



ΔΙΕΘΝΕΣ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΤΗΣ ΕΛΛΑΔΟΣ

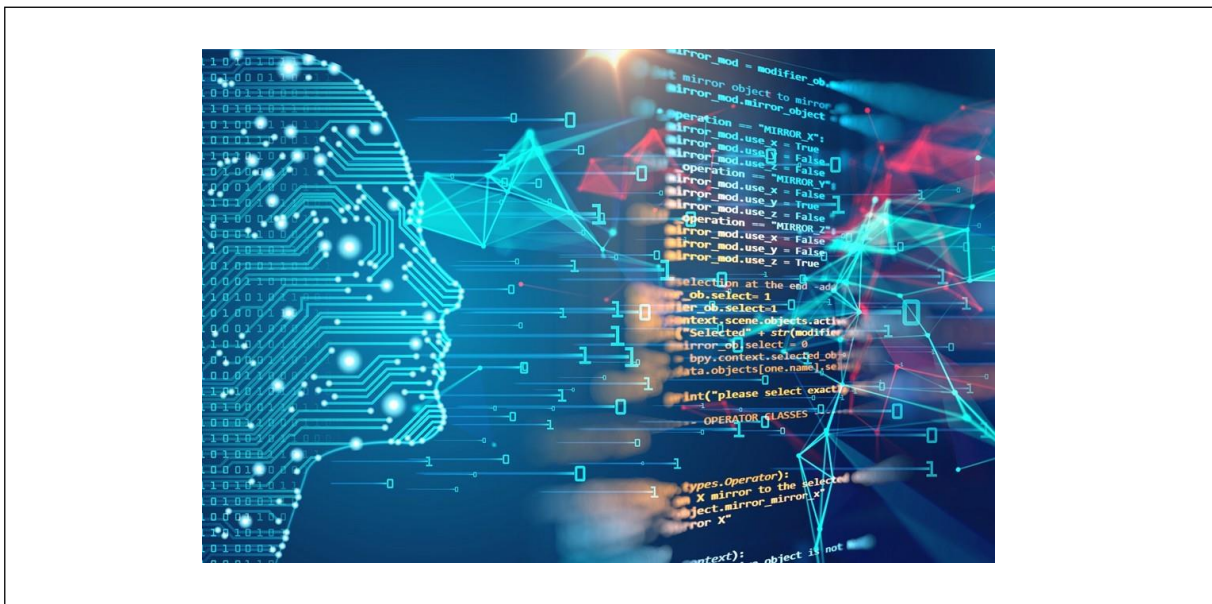
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ  
ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

---

«ΣΥΝΑΙΣΘΗΜΑΤΙΚΗ ΑΝΑΛΥΣΗ ΗΧΗΤΙΚΩΝ  
ΣΗΜΑΤΩΝ ΠΟΥ ΠΡΟΕΡΧΟΝΤΑΙ ΑΠΟ  
ΗΛΕΚΤΡΟΝΙΚΕΣ ΣΥΣΚΕΥΕΣ»



Των φοιτητών  
Ευαγγελίδου Νίκη  
Αρ. Μητρώου: it164660  
Τσαλίκη Ευστράτιου  
Αρ. Μητρώου: it164760

Επιβλέπων  
Κωτσάκης Ρήγας  
Επίκουρος Καθηγητής

**Ημερομηνία 20/01/2023**

Συναισθηματική Ανάλυση Ηχητικών Σημάτων που προέρχονται από Ηλεκτρονικές Συσκευές

Κωδικός Δ.Ε 21348

Ονοματεπώνυμο φοιτητών Ευαγγελίδου Νίκη, Τσαλίκης Ευστράτιος

Ονοματεπώνυμο εισηγητή Κωτσάκης Ρήγας

Ημερομηνία ανάληψης Δ.Ε. 15/10/2021

Ημερομηνία περάτωσης Δ.Ε. 16/1/2023

*Βεβαιώνω ότι είμαι ο συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως πτυχιακή εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.*

*Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία των φοιτητών Ευαγγελίδου Νίκης και Τσαλίκη Ευστράτιου που την εκπόνησαν. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης, ο συγγραφέας/δημιουργός εκχωρεί στο Διεθνές Πανεπιστήμιο της Ελλάδος άδεια χρήσης του δικαιώματος αναπαραγωγής, δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσης της εργασίας διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος. Η ανοικτή πρόσβαση στο πλήρες κείμενο της εργασίας, δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού, ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, πώληση, εμπορική χρήση, διανομή, έκδοση, μεταφόρτωση (downloading), ανάρτηση (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού.*

Η έγκριση της πτυχιακής εργασίας από το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος, δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα, εκ μέρους του Τμήματος.

*«Στην οικογένεια και στους φίλους μας»*

*«Η διδασκαλία είναι κάτι περισσότερο από το να διανέμεις τη γνώση, είναι να εμπνέεις την αλλαγή. Η μάθηση είναι κάτι περισσότερο από το να αποστηθίζεις δεδομένα, είναι να φτάνεις στην κατανόηση.»*

**William Arthur Ward**

## Πρόλογος

Η συγκεκριμένη πτυχιακή εργασία επιλέχθηκε από εμάς με βασικό γνώμονα το θέμα το οποίο επιλέξαμε να άπτεται των ενδιαφερόντων μας και να μπορέσουμε να προσθέσουμε και εμείς ένα λιθαράκι στο συγκεκριμένο γνωστικό αντικείμενο.

Εν συνεχεία και έπειτα από μελέτη της διαθέσιμης βιβλιογραφίας μπροστά μας εμφανίστηκε ένας νέος κόσμος καταρρίπτοντας πολλά από όσα γνωρίζαμε. Αναλάβαμε την πτυχιακή μας εργασία με βεβαιότητα ότι κατείχαμε το μεγαλύτερο κομμάτι του αντικειμένου ωστόσο μετά από αρκετή ενασχόληση το συγκεκριμένο θέμα μας εξέπληξε ευχάριστα καθώς διαπιστώσαμε το φάσμα του.

Στην απόφασή μας αυτή βεβαίως συνέβαλλε και το γεγονός ότι ο κ. Κωτσάκης ο οποίος είναι και ο επιβλέπων καθηγητής μας είναι ένας άνθρωπος με πολλές γνώσεις στον τομέα του Ήχου. Μας βοήθησε να «ξεκλειδωθούμε» εμείς οι ίδιοι αλλά και κατάφερε μέσα σε μικρό χρονικό διάστημα να μας μεταδώσει την αγάπη και το μεράκι που τον διακατέχει όσον αφορά την δουλειά του.

Τέλος, θεωρούμε πως το πιο σημαντικό για εμάς, μετά τις γνώσεις που αποκομίσαμε, ήταν η δύναμη της ομαδικότητας καθώς αντιληφθήκαμε σε μεγάλο βαθμό ότι μέσα από την μεταξύ μας συνεργασία ο καθένας από εμάς κατάφερε να αναπτύξει πολύτιμα εφόδια και γνώσεις που είμαστε σίγουροι ότι θα μας φανούν πολύτιμα στο μέλλον.

## Περίληψη

Η συγκεκριμένη πτυχιακή εργασία έχει ως θέμα την κατηγοριοποίηση των ήχων με βάση το συναίσθημα. Στο θεωρητικό μέρος, γίνεται μία συνοπτική αναφορά στην εξέλιξη της επικοινωνίας ανθρώπου υπολογιστή καθώς και στην δυσκολία της κατηγοριοποίησης των συναισθημάτων. Εν συνεχεία, παρουσιάζεται η έννοια του ήχου και περιγράφονται τα χαρακτηριστικά που τον διέπουν ενώ γίνεται μία καταγραφή της υπάρχουσας κατάστασης στην επιστημονική κοινότητα σε ότι αφορά το θέμα.

Ένα από τα πιο σημαντικά σημεία που αναδείχθηκαν είναι η χρησιμότητα της κατηγοριοποίησης του ήχου με βάση το συναίσθημα στην καθημερινή μας ζωή καθώς μέχρι στιγμής οι περισσότερες έρευνες οι οποίες έχουν δημοσιευθεί επικεντρώθηκαν στο συναίσθημα που προκαλείται από τον λόγο και την μουσική σε αντίθεση με τον μικρό αριθμό ερευνών επάνω στο αντικείμενο που μελετάται.

Κατά το πειραματικό μέρος της εργασίας μας, έγινε προσπάθεια συλλογής και επεξεργασίας ήχων προκειμένου να κατασκευαστεί μια βάση δεδομένων. Επιπλέον, δημιουργήθηκε ένα σύστημα μέσω του οποίου έγινε η εξαγωγή ηχητικών χαρακτηριστικών με χρήση του Matlab βάση των οποίων έγινε και η κατηγοριοποίηση των ήχων σε συνάρτηση με το συναίσθημα χρησιμοποιώντας κατάλληλους αλγόριθμους ταξινόμησης στο Matlab και στο WEKA. Τα χαρακτηριστικά αξιολογήθηκαν με χρήση του WEKA και επιβεβαιώθηκαν μέσα από διαγράμματα στο Excel.

Τέλος, αξίζει να αναφερθεί πως η συγκεκριμένη μελέτη παρουσιάζει αυξημένο ερευνητικό ενδιαφέρον, ενώ στην δική μας εργασία παρουσιάζονται πρώιμα πειράματα καθώς και μοντέλα ταξινόμησης.

EVANGELIDOU NIKI- TSALIKIS EFSTRATIOS

## **Abstract**

The subject of this thesis is the categorization of sounds based on emotion. In the theoretical part, a brief reference is made to the evolution of human-computer communication as well as to the difficulty of categorizing emotions. Then, the concept of sound is presented and the characteristics that govern it are described, while a record is made of the current situation in the scientific community regarding the subject.

One of the most important points that emerged is the usefulness of categorizing sound based on emotion in our daily lives as so far most of the research that has been published has focused on the emotion evoked by speech and music as opposed to the small number of research on the subject being studied.

During the experimental part of our work, an attempt was made to collect and process sounds to build a database. In addition, a system was created through which sound characteristics were extracted using Matlab, based on which the sounds were categorized in relation to emotion using appropriate classification algorithms in Matlab and WEKA. Characteristics were evaluated using WEKA and confirmed through diagrams in Excel.

Finally, it is worth mentioning that this study is of increased research interest, while in our thesis early experiments as well as classification models are presented.

# Ευχαριστίες

Ξεκινώντας νιώθουμε την ανάγκη να ευχαριστήσουμε όλους τους ανθρώπους, που ο καθένας με τον δικό του μοναδικό τρόπο συντέλεσαν στην περάτωση των προπτυχιακών μας σπουδών.

Ευχαριστούμε τον επιβλέποντα καθηγητή μας κ. Κωτσάκη Ρήγα, για την επίβλεψη, την καθοδήγηση καθώς και την αμέριστη βοήθεια που μας προσέφερε απλόχερα όποτε την χρειαζόμασταν. Τόσο οι υποδείξεις του όσο και τα σχόλια του αποτέλεσαν πολύτιμα εφόδια στην διεκπεραίωση αυτής της εργασίας.

Ευχαριστούμε όλους τους καθηγητές μας στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος που χωρίς την δική τους βοήθεια δεν θα φτάναμε ως εδώ.

Πιο πολύ ωστόσο οφείλουμε να ευχαριστήσουμε ο καθένας ξεχωριστά την οικογένειές μας που αποτέλεσαν όλα αυτά τα χρόνια ένα στήριγμα για εμάς.

# Περιεχόμενα

|  |      |
|--|------|
| Πρόλογος.....  | iv   |
| Περίληψη.....  | v    |
| Abstract .....   | vi   |
| Ευχαριστίες .....  | vii  |
| Περιεχόμενα .....  | viii |
| Κατάλογος Εικόνων .....  | xi   |
| Συντομογραφίες.....  | xiii |
| Εισαγωγή.....  | xiv  |
| ΘΕΩΡΗΤΙΚΟ ΜΕΡΟΣ.....   | xv   |
| Κεφάλαιο 1 <sup>ο</sup> Επικοινωνία Ανθρώπου-Υπολογιστή.....             | 1    |
| 1.1 Εισαγωγή.....  | 1    |
| 1.2 Ορισμός.....   | 1    |
| 1.3 Ιστορική αναδρομή.....   | 2    |
| 1.4 Αντίκτυπο του HCI στην κοινωνία και την οικονομία.....               | 4    |
| Κεφάλαιο 2 <sup>ο</sup> : Αναγνώριση Συναισθήματος .....                 | 6    |
| 2.1 Εισαγωγή.....  | 6    |
| 2.2 Ο όρος συναίσθημα .....  | 6    |
| 2.3 Κατηγοριοποίηση Συναισθημάτων .....                                  | 6    |
| 2.4 Συναισθηματική Ανάλυση κατά την Αλληλεπίδραση Ανθρώπου-Μηχανής ..... | 8    |
| Κεφάλαιο 3 <sup>ο</sup> : Ήχος.....                                      | 9    |
| 3.1 Εισαγωγή.....  | 9    |
| 3.2 Περιγραφή ακουστού συστήματος.....                                   | 9    |
| 3.3 Η έννοια του ήχου .....  | 10   |
| 3.4 Ψηφιοποίηση του Ήχου.....  | 11   |
| 3.5 Νευρωνικά Δίκτυα .....   | 13   |
| 3.6 Χαρακτηριστικοί εξαγωγείς.....                                       | 14   |
| 3.6.1 Δυναμικά Χαρακτηριστικά .....                                      | 14   |
| 3.6.2 Χαρακτηριστικά Τέμπο .....   | 14   |
| 3.7 Υψηλού επιπέδου χαρακτηριστικά ήχων .....                            | 14   |
| 3.7.1 Δομή και μορφή .....   | 14   |
| 3.7.2 Στατιστικά Χαρακτηριστικά.....                                     | 15   |
| 3.7.3 Τεχνικές προβλέψεις .....  | 16   |
| 3.7.4 Εξαγωγή .....  | 16   |

|  |   |    |
|--|---|----|
| 3.8  | Επίλογος.....   | 17 |
| Κεφάλαιο 4° : Παρουσίαση του Προβλήματος.....                  |   | 18 |
| 4.1  | Εισαγωγή.....   | 18 |
| 4.2  | Το πρόβλημα της κατηγοριοποίησης.....                   | 18 |
| 4.3  | Βιβλιογραφική ανασκόπηση.....                           | 18 |
| 4.4  | Επίλογος.....   | 21 |
| ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ.....   |   | 22 |
| Κεφάλαιο 5° :Συλλογή και προεπεξεργασία δεδομένων.....         |   | 23 |
| 5.1  | Εισαγωγή.....   | 23 |
| 5.2  | Συλλογή και προεπεξεργασία δεδομένων.....               | 23 |
| 5.3  | Ηχητικά Χαρακτηριστικά.....                             | 24 |
| 5.4  | Κατάτμηση και εξαγωγή χαρακτηριστικών.....              | 30 |
| 5.5  | Αξιολόγηση τιμών.....                                   | 31 |
| 5.6  | Επίλογος.....   | 34 |
| Κεφάλαιο 6° : Πειράματα μηχανικής μάθησης που εκπονήθηκαν..... |   | 35 |
| 6.1  | Εισαγωγή.....   | 35 |
| 6.2  | Matlab.....   | 35 |
| 6.3  | WEKA.....   | 36 |
| 6.4  | Τρόποι επικύρωσης (validation).....                     | 37 |
| 6.5  | Αλγόριθμοι εκμάθησης μοντέλου.....                      | 38 |
| 6.6  | Επίλογος.....   | 40 |
| Κεφάλαιο 7°: Αποτελέσματα πειραμάτων.....                      |   | 41 |
| 7.1  | Εισαγωγή.....   | 41 |
| 7.2  | Μετρικές Αξιολόγησης.....                               | 41 |
| 7.3  | Μήτρα σύγχυσης (confusion matrix).....                  | 43 |
| 7.4  | Αποτελέσματα Αλγορίθμων.....                            | 44 |
| 7.5  | Σχολιασμός αποδόσεων αλγορίθμων και σφαλμάτων τους..... | 61 |
| Κεφάλαιο 8°:Αξιολόγηση Χαρακτηριστικών.....                    |   | 63 |
| 8.1  | Εισαγωγή.....   | 63 |
| 8.2  | Αξιολόγηση χαρακτηριστικών με χρήση WEKA.....           | 63 |
| 8.3  | Αποτελέσματα αξιολόγησης χαρακτηριστικών.....           | 64 |
| 8.4  | Συμπεράσματα.....                                       | 65 |
| Κεφάλαιο 9°: Συμπεράσματα και μελλοντικές κατευθύνσεις.....    |   | 66 |
| 9.1  | Συμπεράσματα.....                                       | 66 |
| 9.2  | Μελλοντικές Κατευθύνσεις.....                           | 66 |



# Κατάλογος Εικόνων

|   |    |
|---|----|
| Εικόνα 1.1: Συνεισφορά του HCI σε διάφορους τομείς .....  | 1  |
| Εικόνα 1.2: Περιπατητική Προσβασιμότητα.....  | 5  |
| Εικόνα 2.1: Συναισθηματικός Χάρτης.....   | 7  |
| Εικόνα 3.1: Ακουστική αντίληψη.....   | 10 |
| Εικόνα 3.2: Μία ημιτονοειδής κυματομορφή.....   | 10 |
| Εικόνα 3.3: Δείγμα μετρήσεων σε τακτά χρονικά διαστήματα.....   | 12 |
| Εικόνα 3.4: Τεχνητά νευρωνικά δίκτυα .....  | 13 |
| Εικόνα 3.5: MIRtoolbox Feature.....   | 17 |
| Εικόνα 4.1: Οι σχολιασμοί των ήχων ανά ομάδα και στο σύνολο. ....   | 19 |
| Εικόνα 4.2: Ακρίβεια ταξινόμησης του τεστ ακρόασης με την εκδήλωση FMC των ήχων.....  | 20 |
| Εικόνα 4.3: Αποτελέσματα πειραμάτων με χρήση Μηχανικής Μάθησης.....   | 20 |
| Εικόνα 4.4: Precision και recall του προτεινόμενου αλγόριθμου .....   | 21 |
| Εικόνα 5.1: Πως φτάσαμε στο mfcc.....   | 25 |
| Εικόνα 5.2: Ρυθμός μηδενικού επιπέδου. Η πορτοκαλί γραμμή είναι το πλάτος και η μπλε γραμμή το ηχητικό σήμα.....  | 25 |
| Εικόνα 5.3: Το 85% της ενέργειας είναι συγκεντρωμένο κάτω από την συχνότητα 5640.53Hz .....   | 26 |
| Εικόνα 5.4: Μέση τιμή φάσματος .....  | 26 |
| Εικόνα 5.5: Τυπική Απόκλιση Σήματος.....  | 27 |
| Εικόνα 5.6: Στο πρώτο σχήμα οι τιμές είναι μαζεμένες στα αριστερά του μέσου όρου ενώ στο δεύτερο στα δεξιά .....  | 27 |
| Εικόνα 5.7: Χαμηλή, κανονική και υψηλή κύρτωση.....   | 28 |
| Εικόνα 5.8: Τα περισσότερα από τα πλαίσια είναι κάτω από το μέσο όρο RMS.....   | 28 |
| Εικόνα 5.9: Το 53,96 % της ενέργειας είναι συγκεντρωμένο πάνω από την συχνότητα 1500Hz.....   | 29 |
| Εικόνα 5.10: Σύγκριση μεταξύ των βασικών εννοιών των συναισθημάτων(happy,sad,tender,anger,fear) και συναισθηματικών διαστάσεων (activity,valence,tension) ..... | 29 |
| Εικόνα 5.11: Κώδικας που εισήχθη στο Matlab .....   | 30 |
| Εικόνα 5.12: Στιγμιότυπο αρχείου xls όπου αποτελεί την βάση αληθείας.....   | 31 |
| Εικόνα 5.13: Διαγράμματα χαρακτηριστικού MFCC για διαφορετικούς συντελεστές .....   | 32 |
| Εικόνα 5.14: Διαγράμματα χαρακτηριστικού Rolloff.....   | 32 |
| Εικόνα 5.15: Διαγράμματα χαρακτηριστικού Brightness .....   | 33 |
| Εικόνα 5.16: Διαγράμματα χαρακτηριστικών συναισθήματος.....   | 33 |
| Εικόνα 6.1: Classification Learner .....  | 35 |
| Εικόνα 6.2: Weka Explorer[49] .....   | 36 |
| Εικόνα 7.1: Παράδειγμα μήτρας σύγχυσης.....   | 43 |
| Εικόνα 7.2: Παράθυρο 3 δευτερολέπτων, 5 Cross Validation.....   | 44 |
| Εικόνα 7.3: Παράθυρο 3 δευτερολέπτων, 10 Cross Validation .....   | 45 |
| Εικόνα 7.4: Παράθυρο 3 δευτερολέπτων, 30 Holdout Validation .....   | 46 |
| Εικόνα 7.5: Παράθυρο 3 δευτερολέπτων, 40 Holdout Validation .....   | 47 |
| Εικόνα 7.6: Παράθυρο 1 δευτερολέπτου, 5 Cross Validation.....   | 48 |
| Εικόνα 7.7: Παράθυρο 1 δευτερολέπτου, 10 Cross Validation.....  | 49 |
| Εικόνα 7.8: Παράθυρο 1 δευτερολέπτου, 30 Holdout Validation.....  | 50 |
| Εικόνα 7.9 : Παράθυρο 1 δευτερολέπτου, 40 Holdout Validation.....   | 51 |
| Εικόνα 7.10: Παράθυρο 3 δευτερολέπτων, 5 Cross Validation.....  | 52 |

|  |    |
|--|----|
| <i>Εικόνα 7.11: Παράθυρο 3 δευτερολέπτων, 10 Cross Validation</i> .....                      | 53 |
| <i>Εικόνα 7.12: Παράθυρο 3 δευτερολέπτων, 30 Holdout</i> .....                               | 53 |
| <i>Εικόνα 7.13: Παράθυρο 3 δευτερολέπτων, 40 Holdout</i> .....                               | 54 |
| <i>Εικόνα 7.14: Παράθυρο 1 δευτερολέπτου, 5 Cross Validation</i> .....                       | 54 |
| <i>Εικόνα 7.15: Παράθυρο 1 δευτερολέπτου, 10 Cross Validation</i> .....                      | 55 |
| <i>Εικόνα 7.16: Παράθυρο 1 δευτερολέπτου, 30 Holdout</i> .....                               | 55 |
| <i>Εικόνα 7.17: Παράθυρο 1 δευτερολέπτου, 40 Holdout</i> .....                               | 56 |
| <i>Εικόνα 7.18: Ποσοστό σφαλμάτων σε παράθυρο 3 δευτερολέπτων, 5 Cross Validation</i> .....  | 56 |
| <i>Εικόνα 7.19: Ποσοστό σφαλμάτων σε παράθυρο 3 δευτερολέπτων, 10 Cross Validation</i> ..... | 57 |
| <i>Εικόνα 7.20: Ποσοστό σφαλμάτων σε παράθυρο 3 δευτερολέπτων, 30 Holdout</i> .....          | 57 |
| <i>Εικόνα 7.21: Ποσοστό σφαλμάτων σε παράθυρο 3 δευτερολέπτων, 40 Holdout</i> .....          | 58 |
| <i>Εικόνα 7.22: Ποσοστό σφαλμάτων σε παράθυρο 1 δευτερολέπτου, 5 Cross Validation</i> .....  | 58 |
| <i>Εικόνα 7.23: Ποσοστό σφαλμάτων σε παράθυρο 1 δευτερολέπτου, 10 Cross Validation</i> ..... | 59 |
| <i>Εικόνα 7.24: Ποσοστό σφαλμάτων σε παράθυρο 1 δευτερολέπτου, 30 Holdout</i> .....          | 59 |
| <i>Εικόνα 7.25: Ποσοστό σφαλμάτων σε παράθυρο 1 δευτερολέπτου, 40 Holdout</i> .....          | 60 |
| <i>Εικόνα 8.1: Κατάταξη αξιολόγησης χαρακτηριστικών σε όλα τα παράθυρα</i> .....             | 64 |

## Συντομογραφίες

|       |   |
|-------|---|
| HCI   | Human-Computer Interaction (Αλληλεπίδραση Ανθρώπου- Υπολογιστή) |
| ΑΑΥ   | Αλληλεπίδραση Ανθρώπου- Υπολογιστή                              |
| Π.Ε.  | Πτυχιακή Εργασία  |
| ΔΙΠΑΕ | Διεθνές Πανεπιστήμιο Ελλάδος                                    |
| MFCC  | Mel-frequency cepstrum  |

# Εισαγωγή

Σκοπός αυτής της πτυχιακής εργασίας είναι η κατηγοριοποίηση των ήχων με βάση τα συναισθήματα που προκαλούν στον χρήστη και η εξαγωγή ενός ασφαλούς συμπεράσματος αν οι συγκεκριμένοι ήχοι επιτυγχάνουν τον σκοπό για τον οποίο δημιουργήθηκαν, να προκαλούν δηλαδή στον χρήστη τα κατάλληλα συναισθήματα.

Σε γενικές γραμμές η ταξινόμηση ήχου έχει πολλές δυνατότητες και βρίσκει εφαρμογή σε πολλούς τομείς. Στην δική μας έρευνα στοχεύσαμε στην κατηγοριοποίηση των ήχων με βάση τα συναισθήματα που προκαλεί στον χρήστη σε τρεις κατηγορίες.

Στο πρώτο κεφάλαιο δίνεται ο ορισμός της Αλληλεπίδρασης Ανθρώπου Μηχανής και γίνεται μία συνοπτική ιστορική αναδρομή. Επιπλέον, αναφερόμαστε στον αντίκτυπο που έχει αυτή η επιστήμη στην κοινωνία και την οικονομία γενικότερα μέσω ενός παραδείγματος εφαρμογής η οποία βελτιώνει την καθημερινότητα ανθρώπων με οπτική αναπηρία.

Στο δεύτερο κεφάλαιο δίνεται ο ορισμός του συναισθήματος και αναφέρονται οι κατηγορίες που διακρίνονται τα συναισθήματα. Στη συνέχεια αναλύεται το συναίσθημα κατά την αλληλεπίδραση ανθρώπου μηχανής.

Στο τρίτο κεφάλαιο αναφερόμαστε στην έννοια του ήχου δίνοντας μία συνοπτική περιγραφή του ανθρώπινου ακουστού συστήματος ενώ ταυτόχρονα αναφέρονται τα χαρακτηριστικά του ήχου και διακρίνονται στις βασικές κατηγορίες τους.

Στο τέταρτο κεφάλαιο περιγράφεται το πρόβλημα που προκύπτει κατά την διαδικασία της κατηγοριοποίησης και γίνεται αναφορά σε μία έρευνα η οποία κατά την άποψη μας σχετίζεται άμεσα με το θέμα μας.

Στο πέμπτο κεφάλαιο περιγράφεται η διαδικασία της έρευνας που ακολουθήθηκε από εμάς ξεκινώντας με την συλλογή και προεπεξεργασία δεδομένων καθώς και μία συνοπτική αναφορά των ηχητικών χαρακτηριστικών που χρησιμοποιήθηκαν

## **ΘΕΩΡΗΤΙΚΟ ΜΕΡΟΣ**

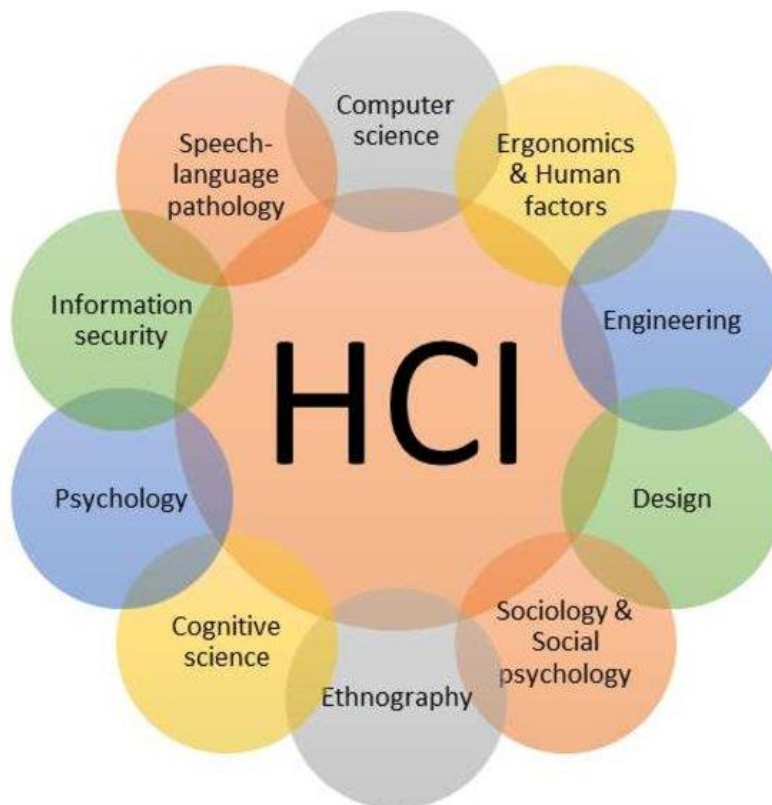
# Κεφάλαιο 1<sup>ο</sup> Επικοινωνία Ανθρώπου-Υπολογιστή

## 1.1 Εισαγωγή

Στο πρώτο κεφάλαιο δίνεται ο ορισμός της επιστήμης της επικοινωνίας ανθρώπου μηχανής και παρατίθεται μία σύντομη ιστορική αναδρομή. Εν συνεχεία, παρουσιάζεται μία καινοτόμα εφαρμογή η οποία συμβάλει άμεσα στην βελτίωση της ποιότητας ζωής των ανθρώπων μέσω της αλληλεπίδρασης τους με κάποια συσκευή.

## 1.2 Ορισμός

Η αλληλεπίδραση ανθρώπου υπολογιστή είναι μία μελέτη η οποία παρουσιάζει τον τρόπο με τον οποίο οι άνθρωποι σχεδιάζουν, υλοποιούν αλλά και χρησιμοποιούν συστήματα υπολογιστών τα οποία είναι διαδραστικά. Αυτή η αλληλεπίδραση εκτυλίσσεται στην επιφάνεια εργασίας του χρήστη και αποτυπώνεται από την συνεργασία του υλικού εξοπλισμού και του λογισμικού. Επιπλέον στην συγκεκριμένη μελέτη παρουσιάζεται και η επιρροή που ασκεί ο υπολογιστής ή κάποια άλλη ηλεκτρονική συσκευή στους ανθρώπους, στους οργανισμούς και γενικότερα σε ολόκληρη την κοινωνία[1].



Εικόνα 1.1: Συνεισφορά του HCI σε διάφορους τομείς

### 1.3 Ιστορική αναδρομή

Πριν την είσοδο των προσωπικών υπολογιστών, αρμόδιοι για τον χειρισμό τους μπορούσαν να είναι μόνο εξειδικευμένοι χρήστες οι οποίοι κατανοούσαν την κατασκευή και την δομή τους. Ωστόσο, μετά το τεχνολογικό άλμα κάτι τέτοιο δεν μπορούσε να συνεχίσει να υφίσταται μιας και πλέον το μεγαλύτερο μέρος του πληθυσμού είναι κάτοχοι μιας φορητής ηλεκτρονικής συσκευής όπως λάπτοπ, κινητά, τάμπλετ και άλλα. Για αυτόν ακριβώς τον λόγο σύμφωνα με το βιβλίο του Don Norman «προϊόντα ή συστήματα αποτύγχαναν εμπορικά ή δεν γίνονταν αποδεκτά στο χώρο εργασίας» [2].

Η πρώτη προσπάθεια για την αρχή του HCI (Human Computer Interaction) ξεκίνησε από τον Vannevar Bush, ο οποίος έπαιξε καθοριστικό ρόλο στη μυστική ανάπτυξη του υπολογιστή στην περίοδο του πολέμου[3]. Εκτός από την συνεισφορά του για την πρακτική ανάπτυξη του ψηφιακού υπολογιστή, η πιο μακροχρόνια επίδραση που άσκησε ήρθε με την μορφή ενός δοκιμίου με τίτλο «As We May Think» [4]. Στόχος του ήταν η απάντηση στο ερώτημα, πώς μπορεί κάποιος να μείνει ενημερωμένος σχετικά με την πληθώρα πληροφοριών που συμβαίνουν σε όλον τον κόσμο. Στο συγκεκριμένο δοκίμιο ο Bush περιέγραφε το όραμα του για το Memex, έναν πρωτοποριακό προσωπικό σταθμό εργασίας που θα έδινε την ευκαιρία σε έμπειρους επαγγελματίες να οργανώνουν καλύτερα, να καταλογίζουν, να ανακαλούν και να μοιράζονται τη δουλειά τους μέσω ενός ευρετηριασμένου και συνδεδεμένου συστήματος μικροφίλμ. Δυστυχώς ο Bush δεν κατάφερε να ολοκληρώσει το έργο του μιας και το Memex αποτέλεσε θύμα του ακριβούς προβλήματος που σχεδιάστηκε να λύσει το σύστημα.

Ο JCR Licklider ήταν μία ακόμα προσωπικότητα στην απαρχή του HCI. Στόχος του Licklider ήταν η δημιουργία ενός υπολογιστή ο οποίος θα ήταν σε θέση να ελέγξει σφάλματα καθοδηγώντας ταυτόχρονα τον άνθρωπο με τέτοιο τρόπο ώστε να επιτευχθεί η λύση του προβλήματος. Προκειμένου να επιτύχει τον σκοπό του, ο Licklider περιέγραψε διάφορους τομείς που απαιτούσαν βελτίωση όπως η ταχύτητα επεξεργασίας, το υλικό, η οργάνωση της μνήμης καθώς και γλωσσικά προβλήματα που στεκόταν εμπόδιο στην επικοινωνία ανθρώπου-υπολογιστή όπως και η εφεύρεση ενός εύχρηστου εξοπλισμού εισόδου και εξόδου[5].

Οι επόμενες δύο δεκαετίες που ακολούθησαν ήταν πολύ σημαντικές για την πρόοδο του HCI. Το 1963 ο Ivan Sutherland ανέπτυξε μία έρευνα η οποία οδήγησε στο πρόγραμμα Sketchpad, το οποίο ονομάστηκε από τον δημιουργό του ως «Ένα γραφικό σύστημα επικοινωνίας ανθρώπου-μηχανής» [6]. Ωστόσο, το σύστημα ήταν δύσκολο στη χρήση με αποτέλεσμα τη μη ευρεία χρησιμοποίησή του. Εντούτοις, είχε τεράστιο αντίκτυπο στο πεδίο. Επιπλέον, το 1965, ο Sutherland μέσα από το δοκίμιο του με τίτλο *The ultimate display* [7] περιέγραψε τους σύγχρονους περιορισμούς των συστημάτων εισόδου και εξόδου κάνοντας ταυτόχρονα πρόταση για μία υποθετική οθόνη η οποία θα μπορούσε να εκμεταλλευτεί πλήρως τις δυνατότητες που προσφέρει ένας υπολογιστής. Αργότερα, ο Sutherland δημιούργησε αυτό που θεωρείται ως το πρώτο σύστημα εικονικής πραγματικότητας το οποίο δεν ήταν άλλο παρά μία οθόνη τοποθετημένη στο κεφάλι που ονομάζεται «The Sword of Damocles». Ο Sutherland έπαιξε επίσης καθοριστικό ρόλο για την εξασφάλιση της χρηματοδότησης του Engelbart ο οποίος μέσα από την εργασία του «Augmenting Human Intellect: A Conceptual Framework» επέφερε πρόοδο στην αναζωογόνηση των φιλοδοξιών του Licklider. Στο τέλος, το έργο του οδήγησε στο σύστημα NLS, ένα σύστημα το οποίο πληροί τις προϋποθέσεις του Engelbart και περιλάμβανε την πρώτη περίπτωση λειτουργικού υπερκειμένου. Έτσι προέκυψε η αντικατάσταση του ακριβού και δύσχρηστου στυλού φωτός που χρησιμοποιούνταν στο Sketchpad, με μία φθηνότερη συσκευή εισόδου, το ποντίκι. Παράλληλα, παρατηρείται η πρώτη παρουσία παραθύρων με σκοπό την ενσωμάτωση κειμένου, γραφικών και περιεχομένου βίντεο σε μία ενιαία οθόνη, η πρώτη παρουσία υπολογιστών δικτύωσης για χρήση από περισσότερους από έναν χρήστες καθώς και το πρώτο παράδειγμα μιας

εφαρμογής επεξεργασίας κειμένου[3], [8], [9]. Δυστυχώς το NLS του Engelbart απέτυχε μιας και ο Engelbart επικεντρώθηκε στη βελτίωση της αποτελεσματικότητας για έμπειρους χρήστες και όχι για γενική χρήση γεγονός που τελικά του κόστισε και την χρηματοδότηση του.

Το 1969, και ενώ η θητεία του Engelbart στο Ερευνητικό Κέντρο του Στάνφορντ έφθανε στο τέλος της, ο Alan Kay τελείωνε το διδακτορικό του. Ο Alan Kay πρότεινε την ιδέα των επικαλυπτόμενων παραθύρων ως τρόπο διαχείρισης τόσο των διαφορετικών οπτικών στοιχείων όσο και εφαρμογών του υπολογιστή φέρνοντας και πάλι στην επιφάνεια τις πρωτοποριακές ιδέες του Engelbart με το NLS. Μετά την ολοκλήρωση του πτυχίου του ο Kay μετεγκαταστάθηκε στο νεοσύστατο για εκείνη την εποχή Xerox PARC, όπου και τέλεσε διευθυντής της Ομάδας Έρευνας και Ανάπτυξης. Λίγα χρόνια αργότερα η συγκεκριμένη ομάδα πρωτοστάτησε σε περισσότερες προόδους στις διεπαφές γραφικών χρήστη (GUI) που χρησιμοποιήθηκαν από τους Engelbart και Sutherland στο NLS και Sketchpad αντίστοιχα βελτιώνοντας το ποντίκι και καθορίζοντας άλλα βασικά χαρακτηριστικά του προσωπικού υπολογιστή[8]. Ωστόσο, η πιο σημαντική συνεισφορά του Kay ήρθε λίγα χρόνια αργότερα όταν μέσα από ένα δοκίμιο με τίτλο «A Personal Computer for Children of All Ages» περιέγραψε το όραμά του για το DynaBook. Το DynaBook σύμφωνα με τον Kay θα ήταν ένα εργαλείο το οποίο θα μπορούσε να αυξήσει αλλά και να βελτιώσει τη μάθηση για τα παιδιά. Σύμφωνα με τον οραματισμό του Kay το DynaBook είναι ένας φορητός, ευέλικτος αλλά και ισχυρός προσωπικός υπολογιστής ο οποίος θα μπορούσε να χρησιμοποιηθεί για αρκετούς διαφορετικούς σκοπούς [10]. Ολοκληρώνοντας το δοκίμιο ο Kay παραθέτει διάφορες τεχνικές προδιαγραφές και προεκτάσεις από σύγχρονα εξαρτήματα δείχνοντας έτσι την «εμπιστοσύνη» που είχε ο ίδιος στην εφαρμογή του.

Κατά τα τέλη της δεκαετίας του 1970 και αρχές του 1980 ξεκίνησε η έλευση μικρότερων αλλά και οικονομικά συμφερότερων μικροϋπολογιστών ενισχυμένων από διαδραστικό λογισμικό παραγωγικότητας που βασίζεται σε κείμενο επιτρέποντας την διαδραστική εξερεύνηση των συνεπειών που έχει η απόφαση του χρήστη. Για παράδειγμα, το VisiCalc μετέτρεψε τον απλό υπολογιστή σε ένα ισχυρό εργαλείο πρόβλεψης [11]. Καθώς η χρήση των προσωπικών υπολογιστών έγινε πιο διαδεδομένη, αναζωπυρώθηκε η ανάγκη για συστήματα προσανατολισμένα στους ανθρώπους και τις δυνατότητες τους. Έτσι το 1982 ιδρύθηκε η Ομάδα Ειδικού Ενδιαφέροντος ACM προκειμένου να δώσει έμφαση στην έρευνα που έχει ως στόχο την δημιουργία ισχυρών δεσμών μεταξύ των χρηστών και της ακαδημαϊκής κοινότητας.

Μέσα σε λίγα χρόνια και έπειτα από σημαντικές δημοσιεύσεις όπως: «The psychology of Human-Computer Interaction»[12], «User centered system design: new perspectives on Human-Computer Interaction»[13] και «Understanding computers and cognition: a new foundation for design»[14], το HCI καθιερώθηκε ως ένας κλάδος που δανείζεται από την επιστήμη των υπολογιστών τον ανθρώπινο παράγοντα, την εργονομία και τη γνωστική ψυχολογία χρησιμοποιώντας όμως τις δικές του θεωρίες, μοντέλα, πλαίσια και στόχους. Ουσιαστικά, είναι μια σχεδιαστική επιστήμη που έχει ως στόχο την κατανόηση και υποστήριξη των ανθρώπων που αλληλεπιδρούν μέσω της τεχνολογίας [15].

