ΤΕΙ ΚΡΗΤΗΣ
ΠΑΡΑΡΤΗΜΑ ΡΕΘΥΜΝΟΥ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΜΟΥΣΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΑΚΟΥΣΤΙΚΗΣ

Αυτόματη αναγνώριση ηχοτοπίων και ηχητικών γεγονότων σε περιβάλλον πόλης

Πτυχιακη Εργασια

του

ΚΑΜΠΙΤΑΚΗ ΜΙΧΑΛΗ

Επιβλέπων: Παναγιώτης Ζέρβας

Ρέθυμνο 2019

ΤΕΙ ΚΡΗΤΗΣ
ΠΑΡΑΡΤΗΜΑ ΡΕΘΥΜΝΟΥ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΜΟΥΣΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΑΚΟΥΣΤΙΚΗΣ

Αυτόματη αναγνώριση ηχοτοπίων και ηχητικών γεγονότων σε περιβάλλον πόλης

Πτυχιακη Εργασια

του

ΚΑΜΠΙΤΑΚΗ ΜΙΧΑΛΗ

Επιβλέπων: Παναγιώτης Ζέρβας

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 16η Ιανουαρίου 2019.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Παναγιώτης Ζέρβας	Σπύρος Κουζούπης	Νίκος Στεφανακης
Επίκουρος Καθηγητής	Επίκουρος Καθηγητής	Επίκουρος Καθηγητής

Ρέθυμνο 2019

ΤΕΙ ΚΡΗΤΗΣ
ΠΑΡΑΡΤΗΜΑ ΡΕΘΥΜΝΟΥ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΜΟΥΣΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΑΚΟΥΣΤΙΚΗΣ

Υπεύθυνη Δήλωση

Βεβαιώνω ότι είμαι συγγραφέας αυτής της πτυχιακής εργασίας, και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην πτυχιακή εργασία. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επίσης, βεβαιώνω ότι αυτή η πτυχιακή εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για τις απαιτήσεις του προγράμματος σπουδών του τμήματος Μηχανικών Μουσικής Τεχνολογίας και Ακουστικής του ΤΕΙ Κρήτης.

(Υπογραφή)

.....
Μηχάλης Καμπιτάκης

Περίληψη

Η εργασία αυτή, έχει σαν σκοπό την θεωρητική μελέτη συστημάτων αυτόματης αναγνώρισης ηχητικών γεγονότων μέσω μηχανικής μάθησης, την ανάπτυξη δικού μας συνόλου δεδομένων από ηχογραφήσεις σε σημεία της πόλης του Ρεθύμνου, καθώς και υλοποίηση ενός συστήματος για αναγνώριση και κατηγοριοποίηση ηχοτοπίων και ηχητικών γεγονότων με την χρήση της βάσης που δημιουργήθηκε.

Στο πρώτο κεφάλαιο, γίνεται η περιγραφή του ορισμού της μηχανικής μάθησης, της ηχητικής αναγνώριση - ταξινόμησης ως πρόβλημα, τον λόγο δημιουργίας και υλοποίησής του, καθώς και της λειτουργίας του. Επιπλέον, γίνεται αναφορά του σκοπού συγγραφής της παρούσας πτυχιακής. Στο δεύτερο κεφάλαιο, παρουσιάζονται βάσεις δεδομένων και σύνολα αρχείων (data bases – data sets) από αναπτυγμένες βάσεις (είτε από μελέτες, είτε από διαγωνισμούς). Συγχρόνως, περιγράφεται το σύνολο επιλογής ως βάση σύγκρισης, καθώς και την διαδικασία ανάπτυξης της δικής μας βάσης. Η θεματική ενότητα του τρίτου κεφαλαίου αφορά τον ορισμό και την ανάλυση της στατιστικής της μηχανικής μάθησης και τις συσχετιζόμενες μαθηματικές σχέσεις. Ταυτοχρόνως, εξετάζεται η λειτουργία των πιο συνηθισμένων και σύγχρονων αλγορίθμων μηχανικής μάθησης (όπως ο SVM και KNN). Ταυτόχρονα, στο κεφάλαιο αυτό, αναλύονται διαδικασίες για την εκτίμηση των μοντέλων μηχανικής μάθησης, και τρόποι για μείωση των διαστάσεων των χαρακτηριστικών (clustering, Feature Selection). Το τέταρτο κεφάλαιο, είναι αφιερωμένο στο κομμάτι επεξεργασίας σήματος, της συνολικής διαδικασίας που χρειάζεται για την παραγωγή χρήσιμων ακουστικών χαρακτηριστικών από το σήμα του ήχου. Ακολούθως, στο πέμπτο κεφάλαιο γίνεται εκτενής περιγραφή της υλοποίησης και ανάπτυξης του συστήματός μας, καθώς και της λογικής που σχετίζεται με αυτό. Τα αποτελέσματα των πειραμάτων παρουσιάζονται στο έκτο κεφάλαιο, ενώ στο έβδομο και τελευταίο, επισημαίνονται τα συμπεράσματα.

Λέξεις Κλειδιά

Αυτόματη ταξινόμηση, μηχανική μάθηση, ηχοτοπία, ηχητικά γεγονότα, SVM, feature selection

Abstract

This thesis performs a theoretical study of automatic recognition systems on audio events, due to machine learning, developing of our own data set of recordings in locations on town Rethymno, and also implementation of our own system for recognition and classification soundscapes and audio events using the created base

In the first chapter, is a description, on what machine learning is, the audio recognition and classification as problem, the reason on creation and implementation, and how does it works. Also there is description of our own work and mention of the purpose of writing this thesis. In the second chapter, are presented data bases and data sets, from developed bases (either from studies or competitions), At the same time, the selection set for reference choice is described and also the implementation process of our own data base. The thematic section of the third chapter, is about analysis of statistics of machine learning and their related mathematics. At the same time, is being considered the operation of the most ordinary and modern machine learning algorithms (like SVM and KNN). Also in this chapter, analyzes the processes of the machine learning model evaluation, and dimensional reduction methods for features (clustering, feature selection). The forth chapter, is dedicated to signal processing part, the whole process needed for useful acoustic features creation from audio signal. Subsequently, the fifth chapter is the extensive description of the implementation and development of our system, and also all the logic behind it. Experiment results are in the sixth chapter and last but not least, in the seventh chapter are highlight the conclusions.

Keywords

Automatic classification, machine learning, soundscapes, audio events, SVM, feature selection

στους γονείς μου

Ευχαριστίες

Αυτή η εργασία, δεν θα είχε ολοκληρωθεί, εάν δεν ήταν ο υπευθύνος καθηγητής μου, Παναγιώτης Ζέρβας, τον οποίο ευχαριστώ πολύ για όλη την καθοδήγηση που μου έδωσε κατά τη διάρκεια της έρευνας, ώστε να κατανοήσω ένα θέμα, το οποίο ήταν εντελώς άγνωστο σε εμένα. Επίσης θα ήθελα να ευχαριστήσω την Ανδριάνα, τον Γιάννη, τον Παναγιώτη και την Ελένη, για την βοήθεια με την εργασία και την στήριξή τους. Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου, που με στήριξαν όλα αυτά τα χρόνια.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

Η εργασία αυτή πραγματεύεται το ζήτημα της ταξινόμησης ηχητικής πληροφορίας μέσω τεχνικών μηχανικής μάθησης. Για να γίνει αυτό, έχει προηγηθεί δημιουργία ηχητικού συνόλου δεδομένων, ηχογραφημένο στην πόλη του Ρεθύμνου, και πάνω σε αυτήν γίνεται η εκπαίδευση τριών μοντέλων, τα οποία, φορτωμένα σε καταλλήλως διαμορφωμένο κόμβο (υπολογιστή κάρτας Raspberry Pi), δίνουν την δυνατότητα στον χρήστη να λαμβάνει πληροφορία σε πραγματικό χρόνο, για τον τύπο των ηχητικών δεδομένων που λαμβάνει ο κόμβος.

1.1 Ηχοτοπία

Ο όρος ηχοτοπίο [1], καθώς και η ανάγκη ανάδειξης της σημασίας του για περαιτέρω μελέτη, προέρχεται από τον R. Murray Schafer και την ερευνητική του ομάδα, έπειτα από έρευνες που διεξήχθησαν στα τέλη της δεκαετίας του 60 [2]. Πρόκειται, όπως ο ίδιος ορίζει, για τοπίο με ήχους. Οι ήχοι που απαρτίζουν ένα ηχοτοπίο είναι το σύνολο των γεωλογικών, των βιολογικών και των ανθρωπογενών ήχων. Γεωλογικοί ήχοι ονομάζονται αυτοί οι οποίοι παράγονται από τη ίδια τη γη (φύση), όπως ο ήχος του αέρα, της θάλασσας, οι υπόηχοι από τις μικροδονήσεις του εδάφους και άλλοι. Βιολογικοί ονομάζονται όλοι οι ήχοι οι οποίοι παράγονται από ζωντανούς οργανισμούς, όπως δηλαδή αυτοί που παράγονται από τα ζώα και τους ανθρώπους. Από την άλλη μεριά, οι ανθρωπογενείς ήχοι είναι εκείνοι που προέρχονται από την ανθρώπινη παραγωγή, όπως οι ήχοι εργοστασίων, πόλεων, αυτοκινήτων, αεροπλάνων. Η αύξηση της στάθμης της ηχητικής πίεσης θεωρείται πηγή μόλυνσης λόγω επικείμενων επιδράσεων που φέρουν δυσμενείς επιπτώσεις στον άνθρωπο, προκαλώντας του διαταραχές ύπνου και δυσκολία στην ανθρώπινη επικοινωνία, ενώ η μακρόχρονη έκθεση σε έντονο ηχητικό περιβάλλον ή θόρυβο μπορεί να ανεβάσει τα επίπεδα στρες, άγχους και άλλα. Παράλληλα, ο θόρυβος έχει επίδραση και στο ζωικό βασίλειο, στο οποίο κατά τα τελευταία χρόνια έχει παρατηρηθεί αύξηση της έντασης των ήχων που παράγουν, στην προσπάθεια για την μεταξύ τους επικοινωνία.

Πιο συγκεκριμένα, στην εργασία αυτή γίνεται μελέτη των κατηγοριών του αστικού ηχοτοπίου (Urban Sound Scape). Ένα αστικό ηχοτοπίο αποτελείται κυρίως από ανθρωπολογικούς ήχους. Με την πάροδο, όμως, των χρόνων και την εξέλιξη της τεχνολογίας αυξήθηκε η

συνολική ηχητική στάθμη, ενώ παράλληλα άλλαξε η μορφή της πιστότητας (συχνοτικό περιεχόμενο) του ήχου. Ακόμη, οι ηχητικοί περίπατοι σε αστικά ηχοτοπία ως διδακτική πρόταση [3], εντοπίζονται σε ποικίλες λογοτεχνικές πηγές όπως στην νουβέλα "Κάπου περνούσε μια φωνή" του Ναπολέοντα Λαπαθιώτη [4], όπου περιγράφεται ηχοπερίπατος σε δρόμους της Αθήνας το 1915. Μέσω αυτού δύνανται να διεξαχθούν συμπεράσματα για την αντίθεση του φάσματος που επικρατούσε στην Αθήνα στις αρχές του 20ου αιώνα σε σύγκριση με την σημερινή. Στο βιβλίο αναφέρεται ότι στην Αθήνα του 1915 επικρατούσαν ήχοι υψηλής πιστότητας (Hi-Fi) με ήχους από κάρα, ήχους ζώων, περιπλανώμενων πωλητών και ομιλίες μεταξύ ανθρώπων. Εν αντιθέσει με τότε, στην Αθήνα της σύγχρονης εποχής επικρατούν ήχοι χαμηλής πιστότητας (Lo-Fi), παραγόμενοι από μηχανές, αυτοκίνητα, εργοστάσια, μηχανήματα που δουλεύουν κι άλλα μέσα τα οποία παράγουν θόρυβο, με αποτέλεσμα οι ήχοι υψηλής πιστότητας που υπάρχουν στο περιβάλλον να είναι καλυμμένοι πλήρως και να μην γίνονται αντιληπτοί. Σε περιοχές με αυξημένο θόρυβο έχουν γίνει προσπάθειες για τον περιορισμό της στάθμης έντασης του θορύβου που λαμβάνει ο άνθρωπος (π.χ με ηχοφράγματα), ενώ τα τελευταία χρόνια γίνεται προσπάθεια για την καταγραφή του, προκειμένου να αντιμετωπιστεί με επιτυχία.

Λόγω της φύσης των σημερινών μεγαλουπόλεων και του μεγέθους διαφορετικών τοπίων που μπορεί να συναντήσει κανείς, υπάρχει αντιστοίχως μεγάλη ποικιλία ηχοτοπίων και κατηγοριών. Στη συνέχεια, παρουσιάζεται η λογική με την οποία γίνεται η κατηγοριοποίηση (ή αλλιώς ταξινόμηση) των ήχων ενός ηχοτοπίου, καθώς και οι λόγοι για τους οποίους υπάρχει η ανάγκη συλλογής τέτοιας πληροφορίας.

1.2 Μηχανική μάθηση και η Ακουστική Αναγνώριση

Η Μηχανική μάθηση, ανήκει σε ένα ευρύτερο επίπεδο, το οποίο ονομάζεται Artificial Intelligence (AI) [5] [6]. Πρόκειται για μια τεχνολογία στην επιστήμη των υπολογιστών, με την οποία μπορεί κάποιος να εκπαιδεύσει με κατάλληλα δεδομένα ένα μοντέλο, με σκοπό την βελτιστοποίηση του αλγόριθμου. Πρόκειται για την τομή μεταξύ των επιστημονικών πεδίων της επιστήμης των υπολογιστών, της μηχανικής και της στατιστικής. Σκοπός της μηχανικής μάθησης είναι η γενίκευση αναγνωρίσιμων προτύπων ή η δημιουργία αγνώστων κανόνων από τα δεδομένα εισόδου.

Ο τρόπος με τον οποίο ένα μοντέλο μπορεί να δημιουργηθεί, χωρίζεται σε τρεις κατηγορίες. Η πρώτη κατηγορία δημιουργίας μοντέλων είναι η επιβλεπόμενη εκπαίδευση (Supervised learning). Πρόκειται για έναν τρόπο εκπαίδευσης ο οποίος μοιάζει με τον τρόπο που ένας δάσκαλος διδάσκει τους μαθητές του· δηλαδή με καθοδήγηση και συνακόλουθη αξιολόγηση σύμφωνα με την απόδοσή τους. Με την τεχνική αυτή, μπορεί να δημιουργηθούν δύο ειδών μοντέλα. Το πρώτο είναι μοντέλο ταξινόμησης (classification), το οποίο μπορεί να διαχειρίζεται διακριτές τιμές και χρησιμοποιείται για να επιστρέφει πιθανότητες της εισόδου σαν κλάση, ετικέτα ή κάποιο άλλο tag. Το δεύτερο είναι μοντέλο παλινδρόμησης (regression), το οποίο διαχειρίζεται συνεχόμενες τιμές και χρησιμοποιείται για την εξαγωγή ποσοστιαίου αποτελέσματος από την είσοδο. Η δεύτερη κατηγορία δημιουργίας μοντέλων είναι η μη-επιβλεπόμενη μάθηση (Unsupervised learning), όπου ο αλγόριθμος αυτός, μαθαίνει μόνος του, χωρίς καθοδήγηση

και δίχως να έχει κάποια μεταβλητή σαν στόχο. Ο αλγόριθμος αυτός ψάχνει για κρυμμένα μοτίβα στα δεδομένα που αναλύει και τις σχέσεις μεταξύ αυτών. Τα είδη μοντέλων που έχουν τη δυνατότητα να δημιουργηθούν με αυτόν τον τρόπο είναι μοντέλα μείωσης διαστάσεων (Dimensionality reduction) και μοντέλα τμηματοποίησης (Clustering). Επιπροσθέτως, η τρίτη κατηγορία είναι η ενισχυμένη εκπαίδευση Reinforcement learning, κατά την οποία το μοντέλο μαθαίνει να συμπεριφέρεται μέσα από την αξιολόγησή του από το ίδιο το περιβάλλον. Στην εκπαίδευση αυτή, το μοντέλο παίρνει μια σειρά από αποφάσεις χωρίς επίβλεψη και στο τέλος αξιολογεί τις αποφάσεις αυτές με $+1$ ή -1 αντίστοιχα, δημιουργώντας το μονοπάτι του μέσω της τελικής ανταμοιβής των αποφάσεών του. Το είδος αυτό είναι πολύ πιο κοντά στην τεχνητή νοημοσύνη, παρά στην ίδια τη μηχανική μάθηση. Σε μερικές περιπτώσεις των προβλημάτων της μηχανικής μάθησης, εφαρμόζεται πρώτα μη-επιβλεπόμενη μέθοδος για την μείωση των διαστάσεων του προβλήματος αυτού και στην συνέχεια ακολουθεί η επιβλεπόμενη μέθοδος εκπαίδευση για την λήψη της επιθυμητής απόφασης.

Όπως προαναφέρθηκε, η επιβλεπόμενη μηχανική μάθηση (Supervised Machine Learning) που απασχολεί το υφιστάμενο προς μελέτη πρόβλημα, χωρίζεται σε δύο κατηγορίες: regression και classification. Στην παρούσα εργασία , θα εξεταστεί μονάχα το classification, το οποίο είναι μια σειρά από διαδικασίες, μέσα από τις οποίες μια μηχανή μπορεί να μάθει να αναγνωρίζει κοινά πρότυπα χαρακτηριστικών, που δύνανται να διεξαχθούν από την είσοδο των δειγμάτων μας (εικόνα, ήχος, κείμενο κ.α.), μέσω ετικετών που του έχουν οριστεί για αυτά τα δεδομένα. Έπειτα γίνεται σύγκριση αυτών με σκοπό την δημιουργία διαχωρισμένων τμημάτων χαρακτηριστικών που ανήκουν σε κάθε μία από τις εν λόγω κατηγορίες (- classes) ή ετικέτες που επιθυμούμε να κατηγοριοποιήσουμε.

Αναφορικά με το κομμάτι του ήχου, τα τελευταία χρόνια με την αύξηση της ηχητικής πληροφορίας, ολοένα και περισσότερο εμφανίζεται η ανάγκη για αυτοματοποιημένα συστήματα αναγνώρισης πληροφορίας. Τόσο για αναγνώριση της ομιλίας ή της ταυτότητας του ομιλητή [7] [8], αναγνώριση συναισθημάτων ανθρώπου [9] (με το σύστημα openSMILE της audEERING, που θα παρουσιαστεί παρακάτω), εφαρμογές αυτόματης αναγνώρισης κομματιών μουσικής (Shazam [10] κ.α.), συστήματα για εξωτερικούς ήχους περιβάλλοντος ή ήχους πόλης με συνεχείς λήψεις για ακουστική εικόνα (SmartSantander [11] [12]), οργανώσεις, μετρήσεις, διαχειρίσεις και αναλύσεις αυτών. Οι περιπτώσεις είναι πολλές ενώ η χρησιμότητά τους ποικίλει: από ψυχαγωγικούς σκοπούς (αναγνώριση μουσικής), επιστημονικούς σκοπούς (μελέτες περιβαλλόντων και ηχοτοπίων) μέχρι και σε θέματα υγείας (όπως π.χ. αυτόματη αναγνώριση ατυχημάτων σε περιβάλλοντα έξυπνης πόλης - SmartSantander, διάγνωση νόσων υγείας και βοήθεια ατόμων με ειδικές ανάγκες [13] [14]).

Η ηχητική αναγνώριση (Sound Recognition) είναι η διαδικασία κατά την οποία αναγνωρίζεται ένα ηχητικό περιεχόμενο, το οποίο έχει καθορισθεί μέσα από ένα σύνολο μιας ηχογράφησης ή άλλου αρχείου ήχου. Για να επιτευχθεί αυτό, θα πρέπει να υπάρξει πρώτα η διαδικασία της συλλογής μιας βάσης δεδομένων (data base ή αλλιώς σύνολο δεδομένων - dataset) και ο καθορισμός των περιεχομένων που βρέθηκαν και τα οποία επιθυμούμε να κατηγοριοποιήσουμε μέσω επισήμανσης ετικετών και ορισμού όμοιων ήχων σαν κλάσεις. Έπειτα, την συλλογή δε-

δομένων ακολουθεί η διαδικασία προετοιμασίας των αρχείων, προκειμένου να είναι ίσα ως προς το πλήθος και παρόμοια ως προς την μορφή (normalization) για όλες τις κατηγορίες. Κατόπιν, έρχεται η εξαγωγή των χαρακτηριστικών. Συγκεκριμένα, τα χαρακτηριστικά του ήχου ποικίλουν, όμως, τα πιο συνηθισμένα είναι τα MFCC [15], όπου θα αναλυθούν αργότερα. Στην συνέχεια, συναντάται η εκπαίδευση (training) του συστήματος, κατά την οποία γίνεται ο διαχωρισμός του περιεχομένου και με βάση αυτό το σύστημα μαθαίνει από τα χαρακτηριστικά του κάθε ηχητικού περιεχόμενου, έτσι ώστε να μπορεί να το αναγνωρίσει όταν αργότερα συναντήσει τμήμα δείγματος με όμοια χαρακτηριστικά. Αυτό γίνεται μέσω μιας σειράς από μαθηματικές πράξεις, κατά τις οποίες ομαδοποιούνται τα χαρακτηριστικά της κάθε κατηγορίας, δημιουργώντας έναν χάρτη χαρακτηριστικών και την περιοχή όπου ανήκουν, βάση του οποίου λαμβάνεται η πιθανότητα να ανήκει το δοκιμαζόμενο τμήμα στην αντίστοιχη κατηγορία. Εν συνεχεία, το σύστημα περνάει από μια σειρά δοκιμές (Test ή αλλιώς Evaluation), κατά τις οποίες εκτιμάται η απόδοσή του (ως προς την ακρίβεια, την ταχύτητα, την σχέση μεταξύ δεδομένων εισόδου – αποτελεσμάτων, ενεργειακής κατανάλωσης κ.α.) και αξιολογείται για το αν και κατά πόσο χρειάζεται βελτίωση. Κατά τη διαδικασία, λοιπόν, της ταξινόμησης (classification), το σύστημα είναι σε θέση να αναγνωρίσει ηχητικό περιεχόμενο (με αρκετά υψηλό ποσοστό επιτυχίας) και να το κατατάξει στην αντίστοιχη κλάση ή κατηγορία, σύμφωνα με την διαδικασία μάθησης που αναφέρθηκε προηγουμένως. Αξιοσημείωτο είναι το γεγονός πως ακόμη και ο άνθρωπος δεν είναι ικανός να αναγνωρίσει ηχητικό περιεχόμενο στο 100%, αφού πάντα υπάρχει ένα μικρό σφάλμα, ανάλογο της ηλικίας, της κατάστασης της υγείας του και άλλων παραγόντων.

Στην παρούσα εργασία, το ηχητικό περιεχόμενο που κατατάσσεται είναι ηχητικά γεγονότα (Events) από ηχογραφήσεις στην πόλη του Ρεθύμνου και ηχοτοπία (Soundscapes) περιοχών της πόλης. Περαιτέρω ανάλυση για τα εν λόγω ζητήματα, θα γίνει σε επόμενο κεφάλαιο.

1.3 Σχετική Εργασία

Αφού εξετάστηκε τι είναι το ηχοτοπίο και πώς χρησιμοποιείται η ηχητική κατηγοριοποίηση και αναγνώριση, γίνεται η εξέταση κάποιων υλοποιήσεων και εφαρμογών σύγχρονης τεχνολογίας αιχμής (συγκεκριμένα κάποιες από αυτές αναφέρθηκαν ήδη προηγουμένως).

Project EAR-IT : Το πρότζεκτ αυτό [11] [16] [12], είναι μια υλοποίηση δικτύου με ακουστικούς αισθητήρες (μικρόφωνα), το οποίο έχει σκοπό τον συνεχή ακουστικό έλεγχο σε αληθινό χρόνο. Το δίκτυο αυτό λειτουργεί μέσω πλατφόρμας Internet of Thins (IoT) και αποτελεί μέρος δύο μεγαλύτερων project. Το πρώτο είναι το HOBNet (Holistic Platform Design for Smart Buildings of the Future Internet) [16], πλαίσιο με το οποίο γίνεται η σχεδίαση έξυπνων κτιρίων, με σκοπό την ανάπτυξη εφαρμογών αυτοματισμού και ενεργειακής απόδοσης. Στο κομμάτι του ήχου, το Project EAR-IT προσφέρει ακουστική άνεση στον εσωτερικό χώρο, ασφάλεια μέσω ηχητικής παρακολούθησης και καλύτερη διαχείριση της ενέργειας, μέσω αναγνώρισης ανθρώπινων δραστηριοτήτων στο εσωτερικό περιβάλλον. Το δεύτερο είναι το poject SmartSantander [11], για το οποίο έγινε λόγος πριν. Πρόκειται για

ένα πλαίσιο αισθητήρων (πάλι μέσω πλατφόρμας IoT), το οποίο επεκτείνεται σε μια ολόκληρη πόλη, την Santander της Ισπανίας. Μέσω αισθητήρων σε όλη την πόλη υπάρχει η δυνατότητα δημιουργίας εφαρμογών αυτοματισμών για την διευκόλυνση και την υγεία των πολιτών και καλύτερης ενεργειακής απόδοσης ολόκληρης της πόλης, όπως και στο προηγούμενο. Ως προς τον ήχο, το Project EAR-IT μετράει την πυκνότητα της κίνησης στους δρόμους για καλύτερη ρύθμιση της κυκλοφορίας, αναγνωρίζει περιπτώσεις έκτακτης ανάγκης και δημιουργεί χάρτες θορύβου (noise map monitoring). Επιπλέον, αξιοσημείωτο σε αυτήν την περίπτωση είναι ότι τα γεγονότα ταξινομούνται απομακρυσμένα, σε κεντρικό υπολογιστή και όχι τοπικά στον κάθε αισθητήρα.

Recognition of Speakers' Emotions and Characteristics : Στην έρευνα αυτή [17], περιγράφεται ένα υλοποιημένο σύστημα ηχητικής κατηγοριοποίησης σε ανθρώπινη ομιλία, με σκοπό την κατανόηση των συναισθημάτων. Αυτό γίνεται εφικτό μέσω μιας τροποποιημένης έκδοσης του πακέτου εξαγωγής ακουστικών χαρακτηριστικών της audEERING, το openSMILE [9] (αναλύεται παρακάτω), το οποίο εξάγει πάνω από 6.000 χιλιάδες διαφορετικά είδη ακουστικών χαρακτηριστικών. Η κατηγοριοποίηση αυτών γίνεται μέσω συνδυασμού δύο μεθόδων: της Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN), αλγόριθμος δηλαδή ο οποίος αναλαμβάνει να βρει τα τμήματα στα οποία υπάρχει φωνή, και της κατηγοριοποίησης με αλγορίθμους SVM (Support-Vector Machine) για αναγνώριση συναισθηματικών κατηγοριών. Η όλη διαδικασία είναι ικανή να τρέξει σε κινητό τηλέφωνο και επεξεργαστή τύπου ARM χωρίς να υπάρχει η ανάγκη άλλων μηχανημάτων (όπως μέσω streaming ήχου σε ξεχωριστό μηχάνημα μόνο για αυτή τη δουλειά) όπως στο προηγούμενο παράδειγμα και όπως ήταν σύνηθες μέχρι πρότινος. Για την υλοποίηση του συστήματος αυτού, πάρθηκαν υπόψιν αποτελέσματα που εξήχθησαν σε χρονικά benchmarks με πληθώρα χαρακτηριστικών, συγκεκριμένα από λιγότερα των 100 έως και μερικές χιλιάδες, ώστε το σύστημα να είναι όχι μόνο αρκετά γρήγορο, αλλά και ελαφρύ ώστε να μπορεί να τρέξει σε κινητό τηλέφωνο.

HomeSound : Στην παρούσα έρευνα [13], η οποία δημοσιεύτηκε στην ιατρική ιστοσελίδα της PubMed Central (PMC), παρουσιάζεται μια μέθοδος για ηχητική ιατρικής παρακολούθησης ασθενών ή ηλικιωμένων στα σπίτια τους, με σκοπό την αποσυμφόρηση των νοσοκομείων από ασθενείς. Μέσω του συστήματος που παρουσιάζεται, δίνεται η δυνατότητα αναγνώρισης ηχητικών γεγονότων, τόσο για την συνεχή παρακολούθηση της κατάστασης του ασθενή εξ αποστάσεως, όσο και για την έγκαιρη βοήθεια από τους γιατρούς σε περίπτωση ανάγκης. Το σύστημα αυτό τρέχει σε έναν επεξεργαστή γραφικών NVIDIA (Graphical Processing Unit) Jetson TK1 και αναγνωρίζει τα γεγονότα τοπικά, με έως και 82% ακρίβεια στην πρόβλεψή τους. Τα χαρακτηριστικά που εξάγονται από τον ήχο που λαμβάνουν οι αισθητήρες είναι Mel frequency-cepstral coefficients, ενώ η μέθοδος κατηγοριοποίησής τους γίνεται με Deep Neural Networks (DNNs), το οποίο, σε σύγκριση με τα Gaussian Mixture Models (GMMs) και Support Vector Machines (SVMs) που δοκιμάστηκαν, είναι πιο αποδοτικό ως προς την ακρίβεια, αλλά υστερεί στο κομμάτι του συνολικού φόρτου πληροφορίας που μπορεί να διαχειριστεί. Έτσι λοιπόν, παίρνονται κάποια κεντρικά σημεία (centroids) ανάμεσα στα ακουστικά

χαρακτηριστικά με την μέθοδο της τμηματοποίησης (clastering), τα οποία υπολογίζονται με τον x-mean αλγόριθμο. Τα κεντρικά σημεία αποτελούν τα καινούρια χαρακτηριστικά με τα οποία γίνεται η κατηγοριοποίηση στις κλάσεις.

