

Πανεπιστήμιο Δυτικής Αττικής  
Σχολή Μηχανικών  
Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών  
ΠΜΣ Επιστήμη και Τεχνολογία της Πληροφορικής και των Υπολογιστών



Διπλωματική  
Εργασία

Εξόρυξη Δεδομένων στα Πληροφοριακά Συστήματα της  
Δημόσιας Διοίκησης



Μεταπτυχιακή φοιτήτρια : Ειρήνη Μάγγα  
Αριθμός Μητρώου: mcse19008

Επιβλέποντες καθηγητές: Χρήστος Σκουρλάς, Αικατερίνη Μαρινάγη

Αθήνα Δεκέμβριος 2020

**Μέλη της τριμελούς επιτροπής εξέτασης:**

- 1) Χρήστος Σκουρλάς
- 2) Αικατερίνη Μαρινάγη
- 3) Βασίλειος Μάμαλης

## ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Η κάτωθι υπογεγραμμένη Μάγγα Ειρήνη του Ζήση, με αριθμό μητρώου mcse19008 φοιτήτρια του Προγράμματος Μεταπτυχιακών Σπουδών Επιστήμη και Τεχνολογία της Πληροφορικής και των Υπολογιστών του Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών της Σχολής Μηχανικών του Πανεπιστημίου Δυτικής Αττικής, δηλώνω ότι:

«Είμαι συγγραφέας αυτής της μεταπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Η Δηλούσα



## Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τους επιβλέποντες καθηγητές μου κ. Χρήστο Σκουρλά και κα. Αικατερίνη Μαρινάγη που μου έδωσαν την ευκαιρία να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα, να συνεργαστώ μαζί τους και να αποκομίσω πολύτιμες γνώσεις σε έναν εξαιρετικά σημαντικό τομέα για την πληροφορική αυτόν της Εξόρυξης Δεδομένων. Η πολύτιμη βοήθεια τους, η εύστοχη καθοδήγηση τους, οι συμβουλές τους αλλά και η εμπιστοσύνη που μου έδειξαν με βοήθησαν να εκπονήσω την παρούσα διπλωματική εργασία και να διευρύνω νέους επιστημονικούς και επαγγελματικούς ορίζοντες. Θα ήθελα επίσης να ευχαριστήσω τη Διοίκηση και όλους τους καθηγητές του μεταπτυχιακού προγράμματος Επιστήμη και Τεχνολογία της Πληροφορικής και των Υπολογιστών του Πανεπιστημίου Δυτικής Αττικής για όλες τις γνώσεις και την υποστήριξη που μου προσέφεραν. Τέλος θα ήθελα να ευχαριστήσω την οικογένεια μου, τους συναδέλφους μου και τους συμφοιτητές μου για την στήριξη τους σε όλη τη διάρκεια του μεταπτυχιακού προγράμματος.

## Περίληψη

---

Η εξέλιξη των Τεχνολογιών Πληροφοριών και Επικοινωνιών (ΤΠΕ), ειδικά με την ανάπτυξη του διαδικτύου και τη χρήση Πληροφοριακών Συστημάτων στο δημόσιο τομέα αύξησε σημαντικά τους όγκους δεδομένων, την ανάγκη για αποθήκευση, τον αριθμό του συναλλασσόμενου κοινού αλλά και την ανάγκη για επεξεργασία και ανάλυση δεδομένων. Η Δημόσια Διοίκηση με τη χρήση τεχνικών και εργαλείων εξόρυξης δεδομένων μπορεί να αναλύσει τον συνεχώς αυξανόμενο όγκο δεδομένων εξάγοντας χρήσιμη γνώση που θα την βοηθήσει να λάβει καλύτερες αποφάσεις. Ένας εύκολος και με χαμηλό κόστος τρόπο για να αντλήσει η Δημόσια Διοίκηση δεδομένα και να κατανοήσει καλύτερα τις ανάγκες των πολιτών είναι οι δημόσιες διαβουλεύσεις. Στην παρούσα εργασία, στο πλαίσιο πιλοτικής εφαρμογής, συγκεντρώθηκαν και μελετήθηκαν σχόλια που εμφανίζονται στον Διαδικτυακό Τόπο Διαβουλεύσεων του Υπουργείου Ψηφιακής Διακυβέρνησης και αφορούν στο σχέδιο διατάξεων του Γενικού Εμπορικού Μητρώου. Στις διατάξεις περιγράφεται η βελτίωση του Πληροφοριακού Συστήματος του Γενικού Εμπορικού Μητρώου (ΓΕ.Μ.Η.). Τα δεδομένα στη συνέχεια επεξεργάστηκαν με τους αλγορίθμους J48, Apriori, K-Means και Naive Bayes μέσω του λογισμικού Weka σε μια προσπάθεια να αναλυθεί το είδος του σχολιασμού, το συναίσθημα του σχολιαστή και το εάν προτείνει αλλαγές στον Νόμο δηλαδή αν συμμετέχει ενεργά στη βελτίωση του Νόμου βοηθώντας έτσι τη Δημόσια Διοίκηση να λάβει καλύτερες αποφάσεις.

**Λέξεις Κλειδιά:** Εξόρυξη Δεδομένων, Μεγάλα Δεδομένα, Πληροφοριακά Συστήματα, Δημόσια Διοίκηση, Εξόρυξη γνώμης, Ανάλυση συναισθήματος, Δημόσια Διαβούλευση, Weka, J48, Apriori, K-Means, Naive Bayes.

## Abstract

---

The evolution of Information and Communication Technologies (ICT), especially with the development of the Internet and the use of Information Systems in the public sector, has significantly increased the volume of data, the need for storage, the number of civilians' transactions and the need for data processing and analysis. Public Administration using data mining techniques and tools can analyze the ever-increasing volume of data by exporting useful knowledge that will help make better decisions. An easy and low cost way for the Public Administration to extract data and better understand the needs of the citizens is the public consultation. In the present work, in the context of a pilot application, comments that appear on the Consultation Website of the Ministry of Digital Government that concern the draft provisions of the Hellenic Business Registry were collected and studied. The provisions describe the improvement of the Information System of the Hellenic Business Registry. The data were then processed with the J48, Apriori, K-Means and Naive Bayes algorithms through Weka software in an attempt to analyze the type of commentary, the commentator's emotion and whether he/she proposes changes to the Law, ie whether the commentator actively participates in the Information Systems' improvement in order to help the Public Administration make better decisions.

**Keywords:** Data Mining, Big Data, Information Systems, Public Administration, Opinion Mining, Emotion Analysis, Public Consultation, Weka, J48, Apriori, K-Means, Naive Bayes.

## Περιεχόμενα

---

<b>1. Εισαγωγή.....</b>	<b>9</b>
1.1 Στόχοι της εργασίας και μεθοδολογία έρευνας .....	10
1.2 Θεματική Παρουσίαση Κεφαλαίων .....	11
<b>2. Τα Μεγάλα Δεδομένα στη Δημόσια Διοίκηση .....</b>	<b>12</b>
2.1 Μεγάλα Δεδομένα στη Δημόσια Διοίκηση.....	12
2.2 Προκλήσεις για τη Δημόσια Διοίκηση .....	13
2.3 Εφαρμογές Μεγάλων Δεδομένων και οφέλη για τη Δημόσια Διοίκηση .....	14
<b>3. Πληροφοριακά Συστήματα στη Δημόσια Διοίκηση.....</b>	<b>15</b>
3.1 Κατηγορίες Πληροφοριακών Συστημάτων στον Δημόσιο Τομέα .....	15
3.1.1 Εσωτερικά Πληροφοριακά Συστήματα .....	15
3.1.2 Ολοκληρωμένα Πληροφοριακά Συστήματα .....	15
3.1.3. Συστήματα Υποστήριξης Αποφάσεων .....	16
3.1.4 Εξωστρεφή Πληροφοριακά Συστήματα .....	16
3.1.5 Ταξινόμηση Πληροφοριακών Συστημάτων Δημόσιας Διοίκησης .....	16
3.2 Το Πληροφοριακό Σύστημα Γ.Ε.ΜΗ.....	17
3.3 Οφέλη για τη Δημόσια Διοίκηση και τους πολίτες από τη χρήση Πληροφοριακών Συστημάτων .....	18
3.4 Εμπόδια εφαρμογής των Πληροφοριακών Συστημάτων στη Δημόσια Διοίκηση.....	19
<b>4. Εξόρυξη Γνώμης – Ανάλυση Συναισθήματος.....</b>	<b>22</b>
4.1 Εξόρυξη Γνώμης Πολιτών.....	22
4.2 Πλεονεκτήματα της Εξόρυξης Γνώμης από τα Ηλεκτρονικά Μέσα.....	22
4.3 Προκλήσεις για τη Δημόσια Διοίκηση κατά την Εξόρυξη Γνώμης – Ανάλυση Συναισθήματος .....	23
<b>5. Ηλεκτρονική Διακυβέρνηση και Δημόσια Διαβούλευση .....</b>	<b>25</b>
5.1 Εισαγωγή.....	25
5.2 Ηλεκτρονική Διακυβέρνηση.....	25
5.3 Δημόσια Διαβούλευση .....	26
<b>6. Εξόρυξη δεδομένων με το λογισμικό WEKA – Πειραματικό μέρος .....</b>	<b>27</b>
6.1 Εισαγωγή στο πειραματικό μέρος .....	27
6.2 Μεθοδολογία συλλογής και επεξεργασίας δεδομένων.....	28
6.3 Το λογισμικό WEKA.....	29
6.4 Η δομή του αρχείου δεδομένων.....	30
6.5 Έναρξη του λογισμικού WEKA .....	32
6.6 Επεξεργασία δεδομένων με το λογισμικό WEKA .....	33
6.6.1 Άνοιγμα αρχείου .....	33
6.6.2 Εφαρμογή αλγορίθμου J4.8 .....	35
6.6.3 Εφαρμογή αλγορίθμου Apriori.....	41
6.6.4 Εφαρμογή αλγορίθμου K-Means .....	47
6.6.5 Εφαρμογή αλγορίθμου Naive Bayes .....	50
<b>Συμπεράσματα .....</b>	<b>52</b>

## Βιβλιογραφία .....54

i) Αρχικό αρχείο arff.....	58
ii) Αρχείο arff μετά από αφαίρεση μετά από αφαίρεση των attributes Commentator και CommentatorGender:.....	61

### Ευρετήριο Εικόνων:

Εικόνα 1 : Σελίδα Διαδικτυακού Τύπου Διαβουλεύσεων .....	27
Εικόνα 2 : Πλοήγηση στη Διαβούλευση.....	28
Εικόνα 3 : Λογότυπο Weka.....	30
Εικόνα 4 : Απόσπασμα αρχείου arff.....	31
Εικόνα 5 : Έναρξη WEKA Version 3.8.4 .....	32
Εικόνα 6 : WEKA Explorer .....	33
Εικόνα 7 : Άνοιγμα αρχείου comments i.....	34
Εικόνα 8 : Ιστογράμματα όλων των χαρακτηριστικών.....	34
Εικόνα 9 : Εφαρμογή αλγορίθμου J4.8 για την κλάση comment (χωρίς φίλτρο).....	36
Εικόνα 10 : Οπτικοποίηση δέντρου αλγορίθμου J4.8 για την κλάση comment.....	37
(χωρίς φίλτρο) .....	37
Εικόνα 11 : Εφαρμογή φίλτρου attribute selection .....	37
Εικόνα 12 : Εφαρμογή αλγορίθμου J4.8 (φίλτρο attribute selection) .....	38
Εικόνα 13 : Οπτικοποίηση δέντρου αλγορίθμου J4.8 (φίλτρο attribute selection) .....	39
Εικόνα 14 : Εφαρμογή αλγορίθμου J4.8 για την κλάση emotion (χωρίς φίλτρο) .....	40
Εικόνα 15 : Οπτικοποίηση δέντρου αλγορίθμου J4.8 για την κλάση emotion.....	41
(χωρίς φίλτρο) .....	41
Εικόνα 16 : Εκτέλεση αλγορίθμου Apriori με default τιμές.....	42
Εικόνα 17 : Αποτελέσματα εκτέλεσης αλγορίθμου Apriori με default τιμές .....	43
Εικόνα 18 : Εκτέλεση αλγορίθμου Apriori με min support 0.2 και min metric 0.8 .....	44
Εικόνα 19 : Αποτελέσματα εκτέλεσης αλγορίθμου Apriori με min support 0.2 και min metric 0.8.....	45
Εικόνα 20 : Εκτέλεση αλγορίθμου Apriori με default τιμές και εξαίρεση των attributes Commentator και CommentatorGender .....	46
Εικόνα 21 : Εκτέλεση αλγορίθμου K-Means .....	48
Εικόνα 22 : Εκτέλεση αλγορίθμου K-means με εξαίρεση των attributes Commentator και CommentatorGender .....	49
Εικόνα 23 : Εκτέλεση αλγορίθμου Naive Bayes.....	50



# 1 Εισαγωγή

---

Τα Μεγάλα Δεδομένα (Big Data), έστω και αν δεν το αντιλαμβανόμαστε, έχουν συνδεθεί με πολλές από τις δραστηριότητες της καθημερινής ζωής. Δεδομένα πλέον συλλέγονται από διάφορες πηγές όπως διαδικτυακές αγορές, κοινωνικά μέσα δικτύωσης, οικονομικές συναλλαγές, κλικ στο διαδίκτυο, δοσοληψίες με τον δημόσιο τομέα κ.α. Τα δεδομένα αυτά φανερώνουν ατομικές συμπεριφορές, ενέργειες και προτιμήσεις που με την κατάλληλη επεξεργασία μπορούν να εξάγουν πολύτιμη γνώση.

Η ανάλυση των Μεγάλων Δεδομένων έχει απασχολήσει οργανισμούς σε διάφορους τομείς οι οποίοι θέλουν να αξιοποιήσουν τα δεδομένα τους στο έπακρο έτσι ώστε να βελτιώσουν τις υπηρεσίες τους, να αυξήσουν την παραγωγικότητα, να κατανοήσουν καλύτερα τις ανάγκες των πελατών, να στοχεύσουν σε συγκεκριμένο καταναλωτικό κοινό, να βελτιώσουν τις έρευνες τους, να κάνουν καλύτερη κατανομή των πόρων τους και να αναδείξουν νέες ιδέες και τεχνολογίες. Πολλές μελέτες πραγματοποιήθηκαν προκειμένου να καλυφθούν αυτές οι ανάγκες όμως αρκετοί μελετητές συνειδητοποίησαν ότι εκτός από τον ιδιωτικό τομέα τέτοιου είδους ανάγκες εντοπίζονται και στον δημόσιο.

Τα τελευταία χρόνια η Δημόσια Διοίκηση προκειμένου να απλουστεύσει και να επιταχύνει διαδικασίες, να διασφαλίσει τα συμφέροντα των πολιτών, να ενισχύσει τη διαφάνεια και να μειώσει τα λειτουργικά έξοδα, αντικατέστησε πολλές χειρόγραφες διαδικασίες με ηλεκτρονικές ή βελτίωσε τις ήδη υπάρχουσες ηλεκτρονικές υπηρεσίες. Αυτό είχε ως αποτέλεσμα τη δημιουργία πολλών Πληροφοριακών Συστημάτων στο δημόσιο τομέα. Η ύπαρξη αυτών των συστημάτων οδήγησε στη συλλογή τεράστιων όγκων δεδομένων μεγάλης κλίμακας (Big Data). Προκειμένου να βελτιωθούν οι παρεχόμενες υπηρεσίες και να βελτιστοποιηθούν οι αποφάσεις που λαμβάνει η Δημόσια Διοίκηση χρησιμοποιούνται στα δεδομένα αυτά τεχνικές εξόρυξης δεδομένων.

Μια καλή πρακτική για την κατανόηση των αναγκών των πολιτών είναι η ύπαρξη ενός ανοιχτού διαύλου επικοινωνίας με τους πολίτες όπου θα μπορούν να εκφέρουν τη γνώμη τους, θα προτείνουν τρόπους βελτίωσης των υπηρεσιών και θα επικοινωνήσουν τις τυχόν δυσκολίες που συναντούν στη χρήση ηλεκτρονικών υπηρεσιών. Κάτι τέτοιο μπορεί να επιτευχθεί με τις δημόσιες διαβουλεύσεις.

## 1.1 Στόχος της εργασίας και μεθοδολογία έρευνας

Ο στόχος της παρούσας εργασίας είναι να παρουσιαστεί ο ρόλος των μεγάλων δεδομένων και της εξόρυξης γνώσης στη δημόσια διοίκηση και στη συνέχεια να εφαρμοστούν μέθοδοι εξόρυξης δεδομένων για την εξόρυξη της γνώμης των πολιτών που συμμετέχουν σε δημόσια διαβούλευση της δημόσιας διοίκησης.

Συγκεκριμένα οι άξονες ανάπτυξης της εργασίας παρουσιάζονται συνοπτικά παρακάτω:

- **Αναζήτηση - Έρευνα:**

Βιβλιογραφική αναζήτηση σε έργα μελετητών και συγγραφέων σχετικά με τα Μεγάλα Δεδομένα στη δημόσια διοίκηση, τα Πληροφοριακά Συστήματα στη Δημόσια Διοίκηση, την εξόρυξη γνώμης στη Δημόσια Διοίκηση, και τέλος τις τεχνικές και αλγορίθμους εξόρυξης δεδομένων.

- **Εφαρμογή μεθόδων εξόρυξης δεδομένων:**

Εφαρμογή επιλεγμένων αλγορίθμων που υλοποιούν διαφορετικές μεθόδους εξόρυξης δεδομένων για την επεξεργασία των σχολίων των πολιτών που συμμετέχουν σε δημόσια διαβούλευση της δημοσίας διοίκησης. Σκοπός είναι να ταξινομηθούν οι γνώμες των πολιτών με τη μέθοδο της ταξινόμησης (classification), να αποκαλυφθούν ομάδες παρόμοιων παρατηρήσεων με τη μέθοδο της συσταδοποίησης (clustering) και να εντοπιστούν μοτίβα συμπεριφορών μέσω κανόνων συσχέτισης (association rules). Τα αποτελέσματα αποτυπώνουν τη γνώμη των πολιτών και μπορούν να βοηθήσουν τη δημόσια διοίκηση να λάβει αποφάσεις και να διαμορφώσει κατάλληλα τις δράσεις της.

### Μεθοδολογία έρευνας:

Στο πειραματικό μέρος της παρούσας εργασίας συγκεντρώθηκαν τα σχόλια των πολιτών που εμφανίζονται στον Διαδικτυακό Τύπο Διαβουλεύσεων του Υπουργείου Ψηφιακής Διακυβέρνησης της Ελλάδας [1]. Τα σχόλια αφορούν τη Δημόσια Διαβούλευση των διατάξεων του Νόμου «Σχέδιο διατάξεων για το Γενικό Εμπορικό Μητρώο» που περιγράφει τη βελτίωση του Πληροφοριακού Συστήματος του Γενικού Εμπορικού Μητρώου (ΓΕ.Μ.Η.). Συνολικά αντλήθηκαν και μελετήθηκαν 153 σχόλια. Τα σχόλια κατηγοριοποιήθηκαν χειροκίνητα ως προς τα άρθρα του Νόμου, το είδος του σχολιαστή, το φύλο του σχολιαστή, το είδος σχολιασμού, το συναίσθημα και αν ο σχολιαστής προτείνει κάποια βελτίωση στις διατάξεις του Νόμου.

Τα δεδομένα επεξεργάστηκαν με το λογισμικό Weka και αξιολογήθηκαν τα αποτελέσματα τους βάσει των ακόλουθων αλγορίθμων: α) J48 και Naive Bayes για τη μέθοδο της ταξινόμησης (classification), β) Apriori που υλοποιεί τη μέθοδο των Κανόνων Συσχέτισης (Association Rules) και γ) K-Means που υλοποιεί τη μέθοδο της συσταδοποίησης (clustering). Συγκεκριμένα:

Ο αλγόριθμος C4.5/J4.8 χρησιμοποιήθηκε για να ταξινομηθούν τα νέα δεδομένα σε κατηγορίες, για να αποκαλυφθεί πως επηρεάζει άλλα χαρακτηριστικά το συναίσθημα των σχολιαστών και για να παραχθεί ένα οπτικοποιημένο δέντρο αποφάσεων που θα

είναι κατανοητό ακόμη και από μη τεχνικούς π.χ. στελέχη της Δημόσιας Διοίκησης. Στο ίδιο σύνολο δεδομένων εφαρμόστηκε ο αλγόριθμος Naive Bayes προκειμένου να δοκιμαστεί ο αλγόριθμος C4.5/J4.8 συγκρίνοντας τα αποτελέσματα δεδομένων με γνωστές τιμές-στόχους με αποτελέσματα δεδομένων που έχουν προγνωστικές τιμές. Ο αλγόριθμος Apriori εφαρμόστηκε προκειμένου να αποκαλυφθούν κανόνες συσχέτισης και για να αναδειχθούν οι σχέσεις μεταξύ των δεδομένων. Ο αλγόριθμος K-means χρησιμοποιήθηκε για να αποκαλυφθούν ομάδες μέσα στο σύνολο των δεδομένων.

## 1.2 Θεματική Παρουσίαση Κεφαλαίων

Η διάρθρωση της παρούσας εργασίας, όπως αυτή λαμβάνει χώρα, καταγράφεται παρακάτω:

Στο **2<sup>ο</sup> Κεφάλαιο** γίνεται αναφορά στα Μεγάλα Δεδομένα στη δημόσια διοίκηση. Παρουσιάζονται οι προκλήσεις που αντιμετωπίζει η δημόσια διοίκηση στη διαχείριση και εξόρυξη των Μεγάλων Δεδομένων και αναλύονται τα οφέλη που αποκομίζει η δημόσια διοίκηση από την επεξεργασία των Μεγάλων Δεδομένων.

Στο **3<sup>ο</sup> Κεφάλαιο** καταγράφονται οι κυριότερες κατηγορίες Πληροφοριακών Συστημάτων που χρησιμοποιούνται στο δημόσιο τομέα, παρουσιάζεται το Πληροφοριακό Σύστημα Γ.Ε.ΜΗ. και αναλύονται τα οφέλη και τα εμπόδια που προκύπτουν από τη χρήση Πληροφοριακών Συστημάτων για τη δημόσια διοίκηση.

Στο **4<sup>ο</sup> Κεφάλαιο** γίνεται λόγος για την εξόρυξη γνώμης των πολιτών, αναλύονται τα πλεονεκτήματα της εξόρυξης γνώμης από τα ηλεκτρονικά μέσα και παρουσιάζονται οι προκλήσεις για τη δημόσια διοίκηση κατά την εξόρυξη γνώμης – ανάλυση συναισθήματος.

Στο **5<sup>ο</sup> Κεφάλαιο** καταγράφεται ο ορισμός της ηλεκτρονικής διακυβέρνησης και ο στόχος της, παρουσιάζεται ο ορισμός της δημόσιας διαβούλευσης και αναλύεται ο σκοπός της.

Στο **6<sup>ο</sup> Κεφάλαιο** παρουσιάζονται οι αλγόριθμοι που χρησιμοποιήθηκαν και το πειραματικό μέρος της εργασίας.

**Τέλος**, παρουσιάζονται τα συμπεράσματα της εργασίας, καταγράφεται η βιβλιογραφία που χρησιμοποιήθηκε για την ανάπτυξη της παρούσας εργασίας και παρατίθεται το παράρτημα.