Το 1984, ο Loughborough ξεκίνησε ένα πρόγραμμα βασισμένο στον ανθρώπινο παράγοντα, τη βιομηχανική μηχανική και την επιστήμη των υπολογιστών.

Κατά τις δεκαετίες 1980 και 1990 η επιστήμη του HCI προσπάθησε να δημιουργήσει συστήματα τα οποία να ήταν εύχρηστα ενώ ταυτόχρονα να αντιστοιχίσει το φυσικό περιβάλλον του γραφείου σε διεπαφές υπολογιστή. Στις μέρες μας κάτι τέτοιο ίσως να φαντάζει αρκετά απλό ωστόσο για εκείνη την εποχή ήταν ένα σπουδαίο κατόρθωμα. Η συγκεκριμένη εποχή αποτέλεσε έναυσμα στο να μπορούμε να κατανοήσουμε τι μπορούσαν και τι όχι οι άνθρωποι κατά την ολοκλήρωση εργασιών σε έναν υπολογιστή αλλά και να αντιληφθούμε ότι η προσωπική πληροφορική αποτελούσε το μέλλον.

Στην επόμενη δεκαετία που ακολούθησε (1990-αρχές 2000) διαφοροποιήθηκε ο ρόλος των υπολογιστών σε εργαλεία επικοινωνίας. Το ηλεκτρονικό ταχυδρομείο κέρδισε την δημοτικότητα, γεγονός που σήμαινε ότι οι άνθρωποι είχαν την δυνατότητα πλέον να αλληλεπιδρούν μεταξύ τους και απομακρυσμένα μέσω των υπολογιστών.

Στην συνέχεια (μέσα 2000-2010), η σκυτάλη πέρασε στον σχεδιασμό έχοντας ως γνώμονα την αξία και την σκέψη για το πως η τεχνολογία μπορεί να επηρεάσει την ζωή των ανθρώπων στρέφοντας την προσοχή σε προσωπικές συσκευές σε κοινωνικά περιβάλλοντα. Ταυτόχρονα, δόθηκε και έμφαση στις σύνθετες αλληλεπιδράσεις μεταξύ των ανθρώπων, των χώρων και των τεχνολογιών καθώς και στον σχεδιασμό εφαρμογών βάζοντας ως προτεραιότητα την ευχαρίστηση, την απόλαυση του παιχνιδιού και την εξερεύνηση του χρήστη <sup>1</sup>.

Παρά το γεγονός ότι το HCI αρχικά απευθυνόταν στην επιστήμη των υπολογιστών και βασιζόταν σε εφαρμογές προσωπικής παραγωγικότητας όπως η επεξεργασία κειμένου και τα υπολογιστικά φύλλα εν τέλει το πεδίο διαφοροποιήθηκε εντελώς. Η ταχεία επέκτασή του σε τομείς που περιλάμβαναν την οπτικοποίηση, τα συστήματα πληροφοριών, τη διαδικασία ανάπτυξης του συστήματος καθώς και πολλούς τομείς σχεδιασμού, έδωσε την ευκαιρία στην εισαγωγή όλο και περισσότερων μαθημάτων σε πολλές σχολές με διαφορετικά γνωστικά αντικείμενα από αυτά της πληροφορικής. Όλα τα παραπάνω συνετέλεσαν ώστε το HCI να συμπεριλάβει την κοινωνική αλλά και οργανωτική πληροφορική, την προσβασιμότητα για τους ανθρώπους της τρίτης ηλικίας και τα γνωστικά και σωματικά προβλήματα των ανθρώπων. Απόρροια αυτού ήταν να καταφέρει να «ξεφύγει» από τον υπολογιστή και να επεκταθεί σε περισσότερες ηλεκτρονικές συσκευές<sup>2</sup>. Σήμερα οι χρήστες λειτουργούν σε ετερογενή περιβάλλοντα χάρη στην χρήση διαφόρων ηλεκτρονικών συσκευών όπως τα laptops, tablets και τα smartphones.

## 1.4 Αντίκτυπο του HCI στην κοινωνία και την οικονομία

Λαμβάνοντας ως παράδειγμα τους ανθρώπους που αντιμετωπίζουν προβλήματα όρασης μπορούμε να αντιληφθούμε σε μεγάλο βαθμό την συνεισφορά του HCI στην ανάπτυξη της κοινωνικότητας και της ζωής τους γενικότερα. Παραδείγματος χάριν, η επίσκεψη στο σούπερ μάρκετ αποτελεί μία ρουτίνα, η οποία είναι εύκολη για τους ανθρώπους που δεν αντιμετωπίζουν κάποιο πρόβλημα όρασης. Αντίθετα, για τους ανθρώπους με οπτική αναπηρία, το σούπερ μάρκετ είναι μία εμπειρία η οποία τους φέρνει σε δύσκολη θέση μιας και δεν μπορούν εύκολα να αυτοεξυπηρετηθούν. Την λύση σε αυτό το πρόβλημα έφερε η εφαρμογή e-vision <sup>3</sup> μια υβριδική προσέγγιση, που ενώνει την ελευθερία κινήσεων μιας εξωτερικής κάμερας με την επεξεργαστική ισχύ και το ποσοστό διείσδυσης των κινητών τηλεφώνων. Ωστόσο, πέρα από τον σχεδιασμό του συστήματος η καινοτομία βρίσκεται και στον σχεδιασμό της εφαρμογής η οποία έχει την ικανότητα αναγνώρισης του περιβάλλοντος. Δηλαδή η δομή της βασίζεται σε ένα συγκεκριμένο πλαίσιο. Το προτεινόμενο σύστημα αποτελείται από τρεις αισθητήρες. Μία κάμερα η οποία προσαρτάται στο κεφάλι ή σε κάποιο άλλο μέρος του σώματος, μία κινητή συσκευή με μία συνοδευτική εφαρμογή που λαμβάνει τις εικόνες που βλέπει η κάμερα και ακουστικά που συνδέονται με την εφαρμογή και παρέχουν ακουστική ανατροφοδότηση στον χρήστη κάνοντας μετατροπή του ήχου σε ομιλία. Μάλιστα η συγκεκριμένη εφαρμογή έχει βρει εφαρμογή στην ελληνική εταιρεία σούπερ μάρκετ Μασούτης<sup>4</sup>.

<sup>1</sup> <https://blog.prototypr.io/the-rise-of-human-computer-interaction-hci-823dd6286e1d>

<sup>2</sup> <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-computer-interaction-brief-intro>

<sup>3</sup> <https://eivision-project.gr/>

<sup>4</sup> <https://www.masoutis.gr/>

Προκειμένου να αντιληφθούμε καλύτερα την εφαρμογή είναι σκόπιμο να χωρίσουμε την επίσκεψη στο σούπερ μάρκετ και να δούμε την χρησιμότητα της εφαρμογής σε κάθε τμήμα. Πιο συγκεκριμένα, μπορούμε να διακρίνουμε τέσσερα επίπεδα το προϊόν, το ράφι, την διαδρομή και τα άλλα.

**Προϊόν:** Ο χρήστης έχει κάποιο προϊόν στα χέρια του και περιμένει από το σύστημα να του το αναγνωρίσει. Από πλευράς συστήματος υπάρχει ανταπόκριση ως προς τον χρήστη αναφέροντας του την ονομασία του προϊόντος .

**Ράφι:** Ο χρήστης βρίσκεται μπροστά σε ένα ράφι με περιορισμένο αριθμό προϊόντων, τότε το σύστημα του αναφέρει την κατηγορία των προϊόντων που έχει μπροστά του.

**Διαδρομή:** Ο χρήστης βρίσκεται μπροστά σε ένα μονοπάτι με μεγάλο αριθμό σχετικών προϊόντων και θέλει το σύστημα να τον κατατοπίσει για την γενική κατηγορία προϊόντων τότε η απάντηση που θα λάβει ο χρήστης είναι ότι βρίσκεται στον διάδρομο με τα κατεψυγμένα προϊόντα.

**Άλλα:** Σε αυτήν την περίπτωση το άτομο βρίσκεται στην είσοδο/έξοδο του καταστήματος τότε το σύστημα τον πληροφορεί ότι δεν βρίσκεται μπροστά σε διάδρομο ή κάποιο ράφι [16].

Επομένως λαμβάνοντας υπόψιν όλα τα παραπάνω καταλαβαίνουμε πόσο μπορεί να διαφοροποιηθεί προς το καλύτερο μία τυπική επίσκεψη στο σούπερ μάρκετ και το μεγάλο αντίκτυπο που έχει στον άνθρωπο που αντιμετωπίζει προβλήματα όρασης.

Αντίστοιχα, όσον αφορά τον αντίκτυπο στην οικονομία, οι άνθρωποι με προβλήματα όρασης μπορούν να ψωνίζουν πολύ πιο γρήγορα και αποτελεσματικά αυξάνοντας ταυτόχρονα και την αυτοπεποίθηση τους αφού μπορούν να λειτουργούν πιο αυτόνομα. [17]



Εικόνα 1.2: Περιπατητική Προσβασιμότητα<sup>5</sup>

<sup>5</sup> [https://www.imerisia.gr/politiki/48740 k-hatzidakis-axio-mimisis-katastima-masoytis-gia-tyfloyis](https://www.imerisia.gr/politiki/48740-k-hatzidakis-axio-mimisis-katastima-masoytis-gia-tyfloyis)

## Κεφάλαιο 2<sup>ο</sup>: Αναγνώριση Συναισθήματος

### 2.1 Εισαγωγή

Παλαιότερα, ο κλάδος της πληροφορικής και της ψυχολογίας αποτελούσαν δύο ξεχωριστούς τομείς. Κανένας δεν μπορούσε να αντιληφθεί εκείνη την εποχή πως με την είσοδο νέων τεχνολογιών αλλά και με την ανάπτυξη του τομέα του HCI, ότι ο χρήστης όχι μόνο θα αλληλεπιδρά με το σύστημα αλλά και το σύστημα από την πλευρά του θα μπορεί να αναγνωρίζει τα συναισθήματα του χρήστη. Εδώ και σχεδόν μισό αιώνα έχει δημιουργηθεί η προοπτική των υπολογιστών και γενικότερα των μηχανών να κατανοήσουν την συναισθηματική κατάσταση του χρήστη παρόλα αυτά μόλις το 1995 ο Picard επινόησε τον όρο «Affective Computing»[18] και γεννήθηκε ένα ερευνητικό πεδίο. Πλέον, υπάρχει μία ολόκληρη κοινότητα στην οποία οι ερευνητές εργάζονται προκειμένου να κατανοήσουν την έννοια του συναισθήματος και του τρόπου με τον οποίο οι υπολογιστές μπορούν να συλλέξουν και να εμπεδώσουν τη συναισθηματική κατάσταση των χρηστών[19].

### 2.2 Ο όρος συναίσθημα

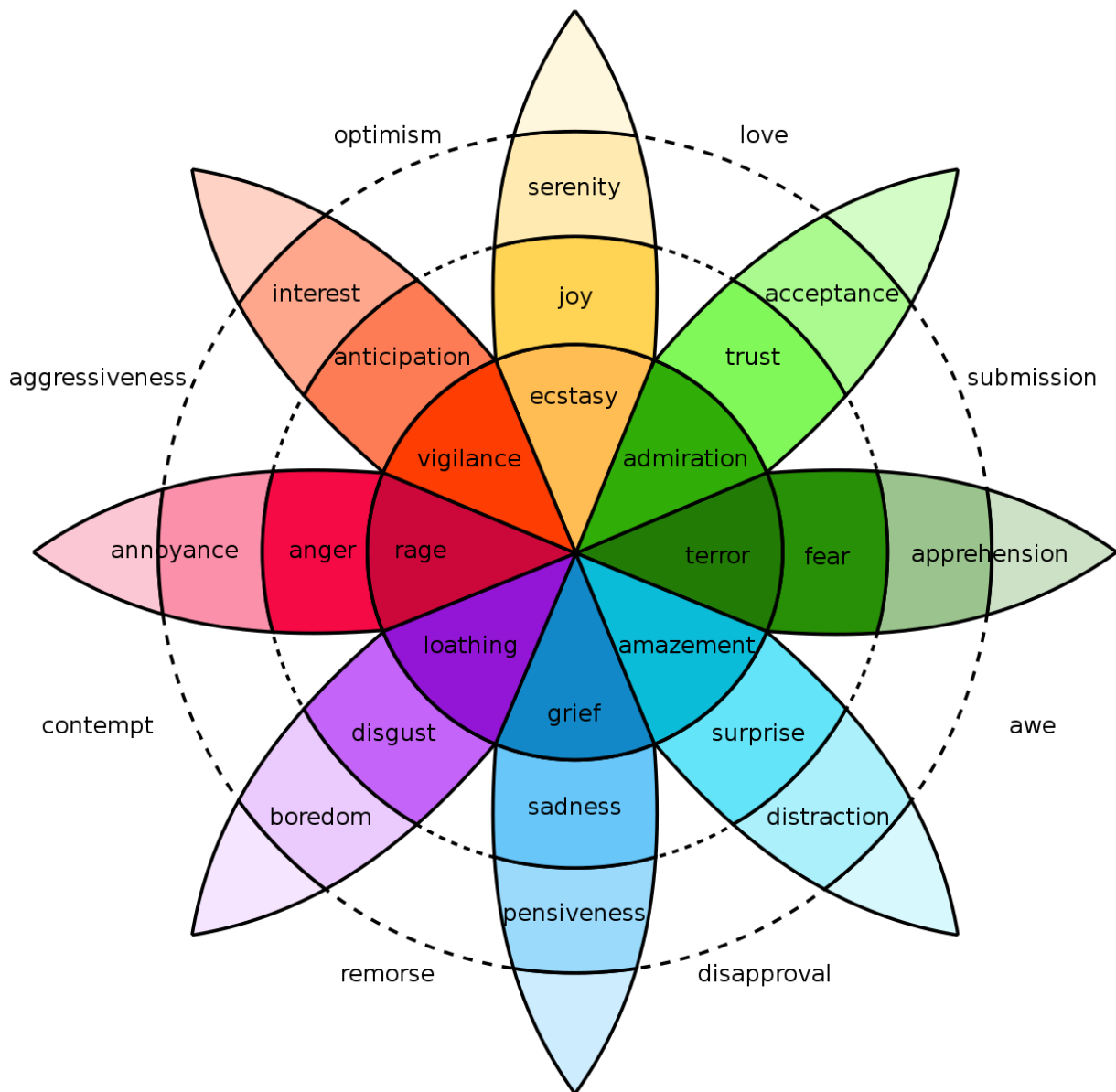
Αν και είναι πολλές οι προσπάθειες που έχουν γίνει μέχρι και σήμερα προκειμένου να διατυπωθεί ένας ακριβής ορισμός του συναισθήματος ακόμα δεν έχει επιτευχθεί. Βέβαια σύμφωνα με τον Griffiths κάτι τέτοιο είναι πολύ σημαντικό προσθέτοντας αργότερα ,το 1997,το δικό του λιθαράκι στην αποτύπωση του ορισμού θεωρώντας το ως ένα «διασπαστικό κίνητρο ανώτερης γνώσης» [20]. Το 1981 τα αδέρφια Kleinginna σε μία προσπάθεια απόδοσης ενός ακριβή ορισμού του συναισθήματος απαρίθμησαν 92 διαφορετικούς ορισμούς του συναισθήματος και εννέα σκεπτικιστικές δηλώσεις παραθέτοντας ταυτόχρονα και κάποιες δικές τους [21]. Σύμφωνα με τον ορισμό που έχει αποδοθεί από το λεξικό Larousse το 1990 το συναίσθημα χαρακτηρίζεται ως ένα « ξαφνικό πρόβλημα, παροδική ταραχή που προκαλείται από έντονο συναίσθημα φόβου, έκπληξης, χαράς κλπ.»<sup>6</sup>. Ούτε η Γνωσιακή Ψυχολογία[22] δεν ορίζει το συναίσθημα επίσημα αλλά ένας ορισμός ο οποίος μπορεί να είναι πιο λειτουργικός προκύπτει από τις πέντε σελίδες του οι οποίες αναφέρονται στο συναίσθημα ως μια ψυχική κατάσταση. Δεν είναι βέβαια ασήμαντο καθώς το συναίσθημα θεωρείται σαν μια απλή σωματική απάντηση [23]. Κατά την πάροδο του χρόνου αναπτύχθηκαν και πολλοί άλλοι ορισμοί ωστόσο θεωρήσαμε σκόπιμο να αναφέρουμε μόνο κάποιους από αυτούς.

### 2.3 Κατηγοριοποίηση Συναισθημάτων

Κάθε άνθρωπος έχει την ικανότητα να αναπτύσσει συναισθήματα με μοναδικό τρόπο ανάλογα με τα βιώματά του. Τα συναισθήματα διακρίνονται σε δύο κατηγορίες. Τα πρωτεύοντα ή βασικά και τα δευτερεύοντα ή παραγόμενα. Ως βασικά συναισθήματα χαρακτηρίζονται αυτά για τα οποία παρατηρείται διαπολιτισμική καθολικότητα στις μη λεκτικές εκφράσεις τους. Αυτού του είδους τα συναισθήματα κάνουν την εμφάνιση τους κατά τη διάρκεια του πρώτου χρόνου της ζωής ενός ανθρώπου [24]. Αντίστοιχα, τα δευτερεύοντα ή παραγόμενα συναισθήματα είναι ένας συνδυασμός των βασικών συναισθημάτων. Σε αυτήν την κατηγορία μπορούμε να εντάξουμε συναισθήματα όπως η υπερηφάνεια, η ευγνωμοσύνη, η ενσυναίσθηση, η ενοχή, η αμηχανία, η ζήλεια, η περιφρόνηση, η

<sup>6</sup> <https://www.larousse.fr/dictionnaires/francais/%C3%A9motion/28829>

ντροπή[25]. Ωστόσο, υπάρχουν έξι βασικά συναισθήματα σύμφωνα με τον Paul Eckman που παρατηρήθηκε ότι βιώθηκαν παγκοσμίως σε κάθε ανθρώπινο πολιτισμό. Τα συναισθήματα που εντόπισε ήταν η ευτυχία, η έκπληξη, η λύπη, η αηδία, ο φόβος και τέλος ο θυμός<sup>7</sup>.



Εικόνα 2.1: Συναισθηματικός Χάρτης

<sup>7</sup> <https://www.verywellmind.com/an-overview-of-the-types-of-emotions-4163976>

## 2.4 Συναισθηματική Ανάλυση κατά την Αλληλεπίδραση Ανθρώπου-Μηχανής

Τα τελευταία χρόνια έχει ενσωματωθεί σε αυτήν την αλληλεπίδραση και το συναίσθημα στην διεπαφή. Πριν από αρκετά χρόνια η Rosalind Picard επινόησε την φράση «affective computing» για υπολογισμούς οι οποίοι επηρεάζουν είτε σκόπιμα είτε όχι το συναίσθημα [18]. Προκειμένου να είμαστε σε θέση να αντιληφθούμε καλύτερα την αναγκαιότητα της αναγνώρισης συναισθήματος κατά την αλληλεπίδραση μπορούμε να λάβουμε ένα παράδειγμα από την καθημερινή μας ζωή. Κατά την διάρκεια της ημέρας ερχόμαστε σε επαφή με αρκετούς ανθρώπους. Η επαφή μας μαζί τους, μας επιφέρει κάποια συναισθήματα όπως αντίστοιχα δημιουργούνται κάποια συναισθήματα και στον άνθρωπο που έχουμε απέναντι μας [26]. Το γεγονός ότι ο άνθρωπος είναι σε θέση να αντιλαμβάνεται μεγάλο μέρος των συναισθημάτων είτε πηγάζουν από εκείνον είτε από τον απέναντι του αδιαμφισβήτητα μπορεί να επιφέρει μεγάλο όφελος στις διαπροσωπικές μας σχέσεις. Επιπλέον, μπορεί και επηρεάζει τις αντιλήψεις μας, την μνήμη καθώς και την ικανότητα μας σχετικά με την λήψη αποφάσεων. Κατά τους Reeves και Nass αντίστοιχη συμπεριφορά αναπτύσσει ο άνθρωπος και κατά την διάρκεια της αλληλεπίδρασης του με ένα μηχάνημα νομίζοντας ότι το δεύτερο έχει συναισθηματική ευφυΐα. Σύμφωνα με την συγκεκριμένη μελέτη κατά την διάρκεια της διεπαφής τα αντικείμενα όπως και τα συμβάντα στην οθόνη μπορούν να ενεργοποιήσουν πρωτόγονες συναισθηματικές αντιδράσεις [27]. Ωστόσο, σύμφωνα με τον Wegner ένα συναισθηματικό ερέθισμα το οποίο είναι αρκετά διεγερτικό έχει σαν αποτέλεσμα ο χρήστης να μην μπορεί να αγνοήσει το κίνητρο και να οδηγήσει στην αποτυχία της ρύθμισης μέσω της επιλεκτικής προσοχής [28].

Από την άλλη πλευρά, μία αντίστοιχη έρευνα των Bower & Forgas έδειξε ότι οι άνθρωποι δίνουν μεγαλύτερη προσοχή σε ερεθίσματα καθώς και σκέψεις οι οποίες να έχουν κάποια σχέση με την συναισθηματική κατάσταση στην οποία βρίσκονται εκείνη την στιγμή [29].

Η ανάγκη για συναισθηματική ανάλυση κατά την διάρκεια της αλληλεπίδρασης του Ανθρώπου με κάποια ηλεκτρονική συσκευή μπορεί να επιφέρει πολλά οφέλη. Πιο συγκεκριμένα, μία διεπαφή η οποία διευκολύνει ή δυσκολεύει τους σκοπούς ενός χρήστη επιδράει άμεσα στην συναισθηματική του κατάσταση. Μία διεπαφή όμως η οποία έχει την ικανότητα να αντιληφθεί το συναίσθημα, θα μπορούσε να αξιοποιήσει τις πληροφορίες που έχει συλλέξει προκειμένου να δώσει μία ανατροφοδότηση σχετικά με το αν είναι σε θέση να ικανοποιήσουν τις ανάγκες του χρήστη αλλά και στην αντίθετη περίπτωση να μπορέσουν να τροποποιηθούν για να επιτύχουν τον σκοπό τους [30].

Μία από τις επιλογές, προκειμένου να κάνουμε την συνύπαρξη του ανθρώπου με το μηχάνημα πιο ομαλή και ωφέλιμη θα ήταν να πραγματοποιήσουμε την επιθυμία του ανθρώπου και να δώσουμε την ικανότητα στο μηχάνημα να αντιλαμβάνεται συναισθήματα. Κάτι τέτοιο δεν είναι ανέφικτο μιας και πλέον έχουν γίνει πολλές έρευνες επάνω στην συναισθηματική ανάλυση με χαρακτηριστική αυτή που προέρχεται από τον ήχο. Ωστόσο, προκειμένου να μπορούμε να αντιληφθούμε καλύτερα κάποιες έννοιες θα ήταν σκόπιμο να γίνει μία συνοπτική αναφορά στον ήχο και τα χαρακτηριστικά που τον διέπουν.

## Κεφάλαιο 3<sup>ο</sup> : Ήχος

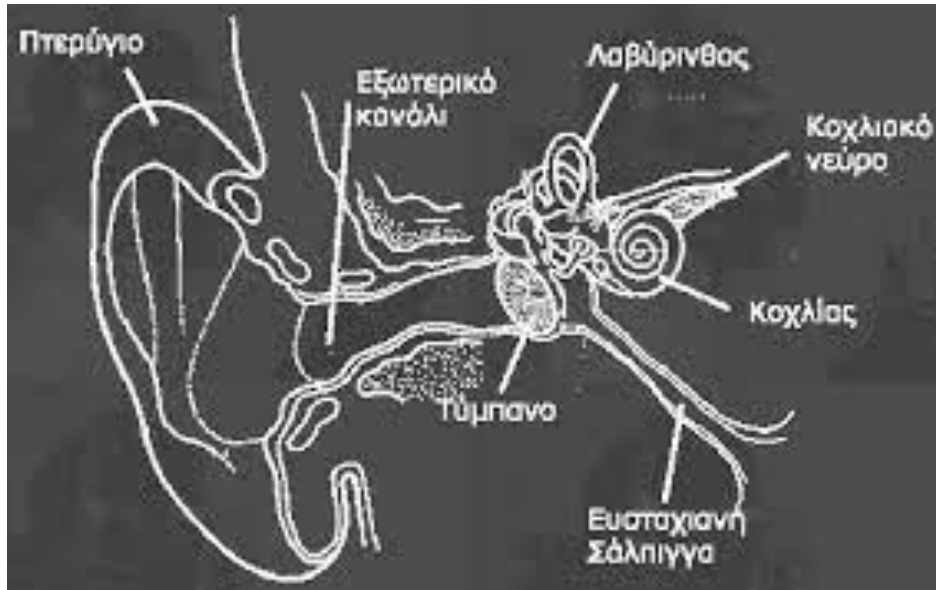
### 3.1 Εισαγωγή

Παραπάνω έγινε προσπάθεια μίας συσχέτισης του τομέα της πληροφορικής και της ψυχολογίας μέσω του ΗCI. Αντίστοιχη συσχέτιση μπορούμε να διακρίνουμε και στον Ήχο. Είναι κοινώς αποδεκτό και έχει επιβεβαιωθεί από αρκετές μελέτες ότι ο άνθρωπος κατά το άκουσμα ενός ήχου, είτε αυτός προέρχεται από ομιλία, είτε από μουσική, είτε από κάποια άλλη πηγή, αναπτύσσει κάποια συναισθήματα. Αυτά τα συναισθήματα όπως είναι φυσικό μπορεί να διαφέρουν από άνθρωπο σε άνθρωπο αλλά και σε κάποιες περιπτώσεις να είναι κοινά. Παρακάτω, παρουσιάζεται μία συνοπτική περιγραφή του ακουστού μας συστήματος και διακρίνονται κάποια βασικά ηχητικά χαρακτηριστικά τα οποία μας βοήθησαν στην εξαγωγή συμπερασμάτων στο πειραματικό μας μέρος.

### 3.2 Περιγραφή ακουστού συστήματος

Μία από τις πέντε βασικές αισθήσεις του ανθρώπου είναι η ακοή. Όργανο αντίληψης αποτελούν τα αυτιά ενώ το αντικείμενο αντίληψης είναι ο ήχος. Η ακοή, θεωρείται η πιο σημαντική από τις υπόλοιπες αισθήσεις καθώς χάρη σε αυτήν γίνεται άμεσα αντιληπτός ο εξωτερικός χώρος, διευκολύνεται η επικοινωνία και η εκπαίδευση<sup>8</sup>. Το ανθρώπινο ακουστικό σύστημα αποτελείται από δύο αυτιά τα οποία βρίσκονται στην αριστερή και την δεξιά πλευρά του κεφαλιού, το αιθουσαίο-κοχλιακό νεύρο και το κεντρικό ακουστικό νευρικό σύστημα το οποίο αποτελείται από ακουστικά κέντρα στον εγκέφαλο και τις συνδετικές οδούς στο εγκεφαλικό στέλεχος. Κάθε αυτί χωρίζεται σε τρία επιπλέον λειτουργικά μέρη: το εξωτερικό, το μέσο και το εσωτερικό αυτί. Το εσωτερικό αυτί αποτελείται από τρία μέρη: τον προθάλαμο, τα ημικυκλικά κανάλια και τον κοχλία ο οποίος χρησιμεύει και ως «προστατευτικό» για το όργανο της ακοής και της ισορροπίας. Τα μέρη του οργάνου της ισορροπίας περιέχονται στον προθάλαμο και στα ημικυκλικά κανάλια. Αντίστοιχα, το όργανο της ακοής, το όργανο του Corti, βρίσκεται στον κοχλία. Το γεγονός ότι οι θέσεις των αισθήσεων της ακοής και της ισορροπίας είναι γειτονικές προκαλεί ορισμένες αλληλεπιδράσεις μεταξύ των δύο αισθήσεων. Οι δομές του ανθρώπινου αυτιού είναι ενσωματωμένες στο κροταφικό οστό του κρανίου με μόνο ορατό σημείο από το ανθρώπινο μάτι το εξωτερικό μέρος του αυτιού, τα λεγόμενα πτερύγια [31]. Το ανθρώπινο αυτί έχει δυναμικό εύρος ακοής περίπου τα 140 dB και μπορεί να ακούσει σε ένα εύρος ζώνης από πολύ χαμηλές συχνότητες (κάποιες δεκάδες Hz) μέχρι περίπου τα 15 KHz με 20 kHz.

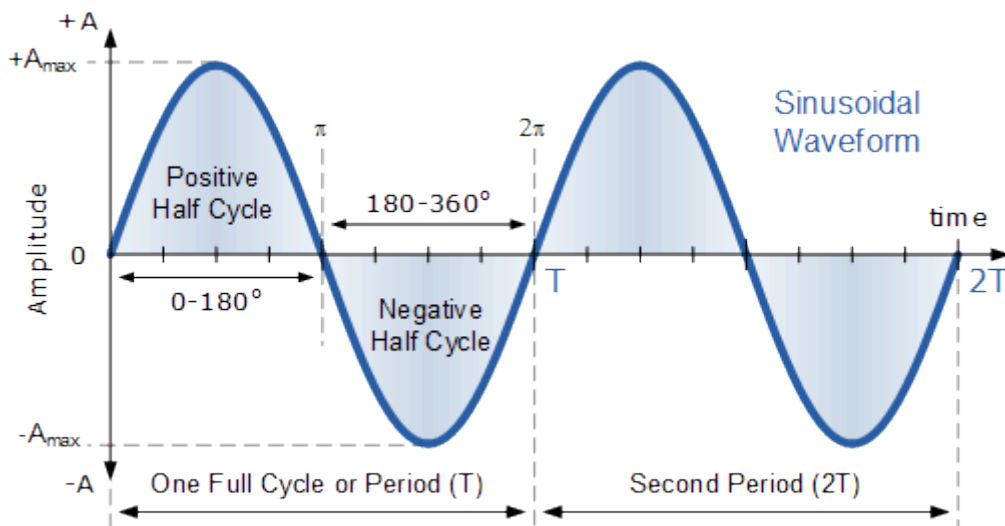
<sup>8</sup> <https://el.wikipedia.org/wiki/%CE%91%CE%BA%CE%BF%CE%AE>



Εικόνα 3.1: Ακουστική αντίληψη

### 3.3 Η έννοια του ήχου

Ένα ηχητικό σήμα παράγεται από τις διακυμάνσεις της πίεσης του αέρα. Τα ηχητικά σήματα επαναλαμβάνονται σε τακτά χρονικά διαστήματα με τέτοιο τρόπο ώστε το κάθε κύμα να έχει το ίδιο σχήμα. Το ύψος, που είναι γνωστό και ως πλάτος, αντιπροσωπεύει την ένταση του ήχου. Ο χρόνος που απαιτείται ώστε να ολοκληρωθεί ένα πλήρες κύμα από το σήμα ονομάζεται περίοδος ενώ αντίστοιχα ο αριθμός των κυμάτων που δημιουργεί το σήμα σε ένα δευτερόλεπτο ονομάζεται συχνότητα. Η συχνότητα είναι αντίστροφη της περιόδου και έχει σαν μονάδα μέτρησης τα Hertz. Ωστόσο, οι ήχοι που συναντάμε τις περισσότερες φορές μπορεί να μην ακολουθούν τόσο απλά και κανονικά περιοδικά μοτίβα [32].



Εικόνα 3.2: Μία ημιτονοειδής κυματομορφή

### 3.4 Ψηφιοποίηση του Ήχου

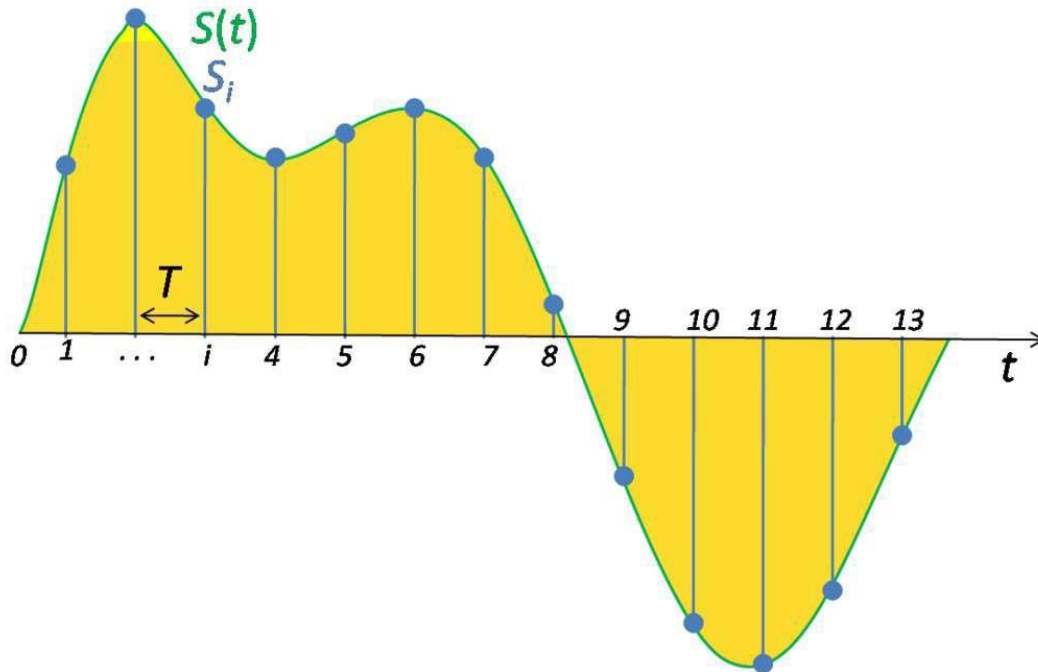
Το Ηχητικό Σήμα είναι αναλογικό. Έτσι, για να μπορεί να αναπαρασταθεί στον υπολογιστή και στην συνέχεια να μπορέσουμε να ταξινομήσουμε τον ήχο θα πρέπει να περάσει από την διαδικασία της ψηφιοποίησης. Κατά την διάρκεια της ψηφιοποίησης, το αναλογικό σήμα μετατρέπεται σε ψηφιακό. Ωστόσο προκειμένου να κωδικοποιηθεί ο ήχος και να αποθηκευτεί σε ψηφιακή μορφή σε έναν ηλεκτρονικό υπολογιστή πρέπει πρώτα να ολοκληρωθεί η διαδικασία της δειγματοληψίας.

Κατά την διάρκεια της δειγματοληψίας γίνεται παροδική λήψη δειγμάτων της έντασης του ηλεκτρικού σήματος στο οποίο έχει μετατραπεί ένας αναλογικός ήχος. Το πλήθος των δειγμάτων που παίρνουμε ανά μονάδα χρόνου μετρώντας σε δευτερόλεπτα ονομάζεται συχνότητα δειγματοληψίας (sampling frequency) ή αλλιώς ρυθμός δειγματοληψίας (sampling rate). Ο ρυθμός δειγματοληψίας παίζει καθοριστικό ρόλο στην πιστότητα της αναπαραγωγής ήχου αλλά και στο μέγεθος των δεδομένων που θα δημιουργηθούν. Ο χρόνος που μεσολαβεί μεταξύ των δειγμάτων λέγεται περίοδος δειγματοληψίας και είναι αντιστρόφως ανάλογη της συχνότητας δειγματοληψίας. Προκειμένου να επιτευχθεί μιας υψηλής ποιότητας αναπαραγωγή ήχου θα πρέπει να επιλεγεί ο ρυθμός δειγματοληψίας με τέτοιο τρόπο, ώστε να καταγραφεί όλο το φάσμα των ακουστικών συχνοτήτων. Σύμφωνα με το θεώρημα Δειγματοληψίας Nyquist-Shannon απαιτείται ελάχιστος ρυθμός δειγματοληψίας τα 40 KHz, δηλαδή 40.000 δείγματα το δευτερόλεπτο [33].

Το επόμενο στάδιο είναι αυτό της κβαντοποίησης. Η κβαντοποίηση σε συνδυασμό με την δειγματοληψία δημιουργεί την ψηφιοποίηση. Το συγκεκριμένο στάδιο είναι πολύ σημαντικό προκειμένου να αναπαρασταθεί κάθε δείγμα με την μορφή μιας δυαδικής σειράς από bits, αφού πρέπει να τεθεί μία διακριτή τιμή στο συνεχώς μεταβαλλόμενο πλάτος της τάσης του αναλογικού σήματος[34]. Κατά την διάρκεια ωστόσο της κβαντοποίησης μπορεί να προκύψουν διάφορα προβλήματα. Ο θόρυβος κατά την διάρκεια της κβάντισης είναι η κύρια πηγή παραμόρφωσης των ηχητικών σημάτων. Απόρροια αυτού αποτελεί το γεγονός ότι η ποιότητα του ήχου θα πρέπει να «συμβιβαστεί» με τον αριθμό των bits που αποθηκεύονται για κάθε δείγμα. Προκειμένου να μπορεί να αντιμετωπιστεί όσο πιο ομαλά γίνεται ο θόρυβος κβάντισης, θα πρέπει να τοποθετηθεί σε τέτοια θέση στις περιοχές του σήματος ώστε να επηρεάζει λιγότερο την πιστότητα του σήματος και ο ρυθμός παροχής των δεδομένων να ελαττώνεται χωρίς όμως να προκαλεί ακουστική παραμόρφωση[35].

Τέλος, κατά την διάρκεια της ψηφιοποίησης μπορεί να εμφανιστεί και το πρόβλημα της αναδίπλωσης. Πρώτος επιστήμονας που ανακάλυψε πως η μέγιστη συχνότητα αναλογικού σήματος που μπορεί να αποδοθεί χωρίς να υπάρξει αλλοίωση, είναι το μισό της συχνότητας δειγματοληψίας ήταν ο H.Nyquist. Στην περίπτωση που κάποιες συχνότητες είναι μεγαλύτερες από τα μισά της συχνότητας ψηφιοποίησης εμφανίζονται λανθασμένα σαν χαμηλές συχνότητες με αποτέλεσμα να δημιουργείται το φαινόμενο του ελλιπούς ρυθμού δειγματοληψίας γνωστό ως αναδίπλωση (aliasing). Προκειμένου να αποφευχθεί η εμφάνιση συχνοτικών ειδωλίων που θα παραμορφώσουν το σήμα μετά την δειγματοληψία οι

ψηφιοποιητές φιλτράρουν όλες τις συχνότητες που είναι μεγαλύτερες από το μισό της συχνότητας δειγματοληψίας [34]



Εικόνα 3.3: Δείγμα μετρήσεων σε τακτά χρονικά διαστήματα<sup>9</sup>

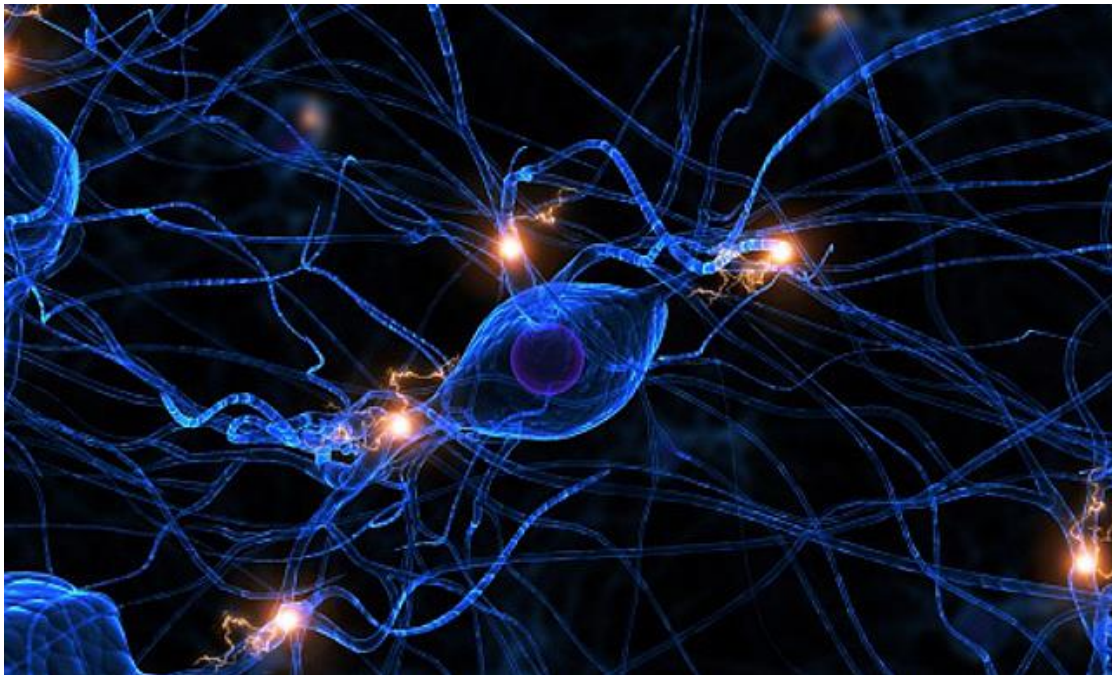
<sup>9</sup> [https://commons.wikimedia.org/wiki/File:Signal\\_Sampling.png](https://commons.wikimedia.org/wiki/File:Signal_Sampling.png)

### 3.5 Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα είναι ένας τομέας που έχει καταφέρει να εισχωρήσει βαθιά στην επιστημονική κοινότητα για δύο λόγους. Ο πρώτος είναι η ανάπτυξη των υλικών και ο δεύτερος η υψηλή διαθεσιμότητα των δεδομένων.

