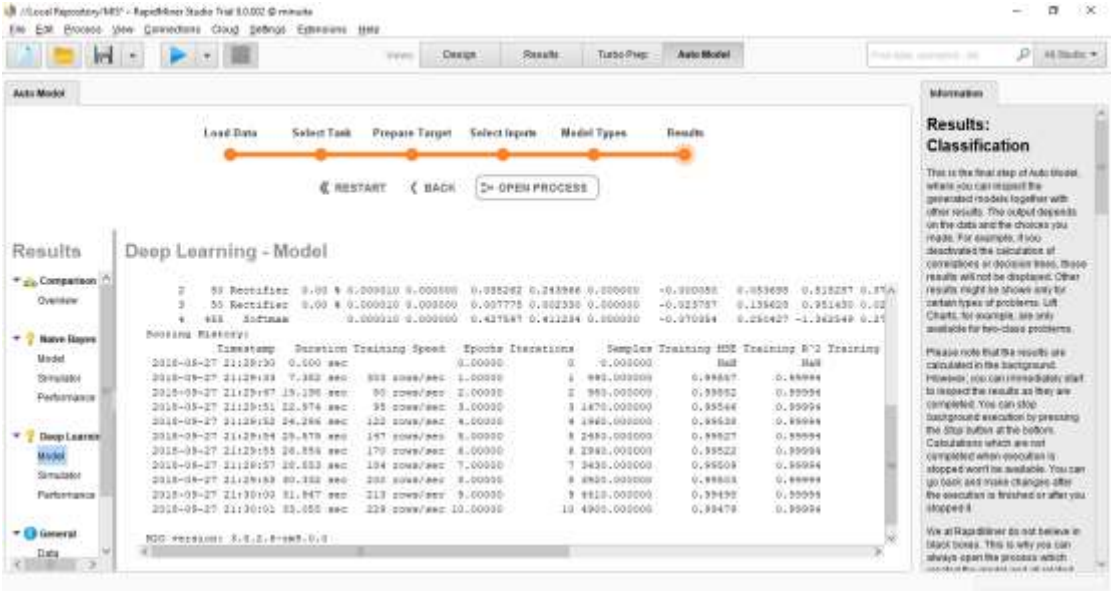


Εικόνα 38: Αποτελεσματικότητα αλγόριθμων Naïve Bayes και Deep Learning

Στην καρτέλα model των αλγόριθμων Naive Bayes και Deep Learning, έχουμε μία γραφική αναπαράσταση του μοντέλου μας (εικόνες 39 & 40).

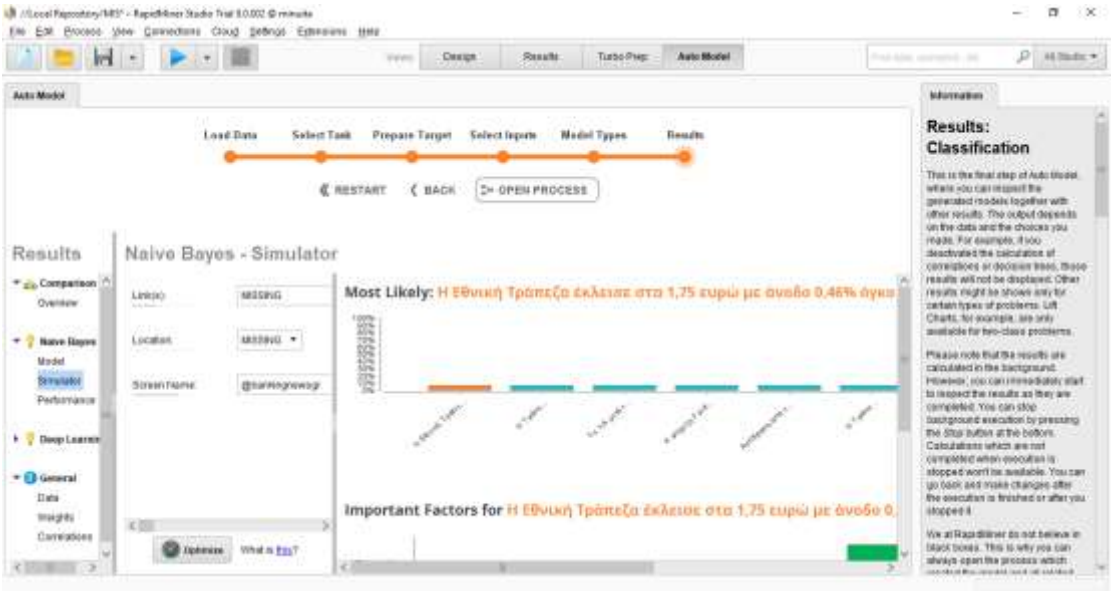


Εικόνα 39: Γραφική αναπαράσταση μοντέλου με τη χρήση του αλγόριθμου Naive Bayes

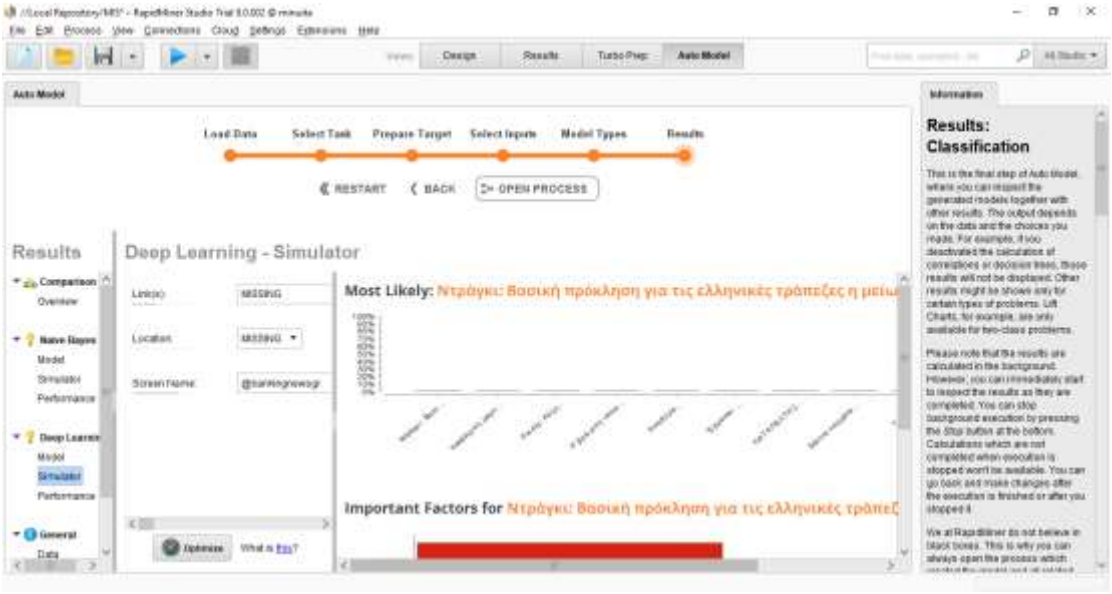


Εικόνα 40: Γραφική αναπαράσταση μοντέλου με τη χρήση του αλγόριθμου Deep Learning

Η οθόνη simulator μας παρέχει μία εύχρηστη και σε πραγματικό χρόνο διασύνδεση, για να αλλάξουμε τις εισόδους στο μοντέλο και να δούμε την έξοδο. Εμφανίζει προβλέψεις, πληροφορίες και επεξηγήσεις για τις εισροές δεδομένων. Στην περίπτωση μας βλέπουμε ότι σύμφωνα με τον αλγόριθμο Naïve Bayes, το πιο σχετικό άρθρο με την αναζήτηση μας αφορά στην τιμή κλεισίματος της μετοχής της Εθνικής Τράπεζας με άνοδο 0,46%, ενώ σύμφωνα με τον αλγόριθμο Deep Learning, αφορά στις δηλώσεις του κ. Ντράγκι για τις προκλήσεις που καλούνται να αντιμετωπίσουν οι ελληνικές τράπεζες (εικόνα 41 & 42).

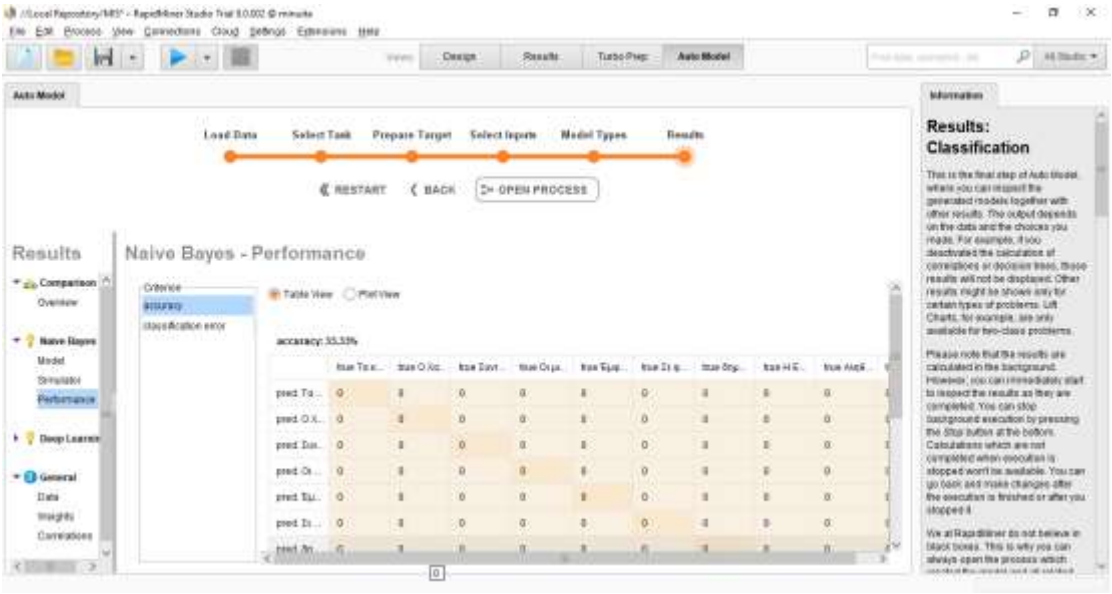


Εικόνα 41: Πλήθος tweets ανά περιεχόμενο κειμένου με τη χρήση του αλγόριθμου Naïve Bayes

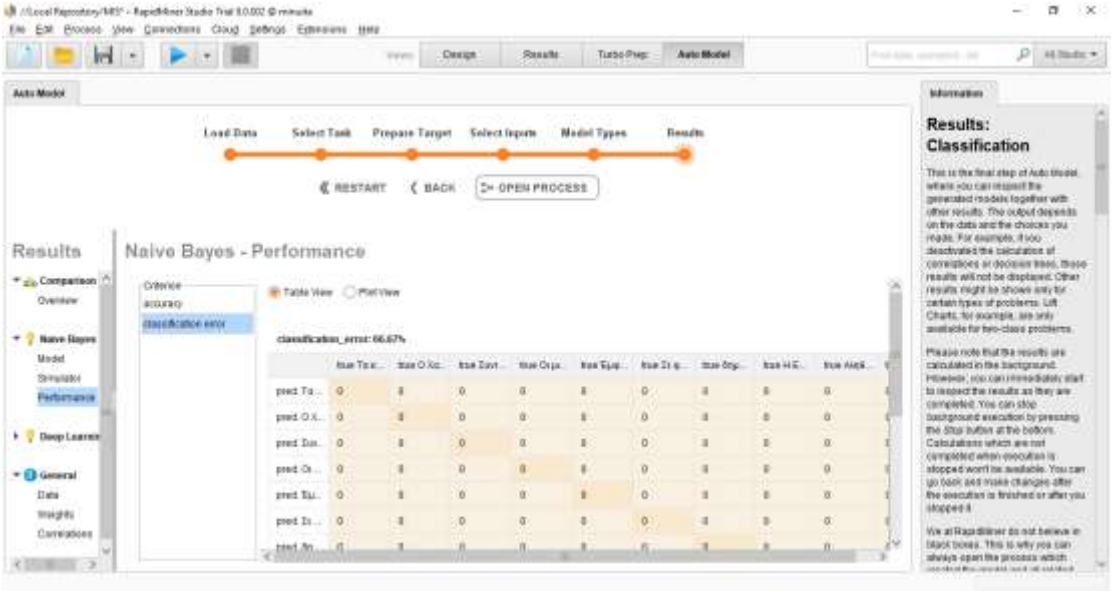


Εικόνα 42: Πλήθος tweets ανά περιεχόμενο κειμένου με τη χρήση του αλγόριθμου Deep Learning