# 2

## Τα Μεγάλα Δεδομένα στη Δημόσια Διοίκηση

### 2.1 Μεγάλα Δεδομένα στη Δημόσια Διοίκηση

Τα τελευταία χρόνια ο όρος Μεγάλα Δεδομένα (Big Data) συναντάται ολοένα και πιο συχνά στο δημόσιο τομέα καθώς οι ανάγκες για αρχειοθέτηση, επεξεργασία και ανάλυση μεγάλων όγκων δεδομένων έχει αυξηθεί.

Οι όγκοι δεδομένων που καλείται πλέον η Δημόσια Διοίκηση να διαχειριστεί είναι τεράστιοι καθώς ψηφιοποιήθηκαν πολλά έντυπα έγγραφα, η χρήση αποκλειστικά ψηφιακών εγγράφων και αιτήσεων έχει καθιερωθεί στις περισσότερες υπηρεσίες του δημοσίου, αναπτύχθηκαν πολλά πληροφοριακά συστήματα και η αυξημένη χρήση ηλεκτρονικών συσκευών και διαδικτύου από τους πολίτες συνέβαλλε στην έγκαιρη ενημέρωση τους και στη συμμετοχή τους σε περισσότερα προγράμματα του δημοσίου (π.χ. επιδοτήσεις, επιχορηγήσεις κλπ) αυξάνοντας έτσι τον αριθμό του συναλλασσόμενου κοινού.

Με την εξέλιξη των ΤΠΕ ( Τεχνολογιών Πληροφοριών και Επικοινωνιών) οι όγκοι των δεδομένων καθώς και η ανάγκη για αποθήκευση και επεξεργασία αυξήθηκε σημαντικά. Η εξέλιξη στις ΤΠΕ βοήθησαν στη δημιουργία περισσότερων δεδομένων και οι τεχνολογίες υπολογιστικού νέφους συνέβαλλαν στην εύκολη πρόσβαση και αποθήκευση τους [2]. Η Δημόσια Διοίκηση χρησιμοποιεί ΤΠΕ προκειμένου να διαχειριστεί πολύπλοκες εσωτερικές διαδικασίες αλλά και να λάβει προτάσεις βελτίωσης από τους πολίτες.

Με την ραγδαία ανάπτυξη του διαδικτύου η Δημόσια Διοίκηση έγινε περισσότερο εξωστρεφής και αφουγκράζεται τη γνώμη των πολιτών οι οποίοι πλέον μπορούν να συμμετέχουν ενεργά. Η συμμετοχή των πολιτών σε αρκετές περιπτώσεις έχει δώσει λύσεις σε προβλήματα που απασχολούσαν τη Δημόσια Διοίκηση για χρόνια. Με την αντικατάσταση του web 1.0 με το web 2.0 και την ανάπτυξη των μέσων κοινωνικής δικτύωσης περισσότεροι πολίτες μπορούν να συμμετέχουν ενεργά πάνω σε μια ευρεία γκάμα θεμάτων ακόμη και να θέτουν θέματα οι ίδιοι. Οι διαδικασίες λήψης αποφάσεων στη Δημόσια Διοίκηση έχουν υποστεί αλλαγές από τη στιγμή που οι πολίτες συμμετέχουν σε αυτές [3].

Οι παραδοσιακές στατιστικές μέθοδοι δεν μπορούν να ικανοποιήσουν τις απαιτήσεις για ανάλυση μεγάλων όγκων δεδομένων που παράγονται δυναμικά σε πραγματικό χρόνο. Η εξόρυξη δεδομένων μπορεί να ξεπεράσει αυτές τις προκλήσεις. Στην εξόρυξη δεδομένων υπάρχουν δύο τύποι αλληλοκαλυπτόμενων μοντέλων, αυτά που ταξινομούν και αυτά που προβλέπουν [30]. Μερικοί από τους διασημότερους αλγόριθμους είναι οι Bayesian Networks, Naive Bayes, Neural Networks, Support Vector Machines, C4.5, K-Nearest Neighbors, K-means και Apriori [31,32].

Με τη χρήση τεχνικών και εργαλείων εξόρυξης δεδομένων η Δημόσια Διοίκηση μπορεί να αντλήσει και να επεξεργαστεί δεδομένα από τον ιστό, τα μέσα κοινωνικής δικτύωσης, τα ιστολόγια και τα forum έτσι ώστε να λάβει κατάλληλες αποφάσεις [4]. Ο δημόσιος τομέας μπορεί να αυξήσει σημαντικά την παραγωγικότητα του, να βελτιώσει την αποδοτικότητα του και να αυξήσει τη διαφάνεια αν αξιοποιήσει τα Μεγάλα Δεδομένα τα οποία συγκεντρώνει [5].

Προκειμένου να βελτιωθούν οι παρεχόμενες υπηρεσίες, να επιτευχθεί καλύτερη κατανομή πόρων, να αναδειχθούν νέες ιδέες και τεχνολογίες και να βελτιστοποιηθούν οι αποφάσεις που λαμβάνει η Δημόσια Διοίκηση χρησιμοποιούνται στα δεδομένα αυτά τεχνικές εξόρυξης δεδομένων.

## 2.2 Προκλήσεις για τη Δημόσια Διοίκηση

Η εξόρυξη δεδομένων είναι μια ζωντανή επιστήμη η οποία εξελίσσεται διαρκώς και παράλληλα με τις τεχνολογίες. Όπως πολλές επιστήμες έτσι και αυτή αντιμετωπίζει αρκετές προκλήσεις ειδικά στον τομέα της Δημόσιας Διοίκησης όπου πρέπει να εφαρμόζονται νόμοι, να εξυπηρετείται το δημόσιο συμφέρον, να λαμβάνονται υπόψη οι ανάγκες των πολιτών, να διασφαλίζεται η προστασία των προσωπικών δεδομένων των πολιτών και να ικανοποιούνται κοινωνικοί και ηθικοί περιορισμοί.

Η μεγαλύτερη τεχνική πρόκληση που πρέπει να αντιμετωπιστεί είναι αυτή του συνεχώς αυξανόμενου όγκου [6] τόσο στην αποθήκευση όσο και στην επεξεργασία. Εκτός από τον όγκο, προκλήσεις για τη διαχείριση των Μεγάλων Δεδομένων, αποτελούν και άλλα χαρακτηριστικά όπως η ταχύτητα και η ποικιλία [7] καθώς η ταχύτητα με την οποία τα δεδομένα παράγονται και οι διαφορετικοί τύποι των δεδομένων δημιουργούν προβλήματα. Προκειμένου να αντιμετωπιστούν τέτοιου είδους προβλήματα αναπτύχθηκαν αρκετές μέθοδοι όπως η εξόρυξη δεδομένων, η ελάττωση δεδομένων και η τεχνητή νοημοσύνη [8].

Στις Βάσεις Δεδομένων μεγάλης κλίμακας θα πρέπει να αντιμετωπιστεί η πρόκληση της οπτικοποίησης των δεδομένων [9] καθώς αρκετές φορές τα δεδομένα προέρχονται από διαφορετικές πηγές, με διαφορετική δομή και διαφορετική ανάλυση.

Η Δημόσια Διοίκηση έχει επίσης να αντιμετωπίσει την πρόκληση της μετατροπής των Μεγάλων Δεδομένων σε πολύτιμη γνώση αλλά και την σωστή αξιοποίηση των αποτελεσμάτων της ανάλυσης. Η πρόκληση αυτή αφορά στη διαθεσιμότητα των δεδομένων, στην ανάλυση και στην ανακάλυψη των σχέσεων που υπάρχουν ανάμεσα στα δεδομένα έτσι ώστε να παραχθεί πολύτιμη γνώση [10].

Μια ακόμη σημαντική πρόκληση είναι αυτή της ενοποίησης δεδομένων πληροφοριακών συστημάτων διαφορετικών υπηρεσιών του δημοσίου καθώς και η απουσία σε πολλές περιπτώσεις της κατάλληλης τεχνογνωσίας για ένα τέτοιο εγχείρημα. Το ίδιο πρόβλημα παρατηρείται και στα δεδομένα που προέρχονται από διαφορετικές πηγές όπως για παράδειγμα τον ιστό ή/και τα κοινωνικά μέσα. Προκειμένου να αποφευχθεί κάτι τέτοιο και να επιτευχθεί διαλειτουργικότητα ανάμεσα στα συστήματα του δημοσίου τομέα, η Δημόσια Διοίκηση θα πρέπει να τυποποιήσει έγγραφα, να δημιουργήσει πρωτόκολλα επικοινωνίας μεταξύ των υπηρεσιών του δημοσίου, να μειώσει την πολυπλοκότητα των διοικητικών διαδικασιών, να καταρτίσει τεχνολογικά τους εμπλεκόμενους που ασχολούνται με τη διαδικασία συγχώνευσης των πληροφοριακών συστημάτων, τα πληροφοριακά συστήματα θα πρέπει να σχεδιάζονται έτσι ώστε να είναι δυνατή η μεταφεριμότητα των δεδομένων χωρίς αυτά να αλλοιώνονται και θα πρέπει να δίνεται στην κάθε δημόσια υπηρεσία και η ανάλογη άδεια χρήσης δεδομένων μέσα στα πλαίσια πάντα των ευαίσθητων και προσωπικών δεδομένων.

## 2.3 Εφαρμογές Μεγάλων Δεδομένων και οφέλη για τη Δημόσια Διοίκηση

Η Δημόσια Διοίκηση συνειδητοποιεί όλο και περισσότερο τα οφέλη που μπορεί να αποκομίσει από την επεξεργασία των Μεγάλων Δεδομένων. Πλέον συλλέγονται τεράστιες ποσότητες δεδομένων που αφορούν σε μισθούς, συντάξεις, επιδόματα, επιχορηγήσεις, φόρους, συστήματα υγείας, συστήματα εκπαίδευσης, περιουσιακά στοιχεία κλπ.

Επίσης έχει γίνει κατανοητό ότι η ανοιχτή διακυβέρνηση και η συλλογή δεδομένων από την επικοινωνία με τους πολίτες μπορεί να φέρει πολύ θετικά αποτελέσματα όπως η καλύτερη των παρεχόμενων υπηρεσιών και η βελτίωση των αποφάσεων που λαμβάνει η Δημόσια Διοίκηση αυξάνοντας μάλιστα τη διαφάνεια και την εμπιστοσύνη των πολιτών.

Δεδομένα συγκεντρώνονται από αισθητήρες για την μέτρηση της κυκλοφορίας, την ρύπανση κλπ έτσι ώστε να ενημερώνονται οι πολίτες και να λαμβάνονται τα κατάλληλα περιβαλλοντολογικά μέτρα.

Η ανάλυση των Μεγάλων Δεδομένων μπορεί να βοηθήσει τη Δημόσια Διοίκηση να εντοπίσει και να αναγνωρίσει μοτίβα όπως για παράδειγμα να εντοπίσει μια οικονομική απάτη, να εποπτεύσει και να επιβάλλει πρόστιμα σε ιδιώτες, να ανιχνεύσει κυβερνοεγκλήματα, να αυξήσει την ασφάλεια των δεδομένων της ανιχνεύοντας συμπεριφορές χρήστη κλπ.

Όσο αποτελεσματικότερος γίνεται ο δημόσιος τομέας με τη χρήση των Μεγάλων Δεδομένων τόσο καλύτερα διαβιώνουν οι πολίτες αφού απαιτούνται λιγότεροι οικονομικοί πόροι και συνεπώς φόροι για τη λειτουργία του και αυτό έχει θετικό αντίκτυπο στην οικονομία και στην κοινωνία [11].

Με την συμβολή του Διαδικτύου των Πραγμάτων (IOT) οι έξυπνες πόλεις (smart cities) βελτιώνουν την ποιότητα ζωής των πολιτών [12].

Τα οφέλη από την χρήση Μεγάλων Δεδομένων είναι πολλά όμως επίσης δημιουργούν προκλήσεις οι οποίες μπορούν να κατευθύνουν το μέλλον τους [13]. Προκειμένου τα Μεγάλα Δεδομένα να χρησιμοποιούνται στο έπακρο τους θα πρέπει οι Δημόσιες Διοικήσεις όλων των χωρών να έρθουν σε μια κοινή συνεννόηση όσον αφορά τα νομικά θέματα που προκύπτουν από τα προσωπικά δεδομένα έτσι ώστε να μπορούν και μεταξύ τους να ανταλλάσσουν πληροφορίες, τεχνογνωσία και πρακτικές διοίκησης.



# 3

## Πληροφοριακά Συστήματα στη Δημόσια Διοίκηση

---

### 3.1 Κατηγορίες Πληροφοριακών Συστημάτων στον Δημόσιο Τομέα

Οι δύο βασικές κατηγορίες πληροφοριακών συστημάτων του δημόσιου τομέα είναι τα εσωτερικά και τα εξωτερικά πληροφοριακά συστήματα. Τα εσωτερικά πληροφοριακά συστήματα υποστηρίζουν εσωτερικές λειτουργίες μιας υπηρεσίας ενώ τα εξωτερικά δίνουν τη δυνατότητα πρόσβασης σε πολίτες ή σε άλλους δημόσιους φορείς.

#### 3.1.1 Εσωτερικά Πληροφοριακά Συστήματα

Τα συστήματα αυτά υποστηρίζουν τις εσωτερικές εργασίες των δημόσιων οργανισμών. Στα εσωτερικά πληροφοριακά συστήματα μπορούν να καταχωρηθούν δεδομένα, να επεξεργαστούν και να εξαχθούν συμπεράσματα με βάση συγκεκριμένα κριτήρια προκειμένου να παρουσιαστούν στα ανώτερα στελέχη. Επίσης μπορούν να αυτοματοποιηθούν ορισμένες τυποποιημένες εργασίες που κάποια δημόσια υπηρεσία εκτελεί συχνά. Ακόμη είναι δυνατή η παρακολούθηση των εργασιών και ο έλεγχος της παραγωγικότητας των υπαλλήλων από τους Προϊσταμένους. Συνηθισμένες λειτουργίες αυτών των συστημάτων είναι η διαχείριση του ανθρώπινου δυναμικού, η έκδοση μισθοδοσίας των υπαλλήλων, η διεκπεραίωση αιτημάτων πολιτών κ.α.

#### 3.1.2 Ολοκληρωμένα Πληροφοριακά Συστήματα

Τα Ολοκληρωμένα Πληροφοριακά Συστήματα (ΟΠΣ) αποτελούνται από διάφορα υποσυστήματα τα οποία εκτελούν μια σειρά εργασιών. Τα περισσότερα Ολοκληρωμένα Πληροφοριακά Συστήματα του δημοσίου δημιουργήθηκαν μετά από την ανάγκη ενοποίησης διάφορων μικρότερων συστημάτων που εκτελούσαν διαφορετικές εργασίες και ικανοποιούσαν τις ανάγκες της δημόσιας υπηρεσίας χωρίς όμως να έχουν διαλειτουργικότητα μεταξύ τους. Με την ένωση αυτών των συστημάτων συγκεντρώθηκαν και ενοποιήθηκαν δεδομένα, τα οποία σε αρκετές περιπτώσεις ήταν κοινά, με αποτέλεσμα να μειωθεί ο όγκος των δεδομένων, να είναι ευκολότερη η διαχείριση και η εξαγωγή πληροφοριών, να μειωθούν τα λειτουργικά έξοδα και να βελτιωθεί η αποτελεσματικότητα της υπηρεσίας. Τα Συστήματα Διαχείρισης Επιχειρησιακών Πόρων (Enterprise Resource Planning Systems - ERP) είναι παράδειγμα Ολοκληρωμένων Πληροφοριακών Συστημάτων. Στο άρθρο [40] τίθενται ζητήματα που αφορούν την αξιοποίηση των ERP συστημάτων στη Δημόσια Διοίκηση.

### **3.1.3. Συστήματα Υποστήριξης Αποφάσεων**

Τα Συστήματα Υποστήριξης Αποφάσεων είναι πληροφοριακά συστήματα που αναλύουν μεγάλους όγκους δεδομένων που βρίσκονται μέσα σε αποθήκες δεδομένων και βοηθούν τη Δημόσια Διοίκηση να λάβει αποφάσεις. Σκοπός αυτών των συστημάτων είναι να τροφοδοτήσουν τους αποφασίζοντες με πληροφορίες που σε συνδυασμό με τις εμπειρίες και τις προσωπικές τους γνώσεις να οδηγηθούν στο να λάβουν τις καλύτερες δυνατές αποφάσεις. Μέσω αυτών των συστημάτων μπορούν να εξαχθούν αποτελέσματα βασισμένα σε διάφορες απαιτήσεις π.χ. χρονικές, γεωγραφικές, συγκριτικές κ.α.

### **3.1.4 Εξωστρεφή Πληροφοριακά Συστήματα**

Τα εξωστρεφή πληροφοριακά συστήματα κάνουν δυνατή τη συναλλαγή της Δημόσιας Διοίκησης με τους πολίτες, τις επιχειρήσεις και άλλους φορείς του δημοσίου. Μέσω των συστημάτων αυτών βελτιώνεται η αποτελεσματικότητα των δημόσιων υπηρεσιών, ενισχύεται η εξωστρέφεια, γίνεται εύκολη η εξυπηρέτηση των πολιτών, καθίσταται δυνατή η συνεργασία μεταξύ των υπηρεσιών και συγκεντρώνονται δεδομένα από πολίτες και φορείς τα οποία αργότερα μπορούν να αξιοποιηθούν. Τέτοιου είδους συστήματα είναι τα: Συστήματα Ηλεκτρονικής Πληροφόρησης, Συστήματα Ηλεκτρονικών Συναλλαγών και τα Συστήματα Ηλεκτρονικής Συνεργασίας Δημόσιων Οργανισμών [14].

### **3.1.5 Ταξινόμηση Πληροφοριακών Συστημάτων Δημόσιας Διοίκησης**

Η ταξινόμηση των Πληροφοριακών Συστημάτων της Δημόσιας Διοίκησης περιγράφεται στο άρθρο [48] όπου τα ΠΣ ταξινομούνται ως εξής:

- Πληροφοριακά Συστήματα Ροών Εργασίας - «Workflow Management Systems-WfMs»
- Συστήματα Υποστήριξης Επιχείρησης - «Enterprise Information Systems- EIS»
- Ολοκλήρωση Επιχειρησιακών Εφαρμογών - «Enterprise Application Integration- EAI»
- Συστήματα Διαχείρισης Επιχειρησιακών Διαδικασιών - «Business Process Management Systems - BPMS»
- Επιχειρηματικής Ευφυΐας - «Business Intelligence - BI»
- Επιχειρηματικές Διαδικτυακές Πύλες - «Enterprise Portals»



### 3.2 Το Πληροφοριακό Σύστημα Γ.Ε.ΜΗ.

Για τις ανάγκες της παρούσας εργασίας μελετήθηκαν τα σχόλια της Δημόσιας Διαβούλευσης των διατάξεων του Νόμου που αφορούν στο Γενικό Εμπορικό Μητρώο. Στις διατάξεις περιγράφεται η βελτίωση του Πληροφοριακού Συστήματος του Γενικού Εμπορικού Μητρώου (Γ.Ε.Μ.Η.) και η επεξεργασία τους και τα αποτελέσματα της έρευνας παρουσιάζονται αναλυτικά στο κεφάλαιο 6.

Τα αρχεία του Γενικού Εμπορικού Μητρώου (Γ.Ε.Μ.Η.) τηρούνται ως βάση δεδομένων που έχει δημιουργηθεί και αποθηκευθεί σε πληροφοριακό σύστημα διαχείρισης το οποίο διανέμεται μέσω διαδικτυακού ιστότοπου.

Το πληροφοριακό Σύστημα Γ.Ε.ΜΗ. απαρτίζεται από:

α) Το Γενικό Ευρετήριο Επωνυμιών στο οποίο καταχωρίζονται οι εμπορικές ή εταιρικές επωνυμίες και οι διακριτικοί τίτλοι.

β) Την Μερίδα του κάθε υπόχρεου όπου καταχωρίζονται πράξεις, στοιχεία ή δηλώσεις για τα οποία υπάρχει υποχρέωση δημοσίευσης.

γ) Τον Φάκελο όπου υποβάλλονται οι αιτήσεις και τα δικαιολογητικά του υπόχρεου.

Πρόσβαση στο πληροφοριακό σύστημα έχουν πιστοποιημένοι χρήστες των Υ.Γ.Ε.ΜΗ, πιστοποιημένοι συμβολαιογράφοι Γ.Ε.ΜΗ και εξωτερικοί χρήστες π.χ. υπόχρεοι ή τρίτοι με εξουσιοδότηση για λήψη εγγράφων.

Το πληροφοριακό σύστημα Γ.Ε.ΜΗ. επικοινωνεί και με άλλα εθνικά μητρώα προκειμένου να:

- Λάβει πληροφορίες σχετικά με κεφαλαιουχικές εταιρίες έδρα των οποίων βρίσκεται σε άλλα κράτη - μέλη της Ευρωπαϊκής Ένωσης.
- Χορηγήσει πληροφορίες για κεφαλαιουχικές εταιρίες με έδρα την ημεδαπή.
- Απαντήσει σε ερωτήματα χρηστών που λαμβάνει στο Ευρωπαϊκό ενιαίο σημείο ηλεκτρονικής πρόσβασης του συστήματος διασύνδεσης μητρώων.

Επίσης υπάρχει διασύνδεση του συστήματος Γ.Ε.ΜΗ. με το Ολοκληρωμένο Πληροφοριακό Σύστημα Φορολογίας (Ο.Π.Σ.Φ.) TAXIS προκειμένου να λαμβάνονται και να διαβιβάζονται δεδομένα τα οποία έχουν καταχωρηθεί και στα δύο πληροφοριακά συστήματα [15].

### 3.3 Οφέλη για τη Δημόσια Διοίκηση και τους πολίτες από τη χρήση Πληροφοριακών Συστημάτων

Παλαιότερα με τη χρήση χειρόγραφων συστημάτων:

- Η δυνατότητα αναζήτησης της πορείας ενός εγγράφου ήταν δύσκολη έως ακατόρθωτη και εξαιρετικά χρονοβόρα.
- Υπήρχε πάντα η πιθανότητα παραποίησης των εγγράφων.
- Η πρόσβαση σε έγγραφα από μη εξουσιοδοτημένα άτομα ήταν δυνατή.
- Υπήρχε μεγάλη καθυστέρηση στη διακίνηση εγγράφων και απασχολούνταν μεγάλος αριθμός υπαλλήλων για αυτή.
- Η έγκριση των εγγράφων ήταν χρονοβόρα αφού η διακίνηση στην ιεραρχία γινόταν χέρι με χέρι.
- Το κόστος διακίνησης των εγγράφων εκτός υπηρεσίας ήταν υψηλό.
- Η κατανάλωση αναλώσιμων ήταν μεγάλη.
- Υπήρχε έλλειψη ασφάλειας των προσωπικών δεδομένων και της εμπιστευτικότητας αφού το έγγραφο άλλαζε χέρια πολλές φορές.
- Το ογκώδες έντυπο αρχείο καταλάμβανε μεγάλο χώρο στις υπηρεσίες του Δημοσίου οι οποίες πολλές φορές για την αποθήκευση των εγγράφων τους δαπανούσαν χρήματα για την φύλαξη του αρχείου σε ιδιώτες.