Audio-only bird classification using unsupervised feature learning : Πρόκειται για μια υλοποίηση ταξινομητή [18], η οποία χρησιμοποιεί ανεπεξέργαστη είσοδο ηχητικού σήματος για την αναγνώριση διαφόρων ειδών πουλιών. Τα χαρακτηριστικά που εξάγονται και εδώ είναι τύπου MFCC και τα οποία, αν και σχεδιασμένα για αναγνώριση ανθρώπινης ομιλίας, παρουσιάζουν μεγάλη ακρίβεια και σε ηχητικά ακούσματα πουλιών, πράγμα που οφείλεται στο εν μέρει κοινό συχνοτικό φάσμα. Η ταξινόμηση γίνεται μέσω αλγορίθμου k-mean και η προγραμματιστική υλοποίηση έχει γίνει σε περιβάλλον python. Η διαδικασία της ταξινόμησης γίνεται απομακρυσμένα μέσω audio streaming αντί για τοπικά στον ίδιο τον αισθητήρα.

1.4 Σκοπός πτυχιακής εργασίας

Η εργασία αυτή έχει σκοπό την δημιουργία βάσης δεδομένων από ηχοτοπία της πόλης του Ρεθύμνου και ηχητικών γεγονότων (που εξήχθησαν από τις ηχογραφήσεις), καθώς και υλοποίηση συστήματος για την αυτόματη αναγνώρισή τους μέσω αλγορίθμων ταξινόμησης, υλοποιημένο με οικονομικό εξοπλισμό. Συγκεκριμένα, χρησιμοποιείται μια πλακέτα υπολογιστή Raspberry Pi 2B, η οποία τρέχει μια ειδική έκδοση λογισμικού για την συγκεκριμένη συσκευή σε linux (Raspbian) και στην οποία καλείται ο αλγόριθμος ταξινόμησης, σχεδιασμένος σε γλώσσα Python.

Η παρούσα εργασία, χωρίζεται σε δύο μέρη. Το πρώτο μέρος είναι δημιουργία ακουστικής βάσης δεδομένων με σκοπό την εξαγωγή πληροφορίας από αυτήν, η οποία εκφράζεται σε ακουστικά χαρακτηριστικά, τιμές δηλαδή χρήσιμες για την διαδικασία του μοντέλου μηχανικής μάθησης, καθώς και η όλη διαδικασία εκπαίδευσης και εκτίμησης των μοντέλων ταξινόμησης. Τα χαρακτηριστικά που εξάγονται κάνουν εφικτή την εκπαίδευση μοντέλων μηχανικής μάθησης, με σκοπό την όσο το δυνατόν μεγαλύτερη επιτυχία στην πρόβλεψη ηχητικών γεγονότων και κατηγοριών ηχοτοπίων. Επιπροσθέτως, η βάση της εργασίας χωρίζεται και αυτή σε δύο μέρη. Το πρώτο αποτελεί τις κατηγορίες ηχοτοπίων που συναντήσαμε στην πόλη του Ρεθύμνου, όπου και έγιναν οι ηχογραφήσεις για την δημιουργία της βάσης αυτής. Το δεύτερο αφορά τα ηχητικά γεγονότα, όλα τα ακούσματα κλάσεων ήχου δηλαδή που υπήρξαν στις ηχογραφήσεις των ηχοτοπίων, ενοποιημένα και κατηγοριοποιημένα με τέτοιο τρόπο, ώστε να μην παίζει ρόλο ο χώρος από τον οποίο προέρχονται (ποιο ηχοτοπίο), αλλά το ίδιο το περιεχόμενό τους. Για το κομμάτι των ηχογραφήσεων της εργασίας, σχεδιάστηκε ένα φορητό καταγραφικό σύστημα, με την χρήση του Raspberry Pi, ενός Blue Snowflake πυκνωτικού μικροφώνου (συνήθης επιλογή για χρήση σε κλήσεις ηλεκτρονικού υπολογιστή, με κύριο χαρακτηριστικό του το αρκετά χαμηλό κόστος του) με ενσωματωμένο ψηφιακό μετατροπέα και προενισχυτή, καθώς και ένα powerbank χωρητικότητας 10.000 mAh το οποίο χρησιμοποιείται για την τροφοδότηση όλου του συστήματος (γιατί όπως αναφέρθηκε το σύστημα είναι φορητό). Για την καταγραφή, χρησιμοποιείται λογισμικό το οποίο σχεδιάστηκε με bash εντολές με την βο-

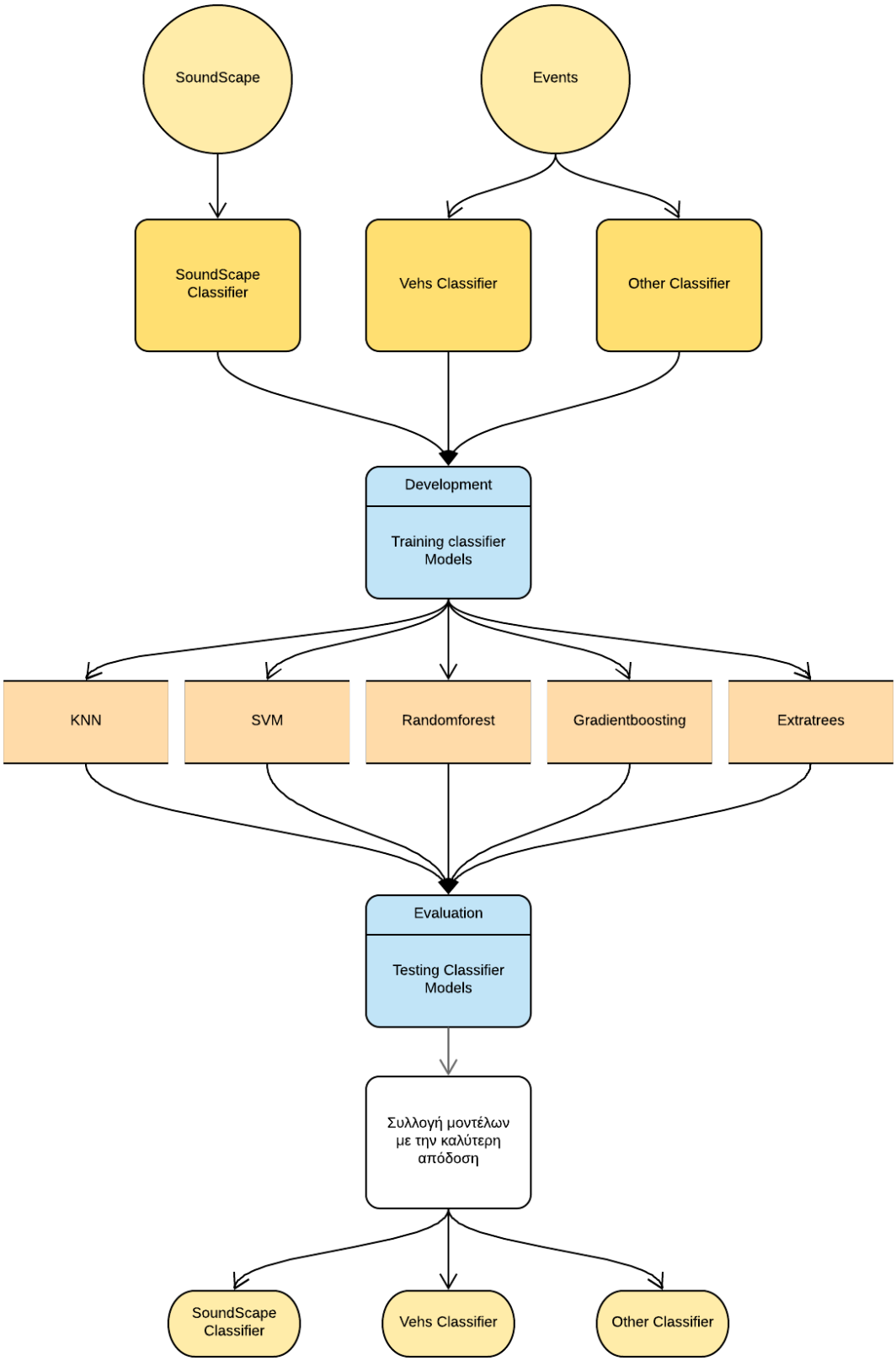
ήθεια της alsa βιβλιοθήκης που συνοδεύεται με τα λειτουργικά συστήματα linux (όπως είναι το Rasbian).

Την ολοκλήρωση των ηχογραφήσεων ακολουθεί η κατάτμηση των κατηγοριών και γεγονότων ήχου από το σύνολο των τοποθεσιών που πραγματοποιήθηκαν οι ηχογραφήσεις. Μετά την συλλογή αυτών, έχει πλέον ολοκληρωθεί η ανάπτυξη του συνόλου δεδομένων της εργασίας, πάνω στο οποίο μπορεί να πραγματοποιηθούν τα πειράματα για την δημιουργία των μοντέλων ταξινομητών μηχανικής μάθησης, ώστε να επιλέξουμε αυτά τα οποία αποδίδουν καλύτερα στην παρούσα περίπτωση. Για να γίνει αυτό, πραγματοποιήθηκε εξαγωγή χαρακτηριστικών με την βοήθεια του ανοικτού λογισμικού openSMILE της audEERING [9]. Η λογική πίσω από αυτό εξηγείται στα επόμενα κεφάλαια. Η διαδικασία της εκπαίδευσης και της εκτίμησης των αποτελεσμάτων φαίνεται στο διάγραμμα 1.1

Η διαδικασία, επαναλαμβάνεται για κάθε ένα από τα μοντέλα ταξινομητών. Στην εργασία αυτή, τα μοντέλα είναι τρία (όπως φαίνεται και στο σχήμα): το πρώτο είναι αποκλειστικά για τα ηχοπεριβάλλοντα και τα άλλα δύο αναλαμβάνουν διαφορετικά είδη ηχητικών γεγονότων (τα οποία χωρίζονται για ευκολία της υλοποίησης σε δύο κατηγορίες Vehs, όπου υπάρχουν κλάσεις οχημάτων και Others, όπου υπάρχουν όλες οι υπόλοιπες). Όταν πλέον υπάρχουν τα μοντέλα των ταξινομητών, τα οποία, σε συνεργασία με τα κατάλληλα χαρακτηριστικά, αποδεικνύουν ότι είναι τα ιδανικά για την κάθε περίπτωση, μπορεί να αρχίσει η υλοποίηση του δεύτερου μέρους της εργασίας αυτής, το οποίο είναι το πρακτικό κομμάτι για την υλοποίηση συστήματος κόμβου, ο οποίος θα αποτελεί ένα αυτόνομο τρόπο αυτόματης ταξινόμησης ηχητικών κλάσεων.

Για το δεύτερο μέρος λοιπόν, πραγματοποιείται βάσει της δικής μας υλοποίησης του συστήματος που αναφέρθηκε με τους ταξινομητές που περιγράφηκαν προηγουμένως και τα γεγονότα που εξάχθηκαν από την ηχητική βάση της πόλης του Ρεθύμνου. Για επεξεργαστική πλακέτα χρησιμοποιήθηκε η ίδια συσκευή Raspberry Pi, στην οποία έγινε υλοποίηση λογισμικού για ηχογράφηση και αναγνώριση τόσο ηχοτοπίων της πόλης, όσο και ηχητικών γεγονότων (όπως ακριβώς δηλαδή και στην διαδικασία της δημιουργίας την ηχητικής βάσης δεδομένων), με βάση τους ταξινομητές της προηγούμενης διαδικασίας, οι οποίοι φορτώνονται στην συσκευή. Η λήψη ήχου γίνεται μέσω USB μικροφώνου συνδεμένου στην συσκευή, το οποίο είναι υπεύθυνο και για την ψηφιοποίηση του ηχητικού σήματος. Στην συνέχεια, ακολουθεί η επεξεργασία του ψηφιακού σήματος, καθώς και η εξαγωγή χαρακτηριστικών από το σήμα μέσω του λογισμικού openSMILE που αναφέρθηκε. Έπειτα, τα χαρακτηριστικά περνάνε στον αλγόριθμο κατηγοριοποίησης, ο οποίος επιστρέφει προβλέψεις κλάσεων· σε ποια κλάση δηλαδή υπάρχει μεγαλύτερη πιθανότητα να ανήκει το ηχητικό τμήμα εκείνο από το οποίο εξάγθησαν τα χαρακτηριστικά αυτά. Για να γίνει αυτό, ο ήχος χωρίζεται σε μικρά τμήματα (παραθυροποίηση), στα οποία υπάρχει δυνατότητα για πιο εστιασμένη μελέτη στο σήμα. Τέλος, όλη η πληροφορία των προβλέψεων όλων των ταξινομητών που τρέχουν ταυτόχρονα, φτάνει στον χρήστη μέσω σύνδεσης σε τοπικό δίκτυο το οποίο δημιουργεί η συσκευή, εάν δεν υπάρχει δυνατότητα για σύνδεση σε υπάρχον.

Με λίγα λόγια σκοπός της εργασία αυτής, είναι περισσότερο η υλοποίηση του όλου συστήματος, και όχι τόσο τα ίδια τα γεγονότα προς αναγνώριση. Με αυτόν τον τρόπο όμως, μπορούν τόσο να μελετηθούν τα ηχοτοπία του Ρεθύμνου και τα ηχητικά γεγονότα που βρίσκο-



Σχήμα 1.1: Ροή εκπαίδευσης

νται σε αυτά, όσο και η πραγματοποίηση σταδιακής αύξησης στην απόδοση του αλγορίθμου, με προϋπόθεση την μελλοντική αύξηση στο μέγεθος της βάσης, αυξάνοντας παράλληλα τις δυνατότητες του ταξινομητή ως προς την ακριβέστερη αναγνώριση ηχοτοποίων και γεγονότων της εργασία αυτής και ως προς το πλήθος αυτών, με την προσθήκη δηλαδή νέων κλάσεων. Η εργασία αυτή δημιουργεί μια βάση, πάνω στα θεμέλια της οποίας θα μπορούσε να πατήσει μετέπειτα εργασία, με ιδέες πιο συγκεκριμένων συστημάτων και υλοποιήσεων. Με αυτόν τον τρόπο, η εργασία θα μπορούσε να συνεισφέρει σε project μεγαλύτερων διαστάσεων για υλοποίηση στην πόλη του Ρεθύμνου, όπως κάτι ανάλογο με το project EAR-IT που αναφέρθηκε στην πόλη Σανταντέρ της Ισπανίας. Ταυτόχρονα, το γεγονός ότι στο σύστημα αυτό γίνεται η πρόβλεψη στην ίδια την πλακέτα η οποία είναι υπεύθυνη για την λήψη του ήχου, απλουστεύει σε μελλοντικό στάδιο την κάλυψη περισσότερων σημείων μέσα στην πόλη και έξω από αυτήν, με την προσθήκη περισσότερων οικονομικών κόμβων σαν αυτόν που σχεδιάζεται στην εργασία και διευρύνοντας ένα δίκτυο. Μέσα στο δίκτυο οι κόμβοι αυτοί θα μπορούσαν να ανταλλάζουν πληροφορία, και το σύνολο αυτής, να καταλήγει στον ίδιο δέκτη μέσω του δικτύου, ο οποίος θα αναλάβει την συλλογή και διαχείρισή της. Με συνεχή λειτουργία ενός κόμβου σε μια συγκεκριμένη περιοχή, μπορεί να πραγματοποιηθεί αναλυτική μελέτη των γεγονότων που συναντήθηκαν σε αυτήν, και την σύνδεση των γεγονότων αυτών με μια πράξη αυτοματισμού, καθώς και με μελλοντική αναβάθμιση του αλγορίθμου για περισσότερη ακρίβεια και με την προσθήκη ποικιλίας επιλογών για αυτοματισμούς. Επιπλέον, με αυτοματοποιημένες ηχητικές μετρήσεις (για εξαγωγή πληροφορίας όπως ένταση, συχνοτικό φάσμα, χρονικά επαναλαμβανόμενα ηχητικά μοτίβα μέσα στην μέρα κ.α.), επιτρέπεται η δημιουργία ηχοχάρτη ο οποίος περιέχει την ηχητική κατάσταση της πόλης.

Κεφάλαιο 2

Σύνολο Δεδομένων

Επειδή ο αλγόριθμος ταξινόμησης είναι τόσο καλός όσο και η βάση δεδομένων με την οποία έγινε η εκπαίδευσή του [19], γίνεται ξεκάθαρο πως η βάση αυτή πρέπει να είναι όσο πιο σωστά δομημένη και ολοκληρωμένη γίνεται. Αυτό σημαίνει πως πρέπει να περιέχει υλικό για όλες τις κλάσεις προς αναγνώριση και ταυτόχρονα το υλικό αυτό να είναι αρκετό (όσο μεγαλύτερο γίνεται από άποψη ποσότητας και χρόνου), σωστά μετρημένο (ώστε όλες οι κλάσεις να έχουν τον ίδιο αριθμό δεδομένων για την εκπαίδευση) και επίσης πρέπει όλα τα αρχεία να έχουν ίδια μορφή (ίδιο τύπο αρχείου) ώστε να υπάρχει ομοιομορφία στο ψηφιακό υλικό που θα γίνει μετέπειτα η μελέτη. Στην περίπτωση των ηχοτοπίων και των ηχητικών γεγονότων μελέτης αυτής της εργασίας, τα δεδομένα είναι αρχεία ήχου και τα χαρακτηριστικά της μορφής τους, ως προς τα οποία πρέπει να υπάρξει προσοχή, ώστε να μην διαφέρουν, είναι ο τύπος αρχείου (.wav, .ogg, .mp3 κ.ο.κ), η συχνότητα δειγματοληψίας (σε Herz, 1600Hz, 22050Hz, 44100Hz κ.ο.κ), ο αριθμός των καναλιών, η ανάλυση των δειγμάτων (σε bit, 8, 16 κ.ο.κ), και τέλος, αναλόγως την εφαρμογή του αλγορίθμου, το ίδιο το μέγεθος του αρχείου σε δευτερόλεπτα (κανονικοποίηση χρόνου). Στην ενότητα αυτή, γίνεται περιγραφή μερικών από τις δημοσιευμένες βάσεις που υπάρχουν και μελετήθηκαν ενώ στην συνέχεια θα γίνει περιγραφή της βάση δεδομένων αυτής της εργασίας (rpi_Reth), η οποία περιέχει δεδομένα από ηχοτοπία και ηχητικά γεγονότα που μπορεί κανείς να συναντήσει στην πόλη του Ρεθύμνου, στην περιοχή δηλαδή όπου και πραγματοποιήθηκαν οι ηχογραφήσεις.

2.1 Βάσεις αναφοράς

Στην εργασία αυτή, βάσεις αναφοράς ονομάζονται τα σύνολα δεδομένων τα οποία έχουν δημοσιευτεί στο διαδίκτυο [20] και τα οποία έχουν είτε εκπαιδευτικό σκοπό, είτε ερευνητικό, είτε είχαν χρησιμοποιηθεί σε διαγωνισμούς μεγαλύτερης ακρίβειας αλγορίθμων ταξινομητών (όπως θα αναλυθεί και στην συνέχεια). με την βοήθεια σχετικής έρευνας αντιστοίχων ερευνών [20], γίνεται εύκολα δυνατή η εύρεση μαζεμένων πηγών από άλλες έρευνες, καθώς και των αντίστοιχων βάσεων δεδομένων τους μαζί με τα γεγονότα / tags που χρησιμοποίησαν. Στον πίνακα 2.1 παρουσιάζονται κάποια περιγραφικά στοιχεία που συλλέχθηκαν από τις βάσεις αυτές. Τα στοιχεία αυτά είναι ο πάροχος / οργανισμός που οργάνωσε την κάθε βάση, το όνομα της

βάσης, ο τύπος ηχογράφησης (ηχογραφημένος ζωντανά ή σε εργαστηριακό περιβάλλον), ο τύπος σημείωσης (annotation), η διάρκεια ηχογραφήσεων, το πλήθος των αρχείων, το πλήθος περιεχομένου και τέλος το πλήθος των γεγονότων της database. Στις παρακάτω παραγράφους της επόμενης ενότητας 2.1.1, σχολιάζονται σύντομα κάποιες από τις έρευνες αυτές, οι οποίες χρησιμοποίησαν τις βάσεις που αναφέρθηκαν με αλγορίθμους μηχανικής μάθησης για την δημιουργία των μοντέλων τους ενώ παράλληλα πραγματοποιείται μελέτη του τρόπου και της λογικής κατά την δημιουργία της αντίστοιχης βάσης, τον σκοπό της δημιουργίας της, καθώς και πώς πραγματοποίησαν την διαδικασία τους.

2.1.1 Data Sets

CICESE Σκοπός των ηχογραφήσεων αυτών [21] [22] του πανεπιστημίου CICESE είναι η μάθηση και δοκιμή συστήματος για αναγνώριση ηχητικών γεγονότων (events) τα οποία δεν είναι απομονωμένα αλλά μέσα σε ηχητικό σύνολο. Οι ηχογραφήσεις για την εκμάθηση έγιναν με κινητό τηλέφωνο, ενώ οι ηχογραφήσεις για την βάση για σύγκριση έγιναν με πυκνωτικό μικρόφωνο. Τα χαρακτηριστικά που εξάγουν για τα μοντέλα τους είναι τύπου Mel Frequency Cepstral Coefficients και ο αλγόριθμος Non Negative Matrix Factorization. Οι ήχοι που περιέχονται στο σετ αυτό είναι ήχοι οι οποίοι προκύπτουν από ανθρώπινη δραστηριότητα, όπως *washing hands, typing, brushing teeth* κ.α. Συνολικά περιέχονται 20 κλάσεις ενώ τα γεγονότα φτάνουν σύνολο τα 1367.

DARES Η βάση αυτή [23] χρησιμοποιείται σε δύο σκέλη, με τα οποία καταπιάστηκα σε δύο ξεχωριστές δημοσιευμένες αναφορές (papers).

Στο πρώτο paper [24], σκοπός των ηχογραφήσεών τους, είναι η αξιοποίησή τους σε αυτοματοποιημένη υλοποίηση ετικετοποίησης re-labeling ηχητικών δεδομένων σε βάσεις ηχητικών γεγονότων για την εξάλειψη σημασιολογικά όμοιων ετικετών από παρόμοια ηχητικά σήματα.

Στο δεύτερο paper [25] εξετάζεται μια μέθοδος κατά την οποία συνδυάζεται η ηχητική ομοιότητα ήχων και η σημασιολογική ομοιότητα σε μία μόνο μέτρηση με την δυνατότητα ανάκτησης ηχητικών δεδομένων που είναι παρόμοια σε περιεχόμενο για την σωστή κατηγοριοποίησή τους.

ECS Το αρχείο της βάσης αυτής [26] [27] αποτελείται από δύο σετ. Το πρώτο περιέχει ένα σύνολο 2.000 μικρών ήχων, χωρισμένων σε συνολικά 50 κατηγορίες, και το δεύτερο ένα σύνολο από 250.000 αταξινόμητα αρχεία, όλα από την ιστοσελίδα freesound. Η βάση αυτή αποτελείται από πέντε κατηγορίες όπου η κάθε μία από αυτές περιέχει 10 κλάσεις: συγκεκριμένα animal sound, nature soundscapes and water sounds, human (non-speech) sound, interior/domestic, exterior/urban noises. Στο άρθρο στο οποίο γίνεται η παρουσίαση των σετ αυτών [26], εξετάζεται και η δυνατότητα (ακρίβεια) αξιολόγησης ήχων περιβάλλοντος από τον άνθρωπο και γίνεται σύγκριση σε σχέση με αυτά τα αποτελέσματα που βγαίνουν από το σύστημα ταξινόμησης, το οποίο χρησιμοποιεί χαρακτηριστικά τύπου MFCC που εξάγουν από την βάση και zero-crossing rate αλγόριθμο για την επεξεργασία του σήματός τους.

ELRA Η βάση δεδομένων τους [28] δημιουργήθηκε με σκοπό την προσέγγιση μοντέλου αναγνώρισης ηχητικών γεγονότων τα οποία χαρακτηρίζουν το προσωρινό ηχητικό περιεχόμενο. Η βάση δημιουργήθηκε για να μπορεί να χρησιμοποιηθεί με υλοποίηση εξαγωγής χαρακτηριστικών και χρησιμοποιεί Convolutive non-negative matrix factorization (NMF) μοντέλο χρήσιμο για εύρεση part-base decomposition δεδομένων. Η σύγκριση με την βάση δεδομένων γίνεται με εξαγωγή MFCC χαρακτηριστικών.

Freiburg Η παρούσα βάση δεδομένων [29] δημιουργήθηκε με σκοπό την ανάπτυξη μοντέλου αυτόματης αναγνώρισης βασισμένου στην ιδέα της ηχητικής φράσης, υλοποιώντας την μέθοδο bag of words κατά την οποία υπάρχουν καλύτερα αποτελέσματα της απομονωμένης ηχητικής λέξης για την δημιουργία μιας πιο σημασιολογικής περιγραφής της. Το σετ αυτό είναι βασισμένο σε ήχους από ανθρώπινες δραστηριότητες (ήχοι από μπάνιο, κουζίνα κ.α.).

IEEE Η βάση αυτή [30] δημιουργήθηκε αποκλειστικά στα πλαίσια πρόκλησης διαγωνισμών για σχεδιασμό συστημάτων αναγνώρισης των ηχητικών γεγονότων που παρέχονται σε αυτήν, με σκοπό την όσο το δυνατόν καλύτερη απόδοση.

INRIA Το σύνολο τον ήχων σε αυτήν την βάση [31] ηχογραφήθηκε με το ανθρωποειδές ρομπότ NAO σε φυσικό περιβάλλον. Σκοπός την βάσης αυτής είναι η χρήση της για μοντέλο διαχωρισμού ήχων εσωτερικών χώρων από τον θόρυβο περιβάλλοντος και αναγνώριση των γεγονότων σε αυτό. Οι ήχοι παρουσιάζονται σε φασματοχρονικό πεδίο ορισμού χρησιμοποιώντας stabilized auditory image (SAI).

MIVIA Η παρούσα database [32] δημιουργήθηκε για την δοκιμή του συστήματος αναγνώρισης ηχητικών γεγονότων σε ρεαλιστικές συνθήκες. Το σύστημα, είναι βασισμένο στην προσέγγιση bag of words, και πρέπει να είναι σε θέση να αναγνωρίζει τόσο μικρής όσο και μεγάλης διάρκειας ήχους σε πραγματικές συνθήκες.

QMUL Αυτή η βάση δεδομένων [33] σχεδιάστηκε για χρήση σε μελλοντικές μελέτες και έρευνες σχετικά με data mining σε αρχεία ήχου από ηχογραφήσεις ηχοτοπίων. Οι κατηγορίες που επιλέχθηκαν εδώ είναι οι επικρατέστερες σε ποσοστό μέσα από μία τεράστια βάση η οποία προσπαθεί να καλύψει ένα ευρύ πλαίσιο ήχων.

Sound Ideas and BBC Sound Effects Library Πρόκειται για μια βιβλιοθήκη ήχου [34] [35], η οποία περιέχει ηχητικές πηγές παραγωγής του BBC από το 1990 και μετά, ηχογραφημένες από κορυφαίους μηχανικούς ήχου απ΄ όλον τον κόσμο. Η βιβλιοθήκη αυτή χρησιμοποιήθηκε για αρκετές έρευνες ανάπτυξης μοντέλων μηχανικής μάθησης, όπως της ταξινόμησης κλιπ ήχου με την χρήση μοντέλων παλινδρόμησης (Regression Models) ή για την δημιουργία εποπτευόμενων μοντέλων ακουστικού θέματος για μη-δομημένη ταξινόμηση ήχου.

TUT Η βιβλιοθήκη ήχων αυτή [36] [37] αποτελεί μια έτοιμη βάση για μελέτη ανάπτυξης μοντέλων μηχανικής μάθησης από ερευνητές. Κάθε χρόνο δίδεται μια καινούρια βάση στην δημοσιότητα, με σκοπό την δημιουργία μοντέλων για αυτήν. Η βάση είναι χωρισμένη για τα απαραίτητα σετ και τμήματα για όλη την διαδικασία της εκπαίδευσης και εκτίμησης.

2.1.2 Στατιστικά και Στοιχεία Βάσεων Αναφοράς

Από τα δεδομένα αυτά, τα γεγονότα τα οποία κατηγοριοποιούν οι ερευνητές των οργανισμών που αναφέρθηκαν είναι:

- Dog
- Rooster
- Pig
- Cow
- Frog
- Cat
- Hen
- Insects
- Sheep
- Crow
- Rain
- Sea waves
- Crackling fire
- Crickets
- Chirping bird
- Water drops
- Wind
- Pouring water
- Toilet Flush
- Thunderstorm
- Crying baby
- Sneezing
- Clapping
- Breathing
- Coughing
- Footsteps
- Laughing
- Brushing teeth
- Snoring
- Drinking-sipping
- Door
- Knock
- Mouse click
- Keyboard typing
- Door-wood creaks
- Can opening
- Washing machine
- Vacuum cleaner
- Clock alarm
- Clock tick
- Glass breaking
- Helicopter
- Chainsaw
- Siren
- Car horn
- Engine
- Train
- Church bells
- Airplane
- Fireworks
- Hand saw
- Background
- Food bag opening
- Blender eating
- Flatware sorting
- Food sorting
- Hair dryer
- Microwave plates sorting
- Stirring cup water
- Boiler
- Washing Machine
- Air condition
- Children playing

- Drilling engine
- Idling
- Gun shot
- Jackhammer
- Street music
- Birdsong
- City
- Nature
- People voice
- Basketball
- Beach
- Bus
- Car
- Hallway
- Office
- Restaurant
- Shop
- Street
- Track & Field

Συνολικά το σε ποιες αναφορές χρησιμοποιήθηκαν οι εκάστοτε βάσεις που αναφέρθηκαν, το πλήθος των αρχείων ήχου καθώς και η διάρκειά τους μπορούν να φανούν στον συγκεντρωτικό πίνακα 2.1.