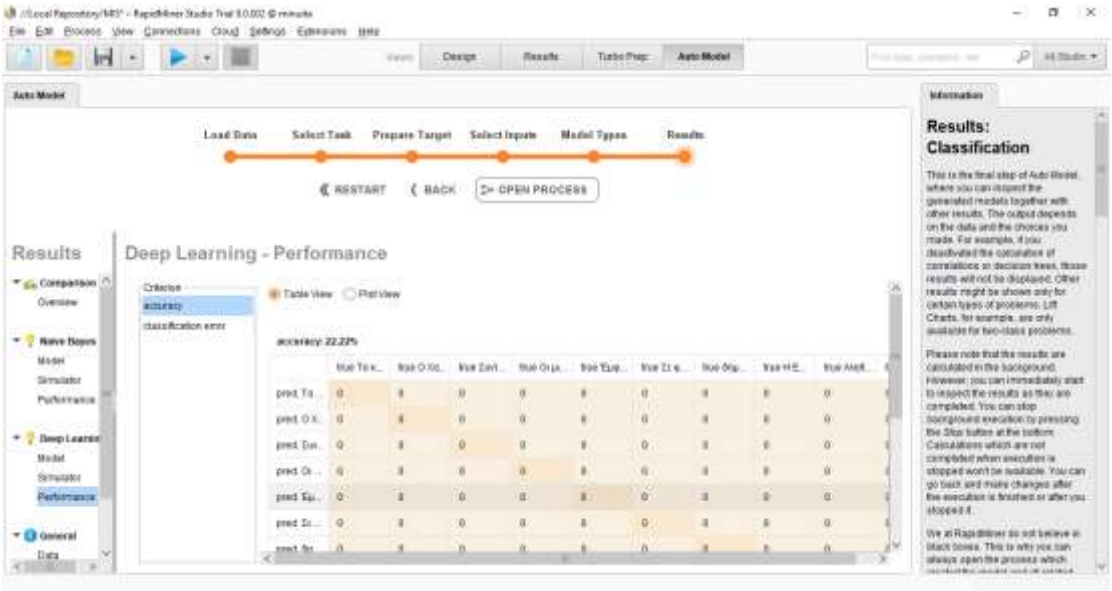
Η οθόνη performance παραθέτει την ακρίβεια πρόβλεψης του μοντέλου και άλλα κριτήρια απόδοσης, ανάλογα με το είδος του προβλήματος ταξινόμησης (εικόνες 43 έως 46).



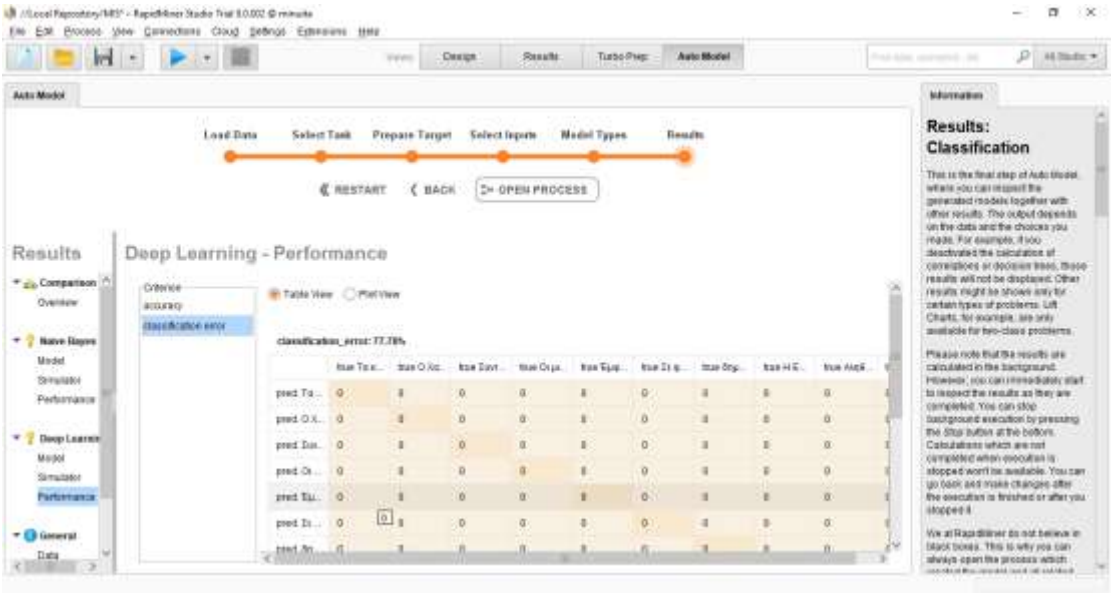
Εικόνα 43: Αποτελεσματικότητα του μοντέλου με τη χρήση του αλγόριθμου Naïve Bayes



Εικόνα 44: Κριτήρια απόδοσης του μοντέλου με τη χρήση του αλγόριθμου Naïve Bayes



Εικόνα 45: Αποτελεσματικότητα του μοντέλου με τη χρήση του αλγόριθμου Deep Learning



Εικόνα 46: Κριτήρια απόδοσης του μοντέλου με τη χρήση του αλγόριθμου Deep Learning

Η υπερπληροφόρηση η οποία χαρακτηρίζει το διαδίκτυο, δημιουργεί αλλοίωση του σήματος και βρίσκεται στη βάση της δημιουργίας φήμης και μη στοχευόμενης πληροφορίας στην αγορά. Αυτή η αλλοιωμένη πληροφορία, αρχικά δημιουργεί προβλήματα σε corporate level, καθώς οι εταιρείες δε μπορούν να οδηγηθούν σε ασφαλή συμπεράσματα από την ανάλυση συναισθήματος. Η συγκέντρωση πληροφοριών όμως μπορεί να επηρεάσει και την επενδυτική συμπεριφορά οικονομούντων ατόμων, όπου συμπεριλαμβάνονται και τα χρηματοοικονομικά, ασφαλιστικά και επενδυτικά προϊόντα.

Στόχος μας στο κεφάλαιο 2, είναι να ελέγξουμε τις προεκτάσεις της διόγκωσης της πληροφορίας στη λιανική τραπεζική.

Κεφάλαιο 2: Συμπεριφορική Χρηματοοικονομική

2.1. Γενικά περί Συμπεριφορικής Χρηματοοικονομικής

Τις τελευταίες δεκαετίες, οι έρευνες των χρηματοοικονομικών βασιζόταν στην υπόθεση της αποτελεσματικότητας της αγοράς και των ορθολογικών επενδυτών, χωρίς να λαμβάνουν υπόψη την ψυχολογία τους κατά τη λήψη επενδυτικών αποφάσεων. Ωστόσο η άποψη αυτή τέθηκε υπό αμφισβήτηση από την πλευρά των ψυχολόγων, οι οποίοι υποστήριξαν ότι οι επενδυτές είναι σε γενικές γραμμές μη ορθολογικοί, λόγω μη προβλέψιμων και καταστροφικών σφαλμάτων που αποδίδονται σε ψυχολογικούς παράγοντες. Το μέγεθος του κινδύνου που μπορεί να αναλάβει ένας επενδυτής βασίζεται κυρίως στα συναισθήματα του φόβου, της ελπίδας και της απληστίας. Ο φόβος είναι αυτός που ωθεί τους επενδυτές να εστιάσουν σε αρνητικά γεγονότα ή ενδεχόμενα, όμως εν τέλει αντιδρούν υποκινούμενοι περισσότερο από την ελπίδα για ένα θετικό αποτέλεσμα, παρά από την απληστία (Shifrin, 2000).

Έτσι, ακριβώς επειδή η παραδοσιακή χρηματοοικονομική δεν ήταν σε θέση να εξηγήσει ορισμένα φαινόμενα, δημιουργήθηκε ένας νέος κλάδος, η συμπεριφορική χρηματοοικονομική. Πρόκειται για ένα κλάδο που αναπτύσσεται ραγδαία και συνδυάζει την έρευνα της συμπεριφοράς με την χρηματοοικονομική, προσπαθώντας να αιτιολογήσει τις αποφάσεις των επενδυτών ερευνώντας τους κοινωνιολογικούς και συναισθηματικούς παράγοντες που τις επηρεάζουν (Σπύρου, 2009).

2.2. Ψυχολογία και κανόνες λήψης αποφάσεων

Όπως ήδη αναφέραμε, η ψυχολογία αποτελεί τη βάση της μη ορθολογικής αγοράς και θεμέλιο λίθο της συμπεριφορικής χρηματοοικονομικής, σύμφωνα με την οποία η λήψη αποφάσεων από τους επενδυτές βασίζεται σε ορισμένους κανόνες (Barberis & Thaler, 2003). Παρακάτω θα περιγράψουμε τους ψυχολογικούς παράγοντες που επηρεάζουν τη λήψη αποφάσεων.

Υπεραυτοπεποίθηση. Σύμφωνα με έρευνες, τα άτομα έχουν την τάση να υπερεκτιμούν τις δυνατότητες τους και την ικανότητα τους να προβλέπουν γεγονότα. Η υπερβολική εμπιστοσύνη στον εαυτό τους, μπορεί να οδηγήσει τους επενδυτές σε λάθος εκτιμήσεις και κατά συνέπεια στη λήψη λανθασμένων αποφάσεων.

Συντηρητισμός. Αποτελεί πτυχή της προσωπικότητας του ατόμου και μπορεί να επηρεάσει τις ενέργειες του. Πρόκειται για την καθυστερημένη ή ακόμη και αδύνατη αλλαγή απόψεων και αποφάσεων, όταν προκύπτει κάποια νέα πληροφορία και βασίζεται στο φόβο των ατόμων για το άγνωστο.

Συμπεριφορά αγέλης. Πρόκειται για την εναρμόνιση των ατόμων με την πλειοψηφία όσων λαμβάνουν αποφάσεις για το ίδιο θέμα. Σύμφωνα με τον LeBon (1895), η ατομική ψυχολογία επηρεάζεται από την ψυχολογία της μάζας και επίσης, όπως αναφέρει ο Freud στο βιβλίο του «Ψυχολογία των μαζών και ανάλυση του εγώ» (1921), τα άτομα τείνουν να μιμούνται το κοινωνικό σύνολο. Η συμπεριφορά αυτή θα αναλυθεί περισσότερο στο κεφάλαιο 2.4.1. «Θεωρία της ψυχολογίας των μαζών».

Υπεραισιοδοξία. Οι αισιόδοξοι επενδυτές θεωρούν ότι δε θα συμβούν σε αυτούς γεγονότα που θα οδηγήσουν μία επένδυση να αποδειχθεί κακή. Δεν είναι σε θέση να αντιληφθούν ότι μπορεί να συμβεί το αντίθετο από αυτό που περιμένουν, με αποτέλεσμα να οδηγηθούν σε καταστροφικές για το χαρτοφυλάκιο τους αποφάσεις.

Προσκόλληση στις πεποιθήσεις. Τα άτομα εμμένουν απόλυτα και για μεγάλο χρονικό διάστημα στην άποψη που έχουν σχηματίσει και δεν είναι πρόθυμα να ερευνήσουν την πιθανότητα να υπάρχουν στοιχεία που έρχονται σε αντίθεση με τα πιστεύω τους.

2.3 Οικονομική ετερογένεια και είδη επενδυτών

Ως ετερογένεια στην οικονομία θεωρούνται οι διαφορές μεταξύ των οικονομικών μονάδων. Αν για παράδειγμα οι προς μελέτη οικονομικές μονάδες διαφέρουν μεταξύ τους, αν δηλαδή υπάρχουν παράγοντες ετερογένειας, έχουμε ένα διαμέτρου αντίθετο μοντέλο απ’ το αν αυτές είναι ταυτόσημες (Arellano, 2009). Στην πραγματική οικονομία υπάρχει διαφορά στις προσδοκίες μεταξύ των επενδυτών και στα κόστη των συναλλαγών και κατά συνέπεια λαμβάνοντας υπόψη την οικονομική ετερογένεια μπορούμε να κατανοήσουμε καλύτερα τις εμπειρικές ιδιότητες των χρηματιστηριακών τιμών (Chiarella et. al 2007).

Σύμφωνα με τον Muth (1691), οι οικονομικές μονάδες προσπαθούν να αξιοποιήσουν τις διαθέσιμες πληροφορίες με τον πιο αποδοτικό τρόπο, για να σχηματίσουν τις προσδοκίες τους και να λάβουν τις μελλοντικές τους αποφάσεις. Ο John Muth ήταν ο πρώτος που υιοθέτησε τη θεωρία των ορθολογικών προσδοκιών

στην οικονομία, η κεντρική άποψη της οποίας είναι πως το σύνολο των επενδυτών λειτουργεί με πανομοιότυπο τρόπο χωρίς να κάνει συστηματικά λάθη και πως όλες οι πληροφορίες είναι διαθέσιμες σε όλους. Αν και σύμφωνα με τη θεωρία των αποτελεσματικών αγορών, οι διαθέσιμες πληροφορίες ενσωματώνονται πλήρως στις τιμές και άρα η χρηματιστηριακή και η πραγματική αξία των μετοχών είναι ίσες, ωστόσο η επεξεργασία των πληροφοριών διαφοροποιείται μεταξύ των επενδυτών και έτσι προκύπτει σημαντικός βαθμός ετερογένειας (Fama, 1970).