Με τη χρήση των Πληροφοριακών Συστημάτων η Δημόσια Διοίκηση αντικατέστησε πολλές χειρόγραφες διαδικασίες με ηλεκτρονικές ή βελτίωσε τις ήδη υπάρχουσες ηλεκτρονικές υπηρεσίες. Τα οφέλη για τη Δημόσια Διοίκηση και τους πολίτες συνοψίζονται παρακάτω:

- Οι πολίτες έχουν πρόσβαση σε πληροφορίες [41].
- Απλουστεύθηκαν και επιταχύνθηκαν διοικητικές διαδικασίες.
- Ενισχύθηκε η διαφάνεια [46].
- Αυξήθηκε η εμπιστοσύνη των πολιτών προς τους δημόσιους φορείς.
- Μειώθηκαν τα λειτουργικά έξοδα (αναλώσιμα, ταχυδρομικά, μεταφορικά κλπ).
- Τα Πληροφοριακά Συστήματα της Δημόσιας Διοίκησης ακολουθούν την ισχύουσα Νομοθεσία και το ιεραρχικό σύστημα της διοίκησης καλύπτοντας έτσι τις ανάγκες των δημόσιων υπηρεσιών και των πολιτών.
- Δημιουργήθηκαν διαδικτυακές υπηρεσίες [41]. Οι υπηρεσίες είναι διαθέσιμες ηλεκτρονικά και εξυπηρετούν μεγάλο αριθμό πολιτών όλο το 24ωρο χωρίς να απαιτείται η φυσική παρουσία.
- Τα στελέχη της δημόσιας υπηρεσίας δεν αναλώνουν τόσο χρόνο στην προσωπική εξυπηρέτηση του πολίτη αφού αυτός μπορεί να εξυπηρετηθεί ηλεκτρονικά.
- Οι πληροφορίες των Πληροφοριακών Συστημάτων είναι έγκυρες και έγκαιρες [42].

- Παρέχεται ασφάλεια στα προσωπικά δεδομένα και αυξάνεται η εμπιστευτικότητα καθώς μόνο εξουσιοδοτημένοι χρήστες έχουν πρόσβαση και υπάρχουν διακριτοί ρόλοι χρηστών με αντίστοιχες εξουσιοδοτήσεις.
- Τυποποιήθηκαν και επανασχεδιάστηκαν διαδικασίες [41] προκειμένου να ενσωματωθούν στα Πληροφοριακά Συστήματα.
- Διασφαλίζεται η ακεραιότητα των δεδομένων.
- Εξοικονομούνται οικονομικοί και ανθρώπινοι πόροι.
- Ενισχύθηκε η διασφάλιση της ποιότητας [41].
- Εξάγονται στατιστικά συμπεράσματα.
- Η Δημόσια Διοίκηση έγινε πιο υπεύθυνη [46].
- Είναι δυνατή η οπτικοποίηση των δεδομένων.
- Η Δημόσια Διοίκηση μπορεί να αξιοποιήσει τις πληροφορίες που συγκεντρώνει από τα Πληροφοριακά Συστήματα και με κατάλληλη επεξεργασία να λάβει πολύτιμη γνώση η οποία θα την βοηθήσει να βελτιώσει τις αποφάσεις και τις υπηρεσίες της.
- Η εξυπηρέτηση των πολιτών γίνεται με τον καλύτερο τρόπο αφού η Δημόσια Διοίκηση έχει κατανοήσει τις ανάγκες τους και διανέμει καλύτερα τους οικονομικούς και ανθρώπινους πόρους της.
- Ευαισθητοποιήθηκαν οι πολίτες και οι επιχειρήσεις για τη χρήση ηλεκτρονικών υπηρεσιών που προσφέρει η Δημόσια Διοίκηση [46].
- Η διαχείριση έργων βελτιώθηκε [41] διευκολύνοντας την παρακολούθηση της πορείας των έργων και την κατανομή των εργασιών.
- Διαφορετικά τμήματα μιας υπηρεσίας υποστηρίζονται και συνεργάζονται μεταξύ τους [43, 44, 45] αλλά και με άλλες υπηρεσίες. Οι δημόσιες υπηρεσίες επικοινωνούν μεταξύ τους ταχύτερα μειώνοντας την γραφειοκρατία.
- Η διασύνδεση πληροφοριακών συστημάτων διαφορετικών δημόσιων υπηρεσιών μπορεί να πολλαπλασιάσει όλα τα παραπάνω οφέλη.

### **3.4 Εμπόδια εφαρμογής των Πληροφοριακών Συστημάτων στη Δημόσια Διοίκηση**

Παρόλο που τα Πληροφοριακά Συστήματα χρησιμοποιούνται κατά κόρον στον δημόσιο τομέα και έχουν βοηθήσει σημαντικά στη βελτίωση του, υπάρχουν αρκετά εμπόδια στην εφαρμογή τους.

Στο άρθρο [51] αναφέρονται τέσσερα βασικά ζητήματα που πρέπει να λάμβάνουν υπόψη οι κυβερνήσεις όταν εξετάζουν την υλοποίηση έργων ηλεκτρονικής διακυβέρνησης:

A) Παρόλο που ένα έργο Πληροφοριακών Συστημάτων στο δημόσιο τομέα δεν σχετίζεται με τη δημιουργία μετοχικής αξίας ή την κερδοφορία, θα πρέπει να καθορίζονται οι παράγοντες επιτυχίας/αποτυχίας του έργου.

B) Θα πρέπει να καθορίζονται οι κύριες μεταβλητές βάσει των οποίων θα παρακολουθείται η απόδοση του έργου.

Γ) Θα πρέπει να καθοριστεί αν οι εκούσιες ή ακούσιες επιπτώσεις αφορούν το άτομο, τις επιχειρήσεις, την περιφέρεια ή τη χώρα.

Δ) Είναι βασικό να εντοπίζονται όλες οι ομάδες ενδιαφερόμενων που επηρεάζονται, καθώς έχει παρατηρηθεί ότι διαφορετικές ομάδες μπορεί να έχουν διαφορετικούς, συγκρουόμενους στόχους.

Σύμφωνα με το άρθρο [52] οι βασικότερες αιτίες που τα πληροφοριακά συστήματα μπορεί να αποτυγχάνουν σχετίζονται λιγότερο με τεχνικά ζητήματα, και περισσότερο με τις κοινωνικές, συμπεριφορικές και πολιτικές τους διαστάσεις.

Η σχεδίαση και η ανάλυση ενός Πληροφοριακού συστήματος είναι απαιτητική και χρονοβόρα καθώς σε αρκετές περιπτώσεις πρέπει στα Πληροφοριακά Συστήματα να εφαρμοστούν Νόμοι και σε περίπτωση που αντικατασταθεί ή τροποποιηθεί κάποιος Νόμος θα πρέπει να γίνουν οι αντίστοιχες αλλαγές ή ακόμη και να επανασχεδιαστεί εξ ολοκλήρου το Πληροφοριακό Σύστημα με αποτέλεσμα να πρέπει να δαπανηθούν αρκετά χρήματα και εργατώρες.

Οι απαιτήσεις που δίνονται από τα στελέχη της δημόσιας υπηρεσίας για τον σχεδιασμό του Πληροφοριακού Συστήματος θα πρέπει να καθοριστούν και να κατανοηθούν πλήρως από τους ειδικούς της πληροφορικής. Μια τέτοια συνεννόηση είναι πολλές φορές δύσκολο να επιτευχθεί καθώς ενώ τα στελέχη έχουν όλες τις απαραίτητες γνώσεις για τον καθορισμό των απαιτήσεων από την άλλη δεν γνωρίζουν τις δυνατότητες και τους περιορισμούς ενός Πληροφοριακού Συστήματος. Οι ειδικοί της πληροφορικής από την άλλη ενώ έχουν την τεχνογνωσία δεν γνωρίζουν πολλές φορές διοικητικά και νομικά θέματα ή πώς να καλύψουν τις εξειδικευμένες ανάγκες μιας δημόσιας υπηρεσίας.

Αν οι απαιτήσεις του Πληροφοριακού Συστήματος δεν καθοριστούν σωστά τότε αυτό δεν θα σχεδιαστεί σωστά και θα δημιουργήσει αρκετά προβλήματα στη λειτουργία της Δημόσιας Διοίκησης, δεν θα μπορέσει να καλύψει τις ανάγκες της Δημόσιας Διοίκησης και κατ' επέκταση των πολιτών και μπορεί να αποθηκεύει περιττές πληροφορίες ή να παραλείπει σημαντικές πληροφορίες.

Εάν το Πληροφοριακό Σύστημα δεν ικανοποιεί τις απαιτήσεις και τους στόχους της διοίκησης θα οδηγηθεί σε αποτυχία δηλαδή σε μη ικανοποίηση της διοίκησης, των χρηστών του συστήματος και θα έχει οικονομικό αντίκτυπο για το δημόσιο συμφέρον. Όσο πιο νωρίς εντοπιστούν και αντιμετωπιστούν τα προβλήματα τόσο μικρότερο θα είναι το κόστος που θα πρέπει να επωμιστεί ο δημόσιος τομέας. Τα ζητήματα που αφορούν τον καθορισμό των απαιτήσεων ενός έργου ανάπτυξης πληροφοριακού συστήματος αποτελούν αντικείμενο της επιστήμης της τεχνολογίας λογισμικού [53]. Η επιλογή του κατάλληλου μοντέλου κύκλου ζωής λογισμικού, αυξάνει το ποσοστό επιτυχίας του έργου.

Οι αλλαγές και οι αναβαθμίσεις ενός Πληροφοριακού Συστήματος είναι δαπανηρές, μπορεί να δυσαρεστήσουν τους χρήστες που είχαν συνηθίσει να εργάζονται σε ένα συγκεκριμένο περιβάλλον και μπορεί να οδηγήσουν σε ασυμβατότητα με το υλικό, πράγμα που σημαίνει μεγαλύτερο οικονομικό κόστος για τη δημόσια υπηρεσία.

Η συντήρηση ενός Πληροφοριακού Συστήματος έχει και αυτή κόστος, θα πρέπει να είναι συνεχής και μπορεί να απαιτήσει αναβάθμιση του εξοπλισμού. Ένα περίπλοκο Πληροφοριακό Σύστημα μπορεί να μην γίνει αποδεκτό από χρήστες που δεν έχουν την κατάλληλη εκπαίδευση και τεχνογνωσία και αυτό στην περίπτωση των στελεχών σημαίνει μείωση της παραγωγικότητας και στην περίπτωση των πολιτών άρνηση στη χρήση του με αποτέλεσμα να απευθύνεται στη δημόσια υπηρεσία για βοήθεια. Για αυτό κατά τον σχεδιασμό θα πρέπει να λαμβάνεται υπόψη η πολυπλοκότητα, η τεχνογνωσία των χρηστών, η ανθρώπινη συμπεριφορά και η ψυχολογία.

Ένα άλλο εμπόδιο είναι ότι τα Πληροφοριακά Συστήματα είναι τρωτά σε επιθέσεις και αυτό μπορεί να θέσει σε κίνδυνο δεδομένα τα οποία μπορεί να αλλοιωθούν ή να χρησιμοποιηθούν κακόβουλα για αυτό θα πρέπει να δίνεται ιδιαίτερη βαρύτητα στην ασφάλεια των συστημάτων. Οι κυβερνοεπιθέσεις έχει παρατηρηθεί ότι μπορούν να έχουν καταστροφικά αποτελέσματα [50].

Ακόμη ένα σοβαρό ζήτημα είναι αυτό της συλλογής και της επεξεργασίας ευαίσθητων προσωπικών δεδομένων, εργασίες οι οποίες θα πρέπει να γίνονται σύμφωνα με την αντίστοιχη νομοθεσία.

Βασικό είναι και το ζήτημα της πρόσβασης. Το άρθρο [49] αναφέρεται στα γενικότερα εμπόδια αποδοχής και εφαρμογής των Τεχνολογιών Πληροφορίας και Επικοινωνίας από τους πολίτες, όπως δημοσιεύονται και σε διεθνή έρευνα του 2018 [50]. Συγκεκριμένα αρκετοί πολίτες που έχουν πρόσβαση στο διαδίκτυο δεν χρησιμοποιούν υπηρεσίες της Δημόσιας Διοίκησης αλλά χρησιμοποιούν το διαδίκτυο μόνο για ψυχαγωγία και άλλοι πολίτες δεν έχουν πρόσβαση στο διαδίκτυο επειδή δεν έχουν την οικονομική δυνατότητα ή την κατάλληλη εκπαίδευση ή είναι σε προχωρημένη ηλικία ή το μορφωτικό τους επίπεδο δεν είναι το κατάλληλο. Επίσης αρκετές υπηρεσίες δεν έχουν σχεδιαστεί ώστε να είναι προσβάσιμες από Άτομα με Ειδικές Ανάγκες (Α.μ.Ε.Α.).

# 4

## Εξόρυξη Γνώμης – Ανάλυση Συναισθήματος

### 4.1 Εξόρυξη Γνώμης Πολιτών

Η γνώμη των πολιτών απασχολούσε ανέκαθεν τη Δημόσια Διοίκηση καθώς είχε πρωταρχικό ρόλο στη διαμόρφωση δημόσιων πολιτικών. Παλαιότερα με τη χρήση στατιστικών και δημοσκοπήσεων γινόταν ανάλυση της γνώμης των πολιτών και η ανάλυση αυτή σε πολλές περιπτώσεις είχε μεγάλη βαρύτητα π.χ. προεκλογικές δημοσκοπήσεις. Πλέον με την ανάπτυξη του διαδικτύου οι πολίτες μπορούν να εκφέρουν τη γνώμη τους σε ιστότοπους, μέσα κοινωνικής δικτύωσης, ηλεκτρονικούς δημόσιους διαλόγους, δημόσιες ηλεκτρονικές διαβουλεύσεις κ.α.

Η εξόρυξη γνώμης (opinion mining) [33] μπορεί να παρέχει στη Δημόσια Διοίκηση πληροφορίες που αφορούν τις πεποιθήσεις, αντιλήψεις και προτιμήσεις του πολίτη πάνω σε ένα συγκεκριμένο θέμα. Στόχος της εξόρυξης γνώμης είναι να εξαχθούν οι απόψεις των πολιτών, οι προτάσεις τους και το συναίσθημα τους έτσι ώστε η Δημόσια Διοίκηση να λάβει πληροφορίες και να τις αξιοποιήσει προκειμένου να κατανοήσει και να καλύψει τις ανάγκες των πολιτών λαμβάνοντας τις καλύτερες δυνατές αποφάσεις.

Προκειμένου τα αποτελέσματα από την εξόρυξη γνώμης να είναι επιτυχή θα πρέπει η Δημόσια Διοίκηση να παρακινήσει περισσότερους πολίτες να συμμετέχουν ενεργά εκφράζοντας τη γνώμη τους σε ηλεκτρονικά μέσα. Σημαντικό θα ήταν να κατασκευαστούν και να εκπαιδευτούν μοντέλα τα οποία θα μπορούν να εφαρμοστούν και να επαναχρησιμοποιηθούν σε μια πληθώρα θεμάτων που απασχολεί τη Δημόσια Διοίκηση.

### 4.2 Πλεονεκτήματα της Εξόρυξης Γνώμης από τα Ηλεκτρονικά Μέσα

Τα πλεονεκτήματα από την Εξόρυξη Γνώμης από τα ηλεκτρονικά μέσα για τη Δημόσια Διοίκηση είναι:

- Εξοικονόμηση δαπανών , χρόνου και ανθρώπινων πόρων που παλαιότερα απαιτούνταν για έρευνες και δημοσκοπήσεις.
- Οι πολίτες μπορούν να γράψουν ένα ελεύθερο κείμενο που θα εκφράζει τις απόψεις και τις προτάσεις τους σε αντίθεση με τις έρευνες και τις δημοσκοπήσεις που χρησιμοποιούσαν έτοιμες φόρμες με ερωτήσεις που δεχόταν συγκεκριμένες απαντήσεις.
- Η γνώμη των πολιτών μπορεί να εκφραστεί ελεύθερα και όχι υπό πίεση κάποιας έρευνας και συνεπώς τα αποτελέσματα αποκρίνονται περισσότερο στην πραγματικότητα.
- Στοιχεία από την εξόρυξη γνώμης μπορούν να συλλέγονται σε τακτά χρονικά διαστήματα.
- Αυξημένη συμμετοχή των πολιτών αφού πλέον σχεδόν όλοι έχουν πρόσβαση σε ηλεκτρονικά μέσα.



- Οι πολίτες αισθάνονται ότι η Δημόσια Διοίκηση είναι ανοιχτή σε προτάσεις και τους αφουγκράζεται ενισχύοντας έτσι την εμπιστοσύνη των πολιτών.
- Αναβάθμιση της Δημόσιας Διοίκησης με τη συστηματική εξαγωγή δεδομένων.

#### 4.3 Προκλήσεις για τη Δημόσια Διοίκηση κατά την Εξόρυξη Γνώμης – Ανάλυση Συναισθήματος

Η εξόρυξη γνώμης (opinion mining) συχνά αναφέρεται και ως ανάλυση συναισθήματος (sentimental analysis) [33]. Οι δύο αυτές έννοιες έχουν μικρές διαφορές και αυτό οφείλεται στο γεγονός ότι αρχικά μελετήθηκαν σε διαφορετικά πεδία. Η εξόρυξη γνώμης αφορά την άποψη που έχει σχηματίσει κάποιος για ένα θέμα ενώ η ανάλυση συναισθήματος αφορά την στάση που διατηρεί κάποιος βασισμένη στο συναίσθημα. Η διαφορά των δύο εννοιών είναι μικρή και η μια περιέχει στοιχεία της άλλης. Παρακάτω αναφέρονται οι σημαντικότερες από τις προκλήσεις που αντιμετωπίζουν οι δύο αυτές έννοιες.

- **Ανίχνευση spam** [16][17]. Αποτελεί μια χρονοβόρα και κοπιαστική εργασία για τους αναλυτές. Θα πρέπει τέτοιου είδους δεδομένα να εντοπιστούν και να αφαιρεθούν από το σύνολο των δεδομένων έτσι ώστε να μην επηρεάσουν το τελικό αποτέλεσμα.
- **Αξιοπιστία γνώμης**. Αποτελεί μια από τις μεγαλύτερες προκλήσεις για τους αναλυτές καθώς θα πρέπει να αξιολογήσουν την αξιοπιστία της γνώμης [18]. Πολλοί άνθρωποι υποκινούμενοι από άλλους είτε από το προσωπικό τους συμφέρον μπορεί να γράψουν μια μεροληπτική γνώμη.
- **Σχετικότητα γνώμης** [37]. Μερικές φορές οι πολίτες ανταλλάσσουν απόψεις ή αναφέρουν την γνώμη τους παρεκκλίνοντας από το επίμαχο θέμα. Αυτό μπορεί να είναι καταστροφικό γιατί μπορεί και άλλοι πολίτες να συμπαρασυρθούν και να μαζευτεί ένας όγκος απόψεων για θέμα που δεν είναι σχετικό. Επίσης είναι χρονοβόρο για τους αναλυτές που θα πρέπει να αφαιρέσουν όλα αυτά τα δεδομένα.
- **Γραμματικά λάθη** [35]. Αρκετές φορές η έλλειψη σωστής ορθογραφίας και οι αναγραμματισμοί δυσχεραίνουν το έργο των αναλυτών. Υπάρχουν και περιπτώσεις όπου η γνώμη του πολίτη είναι τόσο δυσανάγνωστη ώστε να μην μπορεί να κατανοηθεί και άρα να μην συμπεριληφθεί στα δεδομένα.
- **Έλλειψη εξειδίκευσης** [33]. Οι πολίτες αναμφισβήτητα έχουν το δικαίωμα να εκφέρουν τη γνώμη τους όμως υπάρχουν κάποια θέματα τα οποία απαιτούν κάποια εξειδίκευση π.χ. οικονομικά, λειτουργία και διαδικασίες της δημόσιας διοίκησης κ.α. Αυτό έχει ως αποτέλεσμα ο πολίτης να εκφέρει τη γνώμη του και να κάνει προτάσεις οι οποίες δεν μπορούν να υλοποιηθούν αγανακτώντας μάλιστα με ορισμένες αποφάσεις της δημόσιας διοίκησης. Έτσι καταγράφονται εσφαλμένα αρνητικές γνώμες και συγκεντρώνονται προτάσεις οι οποίες δεν βοηθούν τη δημόσια διοίκηση στο να λάβει αποφάσεις.
- **Σθένος γνώμης** [34]. Πολλές φορές οι πολίτες διατυπώνουν τη γνώμη τους με ουδέτερο τρόπο ίσως επειδή νομίζουν ότι θα έχουν συνέπειες ή επειδή φοβούνται την κριτική από άλλους πολίτες.

- **Ψεύτικα προφίλ.** Αρκετές φορές έχει παρατηρηθεί το φαινόμενο ψεύτικα προφίλ να αφήνουν την κριτική τους. Αυτό ενέχει δύο κινδύνους, είτε το ίδιο πρόσωπο αφήνει κριτικές με διαφορετικά προφίλ είτε να γράφονται κριτικές που δεν ανταποκρίνονται στην πραγματικότητα καθώς το άτομο υποκινείται από κάποιες σκοπιμότητες π.χ. πολιτικές, προσωπικές κ.α.
- **Χρήση καθομιλουμένης γλώσσας [36].** Η χρήση γλωσσικών μεταφορών, παροιμιών και ειρωνείας κάνει το έργο των αναλυτών δύσκολο καθώς η ανάλυση τέτοιων κριτικών δεν μπορεί να γίνει με αυτόματο τρόπο π.χ. text mining ή με χειροκίνητο τρόπο καθώς κάποιες κριτικές είναι δύσκολο να κατηγοριοποιηθούν.
- **Ανίχνευση ψεύτικων ειδήσεων.** Υπάρχουν περιπτώσεις κατά τις οποίες οι πολίτες γράφουν μια προσωπική τους ιστορία η οποία είναι δύσκολο να επιβεβαιωθεί ή διασπείρουν ψεύτικες ειδήσεις είτε επειδή τις πιστεύουν είτε επειδή θέλουν να να επηρεάσουν τις γνώμες άλλων πολιτών.
- **«Η επίπτωση των κοινωνικών διαδικασιών από τα δεδομένα και το πρόβλημα στην προστασία της ιδιωτικής ζωής κατά την ανάλυση» [19].** Συνήθως οι έρευνες χρησιμοποιούν δημόσια δεδομένα όμως οι καλύτερες πηγές δεδομένων είναι αυτές που είναι ιδιωτικές και προστατευμένες π.χ. ηλεκτρονικό ταχυδρομείο, άμεσα μηνύματα, ιδιωτικές συνομιλίες κλπ. Για αυτόν τον λόγο διεξάγονται προσωπικές έρευνες με σκοπό να αντληθούν τέτοια δεδομένα.
- **«Θέματα σθένους κατά την κατηγοριοποίηση συναισθημάτων» [20].** Έρευνες για την ανάλυση συναισθημάτων συνήθως επικεντρώνονται στα θετικά και αρνητικά συναισθήματα παραλείποντας πολλές φορές τα ουδέτερα. Σε αρκετές μελέτες που αφορούν στην εκπαίδευση μοντέλων ταξινόμησης τα ουδέτερα συναισθήματα αγνοούνται.
- **Αμφιβολία στο συναίσθημα που απεικονίζουν οι χρήστες.** Τα συναισθήματα των χρηστών που εκφράζουν τη γνώμη τους σε αρκετές περιπτώσεις είναι ασαφή και έμμεσα [21]. Αυτό δυσκολεύει την κατηγοριοποίηση καθώς αυτές οι γνώμες μπορεί να θεωρηθούν ότι δεν ανήκουν σε καμία κατηγορία.