Τα συνελκτικά νευρωνικά δίκτυα (CNN) ως βαθιά νευρωνικά δίκτυα, προσφέρουν εξαιρετικές λύσεις στα προβλήματα αναγνώρισης εικόνας ωστόσο εφαρμόζονται και στην ταξινόμηση ήχου. Τα CNN χρησιμοποιούνται για εργασίες με ταξινόμηση ομιλίας και μουσικής και είναι περισσότερο βελτιωμένα σε σύγκριση με τις μηχανές διανυσματικής υποστήριξης Support Vector Machine(SVM) και τη γενικευμένη μέθοδο στιγμών Generalized Method of Moments (GMM). Όμως, τόσο ο σχεδιασμός όσο και η εκπαίδευση των βαθιών νευρωνικών δικτύων απαιτούν πολλούς υπολογιστικούς πόρους αλλά και πολλά δεδομένα. Λαμβάνοντας υπόψιν το παραπάνω είναι λογικό ότι τα προ-εκπαιδευμένα βαθιά μοντέλα υπερτερούν συγκριτικά με την βαθιά μάθηση [36]. Ωστόσο, στο πειραματικό μέρος της εργασίας μας επιλέξαμε να χρησιμοποιήσουμε και τις δύο προσεγγίσεις.

Τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN), ασχολούνται με τις χρονικές ακολουθίες πληροφοριών και έχουν την δυνατότητα να μοντελοποιούν χρονικές εξαρτήσεις εισάγοντας τον βρόγχο ανάδρασης μεταξύ εισόδου και εξόδου. Τα δίκτυα μακράς βραχύχρονης μνήμης Long Short -Term Memory (LSTM) που ανήκουν στα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) εισάγουν την έννοια της μνήμης κυψέλης. Η κυψέλη λειτουργεί μαθαίνοντας, κρατώντας και τελικά ξεχνώντας τις πληροφορίες . Λόγω αυτού, τα δίκτυα LSTM αποτελούν ένα πολύ ισχυρό εργαλείο. Στην περίπτωση βέβαια που συνδυάσουμε δύο δίκτυα LSTM και δημιουργήσουμε ένα αμφίδρομο δίκτυο Bidirectional (BLTSM) τότε αυξάνουμε ακόμα περισσότερο το πλεονέκτημα τους. Στα αμφίδρομα δίκτυα, το ένα δίκτυο επεξεργάζεται την διαδρομή προς τα εμπρός και το άλλο δίκτυο την διαδρομή προς τα πίσω. Αυτού του είδους δικτύων έχουν εφαρμοστεί επιτυχώς σε διάφορες εργασίες μοντελοποίησης [37].



Εικόνα 3.4: Τεχνητά νευρωνικά δίκτυα

## 3.6 Χαρακτηριστικοί εξαγωγείς

### 3.6.1 Δυναμικά Χαρακτηριστικά

Η συνολική ενέργεια του σήματος δίνεται από τον τύπο:

$$x_{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$$

- **Κατάτμηση (segment)**

Ως κατάτμηση ορίζεται η ανίχνευση σημείων αλλαγής και χρησιμοποιείται για τον εντοπισμό των χρονικών ορίων των τμημάτων.

- **Χαμηλή ενέργεια (low energy)**

Χρησιμοποιείται προκειμένου να υπολογίσουμε πόσα κομμάτια του σήματος εμφανίζουν χαμηλή ενέργεια. Ως χαμηλή ενέργεια υπολογίζεται οποιαδήποτε ενέργεια είναι μικρότερη από τον μέσο όρο.

### 3.6.2 Χαρακτηριστικά Τέμπο

- **Αναλογία προσήμου-μεταβολών (zerocross):**

Αποτελεί δείκτη θορύβου και συνίσταται προκειμένου να μετρήσουμε πόσες φορές το σήμα διασχίζει τον άξονα τον  $x$ , αλλάζει δηλαδή πρόσημο.

- **Συχνότητα αποκοπής (rolloff):**  
Ενέργεια υψηλής συχνότητας (I)

Έχει οριστεί ως προεπιλεγμένη αναλογία η 0,85. Αναλόγως την ποσότητα σήματος που ανιχνεύεται πάνω από αυτή την αναλογία υπολογίζεται η ποσότητα της υψηλής συχνότητας στο σήμα.

- **Κεντροειδές, διακύμανση, λοξότητα, κύρτωση, επιπεδότητα, εντροπία (centroid, spread, skewness, kurtosis, flatness, entropy):**

Όροι οι οποίοι χρησιμοποιούνται για να περιγράψουν την φασματική κατανομή που μπορεί να χαρακτηριστεί με στατιστικές ροπές.

- **Συντελεστές φάσματος συχνότητας (Mfcc):**

Χρησιμοποιείται προκειμένου να προσφέρει μία περιγραφή του φασματικού σχήματος του ήχου.

## 3.7 Υψηλού επιπέδου χαρακτηριστικά ήχων

### 3.7.1 Δομή και μορφή

Προκειμένου να γίνονται αναλύσεις υψηλότερου επιπέδου χρησιμοποιούνται πιο εξεζητημένα εργαλεία.

- **Πίνακας ομοιότητας (similarity matrix):**

Παρουσιάζονται οι ομοιότητες ανάμεσα σε όλα τα ζεύγη των δεδομένων που εισάγονται.

- **Καμπύλη καινοτομίας(novelty):**

Αντιπροσωπεύει την πιθανότητα κατά μήκος του χρόνου να υπάρχουν μεταβάσεις ανάμεσα σε διαδοχικές καταστάσεις. Αυτό φαίνεται από τα ύψη των κορυφών.

- **Κατάτμηση καινοτομίας(segment “Novelty”):**

Υποδεικνύει την χρονική στιγμή που συμβαίνει ασυνέχεια χαρακτηριστικών προκειμένου να γίνεται σωστότερη κατάτμηση του ήχου.

### 3.7.2 Στατιστικά Χαρακτηριστικά

Παρέχει λειτουργίες και εφαρμογές προκειμένου να περιγράψουμε, αναλύσουμε και να υποδείξουμε δεδομένα.

- **Μέσος όρος (mean):**

Δίνει το μέσο όρο των πλαισίων της χαρακτηριστικής συχνότητας. Αν αυτό κατατμηθεί, το αποτέλεσμα θα είναι μία σειρά σε διαδοχικά τμήματα.

- **Τυπική απόκλιση (standard deviation):**

Επιστρέφει την τυπική απόκλιση στα πλαίσια της χαρακτηριστικής συχνότητας. Η συχνότητα είναι ένας δομημένος πίνακας. Αντίστοιχη δόμηση θα έχουν και τα δεδομένα εξόδων.

- **Διάμεσος (median):**

Επιστρέφει την διάμεσο στα πλαίσια της χαρακτηριστικής συχνότητας.

- **Στατιστικά στοιχεία (stat):**

Χρησιμοποιείται σε οποιοδήποτε αντικείμενο και να δώσει τα στατικά στοιχεία του αντικειμένου σε δομημένη μορφή.

- **Ιστόγραμμα (histo):**

Μπορεί να χρησιμοποιηθεί για οποιοδήποτε στοιχείο και να δώσει το κατάλληλο ιστόγραμμα. Τα δεδομένα αποθηκεύονται σε θέσεις ίσων αποστάσεων.

- **Αναλογία προσήμου-μεταβολών (zerocross):**

Μετράει τις φορές που το σήμα περνάει τον άξονα x, δηλαδή αλλάζει πρόσημο. Αυτό αποτελεί ένδειξη θορύβου.

- **Κεντροειδές (centroid):**

Το Κεντροειδές αναφέρεται στο γεωμετρικό κέντρο ενός σχήματος. Αυτή η επεξεργασία μας δίνει το κεντροειδές ενός σήματος.

- **Διάδοση (spread):**

Επαναφέρει την τυπική απόκλιση των δεδομένων.

- **Λοξότητα (skewness):**

Απαντάει με τον συντελεστή λοξότητας. Μετράει την συμμετρία της κατανομής. Στην περίπτωση που έχει θετική τιμή θεωρείται ως θετική λοξή με τιμές μεγαλύτερες από την μέση ενώ αν οι

περισσότερες τιμές είναι αρνητικές τότε θεωρείται αρνητικά λοξή. Στην περίπτωση που είναι απόλυτα συμμετρική τότε έχει μηδενική λοξότητα.

- **Κύρτωση (Kurtosis):**

Επιστρέφει την κύρτωση δεδομένων, την συγκέντρωση δηλαδή των δεδομένων γύρω από την μέση τιμή.

- **Χαρακτηριστικά (features):**

Το σύνολο των χαρακτηριστικών αποθηκεύεται σε έναν πίνακα δομών και μία μεταβλητή επιστρέφει το αποτέλεσμα.

### 3.7.3 Τεχνικές προβλέψεις

Γίνεται πρόβλεψη των αποτελεσμάτων της γραμμικής παλινδρόμησης

- **Συναίσθημα (emotion):**

Επιχειρεί να προβλέψει μία περιγραφή του συναισθήματος με βάση την ανάλυση του ήχου καθώς και του μουσικού περιεχομένου των ηχογραφήσεων.

- **Ταξινόμηση (classify):**

Η ταξινόμηση μπορεί να επιτευχθεί είτε με την διασταύρωση ενός χαρακτηριστικού είτε ενός συνόλου χαρακτηριστικών.

- **Σύμπλεγμα (Cluster):**

Χρησιμοποιούνται κοινά στοιχεία ποικίλων μοναδικών ακουστικών σημάτων προκειμένου να δημιουργηθεί ένα καινούριο ενιαίο σήμα.

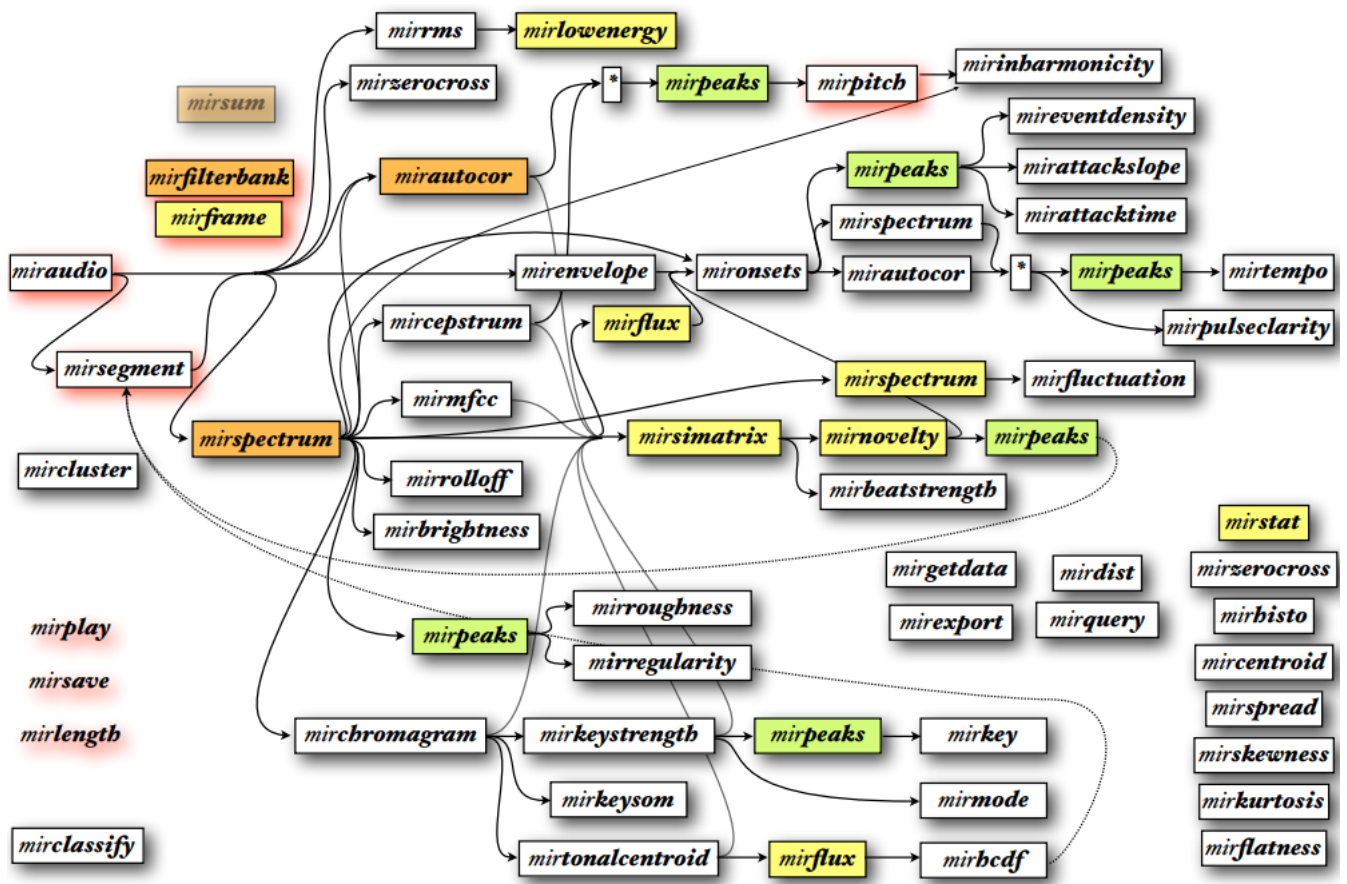
### 3.7.4 Εξαγωγή

- **Επιστροφή αποτελέσματος(getdata)**

Επιστρέφει τα δεδομένα εισόδου σε τέτοια μορφή ώστε να μπορεί να αξιοποιηθεί για υπολογισμούς.

- **Εξαγωγή δεδομένων (export)**

Εξάγει όλα τα δεδομένα σε ένα αρχείο κειμένου.



Εικόνα 3.5: MIRtoolbox Feature

### 3.8 Επίλογος

Στο τρίτο κεφάλαιο, δόθηκε μία περιγραφή του ανθρώπινου ακουστού συστήματος και αναφερθήκαμε στην έννοια του ήχου. Εν συνεχεία, παρουσιάστηκε αναλυτικά η διαδικασία της ψηφιοποίησης και δόθηκαν οι ορισμοί των χαρακτηριστικών που διέπουν τον ήχο αφού πρώτα κατηγοριοποιήθηκαν κατάλληλα. Στόχος της συγκεκριμένης ανάλυσης ήταν να μπορέσουμε να εξοικειωθούμε με ορισμούς αλλά και έννοιες τις οποίες θα συναντήσουμε στα επόμενα κεφάλαια και είναι απαραίτητοι για το πειραματικό μέρος της εργασίας.

## Κεφάλαιο 4<sup>ο</sup> : Παρουσίαση του Προβλήματος

### 4.1 Εισαγωγή

Στο παρακάτω κεφάλαιο αναφερόμαστε στο πρόβλημα που προκύπτει κατά την διαδικασία κατηγοριοποίησης του συναισθήματος συνεπώς και του ήχου με βάση το συναίσθημα και γίνεται μία βιβλιογραφική ανασκόπηση σε έρευνες που ασχολήθηκαν με το αντικείμενο,

### 4.2 Το πρόβλημα της κατηγοριοποίησης

Όπως αναφέρθηκε σε προηγούμενο κεφάλαιο, όταν αναφερόμαστε στο συναίσθημα γνωρίζουμε ότι βαδίζουμε σε αχαρτογράφητα νερά και αυτό γιατί αφενός δεν υπάρχει ένας σαφής ορισμός που να καθορίζει το συναίσθημα και αφετέρου γιατί ο κάθε άνθρωπος βιώνει το συναίσθημα διαφορετικά. Επομένως, η κατηγοριοποίηση δεν γίνεται να είναι απόλυτα ακριβής, παρά μία προσέγγιση. Αντίστοιχα, την ίδια δυσκολία βιώνουμε και κατά την διάρκεια της κατηγοριοποίησης του ήχου καθώς δεν μπορούμε να εξάγουμε ένα ασφαλές συμπέρασμα για την συσχέτιση μεταξύ του συναισθήματος και του ήχου.

Η ταξινόμηση των ήχων με βάση το συναίσθημα μπορεί να γίνει με γνώμονα ήδη καθορισμένα πρότυπα. Σε αυτή την περίπτωση αναφερόμαστε στην εποπτευόμενη μάθηση. Μία άλλη προσέγγιση, αποτελεί η μη εποπτευόμενη μάθηση. Σύμφωνα με αυτήν την προσέγγιση, αναλύονται οι ήχοι εξετάζοντας τα χαρακτηριστικά τους και εν συνεχεία ταξινομούνται σε ομάδες με βάση τις ομοιότητές τους.

Στόχος της συγκεκριμένης πτυχιακής εργασίας είναι να μπορέσουμε να καταλάβουμε κατά πόσο οι ήχοι που προέρχονται από ηλεκτρονικές συσκευές και επιλέχθηκαν για να προκαλούν κάποια συναισθήματα στον χρήστη, επιτυγχάνουν τον σκοπό τους.

### 4.3 Βιβλιογραφική ανασκόπηση

Έπειτα από αρκετή μελέτη της διαθέσιμης βιβλιογραφίας καταλήξαμε στις εργασίες που αναλύονται παρακάτω ως τις πιο άμεσα σχετιζόμενες με το θέμα.

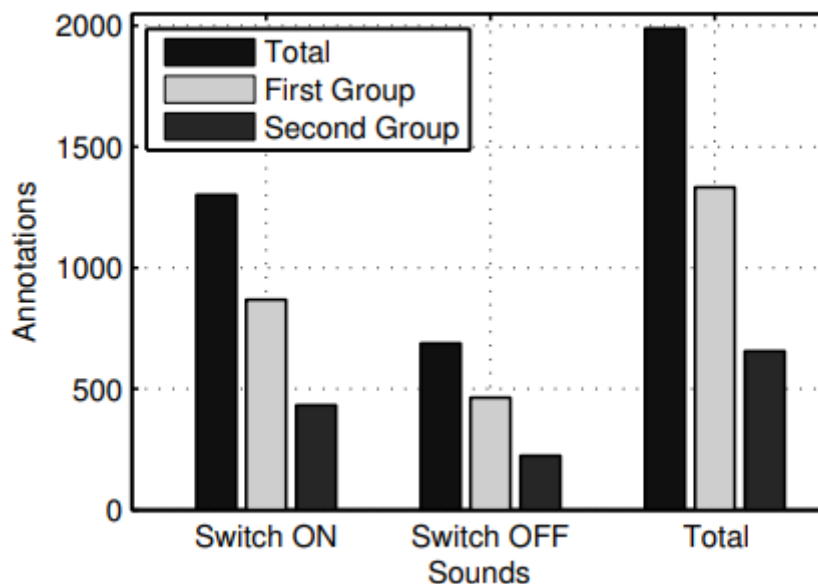
Η έρευνα η οποία παρουσιάστηκε στο 134<sup>ο</sup> Συνέδριο της Ρώμης το 2013 με τίτλο «Investigating Auditory Human-Machine Interaction: Analysis and Classification of Sounds Commonly Used by Consumer Devices» είναι αυτή η οποία εναρμονίζεται σε μεγαλύτερο βαθμό με την παρούσα εργασία. Σύμφωνα με την συγκεκριμένη έρευνα [38] συμπεριλήφθηκαν ήχοι έναρξης καθώς και λήξης διαφόρων καταναλωτικών συσκευών και έγινε μία προσπάθεια προκειμένου να εξαχθεί το συμπέρασμα αν ο χρήστης μίας συσκευής, μπορεί να αντιληφθεί ένα ακουστικό μήνυμα χωρίς να έχει προηγηθεί κάποια εκπαίδευση.

Προκειμένου να επιτευχθεί το παραπάνω, οι συγγραφείς συνδύασαν τεστ ακρόασης και μηχανικής μάθησης σε κάποιους σύντομους ήχους (27 ήχους switch on και 14 ήχους switch off), οι οποίοι συλλέχθηκαν από διάφορες καταναλωτικές συσκευές με συχνότητα δειγματοληψίας τα 44,1 kHz και ανάλυση 16 bit. Όλοι οι ήχοι ήταν προ επεξεργασμένοι και εν συνεχεία κανονικοποιήθηκαν με σκοπό την εξάλειψη πιθανών διαφορών στο αντίστοιχο επίπεδο ηχητικής πίεσης. Έπειτα, εξήγαγαν τα χαρακτηριστικά για το σύνολο του dataset χρησιμοποιώντας το MIR Toolbox στο Matlab.

Μετά το πέρας της διαδικασίας εξαγωγής των χαρακτηριστικών δημιουργήθηκε μία σύνοψη των στατιστικών μέτρων κάθε χαρακτηριστικού με σκοπό να μπορέσει να γίνει μία περιγραφή των διακυμάνσεων που έχουν οι τιμές ανά καρτέ. Σαν αποτέλεσμα αυτή η διαδικασία είχε την δημιουργία ενός συνόλου που αποτελούνταν από 230 χαρακτηριστικά και χρησιμοποιήθηκαν οι στατιστικές μετρήσεις : Mean, Standard Deviation (Std), Flatness, ,Spread, Kurtosis, Skewness, Slope και Centroid.

Όσον αφορά τα τεστ ακρόασης πραγματοποιήθηκαν σε ένα γραφείο εν ώρα εργασίας έχοντας ως στόχο την πιο ρεαλιστική απόδοση του αποτελέσματος και η αναπαραγωγή του ήχου έγινε μέσω ενός συμβατικού φορητού υπολογιστή. Οι συμμετέχοντες χωρίστηκαν σε δύο κατηγορίες. Σε αυτούς που είχαν ευχέρεια στην χρήση διαφόρων συσκευών και στην δεύτερη κατηγορία εντάχθηκαν άνθρωποι ηλικίας άνω των 70 ετών με σκοπό την μικρότερη τριβή με τέτοιου είδους καταναλωτικές συσκευές.

Κάθε συμμετέχοντας έπρεπε να ακούσει τους ήχους και να τους εντάξει με βάση την προσωπική του εκτίμηση σε κάποια από τις δύο κατηγορίες. Είτε στους ήχους έναρξης είτε στους ήχους λήξης. Αυτό οδήγησε σε συνολικά 1989 απαντήσεις εκ των οποίων το 67% λήφθηκε από τους συμμετέχοντες της πρώτης ομάδας ενώ το 33% από αυτούς της δεύτερης. Επιπλέον, το 65,5 % των σχολιασμών αφορούσε τους ήχους έναρξης ενώ το 35,5% αυτούς της λήξης. Παρακάτω παρουσιάζονται αναλυτικά τα αποτελέσματα.



Εικόνα 4.1: Οι σχολιασμοί των ήχων ανά ομάδα και στο σύνολο.

- **Προσέγγιση Μηχανικής Μάθησης**

Σε αυτή την προσέγγιση αρχικά αξιολογήθηκαν τα χαρακτηριστικά με βάση τον αλγόριθμο αξιολόγησης “InfoGainAttributeEval”, ο οποίος αξιολογεί την σημασία κάθε χαρακτηριστικού ξεχωριστά με την εκτίμηση του κέρδους πληροφοριών σε σχέση με την τάξη ,χρησιμοποιώντας τεχνικές εντροπίας. Επιπλέον χρησιμοποιήθηκε ο αλγόριθμος “OneRAttributeEval” ο οποίος δημιουργεί μία απλή τεχνική ταξινόμησης, ένα δέντρο αποφάσεων ενός επιπέδου, για κάθε ένα από τα χαρακτηριστικά και καταγράφει τον αριθμό των διορθωμένων ταξινομημένων σε θέσεις [39], [40]. Στην συνέχεια, εκπονήθηκε η αντίστοιχη εργασία ταξινόμησης στην οποία χρησιμοποιήθηκαν

οι αλγόριθμοι Artificial Neural System (ANS), Logistic Regressions και Sequential Minimal Optimization (SMO).

Η συνολική απόδοση των αλγορίθμων που χρησιμοποιήθηκαν στην Μηχανική μάθηση (ANS, Logistic Regression, SMO) ήταν αρκετά υψηλή με χαμηλότερο ποσοστό το 85,37% όταν χρησιμοποιήθηκε η Logistic Regression και υψηλότερη με ποσοστό 97,56% όταν χρησιμοποιήθηκε το ANS.

Τα αποτελέσματα όλων των παραπάνω συνοψίζονται στους παρακάτω πίνακες.

|                             | “Switch ON” | “Switch OFF” |
|-----------------------------|-------------|--------------|
| <b>1<sup>st</sup> Group</b> | 61,26 %     | 51,07 %      |
| <b>2<sup>nd</sup> Group</b> | 61,57 %     | 45,98 %      |
| <b>Average</b>              | 61,42 %     | 48,52 %      |

Εικόνα 4.2: Ακρίβεια ταξινόμησης του τεστ ακρόασης με την εκδήλωση FMC των ήχων

| <b>ANS</b>                 |                             |                          |                           |
|----------------------------|-----------------------------|--------------------------|---------------------------|
| <b>N. of Folds</b>         | <b>PR<sub>overall</sub></b> | <b>PR<sub>open</sub></b> | <b>PR<sub>close</sub></b> |
| 3-fold                     | 92,68 %                     | 92,59 %                  | 92,86 %                   |
| 7 fold                     | 95,12 %                     | 92,59 %                  | 100 %                     |
| One-Out                    | 97,56 %                     | 96,30 %                  | 100 %                     |
| <b>Logistic Regression</b> |                             |                          |                           |
| 3 fold                     | 85,37 %                     | 85,19 %                  | 85,71 %                   |
| 7 fold                     | 85,37 %                     | 85,19 %                  | 85,71 %                   |
| One-Out                    | 90,24 %                     | 92,59 %                  | 85,71 %                   |
| <b>SMO</b>                 |                             |                          |                           |
| 3 fold                     | 90,24 %                     | 88,89 %                  | 92,86 %                   |
| 7 fold                     | 90,24 %                     | 96,30 %                  | 78,57 %                   |
| One-Out                    | 95,12 %                     | 96,30 %                  | 92,86 %                   |

Εικόνα 4.3: Αποτελέσματα πειραμάτων με χρήση Μηχανικής Μάθησης

Η παραπάνω έρευνα χρησιμοποιεί μία προσέγγιση παρόμοια με αυτή που ακολουθήθηκε στην παρούσα πτυχιακή εργασία. Μία άλλη έρευνα που πραγματοποιήθηκε στο Taiwan με τίτλο “An efficient emotion detection scheme for popular music” [41] χρησιμοποιεί μία διαφορετική διαδικασία ομαδοποίησης μοντέλων, την Gaussian Mixture. Πιο συγκεκριμένα, η βάση δεδομένων αποτελείται από 145 μουσικά κλιπ με ρυθμό δειγματοληψίας τα 44,1 KHz και σαν μοντέλο συναισθημάτων χρησιμοποιείται ένα μοντέλο δύο διαστάσεων του Tayler ενώ για την αναγνώριση των συναισθημάτων χρησιμοποιείται και πάλι ο ανθρώπινος παράγοντας. Στην συνέχεια, εξάγονται τρία χαρακτηριστικά η ένταση, η κανονικότητα του ρυθμού και ο ρυθμός προκειμένου να περιγράψουν ένα μουσικό κλιπ τα οποία εκπαιδεύονται με την κατασκευή Gaussian Mixture Models (GMM). Εφόσον χρησιμοποιήθηκε το μοντέλο δύο διαστάσεων του Tayler κατασκευάστηκαν δύο GMM. Αρχικά, το Arousal GMM (κλάση

1 και 2 ) χρησιμοποιήθηκε για να κατηγοριοποιήσει ένα δοκιμαστικό κλιπ σε υψηλή διέγερση ή χαμηλή διέγερση, ενώ το valence GMM (κλάση 3 και 4) προκειμένου να διακρίνει την ακριβή τάξη των συναισθημάτων. Έπειτα, εξάγονται τέσσερις τιμές για τις τέσσερις κατηγορίες συναισθημάτων. Η κάθε τιμή υποδηλώνει την πιθανότητα μίας κατηγορίας συναισθημάτων. Η κατηγορία με τη μέγιστη τιμή αποτελεί το συναίσθημα του δοκιμαστικού μουσικού κλιπ. Τα αποτελέσματα της παραπάνω έρευνας παρουσιάζονται συνοπτικά στον παρακάτω πίνακα.

| <b>%</b>        | <b>CLASS 1</b> | <b>CLASS 2</b> | <b>CLASS 3</b> | <b>CLASS 4</b> |
|-----------------|----------------|----------------|----------------|----------------|
| <b><i>R</i></b> | 83.33          | 85.71          | 71.43          | 80             |
| <b><i>P</i></b> | 71.43          | 85.71          | 83.33          | 80             |

Εικόνα 4.4: Precision και recall του προτεινόμενου αλγόριθμου

## 4.4 Επίλογος

Στο συγκεκριμένο κεφάλαιο έγινε μία αναφορά στο πρόβλημα που προκύπτει κατά την διαδικασία της κατηγοριοποίησης του ήχου και παρουσιάστηκαν διάφορες έρευνες σχετικές με το αντικείμενο που μελετάται.

## **ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ**

## Πρόλογος

Σε αυτό το μέρος παρουσιάζεται το πειραματικό κομμάτι της πτυχιακής εργασίας. Το πείραμα αφορούσε την κατηγοριοποίηση των συναισθημάτων που προκύπτουν έπειτα από την αναπαραγωγή ήχου από κάποια καταναλωτική συσκευή χρησιμοποιώντας μεθόδους της μηχανικής μάθησης. Στα κεφάλαια που ακολουθούν αναλύεται η διαδικασία και τα κριτήρια συλλογής ηχητικών δεδομένων καθώς και τα βήματα που ακολουθήθηκαν προκειμένου να γίνει εξαγωγή των χαρακτηριστικών από το Matlab και η αξιολόγηση της συνεισφοράς τους στο τελικό αποτέλεσμα. Τέλος, παρουσιάζονται αναλυτικά όλοι οι αλγόριθμοι κατηγοριοποίησης και ακολουθεί ο σχολιασμός των αποτελεσμάτων και των σφαλμάτων τους.

## Κεφάλαιο 5<sup>ο</sup> :Συλλογή και προεπεξεργασία δεδομένων

### 5.1 Εισαγωγή

Στο παρόν κεφάλαιο διατυπώνεται το αρχικό στάδιο του πειράματος που είναι η συλλογή και η επεξεργασία των δεδομένων. Πιο συγκεκριμένα, αναλύεται όλη η μεθοδολογία που ακολουθήθηκε προκειμένου να δημιουργηθεί το σύνολο δεδομένων μας (dataset) καθώς και όλες οι σκέψεις και οι επιλογές που ακολουθήσαμε προκειμένου να γίνει η κατάτμηση σε παράθυρα προκειμένου να καταστούν κατάλληλα για την εξαγωγή των χαρακτηριστικών.

### 5.2 Συλλογή και προεπεξεργασία δεδομένων

Κατά την διάρκεια της συλλογής δεδομένων χρησιμοποιήθηκαν τρεις κατηγορίες ήχων για την διεξαγωγή του πειράματος. Χρησιμοποιήσαμε ήχους που αναπαράγουν οι μηχανές για την ενημέρωση του χρήστη και τους εντάξαμε εμπειρικά σε 3 κατηγορίες error, opening και warning.

Στην πρώτη κατηγορία, στην κατηγορία error, επιλέχθηκαν ήχοι προερχόμενοι από ηλεκτρονικά συστήματα όπως αυτοί των windows ή του PlayStation, οι οποίοι ακούγονται στις περισσότερες περιπτώσεις όταν κάποιος χρήστης εκτελέσει έναν λάθος χειρισμό. Επιπλέον, ενσωματώθηκαν και ήχοι αυτοκινήτων και ανυψωτικών οχημάτων οι οποίοι αναπαράγονται κατά τον εντοπισμό κάποιου προβλήματος.

Στην δεύτερη κατηγορία, την κατηγορία opening, ενσωματώθηκαν ήχοι από startup notifications διαφόρων εταιρειών που κατασκευάζουν κινητά τηλέφωνα καθώς και ποικίλων εφαρμογών όπως messenger Facebook, Viber και Instagram. Η αναπαραγωγή των συγκεκριμένων γίνεται κατά την είσοδο του χρήστη σε κάποια πλατφόρμα ή με την ενεργοποίηση κάποιας συσκευής.

Στην τρίτη και τελευταία κατηγορία, την κατηγορία warning συγκαταλέχθηκαν ήχοι από ανιχνευτές καπνού, ζώνης αυτοκινήτου, συναγερμοί αυτοκινήτου και σπιτιού όπως επίσης και ήχοι που αναπαράγονται από την κινητή συσκευή όταν η μπαταρία είναι έτοιμη να εξαντληθεί.

Η ταξινόμηση των ήχων στηρίχθηκε σε ένα dataset το οποίο αποτελείται από 117 ήχους οι οποίοι προήλθαν είτε από διάφορα dataset τα οποία ήταν διαθέσιμα στο διαδίκτυο είτε από προσωπικές μας ηχογραφήσεις. Ωστόσο, και στις δύο περιπτώσεις προσπαθήσαμε οι ήχοι που συλλέξαμε να είναι στην καλύτερη δυνατή ποιότητα, έτσι ώστε να μην επηρεαστούν οι μετρήσεις μας για παράδειγμα από

περιβαλλοντικούς θορύβους. Επίσης, έγινε μετατροπή όπου αυτό ήταν απαραίτητο σε μορφή wav με ρυθμό δειγματοληψίας τα 44.100 Hz.

Στο τελευταίο στάδιο της προεπεξεργασίας δεδομένων έγινε ο διαχωρισμός του κάθε αρχείου ήχου σε παράθυρα (segmentation). Πιο συγκεκριμένα, το κάθε αρχείο ήχου διασπάστηκε σε διακριτά μέρη προκειμένου από κάθε μέρος να μπορέσουν να εξαχθούν τιμές από τα χαρακτηριστικά που επιλέξαμε και να χρησιμοποιηθούν ως είσοδοι στους αλγορίθμους μηχανικής μάθησης που θα αναλυθούν παρακάτω. Η κατάτμηση έγινε σε παράθυρα τριών καθώς και ενός δευτερολέπτου. Σε αυτό το σημείο αξίζει να αναφερθεί πως όσο μικρότερα είναι τα παράθυρα τόσο αυξάνεται ο όγκος των δεδομένων. Περισσότερες πληροφορίες για την διαδικασία της κατάτμησης αναλύονται σε παρακάτω κεφάλαιο.

### 5.3 Ηχητικά Χαρακτηριστικά

Κάθε ήχος αποτελείται από κάποια χαρακτηριστικά. Στην συγκεκριμένη πειραματική εργασία και έπειτα από μελέτη της διαθέσιμης βιβλιογραφίας προτιμήθηκε η εξαγωγή συμπερασμάτων να γίνει με βάση τα χαρακτηριστικά Mel-frequency cepstrum (Mfcc), Zero crossing rate, Spectral rolloff, Spectral centroid, Spread, Skewness, Kurtosis κ.α. τα οποία θα αναλυθούν εκτενέστερα παρακάτω. Κατά την γνώμη μας τα συγκεκριμένα χαρακτηριστικά μπορούν να μας οδηγήσουν σε ένα ασφαλές συμπέρασμα.

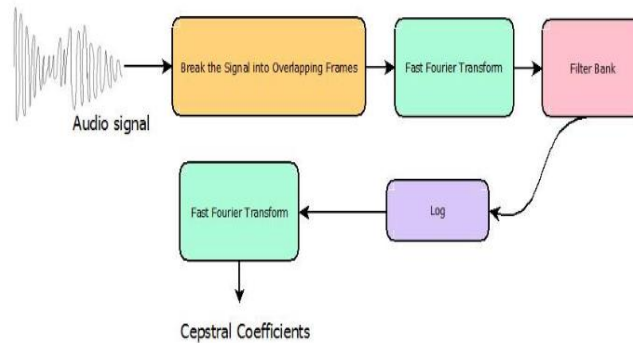
#### Mel-Frequency cepstrum (Mfcc):

Στην επεξεργασία ήχου, το mel-frequency cepstrum (mfc) προσφέρει μία περιγραφή του φασματικού σχήματος του ήχου. Αντίστοιχα, οι εγκεφαλικοί συντελεστές συχνότητας Mel (MFCCs) είναι συντελεστές οι οποίοι αποτελούν συλλογικά ένα MFC. Τα MFCC χρησιμοποιούνται όλο και περισσότερο σε εφαρμογές ανάκτησης μουσικών πληροφοριών όπως είναι η ταξινόμηση ειδών<sup>10</sup>.

Στα παλαιότερα χρόνια κάποιοι ερευνητές ενσωμάτωσαν τους συντελεστές MFCC προκειμένου να μπορούν να προσδιορίσουν το παραγλωσσικό περιεχόμενο. Ωστόσο, κατάφεραν να αναγνωρίσουν επαρκώς μόνο τρία συναισθήματα [42] και άλλα τέσσερα με χαμηλή ακρίβεια αναγνώρισης [43]. Οι εγκεφαλικοί συντελεστές συχνότητας Mel (MFCCs) είναι συντελεστές οι οποίοι αθροιστικά αποτελούν ένα MFC και πηγάζουν από έναν τύπο εγκεφαλικής αναπαράστασης ενός ηχητικού κλιπ. Η ειδοποιός διαφορά τους συναντάται στο γεγονός ότι στο MFC, οι ζώνες συχνοτήτων ισαπέχουν στην κλίμακα Mel, η οποία προσεγγίζει την απόκριση του ανθρώπινου ακουστικού συστήματος πιο στενά από τις γραμμικά αποστασιοποιημένες ζώνες συχνοτήτων που χρησιμοποιούνται στο κανονικό φάσμα. Χάρη σε αυτό, επιτρέπεται η καλύτερη αναπαράσταση του ήχου. Τα MFCC προκύπτουν [44]:

<sup>10</sup> [https://en.wikipedia.org/wiki/Mel-frequency\\_cepstrum](https://en.wikipedia.org/wiki/Mel-frequency_cepstrum)

- Παίρνοντας τον μετασχηματισμό Fourier ενός παραθυρικού αποσπάσματος του σήματος
- Έπειτα, αντιστοιχίζουμε τις δυνάμεις του φάσματος που λήφθηκαν παραπάνω στην κλίμακα Mel, χρησιμοποιώντας τριγωνικά επικαλυπτόμενα παράθυρα ή επικαλυπτόμενα παράθυρα συνημιτόνου.
- Παίρνουμε τα αρχεία καταγραφής των δυνάμεων για κάθε μία από τις συχνότητες Mel.
- Εν συνεχεία, χρησιμοποιούμε τον διακριτό μετασχηματισμό συνημιτόνου της λίστας δυνάμεων mel log, και την διαχειριζόμαστε σαν να ήταν ένα σήμα
- Τέλος, τα πλάτη του φάσματος που προκύπτει, είναι τα MFCC.