Βασικό χαρακτηριστικό της οικονομικής ετερογένειας είναι οι διαφορετικές αντιλήψεις των επενδυτών στο μέγεθος του ρίσκου που μπορούν να αναλάβουν και ως αποτέλεσμα αυτού επηρεάζεται το ασφάλιστρο κινδύνου. Κατά συνέπεια, ο σταθμισμένος μέσος όρος των τιμών ισορροπίας των ετερογενών οικονομικών μονάδων, αντικατοπτρίζεται στις τιμές εξισορρόπησης της αγοράς (Chiarella et. al 2007). Από τα παραπάνω μπορούμε να συμπεράνουμε πως η ετερογένεια των επενδυτών είναι ικανή να επηρεάσει τις τιμές των μετοχών και άρα την απόδοση της αγοράς. Παρόλα αυτά, σύμφωνα με το Hommes (2005), δεν είναι δυνατόν να γνωρίζουμε την τιμή ισορροπίας μιας ετερογενούς αγοράς, διότι δεν είμαστε σε θέση να ξέρουμε τις πεποιθήσεις όλων των υπόλοιπων επενδυτών.

Η περιορισμένη αντίληψη των επενδυτών για την οικονομική αγορά και το μεγάλο κόστος της έρευνας και εν τέλει απόκτησης έγκυρων πληροφοριών, κάνουν ιδιαίτερα δύσκολη τη δημιουργία βέλτιστων χαρτοφυλακίων και τους ωθούν σε εμπειρικές στρατηγικές. Μεγάλη μερίδα επενδυτών σε περιόδους είτε οικονομικής άνθησης, είτε οικονομικής κρίσης, δεν παραμένουν συνεπείς στις επενδυτικές τους επιλογές και δεν εφαρμόζουν οικονομικές αναλύσεις πριν την αγορά αξιογράφων. Οι οικονομικές αγορές διαμορφώνονται από δύο μεγάλες κατηγορίες επενδυτών, τους ορθολογικούς επενδυτές και τους noise traders (Hommes, 2005).

Ορθολογικοί επενδυτές. Θεωρούν ότι η τιμή ενός αξιογράφου εξαρτάται αποκλειστικά και μόνο από την παρούσα αξία των μελλοντικών του μερισμάτων. Είναι πολύ καλοί γνώστες της αγοράς και αγοράζουν κυρίως υποτιμημένα αξιόγραφα πιστεύοντας στις θεμελιώδης οικονομικές αξίες και πως οι τιμές των μετοχών θα κινηθούν προς αυτές. Όταν αντίστοιχα ένα αξιόγραφο είναι υπερτιμημένο, δηλαδή πάνω από τις θεμελιώδης οικονομικές αξίες, προβαίνουν στην πώληση του. Στην πράξη, οι ορθολογικοί επενδυτές επενδύουν όταν ασκούνται αρνητικές πιέσεις στην αγορά και αντιστρόφως. Έτσι, όσο οι τιμές κινούνται γύρω από τις θεμελιώδεις αξίες, οι επενδύσεις τους έχουν καλές αποδόσεις, ενώ αν οι στρατηγικές των υπολοίπων

επενδυτών αλλάζουν διαρκώς, μπορούν να αποβούν ζημιογόνες με αποτέλεσμα να αναγκάζονται και αυτοί να προσαρμοστούν. Άλλωστε είναι πρακτικά αδύνατον ένας επενδυτής να κατέχει την απόλυτα ορθή γνώση των οικονομικών πληροφοριών, την οποία ακόμη κι αν είχε, δε θα ήταν σε θέση να προβλέψει τις στρατηγικές όλων των υπόλοιπων επενδυτών.

Noise traders. Επηρεάζονται από τις στρατηγικές άλλων επενδυτών και παρασύρονται από την εκάστοτε μόδα που επικρατεί στην αγορά. Βασίζονται σε προγνωστικά με ιστορικές τιμές, χωρίς να λαμβάνουν υπόψη τους τις θεμελιώδης αξίες. Με απλά λόγια προσδοκούν ότι οι σημερινές καταστάσεις θα επαναληφθούν στο μέλλον και έτσι αγοράζουν με την άνοδο των τιμών και πωλούν ενώ έχει ήδη αρχίσει η πτώση τους. Συνήθως δεν έχουν επαρκείς γνώσεις των αγορών και προσπαθούν να ενημερωθούν για αυτές από συναλλαγές που έχουν πραγματοποιήσει άλλοι επενδυτές. Πιστεύουν ότι οι πληροφορίες που λαμβάνουν είναι σωστές, οπότε σχηματίζουν λανθασμένες και ανορθολογικές προσδοκίες. Ως αποτέλεσμα, προκύπτουν διακυμάνσεις και ανωμαλίες με χαρακτηριστικά ετερογένειας (Chiarella & He, 2002). Οι noise traders βασιζόμενοι στην ψυχολογία για τη διαμόρφωση της επενδυτικής τους στρατηγικής, έχουν βραχυπρόθεσμες προσδοκίες και κατά συνέπεια, αυξάνεται ο κίνδυνος που προκύπτει από τις συνεχείς αλλαγές των στρατηγικών. Καθώς αυξάνονται οι λανθασμένες προσδοκίες, αυξάνονται και οι αμφιβολίες, γεγονός το οποίο αποθαρρύνει του ορθολογικούς επενδυτές να προχωρήσουν σε συναλλαγές. Το συμπέρασμα αυτών είναι ότι όταν σε μία αγορά συμμετέχουν πολλοί noise traders, αναμένονται και μεγαλύτερες αποδόσεις (Lee et al., 2002).

2.4 Φήμες και χρηματοοικονομική συμπεριφορά

Η αξιολόγηση των πληροφοριών και ο τρόπος με τον οποίο μία νέα πληροφορία επηρεάζει τη σύσταση ενός χαρτοφυλακίου, αποτελούν μία από τις βασικές αρχές της χρηματοοικονομικής. Οι φήμες είναι μία μορφή πληροφορίας, επηρεάζονται από ψυχοκοινωνιολογικούς παράγοντες και μπορούν να οδηγήσουν σε συναισθήματα όπως η υπεραισιοδοξία, ο φόβος ή το άγχος, οδηγώντας τους επενδυτές σε μη ορθολογικές αποφάσεις. Η συμπεριφορική χρηματοοικονομική αναλύει τις συνθήκες κάτω από τις οποίες τόσο ένας μεμονωμένος επενδυτής, όσο και το σύνολο της

αγοράς, λαμβάνει μία απόφαση που δεν είναι ορθολογική (Αλεξάκης & Ξανθάκης, 2008).

Οι αγορές είναι ευαίσθητες στις φήμες, διότι οι συναλλαγές βασίζονται στις πληροφορίες και η πρόσβαση σε προνομιακές πληροφορίες μπορεί να αποφέρει τεράστια κέρδη σε έναν επενδυτή. Μερικοί από τους βασικούς παράγοντες που κάνουν τις αγορές ευαίσθητες στις φήμες είναι ότι οι συμμετέχοντες στις αγορές είναι περιορισμένοι, αξιολογούν τις πληροφορίες που λαμβάνουν, αν και σε συνθήκες πίεσης μπορεί να μην έχουν τη δυνατότητα να τις επαληθεύσουν, οπότε και ρισκάρουν να πάρουν μία απόφαση και επίσης κάθε επενδυτής ίσως λάβει υπόψη του μόνο τις πληροφορίες που υποστηρίζουν τις δικές του απόψεις (Αλεξάκης & Ξανθάκης, 2008).

Οι φήμες μπορούν να κατηγοριοποιηθούν ανάλογα με την πηγή και το περιεχόμενο τους. Όταν π.χ. δε δίνονται σαφείς απαντήσεις σε ανοιχτά θέματα, το κοινό διαμορφώνει τη δική του άποψη επί του θέματος και με την επικοινωνία δημιουργείται μία φήμη. Επίσης μπορεί μία φήμη να δημιουργηθεί επίτηδες, διαδίδοντας πληροφορίες με απώτερο σκοπό το κέρδος επί ενός γεγονότος. Οι φήμες μπορεί να είναι γενικές ή συγκεκριμένες, να αποτελούν αποκύημα της φαντασίας ή να περιέχουν πραγματικές πληροφορίες και μπορεί να επηρεάσουν τις τιμές των μετοχών δημιουργώντας τάσεις στην αγορά (Αλεξάκης & Ξανθάκης, 2008).

2.4.1. Θεωρία της ψυχολογίας των μαζών

Ως μάζα χαρακτηρίζεται η κοινή συμπεριφορά μίας ομάδας ατόμων και ως ψυχολογία της μάζας, η παράλληλη διαμόρφωση ψυχολογίας.

Η θεωρία των μαζών αναπτύχθηκε πρώτη φορά το 1895 από τον Gustave LeBon, στην οποία βασίζεται η επιστήμη της κοινωνικής ψυχολογίας. Σύμφωνα με το LeBon, η ατομική ψυχολογία επηρεάζεται από την ψυχολογία της μάζας, κάνοντας τα άτομα που την αποτελούν να συμπεριφέρονται ως μία οντότητα και να σκέφτονται, να νιώθουν και να αντιδρούν εντελώς διαφορετικά από ότι θα έκαναν ως μονάδες. Οι οντότητες που αποτελούν τη μάζα δεν είναι απαραίτητο ούτε να βρίσκονται στο ίδιο σημείο, ούτε καν να έχουν οποιαδήποτε επικοινωνία μεταξύ τους. Είναι αρκετό να σκέφτονται και να αισθάνονται με τον ίδιο τρόπο (LeBon, 1895).

Σύμφωνα με το Freud (1921), ο άνθρωπος τείνει να είναι μιμητικός και η συμπεριφορά του να επηρεάζεται από το κοινωνικό σύνολο του οποίου είναι μέλος.

Επιπλέον, ως μέλος μίας μάζας μπορεί να εμφανίζει διαφορετικά χαρακτηριστικά και να συμπεριφέρεται εντελώς διαφορετικά από ότι θα έκανε ως αυτόνομη προσωπικότητα.

Βασικό χαρακτηριστικό των μαζών είναι η ευκολοπιστία, ο παρορμητισμός, η έντονη φαντασία και η τάση να γενικεύουν καταστάσεις. Γενικά οι μάζες έχουν πολύ πιο χαμηλό επίπεδο λογικής και κριτικής σκέψης σε σχέση με τα μεμονωμένα άτομα. Αν ένα άτομο διαθέτει κριτική σκέψη ώστε να μπορεί να εκτιμήσει μία κατάσταση και το κατά πόσο επηρεάζεται από αυτή, δεν ανήκει στη μάζα (Freud, 1921).

Μια ισορροπημένη προσωπικότητα έχει μικρότερες πιθανότητες να επηρεαστεί από τη συμπεριφορά των γύρω της. Το μορφωτικό επίπεδο και η ωριμότητα ενός ατόμου είναι άμεσα συνδεδεμένη με την ικανότητα του να αναπτύσσει κριτική σκέψη και να αναλαμβάνει τις ευθύνες που προκύπτουν από τις πράξεις του. Παρόλα αυτά επειδή οι μάζες κατακλύζονται από πολύ έντονα συναισθήματα, μπορεί να επηρεαστεί από αυτά ακόμη και ο πιο ψύχραιμος άνθρωπος (Χόρνεϋ, 1990).

2.5. Εμπειρική Έρευνα – Ανάλυση ερωτηματολογίου

Στην παρούσα ενότητα, θα παρουσιάσουμε την έρευνα στην οποία προβήκαμε ώστε να μπορέσουμε να απαντήσουμε στα ερωτήματα της εργασίας. Στόχος της έρευνας είναι να μελετήσει αν τα σχόλια τρίτων στο διαδίκτυο και κυρίως στα social media επηρεάζουν τις αποφάσεις των επενδυτών.

2.5.1. Μεθοδολογία έρευνας

Για τη συλλογή των δεδομένων μας ακολουθήσαμε τη μέθοδο του ερωτηματολογίου. Μέσω ενός ερωτηματολογίου δίνεται η δυνατότητα μέτρησης συμπεριφορών που δεν είναι εύκολο να παρατηρηθούν. Πριν παρουσιάσουμε τα αποτελέσματα που προέκυψαν από τις απαντήσεις του ερωτηματολογίου, θα κάνουμε μια μικρή αναφορά στη μεθοδολογία που ακολουθήσαμε.