# 5

## Ηλεκτρονική Διακυβέρνηση και Δημόσια Διαβούλευση

---

### 5.1 Εισαγωγή

Οι Τεχνολογίες Πληροφορικής και Επικοινωνιών (ΤΠΕ) εδώ και αρκετά χρόνια έχουν εισέλθει σε πολλούς τομείς της δημόσιας διοίκησης. Η χρήση ΤΠΕ έχει συμβάλει στον εκσυγχρονισμό, στην αύξηση της αποδοτικότητας, στη διαχείριση περίπλοκων εργασιών και στην εξωστρέφεια της δημόσιας διοίκησης. Μέσω της ηλεκτρονικής διακυβέρνησης δίνετε η δυνατότητα επικοινωνίας και συνεργασίας μεταξύ των πολιτών, των επιχειρήσεων και του δημόσιου τομέα. Οι πολίτες πλέον είναι ενημερωμένοι και έχουν άποψη για αρκετές εργασίες που εκτελούνται στον δημόσιο τομέα. Σε αρκετές περιπτώσεις έχουν προταθεί από πολίτες αποτελεσματικές λύσεις για πολύπλοκα ζητήματα που απασχολούσαν τη δημόσια διοίκηση. Η δημόσια διαβούλευση αποτελεί ένα βήμα για τους πολίτες που θέλουν να εκφράσουν τις απόψεις τους ελεύθερα και να συμμετέχουν στη λήψη αποφάσεων και στη χάραξη της πολιτικής ατζέντας.

### 5.2 Ηλεκτρονική Διακυβέρνηση

Η δημόσια διοίκηση με τη χρήση ΤΠΕ, την αξιοποίηση των πληροφοριακών συστημάτων και του διαδικτύου παρέχει υπηρεσίες στους πολίτες μειώνοντας την γραφειοκρατία, αυξάνοντας τη διαφάνεια και ενισχύοντας την εμπιστοσύνη των πολιτών. Στο άρθρο [47] καταγράφεται ο ορισμός της Ηλεκτρονικής Διακυβέρνησης ως «η on-line παροχή κρατικών υπηρεσιών στο κοινό μέσω του διαδικτύου ή μέσω της απαιτούμενης τεχνολογικής υποδομής». Η χρήση της ηλεκτρονικής διακυβέρνησης (e-government) έφερε μια καινούργια μορφή δημόσιας διοίκησης. Στόχος της ηλεκτρονικής διακυβέρνησης δεν είναι η απλή μετατροπή του χειρόγραφου συστήματος σε ηλεκτρονικό αλλά η αξιοποίηση των ΤΠΕ στη δημόσια διοίκηση με τρόπο τέτοιο ώστε να βελτιωθούν οι παρεχόμενες υπηρεσίες και να υπάρξει ένας ανοιχτός διάυλος επικοινωνίας μεταξύ πολίτη και κράτους αλλάζοντας το προφίλ της Κυβέρνησης η οποία μέχρι πριν λάμβανε αποφάσεις απομονωμένη. Οι υπηρεσίες της ηλεκτρονικής διακυβέρνησης μπορούν να προσαρμοστούν στις απαιτήσεις, προτεραιότητες και στόχους που θέτει κάθε φορά η δημόσια διοίκηση. Η ηλεκτρονική διακυβέρνηση άλλαξε σημαντικά τον τρόπο λειτουργίας της Δημόσιας Διοίκησης καθώς εισήγαγε νέες μεθόδους εργασίας και άλλαξε τη σχέση μεταξύ της συνεργασίας των υπηρεσιών και της εξυπηρέτησης του πολίτη. Οι αλλαγές είναι συνεχόμενες και δεν μπορεί να προβλεφθεί που θα οδηγήσουν τη Δημόσια Διοίκηση. Σύμφωνα με το άρθρο [47] η ηλεκτρονική διακυβέρνηση επηρέασε:

**Τις υπηρεσίες με επικέντρο τον πολίτη:** στις υπηρεσίες αυτές απαιτείται επανασχεδιασμός των εργασιών της Δημόσιας Διοίκησης έτσι ώστε να ικανοποιούν τις ανάγκες των πολιτών και να τους βοηθούν να ανταπεξέλθουν στις υποχρεώσεις

τους. Ο σχεδιασμός αυτών των υπηρεσιών ακολουθεί τον πελατοκεντρικό σχεδιασμό που χρησιμοποιείται στον ιδιωτικό τομέα από επιχειρήσεις. Οι υπηρεσίες είναι on-line, συνήθως στηρίζονται στην αυτοεξυπηρέτηση του πολίτη και είναι διαθέσιμες όλο το 24ωρο.

**Τον τρόπο διαχείρισης της πληροφορίας:** Οι πληροφορίες διαχειρίζονται σαν δημόσιος πόρος και έχουν θεσπιστεί νόμοι προστασίας τους. Με την ανάπτυξη της ηλεκτρονικής διακυβέρνησης όλες οι πληροφορίες θεωρούνται πλέον το ίδιο σημαντικές και όχι μόνο αυτές που αφορούν οικονομικά και περιουσιακά στοιχεία, είναι σημαντικός ο κύκλος ζωής των πληροφοριών και η επαναχρησιμοποίησή τους. Οι πληροφορίες επεξεργάζονται μεμονωμένα ή σε συνδυασμό με άλλες πληροφορίες προκειμένου να ληφθούν αποφάσεις. Ένα άλλο θέμα που αναδείχθηκε είναι αυτός της προστασίας των πληροφοριών και των προσωπικών δεδομένων.

**Τις δεξιότητες και σχέσεις:** Η ηλεκτρονική διακυβέρνηση προϋποθέτει ότι οι διαχειριστές αλλά και οι πολίτες είναι γνώστες των νέων τεχνολογιών. Επίσης άλλαξε τον τρόπο συναλλαγής μεταξύ των υπηρεσιών και των πολιτών αφού αντικατέστησε την αυτοπρόσωπη παρουσία του πολίτη και την επικοινωνία με το ηλεκτρονικό ταχυδρομείο ή το τηλέφωνο με on line υπηρεσίες.

**Τα μοντέλα λογοδοσίας και διαχείρισης:** Με την ανάπτυξη υπηρεσιών με επίκεντρο τον πολίτη τα στελέχη της Δημόσιας Διοίκησης κατανόησαν ότι στην πραγματικότητα έχουν να λογοδοτήσουν στους πολίτες και όχι στους πολιτικούς τους προϊσταμένους ή στα στελέχη της κυβέρνησης. Ενδυναμώθηκε ο ρόλος υπαλλήλων που βρίσκονται χαμηλά στην ιεραρχία ή/και στην πρώτη γραμμή εξυπηρέτησης του πολίτη και αυξήθηκε η αυτονομία τους.

### 5.3 Δημόσια Διαβούλευση

Η δημόσια διαβούλευση [38,39] είναι μια πρακτική που ακολουθεί η δημόσια διοίκηση κατά την οποία δημοσιοποιεί νομοθετικές ρυθμίσεις προκειμένου να αφουγκραστεί τους πολίτες και να λάβει καλύτερες αποφάσεις μελετώντας τις απόψεις τους. Κατά τη διαβούλευση το βήμα δίνεται στους πολίτες έτσι ώστε να εκθέσουν τις απόψεις τους, να κάνουν διάλογο μεταξύ τους, να διαπραγματευθούν και να βρουν μια κοινά αποδεκτή λύση. Με τη λήξη της διαβούλευσης συγκεντρώνονται όλες οι ανώνυμες ή επώνυμες απόψεις των πολιτών οι οποίες έχουν εκφραστεί πάνω σε δομημένα κείμενα ή σε ερωτήματα. Στη συνέχεια οι απόψεις μελετώνται και πραγματοποιούνται οι απαιτούμενες αλλαγές στα κείμενα των ρυθμίσεων. Για να συμμετέχουν όμως οι πολίτες θα πρέπει να είναι εξοικειωμένοι με τις νέες τεχνολογίες, να έχουν γνώσεις γύρω από το θέμα που τίθεται σε διαβούλευση και τα κείμενα της διαβούλευσης θα πρέπει να είναι σωστά δομημένα έτσι ώστε να μπορούν να γίνουν κατανοητά. Σκοπός της διαβούλευσης είναι να αξιολογηθούν οι λύσεις που έχουν προταθεί από τους πολίτες έτσι ώστε να ληφθούν ορθότερες αποφάσεις και να παρακινηθούν περισσότεροι πολίτες να συμμετέχουν ενεργά στα κοινά.

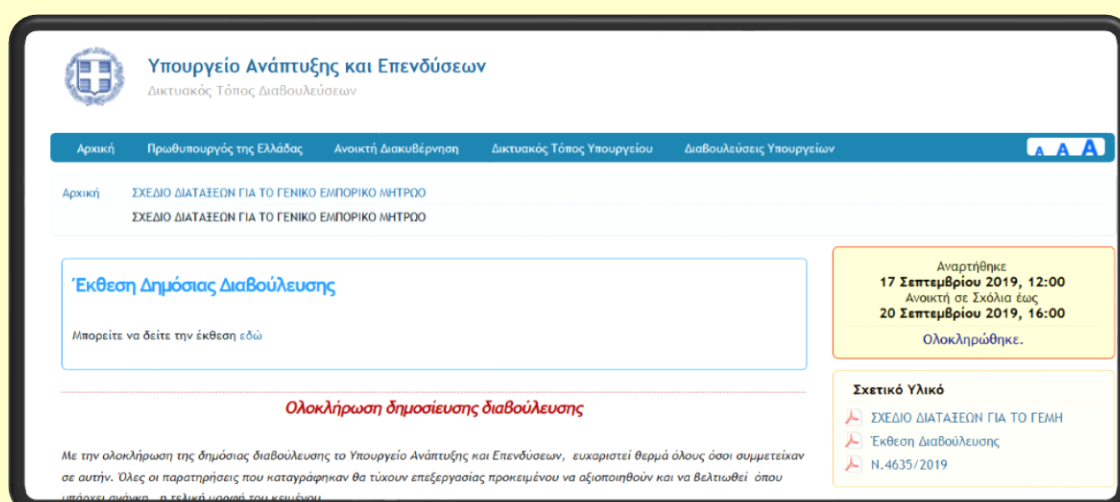
# 6

## Εξόρυξη δεδομένων με το λογισμικό WEKA – Πειραματικό μέρος

### 6.1 Εισαγωγή στο πειραματικό μέρος

Για τις ανάγκες της παρούσας εργασίας μελετήθηκαν τα σχόλια που εμφανίζονται στον Διαδικτυακό Τόπο Διαβουλεύσεων του Υπουργείου Ψηφιακής Διακυβέρνησης στη διεύθυνση <http://www.opengov.gr/γροια/?p=10654>. Τα σχόλια αντλήθηκαν από τη Δημόσια Διαβούλευση για τις διατάξεις Νόμου που αφορούν στο Γενικό Εμπορικό Μητρώο. Στις διατάξεις περιγράφεται η βελτίωση του Πληροφοριακού Συστήματος του Γενικού Εμπορικού Μητρώου (ΓΕ.Μ.Η.). Συνολικά μελετήθηκαν 153 σχόλια.

Για την εξόρυξη δεδομένων έχουν αναπτυχθεί πολλά και διαφορετικά εργαλεία [28]. Επιλέχθηκε για τις ανάγκες της παρούσας εργασίας να χρησιμοποιηθεί το εργαλείο εξόρυξης δεδομένων WEKA [24] για την ανάλυση των δεδομένων καθώς είναι ένα εύχρηστο εργαλείο ανοιχτού κώδικα φιλικό προς τον χρήστη. Τα δεδομένα επεξεργάστηκαν με τους αλγόριθμους C4.5/J4.8, Apriori, K-Means και Naive Bayes. Ο αλγόριθμος C4.5/J4.8 [22] χρησιμοποιήθηκε για την ταξινόμηση των νέων δεδομένων σε κατηγορίες, για να αποκαλυφθεί πως το συναίσθημα των σχολιαστών επηρεάζει άλλα χαρακτηριστικά και για να παραχθεί ένα οπτικοποιημένο δέντρο αποφάσεων που θα είναι κατανοητό ακόμη και από μη τεχνικούς. Για να δοκιμαστεί ο αλγόριθμος C4.5/J4.8 εφαρμόστηκε στα ίδια δεδομένα ο αλγόριθμος Naive Bayes [23] έτσι ώστε να συγκριθούν τα αποτελέσματα δεδομένων με γνωστές τιμές-στόχους με αποτελέσματα δεδομένων που έχουν προγνωστικές τιμές. Ο αλγόριθμος Apriori [25] εφαρμόστηκε προκειμένου να αποκαλυφθούν πολύτιμοι κανόνες συσχέτισης [26] και για να αναδειχθούν και να γίνουν κατανοητές οι σχέσεις μεταξύ των δεδομένων. Ο αλγόριθμος K-means [27] χρησιμοποιήθηκε για να ανακαλυφθούν ομάδες μέσα στο σύνολο των δεδομένων.



Εικόνα 1 : Σελίδα Διαδικτυακού Τόπου Διαβουλεύσεων

Πηγή Εικόνας : Υπουργείο Ψηφιακής Διακυβέρνησης

Πλοήγηση στη Διαβούλευση	
9 Σχόλια	ΚΕΦΑΛΑΙΟ 1ο - ΣΚΟΠΟΣ ΚΑΙ ΠΛΑΙΣΙΟ ΛΕΙΤΟΥΡΓΙΑΣ ΤΟΥ Γ.Ε.ΜΗ. - Άρθρο 1 - Σκοπός του Γενικού Εμπορικού Μητρώου - ορισμοί
11 Σχόλια	Άρθρο 02 - Εγγραφόμενοι στο Γ.Ε.ΜΗ.
25 Σχόλια	Άρθρο 3 - Διάρθρωση Γ.Ε.ΜΗ. & Υπηρεσίες Γ.Ε.ΜΗ.
6 Σχόλια	Άρθρο 4 - Εποπτεία Γ.Ε.ΜΗ.
5 Σχόλια	Άρθρο 5 - Πληροφοριακό Σύστημα του Γ.Ε.ΜΗ.
1 Σχόλιο	Άρθρο 6 Εγγραφή στο Γ.Ε.ΜΗ.- Λήψη Αριθμού Γ.Ε.ΜΗ.
1 Σχόλιο	Άρθρο 7 - Διαλειτουργικότητα Γ.Ε.ΜΗ.
1 Σχόλιο	ΚΕΦΑΛΑΙΟ 2ο - ΕΜΠΟΡΙΚΗ ΔΗΜΟΣΙΟΤΗΤΑ - Άρθρο 8 - Δημοσιότητα του Γενικού Εμπορικού Μητρώου
1 Σχόλιο	Άρθρο 09 - Αποτελέσματα της συστατικής δημοσιότητας

### Εικόνα 2 : Πλοήγηση στη Διαβούλευση

Πηγή Εικόνας : Υπουργείο Ψηφιακής Διακυβέρνησης

## 6.2 Μεθοδολογία συλλογής και επεξεργασίας δεδομένων

Τα δεδομένα συλλέχθηκαν από τον Διαδικτυακό Τόπο Διαβουλεύσεων του Υπουργείου Ψηφιακής Διακυβέρνησης. Στη συνέχεια επεξεργάστηκαν και κατηγοριοποιήθηκαν χειροκίνητα ως προς τα άρθρα, το είδος του σχολιαστή, το φύλο του σχολιαστή, το είδος σχολιασμού, το συναίσθημα και αν ο σχολιαστής προτείνει κάποια βελτίωση στις διατάξεις του Νόμου.

Τα σχόλια διαχωρίστηκαν με βάση τα παρακάτω χαρακτηριστικά:

Χαρακτηριστικό	Όνομα Μεταβλητής	Τιμές Μεταβλητής
Άρθρο Νόμου	Article	{1,2,3,4,5,6,7,8,9,11,12,13,14,15,17,18,19,20, 22, 23, 24, 25, 26, 27, 29, 30, 32}
Σχολιαστής	Commentator	{individual, PublicService}
Φύλο σχολιαστή	CommentatorGender	{Male, Female, NoGender}
Είδος σχολιασμού	Comment	{Positive, Neutral, Negative, Question}
Συναίσθημα σχολιαστή	Emotion	{Hopeful, Excited, Pleading, Angry, Irritated, NoEmotion, Worried}
Αν κάνει πρόταση ο σχολιαστής για τον Νόμο	Recommentation	{Yes, No}

Πίνακας 1: Χαρακτηριστικά σχολίων

Τιμές μεταβλητών	Επεξήγηση τιμών μεταβλητών
Article	Οι αριθμοί των άρθρων στα οποία υπήρχαν σχόλια
individual	Μεμονωμένο πρόσωπο
PublicService	Δημόσια Υπηρεσία
NoGender	Χωρίς φύλο (αφορά τις Δημόσιες Υπηρεσίες)
Positive	Θετικό σχόλιο
Neutral	Ουδέτερο σχόλιο
Negative	Αρνητικό σχόλιο
Question	Ερώτημα για τον Νόμο
Hopeful	Το συναίσθημα του σχολιαστή έδειχνε πως είναι αισιόδοξος
Excited	Το συναίσθημα του σχολιαστή έδειχνε πως είναι ενθουσιασμένος
Pleading	Το συναίσθημα του σχολιαστή δεν ήταν έκδηλο έδειχνε όμως να ζητά κάτι να προστεθεί ή να αλλάξει στο Νόμο
Angry	Το συναίσθημα του σχολιαστή έδειχνε πως είναι θυμωμένος.
Irritated	Το συναίσθημα του σχολιαστή έδειχνε πως είναι εκνευρισμένος.
NoEmotion	Ο σχολιαστής δεν έδειξε κανένα συναίσθημα
Worried	Το συναίσθημα του σχολιαστή έδειχνε πως είναι ανήσυχος.
Yes, No	Ο σχολιαστής προτείνει ή δεν προτείνει κάποια αλλαγή στο Νόμο

Πίνακας 2: Επεξήγηση τιμών μεταβλητών

### 6.3 Το λογισμικό WEKA

Το λογισμικό Weka [24] είναι μια σουίτα λογισμικού που αναπτύχθηκε στο Πανεπιστήμιο Waikato της Νέας Ζηλανδίας. Είναι ιδανικό για ανάλυση δεδομένων και προγνωστική μοντελοποίηση. Υποστηρίζει την μηχανική μάθηση με κατάλληλους αλγόριθμους και εργαλεία. Διαθέτει GUI το οποίο διευκολύνει την πρόσβαση σε όλα τα στοιχεία. Είναι ελεύθερο λογισμικό το οποίο επιτρέπει στους χρήστες να το τροποποιούν και πήρε το όνομα του από ένα υπό εξαφάνιση πουλί της Νέας Ζηλανδίας. Αποτελεί ένα πολύ διαδεδομένο λογισμικό για εξόρυξη δεδομένων. Το WEKA είναι αρκετά δημοφιλές καθώς προσφέρει αρκετές δυνατότητες: διαθέτει μεγάλη ποικιλία μεθόδων που αφορούν στην κατηγοριοποίηση, στην ανάλυση συστάδων, στην παλινδρόμηση και στους κανόνες συσχέτισης, μπορεί να προεπεξεργαστεί δεδομένα, διαθέτει εργαλεία οπτικοποίησης, έχει γραφεί σε γλώσσα προγραμματισμού JAVA και άρα μπορεί να εγκατασταθεί σε πολλές πλατφόρμες, επειδή είναι λογισμικό ανοιχτού κώδικα ο κώδικας είναι διαθέσιμος δημόσια και μπορούν οι χρήστες να τροποποιήσουν και να εξελίξουν τους

αλγορίθμους, το περιβάλλον εργασίας του είναι φιλικό και υπάρχουν διαθέσιμες πολλές βιβλιοθήκες στο διαδίκτυο [29].



**Εικόνα 3 : Λογότυπο Weka**  
Πηγή Εικόνας : Λογισμικό WEKA

#### **6.4 Η δομή του αρχείου δεδομένων**

Το λογισμικό WEKA λαμβάνει ως είσοδο αρχεία δεδομένων με τύπο Attribute Relation File Format (ARFF). Τα αρχεία αυτά περιλαμβάνουν instances (στιγμιότυπα) και attributes (χαρακτηριστικά). Τα αρχεία ARFF αποτελούνται από δύο μέρη, την περιοχή της επικεφαλίδας και την περιοχή των δεδομένων. Στην περιοχή της επικεφαλίδας περιγράφονται τα χαρακτηριστικά όπως μεταβλητές ή ιδιότητες. Συγκεκριμένα περιέχει την έκφραση @relation όπου δηλώνει το όνομα του αρχείου και έπειτα ακολουθούν όλα τα χαρακτηριστικά των παραδειγμάτων του συνόλου ως εξής: @attribute attributename datatype όπου στο attributename δηλώνεται το όνομα του χαρακτηριστικού και στο datatype ο τύπος του. Το WEKA μπορεί να ορίσει τέσσερις τύπους δεδομένων: numerics (αριθμητικούς), nominal (ονομαστικά δεδομένα), date (ημερομηνίας με συγκεκριμένη μορφή) και string (αλφαριθμητικά). Στην περιοχή δεδομένων εμφανίζονται δεδομένα των οποίων τα χαρακτηριστικά είναι ταξινομημένα σε προκαθορισμένη σειρά σύμφωνα με τα όσα έχουν καθοριστεί στην επικεφαλίδα και διαχωρίζονται με κόμμα. Την έναρξη των δεδομένων σηματοδοτεί το @ data.



@RELATION Comments

@ATTRIBUTE article {1,2,3,4,5,6,7,8,9,11,12,13,14,15,17,18,19,20, 22, 23, 24, 25, 26, 27, 29, 30, 32}

@ATTRIBUTE commentator {individual, public service}

@ATTRIBUTE commentator gender {Male, Female, No Gender}

@ATTRIBUTE comment {Positive, Neutral, Negative, Question}

@ATTRIBUTE emotion {Hopeful, Excited, Pleading, Angry, Irritated, No Emotion, Worried}

@ATTRIBUTE recommentation {Yes, No}

@DATA

12, individual, Female, Neutral, Worried, Yes

3, individual, Male, Negative, Irritated, Yes

29, individual, Male, Negative, Angry, No

30, individual, Male, Negative, Irritated, Yes

8, individual, Male, Neutral, No Emotion, Yes

2, individual, Male, Neutral, No Emotion, Yes

19, individual, Male, Neutral, No Emotion, No

#### Εικόνα 4 : Απόσπασμα αρχείου arff

Για την παρούσα εργασία δεν έγινε μετασχηματισμός ούτε καθαρισμός των δεδομένων καθώς δεν υπήρχαν κενές τιμές και είχε ήδη διενεργηθεί από το Υπουργείο φιλτράρισμα στα σχόλια που δεν έχουν σχέση με τη δημόσια διαβούλευση. Σκοπός είναι να παρουσιαστεί η πλήρης εικόνα της διαβούλευσης έτσι όπως εμφανίζεται στους χρήστες και όπως αποτυπώθηκε μετά τη λήξη της στην έκθεση επί αυτής.

Τα περιεχόμενα του αρχείου arff όπου έχουν συγκεντρωθεί όλα τα δεδομένα εμφανίζονται στο παράρτημα της εργασίας στο σημείο i.

## 6.5 Έναρξη του λογισμικού WEKA

Κατά την έναρξη του λογισμικού εμφανίζεται η οθόνη:



Εικόνα 5 : Έναρξη WEKA Version 3.8.4

Πηγή Εικόνας : Λογισμικό WEKA

Ο χρήστης σε αυτό το σημείο έχει την δυνατότητα να επιλέξει μία από τις παρακάτω κύριες εφαρμογές:

**Explorer** : Είναι η δημοφιλέστερη διεπαφή του λογισμικού. Από εδώ ο χρήστης μπορεί να πραγματοποιήσει τις κυριότερες εργασίες της εξόρυξης δεδομένων όπως παλινδρόμηση, προεπεξεργασία δεδομένων, οπτικοποίηση, ανακάλυψη κανόνων συσχέτισης, κατηγοριοποίηση και ανάλυση συστάδων.