2.2 Βάση για Σύγκριση

Η βάση Σύγκρισης ή Reference Data Set όπως θα αναφέρεται από εδώ και στο εξής, είναι η βάση που επιλέχθηκε για σύγκριση των αποτελεσμάτων σε αυτήν, μέσα στην ίδια διαδικασία με την βάση της εργασίας αυτής. Η παρούσα βάση είναι αναπτυγμένη από το κέντρο επιστημονικής έρευνας CICESE [22] στο Μεξικό, με σκοπό την έρευνα πάνω στην αναγνώριση μέσω τεχνολογιών αιχμής και ηχητικών γεγονότων σε περιβάλλον θορύβου (επικαλυπτόμενα ηχητικά γεγονότα). Το κομμάτι της βάσης αυτής, που θα γίνει σύγκριση με την δική μας, περιέχει τις εξής κατηγορίες (στην παρένθεση βρίσκεται η ισπανική μετάφραση, όπως και συναντάται δηλαδή στην παρούσα βάση που έχει δοθεί στην δημοσιότητα):

- bouncing ball (alon)
- tooth brushing (dientes)
- cricket (grillo)
- crying (llanto)
- keys (llaves)
- washing hands (manos)
- keyboard typing (teclado)

Οι ερευνητές που δημιούργησαν την βάση αυτή χωρίζουν τις κατηγορίες αυτές σε τέσσερα πακέτα, τα λεγόμενα testdataclean, testdatamixed, testdataothers και testdataweb. Για την δική μας εργασία, τα πακέτα που θα μας απασχολήσουν είναι τα δύο πρώτα, στα οποία περιέχονται καθαρά γεγονότα (σκέτα, χωρίς τίποτα άλλο να ακούγεται, μη-επικαλυπτόμενα δηλαδή και χωρίς θόρυβο) που, όπως προαναφέρθηκε, ηχογραφήθηκαν με κινητό τηλέφωνο και επικαλυπτόμενα γεγονότα αντίστοιχα τα οποία αυτήν την φορά ηχογραφήθηκαν με πυκνωτικό μικρόφωνο. Σε κάθε ένα από τα πακέτα αυτά περιέχονται 40 αρχεία από την κάθε κατηγορία γεγονότων, χωρισμένα σε τέσσερεις φακέλους (s1, s2, s3, s4, με τον κάθε έναν από αυτούς να περιέχει με την σειρά του 10 αρχεία ίδιας κατηγορίας).

Πίνακας 2.1: Βάσεις Αναφοράς

Provider	*Name*	*Type*	*Annotation*	*Duration*	*Files*	*Contexts*	*Event Count*
CICESE	Sound Events	Isolated	Tags	92 min	1367	1	1367
Dares	G1	Live	Sound events	123 min	123	28	3214
Dares	Amstel	Live	Sound events	54 min	40	1	1002
ELRA	CHIL 2007 Evaluation Pacage	Live	Sound eventss				
ELRA	RWCP Sound Scene Database	Isolated	Tags				
ELRA	FBK-Irst database of isolated meeting-room acoustic events	Isolated	Tags	63 min	288	1	
ESC	ECS-50	Live	Tags	166 min	2000		2000
ESC	ECS-10	Live	Tags	33 min	400		400
ESC	Dataset for Enviromental Classificaion Sound	Live	Tags	20833 min	250000		2000
Freiburg	Freiburg-106, Audio Data Set for Human Activity Recognition	Live	Tags	54 min	1524	1	
IEEE AASP Challenge Detection	Event isolated	Isolated	Tags	19 min	320	1	639
IEEE AASP Challenge Detection	Event live	Live	Sound events	5 min	3	1	205
IEEE AASP Challenge Detection	Event synthetic	Live	Sound events	14 min	9	1	310
INRIA	NAR	Isolated	Tags	8 min	852	4	42
MIVIA	Audio Event Data Set for Surveillance Application	Live	Sound events	2279 min	760	1	
MIVIA	Audio Event Data Set for Surveillance Application	Live	Sound events	60 min	57	1	
NYU	UrbanSound	Live	Sound events	1620 min	1302	1	3075
NYU	UrbanSound8K	Live	Tags	525 min	8732	1	
QMUL	Freefield1010	Live	Tags	1282 min	7690		7690
Sound Ideas	BBC Sound Effects Library Application	Isolated	Descri ption		1655		1655
TU Dortmund	Acoustic event dataset	Isolated	Sound events	34 min	23	1	235
TUT	CASA 2009	Live	Sound events	1133 min	103	10	10326
TUT	CASA 2010	Live	Sound events	535 min	160	16	4173
TUT	TUT Sound events 2016, Development	Live	Sound events	78 min	22	2	954

Ο σκοπός χρήσης της βάσης αυτής στην εργασία είναι η σύγκρισή της κατά την διαδικασία εκπαίδευση και εκτίμηση με την ίδια μέθοδο, τόσο κατά το ποσοστό ακρίβειας και άλλων μετρήσεων που πραγματοποιήθηκαν και θα αναλυθούν σε επόμενο κεφάλαιο, όσο και στην σχέση μεταξύ καθαρών και μεικτών γεγονότων (όπου κατά πλειοψηφία υπάρχουν στο σενάριο των ηχογραφήσεων στην πόλη του Ρεθύμνου) με τους δικούς μας ταξινομητές. Ο λόγος δηλαδή της χρήσης αυτής της βάσης είναι για εξαγωγή συμπερασμάτων, όχι για την ίδια, αλλά για την διαδικασία των δικών μας ταξινομητών.

2.3 Rpi_Reth

Πρόκειται για ένα σύνολο δεδομένων, το οποίο δημιουργήθηκε αποκλειστικά για τις ανάγκες της δικής μας εργασίας. Σκοπός της δημιουργίας του είναι η ανάδειξη του συνόλου των ηχοτοπίων (σημεία στην πόλη) που μπορεί κανείς να συναντήσει μέσα στην πόλη του Ρεθύμνου, καθώς και των ηχητικών γεγονότων που βρίσκονται σε αυτά. Περιέχει κάποια από τα πιο χαρακτηριστικά σημεία της πόλης (σε θεωρητικό επίπεδο αντιπροσωπεύουν καλύτερα ένα ηχητικό περιβάλλον εκάστοτε ηχοτοπίου), στα οποία και έγιναν τελικά οι ηχογραφήσεις. Αρχικά γίνεται η καταγραφή των σημείων αυτών ως σύνολα από ηχοτοπία, κρατώντας δηλαδή αυτούσια τα αρχεία ήχου όσο αφορά την διάρκειά τους, και στην συνέχεια (όπως θα αναλυθεί αργότερα) γίνεται εξαγωγή από αυτά τα ηχοτοπία, όλων των γεγονότων, μικρότερα αρχεία δηλαδή στο μέγεθος της διάρκειας του γεγονότος, που μπορούν να εντοπιστούν μέσα στις αρχικές ηχογραφήσεις που πραγματοποιήθηκαν. Έτσι, όπως αναφέρθηκε, η βάση χωρίζεται σε δύο μέρη. Το πρώτο, αποτελείται από όλες τις ηχογραφήσεις (ολόκληρες) σε όλες τις κατηγορίες ηχοτοπίων που πραγματοποιήθηκαν, ενώ το δεύτερο περιέχει όλα τα γεγονότα που μπόρεσαν να εξαχθούν από το πρώτο μέρος της βάση που αναφέραμε. Τα γεγονότα αυτά, τα κατηγοριοποιούνται βάσει του ίδιου του γεγονότος και όχι βάσει της κατηγορίας που το εντοπίσαμε, ώστε να έχουμε ολοκληρωμένη εικόνα του ίδιου ήχου σε διαφορετικές κατηγορίες ηχοπεριβάλλοντος, ενώ ταυτόχρονα μεγαλώνει η ίδια η βάση, μιας και κατηγορίες ήχων (όπως π.χ. μια κόρνα) από ένα μόνο ηχοπεριβάλλον, δεν θα ήταν αρκετές σε πλήθος για το στάδιο της εκπαίδευσης που θα ακολουθήσει στην συνέχεια.

Τα γεγονότα χωρίζονται και αυτά με την σειρά τους σε δύο υποσύνολα. Στο πρώτο ανήκουν οι κατηγορίες που δεν προέρχονται από μηχανικής φύσης πηγή, όπως είναι ομιλίες, καμπάνες, κόρνες (λόγω της μορφής τους κατατάσσονται εδώ καλύτερα, ασχέτως του ότι η πλειοψηφία αυτών προέρχεται από μηχανάκια), και άλλα. Αντίστοιχα στο δεύτερο ανήκουν όλες οι κατηγορίες που προέρχονται από μηχανικής φύσεως πηγή, λ.χ. μηχανάκια, αυτοκίνητα, φορτηγά, λεωφορεία και άλλα.

2.3.1 Διαδικασία Ηχογράφησης

Όπως αναφέρθηκε, οι ηχογραφήσεις έγιναν με ένα Raspberry Pi model 2b και ένα Blue Snowflake πυκνωτικό μικρόφωνο ενώ η όλη διαδικασία τρέχει σε υλοποιημένο script με την εντολή arecord από το πακέτο της Alsa, η οποία είναι υπεύθυνη για την ηχογράφηση και μορφοποίηση σε wav αρχείο. Η ψηφιοποίηση του σήματος γίνεται με τον ενσωματωμένο ψηφιακό

μετατροπέα στο ίδιο το μικρόφωνο που χρησιμοποιήθηκε. Το μικρόφωνο είναι πυκνωτικό, λειτουργεί δηλαδή με πυκνωτή [38], ο οποίος, με την φυσική κίνηση που παράγεται από το ακουστικό κύμα το οποίο εισέρχεται στην κάψα του μικροφώνου, δημιουργεί πυκνώματα και αραιώματα στο φορτίο του, δημιουργώντας το αναλογικό ήχο σε μορφή σήματος εν εικόνα του ήχου που εισέρχεται. Τα πυκνωτικά μικρόφωνα χαρακτηρίζονται για την ευαισθησία τους ακόμα και σε ήχους πολύ μικρής στάθμης ηχητικής έντασης, αλλά και για την μη αντοχή τους σε υψηλότερες στάθμες. Το πολικό διάγραμμά τους (το χαρακτηριστικό αυτό το οποίο δηλώνει τον χώρο από τον οποίο λαμβάνεται ο ήχος) είναι καρδιοειδές, δηλαδή έχει πολύ μεγάλη ευαισθησία από μπροστά, λιγότερη στα πλάγια και ελάχιστη στην πίσω μεριά. Τέλος, το μικρόφωνο μπορεί να καταγράψει όλο το ηχητικό φάσμα ανάμεσα σε 35 και $20.000Hz$ (το ακουστό φάσμα του ανθρώπου ανέρχεται από 20 εώς 20.000) Οι ηχογραφήσεις έγιναν με συχνότητα δειγματοληψίας 44100 δειγμάτων το δευτερόλεπτο και με bit depth 16 bit ανά δείγμα. Λόγω λάθους κατά την διαδικασία, οι ηχογραφήσεις έγιναν σε μορφή stereo (δύο κανάλια), ενώ το μικρόφωνο έχει μόνο μία είσοδο (μονοφωνικό), έχει διπλασιαστεί το ένα κανάλι δηλαδή, με αποτέλεσμα τα αρχεία μας να καταλαμβάνουν διπλάσιο όγκο χώρου χωρίς να χρειάζεται, πράγμα που όμως διορθώνεται στο pre processing κομμάτι της υλοποίησης, το οποίο θα αναλυθεί αργότερα σε επόμενο κεφάλαιο. Για τις ηχογραφήσεις αυτές, ρυθμίστηκε η ενίσχυση του μικροφώνου (gain) μέσω των ρυθμίσεων της Alsa στο επιθυμητό επίπεδο, όπου και έμεινε σε όλες τις ηχογραφήσεις και για όλα τα σετ με σκοπό την όσο το δυνατόν μεγαλύτερη σταθερότητα των τεχνικών παραγόντων στην παρούσα διαδικασία, ώστε το μόνο αντικείμενο μελέτης να είναι το ίδιο το περιεχόμενο του ήχου.

Επειδή η συσκευή μας δεν περιέχει οθόνη, αλλά ούτε και κάποια άλλη ενσωματωμένη συσκευή εισόδου, έπρεπε να βρεθεί μια λύση ώστε να μπορεί να γίνεται χρήση της συσκευή αυτής με ευκολία στον εξωτερικό χώρο, όπου και έλαβαν μέρος όλες οι ηχογραφήσεις. Αυτό γίνεται εφικτό μέσω προσθήκης ad hoc network [39] στην ίδια την συσκευή, δημιουργώντας πρακτικά, ένα τοπικό δίκτυο κατ΄ απαίτηση και χρησιμοποιώντας την συσκευή σαν οικοδεσπότη (host), ενώ με την χρήση άλλης συσκευής σαν πελάτη (client) μπορεί να πραγματοποιηθεί σύνδεση στον οικοδεσπότη μέσω πρωτοκόλλου SSH (Secure Shell). Όπως φαίνεται και από το όνομά του, το πρωτόκολλο αυτό επιτρέπει την πραγματοποίηση ασφαλούς σύνδεσης σε άλλη συσκευή, με αποτέλεσμα να υπάρχει η δυνατότητα της απομακρυσμένης διαχείρισής της μέσω γραμμής εντολών. Για να επιτευχθεί αυτό, στο σενάριοσ αυτής της εργασίας χρειάζεται έναν SSH πελάτη για κινητή συσκευή. Συγκεκριμένα, επιλέχθηκε ένας τυχαίος (από πληθώρα που υπάρχει στα διαφόρων ειδών app stores) ονόματι Reflection for UNIX SSH Client της Micro Focus το οποίο κυκλοφορεί τόσο για Adroid όσο και IOS συσκευές. Πρακτικά δηλαδή οι ηχογραφήσεις έγιναν με το μικρόφωνο συνδεμένο Raspberry Pi το οποίο με την σειρά του είναι συνδεμένο σε ένα power bank για την παροχή τροφοδοσίας, και με την χρήση μιας WIFI USB κεραίας για διαδίκτυο γίνεται ο έλεγχος ασύρματα μέσω κινητού τηλεφώνου. Οι ηχογραφήσεις αποθηκεύονται σε εξωτερικό USB flash drive για να μην επηρεάζεται η απόδοση της συσκευής κατά την δέσμευση χώρου σε αυτήν, μιας και ο μόνος αποθηκευτικός χώρος ο οποίος διατίθεται σε αυτήν είναι μία micro SD κάρτα, από την οποία τρέχει και το λειτουργικό σύστημα, πράγμα που απαιτεί αρκετούς πόρους από μόνο του.

Το κάθε ένα αρχείο από αυτά που ηχογραφήθηκαν, παίρνει το όνομά του βάσει του ηχοτοπίου κατά το τρέξιμο της εντολής, ενώ ταυτόχρονα παίρνονται υπόψιν και άλλα στοιχεία, όπως η συχνότητα δειγματοληψίας (όπου όπως αναφέρθηκε είναι σταθερή) και ο χρόνος επιθυμητής διάρκειας ηχογράφησης (αναφέρεται παρακάτω ποιος είναι αυτός για κάθε κατηγορία ηχοτοπίων). Ταυτόχρονα χρειάζεται και η ακριβής θέση στην οποία έγινε η ηχογράφηση του αρχείου αυτού και η τοποθεσία μέσω συντεταγμένων (στοιχεία Latitude και Longitude) για την ονομασία των αρχείων. Για τα δύο τελευταία, κατά το τρέξιμο της εντολής δίνονται οι τιμές 1 και 1 αντίστοιχα μιας και δεν υπάρχει κάποιος τρόπος για την λήψη της πραγματικής πληροφορίας αυτής εκείνη την στιγμή με τον εξοπλισμό που διατίθεται στα πλαίσια της εργασίας. Με το πέρας της ηχογράφησης δημιουργείται, εκτός από το αρχείο wav, και ένα txt αρχείο, το οποίο περιέχει όλα εκείνα τα χαρακτηριστικά που αναφέρθηκαν. Με την βοήθεια λοιπόν του Google Maps, το οποίο περιέχει αυτήν την πληροφορία, συμπληρώνονται στο αρχείο αυτό οι συντεταγμένες της κάθε ηχογράφησης. Τέλος, το αρχείο wav παίρνει το όνομά του από όλα αυτά τα χαρακτηριστικά που αναφέρθηκαν (για ευκολότερη διαχείριση και ομαδοποίηση αργότερα), ενώ ακριβώς το ίδιο όνομα (με διαφορετική προφανώς κατάληξη) παίρνει και το txt αρχείο.

Για το πρακτικό κομμάτι της όλης διαδικασίας, χρειάστηκε η προσοχή κάποιων πραγμάτων κατά την ίδια την ηχογράφηση, όπως το μικρόφωνο να μένει ακίνητο καθ΄ όλη την διάρκεια, μιας και, πρώτον η κίνηση αυτή ακούγεται στο μικρόφωνο και δεύτερον θα μπορούσαν να επηρεαστούν παράγοντες όπως η ένταση, το συχνοτικό φάσμα και η ευαισθησία του μικροφώνου σε αυτό. Επίσης, σε κάθε επανάληψη της ηχογράφησης στον ίδιο χώρο το μικρόφωνο έπρεπε να διατηρείται στην ίδια ακριβώς θέση με την προηγούμενη ηχογράφηση. Για παράδειγμα, σε όλες οι ηχογραφήσεις στον Δημοτικό κήπο και συγκεκριμένα στο σημείο της πλατείας (περισσότερα για αυτό αναφέρονται παρακάτω), το μικρόφωνο έμπαινε κάθε φορά στο ίδιο παγκάκι. Με αυτό τον τρόπο, δημιουργείται μια συνέχεια και μια σταθερότητα στις ηχογραφήσεις και στο ηχητικό υλικό ενώ κρατώντας σταθερή την θέση στον χώρο, μπορούν να μελετηθούν ευκολότερα τις αλλαγές στα ηχητικά φαινόμενα που μας ενδιαφέρουν.

2.3.2 Περιγραφή Ηχοτοπίων και Μορφολογικών Χαρακτηριστικών (βάση Ηχοτοπίων)

Εκτός από την ποικιλία σημείων που θα αναφερθούν στην συνέχεια, οι ηχογραφήσεις χωρίστηκαν σε δύο σετ, με σκοπό το μεγαλύτερο ηχητικό εύρος, λόγω των διαφορετικών ηχητικών συνθηκών στην πόλη. Συγκεκριμένα, υπήρχαν διαφορετικές καιρικές συνθήκες, όπως διαφορετική κίνηση στον δρόμο (τόσο λόγω του καιρού, όσο και του τουρισμού, με τους καλοκαιρινούς μήνες να υπάρχει πολύ μεγαλύτερη κίνηση κόσμου) και διαφορετικοί ήχοι από ζώα (τους καλοκαιρινούς μήνες π.χ. περισσότεροι γρύλοι, τζιτζίκια και τιτιβίσματα από πουλιά, σε σχέση με τον χειμώνα που πιο εύκολα συναντάμε ήχους από σκύλους κ.α.).

Για την δημιουργία της βάσης, τέθηκαν σαν αρχικές κατηγορίες των ηχοτοπίων μας κάποια από τα βασικά σημεία της πόλης του Ρεθύμνου, θέλοντας παράλληλα να γίνει κάλυψη όσο το δυνατόν μεγαλύτερου εύρους του αστικού ηχοτοπίου, για να υπάρχει η σιγουριά ότι οι

ηχογραφήσεις μας αντιπροσωπεύουν σαν σύνολο ολόκληρο το ακουστικό φάσμα της πόλης. Οι κατηγορίες λοιπόν που επιλέχθηκαν είναι οι:

- ***Bus*** (Λεωφορείο)
- ***BusStop*** (Στάση Λεωφορείου)
- ***BusyStreet*** (Πολυσύχναστος Δρόμος - Δρόμος με πολλή κίνηση)
- ***CoastialStreet*** (Παραλιακή Οδός)
- ***OpenMarket*** (Λαϊκή Αγορά)
- ***Park*** (Πάρκο)
- ***Pedestrian*** (Πεζόδρομος)
- ***Port*** (Λιμάνι)
- ***QuietStreet*** (Ήσυχος Δρόμος - Δρόμος χωρίς Κίνηση)

Κατά τον σχεδιασμό των σημείων που θα λάμβαναν χώρα οι ηχογραφήσεις, υπήρχε αρχικά μία ακόμα κατηγορία, η οποία περιείχε αρκετά σημεία στην παραλία, αλλά όλες οι ηχογραφήσεις μας εκεί περιέχουν πολύ περισσότερο θόρυβο απ΄ ότι πληροφορία (χαμηλό signal to noise ratio) και έτσι η συγκεκριμένη κατηγορία καταργήθηκε εντελώς.

Bus

Η κατηγορία περιέχει ηχογραφήσεις από διάφορα δρομολόγια αστικού λεωφορείου διαδρομών 10 - 15 λεπτών κατά την διαδρομή τους Περιβόλια - ΚΤΕΛ. Οι ηχογραφήσεις αυτές έγιναν μέσα στο λεωφορείο με σκοπό την καταγραφή της συμπεριφοράς του ηχητικού φάσματος μέσα στο όχημα, τόσο εν κινήσει του λεωφορείου όσο και στις στάσεις που κάνει κατά την διαδρομή αυτή. Η θέση του μικροφώνου στην κατηγορία αυτή παραμένει σταθερή ώστε να αποτυπωθούν οι ίδιες οι αλλαγές που μπορεί να προκύψουν κατά την διάρκεια της ηχογράφησης σε συγκεκριμένο σημείο. Στις ηχογραφήσεις αυτής της κατηγορίας διακρίνεται το ηχόχρωμα του κινητήρα του λεωφορείου καθώς και άλλοι ήχοι μηχανικού χαρακτήρα, κατά πλειοψηφία όμως τα ακούσματα προέρχονται από τους ανθρώπους που κατακλείνουν το όχημα, με φωνές, ομιλίες, ήχους από την κίνησή τους στον χώρο κ.α.

BusStop

Στην κατηγορία αυτή καταγράφεται ο χώρος της στάσης των λεωφορείων σε διάφορες μεγάλες στάσεις της πόλης. Συγκεκριμένα, οι στάσεις που ηχογραφήθηκαν είναι η στάση προς κέντρο στα Περιβόλια, η στάση προς πανεπιστήμιο στην πλατεία των Τεσσάρων Μαρτύρων και τέλος η στάση προς Περιβόλια της οδού Μοάτσου. Όλες οι ηχογραφήσεις που έγιναν στις περιοχές αυτές, είναι μεγέθους 15 λεπτών και όλες έγιναν σε σταθερό σημείο στην κάθε στάση (για όλες τις ηχογραφήσεις που ακολούθησαν). Το ηχοτοπίο που επικρατεί εδώ είναι τελείως

διαφορετικό από της προηγούμενης κατηγορίας, μιας και κυριαρχεί περισσότερο το χαμηλό συχνοτικό εύρος λόγω των περαστικών οχημάτων (όπως και σε επόμενες κατηγορίες) στον δρόμο. Πέρα από τα οχήματα όμως διακρίνουμε μεγάλη ποικιλία από φωνές και ομιλίες (αν και λόγω της θέσης του μικροφώνου μας δεν καταγράφονται τόσο καθαρά) ενώ παράλληλα έχουμε πολύ μεγάλη ποικιλία από οχήματα, τα οποία, όπως θα δούμε στην συνέχεια, τα αναγνωρίζουμε και αυτά, αφού αποτελούν μέρος στο δεύτερο υποσύνολο που θα αναλυθεί στην συνέχεια. Σε σχέση με την κατηγορία BusyStreet που θα αναλυθεί αμέσως μετά, λόγω του αραιού ωραρίου των λεωφορείων στις περιοχές της πόλης, δεν υπάρχει στις ηχογραφήσεις τόσο μεγάλος όγκος καταγραφής των ίδιων των λεωφορείων μέσα σε αυτήν την διάρκεια των 15 λεπτών, με αποτέλεσμα το μόνο ποιοτικό στοιχείο που ξεχωρίζει την BusStop από την BusyStreet να είναι ο συσσωρευμένος κόσμος που είναι μαζεμένος στις στάσεις των λεωφορείων.

BusyStreet

Η καρηγορία αυτή, όπως και η προηγούμενη, περιέχει ηχογραφήσεις από κομβικά σημεία της πόλης τα οποία θεωρητικά περιέχουν την περισσότερη κίνηση τόσο στον ίδιο τον δρόμο, όσο και από πεζούς. Συγκεκριμένα, τα σημεία αυτά που επιλέχθηκαν στην πόλη του Ρεθύμνου, είναι στο δημαρχείο στην οδό Κουντουριώτου και απέναντι από την στάση στους Τέσσερις Μάρτυρες στην ίδια οδό. Όλες οι ηχογραφήσεις μας αυτής της κατηγορίας είναι επίσης 15 λεπτά και η ηχογράφηση έγινε και αυτή με σταθερό το μικρόφωνο, σε ίδιο σημείο για όλες τις ηχογραφήσεις στον εκάστοτε χώρο. Εδώ ηχητικά περιέχεται μια τεράστια ποικιλία από οχήματα στον δρόμο (τα οποία, όπως αναφέρθηκε και πριν, αξιοποιούνται αργότερα για την δημιουργία της βάσης δεδομένων με τα γεγονότα), ενώ δεν λείπει και ένα αρκετά μεγάλο σύνολο πληροφορίας από ομιλίες (εξίσου όμως όχι τόσο καθαρές). Άλλες κατηγορίες ήχου που υπάρχουν είναι πουλιά, καμπάνες, κόρνες και μουσική. Η κατηγορία αυτή είχε την μεγαλύτερη ποικιλομορφία ως προς το ηχητικό περιεχόμενο, μιας και υπήρχε μεγαλύτερη ηχητική κίνηση σε σχέση με τις άλλες και το ηχοπεριβάλλον άλλαζε συνεχώς. Σε κάποια σημεία των ηχογραφήσεω η πληροφορία ήταν τόσο πυκνή που ήταν πολύ δύσκολο να ξεχωρίσουν οι κλάσεις μεταξύ τους και σε κάποια άλλα, παρότι τα σημεία αυτά χαρακτηρίζουν την συγκεκριμένη κατηγορία, κυρίως λόγω της χαρακτηριστικής έντονης κίνησης πεζών και οχημάτων σε αυτήν, η πληροφορία ήταν ελάχιστη (όμοια με QuietStreet).

CoastialStreet

Για την κατηγορία αυτή, επιλέχθηκαν σημεία παραλιακών οδών του Ρεθύμνου, τα οποία αναμένεται να έχουν το μεγαλύτερο ηχητικό ενδιαφέρον. Συγκεκριμένα, οι ηχογραφήσεις πραγματοποιήθηκαν στην παραλιακή της Καλλιθέας (Οδός Άρη Βελουχιώτη, εκτείνεται μέχρι και Περιβόλια) σε τρία σημεία: ένα απέναντι από την καφετέρια Haagen Dazs, ένα στο σημείο διασταύρωσης με την οδό Σαουνάτσου και τέλος, αρχίζοντας από το λιμάνι, με ηχοπερίπατο μέχρι περίπου το ύψος του ΤΕΙ (15 λεπτά διαδρομή). Στις ηχογραφήσεις αυτές, οι δύο πρώτες έγιναν με το μικρόφωνο σε σταθερή θέση ενώ όπως αναφέρθηκε, η τρίτη ήταν ηχοπερίπατος,

δηλαδή το μικρόφωνο κινείται στον χώρο. Σκοπός της τρίτης ηχογράφησης ήταν η καταγραφή του όσο το δυνατόν μεγαλύτερου ηχητικού χώρου, με την κάλυψη σημείων του ηχοτοπίου της κατηγορίας αυτής που δεν ήταν εφικτό να καλυφθεί με το μικρόφωνο σε σταθερή θέση. Το ηχοτοπίο που συναντήθηκε εδώ περιέχει και αυτό αρκετή πληροφορία χαμηλών συχνοτήτων, λόγω του ότι επίσης υπάρχει μεγάλος αριθμός οχημάτων. Επιπλέον, και εδώ είναι έντονες οι ομιλίες, όπως και άλλοι ανθρώπινοι ήχοι, μιας και ο συγκεκριμένος δρόμος έχει πολύ μεγάλο πεζόδρομο με αρκετούς ανθρώπους να συχνάζουν ειδικά την περίοδο των καλοκαιρινών μηνών όπου έγιναν οι ηχογραφήσεις του ενός από τα δύο σετ ηχογραφήσεων για την βάση. Άλλες κατηγορίες επίσης που συναντώνται εδώ είναι κόρνες καθώς και μουσική.