Το ερωτηματολόγιο περιλαμβάνει δομημένες ερωτήσεις στις οποίες οι συμμετέχοντες στην έρευνα, καλούνται να απαντήσουν στο ίδιο σύνολο ερωτήσεων γραπτώς και με συγκεκριμένη σειρά. Στα πλεονεκτήματα της μεθόδου αυτής είναι το χαμηλό κόστος, η δυνατότητα συμπλήρωσης από μεγάλο αριθμό ατόμων, η μη

δυνατή επιρροή του ερευνητή στις απαντήσεις και το μικρό χρονικό διάστημα που απαιτείται για τη συλλογή των απαντήσεων. Επίσης η μέθοδος αυτή κάνει τη διαδικασία των απαντήσεων πιο εύκολη και μειώνει την πιθανότητα να μην απαντηθούν κάποιες εκ των ερωτήσεων. Ακόμη με τη μέθοδο αυτή γίνεται ευκολότερη η ανάλυση και η σύγκριση των απαντήσεων. Από την άλλη πλευρά, κάποιοι εκ των συμμετεχόντων μπορεί να διαφωνούν με τις απαντήσεις και ίσως έτσι περιορίζεται η αυτοέκφραση (Παρασκευόπουλος, 1999).

Κατά τη δημιουργία ενός ερωτηματολογίου πρέπει να ληφθούν υπόψη χαρακτηριστικά όπως η πληρότητα, σαφήνεια και συνοχή των ερωτήσεων. Επίσης στο ερωτηματολόγιο θα πρέπει να συμπεριλάβουμε ερωτήματα ελέγχου ώστε να ελέγξουμε την ορθότητα τον απαντήσεων σε κρίσιμες για την έρευνα μας ερωτήσεις. Αξίζει να σημειωθεί ότι ένα σύντομο ερωτηματολόγιο είναι πιο αποτελεσματικό, καθώς δεν είναι κουραστικό και αυξάνονται οι πιθανότητες να απαντηθεί αν ο ερωτώμενος ξέρει ότι δε θα διαθέσει πολύ χρόνο για να το συμπληρώσει. Τέλος οι ερωτήσεις που τίθενται μπορεί να είναι α) διχοτομικές, δηλαδή θα πρέπει να επιλεγεί μία εκ των δύο απαντήσεων που δίνονται, β) βαθμονόμησης, δηλαδή να επιλεγεί μία από τις υπάρχουσες απαντήσεις, γ) κατάταξης, δηλαδή να επιλεχθούν οι απαντήσεις με σειρά προτεραιότητας, δ) διαβαθμισμένης κλίμακας, δηλαδή να γίνει βαθμολόγηση με συγκεκριμένη κλίμακα στο περιεχόμενο της ερώτησης και ε) ερωτήσεις πολλαπλών επιλογών, δηλαδή να υπάρχει η δυνατότητα επιλογής περισσότερων από μία απαντήσεων (Παρασκευόπουλος, 1999).

Οι μέθοδοι συλλογής του δείγματος χωρίζονται σε δύο κατηγορίες, την αντιπροσωπευτική δειγματοληψία και τη δειγματοληψία κρίσης. Η αντιπροσωπευτική δειγματοληψία είναι ελεγχόμενη ως τις παραμέτρους της και επιτρέπει τη γενίκευση των συμπερασμάτων που προκύπτουν, υπολογίζοντας ένα σφάλμα εκτίμησης. Η συλλογή του δείγματος μπορεί να γίνει με τυχαία δειγματοληψία, με συστηματική δειγματοληψία, με στρωματοποιημένη δειγματοληψία κ.α. Η δειγματοληψία κρίσης χρησιμοποιείται σε περιπτώσεις όπου χρειάζεται να ολοκληρωθεί η έρευνα σε σύντομο χρονικό διάστημα. Τα αποτελέσματα που προκύπτουν δε μπορούν να γενικευθούν και δεν υπάρχει η δυνατότητα να υπολογισθεί το σφάλμα εκτίμησης, οπότε η έρευνα θα πρέπει να είναι περιορισμένης χρήσης και τα αποτελέσματα να χρησιμοποιηθούν προσεκτικά. Η συλλογή του δείγματος μπορεί να γίνει με τεχνικές όπως η δειγματοληψία χιονόμπαλας, ποσοστών κ.α. (Φίλιας, 2001).

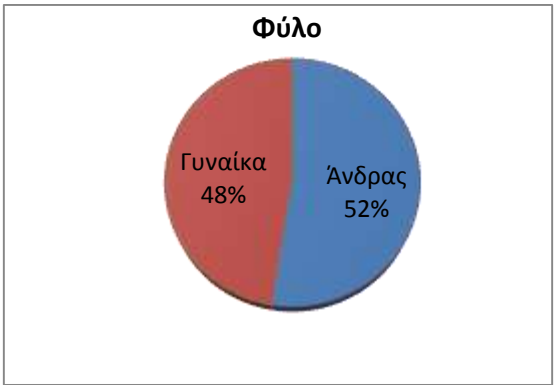
Για τη δική μας έρευνα, αρχικά συντάξαμε το ερωτηματολόγιο στο οποίο επιλέξαμε να υποβάλλουμε ερωτήσεις με κλειστές κωδικοποιημένες απαντήσεις, ορισμένες από τις οποίες είναι μίας επιλογής βαθμονόμησης και ορισμένες διαβαθμισμένης κλίμακας. Οι ερωτήσεις που θέσαμε χωρίζονται σε τρία θεματικά πεδία. Στο πρώτο, περιέχονται ερωτήσεις αναφορικά με τα δημογραφικά στοιχεία των συμμετεχόντων, όπως το φύλλο, η ηλικία, το επάγγελμα, το εισόδημα, το μορφωτικό επίπεδο και η οικογενειακή τους κατάσταση. Στο δεύτερο θέσαμε ερωτήσεις σχετικά με την επενδυτική τους συμπεριφορά. Ο σκοπός των ερωτήσεων αυτών είναι να καθορίσουμε το επενδυτικό προφίλ των ερωτώμενων ώστε να συμπεράνουμε αν είναι ορθολογικοί επενδυτές ή noise traders. Στο τρίτο πεδίο οι ερωτήσεις αφορούν τις αντιδράσεις των επενδυτών στα σχόλια που διαβάζουν στο διαδίκτυο και ειδικότερα στα μέσα κοινωνικής δικτύωσης. Σκοπός μας είναι να εξάγουμε συμπεράσματα για το αν και κατά πόσο οι ερωτώμενοι επηρεάζονται από τα σχόλια τρίτων ως προς τις επενδυτικές τους αποφάσεις. Στο δεύτερο και τρίτο πεδίο ορισμένα από τα ερωτήματα είναι ελέγχου, από το δειγματοληπτικό έλεγχο των οποίων δεν προκύπτουν μη ορθές ή τυχαίες απαντήσεις.

Η συλλογή του δείγματος έγινε με την τεχνική της τυχαίας δειγματοληψίας. Αν και με τη χρήση της τεχνικής αυτής δεν οδηγούμαστε στη δημιουργία αντιπροσωπευτικού δείγματος, καθώς μπορεί να αφήσουμε ακάλυπτες πληθυσμιακές περιοχές και άρα να μην έχουμε πλήρη αντιπροσώπευση των ζητούμενων χαρακτηριστικών, είναι πολύ εύκολη στη χρήση της (Φίλιας, 2001). Απευθύναμε το ερωτηματολόγιο μας σε πολίτες κάθε ηλικίας, επαγγέλματος και μορφωτικού επιπέδου και ο τρόπος με τον οποίο τέθηκε ήταν μέσω email, Viber και social media. Θεωρήσαμε ότι θα λάβουμε ειλικρινέστερες απαντήσεις σε ένα ανώνυμο ερωτηματολόγιο, χωρίς την παρουσία μας. Λάβαμε 177 απαντήσεις, τις οποίες και θα αναπτύξουμε στην επόμενη ενότητα.

2.5.2. Επεξεργασία δεδομένων και ανάλυση αποτελεσμάτων

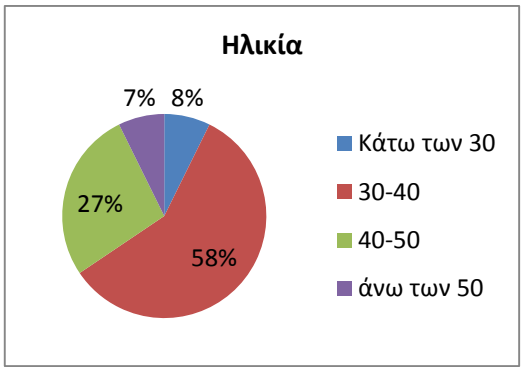
Στο πρώτο πεδίο ερωτήσεων με την ονομασία «Δημογραφικά Στοιχεία», θέσαμε ανάλογα ερωτήματα. Η έρευνα απευθύνθηκε σε ανθρώπους όλων των ηλικιών, επαγγελμάτων, οικογενειακής κατάστασης και μορφωτικού επιπέδου, προσπαθώντας να καλύψει το δείγμα και με εισοδηματικές κατηγορίες.

Όσον αφορά στις ερωτήσεις που συνθέτουν τα δημογραφικά στοιχεία του δείγματος, προκύπτει από τα δοθέντα διαγράμματα ότι απαντήθηκαν όλες οι διαθέσιμες επιλογές, με σχεδόν ισομερή κατανομή στα δύο φύλα.



Διάγραμμα 1: Κατανομή δείγματος ανά φύλο

Στην κλίμακα των ηλικιών που δόθηκε, απαντώνται πολίτες σε όλες τις ηλικίες με το μεγαλύτερο ποσοστό να ανήκει στην παραγωγική δεκαετία των 30 ετών σε ποσοστό 58%, με την αμέσως επόμενη κατηγορία της δεκαετίας των 40 ετών να ακολουθεί με 27%. Το υπόλοιπο ποσοστό μοιράζεται σε νέους κάτω των 30 και ανθρώπους άνω των 50 ετών.

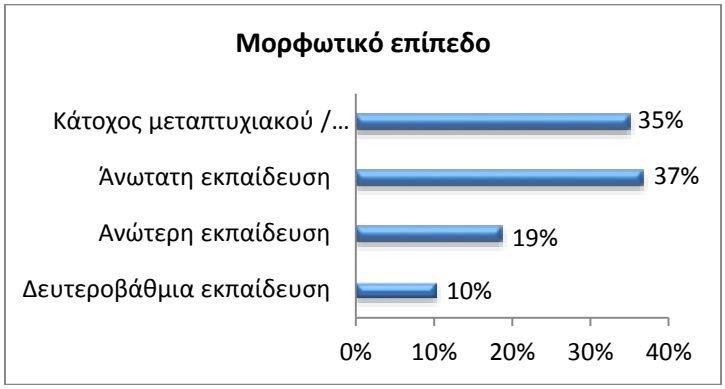


Διάγραμμα 2: Κατανομή δείγματος ανά ηλικία

Όσον αφορά την οικογενειακή κατάσταση του δείγματος, περίπου οι μισοί πολίτες που απάντησαν είναι έγγαμοι με παιδιά (52,50%) με τους άγαμους να επικρατούν έναντι των άλλων επιλογών με ποσοστό 34,50%. Περισσότεροι δηλαδή από τους μισούς, οι οποίοι διαμορφώνουν τα αποτελέσματα της έρευνας, είναι οικογενειάρχες με παιδιά.

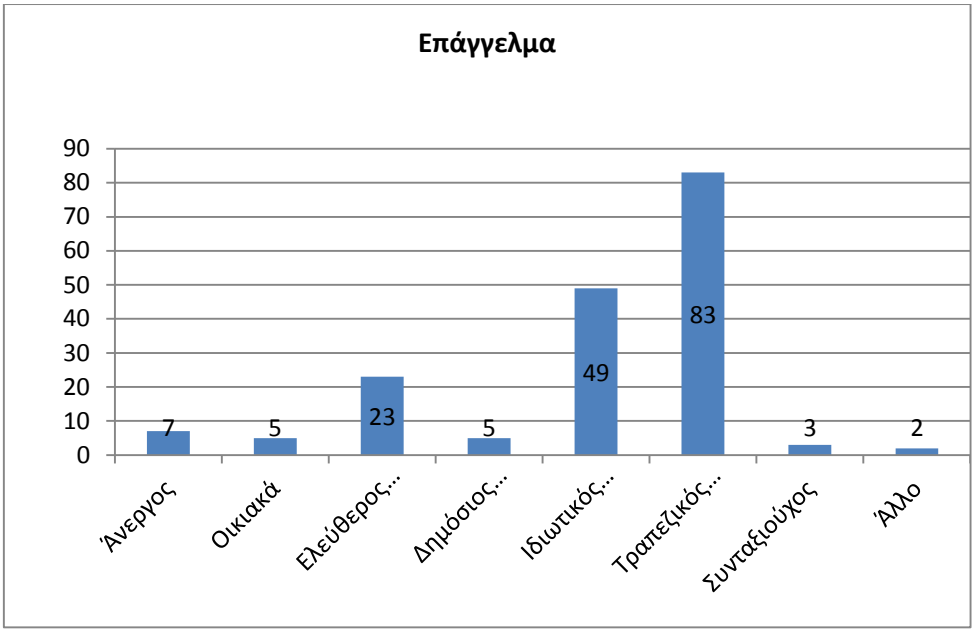
Διάγραμμα 3: Κατανομή δείγματος ανά οικογενειακή κατάσταση

Η μερίδα του λέοντος στο εκπαιδευτικό υπόβαθρο των ερωτηθέντων, ανήκει σε πολίτες με ανωτάτη εκπαίδευση (37%) και κατόχων μεταπτυχιακού / διδακτορικού τίτλου (35%), ποσοστό αθροιστικά που κρίνεται πολύ καλό, καθώς διαμορφώνει σε μεγάλο ποσοστό τα αποτελέσματα της έρευνας. Ακολουθούν ποσοστιαία, η ανώτερη και η δευτεροβάθμια εκπαίδευση.