**Experimenter**: Αποτελεί ένα περιβάλλον στο οποίο μπορούν να διεξαχθούν πειράματα και να αξιολογηθούν οι μέθοδοι της παλινδρόμησης και της κατηγοριοποίησης παρέχοντας μάλιστα και σύγκριση επίδοσης των διαφορετικών μοντέλων. Επίσης μπορεί να παρουσιάσει αποτελέσματα σε πίνακες.

**KnowledgeFlow** : Περιβάλλον όπου μπορούν να διεξαχθούν οι ίδιες εργασίες με τον Explorer αλλά διαθέτει διαφορετικό interface καθώς χρησιμοποιούνται components που συνδέονται μεταξύ τους με γραφικό τρόπο.



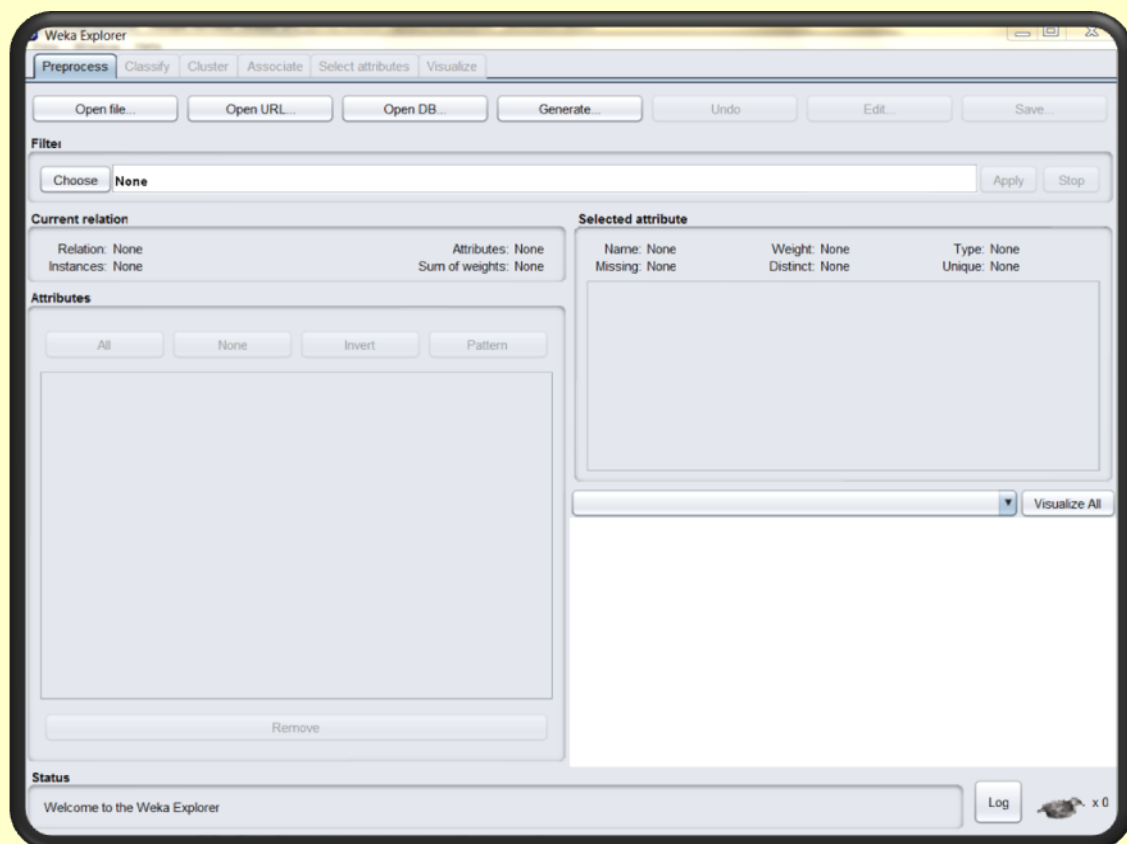
**Workbench:** Αποτελεί μια διεπαφή η οποία ενώνει τις τρεις προηγούμενες εφαρμογές. Μπορεί να διαμορφωθεί από τον χρήστη και του επιτρέπει να διαλέξει ποιες εφαρμογές και ποιες ρυθμίσεις θα εφαρμοστούν.

**Simple CLI:** Παρέχει την δυνατότητα στον χρήστη να εφαρμόσει όλα τα εργαλεία ανακάλυψης γνώσης σε JAVA μέσω μιας διεπαφής με command line.

## 6.6 Επεξεργασία δεδομένων με το λογισμικό WEKA

### 6.6.1 Άνοιγμα αρχείου

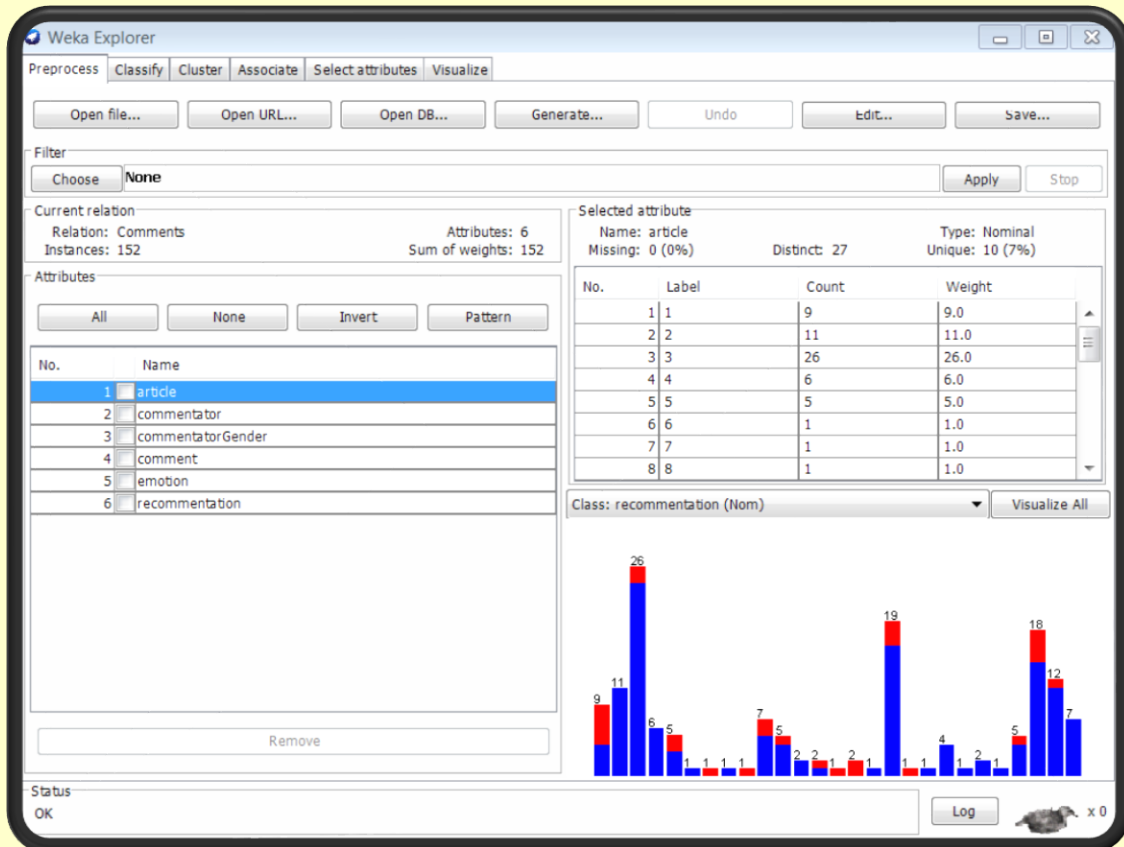
Στην παρούσα εργασία έχει επιλεγεί η διεπαφή Explorer :



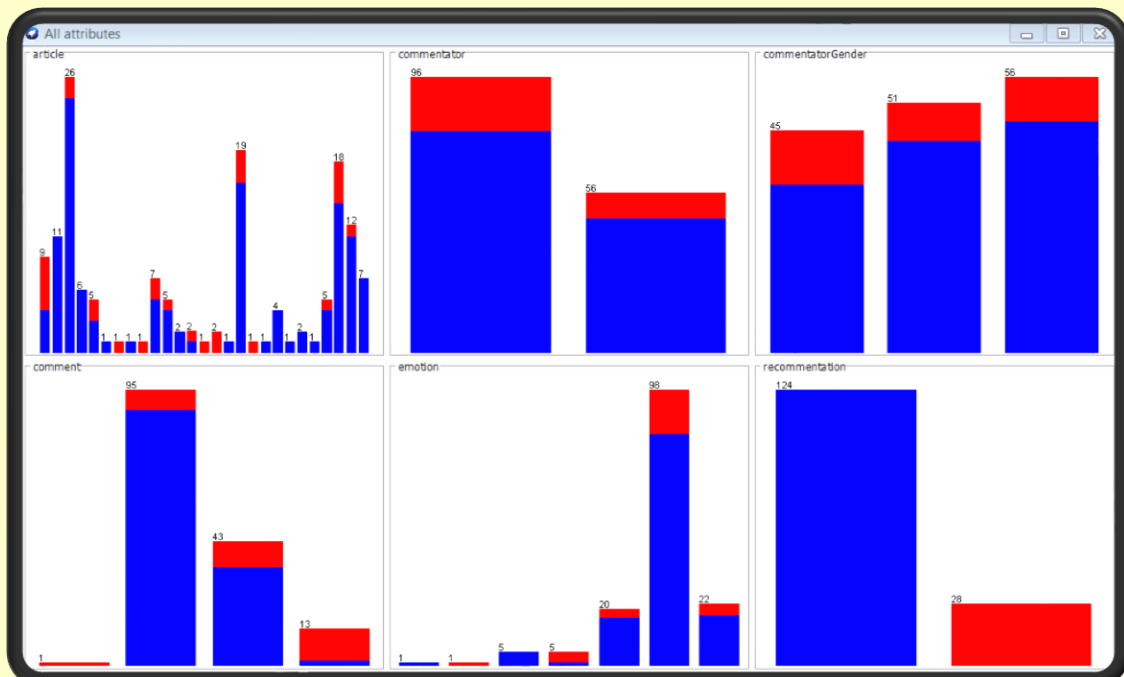
**Εικόνα 6 : WEKA Explorer**  
Πηγή Εικόνας : Λογισμικό WEKA

Έπειτα επιλέχθηκε να ανοίξει το αρχείο arff με τίτλο comments i.

Η καρτέλα Preprocess εκτελεί την φόρτωση του αρχείου arff και εμφανίζει πληροφορίες για το πλήθος των δεδομένων, στατιστικά στοιχεία, πληροφορίες για τα χαρακτηριστικά και εκτελεί μια οπτική απεικόνιση τους σε μορφή ιστογράμματος.



Εικόνα 7 : Άνοιγμα αρχείου comments i  
 Πηγή Εικόνας : Λογισμικό WEKA



Εικόνα 8 : Ιστογράμματα όλων των χαρακτηριστικών  
 Πηγή Εικόνας : Λογισμικό WEKA

### 6.6.2 Εφαρμογή αλγορίθμου J4.8

Ο αλγόριθμος J4.8 αποτελεί μια βελτιωμένη έκδοση του αλγορίθμου C4.5 [22]. Παράγει ένα δέντρο απόφασης και εφαρμόζεται από πάνω προς τα κάτω εκτελώντας επαναληπτικά την μέθοδο διαίρει και βασίλευε. Είναι ένας άπληστος αλγόριθμος ο οποίος ψάχνει για βέλτιστες τοπικά επιλογές έτσι ώστε να ανακαλύψει το ολικό βέλτιστο. Ο αλγόριθμος χρησιμοποιεί το κάθε χαρακτηριστικό των δεδομένων το οποίο είναι ικανό για να ληφθεί μια απόφαση και χωρίζει τα δεδομένα σε υποσύνολα εξετάζοντας το κέρδος των πληροφοριών. Αν κάποιο χαρακτηριστικό έχει μεγαλύτερο ομαλοποιημένο κέρδος τότε το χρησιμοποιεί για να λάβει απόφαση. Η λειτουργία του αλγορίθμου είναι η εξής: Για κάθε χαρακτηριστικό ο αλγόριθμος βρίσκει το ομαλοποιημένο κέρδος των πληροφοριών από τον διαχωρισμό του χαρακτηριστικού. Αν το χαρακτηριστικό αυτό έχει το μεγαλύτερο ομαλοποιημένο κέρδος τότε δημιουργεί έναν κόμβο απόφασης. Έπειτα πηγαίνει στις υπολίστες αυτού του χαρακτηριστικού και προσθέτει τα παιδιά του.

Φορτώθηκε το σύνολο των δεδομένων και επιλέχθηκε ως κλάση η μεταβλητή comment με χρήση της τεχνικής fold cross validation έτσι ώστε να διασπαστεί το σύνολο των δεδομένων σε ίσα σύνολα και να δημιουργηθεί ένα δέντρο απόφασης για αυτή την μεταβλητή. Τα στατιστικά στοιχεία του δέντρου μετά την εκτέλεση του αλγορίθμου χωρίς εφαρμογή κάποιου φίλτρου εμφανίζονται στην παρακάτω εικόνα.

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    Comments
Instances:   152
Attributes:  6
             article
             commentator
             commentatorGender
             comment
             emotion
             recommendation
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

emotion = Hopeful: Neutral (1.0)
emotion = Excited: Positive (1.0)
emotion = Pleading: Neutral (5.0)
emotion = Angry: Negative (5.0)
emotion = Irritated: Negative (20.0)
emotion = NoEmotion
| recommendation = Yes: Neutral (82.0/7.0)
| recommendation = No: Question (16.0/5.0)
emotion = Worried: Negative (22.0/10.0)

Number of Leaves :    8
Size of the tree :    10

Time taken to build model: 0.12 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      129           84.8684 %
Incorrectly Classified Instances    23           15.1316 %
Kappa statistic                    0.7185
Mean absolute error                 0.1116
Root mean squared error             0.2415
Relative absolute error             42.1924 %
Root relative squared error        66.7828 %
Total Number of Instances          152

=== Detailed Accuracy By Class ===

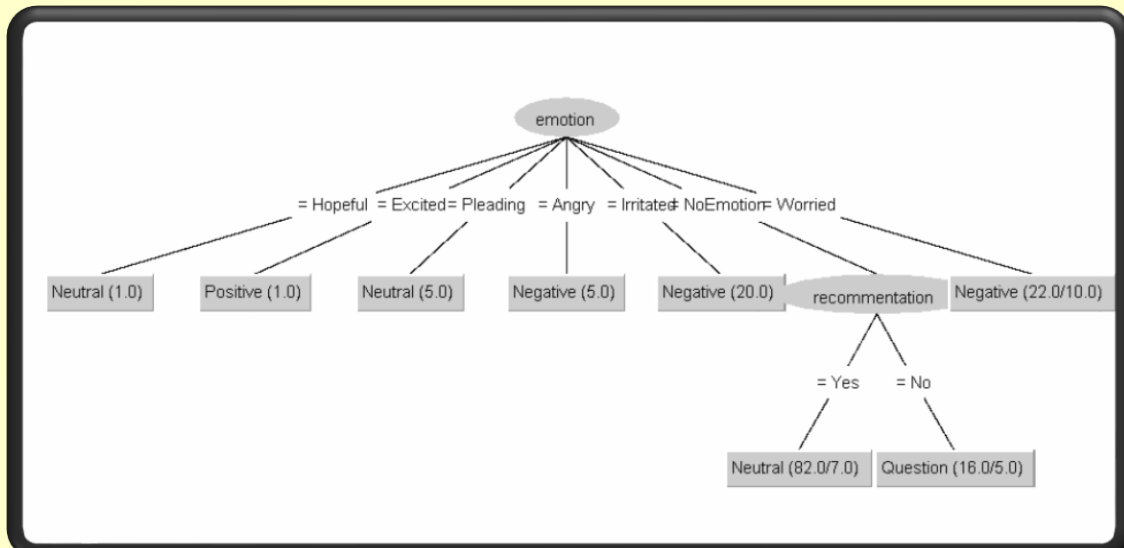
              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
              0,000    0,000    ?          0,000    ?          ?        0,497    0,007    Positive
              0,853    0,140    0,910     0,853    0,880     0,700    0,860    0,865    Neutral
              0,860    0,092    0,787     0,860    0,822     0,749    0,886    0,851    Negative
              0,846    0,036    0,688     0,846    0,759     0,738    0,841    0,491    Question
Weighted Avg.   0,849    0,117    ?          0,849    ?          ?        0,863    0,823

=== Confusion Matrix ===

 a  b  c  d  <-- classified as
 0  1  0  0  | a = Positive
 0 81  9  5  | b = Neutral
 0  6 37  0  | c = Negative
 0  1  1 11  | d = Question

```

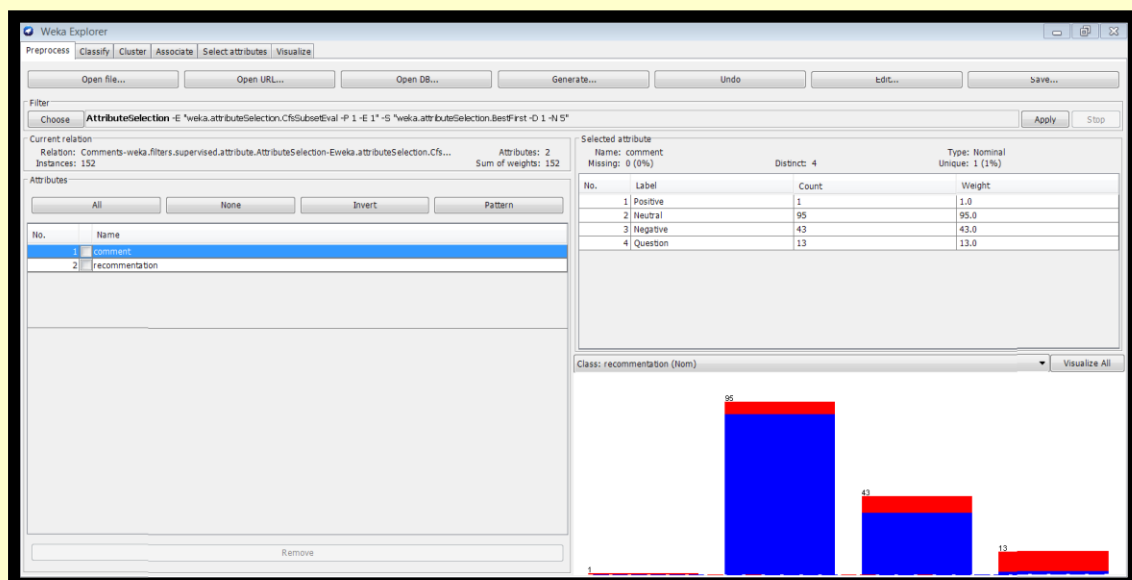
**Εικόνα 9 : Εφαρμογή αλγορίθμου J4.8 για την κλάση comment (χωρίς φίλτρο)**  
 Πηγή Εικόνας : Λογισμικό WEKA



**Εικόνα 10 : Οπτικοποίηση δέντρου αλγορίθμου J4.8 για την κλάση comment (χωρίς φίλτρο)**

Πηγή Εικόνας : Λογισμικό WEKA

Στη συνέχεια στην καρτέλα preprocess έγινε επιλογή του φίλτρου attribute selection. Με αυτό το φίλτρο περιορίστηκε το σύνολο των δεδομένων και διατηρήθηκαν μόνο οι σχετικές μεταβλητές και το μοντέλο απλουσεύθηκε.



**Εικόνα 11 : Εφαρμογή φίλτρου attribute selection**

Πηγή Εικόνας : Λογισμικό WEKA

Εφαρμόστηκε εκ νέου ο αλγόριθμος J4.8 για το νέο τελικό σύνολο δεδομένων με επιλεγμένη την κλάση comments και λάβαμε τα παρακάτω αποτελέσματα:

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    Comments-weka.filters.supervised.attribute.AttributeSelection-Eweka.attributeSelection.CfsSubsetEval
Instances:   152
Attributes:  2
             comment
             recommendation
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

recommendation = Yes: Neutral (124.0/36.0)
recommendation = No: Question (28.0/17.0)

Number of Leaves :    2

Size of the tree :    3

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      97           63.8158 %
Incorrectly Classified Instances    55           36.1842 %
Kappa statistic                    0.2309
Mean absolute error                 0.2362
Root mean squared error             0.3455
Relative absolute error             89.312 %
Root relative squared error         95.5605 %
Total Number of Instances          152

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
              0,000    0,000    ?           0,000    ?           ?         0,421    0,007    Positive
              0,926    0,632    0,710       0,926    0,804       0,368    0,576    0,656    Neutral
              0,000    0,028    0,000       0,000    0,000       -0,089    0,436    0,250    Negative
              0,692    0,115    0,360       0,692    0,474       0,435    0,789    0,293    Question
Weighted Avg.   0,638    0,412    ?           0,638    ?           ?         0,554    0,506

=== Confusion Matrix ===

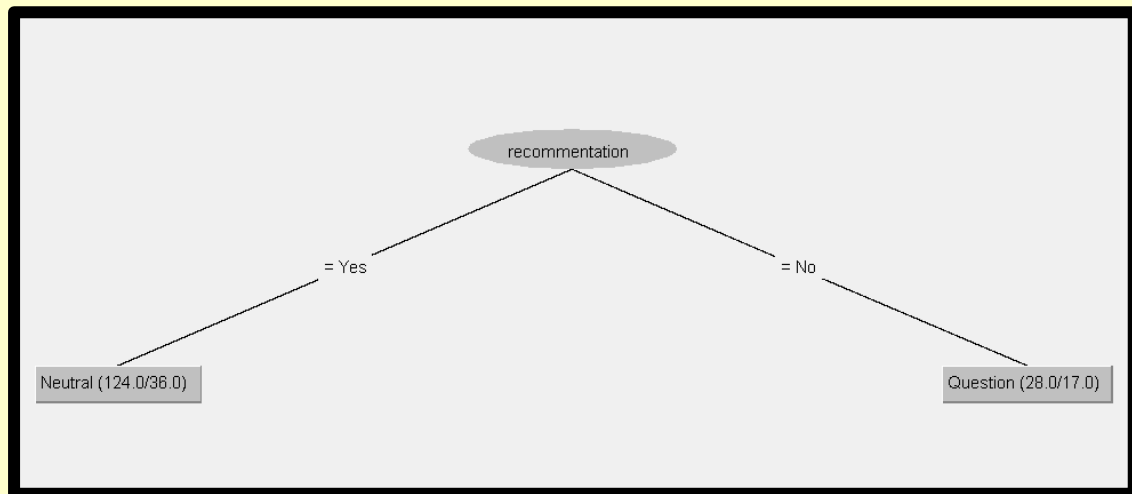
 a  b  c  d  <-- classified as
0  0  0  1 | a = Positive
0 88  1  6 | b = Neutral
0 34  0  9 | c = Negative
0  2  2  9 | d = Question

```

**Εικόνα 12 : Εφαρμογή αλγορίθμου J4.8 (φίλτρο attribute selection)**  
 Πηγή Εικόνας : Λογισμικό WEKA

Παρατηρήθηκε ότι μετά την εφαρμογή του φίλτρου attribute selection το δέντρο που παρήγαγε ο J4.8 ήταν μικρότερο, ο χρόνος δημιουργίας του μικρότερος, το ποσοστό των correctly classified instances μειώθηκε σημαντικά και τα ποσοστά των relative absolute error και των root relative squared error αυξήθηκαν σημαντικά.