OpenMarket

Εδώ εμπεριέχονται ηχογραφήσεις από σημεία της πόλης την ώρα που λαμβάνει χώρα η Λαϊκή Αγορά. Συγκεκριμένα αυτά τα σημεία είναι στο πάρκινγκ απέναντι από τον Δημοτικό Κήπο και στο πάρκινγκ στην διασταύρωση Κολοκοτρώνη και Κωστή Παλαμά. Οι ηχογραφήσεις που πραγματοποιήθηκαν εδώ είναι μεγαλύτερης διάρκειας από τις προηγούμενες, για τον λόγο του ότι υπήρχε η αντικειμενική δυσκολία να γίνει επανάληψη της διαδικασίας στην ίδια θέση, μιας και η Λαϊκή Αγορά πραγματοποιείται σε εβδομαδιαία βάση ανά σημείο και θα έπρεπε να υπάρξει αναμονή μιας βδομάδας για επανάληψη ηχογράφησης στην συγκεκριμένη τοποθεσία. Το ηχητικό περιβάλλον που επικρατεί εδώ αποτελείται μόνο από φωνές και ομιλίες κόσμου, ήχοι δηλαδή οι οποίοι υπερκαλύπτουν άλλους ήχους του περιβάλλοντος, κάνοντας την κατηγορία αυτή όχι τόσο χρήσιμη για εξαγωγή περαιτέρω πληροφοριών για το δεύτερο υποσύνολο της βάσης με τα γεγονότα.

Park

Η κατηγορία αυτή περιέχει ηχογραφήσεις από μεγάλα πάρκα της πόλης. Πιο συγκεκριμένα, έχουν πραγματοποιηθεί ηχογραφήσεις στο παρκάκι στην πλατεία Μικρασιατών, στον Δημοτικό Κήπο και στην πλατεία Μητροπόλεως. Στο πρώτο πραγματοποιήθηκαν ηχογραφήσεις σε δύο σημεία: στο πεζουλάκι μέσα στο πάρκο και στο παγκάκι λίγο πιο έξω (αλλά μέσα στην πλατεία). Στο δεύτερο πραγματοποιήθηκαν τρεις: μία στο παγκάκι πριν την κεντρική πλατεία (το τελευταίο όπως μπαίνει κανείς από την κεντρική πόρτα), μία έξω ακριβώς από την καφετέρια και τέλος, μία πίσω από την παιδική χαρά. Τέλος, στο τρίτο είναι από ένα χαρακτηριστικό σημείο, πάλι δηλαδή από ένα παγκάκι προς την πλευρά της οδού Μανιουδάκη. Και εδώ όλες αυτές οι ηχογραφήσεις έγιναν με το μικρόφωνο σε σταθερή θέση για όλα τα σημεία που αναφέρθηκαν. Όλες οι ηχογραφήσεις έχουν διάρκεια 15 λεπτών. Στο ηχοτοπίο αυτό κυριαρχούν περισσότερο από τις άλλες κατηγορίες περιβαλλοντικοί ήχοι (Ambient), μιας και οι υπόλοιποι ήχοι δεν ήταν τόσο συχνοί. Παρόλα αυτά, σε αυτήν την κατηγορία, οι ηχογραφήσεις μας ήταν αρκετές και έγινε εξαγωγή αρκετού υλικού για το δεύτερο υποσύνολο της βάσης με αρκετές κατηγορίες ήχων, όπως καμπάνες πουλιά, λεωφορεία, γρύλους, σκυλιά, κόρνες, μουσική και ομιλίες, αλλά και για το πρώτο υποσύνολο, όπου υπάρχουν ήχοι από μηχανάκια, αμάξια και άλλα οχήματα (κάποια από τα οποία δεν ήταν δυνατόν να προσδιοριστούν ηχητικά για το τι

ακριβώς είναι).

Pedestrian

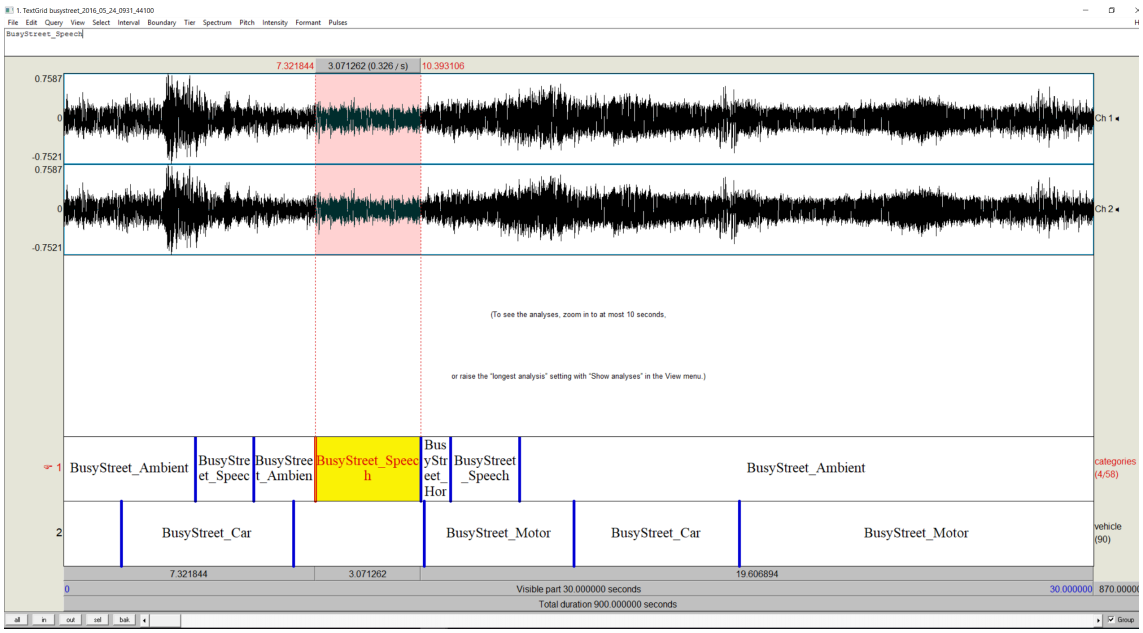
Η κατηγορία περιέχει ηχογραφήσεις από τους πιο πολυσύχναστους πεζόδρομους της πόλης, στην Παλιά Πόλη του Ρεθύμνου (Αρκαδίου και γειτονικά στενάκια). Όλες οι ηχογραφήσεις αυτής της κατηγορίας έγιναν σε μορφή ηχοπερίπατου, ώστε να καλυφθεί όσο το δυνατότερο μεγαλύτερο τμήμα της πόλης και να γίνει καταγραφή όλης αυτής της κίνησης του κόσμου. Στην κατηγορία αυτή δεν απασχολεί εάν το μικρόφωνο είναι σταθερό ή όχι, για το λόγο του ότι, λόγω της μορφής της συγκεκριμένης κατηγορίας θα υπάρχει πολύ περισσότερη πληροφορία εάν οι ηχογραφήσεις μας είναι της μορφής ηχοπεριπάτων. Όλες οι ηχογραφήσεις και εδώ έχουν διάρκεια 15 λεπτών. Όπως είναι αναμενόμενο από ηχητική άποψη, οι περιβαλλοντικοί ήχοι της κατηγορίας μας αποτελούνται από πολλές φωνές, ομιλίες, ήχους από ανθρώπους να κινούνται στον χώρο, ενώ άλλες κατηγορίες ήχων που υπάρχουν εδώ είναι αμάξια και μηχανάκια, καθώς και μουσική (από τα ίδια τα οχήματα ή από μαγαζιά στον δρόμο). Οι ήχοι από οχήματα ήταν πολύ πιο σπάνιοι, μιας και ο δρόμος είναι ανοικτός για τα οχήματα συγκεκριμένες ώρες τις μέρας.

Port

Η κατηγορία περιέχει ηχογραφήσεις από τα δύο λιμάνια της πόλης. Το ένα βρίσκεται στην περιοχή της Καλλιθέας και το άλλο είναι στην Παλιά Πόλη. Οι ηχογραφήσεις που έγιναν ήταν σε 3 διαφορετικά σημεία στο πρώτο (Καλλιθέα): συγκεκριμένα, μία κοντά στην είσοδο (όπως μπαίνει κανείς, δεξιά), μία στα μέσα του λιμανιού και μία στο βάθος, στο τέρμα του λιμανιού, δίπλα σχεδόν στο άγαλμα με τα δελφίνια, ενώ στο δεύτερο σε 2 διαφορετικά σημεία: συγκεκριμένα, μία απέναντι από την καφετέρια και μία στο βάθος. Όλες οι ηχογραφήσεις μας και εδώ έχουν διάρκεια 15 λεπτών. Σχετικά με τον ήχο τώρα, ο περιβάλλοντας ήχος αποτελείται από αρκετό αέρα και ήχο των κυμάτων, ενώ άλλες κατηγορίες που ακούγονται είναι ομιλίες (αν και πιο σπάνιες από τις υπόλοιπες κατηγορίες που αναλύσαμε προηγουμένως), πουλιά, καθώς και κόρνες, αμάξια και μηχανάκια, τα οποία προέρχονται τόσο μέσα από το λιμάνι, όπου και στα δύο υπάρχει μεγάλος χώρος για πάρκινγκ και γενικότερη κίνηση, όσο και έξω από το λιμάνι, όπου ειδικά τα σημεία του πρώτου λιμανιού που βρίσκεται κοντά στην είσοδο είναι πολύ κοντά στην παραλιακή οδό, με αποτέλεσμα το μικρόφωνο να πιάνει αρκετό χώρο και από εκεί.

QuietStreet

Τέλος, αυτή η κατηγορία περιέχει ηχογραφήσεις από δρόμο της πόλης, όπου, σε αντίθεση με την κατηγορία BusyStreet που αναφέρθηκε προηγουμένως, αναμενόταν λιγότερη κίνηση τόσο από πεζούς, όσο και από αυτοκίνητα. Το σημείο ηχογραφήσεων σε αυτήν την κατηγορία είναι μόνο ένα, λόγω της φύσης της κατηγορίας αυτής και των ήχων που ήταν πιθανότερο να υπάρχουν εκεί, με μεγάλη αραίωση δηλαδή, θα αρκούσε μονάχα ένα σημείο. Ο δρόμος αυτός είναι η οδός Κουμουνδούρου, η οποία βρίσκεται πίσω από τον Δημοτικό κήπο. Λόγω της



Σχήμα 2.1: περιβάλλον Praat

ίδιας της φύσης που συναντάμε σε τέτοιου είδους δρόμους, δεν κρίθηκε αναγκαία η προσθήκη περισσοτέρων σημείων μέσα στην πόλη. Η ηχογράφηση και εδώ είναι μεγέθους διάρκειας 15 λεπτών και όλες οι ηχογραφήσεις έγιναν σε σταθερό σημείο. Το ηχοτοπίο εδώ είναι πιο ήσυχο, με κύρια χαρακτηριστικά του να αποτελούν ο αέρας (όταν αυτός υπάρχει) και ήχοι ζώων, όπως γρύλοι (στις βραδινές ηχογραφήσεις), πουλάκια, σκύλοι και άλλα. Παρόλα αυτά, δεν λείπουν ηχητικά γεγονότα όπως οχήματα (αυτοκίνητα και αμάξια) καθώς και κόρνες οχημάτων και ομιλίες.

2.3.3 Διαδικασία κατάτμησης (Praat)

Αφού ακολούθησε όλη η διαδικασία αυτή όπως αναφέρθηκε, έρχεται η σειρά της εξαγωγής πληροφορίας από όλες αυτές τις ηχογραφήσεις των ηχητικών γεγονότων που μπορούν να βρεθούν σε αυτές. Με την βοήθεια ενός ελεύθερου λογισμικού ονόματι Praat [40] σημειώνονται και καταγράφονται κατά μήκος της ηχητικής διάρκειας όλα εκείνα τα γεγονότα που μπορούν να αναγνωριστούν. Η γραμμή αυτή στην οποία σημειώνεται το όνομα του γεγονότος ονομάζεται δεσμίδα (tier) (παράδειγμα τέτοιων δεσμίδων στο σχήμα 2.1). Φορτώνοντας λοιπόν στο λογισμικό τα δικά μας μεγάλα αρχεία των ηχογραφήσεων, γίνεται χρονικό μαρκάρισμα (tag) της πληροφορίας που υπάρχει σε αυτά.

Το Praat είναι ένα ελεύθερο λογισμικό, το οποίο δημιουργήθηκε για επιστημονική ανάλυση ηχητικών σημάτων από τους Paul Boersma και David Weenink από το τμήμα φωνητικής Επιστήμης του Πανεπιστημίου του Amsterdam και αρχικά χρησιμοποιήθηκε για την ανάλυση σημάτων φωνής. Δίνει την δυνατότητα καταγραφής (μαρκάρισμα) γεγονότων στον άξονα του χρόνου πάνω στην εικόνα ενός αρχείου ήχου (αρχικά χρησιμοποιήθηκε για το μαρκάρισμα των χρόνων που μιλούσε ο εκάστοτε ομιλητής), την δεσμίδα που προαναφέραμε και το αποθηκεύει

σε αρχείο τύπου TextGrid. Το λογισμικό αυτό παρέχει και ένα αξιοσημείωτο σύνολο από εργαλεία επεξεργασίας ήχου όπως εξαγωγή χαρακτηριστικών, διαχωρισμός του αρχικού ήχου σε μικρότερα κομμάτια σύμφωνα με το μαρκάρισμα στον χρόνο που προαναφέρθηκε (τμηματοποίηση) και άλλα. Επιπλέον παρέχει τρόπο εισαγωγής script από τον χρήστη για την δυνατότητα αυτοματοποίησης μεγάλων διαδικασιών.

Η μορφή του περιβάλλοντος του προγράμματος, με φορτωμένο δοκιμαστικό ήχο με το αντίστοιχό του TextGrid αρχείο, φαίνεται στο σχήμα 2.1. Όπως φαίνεται, η διαδικασία έγινε σε δύο κανάλια (stereo), πράγμα που δηλώνει πως δεν έχει πραγματοποιηθεί το στάδιο της επεξεργασία ήχου ακόμα. Στο σχήμα επίσης φαίνονται τα δύο layers, όπου εδώ ονομάζονται 1 και 2 αντίστοιχα και αντιπροσωπεύουν τα δύο είδη γεγονότων που υπάρχουν στην βάση, τα Others και Vehs αντίστοιχα.

2.3.4 Σύνολα (Vehs) και (Vehs)

Σε γενική κατεύθυνση, η λογική με την οποία έγινε η τοποθέτηση των γεγονότων (tags) στην χρονική εικόνα του ήχου ήταν η πρόσθεση κάθε ενός γεγονότος που εμφανιζόταν στην παρούσα δεσμίδα (μία μόνο) που αρχικά είχε δημιουργηθεί και λόγω ευκολίας αργότερα στην κοπή και λόγω του ότι τα γεγονότα που θα ακολουθούσαν ήταν άγνωστα (ειδικά στην αρχή). Η αρχική προσέγγιση δηλαδή ήταν ο διαχωρισμός μη-επικαλυπτόμενων γεγονότων, μιας και κάθε tag σταματούσε χρονικά εκεί που ξεκινούσε το άλλο, και αντίστροφα, στην περίπτωση π.χ. των γεγονότων Ambient, έλλειψη δηλαδή κάποιου γεγονότος, με την επικράτηση του περιβάλλοντος ήχου, το μαρκάρισμα της περιοχής χρόνου (tag) ξεκινούσε πάντα εκεί ακριβώς όπου τελείωνε το προηγούμενο. Μετά το πέρας της διαδικασίας εύρεσης των γεγονότων αυτών, όπου πλέον υπάρχει η εικόνα για τα γεγονότα που υπάρχουν στο σύνολο των δεδομένων, έγινε καλύτερος διαχωρισμός των κατηγοριών και έτσι οι κατηγορίες των ηχοτοπίων με δρόμους (BusStop, BusyStreet, CoastialStreet, QuietStreet), προστέθηκαν σε μία ακόμα δεσμίδα, η οποία ήταν δεσμευμένη μόνο για τα οχήματα ενώ στην πρώτη θα παρέμεναν όλα τα υπόλοιπα. Στις κατηγορίες αυτές δηλαδή, κρίθηκε η ανάγκη να γίνει έτσι, λόγω του ότι υπάρχει έντονα το φαινόμενο της επικάλυψης στον ήχο, όταν δηλαδή ένα ηχητικό γεγονός συμπίπτει μαζί με κάποιο άλλο. Η δεύτερη δεσμίδα εξυπηρετεί μεγαλύτερη κάλυψη των ηχητικών γεγονότων ενδιαφέροντος, ακόμα και σε περιπτώσεις που δύο από αυτά συμβαίνουν ακριβώς ταυτόχρονα. Η τελική προσέγγιση δηλαδή που ακολουθήθηκε ήταν ο διαχωρισμός κάποιων επικαλυπτόμενων ηχητικών γεγονότων (και όχι όλων, για ευκολία στην δική μας εργασία, εάν και θα μπορούσε να είχε εφαρμοστεί ξεχωριστή δεσμίδα για κάθε ένα από τα ξεχωριστά ηχητικά γεγονότα, πράγμα όμως που θα μπέρδευε πάρα πολύ τις διαδικασίες).

Κατά την σημείωση των γεγονότων αυτών, ό,τι δεν άνηκε σε κάποια κατηγορία, σημειωνόταν σαν Ambient (όπως αναφέρθηκε προηγουμένως), θεωρώντας δηλαδή πως αυτό το τμήμα ήχου ανήκει στο σύνολο του ηχοπεριβάλλοντος της εκάστοτε κατηγορίας. Αυτή η κίνηση βοηθάει εις διπλούν στα επόμενα στάδια της διαδικασίας. Αρχικά, παρέχεται ευκολία στο στάδιο της εκπαίδευσης, ώστε να προσπεραστούν (προγραμματιστικά) τα αρχεία αυτά, αφού στο επίπεδο των γεγονότων τα τμήματα Ambient θεωρούνται θόρυβος και όχι χρήσιμη πληροφορία

και δεύτερον, κατά το στάδιο της δοκιμής των μοντέλων μας, μπορούν να χρησιμοποιηθούν σαν δοκιμαστικά αρχεία με σκοπό την πρόβλεψη της κατηγορίας ηχοπεριβάλλοντος στην οποία ανήκουν (σε αντίθεση με πριν δηλαδή, στο επίπεδο των ηχοτοπίων, τα Ambient τμήματα θεωρούνται χρήσιμη πληροφορία). Για προγραμματιστικούς λόγους όλα τα γεγονότα (και τα Ambient) αυτών των τμημάτων ήχου τα οποία εξάγουμε από την δεσμίδα του Praat (θα αναλυθεί αργότερα πως), έχουν ονομαστική μορφή 'Soundscape_Event' (όνομα του ηχοτοπίου, κάτω παύλα, όνομα γεγονός, π.χ. Port_Speech).

Αφού ολοκληρωθεί και η διαδικασία αυτή, πρέπει να γίνει η μετατροπή των αρχικών ηχογραφήσεων (μεγάλου μεγέθους που αναφέρθηκαν προηγουμένως) σε τμήματα αρχείων τύπου wav βάσει της διαδικασίας με το Praat που αναλύθηκε. Κάθε μαρκαρισμένο κομμάτι με λίγα λόγια στην δεσμίδα του προγράμματος πρέπει να μετατραπεί σε αυτόνομο κομμάτι ήχου ώστε να μπορεί να χρησιμοποιηθεί στις μετέπειτα διαδικασίες. Για να γίνει αυτό, χρησιμοποιείται ένα πακέτο λογισμικού υλοποιημένο σε Python, το οποίο ονομάζεται praatio [41] και το οποίο αναλαμβάνει με εύκολο τρόπο (αφού είναι σχεδιασμένο να λειτουργεί σε παρόμοια προβλήματα με το δικό μας) όλες τις διαδικασίες διαχείρισης και επεξεργασίας αρχείων τύπου textGrid (τα αρχεία με όλα τα δεδομένα ανά δεσμίδα που δημιουργεί το Praat). Μία από τις συναρτήσεις που περιέχει η βιβλιοθήκη του praatio είναι η αυτόματη κατάτμηση και αποθήκευση σε ανεξάρτητα αρχεία από το αρχικό αρχείο που χρησιμοποιήθηκε, όλων εκείνων των σημείων που έχουν σημειωθεί στην/ις εκάστοτε δεσμίδα/δες.

Λόγω του αρκετά μεγάλου μεγέθους υλικού που υπάρχει διαθέσιμο από τις ηχογραφήσεις, έγινε υλοποίηση προγράμματος σε Python, το οποίο χρησιμοποιεί την βιβλιοθήκη praatio και διαβάζοντας το ψηφιακό μονοπάτι (path) όπου βρίσκονται όλα τα αρχικά αρχεία ηχοπεριβάλλοντος (μεγάλα αρχεία από τις αρχικές ηχογραφήσεις) μαζί με τα textGrid αρχεία τους, τα οποία έχουν τις σημειώσεις που έγιναν με το πρόγραμμα Praat, αυτόματα εξάγει όλα αυτά τα ηχητικά γεγονότα, αποθηκεύοντάς τα στην επιθυμητή μορφή φορμάτ της επιλογής μας (στην εργασία αυτή, βόλευε σε ".wav") και τα οργανώνει σε κατάλληλους φακέλους σύμφωνα με την αντίστοιχη κατηγορία τους. Όλα αυτά τα αρχεία καταλήγουν μέσα στον αντίστοιχο φάκελο με το όνομα της κατηγορίας ηχοπεριβάλλοντος στην οποία ανήκουν. Το τελευταίο γίνεται απλά για ευκολότερη καταγραφή των γεγονότων μας, μιας και στην συνέχεια ενοποιούνται ανά κατηγορία, με το ηχοτοπίο στο οποίο βρέθηκαν να μην παίζει πια κανένα ρόλο. Το στάδιο αυτό όμως, όπου τα αρχεία βρίσκονται στον φάκελο της κατηγορίας ηχοτοπίου όπου βρέθηκαν αρχικά, βολεύει στην στατιστική μελέτη τους για εξαγωγή στατιστικών στοιχείων από αυτά πριν προχωρήσει η όλη διαδικασία.

2.3.5 Στατιστικά και στοιχεία Rpi_Reth

Αναλυτικότερα παρουσιάζονται όλα τα χαρακτηριστικά της βάσης μας στον πίνακα 2.2 που ακολουθεί. Στον πίνακα αυτό φαίνονται όλες οι κατηγορίες ηχοτοπίων (Class), ο αριθμός των αρχείων των ηχοτοπίων που ηχογραφήθηκαν για κάθε κατηγορία (Number of files) από αυτές, η συνολική διάρκεια αυτών (Total Length), η μέση διάρκεια (συνολική δια αριθμό αρχείων, Mean Length), ονομαστικά όλα τα γεγονότα που εξήχθησαν από την συγκεκριμένη

κατηγορία ηχοτοπίων (Events), το σύνολο των τοποθεσιών στις οποίες έγινα ηχογραφήσεις ίδιου ηχοτοπίου (Total Locations) και τέλος, το σύνολο όλων των γεγονότων που εξήχθησαν από το ηχοτοπίο αυτό (Total Files).

Πίνακας 2.2: Rpi_Reth: Σύνολα αρχείων

Class	*Number of files*	*Total Length*	*Mean Length*	*Events*	*Total Locations*	*Total Files*
Bus	3	25	8.3	Ambient, Speech	3	94
Bus_Stop	3	35	11.7	Ambient, Bus, Car, Motor, Speech, Horn	3	219
BusyStreet	8	120	15	Ambient, Bird, Bus, Car, ChurchBell, Horn, Motor, Music, Speech	2	815
CoastalStreet	6	40	6.6	Ambient, Car, Horn, Motor, Speech, Ambient, Music	3	236
OpenMarket	2	25	12.5	Ambient, Car, Horn, Motor, Music, Speech	2	–
Park	16	240	15	Ambient, Bell, Bird, Bus, Car, Churchbell, Cricket, Dog, Horn, Motor, Music, Speech, Vehicle	3	1301
Pedestrian	3	25	8.3	Ambient, Car, Motor, Music, Speech	3	123
Port	7	70	10	Ambient, Bird, Car, Horn, Motor, Speech	1	179
QuietStreet	4	60	15	Ambient, Bird, Car, Cricket, Dog, Horn, Motor, Speech	1	417
Square	3	25	8.3	Ambient, Car, Horn, Motor, Music, Speech	1	39

Κεφάλαιο 3

Στατιστική και Μηχανική Μάθηση

Όπως επισημάνθηκε και στην εισαγωγή της εργασίας, η μηχανική μάθηση είναι ένα πολύ μεγάλο κεφάλαιο, το οποίο στην ουσία συμπεριλαμβάνει τρεις επιπλέον επιστημονικούς τομείς: της επιστήμης των υπολογιστών, της μηχανικής και της στατιστικής. Επίσης, όπως ήδη αναφέρθηκε, υπάρχουν δύο διαφορετικά είδη προβλημάτων μηχανικής μάθησης, με το κάθε ένα από αυτά να απαιτεί διαφορετική μέθοδο αντιμετώπισης. Συγκεκριμένα, οι δύο αντίστοιχες μέθοδοι για αυτά τα προβλήματα είναι η μέθοδος της ταξινόμησης (classification) και η μέθοδος της παλινδρόμησης (regression). Εν συντομία, η πρώτη αναλύει τα δεδομένα και επιστρέφει πιθανότητα για αυτά να ανήκουν στην εκάστοτε κλάση (ποσοστιαία), ενώ, μόλις ολοκληρώσει τον έλεγχο σε όλες τις κλάσεις που εξετάζει, επιστρέφει την επικρατέστερη, αυτή δηλαδή που έχει μεγαλύτερο ποσοστό πιθανότητας. Η δεύτερη αναλύει τα δεδομένα και επιστρέφει από αυτά ένα μαθηματικό αποτέλεσμα. Το πρόβλημα που εξετάζεται στην δική μας εργασία είναι πρόβλημα ταξινόμησης, οπότε αυτή η εργασία θα ασχοληθεί μόνο με αυτό. Στο κεφάλαιο αυτό λοιπόν αναλύονται οι βασικές έννοιες και λειτουργίες κάποιων από τους πιο διαδεδομένους αλγορίθμους μηχανικής μάθησης.

3.1 Τεχνικές Μηχανικής Μάθησης

Σε εφαρμογές αναγνώρισης ήχου (audio classification) συνηθίζεται να χρησιμοποιούνται αλγόριθμοι μηχανικής μάθησης τύπου εποπτευόμενης μάθησης ή αλλιώς μάθησης με επίβλεψη (supervised learning). Το είδος εκπαίδευσης αυτό ονομάζεται έτσι για τον λόγο του ότι χρειάζεται ταμπέλες (π.χ. τα tags που αναφέρονται στην ενότητα 2.3.3) με τα ονόματα των κλάσεων τα οποία καλούνται για πρόβλεψη. Αυτό πρακτικά σημαίνει πως ο αλγόριθμος κάνει την αντιστοίχιση ονόματος κλάσης - χαρακτηριστικών και με βάση αυτήν την σχέση υπολογίζει την πιθανότητα ενός συνόλου χαρακτηριστικών να ανήκει στην κλάση αυτή ή όχι, κρίνοντας ποσοστιαία την πιθανότητα αυτή. Αντίθετα, έχουμε και τους αλγόριθμους unsupervised learning (μη εποπτευόμενη μάθηση ή μάθηση χωρίς επίβλεψη), οι οποίοι χωρίς να έχουν στοιχεία για το τι χαρακτηριστικά περιέχει μια κλάση, προσπαθούν να υπολογίσουν

ομάδες από αυτά τα οποία έχουν κοινά γνωρίσματα και αρχίζουν να δημιουργούν αντίστοιχες σχέσεις μεταξύ των χαρακτηριστικών μόνοι τους. Το πρόβλημα αυτής της εργασίας ανήκει στη μέθοδο της μάθησης με επίβλεψη, καθώς η βάση δεδομένων εμπεριέχει αρχεία ετικετών που ορίζουν στον ταξινομητή κατά την εκπαίδευση ακριβώς ποια χαρακτηριστικά ανήκουν σε ποια κλάση.

3.1.1 Αλγόριθμοι ταξινόμησης

Ο Αλγόριθμος αυτός έχει σχεδιαστεί ώστε να μπορεί να αναλύσει ένα δείγμα προς μελέτη και να επιστρέψει τις πιθανότητες αυτού του δείγματος για το αν ανήκει σε κάθε μία από τις ετικέτες δειγμάτων εκπαίδευσης που έχουν οριστεί σε αυτόν κατά την εκπαίδευση ενώ επιλέγει την ετικέτα αυτή που έχει την μεγαλύτερη πιθανότητα σαν πρόβλεψη.

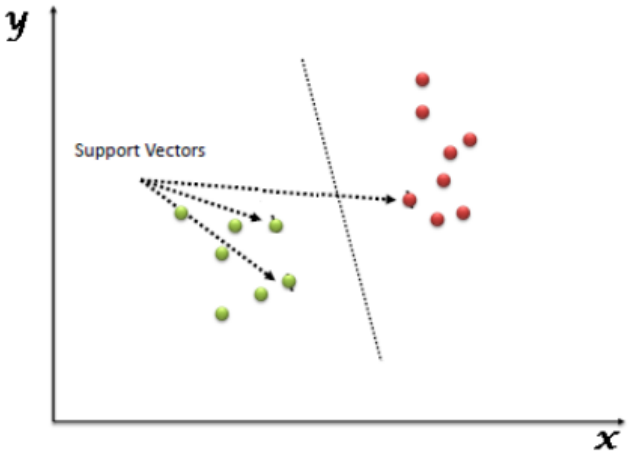
Χαρακτηριστική σημασία στο τελικό αποτέλεσμα της αναγνώρισης των κλάσεων έχει ο ίδιος ο αλγόριθμος ταξινόμησης (Classifier) που θα επιλεχθεί. Αποτελεί τον τρόπο και την τεχνική με την οποία γίνεται η ταξινόμηση μιας κλάσης κρίνοντας στατιστικά το πού έχει περισσότερες πιθανότητες να ανήκει. Ο κάθε ένας από αυτούς δηλαδή έχει διαφορετικό τρόπο με τον οποίο καταλήγει στο τελικό συμπέρασμα. Παρακάτω περιγράφεται η λογική αλγόριθμων οι οποίοι είναι κάποιοι από τους πιο συνήθεις για εφαρμογές μηχανικής μάθησης σε ήχο.