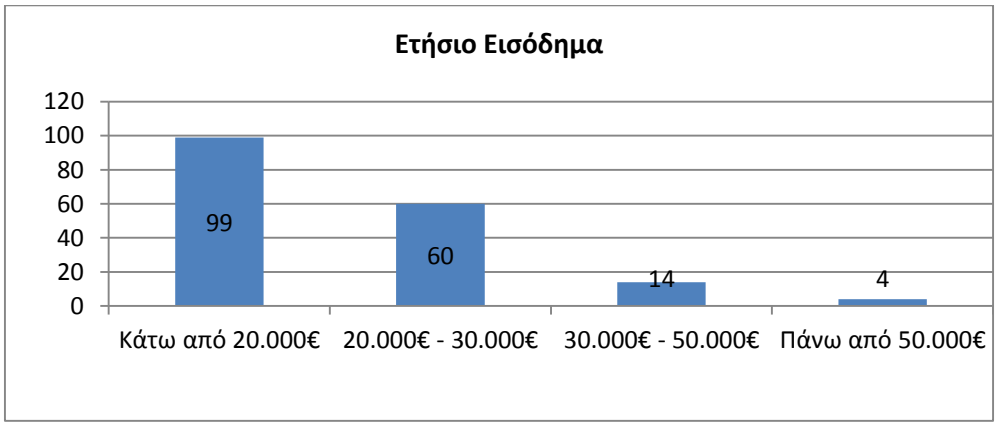
Διάγραμμα 4: Κατανομή δείγματος ανά μορφωτικό επίπεδο

Η επαγγελματική κατάσταση των συμμετεχόντων στην έρευνα, εστιάζεται κυρίως σε μισθωτούς (137 άτομα με ποσοστό 77%) και ελεύθερους επαγγελματίες (ποσοστό 13%). Το σύνολο των μισθωτών, διακρίνεται σε υπαλλήλους στον τραπεζικό χώρο, στον ιδιωτικό και δημόσιο τομέα.



Διάγραμμα 5: Κατανομή δείγματος ανά επάγγελμα

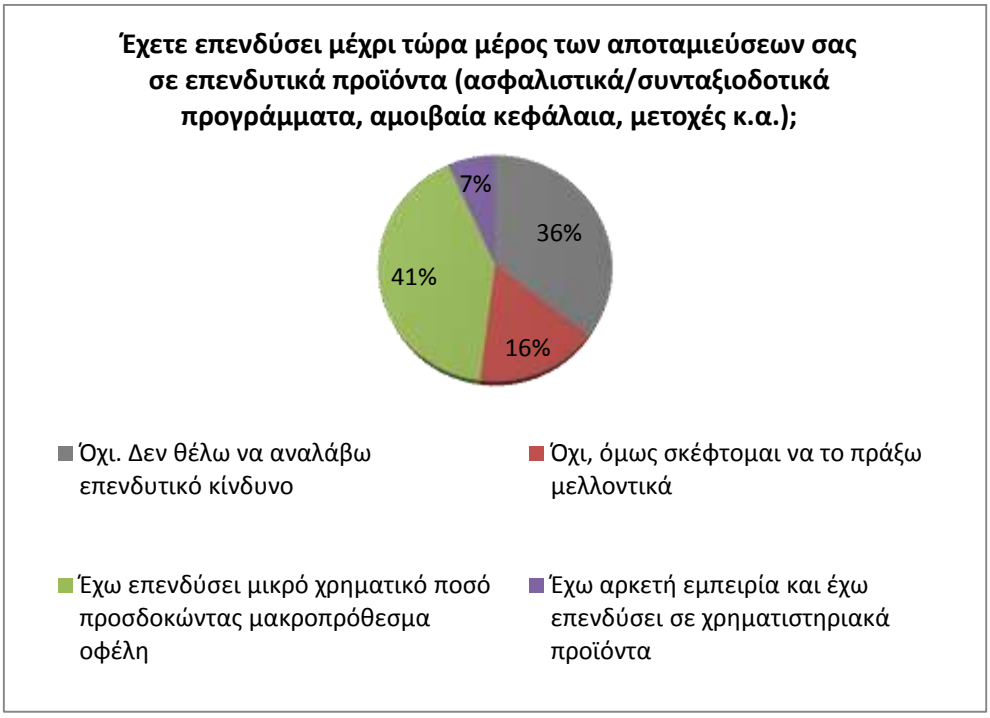
Τέλος, περιλάβαμε και το ετήσιο εισόδημα, ως δημογραφικό στοιχείο, σαν ένα επιπλέον κριτήριο διαστρωμάτωσης και κατηγοριοποίησης του δείγματος, ώστε να εξαχθούν μέσω των απαντήσεων στις ερωτήσεις της έρευνας, η όσο το δυνατόν καλύτερη ερμηνεία της συμπεριφοράς των συμμετεχόντων, με το όριο των 20.000 ευρώ ως ετήσιο εισόδημα, να συναντάται στους περισσότερους από τους μισούς.



Διάγραμμα 6: Κατανομή δείγματος ανά ετήσιο εισόδημα

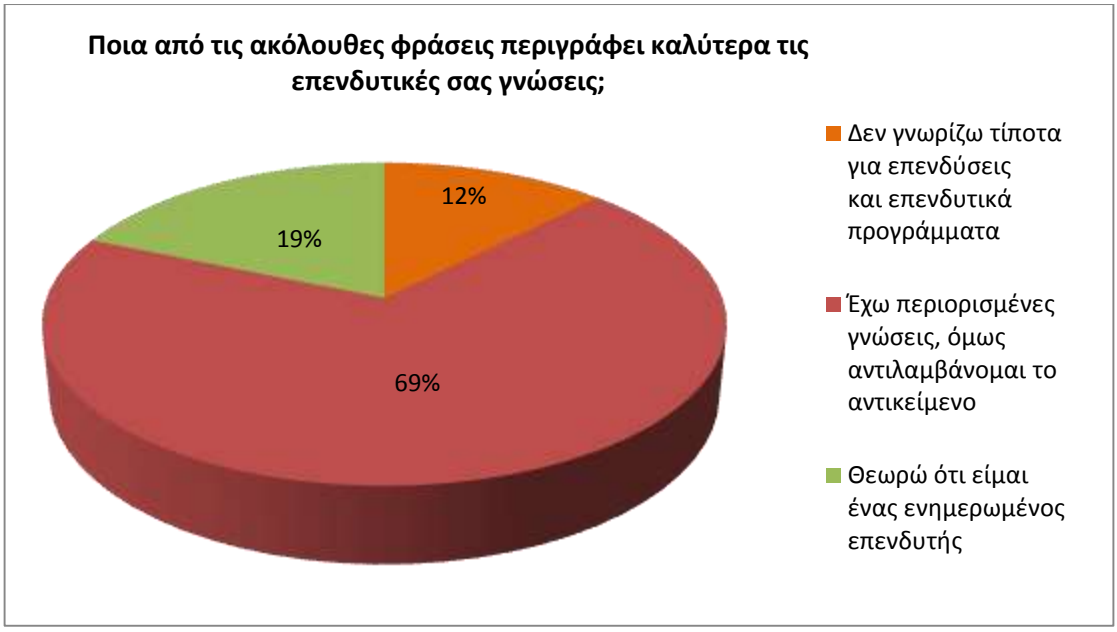
Αφού καταγράψαμε τα δημογραφικά στοιχεία, συνεχίσαμε με το δεύτερο πεδίο ερωτήσεων με την ονομασία «Επενδυτικό Προφίλ». Στο πεδίο αυτό, όπως αναφέραμε και παραπάνω θέσαμε ερωτήματα, σχετικά με την επενδυτική συμπεριφορά των ατόμων, ώστε να εξετάσουμε το επενδυτικό τους προφίλ.

Στην πρώτη ερώτηση, η οποία αφορά ύπαρξη ή μη χαρτοφυλακίου, αξιοσημείωτο είναι το γεγονός ότι ένα αρκετά μεγάλο ποσοστό της τάξης των 35,6% απάντησε ότι δεν έχει επενδύσει ποτέ διότι δε θέλει να αναλάβει τον επενδυτικό κίνδυνο. Ένα 16,4% των ατόμων δηλώνει ότι ναι μεν δεν έχει επενδύσει έως τώρα, όμως προτίθεται να το κάνει στο μέλλον και μόλις το 6,8% των ερωτηθέντων απάντησε ότι έχει αρκετή εμπειρία και διαθέτει χαρτοφυλάκιο με χρηματιστηριακά προϊόντα. Το μεγαλύτερο διαμορφωμένο ποσοστό του 41,2%, αφορά απάντηση «έχω επενδύσει μικρό χρηματικό ποσό προσδοκώντας μακροπρόθεσμα οφέλη».



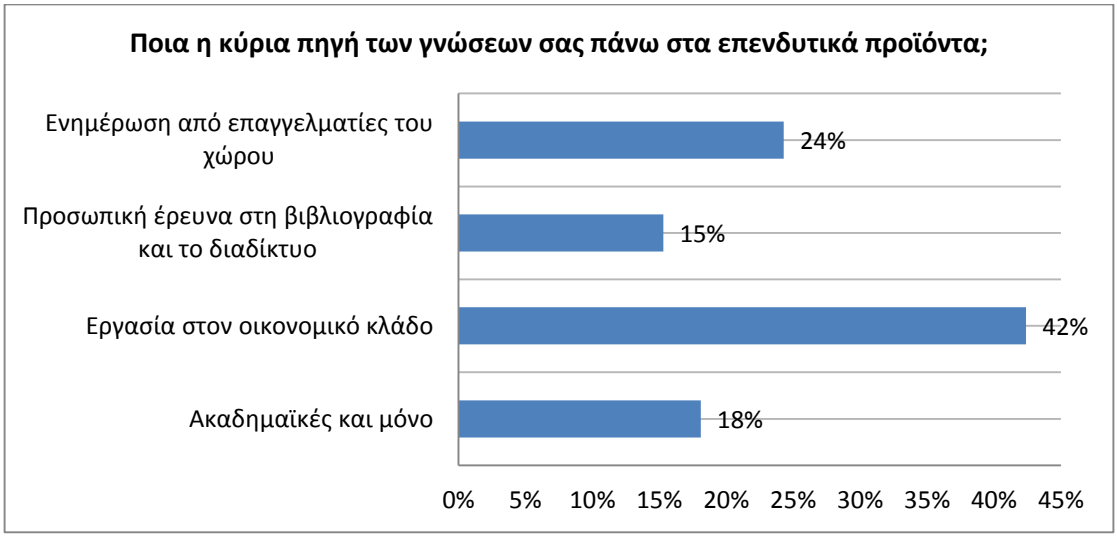
Διάγραμμα 7: Πρόθεση ανάληψης κινδύνου

Στο επόμενο ερώτημα, στο πως κρίνουν οι ερωτηθέντες τις γνώσεις τους πάνω στις επενδύσεις, οι περισσότεροι (68,9%) απάντησαν ότι έχουν περιορισμένες γνώσεις, όμως αντιλαμβάνονται το αντικείμενο. Το 18,6% θεωρεί τον εαυτό του ως ενημερωμένο επενδυτή, ενώ το 12,4% απάντησε ότι δε γνωρίζει τίποτα για επενδύσεις και επενδυτικά προγράμματα.