το αντίστοιχο δέντρο που δημιουργήθηκε είναι το παρακάτω:



**Εικόνα 13 : Οπτικοποίηση δέντρου αλγορίθμου J4.8 (φίλτρο attribute selection)**

Πηγή Εικόνας : Λογισμικό WEKA

Φορτώθηκε εκ νέου το σύνολο των δεδομένων και επιλέχθηκε ως κλάση η μεταβλητή emotion με χρήση της τεχνικής fold cross validation έτσι ώστε να διασπαστεί το σύνολο των δεδομένων σε ίσα σύνολα και να δημιουργηθεί ένα δέντρο απόφασης για αυτή την μεταβλητή. Τα στατιστικά στοιχεία του δέντρου μετά την εκτέλεση του αλγορίθμου χωρίς εφαρμογή κάποιου φίλτρου εμφανίζονται στην παρακάτω εικόνα. Παρατηρήθηκε ότι 112 χαρακτηριστικά ταξινομήθηκαν σωστά και 40 λάθος.

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    Comments
Instances:   152
Attributes:  6
             article
             commentator
             commentatorGender
             comment
             emotion
             recommendation
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----
comment = Positive: Excited (1.0)
comment = Neutral: NoEmotion (95.0/15.0)
comment = Negative: Irritated (43.0/23.0)
comment = Question: NoEmotion (13.0/1.0)

Number of Leaves :    4

Size of the tree :    5

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      112           73.6842 %
Incorrectly Classified Instances     40           26.3158 %
Kappa statistic                     0.4741
Mean absolute error                  0.1106
Root mean squared error              0.2384
Relative absolute error              69.1805 %
Root relative squared error          85.3977 %
Total Number of Instances           152

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0,000  0,000  ?         0,000  ?         ?         0,215    0,007    Hopeful
0,000  0,000  ?         0,000  ?         ?         0,500    0,007    Excited
0,000  0,000  ?         0,000  ?         ?         0,528    0,041    Pleading
0,000  0,000  ?         0,000  ?         ?         0,799    0,091    Angry
1,000  0,174  0,465    1,000  0,635    0,620    0,884    0,394    Irritated
0,939  0,315  0,844    0,939  0,889    0,663    0,769    0,813    NoEmotion
0,000  0,000  ?         0,000  ?         ?         0,546    0,187    Worried
Weighted Avg.   0,737  0,226  ?         0,737  ?         ?         0,739    0,607

=== Confusion Matrix ===

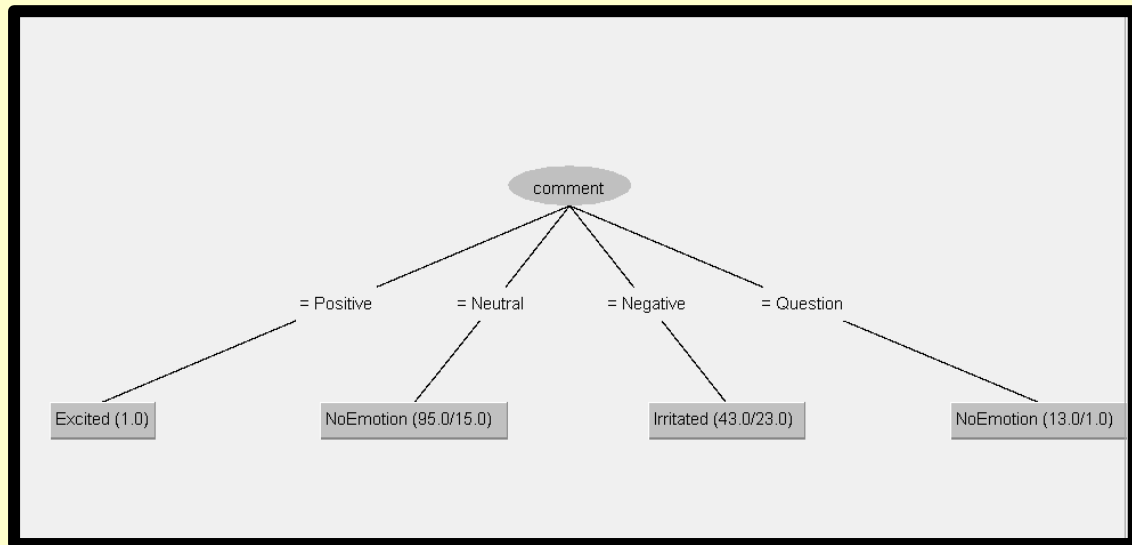
 a  b  c  d  e  f  g  <-- classified as
0  0  0  0  0  1  0 | a = Hopeful
0  0  0  0  0  1  0 | b = Excited
0  0  0  0  0  5  0 | c = Pleading
0  0  0  0  5  0  0 | d = Angry
0  0  0  0  20  0  0 | e = Irritated
0  0  0  0  6  92  0 | f = NoEmotion
0  0  0  0  12  10  0 | g = Worried

```

**Εικόνα 14 : Εφαρμογή αλγορίθμου J4.8 για την κλάση emotion (χωρίς φίλτρο)**

Πηγή Εικόνας : Λογισμικό WEKA





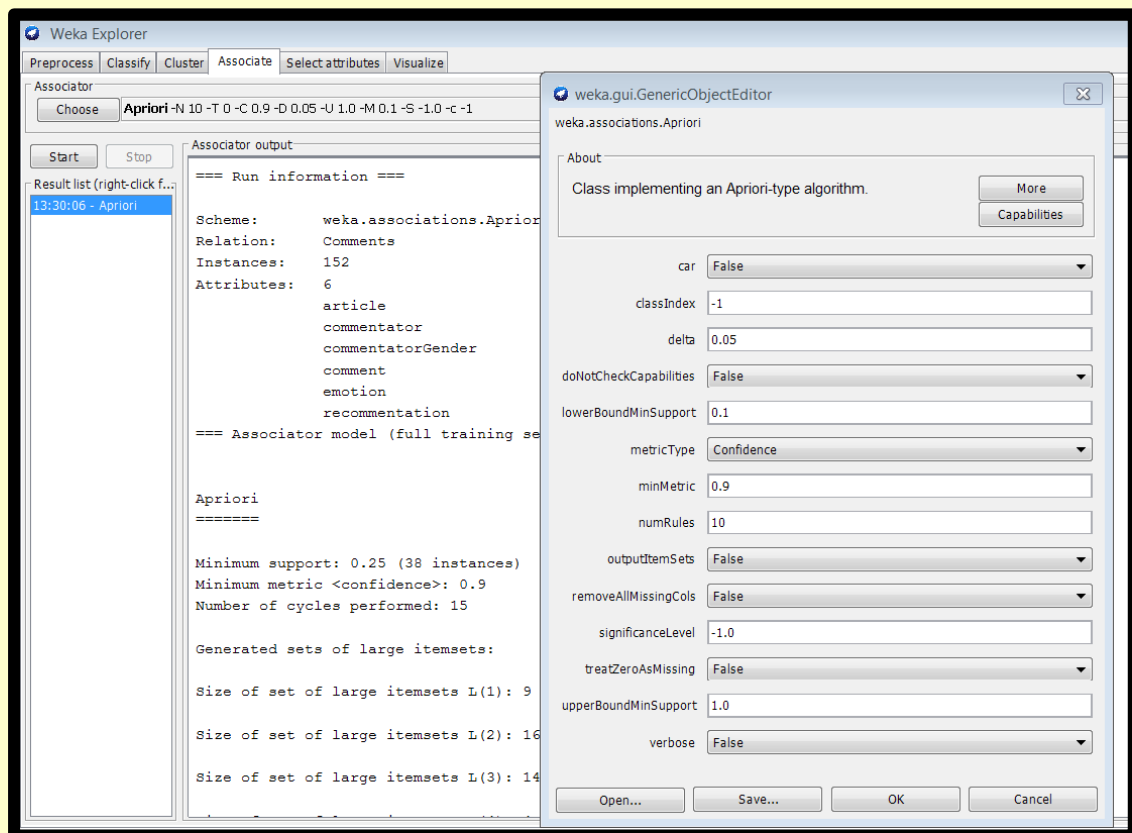
**Εικόνα 15 : Οπτικοποίηση δέντρου αλγορίθμου J4.8 για την κλάση emotion (χωρίς φίλτρο)**

Πηγή Εικόνας : Λογισμικό WEKA

### 6.6.3 Εφαρμογή αλγορίθμου Apriori

Στην εξόρυξη δεδομένων οι κανόνες συσχέτισης [26] είναι μια από τις σημαντικότερες τεχνικές καθώς οι πληροφορίες που παράγονται με την ανάλυση των συσχετίσεων των δεδομένων είναι πολύ χρήσιμες για αρκετούς τομείς. Για να εξαγάγουμε κάποιον κανόνα συσχέτισης θα πρέπει να οριστούν κατώτατα όρια για την υποστήριξη και την εμπιστοσύνη. Για να εξαχθεί ένας κανόνας συσχέτισης ο οποίος έχει ενδιαφέρον θα πρέπει να έχει υποστήριξη μεγαλύτερη από το όριο ελάχιστης υποστήριξης και εμπιστοσύνη μεγαλύτερη από την ελάχιστη εμπιστοσύνη. Οι κανόνες συσχέτισης αναπτύσσονται σε δύο στάδια. Στο πρώτο στάδιο βρίσκονται όλα τα συχνά στοιχειοσύνολα και στο δεύτερο εξαγονται οι κανόνες συσχέτισης που έχουν όριο μεγαλύτερο από την ελάχιστη εμπιστοσύνη. Για το πρώτο στάδιο έχουν αναπτυχθεί αρκετοί αλγόριθμοι. Ο πιο σημαντικός από αυτούς είναι ο αλγόριθμος Apriori. Ο αλγόριθμος Apriori αρχικά διαπερνά όλο τον πίνακα δεδομένων και υπολογίζει ποια δεδομένα ικανοποιούν την ελάχιστη υποστήριξη. Έπειτα σε κάθε πέρασμα υπολογίζονται νέα item sets με βάση τα προηγούμενα. Τα item sets που προκύπτουν είναι υποψήφια item sets επειδή δεν είναι γνωστή η υποστήριξη και συνεπώς δεν γνωρίζουμε αν είναι συχνά. Το εάν είναι συχνά θα καθοριστεί στο τέλος για να γίνει χρήση τους στην επόμενη φάση. Στην παρούσα εργασία έγινε εφαρμογή του αλγορίθμου Apriori [25] και με τις προεπιλεγμένες παραμέτρους αλλά και με μεταβολή των τιμών υποστήριξης και εμπιστοσύνης.

Αρχικά εκτελέστηκε ο αλγόριθμος Apriori με τις default τιμές του Weka στην καρτέλα Associate όπως φαίνεται παρακάτω:



**Εικόνα 16 : Εκτέλεση αλγορίθμου Apriori με default τιμές**  
Πηγή Εικόνας : Λογισμικό WEKA

```

=== Run information ===

Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    Comments
Instances:   152
Attributes:  6
             article
             commentator
             commentatorGender
             comment
             emotion
             recommendation

=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.25 (38 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 9
Size of set of large itemsets L(2): 16
Size of set of large itemsets L(3): 14
Size of set of large itemsets L(4): 4

Best rules found:

1. commentatorGender=NoGender 56 ==> commentator=PublicService 56 <conf:(1)> lift:(2.71) lev:(0.23) [35] conv:(35.37)
2. commentator=PublicService 56 ==> commentatorGender=NoGender 56 <conf:(1)> lift:(2.71) lev:(0.23) [35] conv:(35.37)
3. commentatorGender=Female 51 ==> commentator=individual 51 <conf:(1)> lift:(1.58) lev:(0.12) [18] conv:(18.79)
4. commentatorGender=NoGender recommendation=Yes 47 ==> commentator=PublicService 47 <conf:(1)> lift:(2.71) lev:(0.2) [29] conv:(29.68)
5. commentator=PublicService recommendation=Yes 47 ==> commentatorGender=NoGender 47 <conf:(1)> lift:(2.71) lev:(0.2) [29] conv:(29.68)
6. commentatorGender=Male 45 ==> commentator=individual 45 <conf:(1)> lift:(1.58) lev:(0.11) [16] conv:(16.58)
7. commentatorGender=NoGender comment=Neutral 44 ==> commentator=PublicService 44 <conf:(1)> lift:(2.71) lev:(0.18) [27] conv:(27.79)
8. commentator=PublicService comment=Neutral 44 ==> commentatorGender=NoGender 44 <conf:(1)> lift:(2.71) lev:(0.18) [27] conv:(27.79)
9. commentatorGender=Female recommendation=Yes 43 ==> commentator=individual 43 <conf:(1)> lift:(1.58) lev:(0.1) [15] conv:(15.84)
10. commentatorGender=NoGender emotion=NoEmotion 43 ==> commentator=PublicService 43 <conf:(1)> lift:(2.71) lev:(0.18) [27] conv:(27.16)

```

### Εικόνα 17 : Αποτελέσματα εκτέλεσης αλγορίθμου Apriori με default τιμές

Πηγή Εικόνας : Λογισμικό WEKA

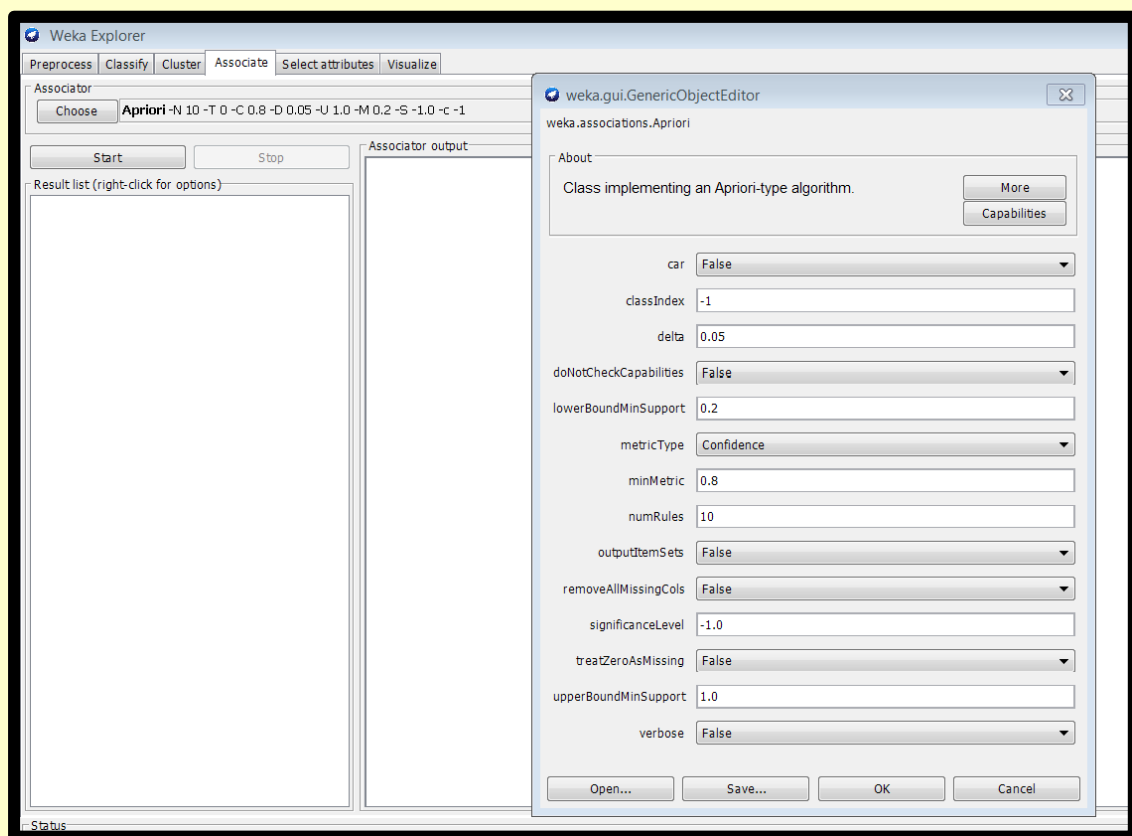
Με minimum support 0.25 και min metric 0,9 και αναζήτηση σε 38 περιπτώσεις μετά από εκτέλεση 15 κύκλων ο αλγόριθμος παράγαγε 4 μεγάλα item sets και τους παρακάτω 10 κανόνες (default όριο Weka):

- Αν ο σχολιαστής δεν έχει φύλο τότε σε 56 περιπτώσεις ο σχολιαστής είναι δημόσια υπηρεσία.
- Αν ο σχολιαστής είναι δημόσια υπηρεσία τότε σε 56 περιπτώσεις ο σχολιαστής δεν έχει φύλο.
- Αν ο σχολιαστής είναι γυναίκα τότε σε 51 περιπτώσεις ο σχολιαστής είναι μεμονωμένο πρόσωπο.
- Αν ο σχολιαστής δεν έχει φύλο και προτείνει αλλαγές για τον Νόμο τότε σε 47 περιπτώσεις ο σχολιαστής είναι δημόσια υπηρεσία.
- Αν ο σχολιαστής είναι δημόσια υπηρεσία και προτείνει αλλαγές για τον Νόμο τότε σε 47 περιπτώσεις ο σχολιαστής δεν έχει φύλο.
- Αν ο σχολιαστής είναι άντρας τότε σε 45 περιπτώσεις ο σχολιαστής είναι μεμονωμένο πρόσωπο.
- Αν ο σχολιαστής δεν έχει φύλο και το σχόλιο του είναι ουδέτερο τότε σε 44 περιπτώσεις ο σχολιαστής είναι δημόσια υπηρεσία.
- Αν ο σχολιαστής είναι δημόσια υπηρεσία και το σχόλιο του είναι ουδέτερο τότε σε 44 περιπτώσεις ο σχολιαστής δεν έχει φύλο.

- Αν ο σχολιαστής είναι γυναίκα και προτείνει αλλαγές για τον Νόμο τότε σε 43 περιπτώσεις ο σχολιαστής είναι μεμονωμένο πρόσωπο.
- Αν ο σχολιαστής δεν έχει φύλο και το σχόλιο του δεν έχει κανένα συναίσθημα τότε σε 43 περιπτώσεις ο σχολιαστής είναι δημόσια υπηρεσία.

Οι περισσότεροι κανόνες επικεντρώνονται στο φύλο του σχολιαστή και στο εάν ο σχολιαστής είναι μεμονωμένο πρόσωπο ή δημόσια υπηρεσία και άρα δεν είναι ικανοποιητικοί αφού το ζητούμενο είναι να αναλυθεί το είδος του σχολιασμού, το συναίσθημα του σχολιαστή και το εάν προτείνει ο σχολιαστής αλλαγές στον Νόμο συμβάλλοντας έτσι στη βελτίωση του.

Εκτελέστηκε ξανά ο αλγόριθμος Apriori με αλλαγή στις default τιμές του Weka στην καρτέλα Associate όπως φαίνεται παρακάτω:



**Εικόνα 18 : Εκτέλεση αλγορίθμου Apriori με min support 0.2 και min metric 0.8**  
 Πηγή Εικόνας : Λογισμικό WEKA

```

=== Run information ===

Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.8 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1
Relation:    Comments
Instances:   152
Attributes:  6
             article
             commentator
             commentatorGender
             comment
             emotion
             recommendation

=== Associator model (full training set) ===
Apriori
=====

Minimum support: 0.35 (53 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 13

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6
Size of set of large itemsets L(2): 6
Size of set of large itemsets L(3): 1

Best rules found:

1. commentatorGender=NoGender 56 ==> commentator=PublicService 56 <conf:(1)> lift:(2.71) lev:(0.23) [35] conv:(35.37)
2. commentator=PublicService 56 ==> commentatorGender=NoGender 56 <conf:(1)> lift:(2.71) lev:(0.23) [35] conv:(35.37)
3. comment=Neutral emotion=NoEmotion 80 ==> recommendation=Yes 75 <conf:(0.94)> lift:(1.15) lev:(0.06) [9] conv:(2.46)
4. comment=Neutral 95 ==> recommendation=Yes 88 <conf:(0.93)> lift:(1.14) lev:(0.07) [10] conv:(2.19)
5. emotion=NoEmotion recommendation=Yes 82 ==> comment=Neutral 75 <conf:(0.91)> lift:(1.46) lev:(0.16) [23] conv:(3.84)
6. comment=Neutral recommendation=Yes 88 ==> emotion=NoEmotion 75 <conf:(0.85)> lift:(1.32) lev:(0.12) [18] conv:(2.23)
7. comment=Neutral 95 ==> emotion=NoEmotion 80 <conf:(0.84)> lift:(1.31) lev:(0.12) [18] conv:(2.11)
8. emotion=NoEmotion 98 ==> recommendation=Yes 82 <conf:(0.84)> lift:(1.03) lev:(0.01) [2] conv:(1.06)
9. emotion=NoEmotion 98 ==> comment=Neutral 80 <conf:(0.82)> lift:(1.31) lev:(0.12) [18] conv:(1.93)
10. commentator=individual 96 ==> recommendation=Yes 77 <conf:(0.8)> lift:(0.98) lev:(-0.01) [-1] conv:(0.88)

```

**Εικόνα 19 : Αποτελέσματα εκτέλεσης αλγορίθμου Apriori με min support 0.2 και min metric 0.8**

Πηγή Εικόνας : Λογισμικό WEKA

Με minimum support 0.2 και min metric 0,8 και αναζήτηση σε 152 περιπτώσεις μετά από εκτέλεση 13 κύκλων ο αλγόριθμος παρήγαγε 3 μεγάλα item sets και τους παρακάτω 10 κανόνες (default όριο Weka):

- Αν ο σχολιαστής δεν έχει φύλο τότε σε 56 περιπτώσεις ο σχολιαστής είναι δημόσια υπηρεσία.
- Αν ο σχολιαστής είναι δημόσια υπηρεσία τότε σε 56 περιπτώσεις ο σχολιαστής δεν έχει φύλο.
- Αν το σχόλιο είναι ουδέτερο και ο σχολιαστής δεν δείχνει κανένα συναίσθημα σε 80 περιπτώσεις τότε σε 75 περιπτώσεις ο σχολιαστής προτείνει αλλαγές για τον Νόμο.
- Αν το σχόλιο είναι ουδέτερο σε 95 περιπτώσεις τότε σε 88 περιπτώσεις ο σχολιαστής προτείνει αλλαγές για τον Νόμο.
- Αν ο σχολιαστής δεν δείχνει συναίσθημα και προτείνει αλλαγές στον Νόμο σε 82 περιπτώσεις τότε σε 75 περιπτώσεις το σχόλιο είναι ουδέτερο.
- Αν το σχόλιο είναι ουδέτερο και ο σχολιαστής προτείνει αλλαγές στον Νόμο σε 88 περιπτώσεις τότε σε 75 περιπτώσεις ο σχολιαστής δεν δείχνει συναίσθημα.

- Αν το σχόλιο είναι ουδέτερο σε 95 περιπτώσεις τότε σε 80 περιπτώσεις ο σχολιαστής δεν δείχνει συναίσθημα.
- Αν ο σχολιαστής δεν δείχνει συναίσθημα σε 98 περιπτώσεις τότε σε 82 περιπτώσεις ο σχολιαστής προτείνει αλλαγές στον Νόμο.
- Αν ο σχολιαστής δεν δείχνει συναίσθημα σε 98 περιπτώσεις τότε σε 80 περιπτώσεις το σχόλιο είναι ουδέτερο.
- Αν ο σχολιαστής είναι μεμονωμένο άτομο σε 96 περιπτώσεις τότε σε 77 περιπτώσεις ο σχολιαστής προτείνει αλλαγές στον Νόμο.

Οι κανόνες που παρήχθησαν είναι σαφώς καλύτεροι σε σχέση με πριν όπου εφαρμόστηκαν οι default τιμές του Weka γιατί πλέον εξήχθησαν περισσότεροι κανόνες που αφορούν το είδος του σχολιασμού, το συναίσθημα του σχολιαστή και το εάν προτείνονται αλλαγές στον Νόμο.

Εκτελέστηκε ξανά ο αλγόριθμος Apriori με τις default τιμές του Weka με παράλειψη όμως από το αρχείο δεδομένων arff τα δεδομένα που αφορούν στο φύλο του σχολιαστή και στο εάν ο σχολιαστής είναι μεμονωμένο άτομο ή δημόσια υπηρεσία.

Τα περιεχόμενα του αρχείου arff εμφανίζονται στο παράρτημα της εργασίας στο σημείο ii.