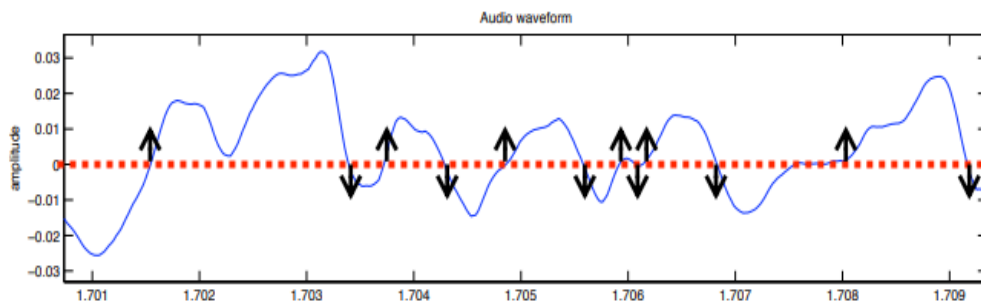


Εικόνα 5.1: Πως φτάσαμε στο mfcc

### Zero crossing rate:

Όταν αναφερόμαστε στον μηδενισμό στην ουσία εντοπίζουμε την στιγμή κατά την οποία δύο διαδοχικά στιγμιότυπα (samples) έχουν διαφορετικά πρόσημα. Επομένως, οι μηδενισμοί επίσης αποτελούν μέτρο που περιγράφει την συχνότητα. Όταν αναφερόμαστε στα απλά σήματα ο ρυθμός μηδενισμού (zero crossing rate) συσχετίζεται άμεσα με τη θεμελιώδη συχνότητα  $f_0$ . Η συχνότητα, μπορεί να υπολογιστεί αν διαιρέσουμε το πλήθος μηδενισμού με το 2. Αυτό συμβαίνει διότι το ημιτονοειδές σήμα τέμνει τον άξονα του χρόνου δύο φορές σε κάθε κύκλο. Ωστόσο υπάρχουν περιπτώσεις όπου τα σήματα μπορεί να τέμνουν τον άξονα του χρόνου περισσότερες από δύο φορές περιπλέκοντας έτσι την κατά τα άλλα απλή διαδικασία υπολογισμού που προαναφέρθηκε. Ενώ αντίστοιχα ο μαθηματικός του υπολογισμός προκύπτει από την σχέση:

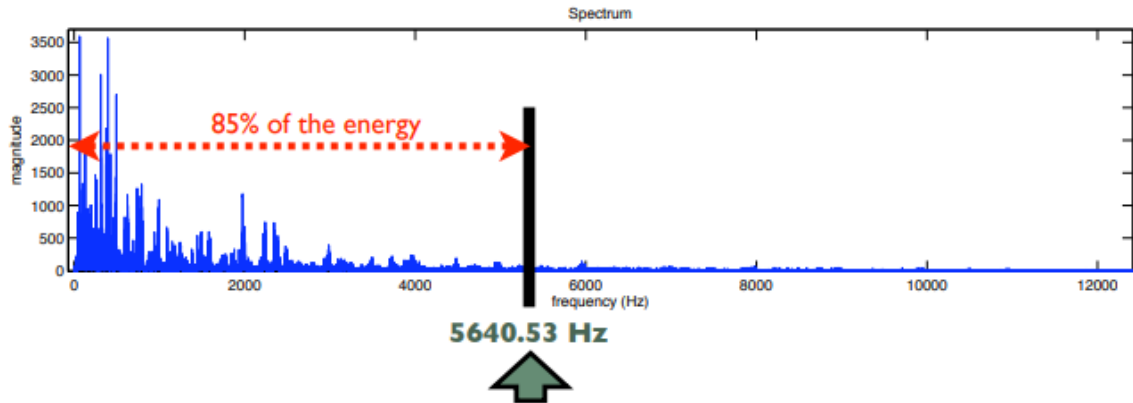
$$zcr = \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbf{1}_{\mathbb{R}<0}(s_t s_{t-1})$$



Εικόνα 5.2: Ρυθμός μηδενικού επιπέδου. Η πορτοκαλί γραμμή είναι το πλάτος και η μπλε γραμμή το ηχητικό σήμα.

**Spectral roll-off:**

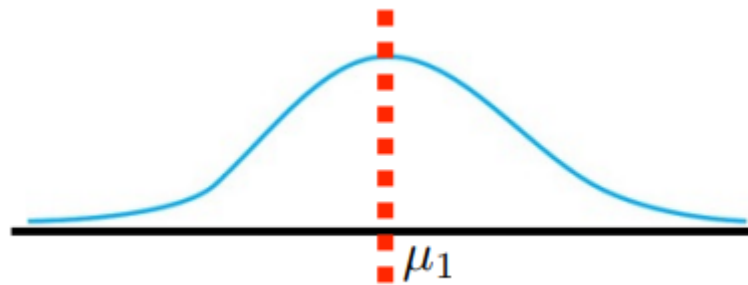
Είναι ένα μέτρο του σχήματος του σήματος. Αντιπροσωπεύει τη συχνότητα κάτω από την οποία βρίσκεται ένα καθορισμένο ποσοστό της συνολικής φασματικής ενέργειας. Προκειμένου να υπολογιστεί η συχνότητα roll-off για κάθε πλαίσιο σε ένα σήμα χρησιμοποιείται η συνάρτηση `librosa.feature.spectral_rolloff`<sup>11</sup>.



Εικόνα 5.3: Το 85% της ενέργειας είναι συγκεντρωμένο κάτω από την συχνότητα 5640.53Hz

**Spectral Centroid**<sup>12</sup>:

Ο φασματικός μέσος όρος είναι το κέντρο «βαρύτητας» του φάσματος ενός σήματος. Το φασματικό κέντρο είναι ένας καλός προγνωστικός δείκτης της «φωτεινότητας» [45] ενός ήχου και είναι ευρέως διαδεδομένη η χρήση του στην ψηφιακή επεξεργασία ήχου και μουσικής ως αυτόματη μέτρηση της μουσικής χροιάς.



Εικόνα 5.4: Μέση τιμή φάσματος

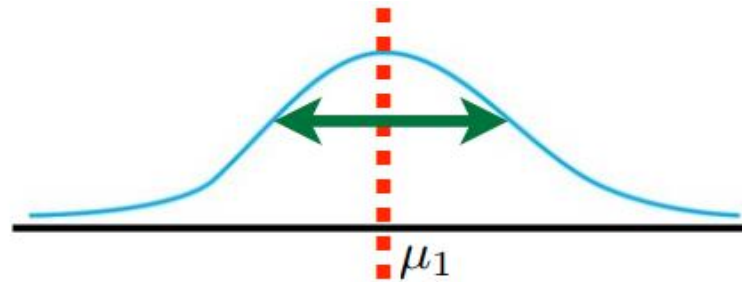
<sup>11</sup> [https://librosa.org/doc/main/generated/librosa.feature.spectral\\_rolloff.html](https://librosa.org/doc/main/generated/librosa.feature.spectral_rolloff.html)

<sup>12</sup> [https://en.wikipedia.org/wiki/Spectral\\_centroid](https://en.wikipedia.org/wiki/Spectral_centroid)

**Spread:**

Λόγω του ότι είναι η τετραγωνική απόκλιση της τυχαίας μεταβλητής από τη μέση τιμή της, η διακύμανση είναι πάντα θετική και είναι ένα μέτρο της διασποράς ή της εξάπλωσης της κατανομής. Η τετραγωνική ρίζα της διακύμανσης ονομάζεται τυπική απόκλιση και είναι ιδιαίτερα σημαντική για την περιγραφή της φύσης της κατανομής διότι έχει τις ίδιες μονάδες με την τυχαία μεταβλητή.

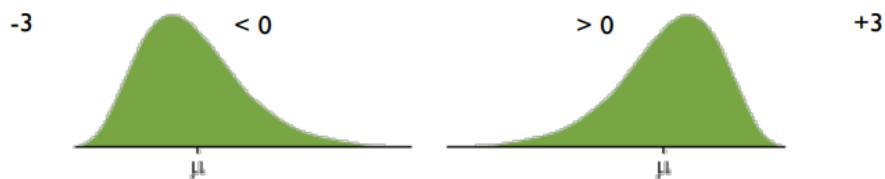
$$\sigma^2 = \mu_2 = \int (x - \mu_1)^2 f(x) dx$$



Εικόνα 5.5: Τυπική Απόκλιση Σήματος

**Skewness:**

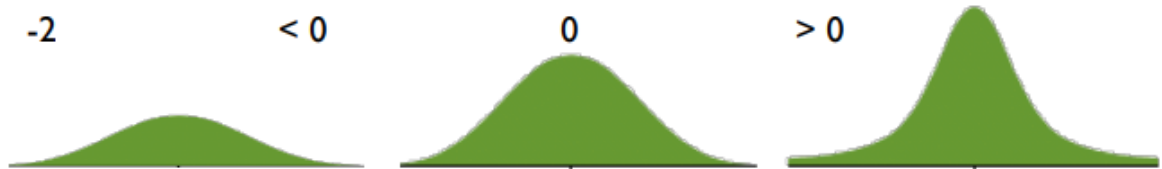
Μετράει την συμμετρία γύρω από το κεντροειδές. Η φασματική κλίση χρησιμοποιείται σε συνδυασμό με άλλες φασματικές ροπές προκειμένου να διακρίνει τον τύπο της άρθρωσης. Για αρμονικά σήματα προσδιορίζει τη σχετική ισχύ των υψηλότερων και χαμηλότερων αρμονικών.



Εικόνα 5.6: Στο πρώτο σχήμα οι τιμές είναι μαζεμένες στα αριστερά του μέσου όρου ενώ στο δεύτερο στα δεξιά.

**Kurtosis:**

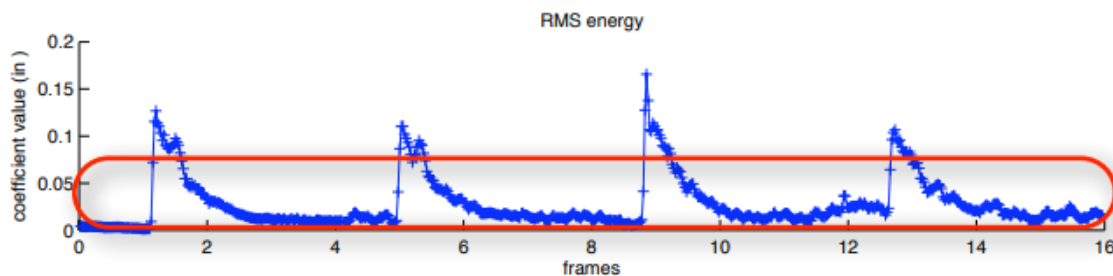
Η φασματική κύρτωση μετρά την επιπεδότητα ή μη σε σχέση με την καμπύλη Gauss του φάσματος γύρω από το κέντρο. Από την άλλη χρησιμοποιείται για να δείξει την κορύφωση ενός φάσματος. Παραδείγματος χάριν, όσο ο λευκός θόρυβος αυξάνεται στο σήμα της ομιλίας τόσο η κύρτωση μειώνεται υποδεικνύοντας έτσι ένα λιγότερο μέγιστο φάσμα.



Εικόνα 5.7: Χαμηλή, κανονική και υψηλή κύρτωση

**Low Energy:**

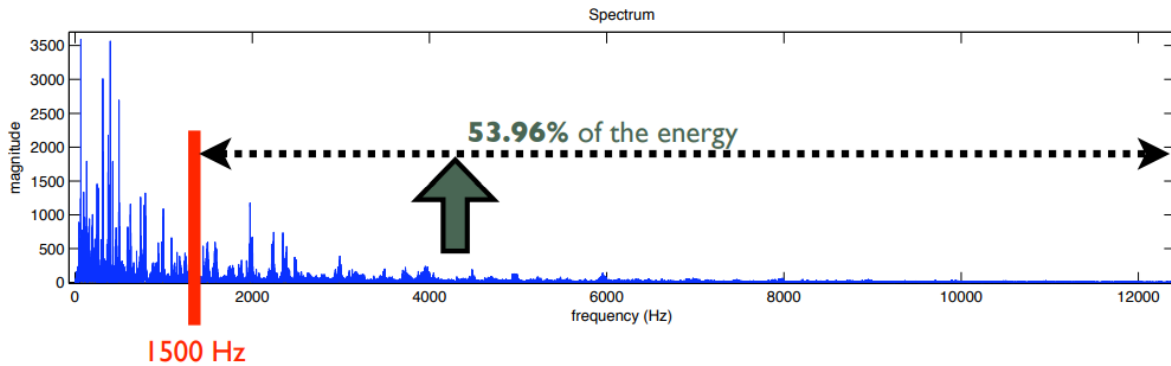
Η καμπύλη ενέργειας χρησιμοποιείται προκειμένου να ληφθεί μία εκτίμηση της χρονικής κατανομής της ενέργειας για να δούμε αν παραμένει σταθερό σε όλο το ηχητικό σήμα ή αν ορισμένα πλαίσια έχουν μεγαλύτερη αντίφαση σε σχέση με άλλα. Ένας τρόπος για να εκτιμηθεί αυτό είναι ο υπολογισμός του χαμηλού ρυθμού ενέργειας δηλαδή το ποσοστό των πλαισίων που εμφανίζουν λιγότερη από την μέση ενέργεια[45].



Εικόνα 5.8: Τα περισσότερα από τα πλαίσια είναι κάτω από το μέσο όρο RMS

**Brightness:**

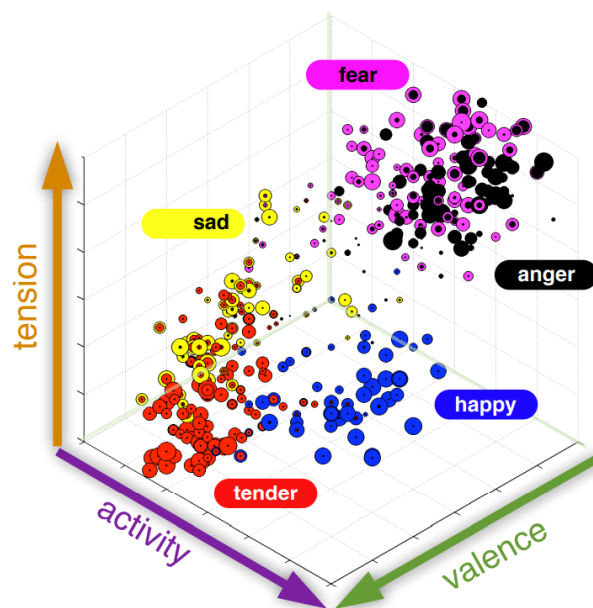
Συνίσταται στον καθορισμό της συχνότητας αποκοπής και στην μέτρηση της ποσότητας της ενέργειας πάνω από αυτή τη συχνότητα[46]. Το αποτέλεσμα εκφράζεται με μία τιμή μεταξύ του 0 και του 1 και χρησιμοποιείται ευρέως σε αντικείμενα miraudio.



Εικόνα 5.9: Το 53,96 % της ενέργειας είναι συγκεντρωμένο πάνω από την συχνότητα 1500Hz

**Emotion:**

Προσπαθεί να προβλέψει μία περιγραφή του συναισθήματος με βάση την ανάλυση του ήχου και του μουσικού περιεχομένου των ηχογραφήσεων. Η έξοδος του miremotion είναι ο υποκειμενικός εντοπισμός του συναισθηματικού περιεχομένου εντός των 5 βασικών τάξεων και εντός των 3 διαστάσεων. Οι τιμές που παίρνει είναι μεταξύ του 1 και του 7 όπου το 1 αντιστοιχεί σε πολύ χαμηλή τιμή και η 7 σε πολύ υψηλή. Οι παραπάνω τιμές βασίζονται στις συναισθηματικές βαθμολογίες των ακροατών και συγκεντρώθηκαν με τον ίδιο ακριβώς τρόπο, καθώς αντιστοιχεί σε μία κλίμακα Libert που χρησιμοποιείται στον τομέα της πειραματικής ψυχολογίας[47].



Εικόνα 5.10: Σύγκριση μεταξύ των βασικών εννοιών των συναισθημάτων(happy,sad,tender,anger,fear) και συναισθηματικών διαστάσεων (activity,valence,tension)

## 5.4 Κατάτμηση και εξαγωγή χαρακτηριστικών

Στην παρακάτω εικόνα (εικόνα 5.11) παρουσιάζεται η αλγοριθμική λογική που χρησιμοποιήθηκε για κατάτμηση σε παράθυρο 1 δευτερολέπτου. Κάθε ηχητικό χαρακτηριστικό για κάθε ένα από τα αρχεία ήχου που επιλέχθηκαν εξήχθη στο εκάστοτε παράθυρο. Εν συνεχεία, οι τελικές τιμές των χαρακτηριστικών που υπολογίστηκαν αποθηκεύτηκαν σε ένα xls αρχείο.

```

1
2
3   dirName=pwd;
4   files = dir( fullfile(dirName, '*.wav') );
5   win=1;
6   k=1;
7
8   for f=1:length(files)
9       fullfileName = fullfile(dirName, files(f).name)
10      duration=mirgetdata(mirlength(miraudio(fullfileName)));
11      for i=0:ceil((duration/win)-1)
12          audio=miraudio(fullfileName, 'Extract', i*win, (i+1)*win);
13
14          AF(k,1)=mirgetdata(mirLowenergy(audio));
15          AF(k,2)=mirgetdata(mirzerocross(audio));
16          AF(k,3)=mirgetdata(mirrolloff(audio, 'Threshold', 0.3));
17          AF(k,4)=mirgetdata(mirrolloff(audio, 'Threshold', 0.5));
18          AF(k,5)=mirgetdata(mirrolloff(audio, 'Threshold', 0.7));
19          AF(k,6)=mirgetdata(mirrolloff(audio, 'Threshold', 0.9));
20          AF(k,7)=mirgetdata(mircentroid(audio));
21          AF(k,8)=mirgetdata(mirspread(audio));
22          AF(k,9)=mirgetdata(mirskewness(audio));
23          AF(k,10)=mirgetdata(mirkurtosis(audio));
24
25          %statistics
26
27
28          mfccs=mirgetdata(mirmfcc(audio));
29          AF(k,11)= mfccs(1);
30          AF(k,12)= mfccs(2);
31          AF(k,13)= mfccs(3);
32          AF(k,14)= mfccs(4);
33          AF(k,15)= mfccs(5);
34          AF(k,16)=mfccs(6);
35          AF(k,17)=mfccs(7);
36          AF(k,18)=mfccs(8);
37          AF(k,19)=mfccs(9);
38          AF(k,20)=mfccs(10);
39          AF(k,21)=mfccs(11);
40          AF(k,22)=mfccs(12);
41          AF(k,23)=mfccs(13);
42
43          AF(k,24)=mirgetdata(mirbrightness(audio, 'Cutoff', 500));
44          AF(k,25)=mirgetdata(mirbrightness(audio, 'Cutoff', 1000));
45
46          happy=mirgetdata(miremotion(audio), 'Happy');
47          sad=mirgetdata(miremotion(audio), 'Sad');
48          tender=mirgetdata(miremotion(audio), 'Tender');
49          anger=mirgetdata(miremotion(audio), 'Anger');
50          fear=mirgetdata(miremotion(audio), 'Fear');
51          AF(k,26)=happy(1);
52          AF(k,27)=sad(2);
53          AF(k,28)=tender(3);
54          AF(k,29)=anger(4);
55          AF(k,30)=fear(5);
56
57          k=k+1;
58      end
59  end
60
61  xlswrite('ALL_WARNINGcorrectonesecond.xls', AF);
62
63

```

Εικόνα 5.11: Κώδικας που εισήχθη στο Matlab

Μετά την εξαγωγή των χαρακτηριστικών στο αρχείο xls συγχωνεύτηκαν σε ένα ενιαίο αρχείο το οποίο ήταν κατάλληλο για επεξεργασία αλλά και ταξινόμηση τόσο στο Matlab όσο και στο Weka. Στο συγκεκριμένο αρχείο, μέρος αυτού παρουσιάζεται και στην εικόνα παρακάτω, στην πρώτη γραμμή δόθηκαν τα ονόματα των χαρακτηριστικών ενώ στο τέλος κάθε γραμμής υπάρχει η επισήμειωση (label) που καθορίζει την κατηγορία του ήχου. Επαναλήφθηκε η ίδια διαδικασία και για τα δύο είδη κατάτμησης (3 δευτερολέπτων και 1 δευτερολέπτου). Στο αρχείο με την κατάτμηση του ήχου ανά 3 δευτερόλεπτα υπήρχαν 117 γραμμές πέρα από την αρχική των επικεφαλίδων ενώ στο αρχείο του ενός δευτερολέπτου 351, το οποίο είναι λογικό αφού ο κάθε ήχος διαιρέθηκε σε τρία μέρη.

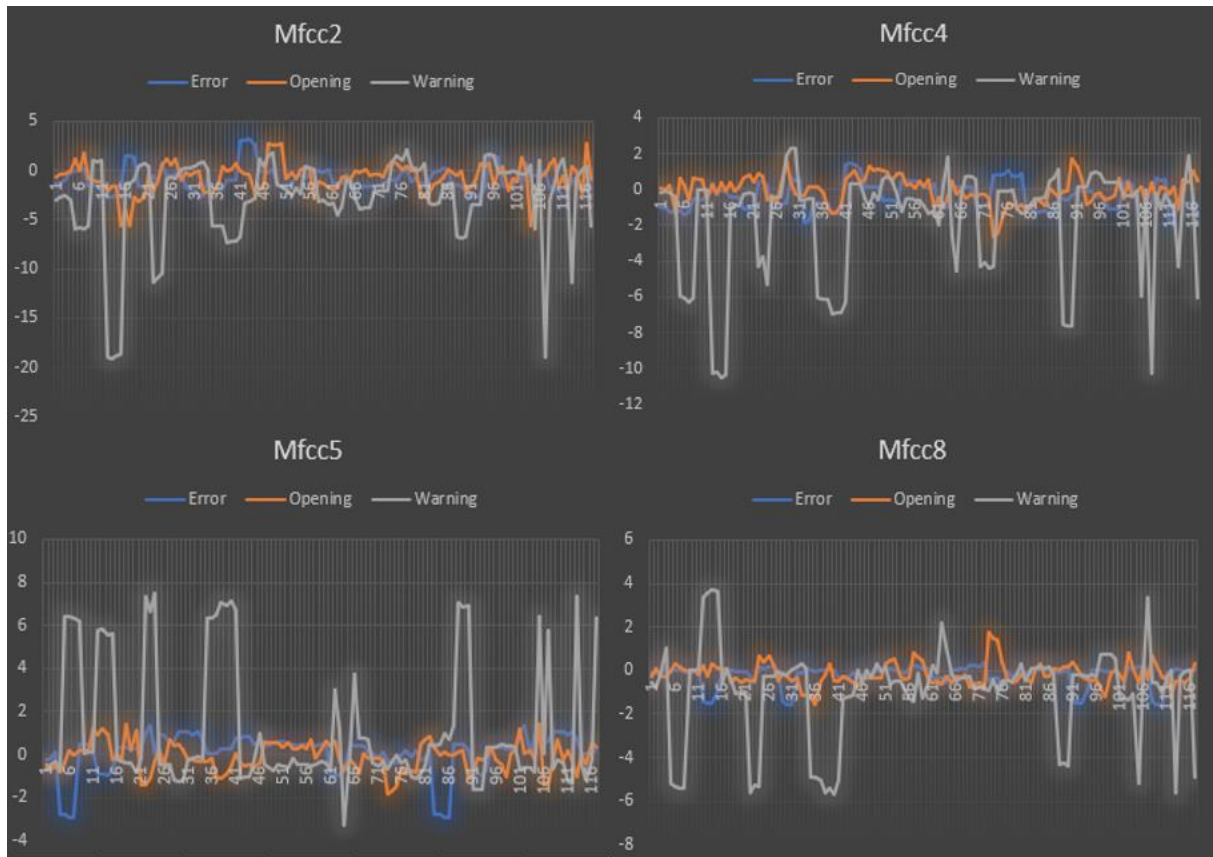
| Mfcc10       | Mfcc11       | Mfcc12       | Mfcc13       | Br500       | Br1000      | Happy       | Sad          | Tender      | Anger       | Fear        | Class |
|--------------|--------------|--------------|--------------|-------------|-------------|-------------|--------------|-------------|-------------|-------------|-------|
| 0.125493638  | 0.289028176  | -0.087361763 | -0.076153643 | 0.2172383   | 0.098708373 | 0.056122123 | 0.197984372  | 0.268768023 | 0.267359219 | 0.258292752 | Error |
| 0.115189087  | 0.333865434  | -0.117799721 | -0.232128105 | 0.271530234 | 0.12938489  | 0.27608572  | 0.292050305  | 0.279309338 | 0.258711751 | 0.261282059 | Error |
| 0.071281199  | 0.371394635  | -0.126500346 | -0.287490981 | 0.329374113 | 0.153454791 | 0.233504749 | 0.214735512  | 0.195489802 | 0.173115093 | 0.13495493  | Error |
| 0.088660689  | 0.243265362  | -0.093636835 | -0.023256742 | 0.194978256 | 0.084072272 | 0.154426922 | 0.155976941  | 0.126162613 | 0.105652659 | 0.096466541 | Error |
| 0.228392484  | 0.00763354   | 0.088131282  | 0.011599102  | 0.837573588 | 0.308606903 | 0.100815929 | 0.089627936  | 0.082956008 | 0.077223519 | 0.061943421 | Error |
| 0.219965006  | 0.007456724  | 0.089791207  | 0.015086159  | 0.837164542 | 0.309337989 | 0.00049501  | 0.000352222  | 0.000215249 | 0.012195433 | 0.084344186 | Error |
| 0.282057504  | 0.034198354  | 0.076801636  | 0.000870264  | 0.835073968 | 0.307417678 | 0.002006932 | 0.001253826  | 0.001104475 | 0.000936099 | 0.000744746 | Error |
| 0.207865226  | -0.069550117 | 0.021070583  | 0.106179891  | 0.564376035 | 0.39555385  | 0.008200611 | 0.004716459  | 0.003721794 | 0.003768245 | 0.00321521  | Error |
| -0.147149206 | -0.041642133 | -0.061800546 | -0.418369267 | 0.688867433 | 0.472367795 | 0.197649192 | 0.174474424  | 0.155867634 | 0.153553902 | 0.146945125 | Error |
| -0.147243495 | -0.040494809 | -0.063357034 | -0.417434618 | 0.688972667 | 0.472621628 | 0.197759884 | 0.174261338  | 0.15589059  | 0.153361774 | 0.146783932 | Error |
| -0.144342926 | -0.041763532 | -0.063352624 | -0.418058747 | 0.689297479 | 0.47323607  | 0.198425428 | 0.173898129  | 0.155222098 | 0.153347189 | 0.146614254 | Error |
| -0.155867912 | -0.031508565 | -0.069301444 | -0.414471127 | 0.689312238 | 0.473341861 | 0.198523372 | 0.173228305  | 0.155320281 | 0.15300756  | 0.147028626 | Error |
| 0.396838314  | 0.010259293  | -0.183665131 | -0.290160564 | 0.894240242 | 0.544992238 | 0.059559034 | 0.048804566  | 0.043892891 | 0.0415903   | 0.053570621 | Error |
| 0.360991855  | 0.034426651  | -0.172603098 | -0.261109729 | 0.893003518 | 0.525434543 | 0.042257328 | 0.042742427  | 0.057662473 | 0.057662473 | 0.056884664 | Error |
| 0.267415248  | 0.04069      | -0.151780517 | -0.203057178 | 0.89069395  | 0.486038684 | 0.041537963 | 0.054164573  | 0.060058117 | 0.051560037 | 0.043877878 | Error |
| 0.415177022  | 0.209742812  | -0.112946423 | -0.406637578 | 0.895013618 | 0.585190195 | 0.057903895 | 0.055991264  | 0.047385148 | 0.039449209 | 0.030527913 | Error |
| -0.093859632 | -0.935862512 | 0.201708507  | 0.177904635  | 0.539377487 | 0.369416724 | 0.039470724 | 0.051127811  | 0.054415532 | 0.056430349 | 0.074586371 | Error |
| -0.131072012 | -0.88284809  | -0.062693604 | -0.176795849 | 0.610371483 | 0.520343464 | 0.043534699 | 0.0493704998 | 0.099744865 | 0.093730604 | 0.085961099 | Error |
| -0.10853117  | -0.906629377 | 0.264707608  | 0.172924028  | 0.603495808 | 0.432998695 | 0.056888646 | 0.046396492  | 0.038232949 | 0.029891545 | 0.024305278 | Error |
| -0.116422962 | -0.561246795 | -0.113318615 | 0.577598485  | 0.265250127 | 0.129785913 | 0.022195891 | 0.022039863  | 0.021729792 | 0.020880512 | 0.019532894 | Error |
| 0.096014351  | 0.087383176  | 0.066431223  | 0.101526085  | 0.127665707 | 0.049742656 | 0.32168338  | 0.265179269  | 0.191203007 | 0.246516442 | 0.331644002 | Error |
| -0.061615985 | 0.029021517  | 0.026659271  | 0.010832161  | 0.514598877 | 0.300954063 | 0.043534699 | 0.041939075  | 0.041737633 | 0.046657602 | 0.048069419 | Error |
| -0.015219061 | -0.036442018 | -0.003164466 | 0.040741433  | 0.288001171 | 0.184726038 | 0.29867739  | 0.394553807  | 0.334952229 | 0.110885259 | 0.067316368 | Error |
| -1.794382538 | 0.896737122  | -0.13007916  | 0.411376806  | 0.971802007 | 0.43074701  | 0.200429111 | 0.193806058  | 0.061089635 | 0.024161849 | 0.007261833 | Error |
| -1.72475048  | 0.86040639   | -0.121066522 | 0.390815193  | 0.967367434 | 0.46143064  | 0.010542477 | 0.014467762  | 0.01299295  | 0.012884331 | 0.01390013  | Error |
| -1.629037744 | 0.709589198  | -0.005469231 | 0.409370279  | 0.978480515 | 0.383209339 | 0.012752397 | 0.015986855  | 0.013247435 | 0.013541442 | 0.017996234 | Error |
| 1.539058655  | 0.778352126  | -0.836824822 | 0.261743385  | 0.992130933 | 0.317989604 | 0.028657947 | 0.039343311  | 0.038056368 | 0.035547248 | 0.033245956 | Error |
| 1.579872471  | 0.835593959  | -0.870103061 | 0.269641253  | 0.990759057 | 0.327078694 | 0.180576206 | 0.258331208  | 0.274999563 | 0.334211413 | 0.36177707  | Error |
| 1.682404828  | 0.868249036  | -0.918159882 | 0.260577772  | 0.990737442 | 0.317727055 | 0.056122123 | 0.197984372  | 0.268768023 | 0.267359219 | 0.258292752 | Error |
| 1.821365623  | 1.101120344  | -0.86257429  | 0.198261874  | 0.992143908 | 0.30535003  | 0.100815929 | 0.089627936  | 0.082956008 | 0.077223519 | 0.061943421 | Error |
| -0.063208689 | 0.027433795  | 0.046829488  | -0.050460354 | 0.460131673 | 0.140865283 | 0.197649192 | 0.174474424  | 0.155867634 | 0.153553902 | 0.146945125 | Error |
| -0.05240144  | 0.00010313   | 0.003511497  | -0.022965859 | 0.485512017 | 0.172477001 | 0.059559034 | 0.048804566  | 0.043892891 | 0.0415903   | 0.053570621 | Error |
| -0.066199403 | 0.029864906  | 0.030394485  | -0.039939647 | 0.470289381 | 0.14942476  | 0.039470724 | 0.051127811  | 0.054415532 | 0.056430349 | 0.074586371 | Error |
| -0.064299612 | 0.042980833  | 0.044065245  | -0.036837947 | 0.445605611 | 0.126579973 | 0.32168338  | 0.265179269  | 0.191203007 | 0.246516442 | 0.331644002 | Error |
| -0.080355896 | -0.241390581 | -0.030585879 | 0.241800191  | 0.976951564 | 0.946283479 | 0.479744227 | 0.668557727  | 0.658859529 | 0.659227125 | 0.658898359 | Error |
| -0.065240509 | -0.245111856 | -0.022276447 | 0.262322614  | 0.977092591 | 0.948534693 | 0.259634845 | 0.26604842   | 0.265706032 | 0.265545176 | 0.267924545 | Error |
| -0.054974507 | -0.254527035 | -0.015781711 | 0.26507894   | 0.977224985 | 0.948618896 | 0.109364365 | 0.11642524   | 0.1050506   | 0.094006706 | 0.09590026  | Error |
| 0.353307184  | -0.41651375  | 0.058280315  | 0.393071172  | 0.975400049 | 0.959365029 | 4.68081E-05 | 4.66262E-05  | 4.61859E-05 | 0.096720255 | 0.154295155 | Error |
| -0.153262366 | -0.589005289 | -0.45657452  | -0.224101458 | 0.992966793 | 0.989423178 | 0.094248736 | 0.156817509  | 0.249176411 | 0.327790972 | 0.367196787 | Error |
| -0.156883462 | -0.611018564 | -0.398019051 | -0.084060186 | 0.993059984 | 0.989754962 | 0.68907563  | 0.891937131  | 0.482758357 | 0.721322685 | 0.395699905 | Error |
| -0.174422704 | -0.596441789 | -0.468857433 | -0.250926875 | 0.992786818 | 0.989059805 | 0.787878788 | 0.895013618  | 0.585190195 | 0.971603515 | 0.870915872 | Error |

Εικόνα 5.12: Στιγμιότυπο αρχείου xls όπου αποτελεί την βάση αληθείας

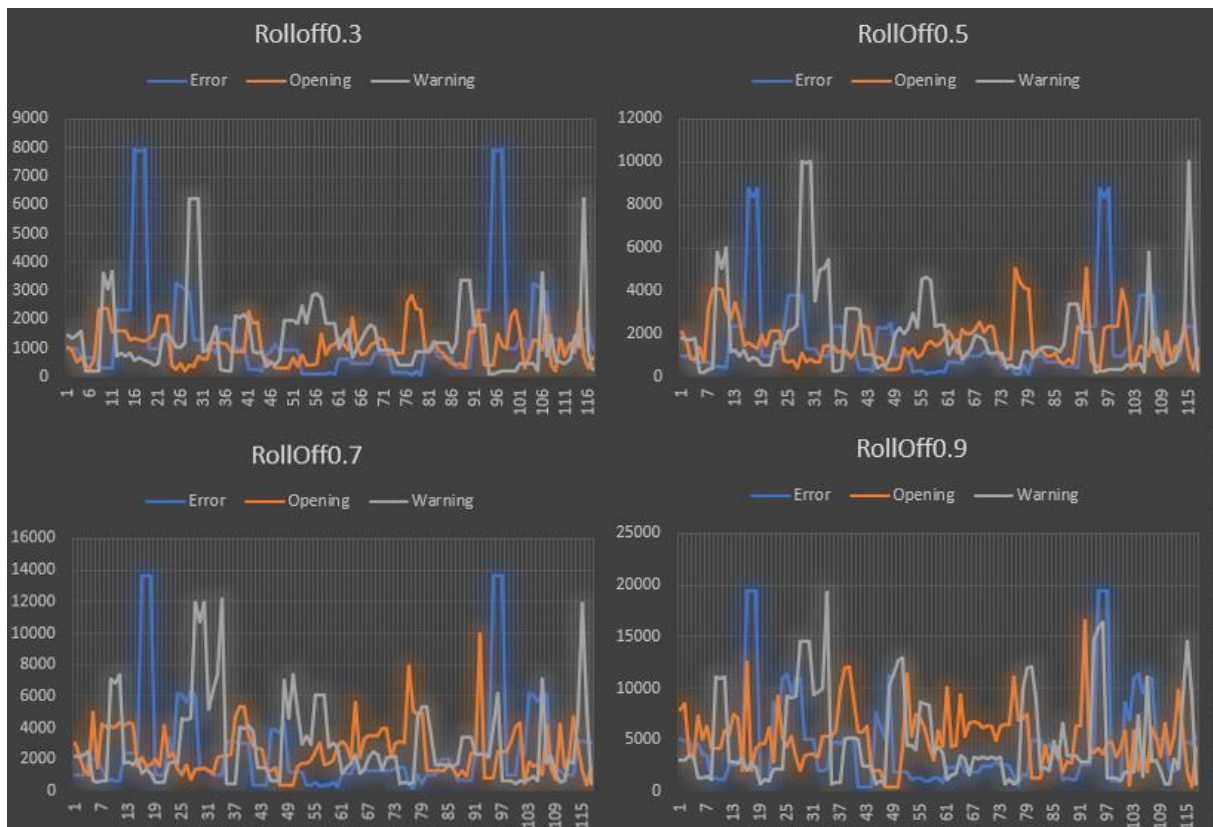
## 5.5 Αξιολόγηση τιμών

Σε αυτό το σημείο θεωρήσαμε σκόπιμο να δούμε αν οι τιμές των χαρακτηριστικών που επιλέξαμε είναι κατάλληλες προκειμένου να εξαχθούν συμπεράσματα μέσα από αυτές. Όταν οι τιμές ενός χαρακτηριστικού προκαλούν πολλές επικαλύψεις μεταξύ των κλάσεων τότε αυτό το χαρακτηριστικό συνεισφέρει λιγότερο στην κατηγοριοποίηση. Αντίστοιχα, αν οι τιμές ενός χαρακτηριστικού δημιουργούν μία πιο ξεκάθαρη εικόνα διαγράμματος σημαίνει πως το συγκεκριμένο χαρακτηριστικό είναι ικανό να βοηθήσει σε μεγάλο βαθμό στην κατηγοριοποίηση.

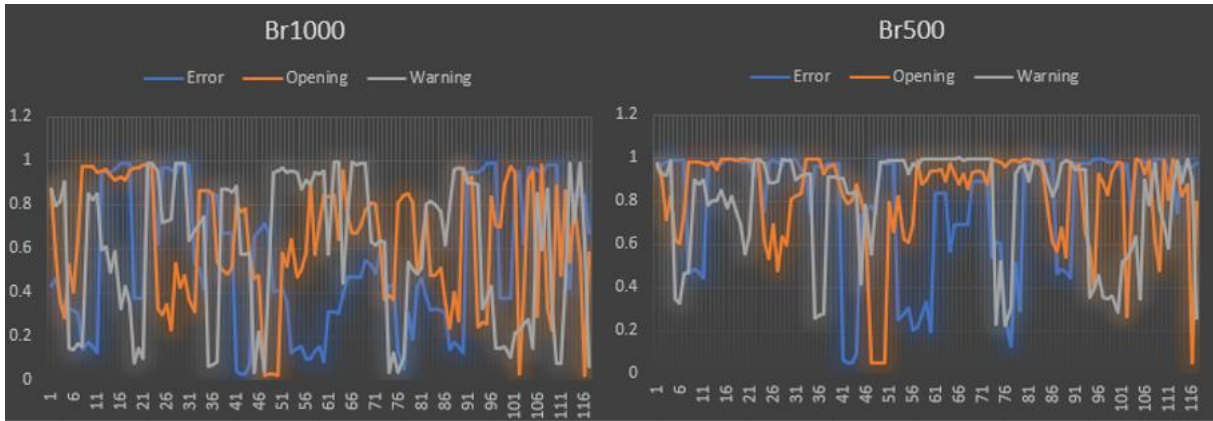
Αυτό που παρατηρήσαμε είναι πως και στις δύο περιπτώσεις (παράθυρο 3 δευτερολέπτων και παράθυρο 1 δευτερολέπτου) δεν παρουσιάζονται πολλές διαφοροποιήσεις. Πιο συγκεκριμένα, το Mfcc για αρκετούς συντελεστές αλλά και η ενεργειακή συχνότητα αποκοπής (Spectral roll-off) είναι αρκετά αξιόπιστα σαν χαρακτηριστικά καθώς δεν δημιουργούν πολλές επικαλύψεις. Αντιθέτως τα χαρακτηριστικά Brightness αλλά και τα χαρακτηριστικά για το συναίσθημα είναι λιγότερο αξιόπιστα λόγω των τιμών τους.