Διάγραμμα 8: Αυτόαξιολόγηση επενδυτικών γνώσεων

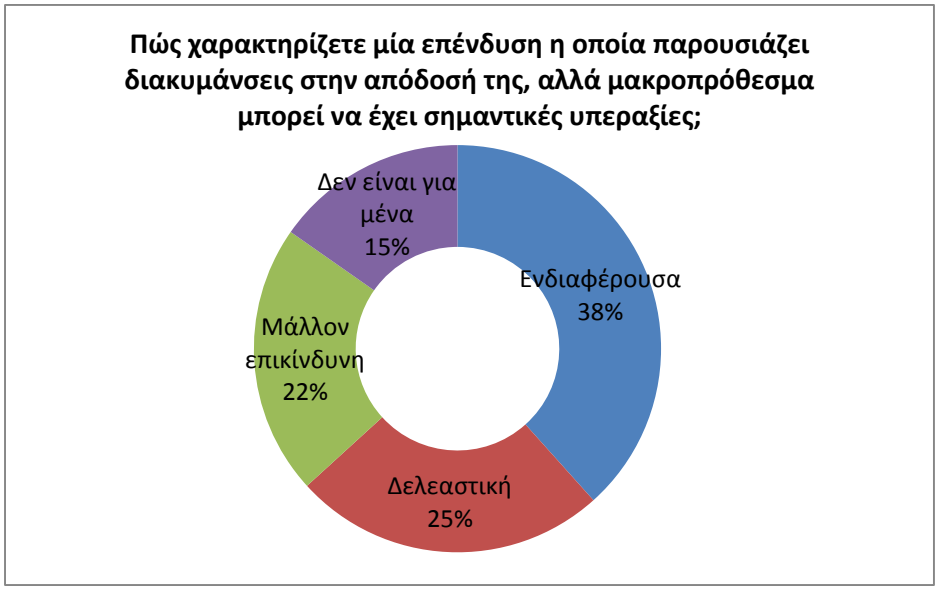
Στην ερώτηση που τέθηκε αναφορικά με την κύρια πηγή γνώσεων των συμμετεχόντων πάνω στα επενδυτικά προϊόντα, η εργασία στον οικονομικό κλάδο κατέχει τα πρωτεία με 42,4% και ακολουθεί η προσωπική έρευνα στη βιβλιογραφία και το διαδίκτυο με 15,3% και οι γνώσεις μόνο λόγω ακαδημαϊκών σπουδών με 18,1%. Εντύπωση δημιουργεί το γεγονός ότι ένα αξιοσημείωτο ποσοστό της τάξης του 24,3% απάντησε ότι ενημερώνεται από επαγγελματίες του χώρου.



Διάγραμμα 9: Κύρια πηγή επενδυτικών γνώσεων

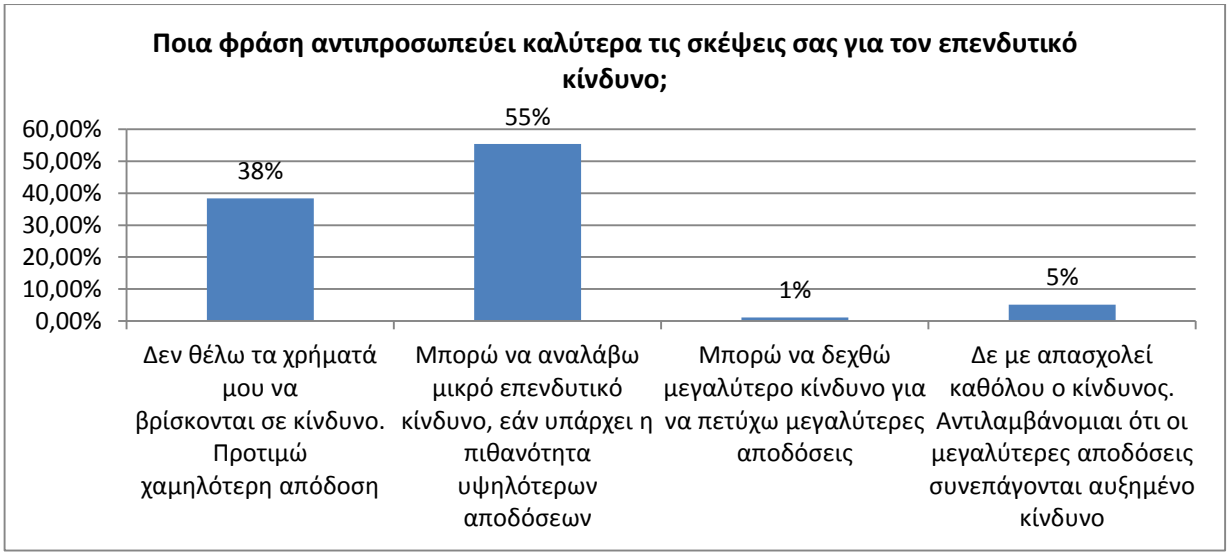
Κατόπιν των προηγούμενων απαντήσεων, ζητήθηκε από τους συμμετέχοντες να χαρακτηρίσουν μια επένδυση με διακυμάνσεις, η οποία όμως δύναται

μακροπρόθεσμα να έχει σημαντικές υπεραξίες. Οι περισσότεροι τη χαρακτήρισαν ως ενδιαφέρουσα (38,4%) ή δελεαστική (24,9%), ενώ 21,5% τη χαρακτήρισε ως επικίνδυνη και 15,3% απάντησε ότι «δεν είναι για μένα».



Διάγραμμα 10: Αξιολόγηση επένδυσης με διακυμάνσεις και πιθανόν σημαντικές μελλοντικές υπεραξίες

Στην τελευταία ερώτηση αυτού του πεδίου, δόθηκαν τέσσερις φράσεις αναφορικά με τον επενδυτικό κίνδυνο και ζητήθηκε ως απάντηση αυτή που τους αντιπροσωπεύει καλύτερα. Το μεγαλύτερο ποσοστό απαντήσεων (55,4%) ήταν «μπορώ να αναλάβω μικρό επενδυτικό κίνδυνο, εάν υπάρχει η πιθανότητα υψηλότερων αποδόσεων». Δύο μόλις από τους συμμετέχοντες απάντησαν ότι μπορούν να δεχθούν μεγαλύτερο κίνδυνο για να πετύχουν μεγαλύτερες αποδόσεις. Εννιά απάντησαν ότι δεν τους απασχολεί καθόλου ο κίνδυνος, καθώς αντιλαμβάνονται ότι οι μεγαλύτερες αποδόσεις συνεπάγονται αυξημένο κίνδυνο. Αξιοσημείωτο είναι ότι ένα πολύ μεγάλο ποσοστό της τάξης του 38,4%, απάντησε ότι δε θέλει τα χρήματα του να βρίσκονται σε κίνδυνο και προτιμά χαμηλότερες αποδόσεις.



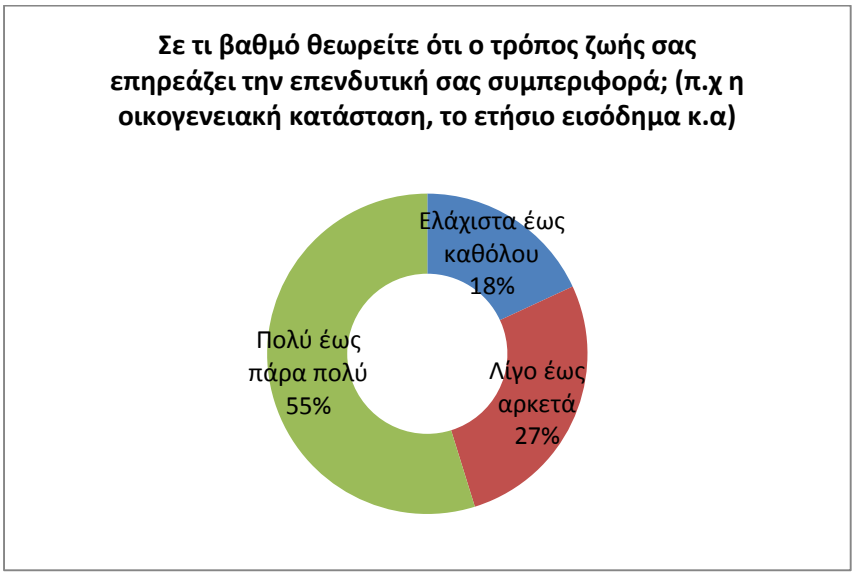
Διάγραμμα 11: Πιθανότητα ανάληψης κινδύνου

Από το πεδίο «Επενδυτικό Προφίλ» προκύπτει ότι οι συμμετέχοντες που απαρτίζουν το δείγμα των 117 ανθρώπων, στην πλειοψηφία τους είναι άπειροι ως επενδυτές και με περιορισμένες γνώσεις πάνω στο αντικείμενο, αν και οι περισσότεροι εργάζονται στον οικονομικό κλάδο. Οι δύο τελευταίες ερωτήσεις ήταν ερωτήσεις ελέγχου. Το γεγονός ότι οι περισσότεροι εκ των ερωτηθέντων, από τη μία χαρακτηρίζουν μία επένδυση με διακυμάνσεις αλλά μελλοντικά σημαντικές υπεραξίες ως ενδιαφέρουσα και δελεαστική, ενώ από την άλλη προτιμούν επενδύσεις με μικρό ή καθόλου κίνδυνο, μας οδηγεί στο συμπέρασμα ότι χαρακτήρισαν την επένδυση χωρίς να λάβουν υπόψη τους το τι θα έκαναν οι ίδιοι στην πραγματικότητα. Με βάση τις απαντήσεις στην τελευταία ερώτηση, όπου το συντριπτικό 93,8% απάντησε ότι δε διατίθεται να αναλάβει καθόλου κίνδυνο ή ότι έστω μπορεί να δεχθεί μικρό επενδυτικό κίνδυνο, η λογική αντίδραση θα ήταν να χαρακτηρίσουν την επένδυση της ερώτησης 4 ως επικίνδυνη ή ότι δεν είναι γι' αυτούς.

Το τρίτο και τελευταίο πεδίο ερωτήσεων, με την ονομασία «Λήψη Αποφάσεων», περιλαμβάνει ερωτήματα σχετικά με τις αντιδράσεις των ατόμων σε σχόλια τρίτων που έχουν αναρτηθεί στα μέσα κοινωνικής δικτύωσης και γενικά στο διαδίκτυο και στο κατά πόσο αυτά μπορεί να επηρεάσουν τις επενδυτικές τους αποφάσεις.

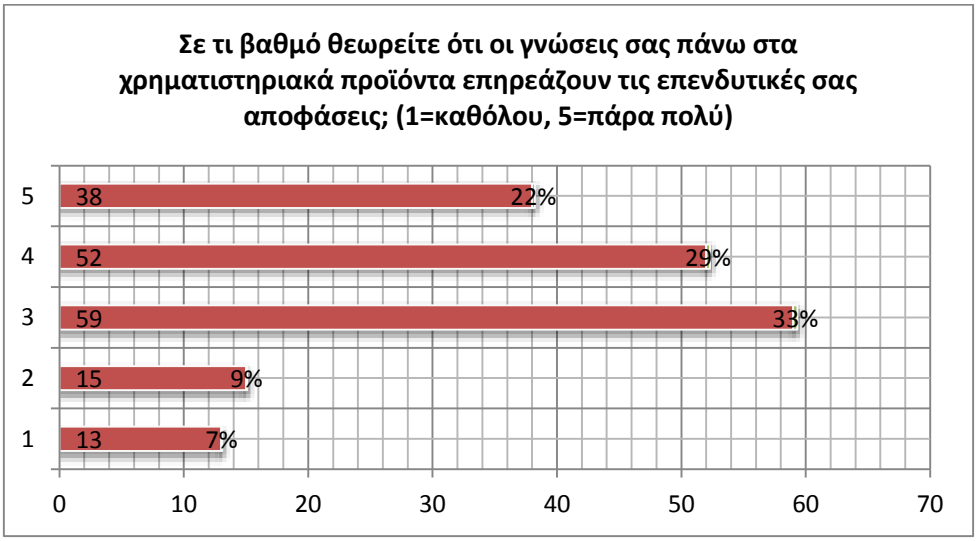
Η πρώτη ερώτηση αυτού του πεδίου, αφορούσε το βαθμό που επηρεάζονται οι επενδυτικές αποφάσεις των ερωτηθέντων από τον τρόπο ζωής τους, όπως π.χ. την οικογενειακή τους κατάσταση, το ετήσιο εισόδημα τους κ.λ.π. Το μεγαλύτερο

(54,8%) απάντησε πως επηρεάζεται πολύ έως πάρα πολύ. Ακολουθεί ποσοστό 18,1% που απάντησε λίγο έως αρκετά και 18,1% που απάντησε ελάχιστα έως καθόλου.