SVM

Support Vector Machine [42] [43], ένας αλγόριθμος μηχανικής μάθησης ο οποίος μπορεί να χρησιμοποιηθεί τόσο σε μεθόδους ταξινόμησης όσο και σε μεθόδους παλινδρόμησης με πιο διαδεδομένη μέθοδο όμως την πρώτη. Στον αλγόριθμο αυτό γίνεται σχεδίαση όλων των δεδομένων σε σημεία σε χώρο n-διαστάσεων (όπου n αριθμός αντίστοιχων των χαρακτηριστικών, τα οποία προσομοιάζονται σε διαστάσεις). Η ταξινόμηση γίνεται με βάση το βέλτιστο διάνυσμα (ή τα βέλτιστα διανύσματα εάν ο αλγόριθμος εξετάζει πάνω από 2 κλάσεις) το οποίο διαφοροποιεί καλύτερα τις δύο (ή περισσότερες) αυτές κλάσεις, όπως φαίνεται στο σχήμα 3.1. Τα πλεονεκτήματά του είναι:

- Μεγάλη αποτελεσματικότητα σε χώρους μεγάλης διάστασης
- Αποτελεσματικότητα σε περιπτώσεις στις οποίες ο αριθμός των διαστάσεων είναι μεγαλύτερος από τον αριθμό των δειγμάτων
- Χρησιμοποιεί ένα υποσύνολο σημείων κατάρτισης κατά την λειτουργία λήψης αποφάσεων (support vector – φορέας υποστήριξης), πράγμα που σημαίνει πως είναι πιο αποδοτικός ως προς την μνήμη
- Μεγάλη ευελιξία: μπορούν να οριστούν διαφορετικές λειτουργίες πυρήνα (SVM-Kernels) για τη λειτουργία λήψης αποφάσεων. Παρέχονται απλοί πυρήνες (γραμμικός πυρήνας, πολυωνυμικός, ακτινικές συναρτήσεις βάσης κ.α.) αλλά υπάρχει και η δυνατότητα καθορισμού προσαρμοσμένων πυρήνων.

Τα μειονεκτήματα όμως του συγκεκριμένου αλγορίθμου είναι:



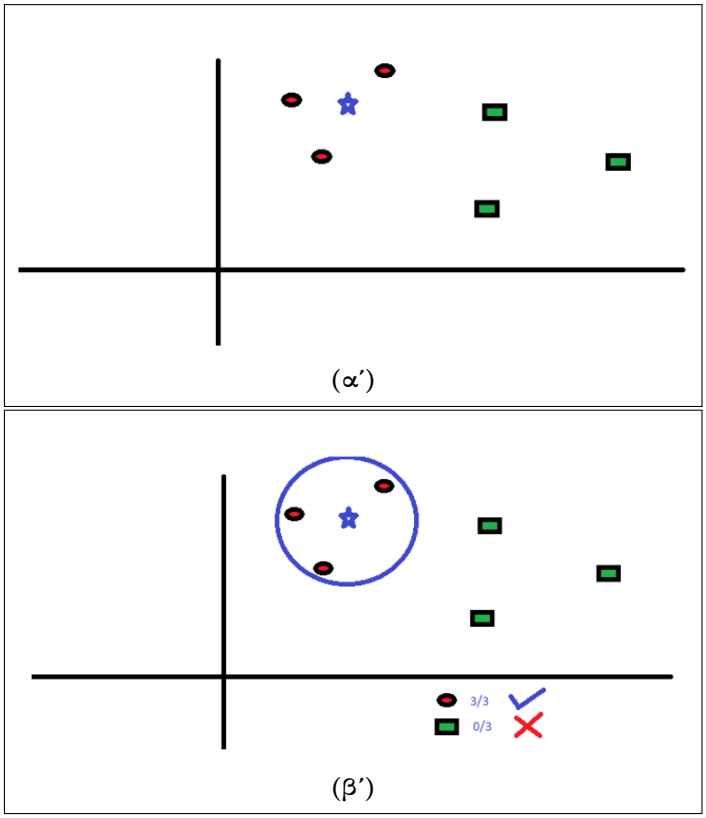
Σχήμα 3.1: Διαχωρισμός χαρακτηριστικών με SVM

- Υπάρχει η πιθανότητα ο αλγόριθμος να παρουσιάσει χαμηλές επιδόσεις εάν ο αριθμός των χαρακτηριστικών είναι πολύ μεγαλύτερος από τον αριθμό των δειγμάτων (παρατηρήσεων)

- Δεν παρέχει απευθείας εκτιμήσεις πιθανοτήτων. Αντιθέτως υπολογίζονται με την χρήση μιας πάρα πολύ δαπανηρής σε ισχύ διαδικασίας πενταπλής διασταυρωμένης επικύρωσης.

KNN

(K – Nearest Neighbors) [44] [45] είναι ένας από τους πιο απλούς αλγορίθμους αλλά παρόλα αυτά παρέχει εξαιρετική ακρίβεια στις προβλέψεις του. Βασίζεται στην ίδια τεχνική χωρικής σχεδίασης δεδομένων όπως και ο SVM (σχήμα 3.2 α΄) με την διαφορά ότι αντί να πραγματοποιεί χρήση διανύσματος για τον χωρισμό των κατηγοριών, χρησιμοποιεί έναν παράγοντα K βάσει του οποίου επιλέγονται τα K γειτονικά δείγματα και υπολογίζονται τα σύνορα της γειτονικής περιοχής των ίδιων χαρακτηριστικών, κατηγοριοποιώντας τα σαν σύνολο (της ίδιας δηλαδή κατηγορίας, σχήμα 3.2 β΄) και ξεχωρίζοντάς τα από τα υπόλοιπα. Όσο μεγαλύτερη είναι αυτή η τιμή τόσο πιο ομαλά είναι τα σύνορα των περιοχών που δημιουργεί ο ταξινομητής μεγαλώνει όμως η πιθανότητα λάθους ενώ αντίθετα εάν η τιμή του παράγοντα ισούται με $K = 1$ τότε το λάθος των εκπαιδευμένων δεδομένων τείνει στο μηδέν λόγω του ότι το ένα πιο κοντινό στοιχείο του κάθε δείγματος είναι πάντα ο εαυτός του. Σε αυτήν την περίπτωση, η βέλτιστη K τιμή είναι πάντα το ένα. Σε πραγματικά σενάρια όμως αυτό δεν ισχύει. Η τεχνική αυτή ονομάζεται μη-γενικευμένη μηχανική μάθηση λόγω του ότι ο αλγόριθμος «θυμάται» όλα τα δείγματα. Όπως και ο SVM, έτσι και ο KNN μπορεί να χρησιμοποιηθεί και σε προβλήματα ταξινόμησης και σε προβλήματα παλινδρόμησης ενώ συγκεκριμένα στην αναγνώριση ήχου έχει συνήθως την υψηλότερη απόδοση επιτυχίας.



Σχήμα 3.2: Διαχωρισμός δεδομένων με KNN

Random Forest

Ο αλγόριθμος αυτός [46] ανήκει στην οικογένεια των αλγορίθμων τύπου "δέντρο" λόγω της χαρακτηριστικής του διακλαδωτής ροής των βαθμίδων αποφάσεων. Διαμορφώνει ένα δέντρο από δοκιμαστικούς κόμβους με την χρήση ενός αντιγράφου εκκίνησης του δείγματος εκμάθησης και τον αλγόριθμο CART (Classification And Regression Tree) μαζί με την τροποποίηση που χρησιμοποιείται στην μέθοδο τυχαίου υποσυνόλου (Random Subspace). Μοιάζει αρκετά με αντίστοιχους αλγορίθμους των δέντρων αποφάσεων (decision tree) και ταξινομητών σάκων (Bagging Classifier), μιας και μοιράζονται τις ίδιες παραμέτρους. Ο συγκεκριμένος αλγόριθμος παράγει τυχαίες διακλαδώσεις, των οποίων τα αποτελέσματα συγχωνεύει με σκοπό την δημιουργία μιας πιο σταθερής και ακριβέστερης πρόβλεψης.

Gradient Boosting

Ο αλγόριθμος αυτός [47] βασίζεται στην απλή αρχή της αδύναμης μάθησης, αναπτύσσοντας περισσότερες αδύναμες μαθήσεις με στόχο την καλύτερη αντιμετώπιση ενός πολύ δύσκολου προβλήματος. Ο Gradient Boosting αλγόριθμος αποτελείται από τρία στοιχεία: την συνάρτηση απώλειας η οποία τον βελτιστοποιεί, μια αδύναμη μάθηση η οποία κάνει την πρόβλεψη και ένα επιπρόσθετο μοντέλο για την προσθήκη αδύναμων μαθήσεων για την ελαχιστοποίηση

της συνάρτησης απώλειας. Η απώλεια αυτή ορίζεται ως εξής:

$$Loss = \sum (y_i - {y_i}^p)^2$$

όπου, y_i η i-οστή στοχευμένη τιμή, ${y_i}^p$ η αντίστοιχη πρόβλεψη και τέλος $L(y_i, {y_i}^p)$ η συνάρτηση απώλειας. Αναλόγως το πρόβλημα μπορεί να χρησιμοποιηθεί και διαφορετική συνάρτηση απώλειας, αρκεί να είναι διαφοροποιήσιμη. Για παράδειγμα μπορεί να χρησιμοποιηθεί τετραγωνικό λάθος για την μέθοδο της παλινδρόμησης ενώ μπορεί να χρησιμοποιηθεί λογαριθμική απώλεια στην μέθοδο ταξινόμησης. Για την αδύναμη μάθηση χρησιμοποιούνται δέντρα αποφάσεων, ένα σύνολο από κομβικά σημεία δηλαδή, τα οποία εξάγουν πραγματικές τιμές για τον διαχωρισμό και οι έξοδοι τους μπορούν να προστεθούν μεταξύ τους. Με τον τρόπο αυτόν συμμετέχουν πολλοί αδύναμοι κόμβοι στην διαδικασία για το αποτέλεσμα και εξαλείφουν αρκετά λάθη μιας και διορθώνονται τυχών υπολείμματα στις προβλέψεις πάνω στην ίδια την διαδικασία. Συνηθίζεται να υπάρχουν περιορισμοί σε αυτήν την διαδικασία, όπως π.χ. τις στρώσεις των κόμβων, τους κόμβους σε κάθε στρώση κ.α. Τέλος, λόγω του ότι τα δέντρα αυτά προσθέτονται και μετά παραμένουν στάσιμα, χωρίς να αλλάζουν τιμή δηλαδή, υπάρχει η διαδικασία καθοδικής διαβάθμισης σαν επιπρόσθετο μοντέλο το οποίο ελέγχει την έξοδο από τα παλαιότερα δέντρα και την ενημερώνει για να ελαχιστοποιήσει το ποσοστό λάθους. Τα πλεονεκτήματα αυτού του αλγορίθμου είναι η ταχύτητα και η απόδοση που έχει σε σχέση με άλλους, με μειονέκτημα όμως την υπερβολική επεξεργαστική ισχύ που ζητάει.

Extra Trees

Ο αλγόριθμος αυτός [48] είναι τελείως διαφορετικός από τους αλγορίθμους απόφασης ως προς τον τρόπο κατασκευής του. Πρόκειται για έναν αλγόριθμο ο οποίος παράγει ένα σύνολο από παλινδρομικά δέντρα, όπως και ο Random Forest και άλλοι αλγόριθμοι τύπου "δέντρο", με την διαφορά όμως ότι χωρίζει τους κόμβους επιλέγοντας εντελώς στην τύχη, ενώ χρησιμοποιεί την όλη διαδικασία εκμάθησης για να μεγαλώσει το δέντρο βάσει των διαχωρισμών των κόμβων που αποδίδουν καλύτερα. Περιέχει δύο παραμέτρους: τον συντελεστή K, ο οποίος υπολογίζεται τυχαία στον κάθε κόμβο και τον αριθμό n_{min} , το ελάχιστο μέγεθος του δείγματος για τον διαχωρισμό του κόμβου.

Βαθιά Νευρονικά Δίκτυα (deep neural network) – Βαθιά Κατανόηση (Deep learning)

Ο αλγόριθμος αυτός [49] [50] παίρνει το όνομά του από τους νευρώνες του ανθρώπινου εγκεφάλου, τους οποίους προσπαθεί να προσομοιώσει. Το μοντέλο αυτό προτάθηκε από τους H. Hubel και Torsten Wiesel το 1959, οι οποίοι το ανέπτυξαν, μελετώντας τους εγκεφαλικούς νευρώνες, οι οποίοι είναι συνδεδεμένοι μεταξύ τους με δεσμούς, πράγμα που τους επιτρέπει την ανταλλαγή πληροφορίας μέσω ηλεκτρισμού και έτσι καθιστά ικανή την μεταξύ τους επικοινωνία και αυτό με την σειρά του την λήψη αποφάσεων και την αποστολή εντολών στο υπόλοιπο σώμα. Έτσι λοιπόν, θέλοντας ο αλγόριθμος να κάνει ακριβώς το ίδιο, αναπτύσσει σύστημα κόμβων διαιρεμένο σε συνδεδεμένα επίπεδα. Το χαρακτηριστικό αυτού του αλγορίθμου είναι η

συνεχής βελτίωσή του κατά την χρήση του, λόγω της ικανότητάς του να προσθέτει καινούρια χαρακτηριστικά, τα οποία δεν είχε συναντήσει παλιότερα και να διαμορφώνει ακόμα καλύτερη εικόνα στα πρότυπα και μοτίβα που έχει αναγνωρίσει. Τα νευρωνικά δίκτυα αυτομάτως ανακαλύπτουν και επεκτείνουν κανόνες με τους οποίους τμηματοποιούν χαρακτηριστικά, ώστε να μάθουν περισσότερα για τις κλάσεις τις που εξερευνούν και να ανεβάσουν την απόδοση ακρίβειας κατά την πρόβλεψη άγνωστης εισόδου. Επειδή όμως το κεφάλαιο των νευρωνικών δικτύων δεν εξετάζεται σαν πιθανή λύση του προβλήματος αυτής της εργασίας λόγω της περιπλοκότητάς του, (βγάλε το κόμμα) δεν θα αναλυθεί περαιτέρω.

3.1.2 Αλγόριθμοι ομαδοποίησης

Εκτός από τους αλγόριθμους τμηματοποίησης, όπου, όπως αναφέρθηκε, η δουλειά τους είναι να δημιουργούν τμήματα από χαρακτηριστικά με κοινά στοιχεία, επιστρέφοντας ποσοστό πιθανότητας, έχουμε και τους αλγόριθμους ομαδοποίησης, οι οποίοι αυτό που κάνουν είναι να ομαδοποιούν με τον καλύτερο δυνατό τρόπο σύνολα χαρακτηριστικών, δημιουργώντας συσπειρώσεις αυτών των συνόλων (clustering), από τις οποίες μπορεί να παρθεί κάποια μέση τιμή και να χρησιμοποιηθεί πλέον αυτή σαν ένα σημείο (παρατήρηση), πράγμα που απλουστεύει π.χ. μετέπειτα διαδικασίες ταξινομητών κ.λ.π. Οι αλγόριθμοι αυτοί ονομάζονται και αλγόριθμοι μείωσης διαστάσεων (Dimensionality Reduction), αφού απλοποιούν μεγάλο σύνολο δεδομένων (πολλές διαστάσεις) σε ένα σημείο (μία διάσταση). Το πλεονέκτημα αυτής της μεθόδου είναι πως μπορεί να χρησιμοποιηθεί για unsupervised training, εκπαίδευση δηλαδή στην οποία δεν έχουμε συγκεκριμένες ταμπέλες σαν κλάσεις. Ένας συνήθης τρόπος αναγνώρισης είναι ένας συνδυασμός αλγορίθμων ομαδοποίησης των χαρακτηριστικών και με τα κεντρικά σημεία που παράγονται από την διαδικασία αυτή να χρησιμοποιούνται σαν είσοδος στον αλγόριθμο ταξινόμησης. Παρακάτω θα αναλυθούν οι δύο πιο διαδεδομένοι αλγόριθμοι clustering.

K-means Clustering : Ο τύπος αυτός της ομαδοποίησης [51]έχει σαν στόχο να χωρίσει σε ομάδες (clusters) τα δεδομένα εισόδου, με τις ομάδες αυτές να είναι ανάλογες του αριθμού συντελεστή K. Κάθε στοιχείο της εισόδου λοιπόν, έπειτα από αρκετές επαναλήψεις της διαδικασίας με σκοπό το όσο καλύτερο μοίρασμα , ταυτίζεται με μια από τις ομάδες αυτές. Στο τέλος κάθε επανάληψης ελέγχεται εάν τα κεντρικά σημεία (centroids) είναι όντως τα κεντρικά σημεία όλων των γειτονικών δεδομένων ή αν με αυτά τα σημεία προκύπτουν άλλα καινούρια γειτονικά δεδομένα στον περίγυρό του. Η διαδικασία επαναλαμβάνεται στην περίπτωση που τα δεδομένα αναδιαμορφώνονται και υπολογίζονται ξανά τα νέα centroids. Η διαδικασία τελειώνει όταν πλέον δεν υπάρχουν άλλες αλλαγές και τα centroids παραμένουν σταθερά για τα ίδια γειτονικά δεδομένα. Στο τέλος της διαδικασίας αυτής ο αλγόριθμος βγάζει σαν έξοδο τα centroids των ομάδων αυτών καθώς και την ταμπέλα, η οποία δηλώνει σε ποιο cluster ανήκει το κάθε σημείο.

X-mean Clustering : Ο αλγόριθμος αυτός [52] αποτελεί βελτιωμένη έκδοση του K-mean αλγορίθμου, για τον λόγο ότι βγάζει την ίδια έξοδο αλλά πιο γρήγορα. Σε αντίθεση με τον

προηγούμενο λοιπόν αλγόριθμο, αυτός έχει την δυνατότητα να κάνει καλύτερη υπολογιστική κλιμάκωση, να υπολογίζει μόνος του τον βέλτιστο παράγοντα K και να μην είναι επιρρεπής σε τοπικές ελάχιστες τιμές.

3.2 Στατιστικές Μετρήσεις Μηχανικής Μάθησης

Οι διαδικασίες αυτές αποτελούν κρίσιμα σημεία για την μηχανική μάθηση. Καθορίζουν τις αρχικές συνθήκες πάνω στις οποίες κάθε μοντέλο μηχανικής μάθησης δημιουργείται. Με τις μετρήσεις αυτές ορίζεται στο μοντέλο τι τιμές (στατιστικά πάντα) περιμένει σαν είσοδο, προσαρμόζοντάς το καταλλήλως.

Πλήθος (Population)

Η τιμή αυτή είναι η ολότητα, ο πλήρης κατάλογος παρατηρήσεων ή όλα τα δεδομένα σχετικά με το υπό μελέτη θέμα.

Δείγμα (Sample)

Η τιμή του δείγματος είναι ένα υποσύνολο του πληθυσμού προς ανάλυση.

Διαφορές μεταξύ Παραμέτρων και Στατιστικών Στοιχείων (Parameter - statistic)

Κάθε μέτρηση που υπολογίζεται από τον πληθυσμό αποτελεί μια παράμετρο ενώ αντίθετα σε ένα δείγμα ονομάζεται στατιστικό στοιχείο.

Μέση Τιμή (Mean)

Πρόκειται για έναν απλό αριθμητικό μέσο όρο, ο οποίος υπολογίζεται λαμβάνοντας το συνολικό άθροισμα των τιμών, διαιρούμενο με το σύνολο αυτών των τιμών. Ο μέσος όρος είναι ευαίσθητος στις αποκλίσεις στα δεδομένα. Η απόκλιση είναι η τιμή ενός συνόλου ή στήλης που είναι εξαιρετικά αποκλίνουσα (πολύ υψηλότερη ή χαμηλότερη τιμή) από το υπόλοιπο σύνολο τιμών στα ίδια δεδομένα.

Median

Αποτελεί το μέσο σημείο των δεδομένων και υπολογίζεται είτε με την οργάνωση της σε αύξουσα είτε σε φθίνουσα σειρά.

Mode

Αποτελεί το πιο επαναλαμβανόμενο σημείο δεδομένων στα συνολικά δεδομένα.

Μετρήσεις Παραλαγής (Measure of variation)

Ή αλλιώς διασπορά, είναι η παραλλαγή των δεδομένων και μετρά τις ασυνέπειες στις τιμές των μεταβλητών στα δεδομένα. Η διασπορά προσομοιάζει την κατάσταση για την εξάπλωση και όχι για τις κεντρικές τιμές (central values).

Εύρος (Range)

Η διαφορά μεταξύ του μέγιστου και του ελάχιστου της τιμής.

Διακύμανση και Τυπική Απόκληση (Variance - Standard Deviation)

Αποτελούν τις μετρήσεις διάδοσης των δεδομένων στο σύνολό τους. Η διακύμανση είναι ο μέσος όρος των τετραγωνικών διαφορών από τον μέσο όρο. Μαθηματικά υπολογίζεται:

$$\sigma^2 = \frac{1}{N}\sum (X - \mu)^2 \tag{3.1}$$

Όπου σ^2 η διακύμανση, N ο αριθμός των παρατηρήσεων, X το ατομικό σύνολο παρατηρήσεων και τέλος μ ο μέσος όρος. Η διακύμανση συμβολίζεται με σ^2 (σίγμα στο τετράγωνο), λόγω του ότι με σ συμβολίζεται η τυπική απόκλιση. Η τυπική απόκλιση δηλαδή υπολογίζεται σαν τετραγωνική ρίζα της τιμής της διακύμανσης:

$$StandardDeviation = \sqrt{Variance} \tag{3.2}$$

Δηλαδή στην ουσία $\sigma = \sqrt{\sigma^2}$.

3.3 Μετρήσεις Ταξινομητών

Καθώς ο σκοπός ενός μοντέλου ταξινομητή [5] είναι η όσο καλύτερη απόδοση γίνεται, εφαρμόζονται και οι αντίστοιχες μετρήσεις, όπου καθορίζουν την απόδοση αυτή καθώς και τον υπολογισμό της πιθανότητας λάθους. Ο τρόπος με τον οποίο γίνεται αυτό είναι ότι η έξοδος του ταξινομητή είναι σε μορφή διακριτών αριθμών, με αποτέλεσμα να μπορεί να οριστεί ακριβώς αν μια πρόβλεψη έχει αποτύχει ή όχι σε δυαδικό αριθμό. Κατά την δοκιμή ενός δείγματος ο ταξινομητής δοκιμάζει μία - μία τις ετικέτες για να εντοπίσει αν το δείγμα ανήκει σε αυτές ή όχι. Εάν εντοπίσει πως ανήκει, τότε το δείγμα ορίζεται σαν ***θετικό*** ενώ σε αντίθετη περίπτωση το δείγμα ορίζεται σαν ***αρνητικό***. Η πρόβλεψη αυτή μπορεί να είναι είτε σωστή, άρα και χαρακτηρίζεται ως ***αληθής***, είτε λανθασμένη, άρα χαρακτηρίζεται σαν ***ψευδής***. Αυτό οδηγεί στους παρακάτω ορισμούς, οι οποίοι θα αναλυθούν στην συνέχεια.

3.3.1 Ακρίβεια (Accuracy)

Η ακρίβεια υπολογίζεται είτε με το κλάσμα είτε με το άθροισμα των αποτελεσμάτων του μοντέλου. Σε ταξινομητές πολλαπλών κατηγοριών το κλάσμα επιστρέφει την τελική ακρίβεια του υποσυστήματος.

Εάν ολόκληρο το σύνολο ετικετών που προβλέφθηκαν για ένα δείγμα ταιριάζει αυστηρά με το πραγματικό σύνολο των ετικετών, τότε η ακρίβεια του υποσυνόλου ισούται με 1.0 ενώ σε αντίθετη περίπτωση ισούται με 0.0. Εάν y η τιμή πρόβλεψης του i δείγματος και $\hat{y}$ η αντίστοιχη αληθινή τιμή το κλάσμα της σωστής πρόβλεψης σε n αριθμό δειγμάτων ορίζεται ως εξής:

$$accuracy(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=1}^{n_{samples}-1} 1(y_i = \hat{y}_i) \tag{3.3}$$

3.3.2 Βαθμός Ακρίβειας (Precision score)

Ο βαθμός ακρίβειας ορίζεται ως εξής:

$$precision = \frac{t_p}{t_p + f_p} \tag{3.4}$$

Όπου εδώ t_p ο αριθμός των αληθινών θετικών προβλέψεων, των προβλέψεων δηλαδή μιας ετικέτας που σωστά επισημάνθηκαν σαν την συγκεκριμένη, ενώ αντίθετα ο f_p είναι ο αριθμός των λανθασμένων θετικών προβλέψεων, των προβλέψεων δηλαδή ενός δείγματος σαν της συγκεκριμένης ετικέτας, χωρίς όμως το δείγμα μας να ανήκει σε αυτήν. Ο βαθμός ακρίβειας δηλαδή είναι η ικανότητα του ταξινομητή να μην επιστρέφει σαν θετική πιθανότητα ένα δείγμα το οποίο είναι αρνητικό και οι τιμές που παίρνει ορίζονται από 1.0 έως 0.0, με την πρώτη να είναι η θετικότερη.

3.3.3 Ποσοστό Ανάκλησης (Recall rate)

Αντίστοιχα, το ποσοστό ανάκλησης ορίζεται ως εξής:

$$recall = \frac{t_p}{t_p + f_n} \tag{3.5}$$

Όπου εδώ t_p ο αριθμός των αληθινών θετικών προβλέψεων και f_n ο αριθμός των ψευδών αρνητικών προβλέψεων, των προβλέψεων δηλαδή που δεν επισημάνθηκαν ως την συγκεκριμένη ετικέτα, παρότι το δείγμα προς ανάλυση άνηκε σε αυτήν. Το ποσοστό ανάκλησης δηλαδή είναι η ικανότητα του ταξινομητή να εντοπίζει σωστά όλα τα θετικά δείγματα. Αντίστοιχα, οι τιμές του ευρύνονται από 1.0 έως 0.0, με το πρώτο την θετικότερη τιμή.

3.3.4 Μέτρηση F (F-measure)

Η μέτρηση αυτή (μετρήσεις F_β και F_1) μπορεί να οριστεί ως ένα ειδικό είδος μέσης τιμής (αρμονικού μέσου βάρους) των τιμών βαθμού ακριβείας και ποσοστού ανάκλησης. Αντίστοιχα με τις άλλες δύο μετρήσεις, η καλύτερη τιμή που μπορεί να επιστρέψει είναι 1.0 ενώ η χειρότερη 0.0. Εάν οριστεί η τιμή $\beta = 1$, τότε οι μετρήσεις F_β και F_1 είναι ισοδύναμες και οι τιμές του βαθμού ακριβείας και του ποσοστού ανάκλησης αποκτούν την ίδια βαρύτητα για την τιμή της μέτρησης F. Η μέτρηση F υπολογίζεται ως εξής:

$$F_\beta = (1 + \beta^2)\frac{precision * recall}{(\beta^2 * precision) + recall} \tag{3.6}$$

3.3.5 Πίνακας Σύγχυσης (Confusion matrix)

Κάθε πράξη ταξινόμησης με μηχανική μάθηση έχει σκοπό την πρόβλεψη ετικετών σε νέα άγνωστα δεδομένα. Ο πιο λειτουργικός τρόπος για τη παρουσίαση της ακρίβειας των προβλέψεων του ταξινομητή είναι μέσω του πίνακα σύγχυσης, όπου φαίνονται τα ζευγάρια ταξινομημένων δειγμάτων και η κατάσταση αλήθειας τους με λεπτομερή προβολή του τι συμβαίνει με τις προβλέψεις.

Ο πίνακας περιέχει δύο άξονες με τις ετικέτες αριθμημένες και στους δύο. Στον κάθετο εμφανίζονται οι κλάσεις των δειγμάτων και στον οριζόντιο οι κλάσεις στις οποίες ταξινομήθηκαν τα δείγματα αυτά. Στην κύρια διαγώνιο παρουσιάζεται ο αριθμός των αληθινών θετικών προβλέψεων, όπου για κάθε μία από αυτές προστίθεται ο βαθμός 1.0.

3.4 Επιλογή Χαρακτηριστικών

Η επιλογή χαρακτηριστικών (Feature Selection ή αλλιώς Attribute Selection), [53] [54] είναι μια διαδικασία κατά την οποία γίνεται η συλλογή των πιο χρήσιμων χαρακτηριστικών από αυτών που μπορούν να εξαγχθούν. Μετά από επαναλαμβανόμενες δοκιμές επιλέγονται τα χαρακτηριστικά εκείνα τα οποία είναι πιο επιρρεπή στην σωστή πρόβλεψη κατά την ταξινόμηση. Τα επιλεγμένα χαρακτηριστικά αυτής της διαδικασίας μπορούν να χρησιμοποιηθούν για την δημιουργία μοντέλων ταξινομητών, τα οποία είναι πιο στοχευμένα στην διαδικασία των κλάσεων, ενώ παράλληλα μειώνεται και το κόστος επεξεργαστικής ισχύος, λόγω της σημαντικής μείωσης των υπολογισμών για χαρακτηριστικά ανά δείγμα. Η διαδικασία της επιλογής χαρακτηριστικών είναι διαφορετική από την διαδικασία μείωσης διαστάσεων. Και οι δύο διαδικασίες προσπαθούν να μειώσουν των αριθμό των σημείων (χαρακτηριστικών) που χρησιμοποιούνται για την διαδικασία, αλλά η διαφορά τους είναι ότι η πρώτη επιλέγει τα υπάρχοντα χαρακτηριστικά εκείνα από ένα σύνολο, τα οποία στατιστικά θα αποδώσουν καλύτερα στην διαδικασία, ενώ η δεύτερη δημιουργεί καινούρια, κάθε ένα από τα οποία εκπροσωπεί ένα μεγαλύτερο σύνολο από τα πρώην υπάρχοντα χαρακτηριστικά.