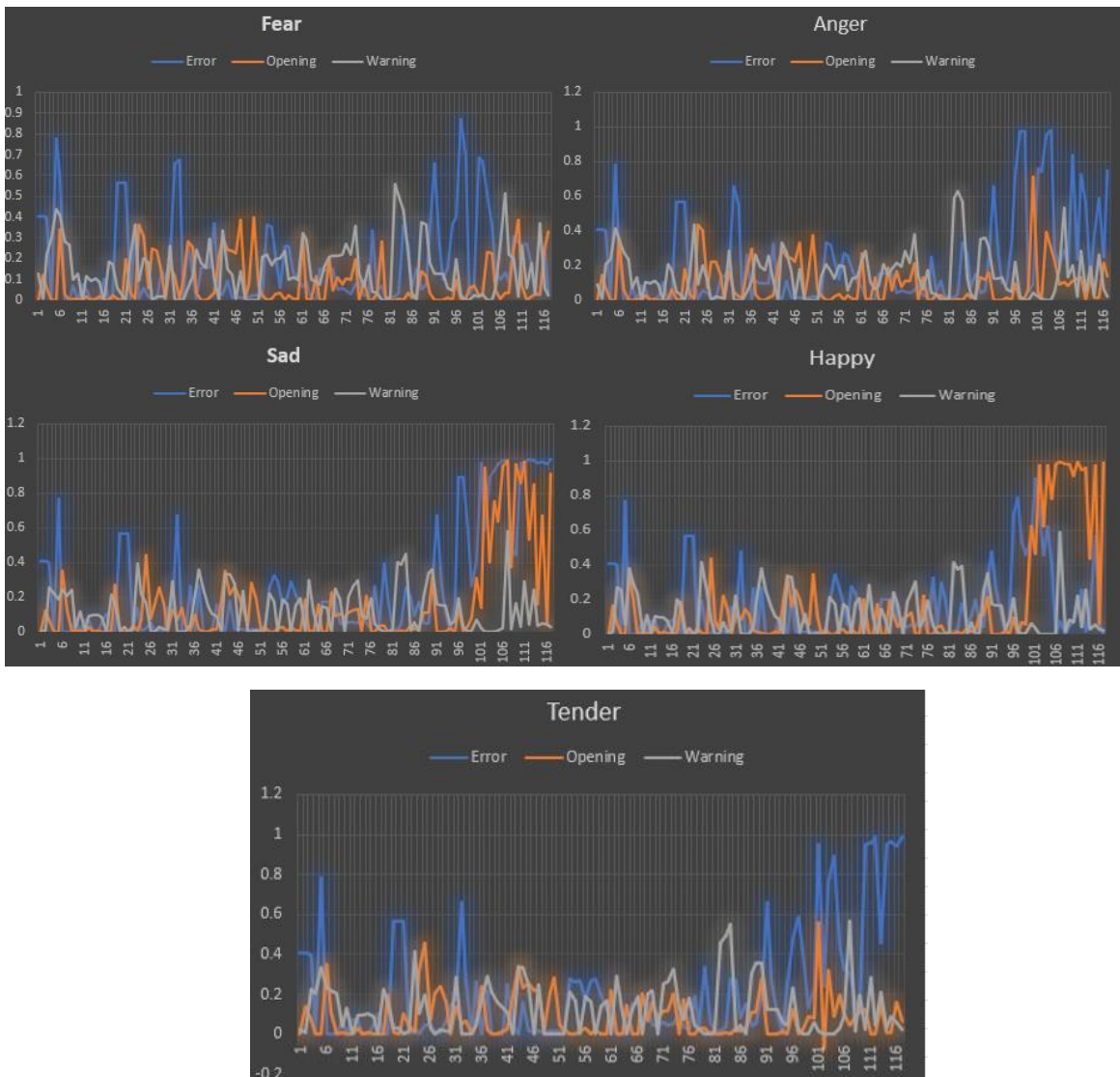
Εικόνα 5.13: Διαγράμματα χαρακτηριστικού MFCC για διαφορετικούς συντελεστές



Εικόνα 5.14: Διαγράμματα χαρακτηριστικού Rolloff



Εικόνα 5.15: Διαγράμματα χαρακτηριστικού Brightness



Εικόνα 5.16: Διαγράμματα χαρακτηριστικών συναισθήματος

## 5.6 Επίλογος

Στο παρόν κεφάλαιο έγινε περιγραφή του αρχικού σταδίου του πειράματος που αφορούσε την επιλογή των ειδών των ήχων, την συλλογή τους και την επιλογή κριτηρίων που χρησιμοποιήθηκαν για την ταξινόμηση. Ακολούθως, παρουσιάστηκαν κάποια συγκεκριμένα χαρακτηριστικά για τα οποία δίνεται μία εικόνα προκειμένου να γίνει σαφέστερη η λειτουργία τους. Εν συνεχεία έγινε η κατάτμηση σε παράθυρα 3 και 1 δευτερολέπτου. Επιπλέον δίνεται ο κώδικας που εισήχθη στο Matlab και το αρχείο Excel όπου εξήχθησαν τα χαρακτηριστικά καθώς και η αξιολόγηση τους.

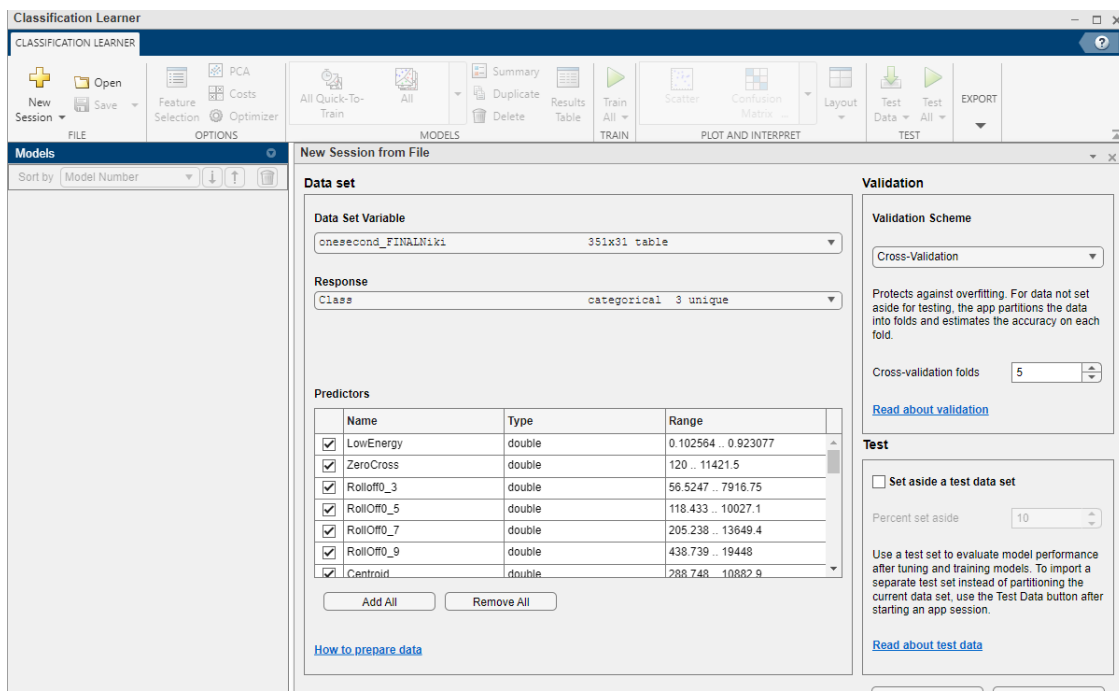
## Κεφάλαιο 6<sup>ο</sup> : Πειράματα μηχανικής μάθησης που εκπονήθηκαν

### 6.1 Εισαγωγή

Μετά την ολοκλήρωση της διαδικασίας εξαγωγής των χαρακτηριστικών έγινε η συγχώνευση τους σε ένα αρχείο xls προκειμένου να γίνει classification στο Matlab, με την χρήση του εργαλείου Classification Learner. Το Classification Learner από μόνο του σαν εργαλείο περιέχει αρκετούς αλγορίθμους ταξινόμησης. Ωστόσο, επιλέξαμε προκειμένου να έχουμε μία πιο εμπειριστατωμένη άποψη να χρησιμοποιήσουμε και το λογισμικό WEKA τόσο για την ταξινόμηση όσο και για την αξιολόγηση των χαρακτηριστικών. Στο παρακάτω κεφάλαιο ακολουθεί μία συνοπτική περιγραφή των εργαλείων και αλγορίθμων που χρησιμοποιήθηκαν καθώς και οι έννοιες των μετρικών αξιολόγησης προκειμένου να γίνει πιο σωστή κατανόηση των αποτελεσμάτων.

### 6.2 Matlab

Το Matlab είναι ένα διαδραστικό πρόγραμμα το οποίο χρησιμοποιείται για αριθμητικούς υπολογισμούς καθώς και για οπτικοποίηση δεδομένων (data visualization)<sup>13</sup> το οποίο αναπτύχθηκε από την MathWorks. Το Matlab δίνει την δυνατότητα στον χρήστη να εκπαιδεύσει μοντέλα προκειμένου να ταξινομήσει δεδομένα χρησιμοποιώντας την εποπτευόμενη μηχανική μάθηση μέσω της εφαρμογής Classification Learner. Επιπλέον ο χρήστης έχει την δυνατότητα να επιλέξει τα χαρακτηριστικά που επιθυμεί για την ταξινόμηση, να καθορίσει σχήματα επικύρωσης όπως το Cross-Validation και το Holdout Validation αλλά και να αξιολογήσει αποτελέσματα<sup>14</sup>.



Εικόνα 6.1: Classification Learner

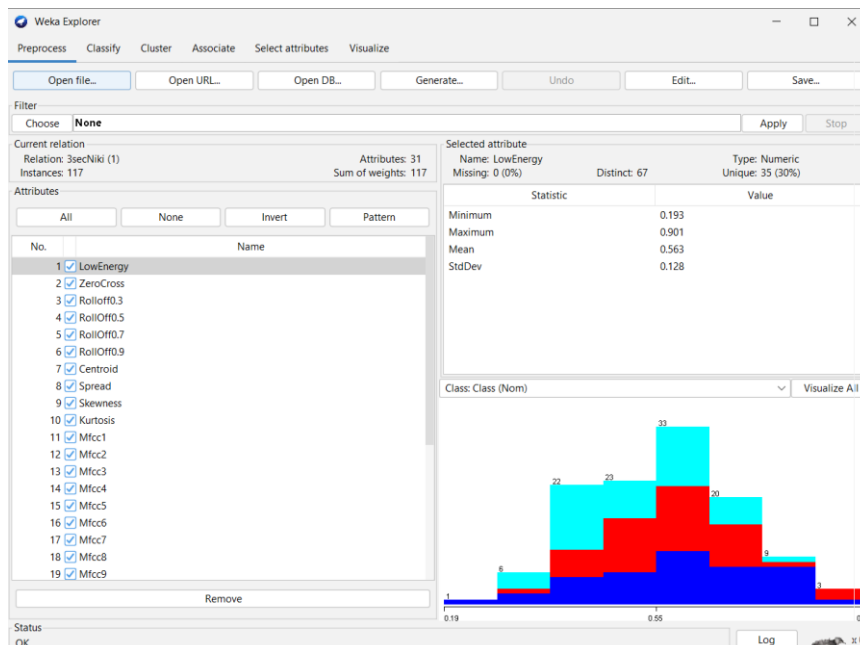
<sup>13</sup> [http://www.mas.ucy.ac.cy/~xenophon/pubs/matlab\\_intro.pdf](http://www.mas.ucy.ac.cy/~xenophon/pubs/matlab_intro.pdf)

<sup>14</sup> <https://www.mathworks.com/help/stats/classificationlearner-app.html>

## 6.3 WEKA

Παράλληλα με την χρήση του Matlab χρησιμοποιήθηκε και το λογισμικό WEKA για την ταξινόμηση αλλά και για την αξιολόγηση των χαρακτηριστικών. Το WEKA είναι ένα ελεύθερο λογισμικό για μηχανική μάθηση και εξόρυξη δεδομένων. Χάρη στο γεγονός ότι έχει μεγάλη γκάμα μεθόδων για εξόρυξη δεδομένων, συνεχή εξέλιξη και υποστήριξη, ελεύθερη διανομή του πηγαίου κώδικα καθώς και δυνατότητα εγκατάστασης σε διαφορετικές πλατφόρμες, το WEKA αποτελεί μέχρι και σήμερα ευρεία αποδοχή τόσο στους ακαδημαϊκούς όσο και στους επιχειρηματικούς κλάδους[48]. Έχει πολλές γραφικές διεπαφές οι οποίες επιτρέπουν την εύκολη πρόσβαση στην υποκείμενη λειτουργικότητα. Το κύριο γραφικό περιβάλλον είναι ο “Explorer”, ο οποίος περιέχει το πάνελ με τους διαφορετικούς πίνακες.

Στον πρώτο πίνακα που ονομάζεται “Προεπεξεργασία” τα δεδομένα μπορούν να φορτωθούν και να μεταφερθούν χρησιμοποιώντας τα εργαλεία προεπεξεργασίας της WEKA, τα λεγόμενα φίλτρα. Ωστόσο, προκειμένου να εισαχθούν τα δεδομένα θα πρέπει το αρχείο να είναι της μορφής .arff. Ο δεύτερος πίνακας δίνει πρόσβαση στους αλγόριθμους ταξινόμησης και παλινδρόμησης του WEKA και ονομάζεται “Ταξινόμηση”. Με τον πίνακα "Cluster" ο χρήστης μπορεί να εκτελέσει έναν αλγόριθμο ομαδοποίησης για τα δεδομένα που έχουν φορτωθεί στον πίνακα της Προεπεξεργασίας. Ο πίνακας "Cluster" παρέχει απλά στατιστικά στοιχεία προκειμένου να αξιολογηθεί η απόδοση ομαδοποίησης. Η αναγνώριση των πιο προβλεπόμενων δεδομένων είναι ίσως από τις πιο σημαντικές διαδικασίες στην εξόρυξη δεδομένων. Για αυτόν τον λόγο ο Explorer διαθέτει έναν ειδικό πίνακα για την επιλογή χαρακτηριστικών που ονομάζεται “Select attributes” το οποίο δίνει πρόσβαση σε πολλούς αλγορίθμους και κριτήρια αξιολόγησης προκειμένου να εντοπιστούν τα πιο σημαντικά χαρακτηριστικά ενός συνόλου δεδομένων. Τέλος, ο πίνακας “Visualize” παρέχει ένα χρωματιστό κωδικοποιημένο διάγραμμα διασποράς matrix μαζί με την επιλογή πρόσβασης σε διαδοχικά βαθύτερα επίπεδα μιας ιεραρχικά οργανωμένης βάσης δεδομένων καθώς και την επιλογή τμημάτων των δεδομένων για οπτικοποίηση.



Εικόνα 6.2: Weka Explorer[49]

## 6.4 Τρόποι επικύρωσης (validation)

Κατά την εκπαίδευση των μοντέλων με την χρήση της εποπτευόμενης Μηχανικής Μάθησης προκειμένου να δημιουργηθούν τα μοντέλα ταξινόμησης πρέπει να γίνει ο διαχωρισμός των δεδομένων εισόδου σε δεδομένα εκπαίδευσης (training set) για να εκπαιδευτεί ο αλγόριθμος κατηγοριοποίησης και δεδομένα ελέγχου (testing set) έτσι ώστε να γίνει ο προσδιορισμός του ποσοστού επιτυχίας του αλγορίθμου [50].

Στις περισσότερες περιπτώσεις στη διαδικασία εποπτευόμενης μηχανικής μάθησης χρησιμοποιούνται δύο τρόποι επικύρωσης, ο k-fold και ο split ή αλλιώς Holdout validation. Ο k-fold, διαχωρίζει τα δεδομένα σε k υποσύνολα. Από αυτά, τα k-1 χρησιμοποιούνται για την εκπαίδευση ενώ το ένα χρησιμοποιείται για έλεγχο. Η συνολική διαδικασία πραγματοποιείται k φορές. Στα συγκεκριμένα πειράματα χρησιμοποιήθηκε k=5 και k=10. Το k=10 διότι αποτελεί μία συνήθη τακτική ενώ το k=5 για ενδεχόμενο γενίκευσης της εκπαίδευσης. Όσο περισσότερο εκπαιδύεται ένα μοντέλο τόσο πιο πολύ ελλοχεύει ο κίνδυνος της υπερ-εκπαίδευσης. Αντίστοιχα, όσο περισσότερο διαιρούνται τα δεδομένα σε folds το μοντέλο “παραμαθαίνει” τα συγκεκριμένα δεδομένα και όταν θα έρθει ένα άγνωστο δεδομένο δεν θα καταφέρει να το κατηγοριοποιήσει.

Ο άλλος τρόπος επικύρωσης είναι το split ή αλλιώς Holdout Validation: ο διαχωρισμός σε δεδομένα ελέγχου και δεδομένα εκπαίδευσης, όπου η εκπαίδευση γίνεται μία μόνο φορά. Ο συγκεκριμένος τρόπος επικύρωσης είναι χρήσιμη στην περίπτωση που έχουμε μεγάλο όγκο βάσης αληθείας (data-set) ή στην περίπτωση στην οποία θέλουμε να χτίσουμε ένα μοντέλο από την αρχή[51]. Στο δικό μας πείραμα οι μετρήσεις πραγματοποιήθηκαν μετρήσεις με ποσοστό δεδομένων (training set) 30% και 40%.

Ανακεφαλαιώνοντας όλα τα παραπάνω για το πείραμα μας χρησιμοποιήσαμε συνολικά 4 μεθόδους επικύρωσης: 5-fold validation, 10-fold validation, 30% Holdout και 40% Holdout.

## 6.5 Αλγόριθμοι εκμάθησης μοντέλου

Στα περισσότερα προβλήματα ταξινόμησης κατά κύριο λόγο χρησιμοποιούνται ο k-Nearest Neighbor(Knn), Support Vector Machines(SVM) ,τα δέντρα αποφάσεων (Decision Trees) και τα Νευρωνικά Δίκτυα (Neural Networks). Επιπλέον, χρησιμοποιήθηκαν ποικίλοι αλγόριθμοι οι οποίοι είναι οι βασικοί αλγόριθμοι μηχανικής μάθησης που προαναφέρθηκαν με κάποιες παραλλαγές τους οι οποίοι είναι διαθέσιμοι από την εφαρμογή του Matlab, το Classification Learner<sup>15</sup> αλλά και κάποιο αλγόριθμοι οι οποίοι είναι διαθέσιμοι από το λογισμικό WEKA.

### Μηχανές Διανυσμάτων Υποστήριξης(Support Vector Machines-SVM)

**Cubic SVM:** Ένα SVM που χρησιμοποιεί τετράγωνο πυρήνα.

**Fine Gaussian SVM:** Κάνει λεπτομερείς διακρίσεις μεταξύ των κλάσεων χρησιμοποιώντας Γκαουσιανό πυρήνα με την κλίμακα πυρήνα ρυθμιζόμενη στην τετραγωνική ρίζα των μεταβλητών εισόδου προς 4.

**Medium Gaussian SVM:** Μεσαίες διακρίσεις, με κλίμακα πυρήνα ρυθμισμένη στην τετραγωνική ρίζα των μεταβλητών εισόδου.

**Quadratic SVM:** Ένας γρήγορος και εύκολος στην ερμηνεία διακριτός ταξινομητής που τοποθετεί ελλειπτικά παραβολικά ή υπερβολικά όρια μεταξύ των κλάσεων.

### Αλγόριθμοι κατηγοριοποίησης πλησιέστερου γείτονα (k-NN)

**Fine KNN:** Ταξινομητής πλησιέστερου γείτονα. Πραγματοποιεί λεπτομερείς διακρίσεις μεταξύ των κλάσεων. Ο αριθμός των γειτόνων έχει οριστεί σε 1.

**Weighted KNN:** Ταξινομητής πλησιέστερου γείτονα μέτριας λεπτομέρειας διακρίσεως μεταξύ κλάσεων χρησιμοποιώντας ένα βάρος απόστασης με αριθμό γειτόνων με ορισμένη τιμή το 10.

**Cosine KNN:** Χρησιμοποιούν μετρική απόστασης συνημιτόνου για τις μέσες διακρίσεις μεταξύ κλάσεων. Ο αριθμός των γειτόνων έχει οριστεί σε 10.

**Subspace KNN:** Κατάλληλο για πολλές προβλέψεις. Αυξάνεται με βάση των αριθμών των εκπαιδευόμενων.

**IBk:** Έχει την δυνατότητα να επιλέξει την κατάλληλη τιμή του k με βάση τη διασταυρούμενη επικύρωση.

**KStar:** Είναι ένας ταξινομητής που βασίζεται σε στιγμιότυπα. Δηλαδή, η κλάση ενός στιγμιότυπου δοκιμής βασίζεται στην κατηγορία αυτών των παρόμοιων στιγμιότυπων εκπαίδευσης όπως αυτό καθορίζεται από κάποια συνάρτηση ομοιότητας.

---

<sup>15</sup> [https://www.mathworks.com/help/stats/choose-a-classifier.html#mw\\_cdbdc267-a5ed-419b-a46f-7217cc1132ef](https://www.mathworks.com/help/stats/choose-a-classifier.html#mw_cdbdc267-a5ed-419b-a46f-7217cc1132ef)

## Δέντρα Αποφάσεων (Decision Trees)

**Bagged Trees:** Τα «Συσκευασμένα Δέντρα» είναι Δέντρα Απόφασης τα οποία χρησιμοποιούν τον αλγόριθμο «τυχαίο δάσος» του Breiman.

**Boosted Tree:** Σε αρκετές περιπτώσεις αποδίδουν καλύτερα σε σχέση με τα Bagged Trees. Αυξάνεται ανάλογα με τον αριθμό των εκπαιδευόμενων ή το μέγιστο αριθμό των split.

**Coarse Tree:** Χρησιμοποιούν μερικά φύλλα προκειμένου να κάνουν χονδροειδείς διακρίσεις μεταξύ των κλάσεων. Ο μέγιστος αριθμός διαχωρισμού είναι 4.

**Fine Tree:** Πολλά φύλλα για να κάνουν λεπτές διακρίσεις μεταξύ των κλάσεων (μέγιστος αριθμός των split είναι 100).

**Medium Tree:** Μεσαίος αριθμός φύλλων που χρησιμοποιείται για λεπτότερες διακρίσεις μεταξύ των κλάσεων (μέγιστος αριθμός των split είναι 20).

**RusBoosted Trees:** Εκτελούνται κατανεμημένα δεδομένα εκπαίδευσης επανειλημμένα σε αδύναμους εκπαιδευόμενους και συνδυάζονται σε έναν ισχυρό ταξινομητή υψηλής ακρίβειας.

## Λογιστική Παλινδρόμηση (Logistic Regression)

**Logistic Regression:** Στην περίπτωση που υπάρχουν δύο κλάσεις, η λογιστική παλινδρόμηση είναι ένας αρκετά δημοφιλής και απλός αλγόριθμος ταξινόμησης ο οποίος είναι εύκολο να ερμηνευτεί. Ο ταξινομητής μοντελοποιεί τις πιθανότητες κλάσης ως συνάρτηση του γραμμικού συνδυασμού των προβλέψεων. Ωστόσο, το συγκεκριμένο μοντέλο δεν έχει καθόλου ευελιξία καθώς δεν μπορεί να αλλαχθεί καμία παράμετρος για τον έλεγχο της προσαρμοστικότητας του μοντέλου.

## Kernel Approximation Classifiers

Χρησιμοποιούνται προκειμένου να εκτελεστούν μη γραμμικής ταξινόμησης δεδομένα με πολλές παρατηρήσεις. Για μεγάλο όγκο δεδομένων στην μνήμη τείνουν να εκπαιδεύονται και να προβλέπουν πιο γρήγορα από τους ταξινομητές SVM με πυρήνες Gaussian. Τα μοντέλα ταξινόμησης πυρήνα Gauss αντιστοιχίζουν τους προγνωστικούς δείκτες σε ένα χώρο υψηλής διάστασης και στη συνέχεια προσαρμόζουν ένα γραμμικό μοντέλο στους μετασχηματισμένους προγνωστικούς παράγοντες στον χώρο υψηλής διάστασης.

**SVM Kernel & Logistic Regression Kernel:** Και οι δύο αλγόριθμοι παρουσιάζουν μέτρια ευελιξία μοντέλου η οποία αυξάνεται όσο μειώνεται η ρύθμιση κλίμακας πυρήνα.

## Neural Network Classifiers

**Narrow Neural Network:** Ένα νευρωνικό δίκτυο με μία πλήρως συνδεδεμένη στρώση μεγέθους 10.

**Medium Neural Network:** Ένα νευρωνικό δίκτυο με μία πλήρως συνδεδεμένη στρώση μεγέθους 25.

**Wide Neural Network:** Ένα νευρωνικό δίκτυο με μία πλήρως συνδεδεμένη στρώση μεγέθους 100.

**Multilayer Perceptron:** Χρησιμοποιεί την αντίστροφη διάδοση για να μάθει ένα πολυστρωματικό perceptron προκειμένου να ταξινομήσει στιγμιότυπα.

## Naive Bayes Classifiers

Οι ταξινομητές Naive Bayes είναι εύκολοι στην ερμηνεία και χρησιμοποιούνται για την ταξινόμηση πολλαπλών κλάσεων.

**Gaussian Naive Bayes:** Σαν μοντέλο έχει χαμηλή ευελιξία καθώς δεν μπορεί να γίνει καμία αλλαγή στις παραμέτρους.

**Kernel Naive Bayes:** Σε αντίθεση με τον Gaussian αν αλλαχθούν οι ρυθμίσεις για τον τύπο του πυρήνα και της υποστήριξης μπορεί να γίνει έλεγχος του τρόπου κατά τον οποίο ο ταξινομητής μοντελοποιεί τις διανομές πρόβλεψης.

## 6.6 Επίλογος

Στο έκτο κεφάλαιο παρουσιάζεται η διαδικασία που ακολουθεί μετά την εξαγωγή των ηχητικών παραμέτρων, δηλαδή τα πειράματα μηχανικής μάθησης. Για τα πειράματα αυτά χρησιμοποιήθηκε το εργαλείο Classification Learner του Matlab καθώς και το WEKA. Τα δεδομένα εισόδου έπρεπε να χωριστούν σε δεδομένα εκπαίδευσης και σε δεδομένα ελέγχου. Κατά την διάρκεια της εκπαίδευσης παρατηρήσαμε ότι η εκπαίδευση δεν πρέπει να ξεπερνάει ένα συγκεκριμένο όριο. Τέλος, απαριθμούνται όλοι οι αλγόριθμοι μηχανικής μάθησης που χρησιμοποιήθηκαν όπως για παράδειγμα οι μηχανές διανυσμάτων υποστήριξης, οι αλγόριθμοι κατηγοριοποίησης πλησιέστερου γείτονα, τα νευρωνικά δίκτυα κ.α.

## Κεφάλαιο 7<sup>ο</sup>: Αποτελέσματα πειραμάτων

### 7.1 Εισαγωγή

Σε αυτό το κεφάλαιο αναλύονται οι μετρικές αξιολόγησης που χρησιμοποιούνται προκειμένου να γίνει η αξιολόγηση των αλγορίθμων μηχανικής μάθησης. Εν συνεχεία, βλέπουμε τα αποτελέσματα των αλγορίθμων που εκτελέστηκαν και εκτελούνται παρατηρήσεις από τα δεδομένα που προκύπτουν.

### 7.2 Μετρικές Αξιολόγησης

Υπάρχουν διάφοροι τρόποι για την λήψη απόδοσης αλλά και την κατανόηση της ακρίβειας ενός μοντέλου. Το cross-validation θεωρείται ένας από τους πιο δημοφιλείς. Κατά την διάρκεια του cross-validation τα δεδομένα χωρίζονται σε δύο κατηγορίες. Την κατηγορία train και την κατηγορία test. Στην κατηγορία test συγκαταλέγεται το 75-80% των δεδομένων το οποίο εν συνεχεία χρησιμοποιείται για την εξάσκηση του classifier. Αντίστοιχα, στην κατηγορία test περιλαμβάνεται το 20-25% των ίδιων δεδομένων και μέσω αυτού λαμβάνονται οι μετρικές απόδοσης. Η συγκεκριμένη διαδικασία επαναλαμβάνεται αρκετές φορές και υπολογίζεται ο μέσος όρος κάθε μετρικής. Επιπλέον το cross-validation χρησιμεύει και στην περίπτωση που το testing set είναι συνέχεια το ίδιο και η ανάλυση έχει εκπαιδευτεί τόσο πολύ στο συγκεκριμένο data set με αποτέλεσμα να μην μπορεί να αναλυθεί κάποιο διαφορετικό data set. Αυτή η περίπτωση είναι γνωστή και ως overtraining.

Προκειμένου να είμαστε σε θέση να υπολογίσουμε την απόδοση ενός classifier χρησιμοποιούνται τρεις μετρικές [52]:

**Precision:** Μετράει πόσα αντικείμενα προβλέφθηκαν σωστά ότι ανήκουν σε μία δεδομένη κατηγορία από όλα τα αντικείμενα που είχαν προβλεφθεί (είτε σωστά είτε εσφαλμένα) ως μέλη της κατηγορίας.

$$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)}$$

**Recall:** Μετράει πόσα αντικείμενα είχαν προβλεφθεί σωστά ως ανήκοντα σε μία δεδομένη κατηγορία από όλα τα αντικείμενα που θα έπρεπε να είχαν προβλεφθεί ότι ανήκουν στην κατηγορία.

$$Recall = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)}$$

**Accuracy:** Μετράει πόσα αντικείμενα έχουν προβλεφθεί σωστά, είτε ανήκουν είτε όχι σε μία δεδομένη κατηγορία από όλα τα αντικείμενα.

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

**F1-score:** Το F1 score είναι μία συνάρτηση της Ακρίβειας (Precision) και της Ανάκλησης (Recall). Το F1 score χρειάζεται στην περίπτωση που θέλουμε να αναζητήσουμε μία ισορροπία μεταξύ της Ακρίβειας και της Ανάκλησης και υπάρχει μία ανομοιόμορφη κατανομή κλάσεων (Μεγάλος αριθμός Πραγματικών Αρνητικών).

$$F1 = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

**Error Rate:** Αποτελεί ένα κριτήριο εξίσου σημαντικό με τα υπόλοιπα και εκφράζει το ποσοστό του αριθμού των δειγμάτων που κατηγοριοποιήθηκαν εσφαλμένα ως προς το συνολικό αριθμό εισόδων στο μοντέλο. Πιο συγκεκριμένα, μας ενδιαφέρει το μερικό ποσοστό σφάλματος το οποίο ορίζεται με βάση την παρακάτω εικόνα.

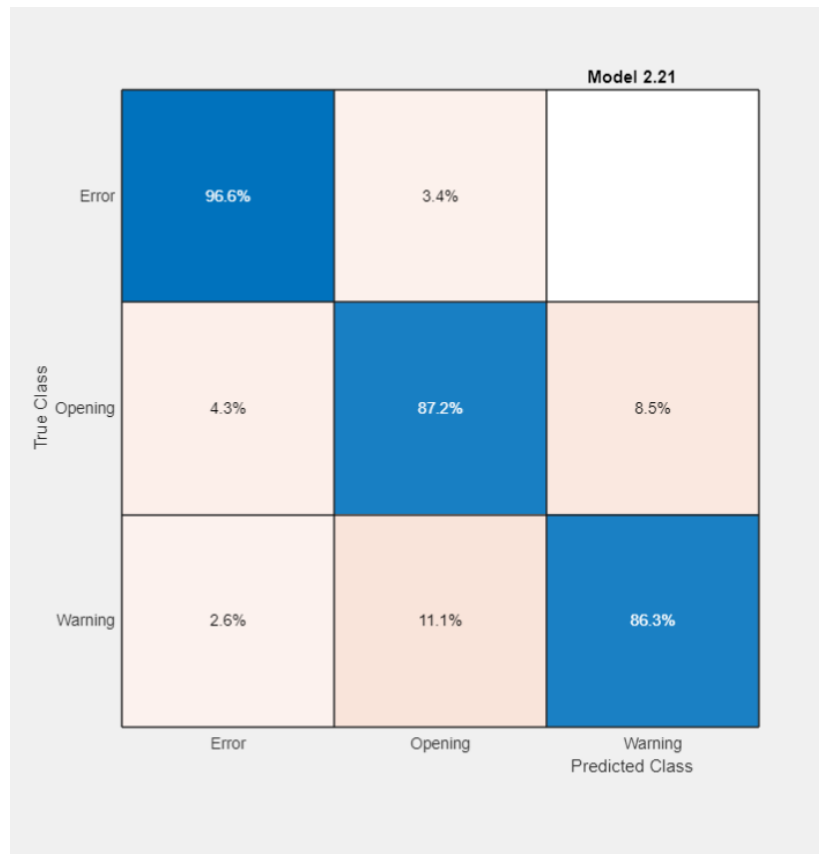
$$FR_{\frac{C_i}{C_j}}(\%) = FR\left(\frac{C_i}{C_j}\right) = \frac{N_{FN}\left(\frac{C_i}{C_j}\right)}{N_{C_i}} * 100$$

Το μερικό ποσοστό σφάλματος FR για την κλάση  $C_i$  ως προς την κλάση  $C_j$  ισούται με τον λόγο του αριθμού των δειγμάτων  $N_{fn}(C_i/C_j)$  της κλάσης  $C_i$  που κατηγοριοποιήθηκαν λανθασμένα στην κλάση  $C_j$  προς τον συνολικό αριθμό δειγμάτων  $N_{C_i}$  όπου περιλαμβάνονται στην κλάση  $C_i$ . Η παραπάνω συνάρτηση αποτελεί ένα πολύ σημαντικό στοιχείο για την ανίχνευση των λανθασμένων κατηγοριοποιήσεων μεταξύ δύο κλάσεων.

Για την δική μας αξιολόγηση χρησιμοποιήθηκαν οι μετρικές αξιολόγησης Accuracy , Precision και το Error Rate.

### 7.3 Μήτρα σύγχυσης (confusion matrix)

Είναι μία σύνοψη των αποτελεσμάτων πρόβλεψης ενός προβλήματος ταξινόμησης. Ο αριθμός σωστών και εσφαλμένων προβλέψεων συνοψίζεται με τις τιμές μετρήσεων και αναλύεται ανά κατηγορία. Παρουσιάζει τους τρόπους που μπερδεύουν το μοντέλο ταξινόμησης κατά την διάρκεια προβλέψεων. Επιπλέον, μας δίνει την δυνατότητα να δούμε τόσο τα σφάλματα που δημιουργούνται από έναν ταξινομητή όσο και τα είδη τους γεγονός που το καθιστά πολύ σημαντικό για την εκπαίδευση του μοντέλου.



Εικόνα 7.1: Παράδειγμα μήτρας σύγχυσης

Στην εικόνα 22 παρουσιάζεται η μήτρα σύγχυσης όπου παράχθηκε από εκπαίδευση μοντέλου με χρήση του αλγορίθμου Bagged Trees σε παράθυρο ανάλυσης 1 δευτερολέπτου με μέθοδο επικύρωσης 10 Cross Validation. Στον κάθετο άξονα παρουσιάζονται οι πραγματικές κλάσεις ενώ στον οριζόντιο άξονα οι κλάσεις οι οποίες προβλέφθηκαν. Στην μπλε διαγώνιο παρουσιάζεται το ποσοστό των δειγμάτων που ταξινομήθηκαν ορθώς. Στα υπόλοιπα πεδία υπάρχουν τα μερικά ποσοστά σφάλματος. Για παράδειγμα, το ποσοστό 2,6% εκφράζει ότι το 2,6% των δειγμάτων ταξινομήθηκαν εσφαλμένα ως Error ενώ η πραγματική τους κλάση ήταν Warning.

## 7.4 Αποτελέσματα Αλγορίθμων

Ακολουθούν όλα τα αποτελέσματα αλγορίθμων σε πίνακες, οι αποδόσεις καθώς και τα σφάλματα για όλους τους αλγορίθμους σε όλα τα παράθυρα ανάλυσης καθώς και με όλους τους τρόπους επικύρωσης.

| Αλγόριθμος                 | Accuracy | P(error) | P(opening) | P(warning) | E(op>err) | E(wa>err) | E(err>op) | E(wa>op) | E(err>wa) | E(op>wa) |
|----------------------------|----------|----------|------------|------------|-----------|-----------|-----------|----------|-----------|----------|
| Bagged Trees               | 70.10%   | 69.20%   | 66.70%     | 74.40%     | 12.80%    | 10.30%    | 25.60%    | 15.40%   | 5.10%     | 20.50%   |
| Fine KNN                   | 66.70%   | 66.70%   | 76.90%     | 56.40%     | 7.70%     | 20.50%    | 20.50%    | 23.10%   | 12.80%    | 15.40%   |
| Quadratic SVM              | 65.80%   | 69.20%   | 74.40%     | 53.80%     | 15.40%    | 20.50%    | 20.50%    | 25.60%   | 10.30%    | 10.30%   |
| Fine Tree                  | 65.00%   | 69.20%   | 61.50%     | 64.10%     | 17.90%    | 17.90%    | 17.90%    | 17.90%   | 12.80%    | 20.50%   |
| Medium Tree                | 65.00%   | 69.20%   | 61.50%     | 64.10%     | 17.90%    | 17.90%    | 17.90%    | 17.90%   | 12.80%    | 20.50%   |
| Coarse Tree                | 64.10%   | 71.80%   | 53.80%     | 66.70%     | 33.30%    | 20.50%    | 17.90%    | 12.80%   | 10.30%    | 12.80%   |
| Cubic SVM                  | 64.10%   | 69.20%   | 74.40%     | 48.70%     | 15.40%    | 25.60%    | 20.50%    | 25.60%   | 10.30%    | 10.30%   |
| RUSBoosted Trees           | 63.20%   | 59.00%   | 66.70%     | 64.10%     | 15.40%    | 20.50%    | 23.10%    | 15.40%   | 17.90%    | 17.90%   |
| Fine Gaussian SVM          | 63.20%   | 46.20%   | 53.80%     | 89.70%     | 5.10%     | 2.60%     | 7.70%     | 7.70%    | 46.20%    | 41.00%   |
| Narrow Neural Network      | 63.20%   | 69.20%   | 69.20%     | 51.30%     | 12.80%    | 25.60%    | 17.90%    | 23.10%   | 12.80%    | 17.90%   |
| Kstar                      | 61.53%   | 62.90%   | 55.10%     | 69.70%     | 17.94%    | 15.38%    | 30.76%    | 25.64%   | 12.82%    | 12.82%   |
| IBK                        | 60.68%   | 59.50%   | 57.80%     | 65.70%     | 15.38%    | 23.07%    | 30.76%    | 17.94%   | 12.82%    | 17.94%   |
| Boosted Trees              | 59.80%   | 74.40%   | 46.20%     | 59.00%     | 38.50%    | 25.60%    | 20.50%    | 15.40%   | 5.10%     | 15.40%   |
| Medium Gaussian SVM        | 59.80%   | 59.00%   | 66.70%     | 53.80%     | 28.20%    | 20.50%    | 17.90%    | 25.60%   | 23.10%    | 5.10%    |
| Wide Neural Network        | 56.40%   | 61.50%   | 59.00%     | 48.70%     | 25.60%    | 28.20%    | 15.40%    | 23.10%   | 23.10%    | 15.40%   |
| Medium Neural Network      | 55.60%   | 56.40%   | 56.40%     | 53.80%     | 25.60%    | 20.50%    | 30.80%    | 25.60%   | 12.80%    | 17.90%   |
| Weighted KNN               | 55.60%   | 53.80%   | 76.90%     | 35.90%     | 12.80%    | 20.50%    | 33.30%    | 43.60%   | 12.80%    | 10.30%   |
| Logistic Regression Kernel | 53.80%   | 43.60%   | 64.10%     | 53.80%     | 12.80%    | 20.50%    | 23.10%    | 25.60%   | 33.30%    | 23.10%   |
| Multilayer Perceptron      | 52.99%   | 53.60%   | 46.80%     | 59.50%     | 17.94%    | 15.38%    | 43.58%    | 20.51%   | 17.94%    | 25.64%   |
| Cosine KNN                 | 52.10%   | 61.50%   | 59.00%     | 35.90%     | 33.30%    | 38.50%    | 25.60%    | 25.60%   | 12.80%    | 7.70%    |
| SVM Kernel                 | 52.10%   | 41.00%   | 56.40%     | 59.00%     | 17.90%    | 30.80%    | 17.90%    | 10.30%   | 41.00%    | 25.60%   |
| Gaussian Naïve Bayes       | 52.10%   | 48.70%   | 74.40%     | 33.30%     | 23.10%    | 20.50%    | 43.60%    | 46.20%   | 7.70%     | 2.60%    |
| SubspaceKNN                | 52.10%   | 51.30%   | 53.80%     | 51.30%     | 30.80%    | 33.30%    | 23.10%    | 15.40%   | 25.60%    | 15.40%   |
| Kernel Naïve Bayes         | 51.30%   | 61.50%   | 43.60%     | 48.70%     | 33.30%    | 20.50%    | 20.50%    | 30.80%   | 1.90%     | 23.10%   |