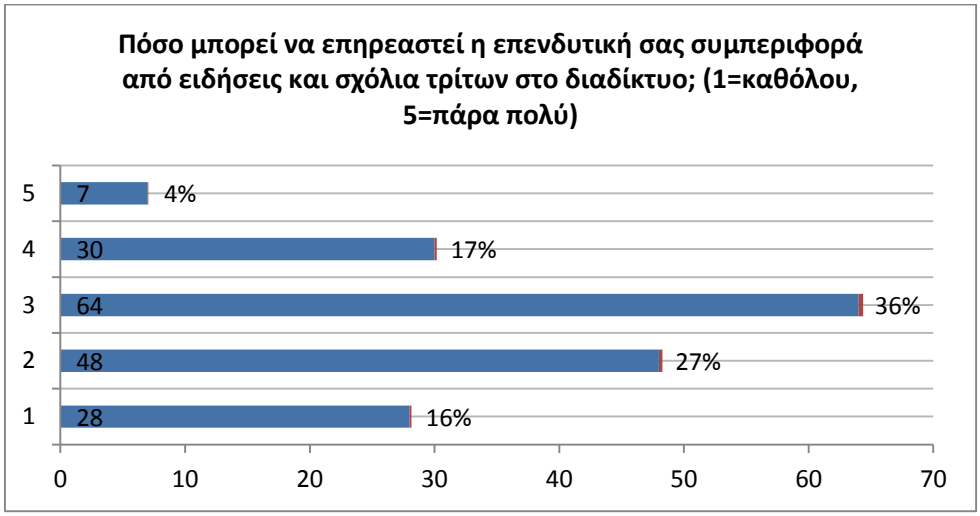
Διάγραμμα 12: Τρόπος ζωής και επενδυτική συμπεριφορά

Αμέσως μετά ζητήθηκε από τους συμμετέχοντες να βαθμολογήσουν με 1 έως 5, κατά πόσο θεωρούν ότι οι επενδυτικές τους αποφάσεις επηρεάζονται από τις γνώσεις τους πάνω στα χρηματιστηριακά προϊόντα. Οι απαντήσεις των περισσοτέρων κινήθηκαν μεταξύ 3 και 4 (συνολικά 62,7%), ενώ δεν ήταν λίγοι και αυτοί που έδωσαν το 5, δηλαδή πάρα πολύ, ως απάντηση (21,5%). Μεταξύ του 1, δηλαδή καθόλου και του 2 κινήθηκε το 15,8% των απαντήσεων.



Διάγραμμα 13: Οικονομικές γνώσεις και επενδυτικές αποφάσεις

Στη συνέχεια ρωτήθηκαν κατά πόσο μπορεί να επηρεαστεί η επενδυτική τους συμπεριφορά από σχόλια τρίτων στο διαδίκτυο (1=καθόλου, 5=πάρα πολύ). Καθόλου απάντησε το 15,8% ενώ πάρα πολύ μόλις το 4%. Με 2 βαθμολόγησε το 27,1% και με 4 το 16,9%. Οι περισσότεροι (36,2%) κινήθηκαν ακριβώς στη μέση και βαθμολόγησαν την ερώτηση με 3.

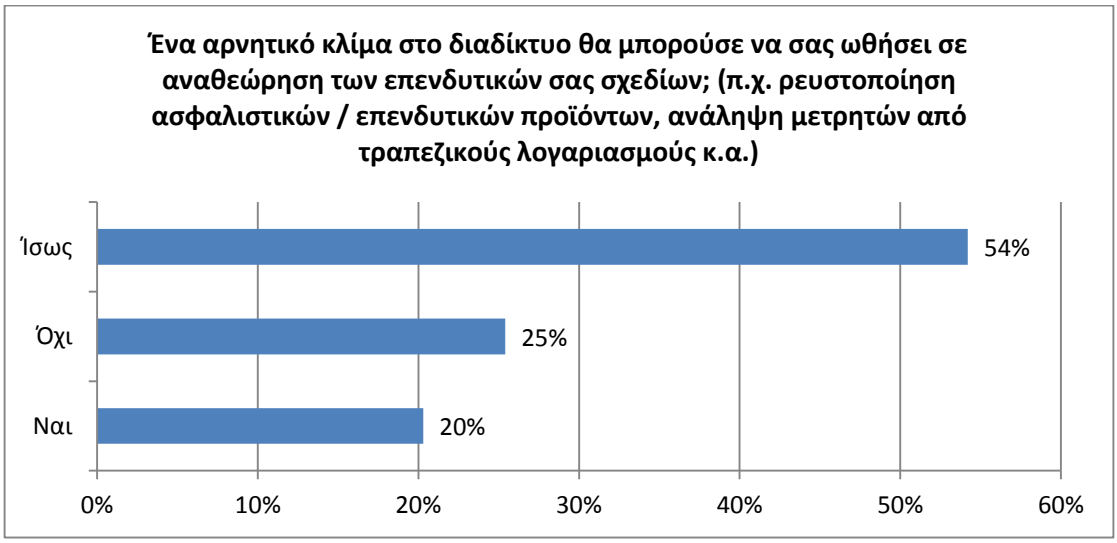


Διάγραμμα 14: Σχόλια στο διαδίκτυο και επενδυτική συμπεριφορά

Στην επόμενη ερώτηση κλήθηκαν να απαντήσουν κατά πόσο πολλά θετικά αρνητικά σχόλια χρηστών του διαδικτύου, μπορεί να επηρεάσουν τις επενδυτικές τους αποφάσεις. Είναι αξιοσημείωτο ότι οι μισοί απάντησαν ότι δε λαμβάνουν υπόψη τους τα σχόλια τρίτων στο διαδίκτυο. Ωστόσο ένα επίσης μεγάλο ποσοστό της τάξης του 29,4% απάντησε ότι λαμβάνει υπόψη του όλα τα σχόλια. Τέλος, το γεγονός ότι το 16,4% απάντησε ότι πιθανόν να λάβει υπόψη του τα σχόλια αν υπάρχουν πολλά αρνητικά, έναντι του 5,1% που απάντησε ότι ίσως τα λάβει υπόψη του αν είναι πολλά τα θετικά, έρχεται σε αντίθεση με την υπεραισιοδοξία που συνήθως χαρακτηρίζει τους επενδυτές.

Διάγραμμα 15: Σχόλια στο διαδίκτυο και λήψη επενδυτικών αποφάσεων

Στη συνέχεια ρωτήθηκαν αν ένα αρνητικό κλίμα στο διαδίκτυο θα μπορούσε να τους ωθήσει στην αναθεώρηση των επενδυτικών τους σχεδίων, όπως π.χ. στη ρευστοποίηση ασφαλιστικών / επενδυτικών προϊόντων ή στην ανάληψη μετρητών από τους τραπεζικούς τους λογαριασμούς. Λίγο περισσότεροι από τους μισούς δήλωσαν ότι ίσως θα επηρεαζόταν από ένα αρνητικό κλίμα στο διαδίκτυο, ενώ ότι σίγουρα θα επηρεαζόταν δήλωσε το 20,3%. Το 25,4% ωστόσο δήλωσε ότι δε θα επηρεαζόταν.



Διάγραμμα 16: Αρνητικό κλίμα στο διαδίκτυο και επενδυτικά σχέδια

Το ίδιο ερώτημα τέθηκε και αντιστρόφως. Ρωτήσαμε δηλαδή αν ένα θετικό κλίμα στο διαδίκτυο θα μπορούσε να τους ωθήσει να προβούν σε μία επένδυση, όπως π.χ. να αγοράσουν μετοχές ή απλά να καταθέσουν σε τραπεζικό λογαριασμό χρήματα που έχουν εκτός τράπεζας. Παρατηρούμε ότι ακριβώς το ίδιο ποσοστό με την προηγούμενη ερώτηση απάντησε ότι ίσως θα επηρεαζόταν. Ωστόσο σε αυτή την ερώτηση, ήταν μεγαλύτερο το ποσοστό που απάντησε ότι δε θα επηρεαζόταν και μικρότερο αυτό απάντησε ότι σίγουρα θα επηρεαζόταν.

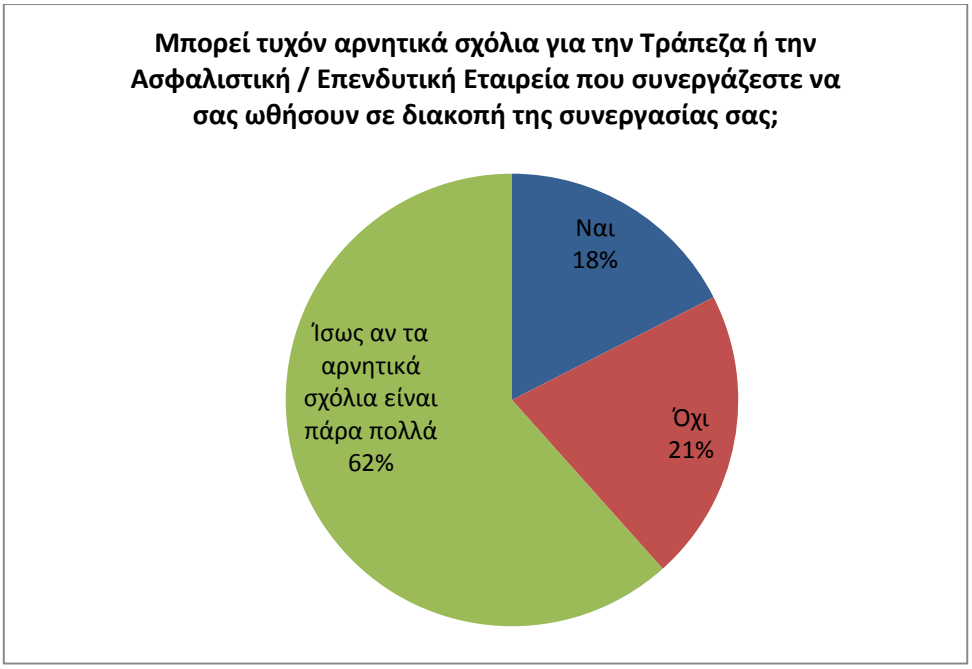
Διάγραμμα 17: Θετικό κλίμα στο διαδίκτυο και επενδυτικά σχέδια

Οι δύο επόμενες ερωτήσεις αφορούσαν τις τυχόν μεταβολές στη σχέση των ερωτηθέντων με την τράπεζα ή την ασφαλιστική / επενδυτική εταιρεία που συνεργάζονται.

Αρχικά ρωτήσαμε κατά πόσο τα σχόλια τρίτων στο διαδίκτυο αναφορικά με μία τράπεζα ή ασφαλιστική / επενδυτική εταιρεία μπορεί να επηρεάσουν την πιθανότητα να συνεργαστούν μαζί της (1=καθόλου, 5=πάρα πολύ). Οι πλειοψηφία των συμμετεχόντων κινήθηκε κάπου στη μέση δίνοντας 37,9% στο 3. Το 21,5% βαθμολόγησε με 4 και ένα 5,1% απάντησε 5, δηλαδή πάρα πολύ. Στον αντίποδα, το 25,4% βαθμολόγησε με 2 και το 10,2% απάντησε 1 δηλαδή καθόλου.

Διάγραμμα 18: Σχόλια στο διαδίκτυο και πιθανότητα συνεργασίας με συγκεκριμένη Τράπεζα ή Ασφαλιστική / Επενδυτική Εταιρεία

Έπειτα θέσαμε την ερώτηση από άλλη οπτική γωνία και τους καλέσαμε να απαντήσουν αν τυχόν αρνητικά σχόλια για την Τράπεζα ή την Ασφαλιστική / Επενδυτική Εταιρεία που συνεργάζονται θα μπορούσαν να τους ωθήσουν σε διακοπή της συνεργασίας τους μαζί της. Το 20,9% απάντησε ότι δε θα τη διέκοπτε και το 17,5% απάντησε ότι ναι θα τη διέκοπτε. Εντύπωση δημιουργεί το γεγονός ότι η συντριπτική πλειοψηφία (61,6%) απάντησε ότι ίσως θα διέκοπτε τη συνεργασία της αν τα αρνητικά σχόλια ήταν πάρα πολλά.



Διάγραμμα 19: Σχόλια στο διαδίκτυο και πιθανότητα διακοπής συνεργασίας με συγκεκριμένη Τράπεζα ή Ασφαλιστική / Επενδυτική Εταιρεία

Το τρίτο πεδίο της «Λήψης Αποφάσεων» ήταν και το τελευταίο του ερωτηματολογίου μας. Το πεδίο αυτό στοχεύει να αποκρυπτογραφήσει μέσω στοχευμένων ερωταπαντήσεων, αν και κατά πόσο τα σχόλια και οι φήμες στο διαδίκτυο μπορούν να επηρεάσουν την επενδυτική συμπεριφορά των ατόμων και κατ' επέκταση τις τιμές των αξιογράφων. Αν δηλαδή μπορούν να επηρεάσουν τις κινήσεις της αγοράς.

Συνοπτικά μπορούμε να πούμε πως από τις απαντήσεις που λάβαμε στο ερωτηματολόγιο μας, συμπεράναμε πως αν και οι περισσότεροι εκ των συμμετεχόντων εργάζονται στον οικονομικό κλάδο και έχουν επενδύσει έστω και μικρά ποσά, έχουν περιορισμένες γνώσεις πάνω στα επενδυτικά προϊόντα. Παρόλα αυτά μόνο το 24,3% των συμμετεχόντων απάντησε ότι ενημερώνεται από επαγγελματίες του χώρου. Διαπιστώσαμε επίσης ότι η πλειοψηφία των ατόμων μπορεί να επηρεαστεί από λίγο έως πάρα πολύ από τα σχόλια που διαβάζει στα μέσα κοινωνικής δικτύωσης, ιδίως τα αρνητικά από αυτά. Το γεγονός αυτό μπορεί να έχει αντίκτυπο στη συνεργασία τους με τράπεζες και ασφαλιστικές / επενδυτικές εταιρείες. Καταλήγουμε λοιπόν στο συμπέρασμα ότι τα άτομα τείνουν περισσότερο να κατηγοριοποιηθούν στην κατηγορία των noise traders, παρά στην κατηγορία των ορθολογικών επενδυτών.