3.4.1 Μέθοδοι Επιλογής Χαρακτηριστικών

Υπάρχουν τρεις μέθοδοι με τις οποίες μπορεί να πραγματοποιηθεί αυτή η διαδικασία, filter method, wrapper method και embedded method. Κάθε μία από αυτές διαφέρει στον τρόπο με τον οποίο επιτυγχάνει αυτήν την επιλογή και οι τρόποι αυτοί περιγράφονται παρακάτω.

Filter Method

Η επιλογή χαρακτηριστικών με φίλτρα εφαρμόζει στατιστικές μετρήσεις σχετικά με την απόδοση του κάθε χαρακτηριστικού (μετράει το σκορ του). Τα χαρακτηριστικά κατατάσσονται σύμφωνα με το σκορ τους και είτε επιλέγονται για να κρατηθούν είτε αφαιρούνται από το σύνολο της βάσης. Η μέθοδος φίλτρου είναι συχνά μονομερής και εξετάζει το χαρακτηριστικό είτε ανεξάρτητα είτε σε σχέση με την εξαρτημένη τιμή μεταβλητής. Μερικά από τα παραδείγματα υπολογισμού της μεθόδου φίλτρων περιλαμβάνουν την δοκιμή τετραγωνικού

χ (χ^2), κέρδος πληροφοριών (Information Gain) και βαθμολογίες συντελεστών συσχέτισης (Correlation Coefficient Scores).

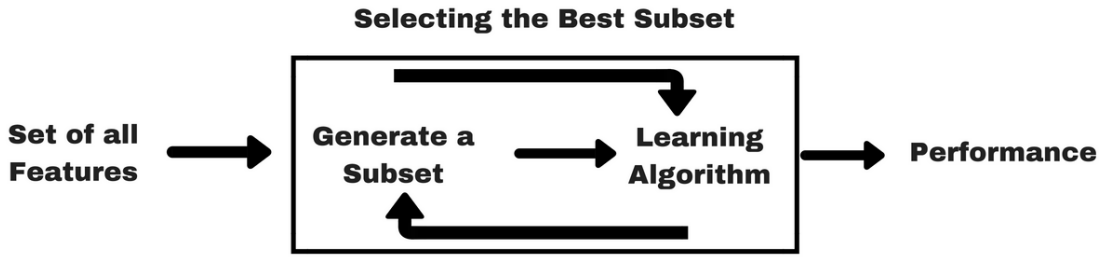
Σχήμα 3.3: Επιλογή χαρακτηριστικών με Μέθοδο Φίλτρου

Wrapper Method

Μέθοδος Περιτύλιξης ονομάζονται οι αλγόριθμοι, οι οποίοι εφαρμόζουν επιλογή ενός σετ χαρακτηριστικών βάσει ενός προβλήματος αναζήτησης, στο οποίο προετοιμάζονται διαφορετικοί συνδυασμοί, εκτιμούνται και συγκρίνονται με άλλους αντίστοιχους συνδυασμούς. Αποτελεί δηλαδή προγνωστικό μοντέλο, το οποίο μπορούμε να χρησιμοποιούμε για να αξιολογήσουμε χαρακτηριστικά και να τους καθορίσουμε ένα σκορ βασισμένο στην ακρίβεια του μοντέλου.

Η διαδικασία της αναζήτησης διαφέρει αναλόγως την μορφή της. Έτσι, υπάρχει η μεθοδική διαδικασία (best-first search), η οποία προσεγγίζει το πρόβλημα διευρύνοντας τον πιθανότερο κόμβο, ο οποίος επιλέγεται σύμφωνα με έναν συγκεκριμένο κανόνα. Η στοχαστική διαδικασία (random hill-climbing algorithm), η οποία ξεκινά με μια αυθαίρετη λύση στο πρόβλημα, προσπαθεί να βρει μια καλύτερη λύση κάνοντας μια σταδιακή αλλαγή στη προηγούμενη. Αν η αλλαγή αυτή παράγει μια καλύτερη λύση, μια άλλη βαθμιαία αλλαγή γίνεται στη νέα λύση και ούτω καθεξής έως ότου δεν υπάρχουν περαιτέρω βελτιώσεις. Τέλος, η ευρετική διαδικασία (heuristics), λειτουργεί με μεγαλύτερη τυχαιότητα, όπως για παράδειγμα με ανακάτεμα των χαρακτηριστικών μπρος - πίσω για προσθήκη και αφαίρεση χαρακτηριστικών.

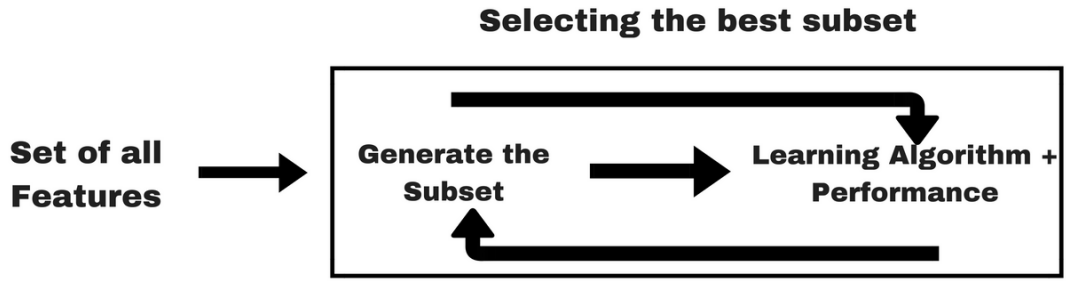
Παράδειγμα αλγορίθμου Περιτύλιξης είναι ο Recursive Feature Elimination (RFE), ο οποίος καθορίζει τιμές βάρους των χαρακτηριστικών βάσει ενός εξωτερικού εκτιμητή, με σκοπό να πετάξει σταδιακά τις τιμές με το χαμηλότερο σκορ μέχρις ότου το σετ των χαρακτηριστικών να γίνει αρκετά μικρό.



Σχήμα 3.4: Επιλογή χαρακτηριστικών με Μέθοδο Περιτύλιξης

Embedded Method

Με την μέθοδο της Ενσωμάτωσης ο αλγόριθμος μαθαίνει ποια χαρακτηριστικά συμβάλλουν καλύτερα στην ακρίβεια του μοντέλου όταν το μοντέλο δημιουργείται. Ο πιο συνηθισμένος τύπος μεθόδου ενσωμάτωσης για την επιλογή χαρακτηριστικών είναι η μέθοδος κανονικοποίησης (Regularization Methods). Η μέθοδος αυτή ονομάζεται επίσης Penalization method (Μεθόδων Επιβολής Κυρώσεων), λόγω του ότι επιβάλλει περιορισμούς στην βελτιστοποίηση ενός αλγορίθμου πρόβλεψης, που οδηγούν το μοντέλο προς την κατώτερη πολυπλοκότητα (λιγότεροι συντελεστές). Παραδείγματα τέτοιου είδους αλγορίθμων είναι οι αλγόριθμοι LASSO, Elastic Net και Ridge Regression.



Σχήμα 3.5: Επιλογή χαρακτηριστικών με Μέθοδο Ενσωμάτωσης

3.4.2 Πρόβλημα Υπερφόρτωσης Χαρακτηριστικών (Overfitting)

Η λανθασμένη επιλογή χαρακτηριστικών μπορεί να οδηγήσει στο φαινόμενο της υπερφόρτωσης Overfitting). Το φαινόμενο αυτό παρουσιάζεται όταν παρατηρείται απότομη μείωση της ακρίβειας ενός μοντέλου σε άγνωστα δεδομένα σε σχέση με γνωστά (σε δεδομένα δηλαδή που έχει εκπαιδευτεί). Το πρόβλημα αυτό δημιουργείται όταν έχουμε δημιουργήσει τόσο σύνθετο το μοντέλο μας, με πάρα πολύ μεγάλη ανάλυση στον διαχωρισμό των χαρακτηριστικών του δηλαδή, με αποτέλεσμα να προβλέπει άριστα τα δεδομένα πάνω στα οποία έχει εκπαιδευτεί, αλλά σε άγνωστα δεδομένα να μην έχει την απαιτούμενη ανοχή στα χαρακτηριστικά που ξεφεύγουν ή είναι στα όρια, με αποτέλεσμα την λάθος πρόβλεψη. Το πρόβλημα δηλαδή προκύπτει λόγω του ότι δεν έχουν διαχωριστεί μόνο τα χρήσιμα χαρακτηριστικά τα οποία θα αντιπροσώπευαν την κλάση αλλά και ο ίδιος ο θόρυβος του σήματος που περιέχεται σε αυτήν. Η αντίθετη περίπτωση είναι όταν συμβαίνει το ανάποδο και υπάρχει τόσο χαμηλή ανάλυση στον διαχωρισμό, με αποτέλεσμα να μειώνεται η πιθανότητα σωστής πρόβλεψης, αλλά παρόλα αυτά να παρατηρείται μεγαλύτερη ανοχή του ταξινομητή σε άγνωστα δεδομένα.

Κεφάλαιο 4

Επεξεργασία σήματος

Στο κεφάλαιο αυτό γίνεται η περιγραφή τρόπων δημιουργίας των χαρακτηριστικών, καθώς και όλη η διαδικασία για την επεξεργασία σήματος· στην δική μας περίπτωση δηλαδή η εξαγωγή από ένα ψηφιακό σήμα ήχου χρήσιμης πληροφορίας, σύμφωνα με την οποία γίνεται η ταξινόμηση του σήματος αυτού ως προς το σε ποια κλάση ανήκει.

Παρόλο που κατά την πάροδο των χρόνων η τεχνολογία έχει αυξηθεί και στο κομμάτι της μηχανικής μάθησης υπάρχουν πλέον περισσότερες δυνατότητες με πιο πλούσια γνώση, έχουν επικρατήσει κάποιες μέθοδοι και τεχνικές για δημιουργία χαρακτηριστικών ως επικρατέστερες λόγω της καλύτερης απόδοσης που έχουν (αναλόγως πάντα το είδος δεδομένων μελέτης). Το είδος των δεδομένων δηλαδή προς αναγνώριση είναι αυτό που θα καθορίσει και τι είδους χαρακτηριστικά θα εξαχθούν από το σήμα. Παρακάτω θα αναφερθούν κάποιες από τις πιο σύγχρονες μεθόδους εξαγωγής της πληροφορίας από ηχητικό σήμα, οι οποίες συγκεκριμένα χρησιμοποιούνται στην περίπτωση της ταξινόμησης ηχητικών αποσπασμάτων και γεγονότων, καθώς και άλλων τύπου κλάσεων ηχητικού σήματος γενικότερα.

4.1 Ακουστικά Χαρακτηριστικά

Με τον όρο χαρακτηριστικά ονομάζουμε τα στοιχεία αυτά τα οποία παρατηρούμε μετά από τακτές λήψεις δειγμάτων ενός συνόλου δεδομένων[1] και όπως φαίνεται και από το όνομά τους «χαρακτηρίζουν» το σύνολο αυτό. Τα χαρακτηριστικά αυτά πρέπει να είναι μετρήσιμα και υπολογίζονται μετά από εύρεση κρυφών μοτίβων που μπορεί να δημιουργούνται στο σύνολο των παρατηρήσεων. Τα χαρακτηριστικά επίσης μπορεί να είναι απλά υπολογίσιμα, όπως η θεμελιώδης συχνότητα του σήματος, η ηχητική στάθμη έντασης κ.α, ή πιο σύνθετα, όπως τα χαρακτηριστικά τύπου MFCC που θα αναλυθούν στην συνέχεια. Στην περίπτωση αυτής της εργασίας, το σύνολο δεδομένων που παρατηρείται είναι το σήμα του ήχου· ψηφιακές τιμές δηλαδή, οι οποίες αναπαριστούν το πλάτος έντασης (A) του σήματος την συγκεκριμένη χρονική στιγμή.

[1] Σε αυτό το κεφάλαιο σύνολο δεδομένων ονομάζεται το αντικείμενο προς μελέτη, το σύνολο δηλαδή όλων των δειγμάτων και των χαρακτηριστικών που αυτά περιέχουν και είναι διαφορετικό από τον όρο ***σύνολο δεδομένων*** όπως χρησιμοποιήθηκε στο κεφάλαιο 2, όπου οριζόταν ως το σύνολο των αρχείων

Για την εξαγωγή πληροφορίας από ένα σύνολο δεδομένων λοιπόν (σε αυτήν την περίπτωση ήχου) χρειάζεται να βρεθεί το στοιχείο, το οποίο χαρακτηρίζει το σύνολο αυτό αυτό (στην περίπτωσή μας τις κλάσεις) που υπάρχει επιθυμία για ταξινόμηση. Με την πληροφορία αυτή μπορεί να γίνει η κατάταξη της εισόδου σε κλάσεις βάσει του περιεχόμενου της. Τα στοιχεία αυτά που το κάνουν εφικτό ονομάζονται χαρακτηριστικά (Features, ακουστικά χαρακτηριστικά στην περίπτωσή μας). Με την χρήση των χαρακτηριστικών γίνεται λιγότερη κατανάλωση υπολογιστικής ισχύος για την σύγκριση και επεξεργασία τους σε σχέση την ίδια διαδικασία που θα χρειαζόταν για επεξεργασία απευθείας στο ακουστικό σήμα και ταυτόχρονα χρειάζονται πολύ λιγότερο αποθηκευτικό χώρο για την καταγραφή τους.

4.1.1 Κατηγορίες Ακουστικών Χαρακτηριστικών

Τα χαρακτηριστικά αυτά χωρίζονται σε κατηγορίες ανάλογα με τον τύπο πληροφορίας που εξάγουν. Συγκεκριμένα, στον ήχο γίνεται η εξαγωγή για τα εξής είδη χαρακτηριστικών που ακολουθούν.

Φασματικά Χαρακτηριστικά

Τα χαρακτηριστικά αυτά εξετάζουν το σήμα ως προς το φάσμα του και εξάγουν πληροφορία σχετικά με το συχνοτικό περιεχόμενο του σήματος. Τέτοια χαρακτηριστικά είναι:

- ***MFCC*** (Mel-Frequency Cepstral Coefficients): Αναπαράσταση βραχυπρόθεσμου δυναμικού φάσματος του σήματος βάσει γραμμικού μετασχηματισμού συνημιτόνου.
- ***Spectral Centroid***: το κέντρο βάρους της ενέργειας του συχνοτικού φάσματος του σήματος, παρουσιάζοντας την συχνοτική περιοχή στην οποία συσσωρεύεται η ενέργεια (χαμηλή, μεσαία, υψηλή).
- ***Spectral Spread***: Βαθμός επέκτασης του συχνοτικού φάσματος, το δεύτερο κεντρικό στιγμιότυπο του φάσματος.
- ***Spectral Flux***: ή αλλιώς η τετραγωνική διαφορά των κανονικοποιημένων μεγεθών φάσματος δύο διαδοχικών στιγμιότυπων/παρατηρήσεων (frames).
- ***Rolloff***: Η συχνότητα κάτω από την οποία συγκεντρώνεται το 90% της κατανομής μεγέθους του φάσματος.
- ***Relative Position of Spectral Max & Minimum***: Μέγιστες και ελάχιστες σχετικές θέσεις φάσματος.

Τονικά Χαρακτηριστικά και Χαρακτηριστικά Ρυθμού

Τα χαρακτηριστικά αυτά αναφέρονται σε περιπτώσεις μουσικής φύσεως, μιας και εξειδικεύονται στην ανάλυση μουσικής, εξετάζοντάς την ως προς τους μουσικούς τόνους και τον ρυθμό.

- ***Chroma***: Εξάγουν μουσικής φύσεως πληροφορία (τόνους, νότες).
- ***Beat-related, Tempo-related***: Χαρακτηριστικά ρυθμού.

Χαρακτηριστικά Κατανομής της ενέργειας

Χαρακτηριστικά που εξετάζουν το τρόπο κατανομής της ενέργειας στο σήμα.

- ***Energy***: Άθροισμα των τετραγώνων των τιμών του σήματος
- ***Entropy of Energy***: Άθροισμα τιμών εντροπίας των υποπλαισίων (sub-frame) της ενέργειας.
- ***Spectral Entropy***: Εντροπία της κανονικοποιημένης φασματικής ενέργειας συνόλου από sub-frames

Voicing Χαρακτηριστικά

Τα χαρακτηριστικά αυτά εξετάζουν το σήμα ως προς το συχνοτικό περιεχόμενο.

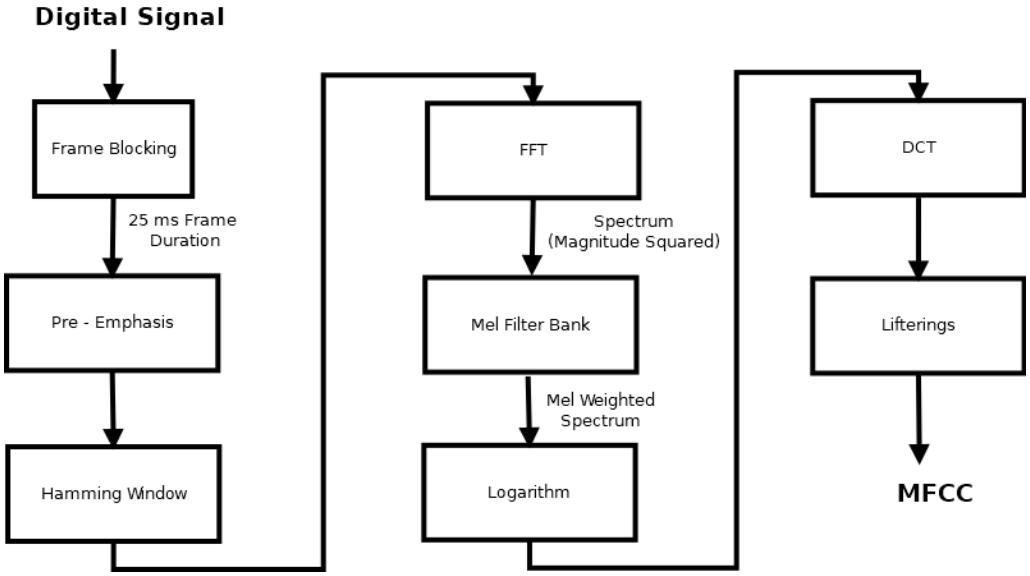
- ***Fundamental Frequency Analysis – F0***: Θεμέλιος συχνότητα
- ***Probability of Voicing***: Μέτρηση αρμονικότητας σε περίοδο σήματος

Πιο συγκεκριμένα, θα αναλυθούν κάποια από τα πιο διαδεδομένα από τα χαρακτηριστικά που συνηθίζεται να εξάγονται σε εφαρμογές ηχητικών ταξινομήσεων.

4.1.2 MFCC Χαρακτηριστικά

Τα Mel-Frequency Cepstral Coefficients [15] χαρακτηριστικά είναι τα πιο διαδεδομένα όσο αφορά την ηχητική αναγνώριση, μιας και προσεγγίζουν τον τρόπο λειτουργίας του ανθρώπινου ακουστικού συστήματος ως προς την ευαισθησία φασματικών περιοχών. Αρχικά τα MFCC χρησιμοποιήθηκαν για την αυτόματη αναγνώριση φωνής, αλλά τα χαρακτηριστικά αυτά βρίσκουν χρησιμότητα σε όλες τις εφαρμογές αναγνώρισης ήχου λόγω της φύσης του σχεδιασμού τους. Η λογική τους είναι ότι λαμβάνουν υπόψιν την μη-γραμμική φύση της αντίληψης των θεμελιωδών συχνοτήτων καθώς και την μη-γραμμική σχέση μεταξύ έντασης και ηχηρότητας, όπως ακριβώς και ο ανθρώπινος εγκέφαλος. Ο υπολογισμός τους περιέχει αρκετά βήματα, τα οποία παρουσιάζονται στο σχήμα 4.1.

Εν συντομία για την διαδικασία εξαγωγής: αρχικά, γίνεται η προέμφαση του σήματος και ακολουθεί η μετατροπή σε πλαίσια (Windowing) διάρκειας συνήθως 20 εώς $40ms$ (το παράδειγμα του διαγράμματος έχει διάρκεια $25ms$) μέσω της χρήσης του φίλτρου Hamming. Στην συνέχεια, κάθε πλαίσιο υποβάλλεται σε γρήγορο μετασχηματισμό Φουριέ διακριτού χρόνου (υπολογισμός περιοδογράμματος, το οποίο δείχνει την φασματική πυκνότητα) και περνάει από μια ομάδα τριγωνικών ζωνοδιαβατών φίλτρων (συχνοτικά φίλτρα Μελ, τα αναλύουμε παρακάτω). Τέλος, η κάθε έξοδος των φίλτρων συμπιέζεται λογαριθμικά και υπολογίζεται πάνω σε αυτές ένας μικρός αριθμός από συνιστώσες.



Σχήμα 4.1: Διαδικασία εξαγωγής MFCC χαρακτηριστικών

Κλίμακα Μελ και Filterbanks

Η κλίμακα Mel (Μελ) [15] είναι σχεδιασμένη με τέτοιο τρόπο ώστε να προσομοιάζει τον τρόπο με τον οποίο ο άνθρωπος αντιλαμβάνεται το συχνοτικό εύρος ως προς την τονική αλλαγή, λόγω του οι τονικές μεταβολές είναι πιο έντονες στις χαμηλές συχνότητες απ' ότι στις υψηλότερες. Η μετατροπή από συχνότητα (Hertz) σε κλίμακα Mel μπορεί να υπολογιστεί με τον τύπο:

$$M(f) = 1125ln(1 + \frac{f}{700}) \tag{4.1}$$

Αντίστροφα, η μετατροπή από κλίμακα Μελ σε συχνότητα υπολογίζεται:

$$M^{-1}(f) = 700^{\frac{m}{1125}-1} \tag{4.2}$$

Για τον υπολογισμό της τράπεζας φίλτρων Μελ (Mel filterbank) επιλέγονται αρχικά τα συχνοτικά άκρα, η κατώτερη δηλαδή και η ανώτερη συχνότητα των φίλτρων (σε Hertz). Για την επιλογή αυτή λαμβάνεται υπόψιν το θεώρημα του Νίκουιστ [55], το οποίο ορίζει πως η συχνότητα ανάλυσης του σήματος πρέπει να είναι τουλάχιστον η διπλάσια από την μέγιστη συχνότητα του σήματος προς ανάλυση ($f_{Nyquist} = 1/2v$, όπου v η μεγαλύτερη συχνότητα της κυματομορφής). Στην συνέχεια, οι δύο συχνότητες αυτές μετατρέπονται σε κλίμακα Μελ. Βάσει αυτών των δύο υπολογίζονται και τα υπόλοιπα (συνήθως το σύνολο των filterbanks είναι 26 – 40) διαιρώντας την απόσταση της κατώτερης – ανώτερης Μελ συχνότητας σε ίσα μέρη). Αφού υπολογιστούν όλες οι απαιτούμενες Μελ συχνότητες, μετατρέπονται ξανά σε συχνότητες (Hertz) και στρογγυλοποιούνται στο κοντινότερο FFT σημείο μέσω του τύπου:

$$f(i) = floor\left((nfft + 1) * \frac{h(i)}{samplerate} \right) \tag{4.3}$$

Όπου h η συχνότητα σε Hertz, samplerate η συχνότητα δειγματοληψίας του σήματος και nfft το μέγεθος (σημείων) του FFT. Τέλος, τα φίλτρα υπολογίζονται με βάση:

$$H_m(k) = \begin{cases} 0, & k < f(m-1), k > f(m+1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)}, & f(m-1) \geq k \geq f(m) \\ \frac{f(m+1)-k}{f(m+1-f(m)}, & f(m) \deg k \deg f(+1) \end{cases} \tag{4.4}$$

Υπολογισμός MFCC

Ένα ηχητικό σήμα [15] συνεχώς αλλάζει στην διάρκεια του χρόνου. Ένας τρόπος λοιπόν να μελετηθεί είναι η κατάτμησή του σε τόσο μικρά πλαίσια, ώστε να μπορεί να θεωρηθεί πως δεν υπάρχει μεγάλη στατιστική μεταβολή σε αυτό. Τα πλαίσια αυτά έχουν διάρκεια συνήθως 20 με $40ms$ λόγω του ότι μέσα σε αυτόν τον χρόνο περιέχονται αρκετά δείγματα για αξιόπιστη εκτίμηση και ταυτόχρονα το σήμα δεν μεταβάλλεται αρκετά σε αυτόν τον χρόνο. Στην συνέχεια, υπολογίζεται το εκτιμώμενο περιοδόγραμμα του δυναμικού φάσματος του κάθε πλαισίου. Το στάδιο αυτό είναι εμπνευσμένο από τον κοχλία του αυτιού ο οποίος δονείται σε διαφορετικά σημεία αναλόγως την συχνότητα του εισερχόμενου ήχου. Συγκεκριμένα, κατά την είσοδο του ήχου στον κοχλία του αυτιού δημιουργείται κίνηση σε συγκεκριμένα τριχίδια τα οποία βρίσκονται μέσα σε αυτόν σε σημείο ανάλογο με το μήκος κύματος της συχνότητας του ήχου που εισέρχεται και τα οποία τριχίδια ενεργοποιούν το αντίστοιχο νεύρο με το οποίο δίνουν σήμα στον εγκέφαλο. Το περιοδόγραμμα προσέγγισης λειτουργεί με παρόμοιο τρόπο, αναγνωρίζοντας ποιες συχνότητες είναι παρούσες κατά το συγκεκριμένο πλαίσιο ήχου.

Στην συνέχεια αποβάλλεται όλη η περιττή πληροφορία χρησιμοποιώντας μια σειρά από φίλτρα Μελ (Mel filterbank). Το πρώτο φίλτρο είναι πολύ στενό και δείχνει το ποσοστό ενέργειας γύρω από τα $0Hz$. Όσο ανεβαίνει η συχνότητα, μεγαλώνει και το πλάτος του φίλτρου για τον λόγο του ότι μικραίνει η ανάγκη για ακρίβεια καθώς στις υψηλότερες συχνότητες έχουμε πιο έντονο το χαρακτηριστικό της συχνοτικής επικάλυψης, της δυσκολίας δηλαδή να αντιληφθούμε διαφορές μεταξύ δύο πολύ κοντινών συχνοτήτων (για αυτόν τον λόγο το ανθρώπινο αυτί παρουσιάζει μη-γραμμική αντίληψη ήχου και δυσκολεύεται να διαχωρίσει τις υψηλότερες συχνότητες). Έτσι υπάρχει ενδιαφέρον μόνο για την προκύπτουσα ποσότητα ενέργειας στην κάθε συχνοτική μπάντα φίλτρου. Ο τρόπος υπολογισμού πλάτους, πλήθους καθώς και απόστασης των φίλτρων δίνεται από την κλίμακα Mel. Μετά από αυτό, το επόμενο βήμα είναι ο υπολογισμός των λογαρίθμων από τα φίλτρα. Αυτό επίσης προσομοιάζει τον τρόπο με τον οποίο λειτουργεί το ανθρώπινο αυτί, μιας και δεν αντιλαμβανόμαστε τις συχνοτικές μεταβολές στάθμης έντασης γραμμικά, αλλά λογαριθμικά λόγω της ευαισθησίας του ανθρώπινου αυτιού σε συγκεκριμένα σημεία του συχνοτικού φάσματος. Με την διαδικασία αυτή, το μοντέλο του ταξινομητή μπορεί πολύ πιο εύκολα να πλησιάσει την ανθρώπινη ακοή με την χρήση των χαρακτηριστικών αυτών, ειδικά σε περιπτώσεις εφαρμογών αναγνώρισης ομιλίας, μιας και το ακουστό φάσμα σχετίζεται άμεσα με το φάσμα της ομιλίας. Παράλληλα, ένα επιπλέον θετικό χαρακτηριστικό της λογαρίθμησης αυτής είναι ότι η λογαριθμική κλίμακα επιτρέπει αφαίρεση μέσης τιμής αντίστροφου μετασχηματισμού Φουριέ του λογαριθμικού προσεγγιστικού φάσματος(cepstral), πράγμα που δεν θα μπορούσε να γίνει με άλλη δομή

συμπίεσης (όπως π.χ. τετραγωνική ρίζα). Τέλος, στο τελευταίο στάδιο γίνεται ο υπολογισμός του DCT (Διακριτός Μετασχηματισμός Συνημιτόνου) του λογαρίθμου των φίλτρων. Ο μετασχηματισμός αυτός υπολογίζεται για το κάθε frame i ως εξής:

$$S_i(k) = \sum_{n=1}^{N} s_i(n) h(n) e^{(\frac{-j2\pi kn}{N})} \tag{4.5}$$

Για το οποίο ισχύει $1 \leq k \leq K$ και όπου $s_i(n)$ ο χρονικός τομέας του αντίστοιχου i δείγματος, $P(k)$ το δυναμικό φάσμα του εκάστοτε frame (ή αλλιώς περιοδόγραμμα), $h(n)$ παράθυρο ανάλυσης μεγέθους N δειγμάτων και τέλος, K είναι το μέγεθος ανάλυσης του DFT. Το δυναμικό φάσμα για κάθε i frame υπολογίζεται ως εξής:

$$P_i(k) = \frac{1}{N} |S_i(k)|^2 \tag{4.6}$$

Όλα τα φίλτρα του DCT επικαλύπτονται μεταξύ τους, με αποτέλεσμα οι ενέργειες των φίλτρων να συσχετίζονται σε κάποιο βαθμό. Ο μετασχηματισμός αυτός διαχωρίζει τις ενέργειες, πράγμα που σημαίνει πως μπορούν να χρησιμοποιηθούν διαγώνιες μήτρες συνδιακύμανσης για την μοντελοποίηση των χαρακτηριστικών του εκάστοτε classifier. Από αυτούς τους 26 συντελεστές DCT μόνο οι 12 κρατιούνται (ο 2ος εώς τον 13ο). Αυτό συμβαίνει γιατί οι υψηλότεροι συντελεστές αντιπροσωπεύουν γρηγορότερες αλλαγές στις ενέργειες των φίλτρων με αποτέλεσμα να υποβιβάζεται το αποτέλεσμα.