Εικόνα 7.2 :Παράθυρο 3 δευτερολέπτων, 5 Cross Validation

| Αλγόριθμος                 | Accuracy | P(error) | P(opening) | P(warning) | E(op>err) | E(wa>err) | E(err>op) | E(wa>op) | E(err>wa) | E(op>wa) |
|----------------------------|----------|----------|------------|------------|-----------|-----------|-----------|----------|-----------|----------|
| Bagged Trees               | 73.50%   | 82.10%   | 71.80%     | 66.70%     | 17.90%    | 17.90%    | 15.40%    | 15.40%   | 2.60%     | 10.30%   |
| Coarse Tree                | 68.40%   | 82.10%   | 53.80%     | 69.20%     | 33.30%    | 23.10%    | 7.70%     | 7.70%    | 10.30%    | 12.80%   |
| Quadratic SVM              | 65.00%   | 66.70%   | 71.80%     | 56.40%     | 17.90%    | 17.90%    | 20.50%    | 25.60%   | 12.80%    | 10.30%   |
| Fine Tree                  | 64.10%   | 64.10%   | 56.40%     | 71.80%     | 23.10%    | 10.30%    | 17.90%    | 17.90%   | 17.90%    | 20.50%   |
| Medium Tree                | 64.10%   | 64.10%   | 56.40%     | 71.80%     | 23.10%    | 10.30%    | 17.90%    | 17.90%   | 17.90%    | 20.50%   |
| Fine KNN                   | 64.10%   | 66.70%   | 17.90%     | 15.40%     | 7.70%     | 23.10%    | 17.90%    | 25.60%   | 15.40%    | 17.90%   |
| Kstar                      | 62.39%   | 63.90%   | 54.90%     | 73.30%     | 17.94%    | 15.38%    | 30.76%    | 28.20%   | 10.25%    | 10.25%   |
| Cubic SVM                  | 61.50%   | 66.70%   | 69.20%     | 48.70%     | 17.90%    | 23.10%    | 17.90%    | 28.20%   | 15.40%    | 12.80%   |
| Boosted Trees              | 61.50%   | 69.20%   | 56.40%     | 59.00%     | 30.80%    | 28.20%    | 20.50%    | 12.80%   | 10.30%    | 12.80%   |
| Medium Neural Network      | 61.50%   | 66.70%   | 61.50%     | 56.40%     | 25.60%    | 25.60%    | 15.40%    | 17.90%   | 17.90%    | 12.80%   |
| RUSBoosted Trees           | 60.70%   | 66.70%   | 48.70%     | 66.70%     | 28.20%    | 17.90%    | 23.10%    | 15.40%   | 10.30%    | 23.10%   |
| SubspaceKNN                | 60.70%   | 61.50%   | 66.70%     | 53.80%     | 15.40%    | 30.80%    | 23.10%    | 15.40%   | 15.40%    | 17.90%   |
| IBK                        | 59.82%   | 61.50%   | 59.50%     | 58.30%     | 15.38%    | 23.07%    | 20.51%    | 23.07%   | 17.94%    | 20.51%   |
| Wide Neural Network        | 59.80%   | 69.20%   | 59.00%     | 51.30%     | 25.60%    | 28.20%    | 23.10%    | 20.50%   | 7.70%     | 15.40%   |
| Multilayer Perceptron      | 58.11%   | 55.00%   | 52.60%     | 66.70%     | 28.20%    | 17.94%    | 30.76%    | 15.38%   | 12.82%    | 20.51%   |
| Weighted KNN               | 57.30%   | 56.40%   | 79.50%     | 35.90%     | 7.70%     | 20.50%    | 30.80%    | 43.60%   | 12.80%    | 12.80%   |
| Fine Gaussian SVM          | 56.40%   | 41.00%   | 48.70%     | 79.50%     | 5.10%     | 7.70%     | 7.70%     | 12.80%   | 51.30%    | 46.20%   |
| Kernel Naïve Bayes         | 54.70%   | 59.00%   | 59.00%     | 46.20%     | 33.30%    | 23.10%    | 17.90%    | 30.80%   | 23.10%    | 7.70%    |
| Medium Gaussian SVM        | 54.70%   | 53.80%   | 64.10%     | 46.20%     | 28.20%    | 23.10%    | 23.10%    | 30.80%   | 23.10%    | 7.70%    |
| Cosine KNN                 | 54.70%   | 59.00%   | 61.50%     | 43.60%     | 28.20%    | 35.90%    | 20.50%    | 20.50%   | 20.50%    | 10.30%   |
| Logistic Regression Kernel | 54.70%   | 43.60%   | 66.70%     | 53.80%     | 15.40%    | 20.50%    | 25.60%    | 25.60%   | 30.80%    | 17.90%   |
| SVM Kernel                 | 53.80%   | 43.60%   | 61.50%     | 56.40%     | 20.50%    | 28.20%    | 17.90%    | 15.40%   | 38.50%    | 17.90%   |
| Narrow Neural Network      | 52.10%   | 66.70%   | 46.20%     | 43.60%     | 30.80%    | 30.80%    | 15.40%    | 25.60%   | 17.90%    | 23.10%   |
| Gaussian Naïve Bayes       | 50.40%   | 46.20%   | 74.40%     | 30.80%     | 20.50%    | 20.50%    | 43.60%    | 48.70%   | 10.30%    | 5.10%    |

Εικόνα 7.3 : Παράθυρο 3 δευτερολέπτων, 10 Cross Validation

| Αλγόριθμος                 | Accuracy | P(error) | P(opening) | P(warning) | E(op>err) | E(wa>err) | E(err>op) | E(wa>op) | E(err>wa) | E(op>wa) |
|----------------------------|----------|----------|------------|------------|-----------|-----------|-----------|----------|-----------|----------|
| Bagged Trees               | 74.30%   | 90.90%   | 75.00%     | 58.30%     | 8.30%     | 16.70%    | 9.10%     | 25.00%   | 0.00%     | 16.70%   |
| Fine Tree                  | 68.60%   | 81.80%   | 66.70%     | 58.30%     | 8.30%     | 8.30%     | 9.10%     | 33.30%   | 9.10%     | 25.00%   |
| Medium Tree                | 68.60%   | 81.80%   | 66.70%     | 58.30%     | 8.30%     | 8.30%     | 9.10%     | 33.30%   | 9.10%     | 25.00%   |
| Quadratic SVM              | 65.70%   | 100.00%  | 66.70%     | 33.30%     | 25.00%    | 33.30%    | 0.00%     | 33.30%   | 0.00%     | 8.30%    |
| Wide Neural Network        | 65.70%   | 100.00%  | 50.00%     | 50.00%     | 50.00%    | 25.00%    | 0.00%     | 25.00%   | 0.00%     | 0.00%    |
| Narrow Neural Network      | 65.70%   | 100.00%  | 58.30%     | 41.70%     | 33.30%    | 33.30%    | 0.00%     | 25.00%   | 0.00%     | 8.30%    |
| Coarse Tree                | 62.90%   | 90.90%   | 50.00%     | 50.00%     | 41.70%    | 25.00%    | 0.00%     | 25.00%   | 9.10%     | 8.30%    |
| Cubic SVM                  | 62.90%   | 90.90%   | 66.70%     | 33.30%     | 33.30%    | 33.30%    | 0.00%     | 33.30%   | 9.10%     | 0.00%    |
| Fine KNN                   | 60.00%   | 81.80%   | 58.30%     | 41.70%     | 25.00%    | 25.00%    | 0.00%     | 33.30%   | 18.20%    | 16.70%   |
| Medium Neural Network      | 60.00%   | 81.80%   | 50.00%     | 50.00%     | 41.70%    | 25.00%    | 18.20%    | 25.00%   | 0.00%     | 8.30%    |
| Kernel Naïve Bayes         | 60.00%   | 90.90%   | 41.70%     | 50.00%     | 58.30%    | 25.00%    | 0.00%     | 25.00%   | 9.10%     | 0.00%    |
| RUSBoosted Trees           | 57.10%   | 72.70%   | 58.30%     | 41.70%     | 16.70%    | 16.70%    | 18.20%    | 41.70%   | 9.10%     | 25.00%   |
| Weighted KNN               | 57.10%   | 81.80%   | 66.70%     | 25.00%     | 33.30%    | 25.00%    | 9.10%     | 50.00%   | 9.10%     | 0.00%    |
| Fine Gaussian SVM          | 57.10%   | 63.60%   | 50.00%     | 58.30%     | 25.00%    | 16.70%    | 0.00%     | 25.00%   | 36.40%    | 25.00%   |
| SVM Kernel                 | 57.10%   | 72.70%   | 75.00%     | 25.00%     | 25.00%    | 50.00%    | 9.10%     | 25.00%   | 18.20%    | 0.00%    |
| SubspaceKNN                | 57.10%   | 54.50%   | 75.00%     | 41.70%     | 16.70%    | 50.00%    | 27.30%    | 8.30%    | 18.20%    | 8.30%    |
| Multilayer Perceptron      | 52.43%   | 47.40%   | 50.00%     | 60.00%     | 10.25%    | 15.38%    | 30.76%    | 17.94%   | 17.94%    | 7.69%    |
| Medium Gaussian SVM        | 51.40%   | 63.60%   | 50.00%     | 41.70%     | 41.70%    | 25.00%    | 0.00%     | 33.30%   | 36.40%    | 8.30%    |
| Gaussian Naïve Bayes       | 51.40%   | 81.80%   | 58.30%     | 16.70%     | 41.70%    | 33.30%    | 18.20%    | 50.00%   | 0.00%     | 0.00%    |
| IBK                        | 51.21%   | 42.10%   | 50.00%     | 60.00%     | 10.25%    | 17.94%    | 33.30%    | 15.38%   | 17.94%    | 7.69%    |
| Cosine KNN                 | 48.60%   | 90.90%   | 33.30%     | 25.00%     | 50.00%    | 50.00%    | 9.10%     | 25.00%   | 0.00%     | 16.70%   |
| Logistic Regression Kernel | 48.60%   | 72.70%   | 75.00%     | 0.00%      | 16.70%    | 41.70%    | 9.10%     | 58.30%   | 18.20%    | 8.30%    |
| Kstar                      | 43.90%   | 30.00%   | 46.20%     | 52.20%     | 15.38%    | 20.51%    | 33.30%    | 20.51%   | 23.07%    | 5.12%    |
| Boosted Trees              | 31.40%   | 100.00%  | 0.00%      | 0.00%      | 100.00%   | 100.00%   | 0.00%     | 0.00%    | 0.00%     | 0.00%    |

Εικόνα 7.4: Παράθυρο 3 δευτερολέπτων, 30 Holdout Validation

| Αλγόριθμος                 | Accuracy | P(error) | P(opening) | P(warning) | E(op>err) | E(wa>err) | E(err>op) | E(wa>op) | E(err>wa) | E(op>wa) |
|----------------------------|----------|----------|------------|------------|-----------|-----------|-----------|----------|-----------|----------|
| Bagged Trees               | 67.40%   | 75.00%   | 60.00%     | 66.70%     | 6.70%     | 20.00%    | 18.80%    | 13.30%   | 6.20%     | 33.30%   |
| Quadratic SVM              | 67.40%   | 68.80%   | 66.70%     | 66.70%     | 13.30%    | 20.00%    | 25.00%    | 13.30%   | 6.20%     | 20.00%   |
| Cubic SVM                  | 65.20%   | 68.80%   | 60.00%     | 66.70%     | 13.30%    | 13.30%    | 25.00%    | 20.00%   | 6.20%     | 26.70%   |
| Fine KNN                   | 65.20%   | 81.20%   | 66.70%     | 46.70%     | 6.70%     | 33.30%    | 12.50%    | 20.00%   | 6.20%     | 26.70%   |
| RUSBoosted Trees           | 65.20%   | 87.50%   | 33.30%     | 73.30%     | 33.30%    | 13.30%    | 12.50%    | 13.30%   | 0.00%     | 33.30%   |
| Fine Tree                  | 63.00%   | 81.20%   | 40.00%     | 66.70%     | 33.30%    | 6.70%     | 6.20%     | 26.70%   | 12.50%    | 26.70%   |
| Medium Tree                | 63.00%   | 81.20%   | 40.00%     | 66.70%     | 33.30%    | 6.70%     | 6.20%     | 26.70%   | 12.50%    | 26.70%   |
| Wide Neural Network        | 60.90%   | 68.80%   | 53.30%     | 60.00%     | 26.70%    | 26.70%    | 31.20%    | 13.30%   | 0.00%     | 20.00%   |
| Medium Neural Network      | 60.90%   | 75.00%   | 46.70%     | 60.00%     | 26.70%    | 26.70%    | 25.00%    | 13.30%   | 0.00%     | 26.70%   |
| Weighted KNN               | 60.90%   | 56.20%   | 80.00%     | 46.70%     | 0.00%     | 20.00%    | 43.80%    | 33.30%   | 0.00%     | 20.00%   |
| Medium Gaussian SVM        | 60.90%   | 56.20%   | 73.30%     | 53.30%     | 20.00%    | 13.30%    | 31.20%    | 33.30%   | 12.50%    | 6.70%    |
| Narrow Neural Network      | 58.70%   | 68.80%   | 46.70%     | 60.00%     | 40.00%    | 20.00%    | 25.00%    | 20.00%   | 6.20%     | 13.30%   |
| Fine Gaussian SVM          | 58.70%   | 56.20%   | 46.70%     | 73.30%     | 0.00%     | 13.30%    | 6.20%     | 13.30%   | 37.50%    | 53.30%   |
| Kernel Naïve Bayes         | 56.50%   | 62.50%   | 46.70%     | 60.00%     | 33.30%    | 26.70%    | 25.00%    | 13.30%   | 12.50%    | 20.00%   |
| Coarse Tree                | 54.30%   | 43.80%   | 60.00%     | 60.00%     | 13.30%    | 6.70%     | 50.00%    | 33.30%   | 6.20%     | 26.70%   |
| Gaussian Naïve Bayes       | 52.20%   | 50.00%   | 73.30%     | 33.30%     | 20.00%    | 33.30%    | 43.80%    | 33.30%   | 6.20%     | 6.70%    |
| Cosine KNN                 | 52.20%   | 56.20%   | 66.70%     | 33.30%     | 26.70%    | 33.30%    | 31.20%    | 33.30%   | 12.50%    | 6.70%    |
| Multilayer Perceptron      | 51.42%   | 46.20%   | 53.80%     | 55.60%     | 15.38%    | 20.51%    | 15.38%    | 15.38%   | 12.82%    | 7.69%    |
| SVM Kernel                 | 50.00%   | 43.80%   | 53.30%     | 53.30%     | 20.00%    | 33.30%    | 25.00%    | 13.30%   | 31.20%    | 26.70%   |
| Logistic Regression Kernel | 50.00%   | 37.50%   | 53.30%     | 60.00%     | 13.30%    | 26.70%    | 25.00%    | 13.30%   | 37.50%    | 33.30%   |
| SubspaceKNN                | 50.00%   | 50.00%   | 66.70%     | 33.30%     | 13.30%    | 33.30%    | 43.80%    | 33.30%   | 6.20%     | 20.00%   |
| IBK                        | 44.28%   | 33.30%   | 46.70%     | 52.60%     | 15.38%    | 20.51%    | 25.64%    | 15.38%   | 15.38%    | 7.69%    |
| Kstar                      | 38.57%   | 28.60%   | 38.20%     | 53.30%     | 20.51%    | 17.94%    | 30.76%    | 23.07%   | 12.82%    | 5.12%    |
| Boosted Trees              | 32.60%   | 0.00%    | 100.00%    | 0.00%      | 0.00%     | 0.00%     | 100.00%   | 100.00%  | 0.00%     | 0.00%    |

Εικόνα 7.5: Παράθυρο 3 δευτερολέπτων, 40 Holdout Validation

| Αλγόριθμος                 | Accuracy | P(error) | P(opening) | P(warning) | E(op>err) | E(wa>err) | E(err>op) | E(wa>op) | E(err>wa) | E(op>wa) |
|----------------------------|----------|----------|------------|------------|-----------|-----------|-----------|----------|-----------|----------|
| Bagged Trees               | 89.50%   | 94.00%   | 85.50%     | 88.90%     | 6.00%     | 0.90%     | 4.30%     | 10.30%   | 1.70%     | 8.50%    |
| SubspaceKNN                | 88.30%   | 89.70%   | 87.20%     | 88.00%     | 3.40%     | 4.30%     | 9.40%     | 7.70%    | 0.90%     | 9.40%    |
| Boosted Trees              | 87.70%   | 94.00%   | 82.10%     | 87.20%     | 10.30%    | 2.60%     | 4.30%     | 10.30%   | 1.70%     | 7.70%    |
| Multilayer Perceptron      | 85.75%   | 86.80%   | 82.50%     | 87.90%     | 11.11%    | 2.56%     | 6.83%     | 10.25%   | 3.41%     | 8.54%    |
| Cubic SVM                  | 85.50%   | 89.70%   | 81.20%     | 85.50%     | 6.80%     | 4.30%     | 5.10%     | 10.30%   | 5.10%     | 12.00%   |
| Kstar                      | 84.90%   | 84.70%   | 80.80%     | 89.70%     | 9.40%     | 6.83%     | 8.54%     | 11.11%   | 1.70%     | 7.69%    |
| Quadratic SVM              | 84.00%   | 85.50%   | 81.20%     | 85.50%     | 9.40%     | 6.00%     | 12.00%    | 8.50%    | 2.60%     | 9.40%    |
| Fine KNN                   | 83.80%   | 84.60%   | 82.90%     | 83.80%     | 6.00%     | 7.70%     | 6.80%     | 8.50%    | 8.50%     | 11.10%   |
| Weighted KNN               | 83.50%   | 85.50%   | 82.90%     | 82.10%     | 9.40%     | 7.70%     | 7.70%     | 10.30%   | 6.80%     | 7.70%    |
| IBK                        | 82.05%   | 83.20%   | 80.70%     | 82.10%     | 10.25%    | 7.69%     | 5.12%     | 13.67%   | 5.98%     | 11.11%   |
| Wide Neural Network        | 81.80%   | 83.80%   | 76.90%     | 84.60%     | 7.70%     | 5.10%     | 10.30%    | 10.30%   | 6.00%     | 15.40%   |
| Medium Neural Network      | 81.80%   | 83.80%   | 77.80%     | 83.80%     | 14.50%    | 4.30%     | 10.30%    | 12.00%   | 6.00%     | 7.70%    |
| Fine Tree                  | 79.80%   | 86.30%   | 73.50%     | 79.50%     | 12.80%    | 7.70%     | 7.70%     | 12.80%   | 6.00%     | 13.70%   |
| SVM Kernel                 | 79.50%   | 81.20%   | 75.20%     | 82.10%     | 6.80%     | 12.00%    | 7.70%     | 6.00%    | 11.10%    | 17.90%   |
| RUSBoosted Trees           | 78.10%   | 82.10%   | 73.50%     | 78.60%     | 14.50%    | 5.10%     | 11.10%    | 16.20%   | 6.80%     | 12.00%   |
| Fine Gaussian SVM          | 78.10%   | 67.50%   | 79.50%     | 87.20%     | 0.90%     | 0.00%     | 12.00%    | 12.80%   | 20.50%    | 19.70%   |
| Medium Tree                | 77.80%   | 80.30%   | 76.10%     | 76.90%     | 11.10%    | 8.50%     | 12.80%    | 14.50%   | 6.80%     | 12.80%   |
| Narrow Neural Network      | 77.80%   | 87.20%   | 65.00%     | 81.20%     | 18.80%    | 6.80%     | 638.00%   | 12.00%   | 6.00%     | 16.20%   |
| Logistic Regression Kernel | 77.20%   | 74.40%   | 76.90%     | 80.30%     | 10.30%    | 7.70%     | 12.80%    | 12.00%   | 12.80%    | 12.80%   |
| Medium Gaussian SVM        | 76.40%   | 78.60%   | 82.10%     | 68.40%     | 11.10%    | 12.00%    | 15.40%    | 19.70%   | 6.00%     | 6.80%    |
| Kernel Naïve Bayes         | 74.40%   | 81.20%   | 81.20%     | 60.70%     | 13.70%    | 8.50%     | 16.80%    | 30.80%   | 0.00%     | 5.10%    |
| Cosine KNN                 | 63.00%   | 78.60%   | 58.10%     | 52.10%     | 28.20%    | 37.60%    | 12.80%    | 10.30%   | 8.50%     | 13.70%   |
| Coarse Tree                | 58.10%   | 70.90%   | 41.00%     | 62.40%     | 41.90%    | 25.60%    | 23.90%    | 12.00%   | 5.10%     | 17.10%   |
| Gaussian Naïve Bayes       | 55.80%   | 53.00%   | 82.10%     | 32.50%     | 16.20%    | 24.80%    | 42.70%    | 42.70%   | 4.30%     | 1.70%    |

Εικόνα 7.6 :Παράθυρο 1 δευτερολέπτου, 5 Cross Validation

| Αλγόριθμος                 | Accuracy | P(error) | P(opening) | P(warning) | E(op>err) | E(wa>err) | E(err>op) | E(wa>op) | E(err>wa) | E(op>wa) |
|----------------------------|----------|----------|------------|------------|-----------|-----------|-----------|----------|-----------|----------|
| Bagged Trees               | 90.00%   | 96.60%   | 87.20%     | 86.30%     | 4.30%     | 2.60%     | 3.40%     | 11.10%   | 0.00%     | 8.50%    |
| SubspaceKNN                | 88.60%   | 92.30%   | 86.30%     | 87.20%     | 2.60%     | 6.00%     | 6.80%     | 6.80%    | 0.90%     | 11.10%   |
| Kstar                      | 87.46%   | 89.20%   | 84.30%     | 89.10%     | 5.12%     | 5.98%     | 5.98%     | 10.25%   | 2.56%     | 7.69%    |
| Cubic SVM                  | 87.20%   | 91.50%   | 86.30%     | 83.80%     | 4.30%     | 4.30%     | 3.40%     | 12.00%   | 5.10%     | 9.40%    |
| Quadratic SVM              | 85.80%   | 87.20%   | 83.80%     | 86.30%     | 7.70%     | 2.60%     | 9.40%     | 11.10%   | 3.40%     | 8.50%    |
| Fine KNN                   | 85.80%   | 87.20%   | 84.60%     | 85.50%     | 6.80%     | 5.10%     | 5.10%     | 9.40%    | 7.70%     | 8.50%    |
| Multilayer Perceptron      | 85.75%   | 87.50%   | 83.20%     | 86.40%     | 7.69%     | 5.12%     | 8.54%     | 7.69%    | 1.70%     | 11.96%   |
| Weighted KNN               | 85.50%   | 90.60%   | 84.60%     | 81.20%     | 7.70%     | 7.70%     | 5.10%     | 11.10%   | 4.30%     | 7.70%    |
| Boosted Trees              | 85.20%   | 89.70%   | 78.60%     | 87.20%     | 12.00%    | 2.60%     | 8.50%     | 10.30%   | 1.70%     | 9.40%    |
| IBK                        | 85.18%   | 88.90%   | 82.20%     | 84.50%     | 7.69%     | 3.41%     | 5.12%     | 12.82%   | 5.98%     | 9.40%    |
| Wide Neural Network        | 84.00%   | 92.30%   | 81.20%     | 78.60%     | 10.30%    | 8.50%     | 4.30%     | 12.80%   | 3.40%     | 8.50%    |
| SVM Kernel                 | 82.60%   | 88.90%   | 76.90%     | 82.10%     | 7.70%     | 8.50%     | 7.70%     | 9.40%    | 3.40%     | 15.40%   |
| Fine Gaussian SVM          | 82.30%   | 72.60%   | 88.00%     | 86.30%     | 0.90%     | 0.00%     | 18.80%    | 13.70%   | 8.50%     | 11.10%   |
| Fine Tree                  | 82.10%   | 89.70%   | 81.20%     | 75.20%     | 11.10%    | 7.70%     | 6.80%     | 17.10%   | 3.40%     | 7.70%    |
| Medium Neural Network      | 82.10%   | 87.20%   | 75.20%     | 83.80%     | 12.80%    | 6.00%     | 10.30%    | 10.30%   | 2.60%     | 12.00%   |
| Narrow Neural Network      | 81.50%   | 85.50%   | 79.50%     | 79.50%     | 11.10%    | 9.40%     | 8.50%     | 11.10%   | 6.00%     | 9.40%    |
| Logistic Regression Kernel | 80.60%   | 80.30%   | 77.80%     | 83.80%     | 8.50%     | 6.80%     | 10.30%    | 9.40%    | 9.40%     | 13.70%   |
| RUSBoosted Trees           | 79.50%   | 86.30%   | 73.50%     | 78.60%     | 16.20%    | 8.50%     | 9.40%     | 12.80%   | 4.30%     | 10.30%   |
| Medium Tree                | 79.20%   | 87.20%   | 74.40%     | 76.10%     | 15.40%    | 10.30%    | 9.40%     | 13.70%   | 3.40%     | 10.30%   |
| Medium Gaussian SVM        | 75.50%   | 78.60%   | 81.20%     | 66.70%     | 11.10%    | 12.80%    | 14.50%    | 20.50%   | 6.80%     | 7.70%    |
| Kernel Naïve Bayes         | 74.10%   | 80.30%   | 82.90%     | 59.00%     | 12.00%    | 6.00%     | 19.70%    | 35.00%   | 0.00%     | 5.10%    |
| Cosine KNN                 | 67.50%   | 87.20%   | 63.20%     | 52.10%     | 25.60%    | 37.60%    | 6.00%     | 10.30%   | 6.80%     | 11.10%   |
| Gaussian Naïve Bayes       | 57.30%   | 53.00%   | 84.60%     | 34.20%     | 13.70%    | 24.80%    | 44.40%    | 41.00%   | 2.60%     | 1.70%    |
| Coarse Tree                | 56.70%   | 70.10%   | 39.30%     | 60.70%     | 46.20%    | 22.20%    | 23.10%    | 17.10%   | 6.80%     | 14.50%   |

Εικόνα 7.7: Παράθυρο 1 δευτερολέπτου, 10 Cross Validation

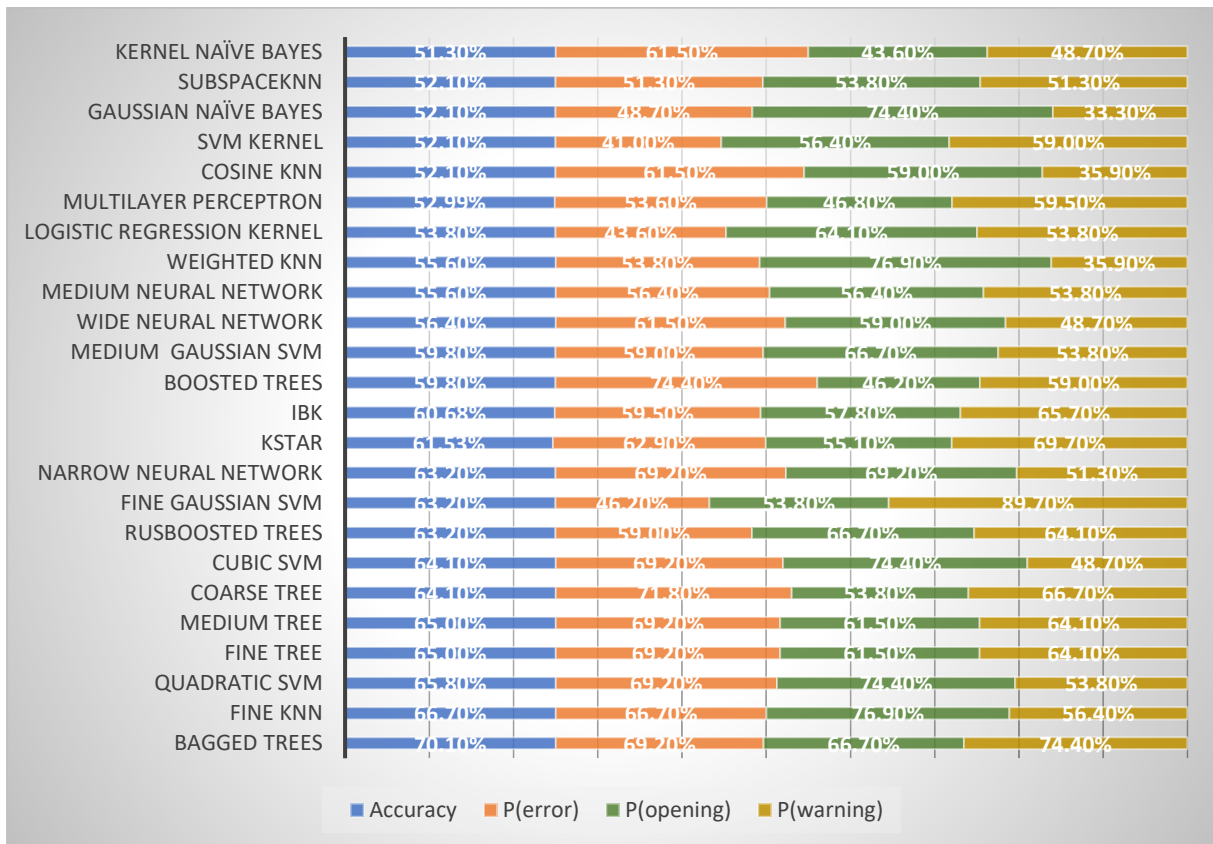
| Αλγόριθμος                 | Accuracy | P(error) | P(opening) | P(warning) | E(op>err) | E(wa>err) | E(err>op) | E(wa>op) | E(err>wa) | E(op>wa) |
|----------------------------|----------|----------|------------|------------|-----------|-----------|-----------|----------|-----------|----------|
| SubspaceKNN                | 95.20%   | 100.00%  | 94.30%     | 91.40%     | 0.00%     | 5.70%     | 0.00%     | 2.90%    | 0.00%     | 5.70%    |
| Bagged Trees               | 93.30%   | 97.10%   | 94.30%     | 88.60%     | 2.90%     | 0.00%     | 2.90%     | 11.40%   | 0.00%     | 2.90%    |
| Boosted Trees              | 90.50%   | 94.30%   | 91.40%     | 85.70%     | 2.90%     | 0.00%     | 5.70%     | 14.30%   | 0.00%     | 5.70%    |
| Cubic SVM                  | 88.60%   | 97.10%   | 91.40%     | 77.10%     | 0.00%     | 5.70%     | 2.90%     | 17.10%   | 0.00%     | 8.60%    |
| Fine KNN                   | 88.60%   | 91.40%   | 88.60%     | 85.70%     | 5.70%     | 2.90%     | 5.70%     | 11.40%   | 2.90%     | 5.70%    |
| Quadratic SVM              | 86.70%   | 88.60%   | 82.90%     | 88.60%     | 8.60%     | 2.90%     | 8.60%     | 8.60%    | 2.90%     | 8.60%    |
| Medium Neural              | 82.90%   | 91.40%   | 71.40%     | 85.70%     | 20.00%    | 2.90%     | 5.70%     | 11.40%   | 2.90%     | 8.60%    |
| Weighted KNN               | 82.90%   | 88.60%   | 82.90%     | 77.10%     | 14.30%    | 11.40%    | 5.70%     | 11.40%   | 5.70%     | 2.90%    |
| Narrow Neural Network      | 81.90%   | 91.40%   | 71.40%     | 82.90%     | 17.10%    | 11.40%    | 5.70%     | 5.70%    | 2.90%     | 11.40%   |
| Fine Gaussian SVM          | 81.90%   | 74.30%   | 97.10%     | 74.30%     | 0.00%     | 0.00%     | 22.90%    | 25.70%   | 2.90%     | 2.90%    |
| Wide Neural Network        | 81.00%   | 85.70%   | 74.30%     | 82.90%     | 14.30%    | 8.60%     | 11.40%    | 8.60%    | 2.90%     | 11.40%   |
| Kernel Naïve Bayes         | 77.10%   | 85.70%   | 80.00%     | 65.70%     | 11.40%    | 2.90%     | 14.30%    | 31.40%   | 0.00%     | 8.60%    |
| SVM Kernel                 | 77.10%   | 71.40%   | 71.40%     | 88.60%     | 14.30%    | 5.70%     | 14.30%    | 5.70%    | 14.30%    | 14.30%   |
| Medium Gaussian SVM        | 75.20%   | 82.90%   | 82.90%     | 60.00%     | 11.40%    | 14.30%    | 14.30%    | 25.70%   | 2.90%     | 5.70%    |
| Fine Tree                  | 74.30%   | 77.10%   | 65.70%     | 80.00%     | 17.10%    | 11.40%    | 22.90%    | 8.60%    | 0.00%     | 17.10%   |
| RUSBoosted Trees           | 74.30%   | 74.30%   | 71.40%     | 77.10%     | 14.30%    | 8.60%     | 25.70%    | 14.30%   | 0.00%     | 14.30%   |
| Medium Tree                | 73.30%   | 77.10%   | 68.60%     | 74.30%     | 17.10%    | 11.40%    | 22.90%    | 14.30%   | 0.00%     | 14.30%   |
| Logistic Regression Kernel | 71.40%   | 60.00%   | 80.00%     | 74.30%     | 8.60%     | 0.00%     | 22.90%    | 25.70%   | 17.10%    | 11.40%   |
| Multilayer Perceptron      | 67.47%   | 68.50%   | 59.10%     | 76.30%     | 11.11%    | 8.54%     | 23.93%    | 8.54%    | 8.54%     | 7.69%    |
| Kstar                      | 65.04%   | 83.90%   | 56.40%     | 62.90%     | 4.27%     | 3.41%     | 19.65%    | 17.94%   | 15.38%    | 12.82%   |
| Cosine KNN                 | 61.90%   | 82.90%   | 62.90%     | 40.00%     | 25.70%    | 48.60%    | 11.40%    | 11.40%   | 5.70%     | 11.40%   |
| IBK                        | 60.56%   | 78.80%   | 53.80%     | 57.30%     | 5.98%     | 3.41%     | 20.51%    | 15.38%   | 19.65%    | 17.94%   |
| Coarse Tree                | 60.00%   | 74.30%   | 68.60%     | 37.10%     | 31.40%    | 40.00%    | 25.70%    | 22.90%   | 0.00%     | 0.00%    |
| Gaussian Naïve Bayes       | 54.30%   | 54.30%   | 80.00%     | 28.60%     | 14.30%    | 34.30%    | 42.90%    | 37.10%   | 2.90%     | 5.70%    |

Εικόνα 7.8 :Παράθυρο 1 δευτερολέπτου, 30 Holdout Validation

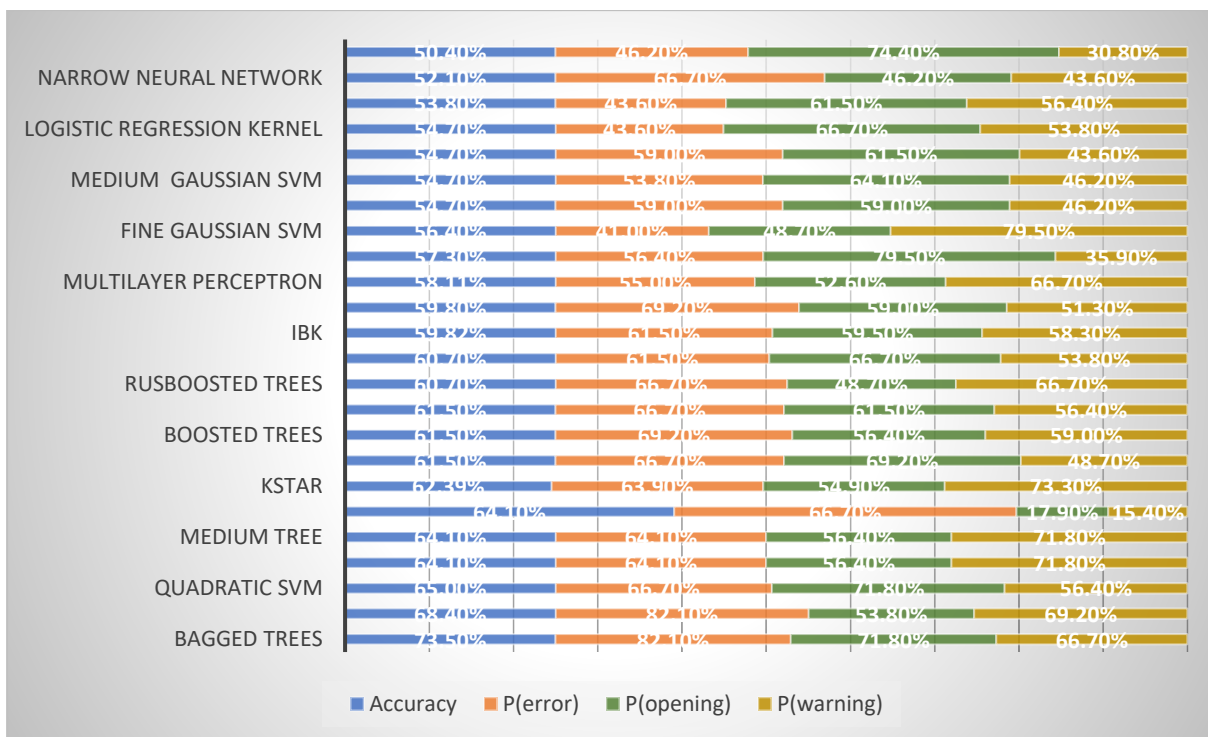
| Αλγόριθμος                 | Accuracy | P(error) | P(opening) | P(warning) | E(op>err) | E(wa>err) | E(err>op) | E(wa>op) | E(err>wa) | E(op>wa) |
|----------------------------|----------|----------|------------|------------|-----------|-----------|-----------|----------|-----------|----------|
| Cubic SVM                  | 85.70%   | 95.70%   | 76.60%     | 85.10%     | 6.40%     | 0.00%     | 2.20%     | 14.90%   | 2.20%     | 17.00%   |
| Boosted Trees              | 84.30%   | 95.70%   | 74.50%     | 83.00%     | 10.60%    | 2.10%     | 4.30%     | 14.90%   | 0.00%     | 14.90%   |
| Quadratic SVM              | 82.90%   | 87.00%   | 74.50%     | 87.20%     | 12.80%    | 2.10%     | 8.70%     | 10.60%   | 4.30%     | 12.80%   |
| Bagged Trees               | 82.10%   | 95.70%   | 68.10%     | 83.00%     | 19.10%    | 4.30%     | 4.30%     | 12.80%   | 0.00%     | 12.80%   |
| SubspaceKNN                | 81.40%   | 84.80%   | 76.60%     | 83.00%     | 8.50%     | 6.40%     | 8.70%     | 10.60%   | 6.50%     | 14.90%   |
| Fine KNN                   | 80.70%   | 84.80%   | 76.60%     | 80.90%     | 4.30%     | 6.40%     | 8.70%     | 12.80%   | 6.50%     | 19.10%   |
| Weighted KNN               | 80.00%   | 78.30%   | 83.00%     | 78.70%     | 2.10%     | 4.30%     | 6.50%     | 17.00%   | 15.20%    | 14.90%   |
| Medium Neural              | 79.30%   | 89.10%   | 61.70%     | 87.20%     | 14.90%    | 6.40%     | 6.50%     | 6.40%    | 4.30%     | 23.40%   |
| Wide Neural Network        | 77.10%   | 82.60%   | 63.80%     | 85.10%     | 12.80%    | 4.30%     | 10.90%    | 10.60%   | 6.50%     | 23.40%   |
| Fine Tree                  | 75.70%   | 87.00%   | 74.50%     | 66.00%     | 14.90%    | 12.80%    | 13.00%    | 21.30%   | 0.00%     | 10.60%   |
| Medium Tree                | 75.70%   | 87.00%   | 74.50%     | 66.00%     | 14.90%    | 12.80%    | 13.00%    | 21.30%   | 0.00%     | 10.60%   |
| RUSBoosted Trees           | 75.70%   | 89.10%   | 63.80%     | 74.50%     | 23.40%    | 8.50%     | 10.90%    | 17.00%   | 0.00%     | 12.80%   |
| Narrow Neural Network      | 72.90%   | 76.10%   | 59.60%     | 83.00%     | 19.10%    | 2.10%     | 6.50%     | 14.90%   | 17.40%    | 21.30%   |
| SVM Kernel                 | 71.40%   | 67.40%   | 66.00%     | 80.90%     | 21.30%    | 10.60%    | 6.50%     | 8.50%    | 26.10%    | 12.80%   |
| Fine Gaussian SVM          | 68.60%   | 71.70%   | 46.80%     | 87.20%     | 0.00%     | 0.00%     | 4.30%     | 12.80%   | 23.90%    | 53.20%   |
| Kernel Naïve Bayes         | 67.90%   | 67.40%   | 76.60%     | 59.60%     | 14.90%    | 6.40%     | 23.90%    | 34.00%   | 8.70%     | 8.50%    |
| Medium Gaussian SVM        | 67.90%   | 65.20%   | 76.60%     | 61.70%     | 12.80%    | 14.90%    | 10.90%    | 23.40%   | 23.90%    | 10.60%   |
| Logistic Regression Kernel | 67.90%   | 69.60%   | 68.10%     | 66.00%     | 21.30%    | 10.60%    | 6.70%     | 23.40%   | 21.70%    | 10.60%   |
| Coarse Tree                | 62.10%   | 65.20%   | 57.40%     | 63.80%     | 27.70%    | 31.90%    | 19.60%    | 4.30%    | 15.20%    | 14.90%   |
| Cosine KNN                 | 60.00%   | 54.30%   | 59.60%     | 66.00%     | 27.70%    | 19.10%    | 17.40%    | 14.90%   | 28.30%    | 12.80%   |
| Gaussian Naïve Bayes       | 55.70%   | 60.90%   | 76.60%     | 29.80%     | 21.30%    | 29.80%    | 39.10%    | 40.40%   | 0.00%     | 2.10%    |
| Multilayer Perceptron      | 70.14%   | 64.50%   | 70.00%     | 78.90%     | 12.82%    | 12.82%    | 11.11%    | 6.83%    | 5.12%     | 5.12%    |
| IBK                        | 67.29%   | 78.70%   | 66.70%     | 59.30%     | 5.98%     | 5.12%     | 7.69%     | 11.96%   | 13.67%    | 14.52%   |
| Kstar                      | 73.45%   | 78.80%   | 68.80%     | 73.80%     | 5.12%     | 6.83%     | 11.11%    | 10.25%   | 6.83%     | 7.69%    |

Εικόνα 7.9 :Παράθυρο 1 δευτερολέπτου, 40 Holdout Validation

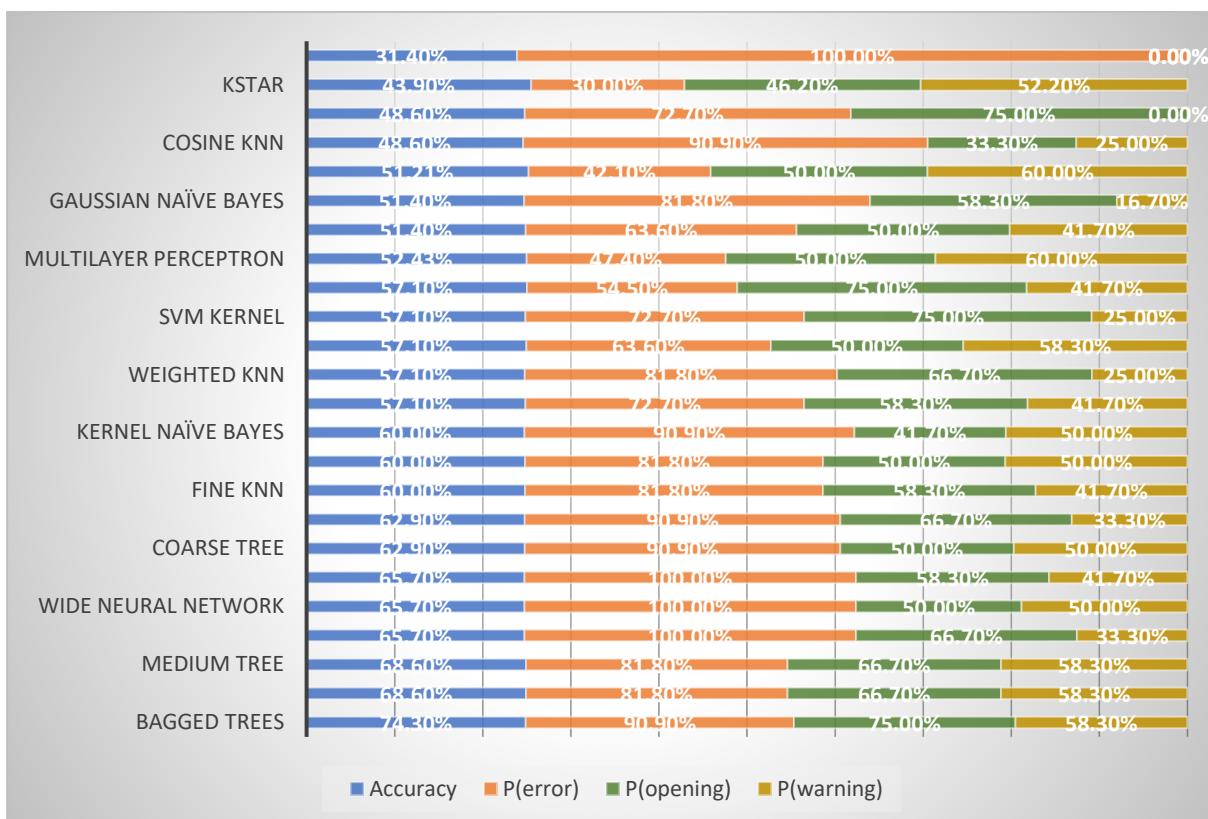
Ακολουθούν τα ποσοστά αποδόσεων σε μορφή ραβδογραμμάτων.