Με βάση τα συμπεράσματα που εξάγαμε από το ερωτηματολόγιο, γίνεται σαφές ότι η ποιότητα της πληροφορίας που λαμβάνουν τα άτομα είναι ιδιαίτερα κρίσιμη καθώς επηρεάζει τις επενδυτικές τους αποφάσεις. Κατανοούμε λοιπόν για ποιο λόγο είναι σημαντικό για τις τράπεζες και τις ασφαλιστικές / επενδυτικές εταιρείες να αποκωδικοποιήσουν τα σχόλια που αναρτώνται στο διαδίκτυο και να τα κατηγοριοποιήσουν ως προς το συναίσθημα που εκφράζουν, προκειμένου να προβλέψουν τις αντιδράσεις των πελατών τους σε αυτά. Έτσι, αντιλαμβανόμαστε ότι το πεδίο έρευνας που αφορά στην ανάλυση συναισθήματος, μελλοντικά μπορεί να αποτελέσει σημαντικό εργαλείο για τη χάραξη στρατηγικής από τις εταιρείες που δραστηριοποιούνται στον οικονομικό κλάδο. Συνεπώς, η περεταίρω βελτίωση των μεθόδων ανάλυσης κειμένου που αναπτύξαμε στο πρώτο κεφάλαιο κρίνεται απαραίτητη.

Συμπεράσματα

Τα μέσα κοινωνικής δικτύωσης αποτελούν ίσως το πιο διαδεδομένο μέσο επικοινωνίας και ανταλλαγής απόψεων. Οι χρήστες πλέον του διαδικτύου τα χρησιμοποιούν τόσο για να εκφράσουν τις απόψεις τους, όσο και για να ενημερωθούν για τις απόψεις των άλλων. Ο όγκος των πληροφοριών που μπορούν να αντληθούν από αυτά είναι τεράστιος και τα αποτελέσματα που μπορούν να διεξαχθούν από την επεξεργασίας τους πολύτιμα.

Το ερώτημα το οποίο έθεσε η παρούσα διπλωματική εργασία είναι αν και κατά πόσο τα άτομα επηρεάζονται ως προς τις επενδυτικές τους αποφάσεις από σχόλια τρίτων στα μέσα κοινωνικής δικτύωσης και στο πως ο τραπεζικός / επενδυτικός κλάδος μπορεί να χρησιμοποιήσει προς όφελος του τις γνώσεις του πάνω σε αυτό.

Στο πλαίσιο αυτό, αφού έγινε αναφορά στα big data και τις μεθόδους της ανάλυσης συναισθήματος και της μηχανικής μάθησης, για να εξετάσουμε τις τεχνικές κατηγοριοποίησης κειμένου, προχωρήσαμε στην εξόρυξη και την επεξεργασία tweets από το κοινωνικό δίκτυο Twitter. Από τα αποτελέσματα που εξήχθησαν διαπιστώσαμε ότι τόσο με τη χρήση του αλγόριθμου Naïve Bayes, όσο και με τη χρήση του αλγόριθμου Deep Learning, η ακρίβεια των αποτελεσμάτων ήταν χαμηλή και το σφάλμα υψηλό. Το γεγονός αυτό μας οδηγεί στο συμπέρασμα ότι χρειάζεται βελτίωση των αλγόριθμων μηχανικής μάθησης, ώστε να έχουμε πιο ορθά αποτελέσματα κατά τη συναισθηματική ανάλυση κειμένων.

Στη συνέχεια έγινε αναφορά στη συμπεριφορική χρηματοικονομική, αναλύθηκαν τα είδη των επενδυτών και η επιρροή της ετερογένειας στις κινήσεις της αγοράς. Αναφερθήκαμε στην ψυχολογία των ατόμων και στο πως μπορεί να επηρεάσει τις επενδυτικές τους αποφάσεις. Για να ερευνήσουμε τις αντιδράσεις των ατόμων αναφορικά με το κλίμα που επικρατεί στο διαδίκτυο, προβήκαμε στη διεξαγωγή ερωτηματολογίου. Από την ανάλυση των απαντήσεων και τα αποτελέσματα που εξάγαμε, συνεπάγεται ότι η πλειοψηφία των ατόμων, επηρεάζεται ως προς τις επενδυτικές της αποφάσεις, από τα σχόλια που διαβάζει στα μέσα κοινωνικής δικτύωσης, ιδίως τα αρνητικά από αυτά. Το γεγονός ότι οι απόψεις τρίτων στο διαδίκτυο μπορούν να επηρεάσουν τη συνεργασία των ατόμων με τράπεζες και ασφαλιστικές / επενδυτικές εταιρείες, επιβεβαιώνει τη θεωρία της ψυχολογίας των μαζών. Οι διαφορετικές απαντήσεις των ατόμων, ως προς το μέγεθος του ρίσκου που

είναι διατεθειμένοι να αναλάβουν, είναι βασικό χαρακτηριστικό οικονομικής ετερογένειας, η οποία μπορεί να επηρεάσει τις τιμές των αξιογράφων και άρα την απόδοση της αγοράς. Επιπρόσθετα έχουμε σημαντικές ενδείξεις ότι η επίδραση του οικονομικού κλίματος που επικρατεί στο διαδίκτυο αποτυπώνεται εμφανώς στη συμπεριφορά των επενδυτών.

Για την καλύτερη τεκμηρίωση των αποτελεσμάτων, απαιτείται περαιτέρω έρευνα με διευρυμένο δείγμα και εφαρμογή περισσότερων και πιο εξειδικευμένων αλγόριθμων μηχανικής μάθησης και συναισθηματικής ανάλυσης κειμένων.

Βιβλιογραφία – Αναφορές

Bing Liu., 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Bo Pang and Lillian Lee., 2008 *Opinion Mining and Sentiment Analysis*. Foundations and Trends in Information Retrieval 2(1-2), pp 1-135.

Peter D. Turney., 2002. *Thumbs up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews.* Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics.

Bo Pang, Lillian Lee and Shivakumar Vaithyanathan., 2002. *Thumbs Up? Sentiment Classification using Machine Learning Techniques.* Proceedings of the ACL-02 conference on Empirical Methods in Natural Language Processing-Volume 10. Association for Computational Linguistics.

Andrea Esuli and Fabrizio Sebastiani., 2006. *SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining.* Proceedings of LREC. Vol. 6.

Taboada et al., 2011. *Lexicon-Based Methods for Sentiment Analysis*. ACL. pp. 267-307.

Kennedy and Inkpen., 2006. *Sentiment classification of movie and product reviews using contextual valence shifters*. Computational Intelligence. pp. 110-125.

Polanyi and Zaenen., 2006. *Contextual valence shifters.* In Janyce Wiebe, editor, Computing Attitude and Affect in Text: Theory and Applications. Springer. pp. 1-10.

Poria et al., 2015. *Sentiment Data Flow Analysis by Means of Dynamic Linguistic Patterns*. In IEEE Computational Intelligence Magazine 10. pp. 26-36.

Phil Simon., 2013. *Too Big to Ignore: The Business Case for Big Data*. Wiley,. ISBN 978-1-118-63817-0.

Bing, L., 2011. *Web Data Mining – Exploring Hyperlinks, Contents and Usage Data*. Springer.

Bishop, C. M., 2006. *Pattern Recognition and Machine Learning*. Springer, ISBN 0-387-31073-8.

Mayer – Schonberger, V., Cukier, K., 2013. *Big data: a revolution that will transform how we live, work, and think.* New York: Eamon Dolan/Houghton Mifflin Harcourt.

Khan, S., Shiraz, M., Abdul Wahab, A., Gani, A., Han, Q., BinAbdul Rahman, Z., 2014. *A comprehensive review on adaptabilityof network forensics frameworks for mobile cloud computing.* The Scientific World Journal.

Wang, H., Xu, Z., Fujita, H., Liu, S., 2016. *Towards felicitous decision making: An overview on challenges and trends of Big Data.* Information Sciences, 367–368(1), pp. 747–765.

Chen, M., Mao, S., Liu, Y., 2014. *Big Data: A survey.* Mobile Netw. Appl., 19:1, pp. 171–209.

Yaqoob, I., Hashem, I., Gani, A., et al., 2016. *Big data: From beginning to future.* International Journal of Information Management, 36(6), pp. 1231–1247.

Βλαχάβας, Ι., Κεφαλάς, Π., Βασειλιάδης, Ν., Ρεφανίδης, Ι., Κόκκορας, Φ., Σακελλαρίου, Η., 2002. *Τεχνητή Νοημοσύνη*. Γαρταγάνη.

G. Tsoumakas, I. Katakis, I. Vlahavas., 2008. *Effective and Efficient Multilabel Classification in Domains with Large Number of Labels.* Proc. ECML/PKDD 2008 Workshop on Mining Multidimensional Data (MMD'08), pages 30-44, Antwerp, Belgium.

M. Hearst., 1999. *Untangling text data minin.* In Proc. of ACL ΄99 the 37th Annual Meating of Association for Caomputational Linguistics,.

Bo Pang and Lillian Lee., 2005. *Seeing Stars: Exploiting class relationships for sentiment categorization with respect to rating scales*. In Proceedings of the 43rd annual meeting on association for computational linguistics. pp. 115-124.

Moshe Koppel and Jonathan Schler., 2005. *The Importance of Neutral Examples for Learning Sentiment".*

Simon Haykin., 2009. *Neural Networks and Learning Machines.*

Alan M. F. Souza, Fabio M. Soares, Bostjan Kaluza, Yusuke Sugomori., 2017. *Deep Learning: Practical Neural Networks with Java*.

Shifrin, H., 2000. *Beyond greed and fear. Understanding behavioral finance and the psychology of investing.* Harvard Business School Press.

Σπύρου, Σ., 2009. *Συμπεριφορική Χρηματοοικονομική*. Μπένου.

Barberis, N.C., and Thaler, R. H., 2003. *A Survey of Behavioral Finance.* In: Constantinides, G. M. and Harris, M. and Stulz, R. M. (eds). Handbook of the 496 Economics of Finance: Volume 1B, Financial Markets and Asset Pricing. Elsevier North Holland, Chapter 18, pp. 1053-1128

LeBon, Gustave., 1895. *Ψυχολογία των μαζών*. Ελλ. Μτφρ., Ι. Σ. Χριστοδούλου., 1990. Ζήτρος.

Freud, Sigmund., 1921. *Ψυχολογία των μαζών και ανάλυση του Εγώ*. Ελλ. Μτφρ., Πατσογιάννης, Βασίλειος., 2014. Πλέθρον.

Arellano, Manuel., 2009. *Static Panel Data Models.* Class notes.

Chiarella, Carl. & Dieci, Roberto. & He, Xue-Zhong., 2007. *Do heterogeneous beliefs diversity market risk?.* The European Journal of Finance, 17:3, p. 241-258.

Muth, John., 1961. *Rational Expectations and thw Theory of Price Movements.* Econometrica, Vol 29, No.3, p. 315-335.

Fama, F. Eugene., 1970. *Efficient Capital Markets: A Review of Theory and Empirical Work.* Journal of Finance, Vol. 25, p. 383-417.

Hommes, Cars., 2005. *Heterogeneous agent models.* Tinbergen Institute Discussion Paper, CeNDEF, Faculty of Economics and Econometrics, University of Amsterdam, and Tinbergen Institute.

Chiarella, C., and He, X., 2002. *An adaptive model on assete pricing and wealth dynamics with heterogeneouw trading strategies.* School of Finance and Economics.

Lee, A., Lee, C., 2006. *Encyclopedia of Finance.* Springer Science and Business Media LLC.

Αλεξάκης, Χ. , Ξανθάκης, Μ., 2008. *Συμπεριφορική χρηματοοικονομική.* Σταμούλης.

Χόρνεϋ, Κάρεν., 1990. *Ο ρόλος της νεύρωσης και η ανέλιξη της προσωπικότητας – Ο αγώνας για την αυτοκαταξίωση.* Το περιβόλι της Παναγιάς.

Παρασκευόπουλος Ι., 1999. *Ερωτηματολόγιο διαπροσωπικής και ενδοπροσωπικής προσαρμογής.* Ελληνικά Γράμματα.

Φίλιας, Β., 2001. *Εισαγωγή στη Μεθοδολογία και τις τεχνικές των Κοινωνικών* Ερευνών. Gutenberg.