4.2 Λειτουργικά Χαρακτηριστικά

Τα λειτουργικά χαρακτηριστικά [56] είναι χαρακτηριστικά εξαγωγής πληροφορίας σε περιπτώσεις πολλών μεταβλητών, σχεδιασμένα με σκοπό την κάλυψη των πιθανών διαστάσεων του προβλήματος. Αποτελούνται από δεδομένα που μπορούν να ληφθούν ανά συγκεκριμένες διακριτές χρονικές στιγμές και αναλύουν την σχέση μεταξύ αυτών των δεδομένων. Συγκεκριμένα, τα χαρακτηριστικά που ανήκουν σε αυτήν την κατηγορία μπορεί να είναι:

- Αναπαράστασης της Κατανομής λειτουργιών (mean, variation, covariation)
- Σχέσης μεταξύ των Λειτουργικών δεδομένων (covariates, responses, other functions)
- Σχέσης μεταξύ παράγωγων των Λειτουργιών
- Στιγμής (χρόνου) γεγονότων στις Λειτουργίες

4.2.1 Στατιστικά Χαρακτηριστικά

Τα χαρακτηριστικά αυτά είναι τιμές που υπολογίζονται βάσει στατιστικών μεθόδων, χρήσιμων σε περιπτώσεις αρκετών μεταβλητών, ή μεγάλου αριθμού παρατηρημένων δειγμάτων. Τέτοιου τύπου μετρήσεις είναι οι εξής:

- ***Arithmetic Mean*** (Αριθμιτικό Μέσο)

- ***Centroid*** (Κέντρο Βάρους)
- ***Root Quadratic Means*** (Τετραγωνική Ρίζα Μέσου)
- ***Number of Non-Zero values*** (Αριθμός των Μη-Μηδενικών Τιμών)
- ***arithmetic mean of non-zero values*** (Αριθμητικό Μέσο Μη-Μηδενικών) Τιμών
- ***Standard Deviation*** (Τυπική Απόκλιση): για την ποσοτικοποίηση της ποσότητας της μεταβολής ή της διασποράς
- ***Zero Crossing Rate*** (Μηδενικού Ποσοστού Διέλευσης): εξάγουν τον ρυθμό της σημαδιακής αλλαγής του σήματος κατά την διάρκεια συγκεκριμένου στιγμιότυπου.
- ***Interquartile Range (IQR)*** (Διατεταρτημοριακό εύρος): στατιστική διασπορά ή διακύμανση του μεσαίου 50% των παρατηρήσεων.

4.2.2 Regression related

Τα χαρακτηριστικά αυτά είναι οι τιμές που δείχνουν τη σχέση μεταξύ εξαρτημένης (στόχου) και ανεξάρτητης μεταβλητής (πρόβλεψης), την αιτιώδη σχέση δηλαδή μεταξύ των μεταβλητών. Τέτοιου τύπου μετρήσεις είναι οι εξής:

- ***Linear Regression Slope*** (Κλίση Γραμμική Παλινδρόμηση)
- ***Offset*** (Αντιστάθμισμα)
- ***Corresponding Approximation Error*** (Αντίστοιχο Σφάλμα Προσέγγισης)
- ***Quadratic Regression Coefficients*** (Τετραγωνικοί Συντελεστές Παλινδρόμησης)

4.2.3 Minima/maxima related

Τα χαρακτηριστικά αυτά έχουν να κάνουν με τα **maxima** και **minima** (τα αντίστοιχα πολλαπλάσια του μέγιστου και του ελάχιστου) μιας συνάρτησης, που είναι γνωστά συλλογικά ως **extrema** (από τον πληθυντικό της λέξης **extremum**). Οι τιμές αυτές είναι η μεγαλύτερη και η μικρότερη τιμή της συνάρτησης είτε μέσα σε ένα δεδομένο εύρος (το τοπικό ή το σχετικό άκρο) είτε σε ολόκληρο τον τομέα μιας συνάρτησης (το σφαιρικό ή το απόλυτο άκρο).

- ***Range*** (Εύρος)
- ***Position of Max/Min*** (Θέση Μέγιστου/Ελάχιστου)
- ***Difference Max/Min*** (Διαφορά Μέγιστου/Ελάχιστου)
- ***Arithmetic Mean*** (Αριθμητικό Μέσο)

Κεφάλαιο 5

Ανάλυση και Υλοποίηση

Στο κεφάλαιο αυτό γίνεται η εκτεταμένη περιγραφή της διαδικασίας με την οποία έγινε η υλοποίηση και η σχεδίαση τόσο του classifiers βάσει του συνόλου δεδομένων (το οποίο αναλύθηκε στο κεφάλαιο 2), όσο και του ίδιου του πρακτικού κομματιού στο Raspberry Pi (το οποίο τρέχει σε πραγματικό χρόνο), καθώς και των κριτηρίων της διαδικασίας αυτής.

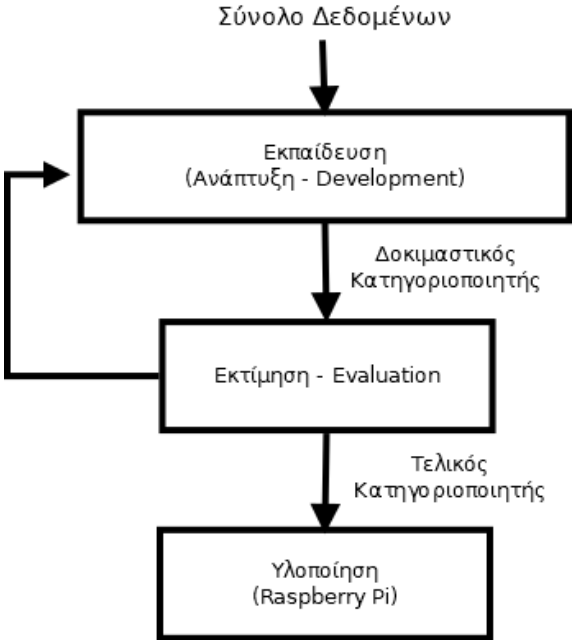
Για την υλοποίηση της εργασίας αυτής χρησιμοποιήθηκε η γλώσσα προγραμματισμού python, λόγω της ευκολίας που παρέχει τόσο κατά τον ίδιο τον προγραμματισμό, λόγω του εύκολου και γρήγορου για την υλοποίηση συντακτικού της, όσο και της μεγάλης ποικιλίας εξωτερικών βιβλιοθηκών, όπως η scikit-learn [57] ή η mlpy [58] (η οποία χρησιμοποιήθηκε για την δημιουργία των μοντέλων ταξινομητών για αυτήν την εργασία, καθώς και περαιτέρω εργαλείων τα οποία θα αναλυθούν στην συνέχεια). Τα παραπάνω είναι ολοκληρωμένα εργαλεία για όλες τις λειτουργίες μηχανικής μάθησης που χρειάζεται κάποιος για την ανάπτυξη σχετικών εφαρμογών.

Οι διαδικασίες για την όλη έρευνα και υλοποίηση, καθώς και η σειρά με την οποία έγιναν, παρουσιάζονται στο σχήμα 5.1.

Για τα τρία κομμάτια που θα αναλυθούν στην συνέχεια έχει προηγηθεί υλοποίηση ξεχωριστού λογισμικού για το κάθε ένα από αυτά, σχεδιασμένο σε γλώσσα Python (έκδοση 2.7). Η λογική με την οποία έγινε η σχεδίαση αυτών των προγραμμάτων είναι βασισμένη στην ευκολότερη μελλοντική ανάπτυξή του με περισσότερο και πιο αναπτυγμένο υλικό, όπως π.χ. για ταξινομητές εκπαιδευμένους με μεγαλύτερες, πιο αναπτυγμένες και προσανατολισμένες βάσεις δεδομένων.

5.1 Εκπαίδευση

Σκοπός της βαθμίδας αυτής, είναι η δημιουργία του τελικού ταξινομητή μέσω της χρήσης των χαρακτηριστικών που εξάγονται από την βάση δεδομένων που σχεδιάστηκε, προσπαθώντας για την επίτευξη του όσο το δυνατόν μεγαλύτερου αποτελέσματος σωστής πρόβλεψης κλάσεων των γεγονότων. Για να γίνει αυτό όμως πρέπει πρώτα το μοντέλο να περάσει μέσα από μία σειρά από διαδικασίες ελέγχου, ώστε να διαμορφωθεί (ως προς τις παραμέτρους του και την είσοδό του) κατάλληλα με σκοπό την απόδοση και την αξιοπιστία του. Το κομμάτι της



Σχήμα 5.1: Διαδικασία ανάπτυξης

ανάπτυξης λοιπόν είναι η αρχή της διαδικασίας, όπου μαζί με την διαδικασία της εκτίμησης, αποτελεί τον κορμό όλης της ανάπτυξης του τελικού μοντέλου. Μετά από πρόχειρες δοκιμές σε όλα τα μοντέλα που αναφέρθηκαν στο σχήμα 1.1, τα μοντέλα των ταξινομητών SVM και KNN είναι αυτά τα δύο τα οποία αποδίδουν καλύτερα στις κατηγορίες της εργασίας. Άρα τα στάδια που ακολουθούν αναφέρονται μόνο σε αυτά τα δύο μοντέλα και όχι σε όλα όσα είχαν αρχικά οριστεί, μια και σκοπός πλέον είναι η βελτίωση της όλης διαδικασίας πάνω στα μοντέλα που αποδίδουν καλύτερα.

5.1.1 Εξαγωγή Χαρακτηριστικών

Αρχικά, η λογική και η όλη διαδικασία της εξαγωγής των χαρακτηριστικών έγινε βάσει της pyAudioAnalysis υλοποίησης [59], η οποία αποτελεί μια ολοκληρωμένη λύση τόσο για την εξαγωγή των χαρακτηριστικών, όσο και για την εκπαίδευση όλων των δημοφιλών μοντέλων ταξινομητών, καθώς και της διαδικασίας εκτίμησης μέσω της χρήσης cross-validation εκπαίδευσης. Η βιβλιοθήκη αυτή, αν και δεν ήταν αυτή που χρησιμοποιήθηκε για την εξαγωγή των χαρακτηριστικών για την τελική υλοποίηση της εργασίας (παρά μόνο για την εκπαίδευση με cross-validation), ήταν ένας μεγάλος οδηγός για την σειρά και την λογική με την οποία γίνονται αρκετά πράγματα, τόσο για την εξαγωγή των χαρακτηριστικών και τον υπολογισμό τους, όσο και κατά την επεξεργασία τους για την είσοδό τους στον ταξινομητή.

Ενώ λοιπόν το αρχικό μοντέλο για την εκπαίδευση ήταν βασισμένο στην pyAudioAnalysis βιβλιοθήκη για την εξαγωγή των χαρακτηριστικών, παρά τις δοκιμές μας, στα αποτελέσματα της εκτίμησης η πιθανότητα σωστής κατηγοριοποίησης ήταν πολύ χαμηλή. Συγκεκριμένα, η βιβλιοθήκη αυτή εξάγει δύο είδη χαρακτηριστικών με το κάθε ένα από αυτά να περιέχει 34 χαρακτηριστικά, σύνολο δηλαδή 68. Τα είδη αυτά είναι τα χαρακτηριστικά τύπου Short-term

(Βραχυπρόθεσμα) και Mid-term (Μεσοπρόθεσμα). Τα είδη αυτά προκύπτουν από τα ίδια αρχικά χαρακτηριστικάαλλά με διαφορετική χρονική στατιστική διαδικασία. Η τελική προσέγγιση του δικού μας προβλήματος έγινε με την βοήθεια του OpenSmile, το οποίο επιτρέπει την παραγωγή πάνω από 6000 διαφορετικών χαρακτηριστικών από το σήμα ήχου. Συγκεκριμένα, η επιλογή των χαρακτηριστικών για εξαγωγή γίνεται μέσω Configuration αρχείου, το οποίο είτε μπορεί να δημιουργηθεί βάσει των αναγκών του εκάστοτε προβλήματος, είτε με την απευθείας χρήση κάποιου έτοιμου από τα ήδη υπάρχοντα. Το αρχείο επιλογής αυτής της υλοποίησης ανήκει στα έτοιμα, ονομάζεται emobase.conf και εξάγει 989 διαφορετικά χαρακτηριστικά ανά λήψη παρατήρησης (Βραχυπρόθεσμα μόνο σε αυτό το στάδιο). Αν και αποτελούν υπερβολικό νούμερο για Real Time υλοποίηση μέσω του Raspberry Pi, κατά την εκτίμηση τα αποτελέσματα είναι αρκετά καλύτερα στην πρόβλεψη των κλάσεων χωρίς περαιτέρω διαδικασίες επεξεργασίας αυτών. Ταυτόχρονα υπάρχει και η επιλογή για την ομαδοποίηση ή επιλογή των πιο κατάλληλων χαρακτηριστικών από το σύνολο αυτών, πράγμα που θα μπορούσε να δώσει πιο στοχευμένη μορφή στα χαρακτηριστικά και να μειώσει και την επεξεργαστική ισχύ που απαιτείται για την διαδικασία της πρόβλεψη.

Σε αντίθεση με την υλοποίηση του pyAudioAnalysis, στο OpenSmile δεν υπάρχει αμεσότητα επικοινωνίας του κυρίως κώδικα με το λογισμικό για την εξαγωγή λόγω διαφορετικών γλωσσών (αντίθετα στην πρώτη περίπτωση τα δεδομένα περνάνε άμεσα μέσω numpy, το οποίο αποτελεί μια ταχύτατη υλοποίηση μαθηματικών και επιστημονικών διαδικασιών λόγω της αρχιτεκτονικής του, γραμμένη σε γλώσσα C/C++ και η οποία είναι πολύ πιο γρήγορη από την python). Ο τρόπος λοιπόν που τα χαρακτηριστικά περνάνε στο main πρόγραμμα εκπαίδευσης είναι με την εγγραφή τους κατά την εξαγωγή σε comma separated values (csv) μορφής αρχείο και διάβασμά του γραμμή - γραμμή μέσω του κώδικά μας. Η επιλογή για εξάγωγή σε csv, γίνεται μέσω του αρχείου διαμόρφωσης (configuration file), όπου μετά από ελάχιστες διαφοροποιήσεις στην αρχική έκδοση του emobase.conf, που όπως αναφέρθηκε χρησιμοποιείται σε αυτήν, αποτελεί μέρος της OpenSmile υλοποίησης για την αναγνώριση συναισθημάτων, αλλά αποδίδει πολύ καλά και στην περίπτωση των κλάσεών μας. Επίσης, στο συγκεκριμένο αρχείο βοηθάει το γεγονός ότι υπάρχει και αντίστοιχο έτοιμο .conf αρχείο για ζωντανή εξαγωγή των χαρακτηριστικών σε πραγματικό χρόνο, πράγμα που μας βοηθάει στο τελευταίο στάδιο της υλοποίησης, στην πρόβλεψη δηλαδή σε πραγματικό χρόνο.

Αφού διαβαστεί λοιπόν η γραμμή με όλα τα χαρακτηριστικά, φορμάρεται σε μορφή για numpy. Με τον τρόπο αυτό, μπορούν να υπολογιστούν τόσο τα mid - term χαρακτηριστικά (μέσο όρο n- χαρακτηριστικών), όσο και οι τιμές MEAN και STD (Standard Deviation) για τον υπολογισμό της μέσης τιμής και την απόκλιση από αυτήν αντίστοιχα. Οι τιμές των χαρακτηριστικών υπολογίζονται ξανά με την χρήση των δύο τιμών αυτών και τροφοδοτούνται στον ταξινομητή για την εκπαίδευσή τους μέσω της χρήσης της μεθόδου K-fold cross-validation, χωρίζοντας τον αριθμό των χαρακτηριστικών ανά κατηγορία σε K ίσα τμήματα και με K αριθμό επαναλήψεων με διαφορετικό τμήμα για δοκιμή κάθε φορά. Αυτό επιτρέπει την καλύτερη εκτίμηση του μοντέλου μέσω της επαφής του ταξινομητή με περισσότερα άγνωστα δεδομένα, μιας και μέσω των K που δημιουργούνται, όλη η βάση περνάει από αυτόν ως άγνωστο δεδομένο. Η όλη διαδικασία του K-fold cross-validation της εργασίας επαναλαμβάνεται για κάθε

μία από τις τιμές των συντελεστών του ταξινομητή. Τα αποτελέσματα των μετρήσεων precision rate (PRE), recall rate (REC) και F1 (βλ. κεφάλαιο 3.3) αποθηκεύονται και επιλέγεται η τιμή του συγκεκριμένου συντελεστή, η οποία είχε την καλύτερη απόδοση για την πρόβλεψη των κλάσεων.

Σε αυτήν την εργασία, δημιουργούνται τρεις ταξινομητές, ένας για την κάθε μία από τις κατηγορίες που αναλύθηκαν στο προηγούμενο κεφάλαιο. Συγκεκριμένα, λόγω της πολυμορφίας του συνόλου της βάσης, χρησιμοποιήθηκε όσο τον δυνατόν μεγαλύτερος αριθμός γεγονότων για την εκπαίδευση με σκοπό την επίτευξη καλύτερων αποτελεσμάτων. Για να γίνει αυτό εξυπηρετούσε να χωριστεί το σύνολο της βάσης στην μορφή ακριβώς που έγινε και η ανάλυσή της (Soundscapes, Vehs, Others) για να μην βαρύνουν μόνο έναν ταξινομητή ο οποίος θα καλούταν να βγάλει ποσοστά πιθανοτήτων για τόσες πολλές κλάσεις. Με αυτόν τον τρόπο, υπάρχει και μεγαλύτερη ευελιξία ως προς την εξισορρόπηση του αριθμού των δεδομένων που θα τροφοδοτηθούν στον κάθε ταξινομητή.

Μετά από την προεπεξεργασία της βάσης λοιπόν έχουν καταταχθεί σε κατηγορίες όλα τα wav αρχεία των επιμέρους τμημάτων της βάσης (Soundscapes, Vehs, και Others) Συγκεκριμένα για την κάθε μία από της κατηγορίες ταξινομητών, η αντίστοιχη βάση που θα χρησιμοποιηθεί για τροφοδότηση αποτελείται από:

- 147 αρχεία για κάθε μία κατηγορία (από τις 9 κλάσεις, 1323 αρχεία σύνολο), μεγέθους 10 δευτερολέπτων το κάθε ένα από αυτά, για την Soundscape κατηγορία
- 319 αρχεία για κάθε μία κατηγορία (από τις 2 κλάσεις, 638 αρχεία σύνολο), μη σταθερού μεγέθους (όλο το γεγονός), για το Vehs υποσύνολο των Events
- 46 αρχεία για κάθε μία κατηγορία (από τις 4 κλάσεις, 184 αρχεία σύνολο) μη σταθερού μεγέθους (όλο το γεγονός), για το Other υποσύνολο των Events

5.1.2 Επιλογή Παραμέτρων

Οι παράμετροι των classifier αποτελούν σημαντικό παράγοντα μιας και καθορίζουν τον τρόπο συμπεριφοράς του classifier. Στον κώδικά μας ο όρος εμφανίζεται γενικά ως «παράμετροι», οι οποίες όμως αντιπροσωπεύουν διαφορετικό στοιχείο ανάλογα με τον τύπο του εκάστοτε classifier. Συγκεκριμένα για τους classifiers SVM και KNN που χρησιμοποιούνται στο κομμάτι της υλοποίησης, υπάρχει η παράμετρος κόστους C για τον πρώτο, η οποία ρυθμίζει το βάρος του ταξινομητή κατά τον χωρικό διαχωρισμό των δειγμάτων που καλείται να ταξινομήσει, ενώ για τον δεύτερο έχουμε τον αριθμό γειτόνων, την τιμή K δηλαδή, η οποία καθορίζει την ομαλότητα των συνόρων που σχεδιάζει ο ταξινομητής. Για την πρώτη περίπτωση του ταξινομητή τύπου SVM (είτε γραμμικού, είτε rbf), ο ίδιος παίρνει τις τιμές $C : 0.001, 0.01, 0.5, 1.0, 5.0$ και 10.0. Αντίστοιχα ο ταξινομητής KNN παίρνει τις K τιμές : $1, 3, 5, 7, 9, 11, 13$ και 15. Μέσω της διαδικασίας της k-fold cross validation εκπαίδευσης επιλέγεται η τιμή αυτή, με την οποία ο ταξινομητής έχει την βέλτιστη απόδοση (πράγμα που φαίνεται από τα στοιχεία PRE, REC και F1 που υπολογίζονται ανά εκπαίδευση και δοκιμή

μέσα στην λούπα) και καταχωρείται μαζί με τα υπόλοιπα στοιχεία του classifier για να μπορούν να τον καλέσουν αργότερα στο στάδιο της υλοποίησης που θα αναλυθεί παρακάτω.

5.1.3 Επιλογή Χαρακτηριστικών

Τέλος, όπως αναφέρθηκε, μέσω της διαδικασίας αυτής επιλέγονται τα χαρακτηριστικά τα οποία θα αποδώσουν καλύτερα κατά την πρόβλεψη ενώ παράλληλα μειώνεται και ο συνολικός αριθμός των χαρακτηριστικών που εισάγονται στο μοντέλο. Όπως αναλύθηκε στο κεφάλαιο 3.4, υπάρχουν τρεις μέθοδοι με τις οποίες μπορεί να γίνει αυτό. Λόγω της δομής του λογισμικού μας αλλά και κάποιων περιορισμών στην βιβλιοθήκη mlpy που χρησιμοποιήθηκε για την δημιουργία των μοντέλων, δεν επιτρέπεται να γίνει αυτή η διαδικασία μέσα στην λούπα της διαδικασίας k-fold cross-validation. Αυτό θα ήταν πιο ορθό, λόγω του ότι τα χαρακτηριστικά που εντοπίζονται σε κάθε ένα πακέτο δεδομένων (του K συνόλου) δοκιμάζονται σε άλλα άγνωστα πακέτα δεδομένων και από όλη αυτήν την διαδικασία θα επιλεγόταν ποια από τα συνολικά χαρακτηριστικά είναι τα κατάλληλα για να χρησιμοποιήσουμε για την τελική εκπαίδευση. Με αυτόν τον τρόπο δηλαδή αποφεύγεται το φαινόμενο του overfitting.

Αντ΄ αυτού η επιλογή των χαρακτηριστικών γίνεται με την βοήθεια του λογισμικού WEKA [60], το οποίο περιέχει ένα μεγάλο σύνολο από εργαλεία για data mining. Ένα από τα εργαλεία που παρέχονται στο λογισμικό αυτό είναι η επιλογή των χαρακτηριστικών. Αν και το λογισμικό αυτό επιτρέπει την επιλογή χαρακτηριστικών μέσω k-fold cross-validation εκτίμησής τους, στην εργασία εφαρμόζεται η τρίτη μέθοδος επιλογής χαρακτηριστικών, η μέθοδος δηλαδή της Ενσωμάτωσης των Χαρακτηριστικών (αλγόριθμος BestFirst Search). Με τη μέθοδο αυτή το λογισμικό ακολουθεί μια σειρά από διαδικασίες, με τις οποίες ελέγχει ποια από τα χαρακτηριστικά που τροφοδοτήθηκαν, έχουν καλύτερη απόδοση στον ίδιο τον ταξινομητή. Το λογισμικό εμπεριέχει όλους τους γνωστούς αλγορίθμους ταξινομητών, τους οποίους χρησιμοποιεί για να κάνει τον έλεγχο των χαρακτηριστικών και να βρει τα βέλτιστα. Για να γίνει αυτό όμως πρέπει πρώτα να αποφασιστεί ποιος είναι ο καλύτερος ταξινομητής για τις δικές μας περιπτώσεις.

5.2 Εκτίμηση

Σε αυτό το στάδιο της διαδικασίας εξετάζεται η απόδοση του ταξινομητή τόσο για την απόδοσή του σε άγνωστα δεδομένα, όσο και σε δοκιμαστικές συνθήκες σαν προσομοίωση τελικού project. Με την εκτίμηση (Evaluation) μπορεί, όπως δηλώνει και το όνομά της, να φανεί εάν ο ταξινομητής που δημιουργήθηκε κατά την διαδικασία που αναλύθηκε, αποδίδει σε δεδομένα που δεν έχουν χρησιμοποιηθεί για εκπαίδευση και αν ναι, σε τι ποσοστό. Το στάδιο αυτό διαφέρει από αυτό του K-fold cross-validation σε αυτή την εργασία (που όπως αναφέρθηκε και εκεί γίνεται εκτίμηση), για τον λόγο ότι τώρα ελέγχεται μεν η όλη διαδικασία του project με σκοπό την καλύτερη απόδοση του κάθε ενός από τους ταξινομητές, αλλά όλο αυτό γίνεται βάσει των βέλτιστων παραμέτρων που ήδη εκτιμήθηκαν ως βέλτιστοι στην K-fold cross-validation διαδικασία. Εδώ ελέγχονται όλες οι άλλες παράμετροι που υπάρχουν στην συνολική διαδικασία για την εξαγωγή, βάσει των οποίων σχεδιάστηκε το πρακτικό κομμάτι αργότερα. Δίνεται βάση στις διαδικασίες τόσο πριν την εξαγωγή των χαρακτηριστικών με

το στάδιο του pre processing των δεδομένων που θα τροφοδοτηθούν (στην περίπτωση της βάσης μας είναι τα αρχεία ήχου και το στάδιο της μελέτης για την συλλογή των βέλτιστων χαρακτηριστικών) όσο και κατά το διάβασμα αυτών από τον ίδιο τον classifier.

Έτσι λοιπόν στο στάδιο αυτό δίνεται βάρος στον ρυθμό δειγματοληψίας (για έλεγχο ανάγκης καλύτερης προεπεξεργασίας σήματος), στον βέλτιστο χρόνο κοπής τμήματος του ήχου (από το οποίο γίνεται η εξαγωγή των χαρακτηριστικών), στο αν αποδίδει καλύτερα η πρόβλεψη με την λήψη των χαρακτηριστικών με επικάλυψη (το μέγεθος του παραθύρου υπολογισμού είναι μεγαλύτερο από το βήμα) ή χωρίς επικάλυψη (το παράθυρο και το βήμα είναι ίσα), στον τύπο (είδος) και στον αριθμό των χαρακτηριστικών που εξυπηρετούν καλύτερα το εκάστοτε μοντέλο καθώς και στην επεξεργασία αυτών και τέλος στις διαφορές ανάμεσα σε short-term - mid-term χαρακτηριστικά.

5.2.1 Τμηματοποίηση Ηχητικών Αποσπασμάτων

Ο βέλτιστος χρόνος κοπής των αποσπασμάτων ήχου για δοκιμή πρόβλεψης της κλάσης παίζει μεγάλο ρόλο στο τελικό αποτέλεσμα, μιας και τα αρχεία της βάσης δεδομένων δεν ήταν ίσα ως προς το μέγεθός τους σε χρόνο, αφού ήταν διαχωρισμένα βάσει των ίδιων των γεγονότων, τα οποία ποικίλουν σε χρόνο τόσο μεταξύ άλλων κλάσεων (π.χ. άλλη διάρκεια έχει το κελάηδημα ενός πουλιού και άλλη διάρκεια ο ήχος που δημιουργείται από το πέρασμα ενός αυτοκινήτου), όσο και μεταξύ ίδιας κλάσης (π.χ. δύο διαφορετικοί ήχοι ίδιας κόρνας αλλά με διαφορετική διάρκεια ο κάθε ένας από αυτούς), με αποτέλεσμα να υπάρχει ανομοιομορφία και να δημιουργείται πρόβλημα κατά το πέρασμα των χαρακτηριστικών στον ταξινομητή. Η λύση λοιπόν στην χρονική διάρκεια του ηχητικού τμήματος (παραθύρου) πρέπει να ικανοποιεί όλες αυτές τις περιπτώσεις, ώστε να υπάρχει αρκετά μεγάλη ακρίβεια κατά την πρόβλεψη μέσω των τιμών των χαρακτηριστικών που προκύπτουν από το τμήμα. Αυτό μπορεί να υπολογιστεί χονδρικά από μέσους όρους των τμημάτων, αλλά παρόλα αυτά, η καλύτερη διάρκεια για την βέλτιστη απόδοση του τμηματοποιητή μπορεί να βρεθεί με πειραματικές δοκιμές και συνεχόμενες εκτιμήσεις αυτών. Σκοπός αυτού του εγχειρήματος στην ουσία είναι η διάρκεια αυτή να περιέχει λιγότερο θόρυβο (ήχος/πληροφορία που δεν ανήκει στην κλάση) και παράλληλα να μην πετάει έξω πληροφορία χρήσιμη για τον χαρακτηρισμό των κλάσεων.