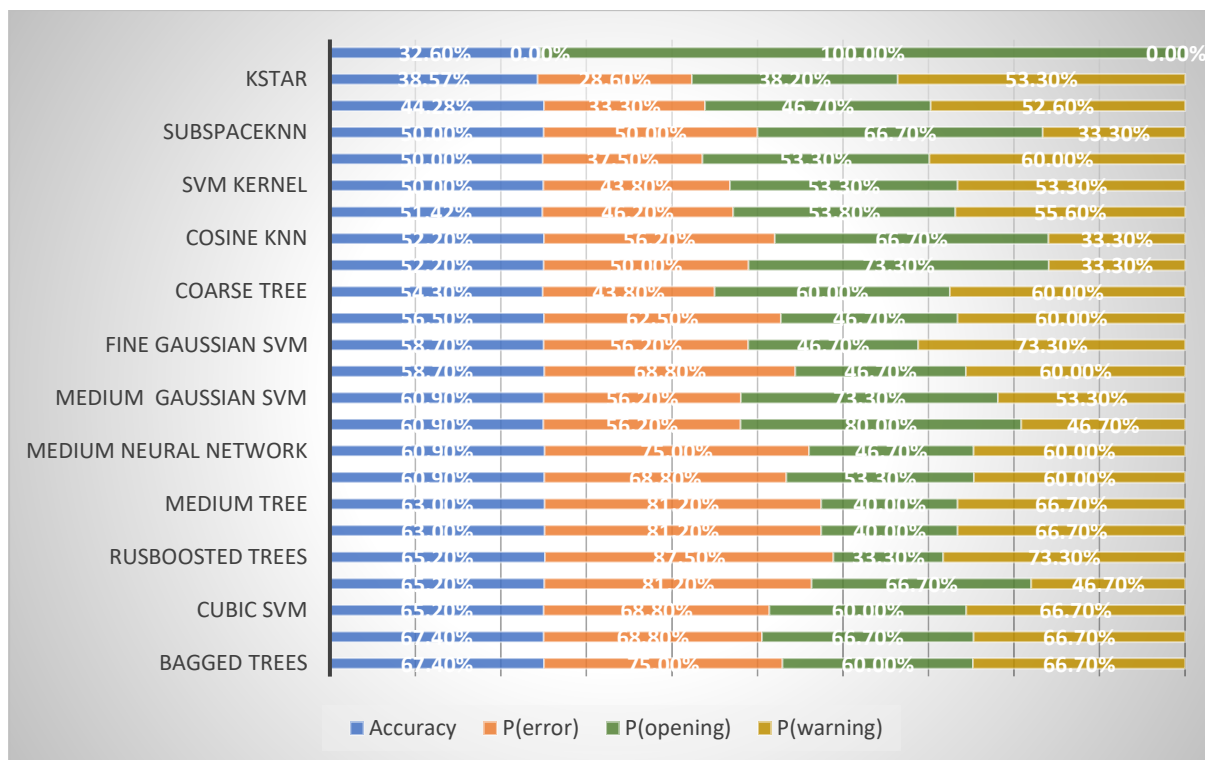
Εικόνα 7.10 : Παράθυρο 3 δευτερολέπτων, 5 Cross Validation



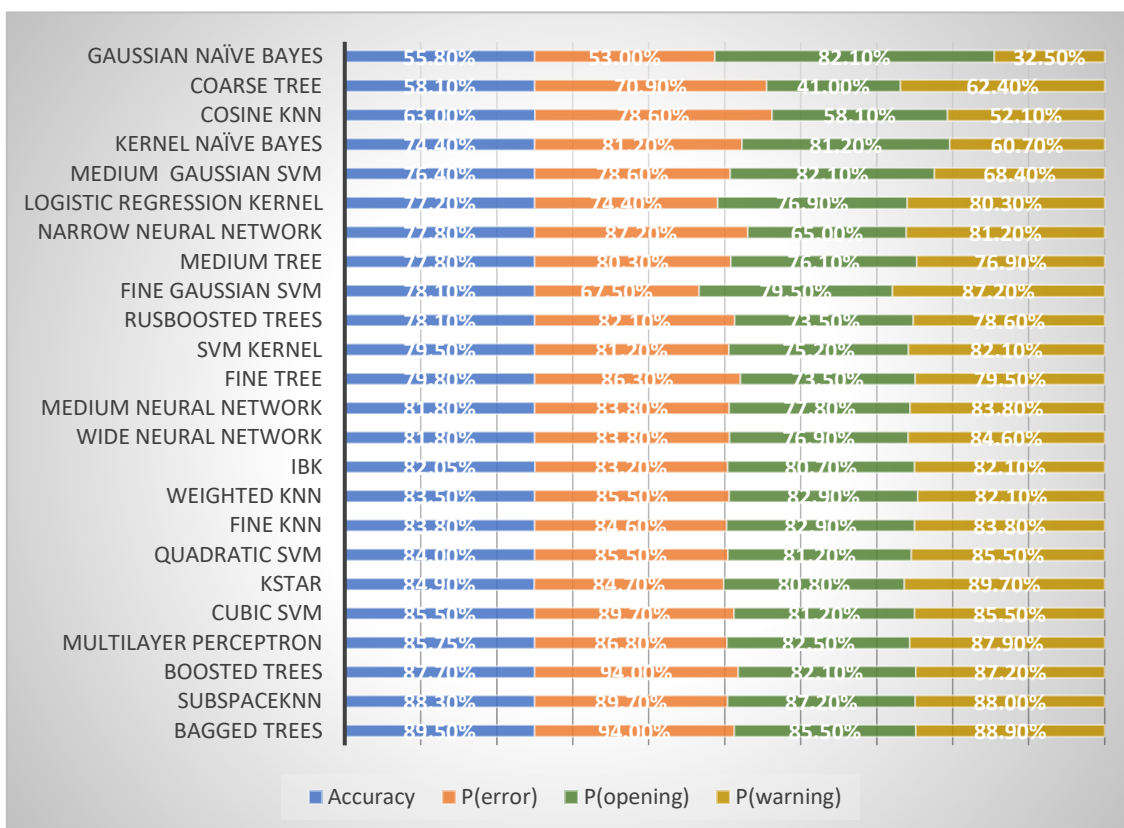
Εικόνα 7.11 : Παράθυρο 3 δευτερολέπτων, 10 Cross Validation



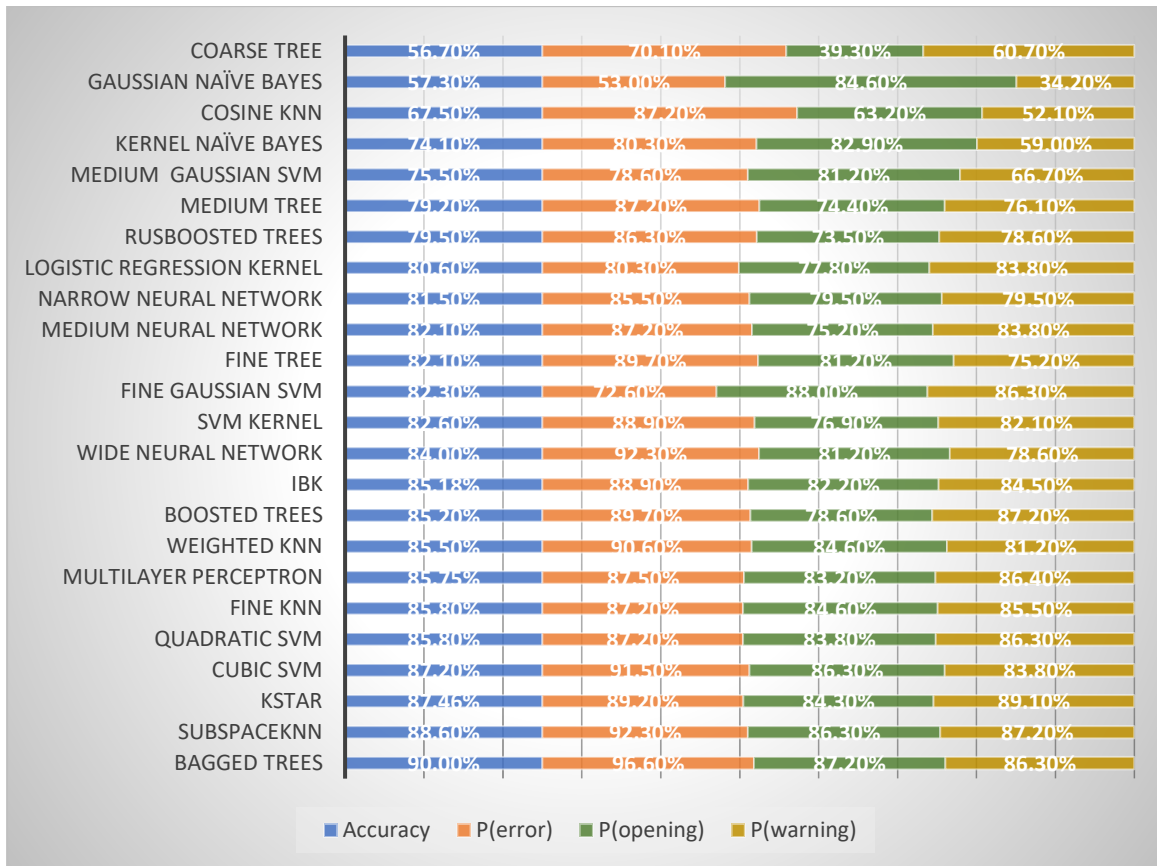
Εικόνα 7.12 : Παράθυρο 3 δευτερολέπτων, 30 Holdout



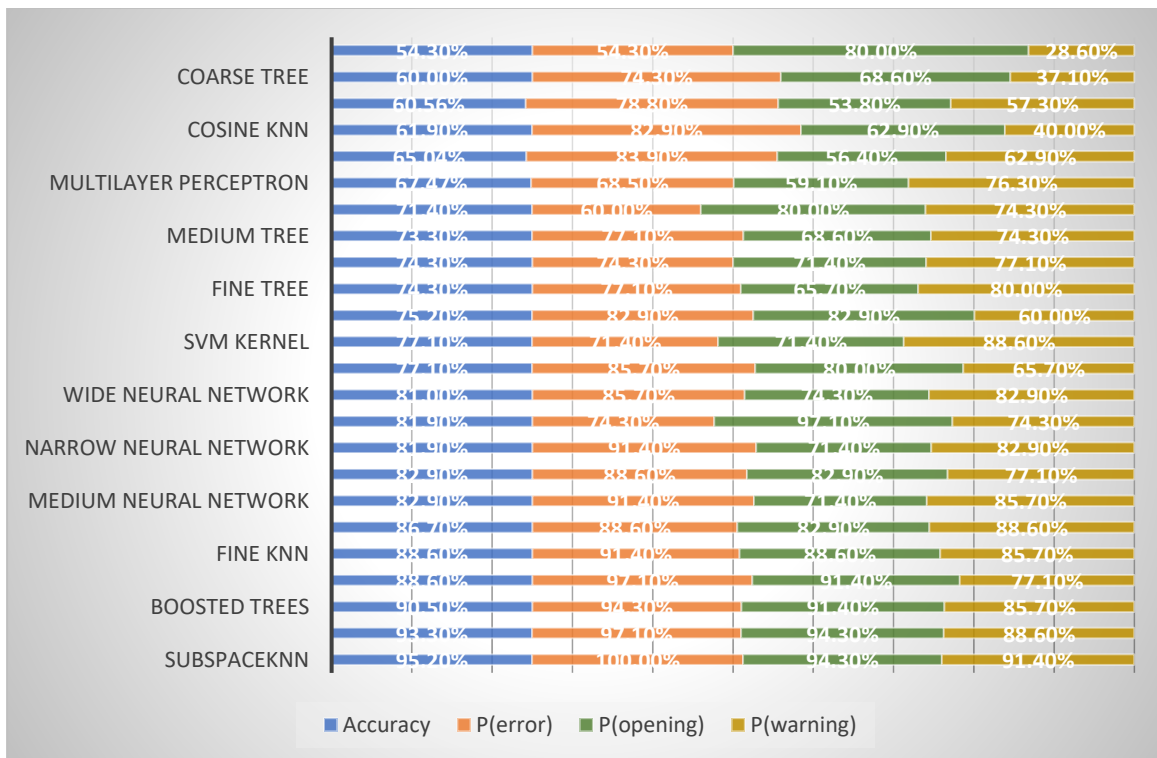
Εικόνα 7.13 : Παράθυρο 3 δευτερολέπτων, 40 Holdout



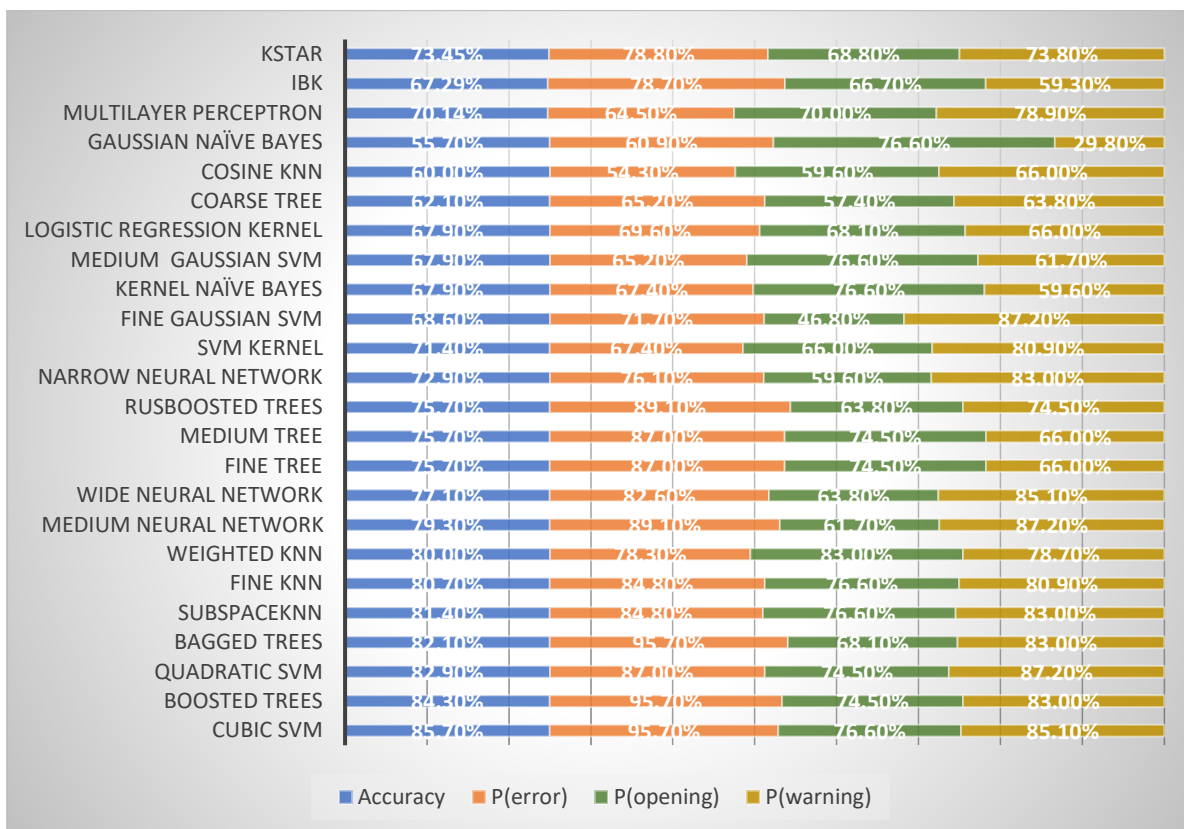
Εικόνα 7.14 : Παράθυρο 1 δευτερολέπτου, 5 Cross Validation



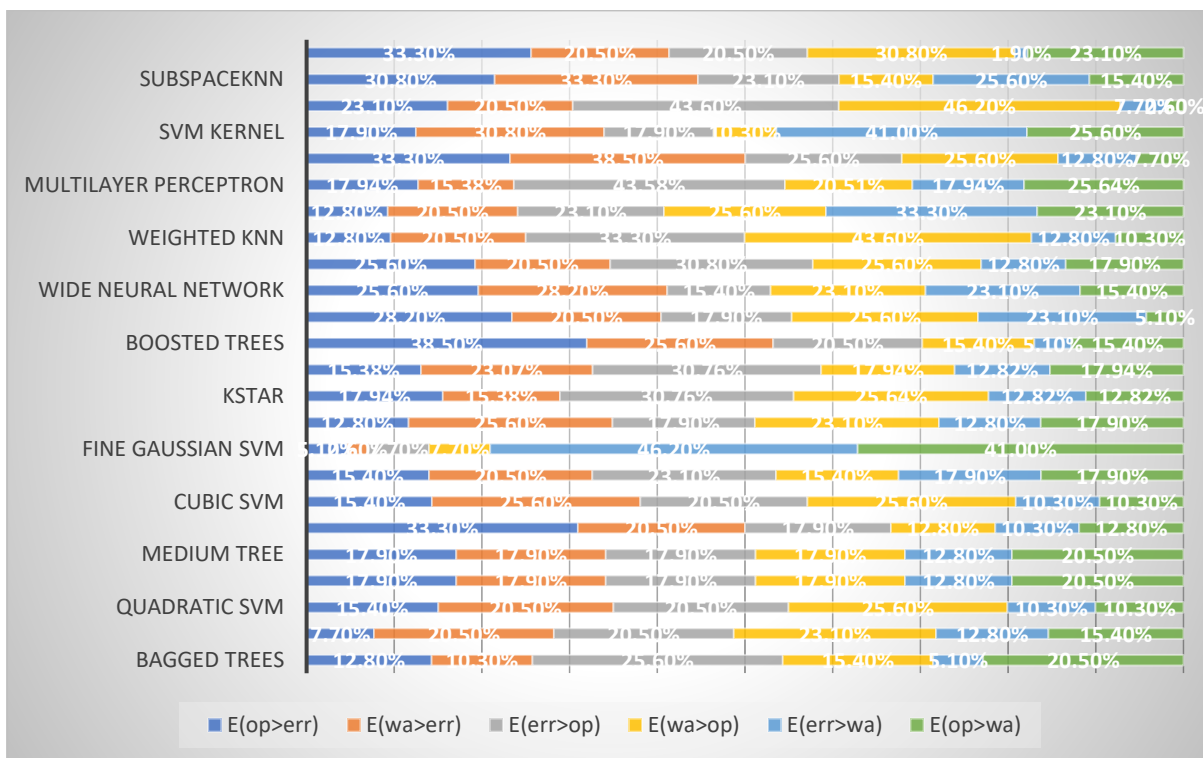
Εικόνα 7.15 : Παράθυρο 1 δευτερολέπτου, 10 Cross Validation



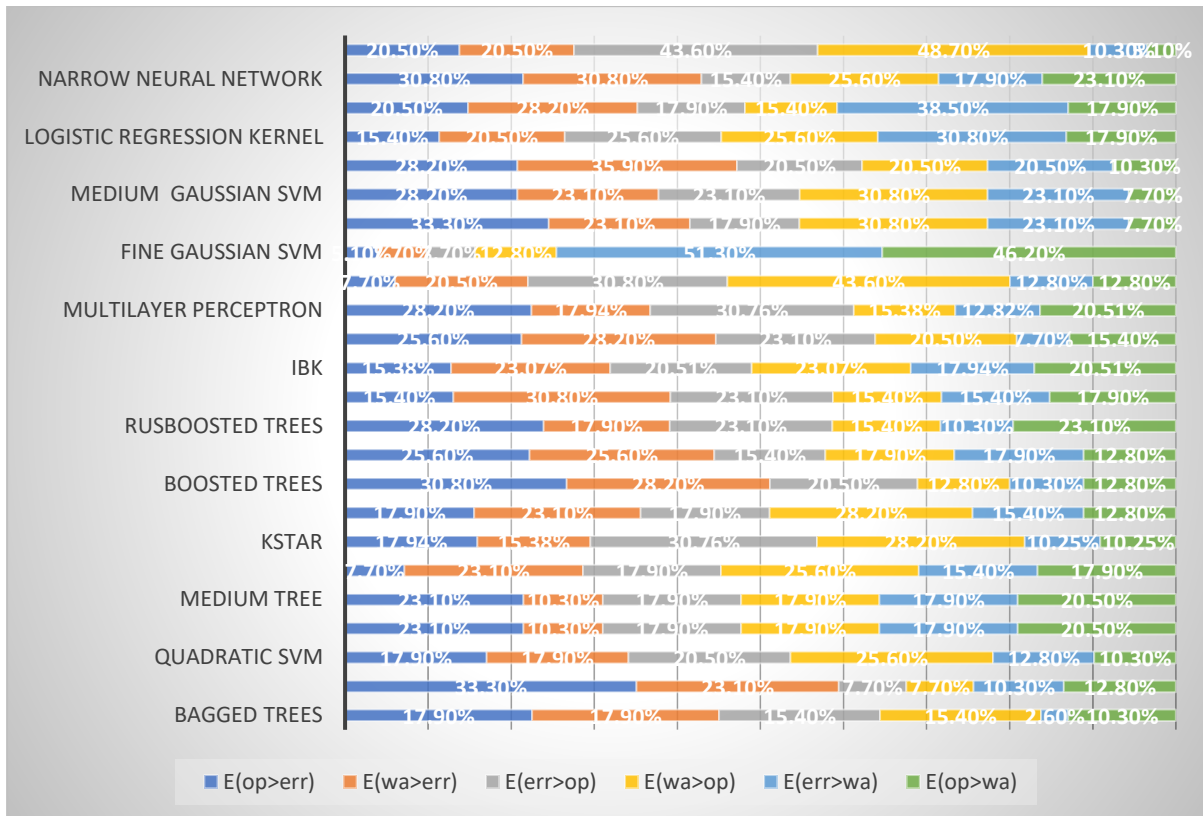
Εικόνα 7.16 : Παράθυρο 1 δευτερολέπτου, 30 Holdout



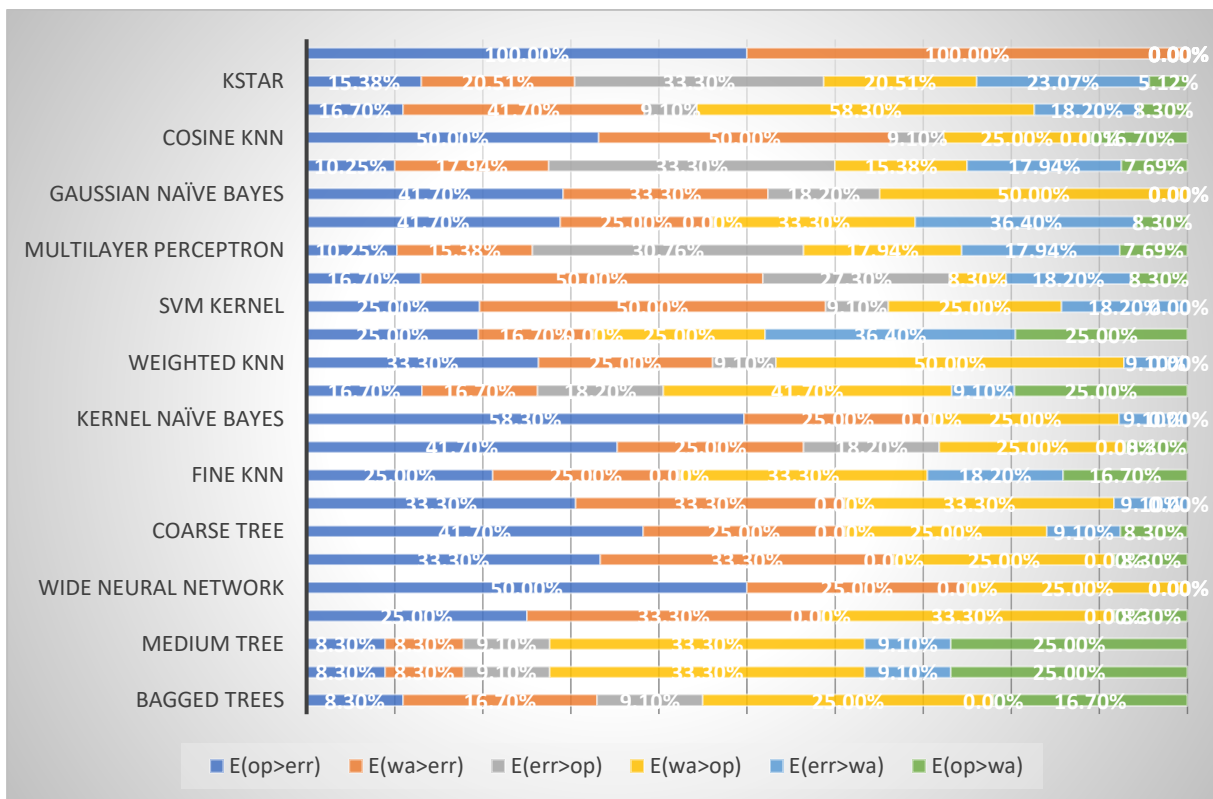
Εικόνα 7.17 : Παράθυρο 1 δευτερολέπτον, 40 Holdout



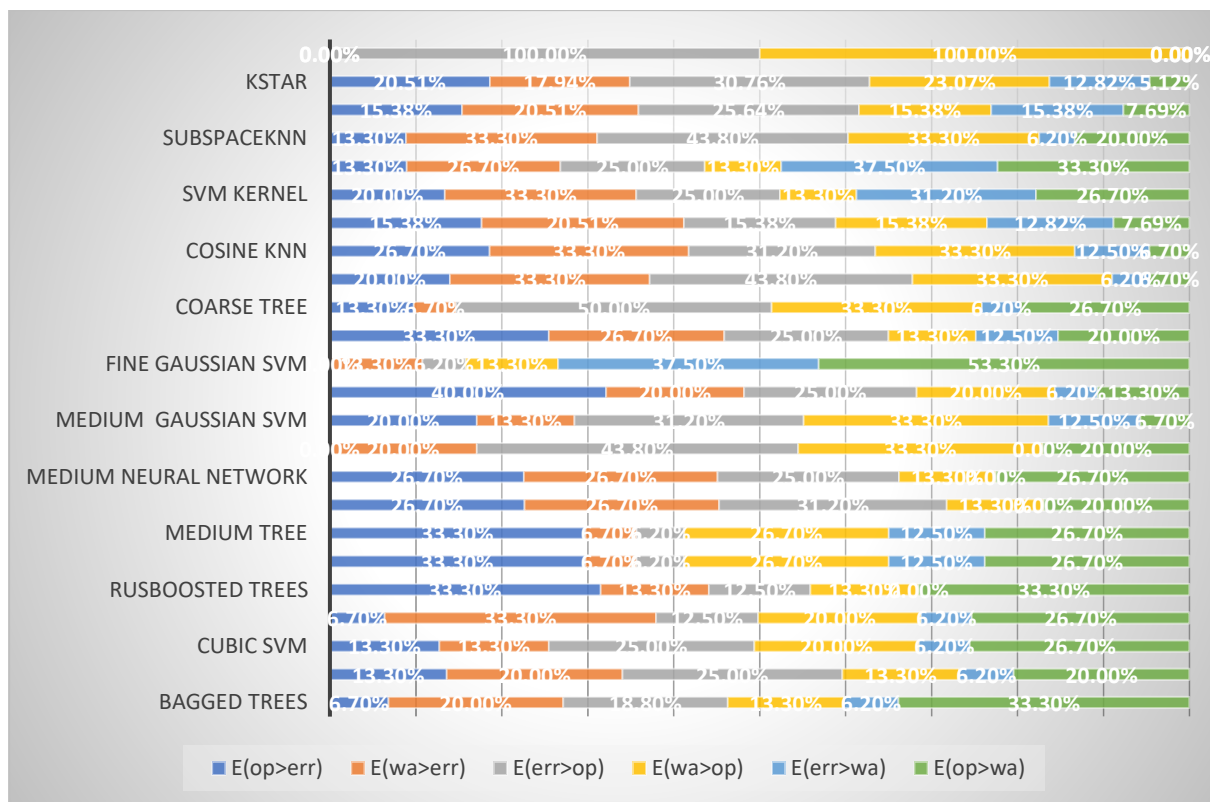
Εικόνα 7.18 : Ποσοστό σφαλμάτων σε παράθυρο 3 δευτερολέπτον, 5 Cross Validation



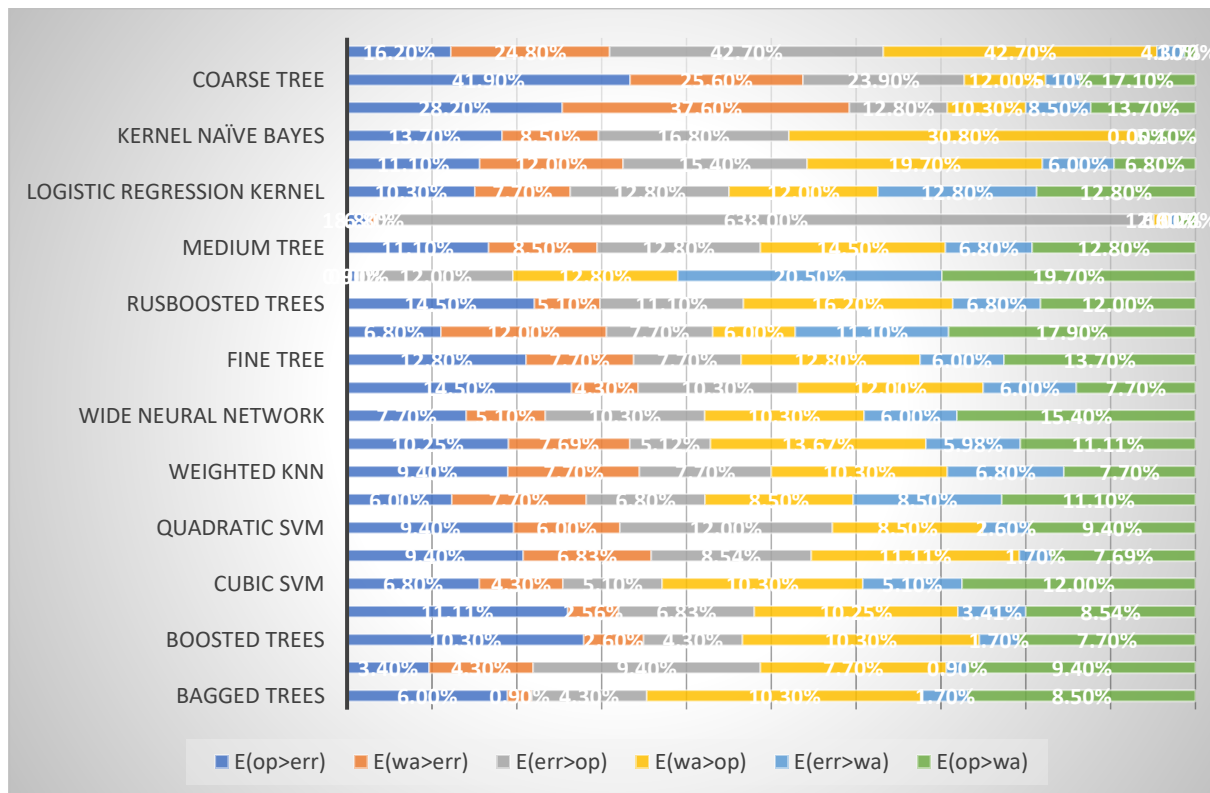
Εικόνα 7.19 : Ποσοστό σφαλμάτων σε παράθυρο 3 δευτερολέπτων, 10 Cross Validation



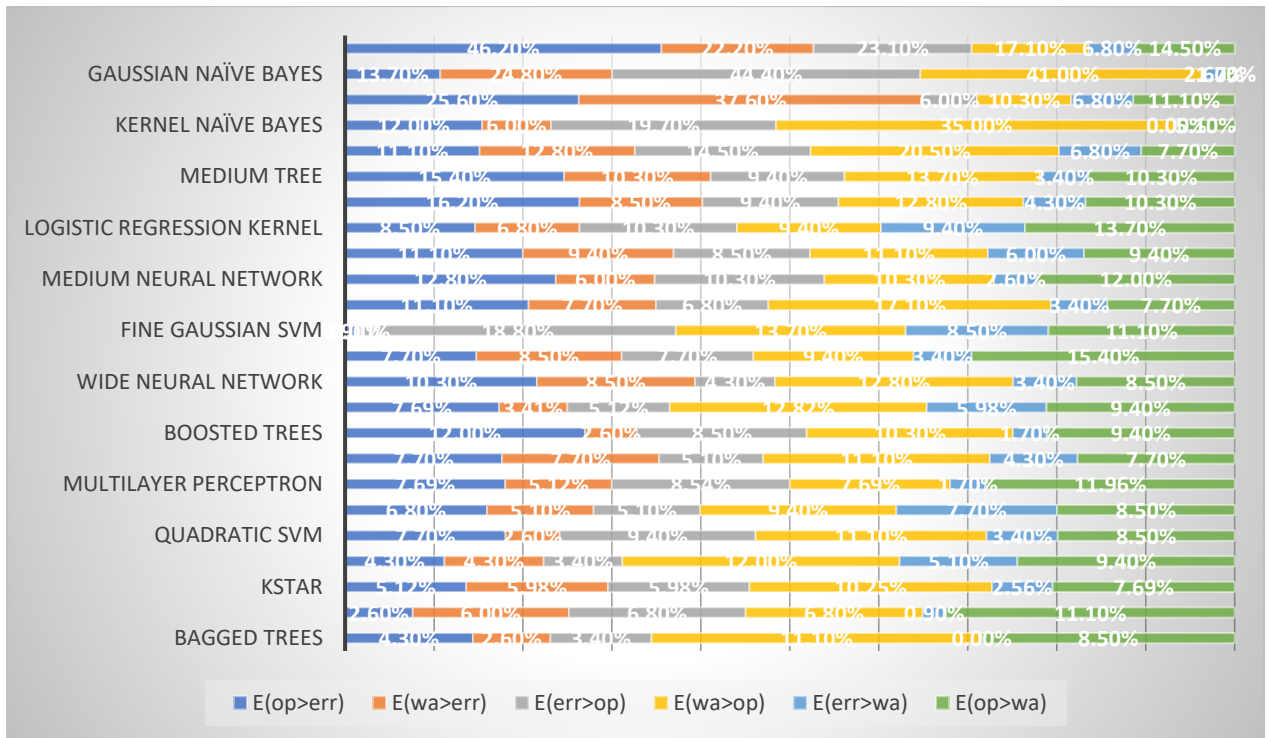
Εικόνα 7.20 : Ποσοστό σφαλμάτων σε παράθυρο 3 δευτερολέπτων, 30 Holdout



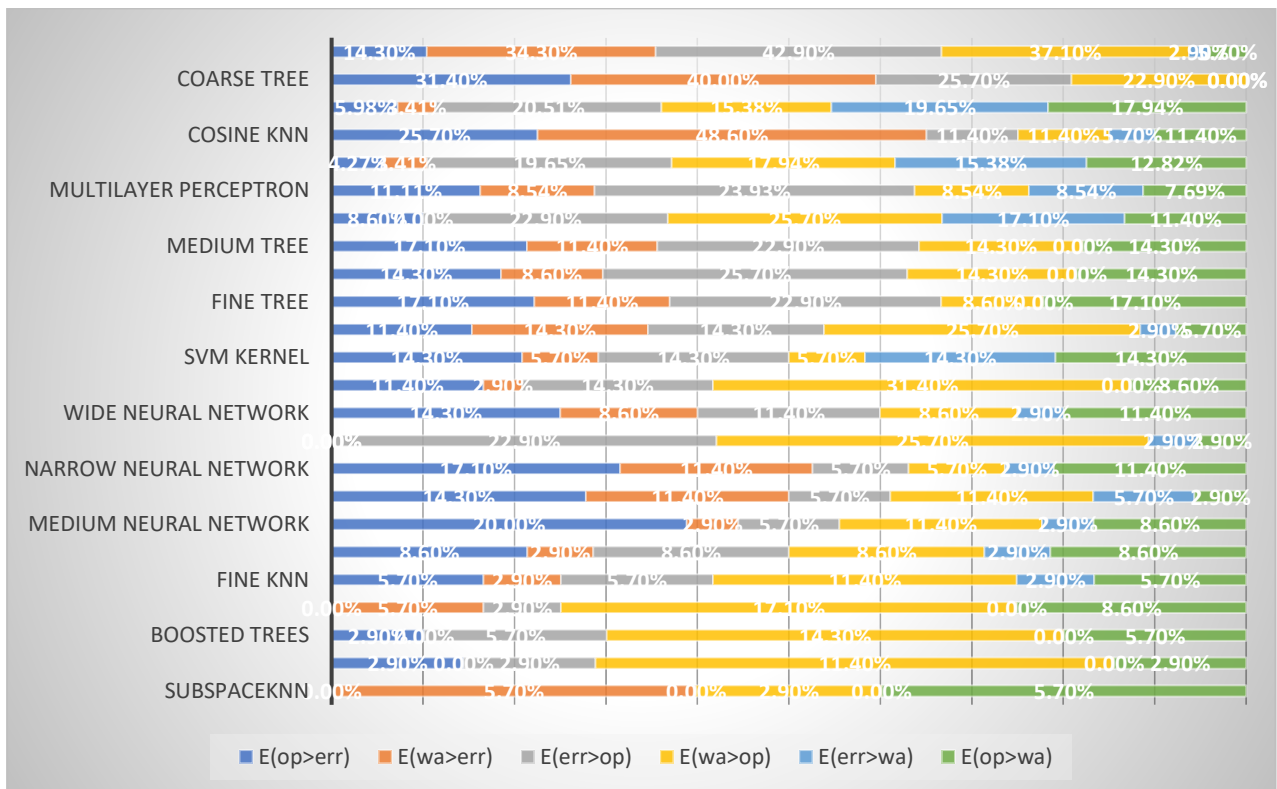
Εικόνα 7.21 : Ποσοστό σφαλμάτων σε παράθυρο 3 δευτερολέπτων, 40 Holdout



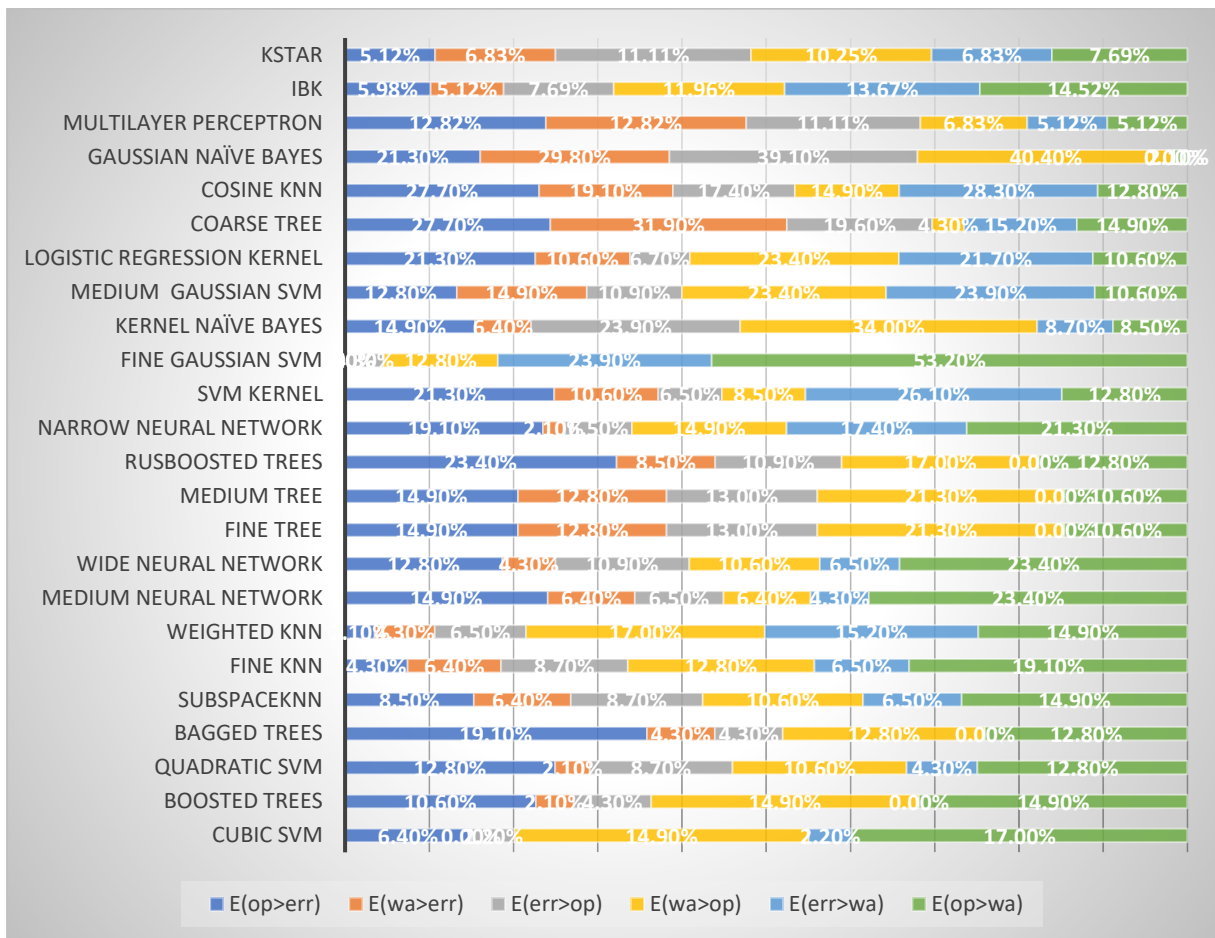
Εικόνα 7.22 : Ποσοστό σφαλμάτων σε παράθυρο 1 δευτερολέπτου, 5 Cross Validation



Εικόνα 7.23 : Ποσοστό σφαλμάτων σε παράθυρο 1 δευτερολέπτου, 10 Cross Validation



Εικόνα 7.24 : Ποσοστό σφαλμάτων σε παράθυρο 1 δευτερολέπτου, 30 Holdout



Εικόνα 7.25 :Ποσοστό σφαλμάτων σε παράθυρο 1 δευτερολέπτου, 40 Holdout

## 7.5 Σχολιασμός αποδόσεων αλγορίθμων και σφαλμάτων τους

Σε γενικές γραμμές αυτό που παρατηρείται είναι ότι οι αποδόσεις των αλγορίθμων σε όλα τα παράθυρα ανάλυσης και σε όλους τους τρόπους επικύρωσης κυμαίνονται στο 60%-80% κατά μέσο όρο ενώ παρατηρούνται και τιμές χαμηλότερες αλλά και υψηλότερες από αυτόν, ποσοστά τα οποία είναι αρκετά ικανοποιητικά και ρεαλιστικά δεδομένου ότι για την ταξινόμηση των ήχων χρησιμοποιήθηκαν και τα συναισθήματα.