```

=== Run information ===

Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    Comments
Instances:   152
Attributes:  4
             article
             comment
             emotion
             recommendation

=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.1 (15 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 10

Size of set of large itemsets L(2): 12

Size of set of large itemsets L(3): 3

Best rules found:

1. emotion=Irritated 20 ==> comment=Negative 20 <conf:(1)> lift:(3.53) lev:(0.09) [14] conv:(14.34)
2. emotion=Irritated recommendation=Yes 17 ==> comment=Negative 17 <conf:(1)> lift:(3.53) lev:(0.08) [12] conv:(12.19)
3. article=3 comment=Neutral 16 ==> recommendation=Yes 16 <conf:(1)> lift:(1.23) lev:(0.02) [2] conv:(2.95)
4. comment=Neutral emotion=NoEmotion 80 ==> recommendation=Yes 75 <conf:(0.94)> lift:(1.15) lev:(0.06) [9] conv:(2.46)
5. comment=Neutral 95 ==> recommendation=Yes 88 <conf:(0.93)> lift:(1.14) lev:(0.07) [10] conv:(2.19)
6. article=3 26 ==> recommendation=Yes 24 <conf:(0.92)> lift:(1.13) lev:(0.02) [2] conv:(1.6)
7. emotion=NoEmotion recommendation=Yes 82 ==> comment=Neutral 75 <conf:(0.91)> lift:(1.46) lev:(0.16) [23] conv:(3.84)

```

**Εικόνα 20 : Εκτέλεση αλγορίθμου Apriori με default τιμές και εξαίρεση των attributes Commentator και CommentatorGender**

Πηγή Εικόνας : Λογισμικό WEKA

Με minimum support 0.1 και min metric 0,9 και αναζήτηση σε 15 περιπτώσεις μετά από εκτέλεση 18 κύκλων ο αλγόριθμος παράγαγε 3 μεγάλα item sets και τους παρακάτω 7 κανόνες:

- Αν το συναίσθημα του σχολιαστή δείχνει ότι είναι εκνευρισμένος τότε σε 20 περιπτώσεις το σχόλιο είναι αρνητικό.
- Αν το συναίσθημα του σχολιαστή δείχνει ότι είναι εκνευρισμένος και προτείνει αλλαγές στον Νόμο τότε σε 17 περιπτώσεις το σχόλιο είναι αρνητικό.
- Αν το σχόλιο αφορά το άρθρο 3 του Νόμου και είναι ουδέτερο τότε σε 16 περιπτώσεις ο σχολιαστής προτείνει αλλαγές στον Νόμο.
- Αν το σχόλιο είναι ουδέτερο και ο σχολιαστής δεν δείχνει κανένα συναίσθημα σε 80 περιπτώσεις τότε σε 75 περιπτώσεις ο σχολιαστής προτείνει αλλαγές για τον Νόμο.
- Αν το σχόλιο είναι ουδέτερο σε 95 περιπτώσεις τότε σε 88 περιπτώσεις ο σχολιαστής προτείνει αλλαγές για τον Νόμο.
- Αν το σχόλιο αφορά το άρθρο 3 του Νόμου σε 26 περιπτώσεις τότε σε 24 περιπτώσεις ο σχολιαστής προτείνει αλλαγές στον Νόμο.
- Αν ο σχολιαστής δεν δείχνει συναίσθημα και προτείνει αλλαγές στον Νόμο σε 82 περιπτώσεις τότε σε 75 περιπτώσεις το σχόλιο είναι ουδέτερο.

Παρατηρήθηκε ότι στους 4 από τους 7 κανόνες η βεβαιότητα έπεσε κάτω από 1. Οι κανόνες που έχουν παραχθεί είναι πολύ καλύτεροι σε σχέση με πριν, επειδή εξήχθησαν περισσότεροι κανόνες που αφορούν στο είδος του σχολιασμού, το συναίσθημα του σχολιαστή και το εάν προτείνει αλλαγές στον Νόμο

#### **6.6.4 Εφαρμογή αλγορίθμου K-Means**

Ο αλγόριθμος K-Means [27] είναι ένας αλγόριθμος ο οποίος εφαρμόζεται συχνά και αποτελεί την ρίζα για πολλούς αλγορίθμους. Παράγει ένα σύνολο συσταδοποιήσεων τα οποία δεν έχουν μεταξύ τους μεγάλη σχέση. Στόχος του αλγορίθμου είναι η βελτιστοποίηση της συνάρτησης κόστους. Ο αλγόριθμος λειτουργεί σε δύο βήματα. Στο πρώτο προσπελούνται τα διανύσματα και για κάθε διάνυσμα υπολογίζεται η απόσταση του από τις ομάδες που ήδη υπάρχουν με τον υπολογισμό της Ευκλείδειας απόστασης από το μέσο διάνυσμα και έπειτα σχηματίζεται ένα σύνολο για κάθε ομάδα (K- σύνολα διανυσμάτων). Στο δεύτερο βήμα τροποποιούνται κατάλληλα τα μέσα διανύσματα αφού υπολογίζονται εκ νέου. Ο αλγόριθμος σταματά όταν οι ενημερωμένες τιμές των διανυσμάτων έχουν απειροελάχιστη διαφορά.

Ο αλγόριθμος εκτελέστηκε για όλα τα attributes:

```
=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t
Relation:    Comments
Instances:   152
Attributes:  6
             article
             commentator
             commentatorGender
             emotion
             recommendation

Ignored:     comment

Test mode:   Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 5
Within cluster sum of squared errors: 225.0

Initial starting points (random):

Cluster 0: 1,individual,Male,Angry,No
Cluster 1: 2,PublicService,NoGender,NoEmotion,Yes
Cluster 2: 1,PublicService,NoGender,Irritated,No

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute          Full Data          Cluster#
                   (152.0)           (84.0)           (56.0)           (12.0)
=====
article            3                  19                3                  1
commentator       individual         individual PublicService individual
commentatorGender NoGender          Female            NoGender          Male
emotion           NoEmotion         NoEmotion         NoEmotion         Angry
recommendation    Yes               Yes               Yes               No

Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      84 ( 55%)
1      56 ( 37%)
2      12 (  8%)

Class attribute: comment
Classes to Clusters:

  0  1  2 <-- assigned to cluster
  0  0  1 | Positive
50 44  1 | Neutral
27  9  7 | Negative
 7  3  3 | Question

Cluster 0 <-- Negative
Cluster 1 <-- Neutral
Cluster 2 <-- Question

Incorrectly clustered instances :      78.0      51.3158 %
```

**Εικόνα 21 : Εκτέλεση αλγορίθμου K-Means**  
Πηγή Εικόνας : Λογισμικό WEKA



Ο αλγόριθμος έτρεξε για 5 επαναλήψεις σε 0.01 sec και είχε ποσοστό incorrectly clustered instances 51.3158%.

Θα εκτελεστεί ξανά ο αλγόριθμος K-Means αλλά αυτή την φορά με εξαίρεση των attributes Commentator και CommentatorGender αφού περισσότερο ενδιαφέρει το είδος του σχολιασμού, το συναίσθημα του σχολιαστή και το εάν προτείνει αλλαγές στον Νόμο και όχι τόσο το φύλο του ή εάν είναι μεμονωμένο πρόσωπό ή δημόσια υπηρεσία.

```
=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25
Relation:    Comments
Instances:   152
Attributes:  4
             article
             emotion
             recommendation

Ignored:     comment
Test mode:   Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 3
Within cluster sum of squared errors: 174.0

Initial starting points (random):

Cluster 0: 1,Angry,No
Cluster 1: 2,NoEmotion,Yes
Cluster 2: 1,Irritated,No

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute      Full Data      Cluster#
              (152.0)      (29.0)      (120.0)      (3.0)
-----
article        3              1              3              1
emotion        NoEmotion NoEmotion NoEmotion Irritated
recommendation Yes          No          Yes          No

Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      29 ( 19%)
1      120 ( 79%)
2       3 (  2%)

Class attribute: comment
Classes to Clusters:

 0 1 2 <-- assigned to cluster
 1 0 0 | Positive
 9 86 0 | Neutral
 7 33 3 | Negative
12 1 0 | Question

Cluster 0 <-- Question
Cluster 1 <-- Neutral
Cluster 2 <-- Negative

Incorrectly clustered instances :      51.0      33.5526 %
```

**Εικόνα 22 : Εκτέλεση αλγορίθμου K-means με εξαίρεση των attributes Commentator και CommentatorGender**

Πηγή Εικόνας : Λογισμικό WEKA

Παρατηρήθηκε ότι ο αλγόριθμος έτρεξε αυτή τη φορά για 3 επαναλήψεις στον ίδιο όμως χρόνο με πριν (0.01 sec) και είχε ποσοστό incorrectly clustered instances 33.5526% σε αντίθεση με την προηγούμενη εκτέλεση όπου το ποσοστό ήταν 51.3158%.

## 6.6.5 Εφαρμογή αλγορίθμου Naive Bayes

Εφαρμόστηκε ο αλγόριθμος Naive Bayes [23] σε όλα τα δεδομένα χωρίς εξαίρεση attributes έτσι ώστε να μπορούν να συγκριθούν τα αποτελέσματα με τα ευρήματα της εκτέλεσης του αλγορίθμου J48.

```
=== Run information ===
Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    Comments
Instances:   152
Attributes:  6
             article
             commentator
             commentatorGender
             comment
             emotion
             recommendation
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier
Attribute          Class
                   Yes  No
                   (0.61) (0.19)
=====
article
 1                5.0  6.0
 2               12.0  1.0
 3               25.0  3.0
 4                7.0  1.0
 5                4.0  3.0
 6                2.0  1.0
 7                1.0  2.0
 8                2.0  1.0
 9                1.0  2.0
11                6.0  3.0
12                5.0  2.0
13                3.0  1.0
14                2.0  2.0
15                1.0  2.0
17                1.0  3.0
18                2.0  1.0
19               17.0  4.0
20                1.0  2.0
22                2.0  1.0
23                5.0  1.0
24                2.0  1.0
25                3.0  1.0
26                2.0  1.0
27                5.0  2.0
29               15.0  5.0
30               12.0  2.0
32                8.0  1.0
[total]          151.0 55.0

commentator
individual         78.0 20.0
PublicService     48.0 10.0
[total]           126.0 30.0

commentatorGender
Male              35.0 12.0
Female            44.0  5.0
NoGender          48.0 10.0
[total]           127.0 31.0

comment
Positive          1.0  2.0
Neutral           89.0  8.0
Negative          35.0 10.0
Question          3.0 12.0
[total]           128.0 32.0

emotion
Hopeful           2.0  1.0
Excited           1.0  2.0
Pleading          6.0  1.0
Angry             2.0  5.0
Irritated        18.0  4.0
NoEmotion        83.0 17.0
Worried          19.0  5.0
[total]           131.0 35.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      134          88.1579 %
Incorrectly Classified Instances    18          11.8421 %
Kappa statistic                    0.5428
Mean absolute error                 0.1843
Root mean squared error             0.3211
Relative absolute error             60.7134 %
Root relative squared error         82.7902 %
Total Number of Instances          152

=== Detailed Accuracy By Class ===
              TP Rate  FP Rate  Precision  Recall  F-Measure  MDC     ROC Area  PRC Area  Class
0.968   0.500   0.896   0.968   0.930   0.561   0.784   0.935   Yes
0.500   0.032   0.778   0.500   0.609   0.561   0.784   0.570   No
Weighted Avg.   0.882   0.414   0.874   0.882   0.871   0.561   0.784   0.868