Η μεγαλύτερη συνολική απόδοση επιτεύχθηκε με τον αλγόριθμο Subspace KNN σε παράθυρο 1 δευτερολέπτου και μέθοδο επικύρωσης 30% Holdout Validation με ποσοστό απόδοσης 95,20%.

Σχετικά με τις μεγαλύτερες μερικές αποδόσεις, δηλαδή την καλύτερη απόδοση για κάθε είδος ήχου, βρέθηκε για το Error 100,00% με τον αλγόριθμο Subspace KNN σε παράθυρο 1 δευτερολέπτου με μέθοδο επικύρωσης 30% Holdout Validation. Για την κατηγορία των ήχων Opening 97,10% με τον Fine Gaussian SVM σε παράθυρο 1 δευτερολέπτου και μέθοδο επικύρωσης 30% Holdout Validation ενώ για την κατηγορία των ήχων warning 91,40% με τον αλγόριθμο Subspace KNN σε παράθυρο 1 δευτερολέπτου με μέθοδο επικύρωσης 30% Holdout Validation. Στην κατηγοριοποίηση των ήχων Opening παρατηρήθηκε μερική απόδοση 100,00% με τον αλγόριθμο Boosted Trees σε παράθυρο 3 δευτερολέπτων με τις μεθόδους επικύρωσης 30% και 40% Holdout Validation παρόλα αυτά είναι ένα αποτέλεσμα το οποίο δεν αποτυπώνεται σαν επιτυχία του συγκεκριμένου αλγορίθμου καθώς τα ταξινόμησε όλα στην ίδια κλάση.

Όσον αφορά την γενική σύγκριση των παραθύρων ανάλυσης παρατηρούνται οι καλύτερες συνολικές αποδόσεις να υπάρχουν στο παράθυρο 1 δευτερολέπτου με μεγάλη διαφορά από το παράθυρο των 3 δευτερολέπτων που είχε αρκετά χαμηλότερα αποτελέσματα. Αυτό συμβαίνει καθώς στα 3 sec το παράθυρο είναι αρκετά μεγάλο ώστε να αναπαρασταθεί με μία μόνο τιμή με αποτέλεσμα να χάνονται οι διακυμάνσεις ή οι μεταβολές του ήχου.

Για τις διαφορετικές μεθόδους επικύρωσης σε συγκεκριμένο παράθυρο γενικά δεν παρατηρήθηκαν ακραίες αποκλίσεις μεταξύ των μεθόδων επικύρωσης. Ο 10 Cross Fold Validation είχε περίπου 3% καλύτερη απόδοση από τον 5 Cross Fold Validation ενώ μεγαλύτερο ήταν το ποσοστό μεταξύ των Holdout Validation καθώς το 30% Holdout Validation είχε περίπου 10% καλύτερη απόδοση σε σχέση με αυτό τον 40% Holdout Validation. Τέλος σε ότι αφορά την σύγκριση μεταξύ των δύο μεθόδων επικύρωσης σε κάθε παράθυρο, οι Holdout Validation είχαν καλύτερα αποτελέσματα (τάξεως 5%) από τους Cross Fold Validation.

Σε γενικές γραμμές καλύτερες αποδόσεις και ομογένεια σε όλους τους αλγορίθμους και τα παράθυρα παρατηρήθηκε να έχει το error μετά το opening και εν συνεχεία το warning. Πιθανότατα ο λόγος που μας οδήγησε σε αυτές τις αποδόσεις είναι καθώς στις κατηγορίες error και opening είναι πιο ξεκάθαροι οι ήχοι σε αντίθεση με τους ήχους warning στους οποίους το άκουσμα τους μπορεί να μπερδεύει.

Όσον αφορά τα σφάλματα κατηγοριοποίησης τα αποτελέσματα είναι αρκετά ικανοποιητικά με μέσο όρο περίπου στο 20%, πλην κάποιων υπερβολικών αποκλίσεων.

Τα μεγαλύτερα ποσοστά σφαλμάτων σε όλα τα παράθυρα και τρόπους επικύρωσης παρουσιάστηκαν στα δείγματα τα οποία κατηγοριοποιήθηκαν σαν opening ενώ η πραγματική κλάση ήταν error.

Αντίθετα, τα μικρότερα ποσοστά σφαλμάτων παρουσιάστηκαν στα δείγματα που λανθασμένα κατηγοριοποιήθηκαν σαν warning ενώ η πραγματική τους κλάση ήταν error.

Ο αλγόριθμος ο οποίος παρουσίαζε τα πιο ασταθή ποσοστά σφαλμάτων στα περισσότερα πειράματα ήταν ο Narrow Neural Network με 63,80% σφάλμα κατηγοριοποίησης σε opening ενώ η πραγματική κλάση ήταν error και σφάλμα κατηγοριοποίησης 6% σε warning ενώ η πραγματική τους κλάση ήταν error. Αξίζει να αναφερθεί και ο αλγόριθμος Boosted Trees ο οποίος είχε 100% σφάλμα κατηγοριοποίησης opening ενώ η πραγματική του κλάση ήταν warning και αντίστοιχα 100% σφάλμα κατηγοριοποίησης opening ενώ η πραγματική του κλάση ήταν error. Ωστόσο δεν θεωρείται αξιόπιστος λόγω του ότι όλους τους ήχους τους ταξινόμησε σαν opening και όπως είναι λογικό παρουσίασε εξαιρετικά μικρό ποσοστό απόδοσης.

Τέλος όσον αφορά τα σφάλματα σε συνάρτηση με τα παράθυρα ανάλυσης αλλά και τους τρόπους επικύρωσης, αναλογικά με τα ποσοστά επιτυχίας ήταν αυτό του ενός δευτερολέπτου καθώς είχε αρκετά 0% και όσα δεν ήταν μηδενικά είχαν χαμηλότερα ποσοστά σφαλμάτων σε σύγκριση με αυτά των τριών δευτερολέπτων ενώ αντίστοιχα ο καλύτερος τρόπος επικύρωσης ήταν ο 30% Holdout Validation.

## Κεφάλαιο 8<sup>ο</sup>: Αξιολόγηση Χαρακτηριστικών

### 8.1 Εισαγωγή

Στο παρόν κεφάλαιο γίνεται η αξιολόγηση των ηχητικών χαρακτηριστικών που χρησιμοποιήθηκαν προκειμένου να ταξινομηθούν τα δεδομένα σχετικά με την σημαντικότητα τους ως προς την συμβολή τους στο τελικό αποτέλεσμα. Προσωπικά επιλέξαμε η αξιολόγηση των χαρακτηριστικών να γίνει με δύο τρόπους. Ο πρώτος τρόπος ήταν με χρήση του λογισμικού WEKA που χρησιμοποιήθηκε και παραπάνω ενώ ο δεύτερος ήταν μέσα από διαγράμματα με χρήση του Excel τα οποία αναλύθηκαν παραπάνω.

### 8.2 Αξιολόγηση χαρακτηριστικών με χρήση WEKA

Έπειτα από την ολοκλήρωση των πειραμάτων μηχανικής μάθησης κρίθηκε απαραίτητο να γίνει εξέταση της αποδοτικότητας των χαρακτηριστικών καθώς είναι κοινώς αποδεκτό το γεγονός ότι δεν γίνεται όλα τα χαρακτηριστικά να έχουν την ίδια συνεισφορά. Προκειμένου να γίνει η παραπάνω αξιολόγηση που προαναφέρθηκε, χρησιμοποιήθηκε το λογισμικό WEKA. Κατά την διαδικασία κατάταξης έγινε χρήση δύο αλγορίθμων, του InfoGain και του OneR. Ο InfoGain μετράει πως κάθε χαρακτηριστικό συνεισφέρει στην μείωση της εντροπίας. Με αυτόν τον τρόπο ταξινομεί με φθίνουσα σειρά τα χαρακτηριστικά που έχουν το μεγαλύτερο κέρδος πληροφορίας. Αντίστοιχα, ο OneR δομεί το μονοεπίπεδο δέντρο αποφάσεως για κάθε ένα από τα χαρακτηριστικά εισόδου και ταξινομεί και πάλι με φθίνουσα σειρά τα χαρακτηριστικά τα οποία έχουν και τα μεγαλύτερα ποσοστά ορθής ταξινόμησης.

### 8.3 Αποτελέσματα αξιολόγησης χαρακτηριστικών

Παρακάτω παρουσιάζονται τα αποτελέσματα αξιολόγησης όλων των χαρακτηριστικών για κάθε ένα από τα παράθυρα ανάλυσης.

| #  | Παράθυρο 3 δευτερολέπτων |            | Παράθυρο 1 δευτερολέπτου |            |
|----|--------------------------|------------|--------------------------|------------|
|    | InfoGain                 | OneR       | InfoGain                 | OneR       |
| 1  | Mfcc5                    | Mfcc5      | Rolloff0.9               | Rolloff0.9 |
| 2  | Mfcc2                    | Skewness   | Mfcc8                    | Mfcc8      |
| 3  | Mfcc4                    | Rolloff0.9 | Mfcc7                    | Mfcc7      |
| 4  | Skewness                 | Mfcc2      | Mfcc5                    | LowEnergy  |
| 5  | Kurtosis                 | Mfcc8      | Mfcc6                    | Rolloff0.5 |
| 6  | Mfcc1                    | Mfcc12     | Rolloff0.7               | ZeroCross  |
| 7  | Mfcc6                    | Kurtosis   | Mfcc12                   | Rolloff0.7 |
| 8  | Mfcc3                    | Mfcc7      | Mfcc10                   | Mfcc5      |
| 9  | Mfcc12                   | Anger      | Mfcc2                    | Mfcc9      |
| 10 | Mfcc13                   | Centroid   | Mfcc9                    | Mfcc2      |
| 11 | Mfcc9                    | Mfcc1      | Rolloff0.3               | Br500      |
| 12 | ZeroCross                | Mfcc6      | Spread                   | Mfcc13     |
| 13 | Mfcc8                    | Sad        | Mfcc4                    | Rolloff0.3 |
| 14 | Mfcc11                   | Mfcc11     | Kurtosis                 | Mfcc12     |
| 15 | Mfcc7                    | Spread     | Mfcc13                   | Centroid   |
| 16 | Rolloff0.3               | Rolloff0.7 | Skewness                 | Mfcc3      |
| 17 | Rolloff0.5               | Mfcc10     | Mfcc11                   | Mfcc6      |
| 18 | Spread                   | Br500      | LowEnergy                | Kurtosis   |
| 19 | Rolloff0.9               | Rolloff0.5 | Mfcc3                    | Skewness   |
| 20 | Rolloff0.5               | Fear       | Centroid                 | Mfcc10     |
| 21 | Centroid                 | Happy      | Mfcc1                    | Spread     |
| 22 | Fear                     | ZeroCross  | Br500                    | Fear       |
| 23 | Anger                    | Tender     | Sad                      | Mfcc4      |
| 24 | Mfcc10                   | Mfcc4      | Tender                   | Mfcc1      |
| 25 | Tender                   | LowEnergy  | ZeroCross                | Mfcc11     |
| 26 | Sad                      | Mfcc13     | Happy                    | Br1000     |
| 27 | Happy                    | Br1000     | Fear                     | Happy      |
| 28 | Br1000                   | Rolloff0.3 | Anger                    | Anger      |
| 29 | Br500                    | Mfcc9      | Rolloff0.5               | Tender     |
| 30 | LowEnergy                | Mfcc3      | Br1000                   | Sad        |

Εικόνα 8.1: Κατάταξη αξιολόγησης χαρακτηριστικών σε όλα τα παράθυρα

## 8.4 Συμπεράσματα

Σε γενικές γραμμές παρατηρείται διαφοροποίηση μεταξύ των παραθύρων ανάλυσης αλλά παρουσιάζεται σχετικά ομοιόμορφη κατανομή των αποτελεσμάτων μεταξύ των αλγορίθμων επικύρωσης στο ίδιο παράθυρο ανάλυσης.

Πιο συγκεκριμένα στο παράθυρο ανάλυσης 3 δευτερολέπτων και οι δύο αλγόριθμοι επικύρωσης αξιολογούν το Mfcc για 5 συντελεστές ως το καλύτερο χαρακτηριστικό. Στην δεύτερη και τρίτη θέση για τον αλγόριθμο InfoGain βρίσκεται και πάλι το χαρακτηριστικό Mfcc για 2 και για 4 συντελεστές αντίστοιχα ενώ για τον αλγόριθμο OneR βρίσκονται τα Skewness και της ενεργειακής συχνότητας αποκοπής (Spectral roll-off) για το 90% της ενέργειας. Στη μέση περίπτωση της συνολικής κατάταξης για τον InfoGain βρίσκεται και πάλι το χαρακτηριστικό Mfcc για 11 και 7 συντελεστές καθώς και η ενεργειακή συχνότητα αποκοπής (Spectral roll-off) για 30% και 50% της ενέργειας. Στον αλγόριθμο επικύρωσης OneR βρίσκεται και πάλι το χαρακτηριστικό Mfcc για 11 συντελεστές στην ίδια θέση κατάταξης με αυτή του InfoGain όπως επίσης το Spread, το Spectral roll-off για 70% της ενέργειας και το Mfcc για 10 συντελεστές. Τέλος την χειρότερη κατάταξη κατείχαν τα χαρακτηριστικά Brightness και LowEnergy για τον InfoGain και για τον OneR το χαρακτηριστικό Mfcc για 9 και 3 συντελεστές όπως επίσης και το Spectral roll-off για 30% της ενέργειας.

Όσον αφορά το παράθυρο ανάλυσης 1 δευτερολέπτου τις πρώτες τρεις θέσεις στην κατάταξη του πίνακα κατέχουν η ενεργειακή συχνότητα αποκοπής (Spectral roll-off) για το 90% της ενέργειας όπως επίσης το χαρακτηριστικό Mfcc για 8 και 7 συντελεστές τα οποία βρίσκονται στην ίδια θέση και για τους δύο αλγόριθμους επικύρωσης. Στη μέση της κατάταξης για τον InfoGain βρίσκονται τα φασματικά χαρακτηριστικά Kurtosis, Skewness όπως επίσης και το Mfcc για 13 και 11 συντελεστές και στον OneR το Mfcc για 12, 3 και 6 συντελεστές όπως επίσης και το centroid. Αντίστοιχα στην βάση της κατάταξης παρουσιάζονται τα χαρακτηριστικά συναισθήματος Fear, Anger καθώς και το Brightness 1000 και το Spectral roll-off για το 50% της ενέργειας ενώ για τον OneR τα χαρακτηριστικά συναισθήματος Happy, Anger, Tender και Sad.

Τέλος αξίζει να σημειωθεί πως τα Emotion παρουσιάζουν χαμηλή θέση σε όλα τα παράθυρα ανάλυσης καθώς και αλγορίθμων επικύρωσης εκτός από την περίπτωση των 3 δευτερολέπτων με χρήση της μεθόδου επικύρωσης OneR στην οποία βρίσκονται διάσπαρτα στον πίνακα κατάταξης ωστόσο το συναίσθημα που διαφοροποιεί αυτούς τους ήχους κατά κύριο λόγο είναι το Anger και το Fear.

## Κεφάλαιο 9<sup>ο</sup>: Συμπεράσματα και μελλοντικές κατευθύνσεις

### 9.1 Συμπεράσματα

Μετά την συλλογή με χρήση αυστηρών κριτηρίων 39 ήχων, για κάθε είδος ήχου για κάθε ένα από τα τρία είδη βρέθηκαν έπειτα από εκτενή μελέτη της διαθέσιμης βιβλιογραφίας ποια χαρακτηριστικά θα πρέπει να εξαχθούν προκειμένου να έχουμε αξιόπιστα αποτελέσματα. Εν συνεχεία, εκπαιδεύτηκαν διάφορα μοντέλα Εποπτευόμενης Μηχανικής Μάθησης σε παράθυρα 3 και 1 δευτερολέπτου με τους τρόπους επικύρωσης 5,10 Cross Validation και 30% , 40% Holdout Validation. Όσον αφορά, το τεχνικό κομμάτι τα αποτελέσματα των παραπάνω πειραμάτων ήταν αρκετά ικανοποιητικά καθώς ήταν παρόμοια με παρεμφερή πειράματα πάνω στο αντικείμενο που μελετάται καθώς και λόγω του ότι μία συνισταμένη αποτελούσε και το συναίσθημα το οποίο από μόνο του είναι δύσκολο να κατηγοριοποιηθεί. Καλύτερα ποσοστά κατηγοριοποίησης είχε το error το οποίο κυμαινόταν κατά μέσο όρο στο 65% για παράθυρο 3 δευτερολέπτων και 85% για παράθυρο 1 δευτερολέπτου. Στην συνέχεια ακολουθεί το opening το οποίο κυμαινόταν κατά μέσο όρο στο 60% για παράθυρο 3 δευτερολέπτων και 80% για παράθυρο 1 δευτερολέπτου και τέλος το warning το οποίο κατά μέσο όρο βρισκόταν στο 58% για παράθυρο 3 δευτερολέπτων και περίπου στο 75% για παράθυρο 1 δευτερολέπτου. Τα σφάλματα κατηγοριοποίησης κυμάνθηκαν στις περισσότερες περιπτώσεις στο 20 % κατά μέσο όρο. Τα δέντρα απόφασης και οι αλγόριθμοι KNN φάνηκαν να έχουν τις καλύτερες αποδόσεις και στην συνέχεια οι SVM και τα νευρωνικά δίκτυα. Όσον αφορά την γενική σύγκριση των παραθύρων ανάλυσης παρατηρείται ότι λειτουργεί καλύτερα η παραθυροποίηση για το 1 sec σε αντίθεση με αυτό των 3 sec καθώς στα 3 sec το παράθυρο είναι αρκετά μεγάλο ώστε να αναπαρασταθεί με μία μόνο τιμή με αποτέλεσμα να χάνονται οι διακυμάνσεις ή οι μεταβολές. Αναφορικά με τους τρόπους επικύρωσης παρατηρούνται οι καλύτερες συνολικές αποδόσεις να υπάρχουν για 30% Holdout Validation. Στην περίπτωση ωστόσο που εξεταστεί μεμονωμένα το κάθε ηχητικό απόσπασμα τότε θα γινόταν χρήση του αλγόριθμου Subspace KNN σε παράθυρο 1 δευτερολέπτου με μέθοδο επικύρωσης 30% Holdout Validation για την κατηγορία Error. Στην κατηγορία Opening τον Fine Gaussian SVM σε παράθυρο 1 δευτερολέπτου και μέθοδο επικύρωσης 30% Holdout Validation και τέλος στην κατηγορία Warning τον αλγόριθμο Subspace KNN σε παράθυρο 1 δευτερολέπτου με μέθοδο επικύρωσης 30% Holdout Validation. Αντίστοιχα, σε ότι αφορά το συναίσθημα αυτό που παρατηρείται σε γενικές γραμμές με βάση την αξιολόγηση των χαρακτηριστικών με χρήση του WEKA είναι ότι τα συναισθήματα που διαφοροποιούν αυτούς τους ήχους είναι το Anger σε παράθυρο 3 δευτερολέπτων ενώ αντίστοιχα το Fear σε παράθυρο ενός δευτερολέπτου. Σε επίπεδο κλάσης, οι ήχοι που αναπαράγονται κατά τον λάθος χειρισμό μίας ηλεκτρονικής συσκευής τείνουν να είναι πιο Anger. Αντίστοιχα οι ήχοι έναρξης πιο Happy ενώ οι ήχοι προειδοποίησης πιο Fear.

### 9.2 Μελλοντικές Κατευθύνσεις

Είναι κοινώς αποδεκτό ότι ο τομέας της μηχανικής μάθησης και συγκεκριμένα των διάφορων εφαρμογών της επάνω στον ήχο αλλά και το συναίσθημα είναι ένα πεδίο το οποίο είναι σε πρώιμα στάδια.

Έχοντας ως έμπνευση την παρούσα πτυχιακή εργασία στόχος μας είναι να την εξελίξουμε στο μέλλον παραπάνω. Θα ήταν αρκετά ενδιαφέρον να διεξαχθεί ξανά το ίδιο πείραμα με συμμετοχή ανθρώπων και χρήση ερωτηματολογίων προκειμένου να δούμε κατά πόσο το σύστημα και ο άνθρωπος συμφωνούν στο αποτέλεσμα. Ταυτόχρονα, θα μπορούσαν να ενταχθούν περισσότερες κατηγορίες ήχων προκειμένου να γίνει σχετική διερεύνηση της συμπεριφοράς τους.

Επιπλέον, θα μπορούσε να υπάρξει συσχέτιση με το Διαδίκτυο των Πραγμάτων (IoT) με πληθώρα εφαρμογών. Για παράδειγμα, στον τομέα του marketing θα μπορούσε να υπάρχει ηχητική ανάλυση κατά την αναπαραγωγή των διαφημίσεων προκειμένου να ανέβει η αγοραστική δύναμη της εκάστοτε εταιρείας. Επιπροσθέτως, θα μπορούσε να αξιοποιηθεί σε χώρους ψυχαγωγίας. Μέσα από έναν αισθητήρα-μικρόφωνο θα συλλέγονται τα ηχητικά δεδομένα, θα ταξινομούνται με χρήση των κατάλληλων αλγορίθμων και θα γίνεται εξαγωγή μίας αναφοράς σχετικά με τα συναισθήματα που προκαλεί ο ήχος σε συνδυασμό με παρουσίαση εναλλακτικών επιλογών για κάθε κατηγορία. Θα μπορούσε να χρησιμοποιηθεί σε υπηρεσίες ροής μουσικής (music streaming services) όπως το Spotify ώστε να μπορεί μέσα από τις επιλογές του χρήστη να αντιλαμβάνεται την τρέχουσα ψυχολογική του κατάσταση και να συνεχίζει την αναπαραγωγή μουσικής στο ίδιο συναισθηματικό μοτίβο. Τέλος, θα ήταν ιδιαίτερα ενδιαφέρον η συγκεκριμένη έρευνα να αποτελέσει την αφετηρία για την παραγωγή ήχων οι οποίοι θα προκαλούν αντίστοιχα συναισθήματα με την λογική human machine interaction.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] B. Myers *et al.*, “Strategic directions in human-computer interaction,” *ACM Comput. Surv.*, vol. 28, no. 4, pp. 794–809, Dec. 1996, doi: 10.1145/242223.246855.
- [2] D. A. Norman, *The design of everyday things*, Revised and Expanded edition. New York, New York: Basic Books, 2013.
- [3] J. Grudin, “A Moving Target””The Evolution of Human-Computer Interaction,” in *Human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications. (3rd edition).*, Human-Computer interaction handbook: Fundamentals, Evolving technologies, and Emerging applications. (3rd edition), Taylor & Francis Group, 2012. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/moving-target-evolution-human-computer-interaction/>
- [4] Bush, Vannevar, “As We May Think,” *Atl. Mon.*, vol. 176, Jul. 1945.
- [5] J. C. R. Licklider, “Man-Computer Symbiosis,” *IRE Trans. Hum. Factors Electron.*, vol. HFE-1, no. 1, pp. 4–11, Mar. 1960, doi: 10.1109/THFE2.1960.4503259.
- [6] Sutherland, Ivan., “Sketchpad: A Man-machine Graphical Communication System,” Republished by: University of Cambridge, 2003, 1963. [Online]. Available: <http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-574.pdf>
- [7] Sutherland, Ivan., “The Ultimate Display,” 1965, [Online]. Available: <http://papers.cumincad.org/data/works/att/c58e.content.pdf>
- [8] L. Manovich, *Software takes command: extending the language of new media*. New York ; London: Bloomsbury, 2013.
- [9] B. A. Myers, “A brief history of human-computer interaction technology,” *Interactions*, vol. 5, no. 2, pp. 44–54, Mar. 1998, doi: 10.1145/274430.274436.
- [10] Kay, Alan C, “A personal computer for children of all ages.,” *Proc. ACM Annu. Conf.*, vol. 1, 2011.
- [11] T. Winograd, Ed., *Bringing design to software*. New York, N.Y. : Reading, Mass: ACM Press ; Addison- Wesley, 1996.
- [12] S. K. Card, T. P. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*, 1st ed. CRC Press, 2018. doi: 10.1201/9780203736166.
- [13] D. A. Norman, Ed., *User centered system design: new perspectives on human-computer interaction*, 9. [print.]. Hillsdale, NJ: Erlbaum, 1986.
- [14] T. Winograd and F. Flores, *Understanding computers and cognition: a new foundation for design*, 24th printing. Boston: Addison-Wesley, 2008.
- [15] J. M. Carroll, “HUMAN-COMPUTER INTERACTION: Psychology as a Science of Design,” *Annu. Rev. Psychol.*, vol. 48, no. 1, pp. 61–83, Feb. 1997, doi: 10.1146/annurev.psych.48.1.61.
- [16] K. Georgiadis, F. Kalaganis, P. Migkotzidis, E. Chatzilaris, S. Nikolopoulos, and I. Kompatsiaris, “A Computer Vision System Supporting Blind People - The Supermarket Case,” in *Computer Vision Systems*, vol. 11754, D. Tzovaras, D. Giakoumis, M. Vincze, and A. Argyros, Eds. Cham: Springer International Publishing, 2019, pp. 305–315. doi: 10.1007/978-3-030-34995-0\_28.
- [17] Arun Kumar Singh and Asst. Professor, College of Computing and I, “THE ACTIVE IMPACT OF HUMAN COMPUTER INTERACTION (HCI) ON ECONOMIC, CULTURAL AND SOCIAL LIFE,” Sep. 2017.
- [18] R. W. Picard, *Affective computing*, 1. paperback ed. Cambridge, Mass.: MIT Press, 2000.
- [19] M. Hurst, M. Glencross, and T. Jackson, “Affecting HCI: a new approach to affective computing through an emotional saliency model,” in *Proceedings of the ACM Symposium on Applied Perception*, Dublin Ireland, Aug. 2013, pp. 115–115. doi: 10.1145/2492494.2501874.
- [20] P. E. Griffiths, *What Emotions Really Are: The Problem of Psychological Categories*. University of Chicago Press, 1997. doi: 10.7208/chicago/9780226308760.001.0001.
- [21] P. R. Kleinginna and A. M. Kleinginna, “A categorized list of emotion definitions, with suggestions for a consensual definition,” *Motiv. Emot.*, vol. 5, no. 4, pp. 345–379, Dec. 1981, doi: 10.1007/BF00992553.
- [22] M. W. Eysenck, Ed., *The Blackwell dictionary of cognitive psychology*, 2. ed. Oxford: Blackwell Reference, 2003.
- [23] W. B. Cannon, “The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory,” *Am. J. Psychol.*, vol. 39, no. 1/4, p. 106, Dec. 1927, doi: 10.2307/1415404.

- [24] Αλεξάνδρα Δεδικούση, Ελευθερία Γωνίδα, and Γρηγόρης Κιοσέογλου, “Κατανόηση συναισθημάτων και συναισθηματική επίγνωση κατά τη σχολική ηλικία: Η συμβολή της γνωστικής επίδοσης των παιδιών και των αντιλήψεών τους για τη γονεϊκή ζεστασιά ή ψυχρότητα,” *Ι΄ Επιστημονική Επετηρίδα Του Τμήματος Ψυχολογίας ΑΠΘ*, 2013.
- [25] Μπανταβή Παναγιώτα and Ιωάννης Ουζουνσαββίδης, “Ανίχνευση των ανθρώπινων συναισθημάτων με τη βοήθεια συστήματος όρασης και τη μελέτη των εκφράσεων του προσώπου του,” Καβάλα, 2013.
- [26] R. Beale and C. Peter, *Affect and Emotion in Human-Computer Interaction: From Theory to Applications*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg Springer e-books, 2008.
- [27] B. Reeves and C. I. Nass, *The media equation: how people treat computers, television, and new media like real people and places*, 1. paperback ed., [Reprint.]. Stanford, Calif: CSLI Publ, 20.
- [28] Wegner, Daniel M., “‘Ironic processes of mental control.’ Psychological review,” vol. 101, no. 1, pp. 34–52, 1994.
- [29] Bower, G. H., & Forgas, J. P., “Affect, memory, and social cognition,” in *Cognition and emotion*, Oxford University Press, 2000, pp. 87–168.
- [30] Scott Brave and Clifford Nass, “Emotion in human-computer interaction,” in *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*, L. Erlbaum Associates Inc., 2002, pp. 81–96.
- [31] C. E. Rash and U.S. Army Aeromedical Research Laboratory, *Helmet-mounted displays: sensation, perception, and cognition issues*. Fort Rucker, Ala.: U.S. Army Aeromedical Research Laboratory, 2009.
- [32] K. Doshi, “Audio Deep Learning Made Simple (Part 1): State-of-the-Art Techniques,” *Towards Data Science*.
- [33] Κολοκυθάς Κωνσταντίνος, “ΗΧΟΣ,” 2016. [Online]. Available: [https://repository-web.kallipos.gr/bitstream/11419/3494/2/02\\_chapters\\_05.pdf](https://repository-web.kallipos.gr/bitstream/11419/3494/2/02_chapters_05.pdf)
- [34] Δημήτρης Πέτρος and Alexey Bukreyev, “Υλοποίηση σταδίων κωδικοποίησης ηχητικού σήματος πλήρους φάσματος στο προγραμματιστικό περιβάλλον Matlab,” Τεχνολογικό Εκπαιδευτικό Ίδρυμα Δυτικής Ελλάδας, Αντίρριο, 2015.
- [35] Μαρκουτσά Δήμητρα, “«Η εξέλιξη των προτύπων κωδικοποίησης ήχου. Από τον ήχο MPEG-1 στον ήχο MPEG-H»,” Αντίρριο, 2015. [Online]. Available: <http://repository.library.teimes.gr/xmlui/bitstream/handle/123456789/5942/%ce%97%20%ce%b5%ce%be%ce%ad%ce%bb%ce%b9%ce%be%ce%b7%20%cf%84%cf%89%ce%bd%20%cf%80%cf%81%ce%bf%cf%84%cf%8d%cf%80%cf%89%ce%bd%20%ce%ba%cf%89%ce%b4%ce%b9%ce%ba%ce%bf%cf%80%ce%bf%ce%af%ce%b7%cf%83%ce%b7%cf%82%20%ce%ae%cf%87%ce%bf%cf%85.pdf?sequence=1&isAllowed=y>
- [36] S. Lu, Z. Lu, and Y.-D. Zhang, “Pathological brain detection based on AlexNet and transfer learning,” *J. Comput. Sci.*, vol. 30, pp. 41–47, Jan. 2019, doi: 10.1016/j.jocs.2018.11.008.
- [37] P. Gimeno, I. Viñals, A. Ortega, A. Miguel, and E. Lleida, “Multiclass audio segmentation based on recurrent neural networks for broadcast domain data,” *EURASIP J. Audio Speech Music Process.*, vol. 2020, no. 1, p. 5, Dec. 2020, doi: 10.1186/s13636-020-00172-6.
- [38] Drossos Konstantinos, K. Rigas, Pappas Panos, Kalliris George, and Floros Andreas, “Investigating Auditory Human-Machine Interaction: Analysis and Classification of Sounds Commonly Used by Consumer Devices.,” Rome, Italy, Dec. 2013. [Online]. Available: [https://www.researchgate.net/publication/237841429\\_Investigating\\_Auditory\\_Human-Machine\\_Interaction\\_Analysis\\_and\\_Classification\\_of\\_Sounds\\_Commonly\\_Used\\_by\\_Consumer\\_Devices](https://www.researchgate.net/publication/237841429_Investigating_Auditory_Human-Machine_Interaction_Analysis_and_Classification_of_Sounds_Commonly_Used_by_Consumer_Devices)
- [39] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software: an update,” *ACM SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009, doi: 10.1145/1656274.1656278.
- [40] R. Kotsakis, G. Kalliris, and C. Dimoulas, “Investigation of broadcast-audio semantic analysis scenarios employing radio-programme-adaptive pattern classification,” *Speech Commun.*, vol. 54, no. 6, pp. 743–762, Jul. 2012, doi: 10.1016/j.specom.2012.01.004.
- [41] C.-H. Yeh, H.-H. Lin, and H.-T. Chang, “An efficient emotion detection scheme for popular music,” in *2009 IEEE International Symposium on Circuits and Systems*, Taipei, Taiwan, May 2009, pp. 1799–1802. doi: 10.1109/ISCAS.2009.5118126.

- [42] S. Bedoya-Jaramillo, E. Belalcazar-Bolanos, T. Villa-Canas, J. R. Orozco-Arroyave, J. D. Arias-Londono, and J. F. Vargas-Bonilla, “Automatic emotion detection in speech using mel frequency cepstral coefficients,” in *2012 XVII Symposium of Image, Signal Processing, and Artificial Vision (STSIVA)*, Medellin, Antioquia, Colombia, Sep. 2012, pp. 62–65. doi: 10.1109/STSIVA.2012.6340558.
- [43] F. S. A., V. K. V.R., R. S. A., A. Jayakumar, and B. A. P., “Speaker Independent Automatic Emotion Recognition from Speech: A Comparison of MFCCs and Discrete Wavelet Transforms,” in *2009 International Conference on Advances in Recent Technologies in Communication and Computing*, Kottayam, Kerala, India, 2009, pp. 528–531. doi: 10.1109/ARTCom.2009.231.
- [44] Md. Sahidullah and G. Saha, “Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition,” *Speech Commun.*, vol. 54, no. 4, pp. 543–565, May 2012, doi: 10.1016/j.specom.2011.11.004.
- [45] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 293–302, Jul. 2002, doi: 10.1109/TSA.2002.800560.
- [46] P. N. Juslin, “Cue utilization in communication of emotion in music performance: Relating performance to perception.,” *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 26, no. 6, pp. 1797–1812, Dec. 2000, doi: 10.1037/0096-1523.26.6.1797.
- [47] E. Tuomas, L. Olivier, and T. Petri, “PREDICTION OF MULTIDIMENSIONAL EMOTIONAL RATINGS IN MUSIC FROM AUDIO USING MULTIVARIATE REGRESSION MODELS,” ISMIR, 2009. [Online]. Available: <https://ismir2009.ismir.net/proceedings/PS4-8.pdf>
- [48] Κύρκος, Ε, “Επιχειρηματική Ευφυΐα & Εξόρυξη Δεδομένων.,” Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, ΕΜΠ, Αθήνα, 2015.
- [49] M. Burkhardt, “Automatic prediction of student’s personality, self-concept and intelligence.,” 2016, doi: 10.13140/RG.2.2.22426.03520.
- [50] P. Κωτσάκης and R. Kotsakis, “Εφαρμογή αλγορίθμων μηχανικής εκμάθησης για εξόρυξη και κατηγοριοποίηση πληροφοριών περιεχομένου στα οπτικοακουστικά μέσα.,” Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης (ΑΠΘ), Τμήμα Δημοσιογραφίας και Μέσων Μαζικής Επικοινωνίας, 2015. doi: 10.12681/eadd/35580.
- [51]. Allibhai E, “Hold-out vs. Cross-validation in Machine Learning,” Oct. 03, 2018. Hold-out vs. Cross-validation in Machine Learning
- [52] Φερφέλης, Θωμάς, “Μηχανική μάθηση-ανάλυση συναισθήματος,” BS thesis, 2019.