=== Confusion Matrix ===
  a  b  <-- classified as
120  4 | a = Yes
 14 14 | b = No
```

Εικόνα 23 : Εκτέλεση αλγορίθμου Naive Bayes  
Πηγή Εικόνας : Λογισμικό WEKA

Το 88.1579% είναι correctly classified instances ενώ το 11.8421% είναι incorrectly classified instances και ο χρόνος που απαιτήθηκε ήταν 0 sec. Παρατηρήθηκε ότι τα αποτελέσματα του αλγορίθμου Naive Bayes ήταν καλύτερα σε σχέση με αυτά του J48 αφού τα correctly classified instances του είχαν ποσοστό 84.8684% και τα incorrectly classified instances ποσοστό 15.1316% και ο χρόνος που απαιτήθηκε για την εκτέλεση ήταν 0.12 sec.

## Συμπεράσματα

---

Η δημόσια διοίκηση με τη συμβολή των Τεχνολογιών Πληροφοριών και Επικοινωνιών (ΤΠΕ) έγινε περισσότερο εξωστρεφής, αφουγκράζεται τη γνώμη των πολιτών, αύξησε σημαντικά την παραγωγικότητα της, βελτίωσε τις παρεχόμενες υπηρεσίες της, ανέδειξε νέες ιδέες και βελτίωσε τις αποφάσεις που λαμβάνει. Από την άλλη πλευρά οι πολίτες με την ενεργή συμμετοχή τους στις αποφάσεις της δημόσιας διοίκησης δεν αισθάνονται αποκομμένοι από τα κοινά, συμμετέχουν ενεργά στις αποφάσεις της Δημόσιας Διοίκησης, αυξήθηκε η εμπιστοσύνη τους προς το κράτος, αισθάνονται ότι η Δημόσια Διοίκηση είναι ανοιχτή σε προτάσεις και μπορούν να ενημερώνονται για αρκετές εργασίες της Δημόσιας Διοίκησης και να εξυπηρετούνται όλο το 24ωρο.

Με τη χρήση τεχνικών και εργαλείων εξόρυξης δεδομένων η Δημόσια Διοίκηση μπορεί να αντλήσει και να αναλύσει δεδομένα όπου θα την βοηθήσουν να εντοπίσει και να αναγνωρίσει μοτίβα π.χ. οικονομική απάτη, να εποπτεύσει και να επιβάλει πρόστιμα σε ιδιώτες, να ανιχνεύσει το κυβερνοέγκλημα και να κατανοήσει τους πολίτες και να λάβει καλύτερες αποφάσεις.

Οι προκλήσεις που καλείται να αντιμετωπίσει η Δημόσια Διοίκηση κατά τη δημιουργία Πληροφοριακών Συστημάτων είναι πολλές. Οι κυριότερες είναι η εφαρμογή Νόμων σε Πληροφοριακά Συστήματα όπου σε περίπτωση τροποποίησης των Νόμων το Πληροφοριακό Σύστημα πρέπει να επανασχεδιαστεί, η εξυπηρέτηση του δημοσίου συμφέροντος, η ικανοποίηση των αναγκών των πολιτών, η διασφάλιση της προστασίας των προσωπικών δεδομένων των πολιτών, η ικανοποίηση κοινωνικών και ηθικών περιορισμών, η διαχείριση και επεξεργασία του συνεχώς αυξανόμενου όγκου δεδομένων και τα δεδομένα που προέρχονται από διαφορετικές πηγές και με διαφορετική δομή και ανάλυση.

Τα πλεονεκτήματα από την εξόρυξη γνώμης από τα ηλεκτρονικά μέσα είναι πολλά καθώς εξοικονομούνται δαπάνες, χρόνος και ανθρώπινοι πόροι που παλαιότερα απαιτούνταν για έρευνες και δημοσκοπήσεις, οι πολίτες εκφράζουν με ελεύθερο κείμενο τις απόψεις και τις προτάσεις τους χωρίς πίεση χρόνου ή ερωτήσεων, τα στοιχεία από την εξόρυξη γνώμης μπορούν να συλλεχθούν σε τακτά χρονικά διαστήματα, η συμμετοχή των πολιτών αυξήθηκε αφού η πρόσβαση στα ηλεκτρονικά μέσα είναι προσβάσιμη και εύκολη για μια μεγάλη μερίδα πολιτών.

Η δημόσια διαβούλευση είναι ένα βήμα για τους πολίτες στο οποίο μπορούν να εκφράσουν τις απόψεις τους ελεύθερα και να συμμετέχουν στη λήψη αποφάσεων. Για τη δημόσια διοίκηση η δημόσια διαβούλευση αποτελεί έναν δίαυλο επικοινωνίας με τους πολίτες και μια καλή πηγή άντλησης των απόψεων των πολιτών που μπορεί να αξιοποιηθεί με τη συλλογή και ανάλυση των δεδομένων.

Η διαβίωση των πολιτών γίνεται καλύτερη όσο ο δημόσιος τομέας γίνεται αποτελεσματικότερος αφού απαιτούνται λιγότεροι οικονομικοί πόροι και φόροι για τη λειτουργία του κράτους.

Στην παρούσα εργασία μελετήθηκαν τα σχόλια που εμφανίζονται στον Διαδικτυακό Τόπο Διαβουλεύσεων του Υπουργείου Ψηφιακής Διακυβέρνησης. Τα σχόλια αντλήθηκαν από τη Δημόσια Διαβούλευση για τις διατάξεις Νόμου που αφορούν στο Γενικό Εμπορικό Μητρώο. Στις διατάξεις περιγράφεται η βελτίωση του

Πληροφοριακού Συστήματος του Γενικού Εμπορικού Μητρώου. Συνολικά μελετήθηκαν 153 σχόλια.

Τα σχόλια επεξεργάστηκαν με τους αλγορίθμους J48, Apriori, K-Means και Naive Bayes με χρήση του λογισμικού Weka, σε μια προσπάθεια να αναλυθεί το είδος του σχολιασμού, το συναίσθημα του σχολιαστή και το εάν προτείνει αλλαγές στον Νόμο δηλαδή αν συμμετέχει ενεργά στη βελτίωση του.

Από τη συλλογή των δεδομένων παρατηρήθηκε ότι τα περισσότερα σχόλια συγκέντρωσε το άρθρο 3 των διατάξεων που αφορά στη διάρθρωση του Γ.Ε.ΜΗ. και τις Υπηρεσίες του. Το 63% των σχολιαστών ήταν μεμονωμένα άτομα που από τη μια φανερώνει την υψηλή συμμετοχή των πολιτών και από την άλλη το αυξημένο ενδιαφέρον των αρμόδιων ή εμπλεκόμενων δημοσίων υπηρεσιών και στελεχών τους. Στην πλειοψηφία τους οι σχολιαστές ήταν γυναίκες. Τα περισσότερα σχόλια όπως ήταν φυσικό ήταν ουδέτερα (62,5%) και δεν περιείχαν συναίσθημα (64,5%) καθώς στην περίπτωση της διαβούλευσης δεν μπορούν να υπάρχουν πολλά θετικά σχόλια και συναισθήματα αφού ο σκοπός της διαβούλευσης είναι οι πολίτες να εκφράσουν τις ενστάσεις τους στις νέες ρυθμίσεις. Το 82% των σχολιαστών έκανε προτάσεις για αλλαγές στον Νόμο γεγονός που φανερώνει την επιτυχία της διαβούλευσης.

Από την επεξεργασία και ανάλυση των δεδομένων με το λογισμικό weka παρατηρήθηκε ότι ο αλγόριθμος J48 εξήγαγε καλύτερα αποτελέσματα όταν χρησιμοποιήθηκε σε όλα τα δεδομένα χωρίς το φίλτρο attribute selection.

Οι κανόνες που εξήχθησαν από τον αλγόριθμο Apriori είναι πολύ σημαντικοί και κάνει τις συσχετίσεις μεταξύ των δεδομένων να γίνουν κατανοητές. Τα δεδομένα που χρησιμοποιήθηκαν, επειδή περιλάμβαναν και τιμές που αφορούσαν το προφίλ του σχολιαστή όπως το φύλο του και το εάν είναι μεμονωμένο άτομο ή δημόσια υπηρεσία, επηρέασαν τους κανόνες οι οποίοι επικεντρώθηκαν σε αυτά τα χαρακτηριστικά. Όταν αυτά τα χαρακτηριστικά (attributes) παραλείφθηκαν οι κανόνες έγιναν πιο ξεκάθαροι και αφορούσαν στο ζητούμενο, δηλαδή το είδος του σχολιασμού, το συναίσθημα του σχολιαστή και το εάν προτείνει αλλαγές στον Νόμο.

Το ίδιο συνέβη και στον αλγόριθμο K-means όπου παραλείφθηκαν τα χαρακτηριστικά (attributes) που αφορούσαν το προφίλ του σχολιαστή και το φύλο του σχολιαστή, αφού περισσότερο ενδιαφέρει το είδος του σχολιασμού, το συναίσθημα του σχολιαστή και το εάν προτείνει αλλαγές στον Νόμο. Τότε ο αριθμός των λανθασμένων ομαδοποιήσεων μειώθηκε.

Επίσης παρατηρήθηκε ότι τα αποτελέσματα του αλγορίθμου Naive Bayes ήταν καλύτερα από αυτά του J48. Συνεπώς όσο καλός και να είναι ένας αλγόριθμος θα πρέπει η είσοδος των δεδομένων να επιλεγεί προσεκτικά και τα δεδομένα να έχουν αξία έτσι ώστε οι εξαγόμενες πληροφορίες να μας οδηγήσουν στη λήψη βέλτιστων αποφάσεων.

## Βιβλιογραφία

---

1. Διαδικτυακός ιστότοπος Ανοιχτής Διακυβέρνησης του Υπουργείου Εσωτερικών ([www.opengov.gr](http://www.opengov.gr)).
2. Chen M., Mao, S. and Liu, Y. (2014). Big Data: A Survey, Mobile Networks and Application, 171-209.
3. Liu, S. M. and Yuan, Q. (2015). The Evolution of Information and Communication Technology In Public Administration, 140-151
4. Ko, A., Gabor, A. and Szabo, Z. (2013) Policy Making Improvement though Social Learning, Technology Enabled Innovation for Democracy. Springer Berlin Haidelberg, 228.
5. Desouza, K. and Benoy, J. (2014). Big Data in the Public Sector: Lessons for Practitioners and Scholars, Administration and Society, 1-23.
6. Raad, E., Bechara, A.B., and Chbeir, R. (2015) Preventing sensitive relationships disclosure for better social media preservation. International Journal of Information Security, 1-22.
7. Jiao, Y., Yinghui, W., Shaohua, Z. Yin, L., Baoming, Y. and Lei, Y. (2013). A cloud approach to unified lifecycle data management in architecture, engineering, construction and facilities management: Integrating BIMs and SNS, Advanced Engineering Informatics, 173-188.
8. Quick, D. and Choo, K. R. (2014). Impacts of increasing volume of digital forensic data: A survey and future research challenges, Digital Investigation, 273-294.
9. El Kadiri, S., Grabot, B., Thoben, K., Hribernik K., Emmanouilidis, C., Von Cieminski, G. and Kiritsis, D. (2015). Current trends on ICT technologies for enterprise information systems. Computers in Industry, 14-15.
10. Jordan, L. (2015). The problem with Big Data in Translational Medicine. A review of where we've been and the possibilities ahead, Applied & Translational Genomics, 1-4
11. Munné, R. (2016). Big Data in the Public Sector. pp. 195-208
12. Kamel Boulos, M. and Al-Shorbaji N. (2014) On the Internet of Things, smart cities and the WHO Healthy Cities, International Journal of Health Geographics, 1-6.
13. Fredriksson, C., Mubarak, F., Tuohimaa, M. and Zhan, M. (2017), Big Data in the Public Sector: A Systematic Literature Review, Scandinavian Journal of Public Administration, Vol 21, No 3.
14. Αποστολάκης Ι., Λουκής Ε., Χάλαρης Ι. (2008) Ηλεκτρονική Δημόσια Διοίκηση, Οργάνωση Τεχνολογία και Εφαρμογές, 105-129.
15. Νόμος 4635/2019 (ΦΕΚ 167/30-10-2019/τ.Α') «Επενδύω στην Ελλάδα και άλλες διατάξεις.» Μέρος ΙΣΤ', Άρθρα 89 και 91.
16. Jindal, N. and Liu, B. (2008). Opinion spam and analysis. Proceedings of the 2008 international conference on web search and data mining. ACM, 219-230.
17. Liu, B. (2012). Opinion mining and sentiment analysis. Sentiment Analysis and Opinion Mining Morgan & Claypool Publishers San Francisco, 14.
18. Av, M.K. and Nandkumar, A. (2020). A Survey on Challenges and Research Opportunities in Opinion Mining. SN Computer Science, 1-6.

19. Kleinberg, J. (2007). Challenges in mining social network data: processes, privacy and paradoxes. Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining, 4-5.
20. Adedoyin-Olowe, M, Gaber, M.M, Stahl F. (2013). A survey of data mining techniques for social media analysis. Journal of data mining digital humanities, 5.
21. Bifet, A., Frank, E. (2010) Sentiment Knowledge discovery in Twitter streaming data. Proceedings of the 13th International Conference, DS 2010, Canberra, Australia, 5.
22. Ross J. Quinlan. (1993). C4.5: Programs for machine learning. San Mateo, CA: Morgan Kaufman.
23. Domingos, P. and Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29(2-3), 103–130.
24. Weka team (2020). *Weka The workbench for machine learning*, Machine Learning Group, University of Waikato, New Zealand. <https://www.cms.waikato.ac.nz/~ml/Weka/>
25. Al-Maolegi, M. and Arkok, B. (2014). An Improved Apriori Algorithm For Association Rules. *International Journal on Natural Language Computing*, 3. Doi:10.5121/ijnlc.2014.3103
26. Agrawal, R., Imielinski, T. and Swami, A. (1993). Mining association rules between sets of items in large databases. ACM SIGMOD Rec. 22. 207-216. Doi:10.1145/170035.170072
27. Raykov, YP. , Boukouvalas, A., Baig, F. (2016) Little MA. What to do when k-meansclustering fails: A simple yet principledalternative algorithm. PLoS One. 2016; 11(9):e0162259. DOI: 10.1371/journal.pone.0162259
28. Solanki, H. (2013). Comparative study of data mining tools and analysis with unified data mining theory. *International Journal of Computer Applications*, 75(16), 23–28. Doi:10.5120/13195-0862
29. Witten, I., Frank, E., Hall, M.A. and Pal, C.J. (2016). *Data mining: Practical machine learning tools and techniques*, 4th edition, chapter 4. Cambridge, MA: Morgan Kaufmann
30. North, M.A. (2012). *Data Mining for the Masses*. Licensed under a Creative Commons Attribution 3.0 License.
31. Adedoyin-Olowe, M., Gaber, M.M. and Stahl, F. (2014). A survey of data mining techniques for social media analysis. *Journal of data mining digital humanities*. arXiv:1312.4617v2
32. Zadari, T. (2015). Data Mining in Social Media. *International Journal of Scientific & Engineering Research*, 6( 7), (July 2015), 152-154.
33. Pang, B. and Lee, L. (2008). *Opinion Mining and Sentiment Analysis (Foundations and Trends (R) in Information Retrieval)*.
34. Kar, A., & Mandal, D. P. (2011). Finding opinion strength using fuzzy logic on web reviews. *International Journal of Engineering and Industries*, 2(1), 37-43.
35. Dhokrat, A., Khillare, S., & Mahender, C. N. (2015). Review on techniques and tools used for opinion mining. *International Journal of Computer Applications Technology and Research*, 4(6), 419-424.



36. Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: analyzing and comparing opinions on the web. In Proceedings of the 14th international conference on World Wide Web (pp. 342-351). ACM.
37. Wang H. (2008). A knowledge management approach to data mining process for business intelligence. *Industrial Management & Data Systems*. 2008; 108(5): 622–34.
38. Schuler, D. (2010). Civic Intelligence and the Evolution of Community Networks. *AI and Society*, 25 (3), 291-307.
39. Hartz-Karp, J. (2005). A Case Study in Deliberative Democracy: Dialogue with the City, *Journal of Public Deliberation*: Vol. 1 (1:6).
40. Nandi, M. & Nayak, N. (2008). ERP Implementation in a Public Sector Organization: A dialectic perspective. In book: *Enterprise systems and business process management: Global best practices*, Macmillan India. Editors: M. Jaiswal & R. K. Garg
41. Ziemia, E. and Obłąk, I. (2014). The survey of information systems in public administration in Poland. *Interdisciplinary Journal of Information, Knowledge, and Management*, 9, 31-56.
42. Cordella, A., and Bonina, C. M. (2012). A public value perspective for ICT enabled public sector reforms: A theoretical reflection. *Government Information Quarterly*, 29, 512-520.
43. Bhuiyan, S. H. (2011). Modernizing Bangladesh public administration through e-governance: Benefits and challenges. *Government Information Quarterly*, 28, 54-65.
44. Torres, L., Pina, V., and Acerete, B. (2005). E-government developments on delivering public services among EU cities. *Government Information Quarterly*, 22, 217-238.
45. Tung, L. L., & Rieck, O. (2005). Adoption of electronic government services among business organizations in Singapore. *Journal of Strategic Information Systems*, 14, 417-440.
46. Pina, V., Torres, L., and Acerete, B. (2007). Are ICTs promoting government accountability? A comparative analysis of e-governance developments in 19 OECD countries. *Critical Perspectives on Accounting*, 18, 583-602.
47. Brown, D. (2005). Electronic government and public administration. *International Review of Administrative Sciences*, Vol.71 (2), 241–254.
48. Ziemia, E. and Kolasa, I. (2014). The Survey of Information Systems in Public Administration in Poland. *Interdisciplinary Journal of Information, Knowledge, and Management*. 9. 10.28945/1939.
49. Ziyadin, S., Malayev, K., Fernández-Plazaola, I., Ismail, G., & Beyzhanova, A. (2020). Digital Modernization of the System of Public Administration: Prerogatives and Barriers, *E3S Web Conf. Volume 159, 05003*. <https://doi.org/10.1051/e3sconf/202015905003>
50. United Nations (2018). E-Government Survey 2018. United Nations Publications (2018). Available at: [https://www.unescap.org/sites/default/files/E-Government%20Survey%202018\\_FINAL.pdf](https://www.unescap.org/sites/default/files/E-Government%20Survey%202018_FINAL.pdf)
51. Dwivedi, Y.K. et al. (2014). Research on information systems failures and successes: Status update and future directions. *Information Systems Frontiers* 17(1), pp. 143-157.

52. Kobayashi, K. (2018). "A Study on the Causes of Information System Failure," 2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI), Yonago, Japan, 2018, pp. 762-764, doi: 10.1109/IIAI-AAI.2018.00157.
53. Pfleeger, S.L. (2003). Τεχνολογία Λογισμικού, Θεωρία και Πράξη. Εκδόσεις Κλειδάριθμος.

## Παράρτημα

---

### i) Αρχικό αρχείο arff

@RELATION Comments

@ATTRIBUTE article {1,2,3,4,5,6,7,8,9,11,12,13,14,15,17,18,19,20, 22, 23, 24, 25, 26, 27, 29, 30, 32}

@ATTRIBUTE commentator {individual, public service}

@ATTRIBUTE commentator gender {Male, Female, No Gender}

@ATTRIBUTE comment {Positive, Neutral, Negative, Question}

@ATTRIBUTE emotion {Hopeful, Excited, Pleading, Angry, Irritated, No Emotion, Worried}

@ATTRIBUTE recommendation {Yes, No}

@DATA

12, individual, Female, Neutral, Worried, Yes  
3, individual, Male, Negative, Irritated, Yes  
29, individual, Male, Negative, Angry, No  
30, individual, Male, Negative, Irritated, Yes  
8, individual, Male, Neutral, No Emotion, Yes  
2, individual, Male, Neutral, No Emotion, Yes  
19, individual, Male, Neutral, No Emotion, No  
1, individual, Male, Negative, Angry, No  
1, individual, Male, Negative, Angry, No  
1, individual, Male, Negative, Angry, No  
19, individual, Female, Negative, Irritated, Yes  
32, individual, Male, Neutral, Worried, Yes  
19, individual, Male, Negative, Irritated, No  
30, individual, Male, Neutral, No Emotion, Yes  
2, individual, Male, Neutral, No Emotion, Yes  
32, individual, Male, Negative, Worried, Yes  
3, individual, Female, Neutral, No Emotion, Yes  
19, individual, Female, Negative, No Emotion, Yes  
29, individual, Male, Negative, Irritated, Yes  
32, individual, Female, Negative, Irritated, Yes  
19, individual, Female, Negative, No Emotion, Yes  
3, individual, Female, Neutral, No Emotion, Yes  
30, individual, Female, Negative, Irritated, Yes  
3, public service, No Gender, Negative, Worried, Yes  
3, public service, No Gender, Negative, Worried, Yes  
5, public service, No Gender, Negative, No Emotion, Yes  
29, public service, No Gender, Negative, Worried, Yes  
19, individual, Female, Negative, Irritated, Yes  
3, individual, Female, Negative, Irritated, Yes

19, individual, Female, Negative, Worried, Yes  
29, individual, Female, Neutral, No Emotion, Yes  
1, individual, Male, Neutral, Pleading, Yes  
32, individual, Female, Neutral, No Emotion, Yes  
32, individual, Female, Negative, Irritated, Yes  
19, individual, Female, Negative, Irritated, Yes  
3, individual, Female, Negative, Irritated, Yes  
19, individual, Female, Negative, Irritated, Yes  
30, individual, Female, Negative, Irritated, Yes  
12, individual, Male, Negative, Irritated, Yes  
2, individual, Female, Neutral, Pleading, Yes  
2, individual, Female, Neutral, Pleading, Yes  
3, individual, Female, Neutral, No Emotion, Yes  
11, individual, Female, Negative, No Emotion, Yes  
14, individual, Female, Question, No Emotion, No  
15, individual, Female, Question, No Emotion, No  
17, individual, Female, Question, No Emotion, No  
19, individual, Female, Negative, No Emotion, Yes  
22, individual, Female, Neutral, No Emotion, Yes  
23, individual, Female, Neutral, No Emotion, Yes  
24, individual, Female, Neutral, No Emotion, Yes  
27, individual, Female, Question, No Emotion, No  
29, individual, Female, Neutral, No Emotion, No  
30, individual, Female, Negative, Worried, No  
19, individual, Female, Negative, Angry, Yes  
30, public service, No Gender, Neutral, No Emotion, Yes  
1, individual, Male, Negative, Worried, Yes  
2, individual, Male, Neutral, No Emotion, Yes  
3, individual, Male, Neutral, Pleading, Yes  
4, individual, Male, Neutral, Pleading, Yes  
5, individual, Male, Neutral, No Emotion, Yes  
6, individual, Male, Neutral, No Emotion, Yes  
7, individual, Male, Question, No Emotion, No  
11, individual, Male, Question, No Emotion, No  
12, individual, Male, Question, No Emotion, No  
20, individual, Male, Negative, Worried, No  
26, individual, Male, Neutral, No Emotion, Yes  
29, individual, Male, Neutral, No Emotion, Yes  
29, individual, Male, Neutral, No Emotion, Yes  
30, individual, Male, Negative, Irritated, Yes  
11, individual, Female, Neutral, No Emotion, Yes  
1, individual, Female, Question, No Emotion, Yes  
2, individual, Female, Neutral, No Emotion, Yes  
2, individual, Female, Neutral, No Emotion, Yes  
11, individual, Female, Neutral, No Emotion, Yes  
12, individual, Female, Neutral, No Emotion, Yes  
19, individual, Female, Negative, Irritated, Yes

23, individual, Female, Neutral, No Emotion, Yes  
32, individual, Male, Negative, Worried, Yes  
25, individual, Female, Neutral, No Emotion, Yes  
2, individual, Male, Neutral, No Emotion, Yes  
27, individual, Female, Neutral, No Emotion, Yes  
3, individual, Male, Negative, Worried, Yes  
29, individual, Female, Neutral, No Emotion, Yes  
30, individual, Female, Neutral, No Emotion, Yes  
30, individual, Female, Neutral, No Emotion, Yes  
19, individual, Male, Negative, Worried, Yes  
3, public service, No Gender, Neutral, No Emotion, Yes  
14, public service, No Gender, Neutral, No Emotion, Yes  
23, public service, No Gender, Neutral, No Emotion, Yes  
29, public service, No Gender, Neutral, No Emotion, Yes  
30, public service, No Gender, Neutral, No Emotion, Yes  
3, public service, No Gender, Negative, Worried, Yes  
29, public service, No Gender, Neutral, No Emotion, Yes  
1, public service, No Gender, Negative, Irritated, No  
1, public service, No Gender, Neutral, No Emotion, Yes  
3, public service, No Gender, Neutral, No Emotion, Yes  
27, public service, No Gender, Neutral, No Emotion, Yes  
2, public service, No Gender, Neutral, No Emotion, Yes  
3, public service, No Gender, Negative, Irritated, No  
29, public service, No Gender, Neutral, No Emotion, Yes  
4, public service, No Gender, Neutral, No Emotion, Yes  
3, public service, No Gender, Neutral, No Emotion, Yes  
4, public service, No Gender, Neutral, No Emotion, Yes  
5, public service, No Gender, Neutral, No Emotion, No  
11, public service, No Gender, Question, No Emotion, No  
29, individual, Female, Neutral, No Emotion, No  
9, public service, No Gender, Question, No Emotion, No  
17, public service, No Gender, Neutral, No Emotion, No  
19, public service, No Gender, Neutral, Worried, Yes  
19, public service, No Gender, Neutral, No Emotion, Yes  
27, public service, No Gender, Neutral, No Emotion, Yes  
29, public service, No Gender, Neutral, No Emotion, Yes  
30, public service, No Gender, Neutral, No Emotion, Yes  
32, public service, No Gender, Neutral, No Emotion, Yes  
3, individual, Female, Neutral, No Emotion, Yes  
4, individual, Female, Neutral, No Emotion, Yes  
27, individual, Female, Neutral, No Emotion, Yes  
2, public service, No Gender, Neutral, No Emotion, Yes  
3, public service, No Gender, Neutral, No Emotion, Yes  
3, public service, No Gender, Neutral, No Emotion, Yes  
4, public service, No Gender, Neutral, No Emotion, Yes  
5, public service, No Gender, Neutral, No Emotion, Yes  
11, public service, No Gender, Neutral, No Emotion, Yes

29, public service, No Gender, Neutral, No Emotion, Yes  
30, public service, No Gender, Neutral, No Emotion, Yes  
3, individual, Female, Question, No Emotion, No  
19, public service, No Gender, Question, No Emotion, No  
11, public service, No Gender, Neutral, No Emotion, Yes  
29, public service, No Gender, Neutral, Worried, No  
5, public service, No Gender, Neutral, Worried, No  
3, individual, Male, Neutral, No Emotion, Yes  
3, public service, No Gender, Neutral, Worried, Yes  
3, individual, Male, Neutral, Worried, Yes  
2, public service, No Gender, Neutral, No Emotion, Yes  
29, public service, No Gender, Neutral, No Emotion, Yes  
18, individual, Male, Neutral, No Emotion, Yes  
13, individual, Male, Neutral, Hopeful, Yes  
4, public service, No Gender, Neutral, No Emotion, Yes  
13, individual, Male, Neutral, No Emotion, Yes  
12, individual, Male, Neutral, No Emotion, Yes  
3, public service, No Gender, Negative, No Emotion, Yes  
29, public service, No Gender, Neutral, No Emotion, Yes  
3, public service, No Gender, Neutral, Worried, Yes  
3, public service, No Gender, Neutral, No Emotion, Yes  
25, individual, Male, Neutral, No Emotion, Yes  
19, public service, No Gender, Negative, Irritated, Yes  
3, public service, No Gender, Neutral, No Emotion, Yes  
19, individual, Male, Question, Worried, Yes  
23, individual, Male, Neutral, No Emotion, Yes  
29, public service, No Gender, Neutral, No Emotion, Yes  
1, individual, Male, Positive, Excited, No  
19, public service, No Gender, Neutral, Worried, Yes

**ii) Αρχείο arff μετά από αφαίρεση μετά από αφαίρεση των attributes Commentator και CommentatorGender:**

@RELATION Comments

@ATTRIBUTE article {1,2,3,4,5,6,7,8,9,11,12,13,14,15,17,18,19,20, 22, 23, 24, 25, 26, 27, 29, 30, 32}

@ATTRIBUTE comment {Positive, Neutral, Negative, Question}

@ATTRIBUTE emotion {Hopeful, Excited, Pleading, Angry, Irritated, NoEmotion, Worried}

@ATTRIBUTE recommendation {Yes, No}

@DATA

12, Neutral, Worried, Yes

3, Negative, Irritated, Yes

29, Negative, Angry, No

30, Negative, Irritated, Yes

8, Neutral, NoEmotion, Yes  
2, Neutral, NoEmotion, Yes  
19, Neutral, NoEmotion, No  
1, Negative, Angry, No  
1, Negative, Angry, No  
1, Negative, Angry, No  
19, Negative, Irritated, Yes  
32, Neutral, Worried, Yes  
19, Negative, Irritated, No  
30, Neutral, NoEmotion, Yes  
2, Neutral, NoEmotion, Yes  
32, Negative, Worried, Yes  
3, Neutral, NoEmotion, Yes  
19, Negative, NoEmotion, Yes  
29, Negative, Irritated, Yes  
32, Negative, Irritated, Yes  
19, Negative, NoEmotion, Yes  
3, Neutral, NoEmotion, Yes  
30, Negative, Irritated, Yes  
3, Negative, Worried, Yes  
3, Negative, Worried, Yes  
5, Negative, NoEmotion, Yes  
29, Negative, Worried, Yes  
19, Negative, Irritated, Yes  
3, Negative, Irritated, Yes  
19, Negative, Worried, Yes  
29, Neutral, NoEmotion, Yes  
1, Neutral, Pleading, Yes  
32, Neutral, NoEmotion, Yes  
32, Negative, Irritated, Yes  
19, Negative, Irritated, Yes  
3, Negative, Irritated, Yes  
19, Negative, Irritated, Yes  
30, Negative, Irritated, Yes  
12, Negative, Irritated, Yes  
2, Neutral, Pleading, Yes  
2, Neutral, Pleading, Yes  
3, Neutral, NoEmotion, Yes  
11, Negative, NoEmotion, Yes  
14, Question, NoEmotion, No  
15, Question, NoEmotion, No  
17, Question, NoEmotion, No  
19, Negative, NoEmotion, Yes  
22, Neutral, NoEmotion, Yes  
23, Neutral, NoEmotion, Yes  
24, Neutral, NoEmotion, Yes  
27, Question, NoEmotion, No



29, Neutral, NoEmotion, No  
30, Negative, Worried, No  
19, Negative, Angry, Yes  
30, Neutral, NoEmotion, Yes  
1, Negative, Worried, Yes  
2, Neutral, NoEmotion, Yes  
3, Neutral, Pleading, Yes  
4, Neutral, Pleading, Yes  
5, Neutral, NoEmotion, Yes  
6, Neutral, NoEmotion, Yes  
7, Question, NoEmotion, No  
11, Question, NoEmotion, No  
12, Question, NoEmotion, No  
20, Negative, Worried, No  
26, Neutral, NoEmotion, Yes  
29, Neutral, NoEmotion, Yes  
29, Neutral, NoEmotion, Yes  
30, Negative, Irritated, Yes  
11, Neutral, NoEmotion, Yes  
1, Question, NoEmotion, Yes  
2, Neutral, NoEmotion, Yes  
2, Neutral, NoEmotion, Yes  
11, Neutral, NoEmotion, Yes  
12, Neutral, NoEmotion, Yes  
19, Negative, Irritated, Yes  
23, Neutral, NoEmotion, Yes  
32, Negative, Worried, Yes  
25, Neutral, NoEmotion, Yes  
2, Neutral, NoEmotion, Yes  
27, Neutral, NoEmotion, Yes  
3, Negative, Worried, Yes  
29, Neutral, NoEmotion, Yes  
30, Neutral, NoEmotion, Yes  
30, Neutral, NoEmotion, Yes  
19, Negative, Worried, Yes  
3, Neutral, NoEmotion, Yes  
14, Neutral, NoEmotion, Yes  
23, Neutral, NoEmotion, Yes  
29, Neutral, NoEmotion, Yes  
30, Neutral, NoEmotion, Yes  
3, Negative, Worried, Yes  
29, Neutral, NoEmotion, Yes  
1, Negative, Irritated, No  
1, Neutral, NoEmotion, Yes  
3, Neutral, NoEmotion, Yes  
27, Neutral, NoEmotion, Yes  
2, Neutral, NoEmotion, Yes

3, Negative, Irritated, No  
29, Neutral, NoEmotion, Yes  
4, Neutral, NoEmotion, Yes  
3, Neutral, NoEmotion, Yes  
4, Neutral, NoEmotion, Yes  
5, Neutral, NoEmotion, No  
11, Question, NoEmotion, No  
29, Neutral, NoEmotion, No  
9, Question, NoEmotion, No  
17, Neutral, NoEmotion, No  
19, Neutral, Worried, Yes  
19, Neutral, NoEmotion, Yes  
27, Neutral, NoEmotion, Yes  
29, Neutral, NoEmotion, Yes  
30, Neutral, NoEmotion, Yes  
32, Neutral, NoEmotion, Yes  
3, Neutral, NoEmotion, Yes  
4, Neutral, NoEmotion, Yes  
27, Neutral, NoEmotion, Yes  
2, Neutral, NoEmotion, Yes  
3, Neutral, NoEmotion, Yes  
3, Neutral, NoEmotion, Yes  
4, Neutral, NoEmotion, Yes  
5, Neutral, NoEmotion, Yes  
11, Neutral, NoEmotion, Yes  
29, Neutral, NoEmotion, Yes  
30, Neutral, NoEmotion, Yes  
3, Question, NoEmotion, No  
19, Question, NoEmotion, No  
11, Neutral, NoEmotion, Yes  
29, Neutral, Worried, No  
5, Neutral, Worried, No  
3, Neutral, NoEmotion, Yes  
3, Neutral, Worried, Yes  
3, Neutral, Worried, Yes  
2, Neutral, NoEmotion, Yes  
29, Neutral, NoEmotion, Yes  
18, Neutral, NoEmotion, Yes  
13, Neutral, Hopeful, Yes  
4, Neutral, NoEmotion, Yes  
13, Neutral, NoEmotion, Yes  
12, Neutral, NoEmotion, Yes  
3, Negative, NoEmotion, Yes  
29, Neutral, NoEmotion, Yes  
3, Neutral, Worried, Yes  
3, Neutral, NoEmotion, Yes  
25, Neutral, NoEmotion, Yes

19, Negative, Irritated, Yes  
3, Neutral, NoEmotion, Yes  
19, Question, Worried, Yes  
23, Neutral, NoEmotion, Yes  
29, Neutral, NoEmotion, Yes  
1, Positive, Excited, No  
19, Neutral, Worried, Yes