Το τμήμα αυτό στο τελικό στάδιο της εργασίας, που είναι η τμηματοποίηση σε αληθινό χρόνο, ορίζεται από το αρχείο config του OpenSmile μέσω της παραμέτρου blocksize το οποίο χαρακτηρίζεται σε αριθμό δειγμάτων (frames) και ορίζει το παράθυρο επεξεργασίας του σήματος για την εξαγωγή. Περαιτέρω επεξεργασία για την χρονική κατάσταση των χαρακτηριστικών μπορεί να γίνει στον κώδικα που υλοποιήθηκε σε Python μέσω του υπολογισμού βραχυπρόθεσμων και μεσοπρόθεσμων χαρακτηριστικών.

5.2.2 Βραχυπρόθεσμα και Μεσοπρόθεσμα χαρακτηριστικά

Παρομοίως παίζουν και αυτά ρόλο για την πρόβλεψη των κλάσεων. Όσο αναφορά τις διαφορές στα short-term και mid-term χαρακτηριστικά, τα δεύτερα στην ουσία υπολογίζονται βάσει των πρώτων, μιας και βγαίνουν στατιστικά μέσα από σύνολο short-term χαρακτηριστι-

κών. Τα short-term χαρακτηριστικά εξάγονται ανά βήμα (το οποίο είναι αποδεκτό συνήθως με τιμές 20 με 100 ms) σε μορφή λίστας, η οποία περιέχει ένα από κάθε μορφή χαρακτηριστικό που έχει επιλεχτεί (π.χ. MFCCs, Energy, Spectral Energy, κ.λ.π.) Τα mid-term χαρακτηριστικά λοιπόν δείχνουν μια πιο κανονικοποιημένη εικόνα του ήχου προς μελέτη, εξαλείφοντας τυχαίες εναλλαγές του σήματος που μπορεί να εμφανίστηκαν κατά την διάρκεια. Το μέγεθος όμως των mid-term χαρακτηριστικών, όπως και στον χρόνο κοπής τμήματος του ήχου που αναφέρθηκε προηγουμένως, παίζει άμεσα καθοριστικό ρόλο για την έξοδο του αποτελέσματος πρόβλεψης. Ταυτόχρονα όμως υπάρχει περιορισμός των πόσων short-term μπορεί να διαχειριστεί ένα σύστημα, της συχνότητας δηλαδή με την οποία αυτό τροφοδοτείται με χαρακτηριστικά τα οποία πρέπει να ταξινομήσει ανάλογα με την επεξεργαστική του ισχύ. Έτσι χρειάζεται να υπάρξει περιορισμός αυτής της παραμέτρου για την εργασία αυτή, λόγω του ότι θα χρησιμοποιηθεί ένα Raspberry Pi model 2B σαν μονάδα επεξεργαστή στην τελική υλοποίηση.

Με λίγα λόγια, στο στάδιο αυτό έχουν εφαρμοστεί όλα τα θεμέλια, ώστε να υπάρχουν οι επιλογές αργότερα για κρίση του τι δουλεύει και τι όχι πάνω στην διαδικασία και ταυτόχρονα αρκετές επιλογές ώστε να μπορούν να γίνουν οι απαραίτητες δοκιμές μέσω της διαδικασίας της εκτίμησης, μιας και το στάδιο αυτό δεν γίνεται μια φορά αλλά μαζί με το στάδιο της εκπαίδευσης επαναλαμβάνεται μέχρι να εξαντληθούν όλες οι επιλογές που τέθηκαν.

5.3 Υλοποίηση

Τέλος, στο κομμάτι αυτό, με βάση όλες τις δοκιμές που πραγματοποιήθηκαν στα προηγούμενα στάδια, μπορεί πλέον να σχεδιαστεί και να υλοποιηθεί το σύστημα αυτό, ώστε να δουλεύει σε πραγματικές συνθήκες και πραγματικό χρόνο. Όπως αναλύθηκε στα προηγούμενα κεφάλαια λοιπόν, το σύστημά αποτελείται από το Raspberry Pi σαν μονάδα επεξεργαστή, ένα USB μικρόφωνο με το οποίο θα γίνεται η καταγραφή του ήχου, μία WI-FI κεραία με την οποία ο χρήστης θα μπορεί να διαβάσει τα αποτελέσματα (στο μοντέλο 2B που χρησιμοποιείται εδώ δεν εμπεριέχεται ενσωματωμένο κι έτσι χρησιμοποιείται εξωτερική κεραία σε μορφή USB) και τέλος ένα σύνολο από scripts με εντολές, το οποίο δημιουργήθηκε για αυτοματισμούς και για την ίδια την διαδικασία της κατηγοριοποίησης.

5.3.1 Hardware

Raspberry Pi

Όπως αναφέρθηκε, το Raspberry pi είναι μια πλακέτα σε μέγεθος πιστωτικής κάρτας η οποία μπορεί να φιλοξενήσει λειτουργικό σύστημα (επιλεγμένες πλατφόρμες προς το παρόν βασισμένες σε Linux λειτουργικό) και περιέχει πλήθος πιν εισόδων - εξόδων, led ενδείξεις για τις λειτουργίες του, καθώς και, ανάλογα το μοντέλο, πλήθος I/O διασυνδέσεων όπως Ethernet για σύνδεση στο δίκτυο, USB για σύνδεση με πληκτρολόγιο, ποντίκι και άλλων περιφερειακών συσκευών, HDMI ή RCA έξοδο για οθόνη καθώς και έξοδο jack 3.5mm για

τον ήχο και τέλος είσοδο standard ή micro SD (στα μοντέλα 2ης γενιάς και μετά) κάρτας όπου γίνεται η αποθήκευση των αρχείων μας και του λειτουργικού συστήματος. Η τροφοδοσία του γίνεται από συμβατό τροφοδοτικό καλώδιο (η εταιρεία συνιστά το δικής της κατασκευής 5V (DC Micro-USB Type B καλώδιο). Δίνεται η δυνατότητα για σύνδεση οποιασδήποτε συσκευής σχεδιασμένης για Arduino, λόγω του ότι το Arduino είναι πιο διαδεδομένο από το Raspberry pi και τα περιφερειακά που είναι φτιαγμένα για τη συσκευή αυτή είναι περισσότερα. Έτσι υπάρχει στο εμπόριο γέφυρα διασύνδεσης, η οποία δίνει τη δυνατότητα προσαρμογής τους σε ένα Raspberry με την χρήση γέφυρας διασύνδεσης Arduino η οποία εύκολα προσαρμόζεται στο Raspberry. Με αυτόν τον τρόπο μπορούν να αξιοποιηθούν καλύτερα οι περισσότερες δυνατότητες που προσφέρει το Raspberry με τα περιφερειακά όμως του Arduino, όπως για παράδειγμα για την υλοποίηση IoT δικτυώσεων κ.α. Για τον ήχο συγκεκριμένα υποστηρίζει το υποκείμενο πλαίσιο Advanced Linux Sound Architecture (ALSA) που χρησιμοποιείται σε linux συστήματα, το οποίο είναι η αρχιτεκτονική ήχου η οποία χρησιμοποιείται τόσο από το Raspberry όσο και από τα USB περιφερειακά που συνδέονται με αυτό και σχετίζονται με ήχο παρέχοντας Kernel drivers.

Μικρόφωνο

Το μικρόφωνο τύπου USB θα μπορούσε να είναι οποιοδήποτε μοντέλο της αγοράς, από την άποψη ότι όλα παρέχουν την βασική λειτουργία που θέλουμε να καλύψουμε, την λήψη δηλαδή του ήχου αυτού καθ΄ αυτού. Βασική προϋπόθεση όμως είναι η ικανοποιητική ποιότητα καταγραφής ήχου. Στην εργασία αυτή γίνεται η χρήση του ίδιου μικροφώνου με το οποίο έγινε η διαδικασία των ηχογραφήσεων, για τον λόγο ότι κάθε κάψουλα (που περιέχεται μέσα στο μικρόφωνο και αποτελεί το εξάρτημα το οποίο είναι υπεύθυνο για την ίδια την λήψη της πληροφορίας) διαφέρει από τις άλλες (άλλων μοντέλων μικροφώνων), με αποτέλεσμα να υπάρχουν αποκλίσεις τόσο στην ταχύτητα με την οποία η κάψουλα κινείται, όσο και στο ίδιο το συχνοτικό φάσμα. Συγκεκριμένα, με την λέξη φάσμα ορίζεται τόσο το συχνοτικό εύρος, το σύνολο δηλαδή όλων των συχνοτήτων σε Hz (hertz) που μπορεί να καταγράψει το μικρόφωνο και μετριέται από την χαμηλότερη συχνότητα έως την υψηλότερη, όσο και η ίδια η απόκριση σε αυτό, η ευαισθησία δηλαδή σε διάφορες συχνοτικές μπάντες του εύρους. Αποτέλεσμα αυτών λοιπόν είναι η πιθανότητα ύπαρξης απόκλισης από μικρόφωνο σε μικρόφωνο, δίνοντας αποκλίσεις και στα ίδια τα χαρακτηριστικά που εξάγονται από το σήμα. Το αν η απόκλιση υπάρχει ή όχι και το πόσο μικρή ή μεγάλη αυτή είναι, δεν θα μας απασχολήσει σε αυτήν την εργασία. Για την τελική μας υλοποίηση χρησιμοποιείται το ίδιο Blue Snowflake, το οποίο, όπως αναφέρθηκε, περιέχει κάψουλα ακριβείας (precision-tuned), ψηφιακό μετατροπέα καθώς και ειδικά σχεδιασμένο ενσωματωμένο προενισχυτή.

5.3.2 Λογισμικό

OpenSmile

Αναλυτικότερα, η σειρά με την οποία λειτουργεί το σύστημα έχει ως εξής: Αρχικά φορτώνονται τα μοντέλα των ταξινομητών (τα οποία τρέχουν ταυτόχρονα) όπως επίσης και οι

STD και MEAN τιμές, οι οποίες είναι διαφορετικές για κάθε μοντέλο, για τον υπολογισμό των αντίστοιχων τιμών της standard deviation διαδικασίας (επίσης διαφορετική διαδικασία ανά μοντέλο), με την οποία υπολογίζονται οι καινούριες τιμές των χαρακτηριστικών, οι οποίες είναι αυτές που θα τροφοδοτηθούν στο σύστημα. Το κάθε μοντέλο περνάει σε δικό του νήμα (thread), πράγμα που επιτρέπει στο πρόγραμμα να υπολογίζει ταυτόχρονα τις διαφορετικές προβλέψεις βάσει των τιμών των χαρακτηριστικών που περνάνε σε αυτό. Επίσης, στο κάθε ένα από αυτά τα νήματα περνάνε και δύο λίστες τύπου Queue (ουρές), οι οποίες επιτρέπουν το πέρασμα δεδομένων από νήμα σε νήμα. Η ουρά είναι τύπου FIFO (First In - First out) και το αντικείμενο που είναι μέσα στην ουρά αφαιρείται από αυτήν κατά το διάβασμά του. Η πρώτη ουρά χρησιμοποιείται για το πέρασμα των χαρακτηριστικών στον ταξινομητή ενώ η δεύτερη για την επιστροφή της κλάσης και της πιθανότητας αυτής που εξάγει το μοντέλο ταξινομητή. Σε κάθε ταξινομητή από αυτούς χρησιμοποιείται διαφορετικό μέγεθος πακέτου χαρακτηριστικών, το οποίο λειτουργεί πρακτικά σαν διαφορετικό μέγεθος παραθύρου εξαγωγής των χαρακτηριστικών, χρησιμοποιώντας τις βέλτιστες τιμές τμηματοποίησης που βρέθηκαν κατά τις δοκιμές του προηγούμενου σταδίου.

Όπως και στις προηγούμενες διαδικασίες, έτσι και εδώ για τα χαρακτηριστικά χρησιμοποιείται μια ελαφρώς τροποποιημένη έκδοση του emobase conf αρχείου για το OpenSmile, σχεδιασμένη ειδικά για ζωντανό, κανονικού χρόνου, πέρασμα των χαρακτηριστικών στο CSV αρχείο. Το παράθυρο για το Raspberry Pi, το χρονικό πλαίσιο δηλαδή που διαβάζεται για να εξάγει τα χαρακτηριστικά, είναι περίπου ίσο με 0.5 sec (blocksize = 500). Στην συνέχεια, διαβάζονται σε πραγματικό χρόνο οι τιμές που γράφονται στο CSV αρχείο μέσω μιας συνάρτησης που υλοποιήθηκε στον κώδικα, η οποία τις περνάει σε πίνακα τιμών τύπου numpy. Ο πίνακας αυτός, όπως και στα προηγούμενα στάδια, πραγματοποιεί συλλογή σε πακέτα, τα οποία προορίζονται για την προώθηση σε παρακάτω νήματα.

Για να γίνει αυτό έχει προηγηθεί η υλοποίηση μιας συνάρτησης, η οποία βάζει σε σειρά όλες τις διαδικασίες που χρειάζονται για την λειτουργία καθ' όλη την διαδικασία της ζωντανής ροής. Αρχικά, τα δεδομένα των χαρακτηριστικών περνάνε μόνο στο μοντέλο που έχει οριστεί σαν *Master Classifier*. Στην περίπτωση της βάσης που αναλύθηκε, το μοντέλο που είναι ορισμένο σαν *master* είναι του ταξινομητή των ηχοτοπίων, μιας και είναι αυτό που θα καθορίσει (από ακουστική άποψη) το είδος των γεγονότων που υπάρχει μεγαλύτερη πιθανότητα να υπάρχουν στον χώρο καταγραφής. Με το αποτέλεσμα κλάσης της εξόδου του ταξινομητή αυτού γίνεται η επιλογή του δεύτερου ταξινομητή *Slave Classifier*, ο οποίος μπορεί να είναι ένας μόνο ή και οι δύο. Η λογική της επιλογής αυτής φαίνεται ότι γίνεται βάσει ιεραρχίας, η οποία (σύμφωνα με την δομή της βάσης) καθορίζει το τι κλάσεις βρίσκονται σε κάθε κατηγορία ηχοτοπίου. Η ιεραρχία αυτή φαίνεται στο σχήμα 5.1. Αυτό συμβαίνει ώστε το σύστημα να κερδίσει επεξεργαστικούς πόρους και μνήμη σε στιγμές που δεν υπάρχει νόημα στην χρήση των δύο Slave Classifier ταυτόχρονα.

Το σύστημα ελέγχει συνεχώς για έξοδο από τον Master Classifier λόγω της μικρής καθυστέρησης που υπάρχει για την δημιουργία των πακέτων και μόνο όταν επιστραφεί κάποια κλάση σαν αποτέλεσμα, γίνεται ο έλεγχος για να καλεστούν και οι επόμενοι ταξηνομητές. Η επιλογή προγραμματιστικά γίνεται μέσω μιας λίστας που ορίζεται στην αρχή του προγράμματος,

Πίνακας 5.1: Ιεραρχία κλάσεων

Other	BusStop	OpenMarket	Port
	Bus	Park	QuietStreet
	CoastalStreet	Pedestrian	
Vehs	BusyStreet	OpenMarket	
	BusStop	Pedestrian	
	CoastalStreet	QuietStreet	

η οποία με την σειρά της καθορίζει τις σχέσεις μεταξύ των ταξινομητών.

Προγραμματισμός νημάτων

Επειδή η συνάρτηση που αναφέρθηκε μπορεί να χρειαστεί να περάσει δεδομένα μέχρι και σε 3 νήματα ταυτοχρόνως, αλλά υπάρχει και η πιθανότητα να χρειαστεί να τα περάσει μόνο σε δύο, τα νήματα με τους ταξιμομητές μας λειτουργούν όσο δέχονται τις τιμές αυτές στις αντίστοιχες ουρές που τους ορίστηκε κατά το φόρτωμά τους, ενώ όταν η ουρά στερέψει, μένουν σε αδράνεια (καλείται στο νήμα η εντολή time.sleep()). Αυτό επιτρέπει την γρήγορη εναλλαγή ταξινομητών χωρίς να υπάρχει η ανάγκη για φόρτωσή τους από την αρχή, πράγμα που θα χρειαζόταν περισσότερο χρόνο. Το μειονέκτημα όμως είναι η παραπάνω μνήμη που ζητείται από το σύστημα, μιας και το νήμα παραμένει ενεργό παρόλο που ο ίδιος ο ταξινομητής «κοιμάται» για κάποια δευτερόλεπτα ανά κύκλο (ίσο με τον χρόνο του παραθύρου των χαρακτηριστικών), κάνοντάς τον στην ουσία αδρανή για το αντίστοιχο διάστημα. Αυτό σημαίνει ότι σε σενάρια που το ηχοτοπίο δεν μένει για οποιοδήποτε λόγο σταθερό, το πλήθος των Slave ταξινομητών μπορεί να αλλάξει ανά πάσα στιγμή, εάν αλλάξει και το ίδιο το ηχοτοπίο.

Τέλος, υπάρχουν δύο ακόμα νήματα τα οποία παίζουν καθοριστικό ρόλο για την λειτουργία του συστήματος. Το πρώτο από αυτά είναι το Κύριο Νήμα, το οποίο είναι αυτό που τρέχει όλο το πρόγραμμα και με αυτό ελέγχονται τα υπόλοιπα καθώς και κάποιες διαδικασίες, όπως η λήξη του λογισμικού (με την χρήση του πληκτρολογίου μέσω Ctr - C). Το τελευταίο νήμα είναι υπεύθυνο για το τύπωμα των εξόδων μέσω μιας ακόμα ουράς, στην οποία αποθηκεύονται όλες οι προβλεπόμενες κλάσεις και οι πιθανότητές τους από όλους τους ταξινομητές. Το νήμα αυτό τυπώνει στην κονσόλα τα αποτελέσματα καθώς και το σε ποιον από τους ταξινομητές ανήκουν. Επίσης, δημιουργεί ένα αρχείο τύπου txt στο οποίο αποθηκεύει αναλυτικότερα όλα αυτά τα δεδομένα για μετέπειτα έλεγχο.

Διάφορα scripts κ.α.

Τα scripts που αναφέρθηκαν είναι αρχεία αποτελούμενα από ακολουθίες γραμμών εντολών (στην περίπτωση της εργασίας μας, για λειτουργικό τύπου Unix (Rasbian)), πράγμα που βοηθάει στον σχεδιασμό αυτοματισμών που χρειάζεται να συμβαίνουν στην συσκευή, όπως για παράδειγμα έναρξη λειτουργιών απαραίτητων για την σωστή λειτουργία του κυρίως προγράμματος που σχεδιάστηκε (όπως η έναρξη του portAudio, βασικού πρωτόκολλου για διαχείριση

του σήματος εισόδου από το μικρόφωνο) ή τον τερματισμό άλλων άχρηστων λειτουργιών για την στιγμή που τρέχει το λογισμικό ταξινόμησης ή προγραμμάτων τα οποία καταναλώνουν πολύτιμους επεξεργαστικούς πόρους στην συσκευή. Τα αρχεία αυτά ποικίλουν ανάλογα τόσο με τον τύπο Unix που χρησιμοποιείται, όσο και με τον τύπο διεργασιών που χρειάζεται να πραγματοποιηθεί. Σε αυτήν την περίπτωση δημιουργούνται αρχεία τύπου *.sh*, scripts δηλαδή για bash εντολές του λειτουργικού συστήματος.

Κεφάλαιο 6

Αποτελέσματα

Στο κεφάλαιο αυτό παρουσιάζονται δύο ειδών αποτελέσματα. Τα πρώτα είναι τα αποτελέσματα της εργασίας μέσω της διαδικασίας που αναλύθηκε στο κεφάλαιο 5 με τα χαρακτηριστικά που αναλύθηκαν (OpenSmile) και τις κλάσεις ανά ταξινομητή. Τα δεύτερα είναι μετρήσεις προβλέψεων, των οποίων η εξαγωγή έγινε χρησιμοποιώντας τα ίδια προγραμματιστικά εργαλεία και τις ίδιες τεχνικές με αυτά που χρησιμοποιήθηκαν για την δική μας βάση πάνω στην βάση αναφοράς.

6.1 RPi Reth

Μετά από όλη την διαδικασία που αναλύθηκε στο κεφάλαιο 5 και τις αντίστοιχες δοκιμές που χρειάστηκαν για να φτάσει το επίπεδο των προβλέψεων να θεωρείται αποδεκτό, παρουσιάζονται οι μετρήσεις που πραγματοποιήθηκαν στα τελικά πειράματα με σκοπό την εύρεση των βέλτιστων αλγορίθμων ταξινόμησης και τις παραμέτρους τους για το πρόβλημα αυτής της εργασίας.

6.1.1 Επιλογή Ταξινομητή και Παραμέτρων

Στους παρακάτω πίνακες παρουσιάζονται τα precision rate (PRE), recall rate (REC), F1 καθώς και η τελική ακρίβεια (ACC) ανά ταξινομητή κατά την διαδικασία της k-fold cross validation εκπαίδευσης, χωρίς την διαδικασία της επιλογής χαρακτηριστικών. Παρακάτω λοιπόν στους πίνακες 6.1, 6.2, 6.3, 6.4, 6.5 και 6.6 βρίσκονται τα αποτελέσματα των τελικών μετρήσεων. Στους πίνακες αυτούς παρουσιάζονται αναλυτικά ανά κατηγορία (τόσο των ηχοτοπίων όσο και των ηχητικών γεγονότων) οι υπολογισμοί των μετρήσεων που αναλύθηκαν στην ενότητα 3.3 ανά τις τιμές των διάφορων παραμέτρων (οι οποίες συμβολίζονται με C). Στους πίνακες αυτούς φαίνεται σε ποιες κατηγορίες υπάρχει η μεγαλύτερη πιθανότητα σωστής πρόβλεψης ενώ παράλληλα διακρίνονται και οι κατηγορίες στις οποίες υπάρχει η χαμηλότερη πιθανότητα. Η επιλογή των παραμέτρων που αποδίδουν καλύτερα για τις προβλέψεις των δικών μας δεδομένων γίνεται με βάση την υψηλότερη τιμή $F1$. Όπως χαρακτηριστικά φαίνεται, για την δική μας βάση ο αλγόριθμος Support Vector Machine αποδίδει καλύτερα απ΄ ότι ο αλγόριθμος KNN. Επίσης, εύκολα διακρίνεται ότι ο αλγόριθμος ταξινόμησης για SoundScapes

λειτουργεί πιο αποδοτικά με τιμή κόστους (όπου συμβολίζεται με την τιμή C για τον αλγόριθμο SVM) ίση με $C = 0.001$ ενώ για τις κατηγορίες Vehs και Other βέλτιστες τιμές κόστους είναι $C = 0.001$ και $C = 0.010$ αντίστοιχα. Αυτές οι τιμές λοιπόν είναι οι τελικές τιμές που θα χρησιμοποιηθούν ως τιμές παραμέτρων για την δημιουργία των τελικών μοντέλων για ταξινόμηση σε πραγματικό χρόνο.

Με τους καινούριους ταξινομητές λοιπόν (με τις βέλτιστες τιμές παραμέτρων) έτοιμους, βγαίνουν τα αποτελέσματα που φαίνονται στους πίνακες σύγχυσης παρακάτω (confusion matrix) [πίνακες 6.7, 6.8, 6.9] για αναλυτικότερη εικόνα των μετρήσεων με τα τελικά μας μοντέλα. Όπως αναφέρθηκε, στους πίνακες αυτούς αναφέρεται πιο αναλυτικά, ποιες κατηγορίες έχουν την καλύτερη απόδοση, αφού φαίνεται ποσοστιαία ποιες κατηγορίες κατηγοριοποιήθηκαν σαν την κλάση που όντως είναι (διαγώνιος από πάνω προς τα κάτω) και σε ποιες κατηγορίες υπάρχει μεταξύ τους σύγχυση, αν μπερδεύτηκαν δηλαδή με άλλη κλάση, και αν ναι, με ποια.

6.1.2 Επιλογή Χαρακτηριστικών

Όπως αναφέρθηκε, για το κομμάτι της επιλογής των χαρακτηριστικών χρησιμοποιήθηκε το λογισμικό WEKA ενώ η μέθοδος βάσει της οποίας έγινε η επιλογή είναι η μέθοδος της ενσωμάτωσης που όπως αναφέρθηκε είναι η επιλογή των χαρακτηριστικών μέσω μιας διαδικασίας ελέγχου για το ποια από αυτά αποδίδουν καλύτερα με τον ταξινομητή με τον οποίο θα δοκιμαστούν. Αυτό γίνεται εφικτό με το λογισμικό WEKA μέσω των έτοιμων υλοποιημένων αλγορίθμων ταξινόμησης που εμπεριέχονται σε αυτό. Εφόσον επιλέχθηκε αλγόριθμος ταξινόμησης που θα χρησιμοποιηθεί για την τελική υλοποίηση θα είναι Support Vector Machine, έγινε επανάληψη της διαδικασίας εύρεσης βέλτιστων χαρακτηριστικών στο λογισμικό (ο αλγόριθμος SVM ονομάζεται SMO στην WEKA) για κάθε μία από τις κατηγορίες διαδικασιών ταξινομητών που θα έχουμε στην δική μας περίπτωση (SoundScapes, Vehs και Other).

Ένα πολύ μεγάλο θέμα που απασχολεί αυτό το κομμάτι της διαδικασίας είναι η αποφυγή του overfitting (βλ. κεφάλαιο 3.4.2), η προσπάθεια δηλαδή επιλογής χαρακτηριστικών τόσο επικεντρωμένων στα δεδομένα των δεδομένων εκπαίδευσης (γνωστά δεδομένα) με αποτέλεσμα το μοντέλο να μην αποδίδει ορθά σε άγνωστα (σε πραγματικά δηλαδή) σενάρια. Αυτό πρακτικά σημαίνει πως εάν εφαρμοστεί η διαδικασία επιλογής των χαρακτηριστικών μόνο σε γνωστά δεδομένα υπάρχει ο κίνδυνος μη σωστής απόκρισης του μοντέλου κατά την δοκιμή σε άγνωστα. Επειδή όμως, όπως εξηγήθηκε στην παράγραφο 2.1.2, δεν υπάρχουν στην σχετική βιβλιογραφία κλάσεις όμοιες με τις κλάσεις της εργασίας μας, υπάρχει δυσκολία να βρεθούν διαφορετικά δεδομένα με σκοπό την αποκλειστική τους χρήση για την εύρεση των βέλτιστων χαρακτηριστικών όπως θα ήταν πιο ορθό, ώστε να χρησιμοποιηθούν τα χαρακτηριστικά αυτά μετά στην δική μας βάση για την εκτίμηση του αν λειτουργούν όπως πρέπει ή όχι. Πρακτικά όμως λόγω της φύσης του συγκεκριμένου αλγορίθμου ταξινόμησης (γραμμικός διαχωρισμός) και του τρόπου που αυτός ελαχιστοποιεί τους περιορισμούς κατά την χωροτοποθέτηση των δεδομένων, έχει το πλεονέκτημα πως μπορεί και αντιστέκεται σε αρκετό βαθμό στο πρόβλημα της υπερφόρτωσης. Παράλληλα, η λογική που έγινε η όλη δόμηση στην εργασία είναι η αλλαγή όσο λιγότερων παραμέτρων γίνεται, από το πρώτο στάδιο που ήταν η διαδικασία των

Πίνακας 6.1: SVM - SoundScapes

	Bus			BusStop			BusyStreet		
C	PRE	REC	F1	PRE	REC	F1	PRE	REC	F1
0.001	97.4	99.3	98.3	82.6	72.7	77.3	71.0	83.3	76.7
0.010	98.0	98.7	98.3	69.5	76.0	72.6	71.0	80.0	75.2
0.500	98.6	97.3	98.0	71.0	73.3	72.1	71.0	75.3	73.1
1.000	97.4	99.3	98.3	73.8	80.7	77.1	76.0	76.0	76.0
5.000	94.3	98.7	96.4	70.3	74.0	72.1	73.6	78.0	75.7
10.00	96.7	97.3	97.0	70.5	78.0	74.1	73.7	74.7	74.2

	CoastialStreet			OpenMarket			Park		
C	PRE	REC	F1	PRE	REC	F1	PRE	REC	F1
0.001	72.6	76.0	74.3	90.8	79.3	84.7	81.6	74.0	77.6
0.010	82.9	77.3	80.0	76.1	82.7	79.2	76.0	74.0	75.0
0.500	82.6	82.0	82.3	79.2	78.7	78.9	75.0	80.0	77.0
1.000	81.2	78.0	79.6	79.5	77.3	78.4	82.1	82.7	82.4
5.000	78.1	80.7	79.3	81.4	78.7	80.0	76.5	78.0	77.2
10.00	79.7	84.0	81.8	80.8	81.3	81.1	79.3	79.3	79.3

	Pedestrian			Port			QuietStreet		
C	PRE	REC	F1	PRE	REC	F1	PRE	REC	F1
0.001	83.5	88.0	85.7	90.8	92.0	91.4	81.3	84.0	82.6
0.010	81.9	75.3	78.5	92.4	89.3	90.8	79.3	71.3	75.4
0.500	80.6	74.7	77.5	89.4	90.0	89.7	80.0	74.7	77.2
1.000	78.1	76.0	77.0	89.0	86.7	87.8	79.3	79.3	79.3
5.000	85.5	74.7	79.7	89.6	86.0	87.8	78.4	77.3	77.9
10.00	80.0	80.0	80.0	92.6	83.3	87.7	82.5	75.3	78.7

	OVERALL	
C	ACC	F1
0.001	83.2	83.2
0.010	80.5	80.6
0.500	80.7	80.7
1.000	81.8	81.8
5.000	80.7	80.7
10.00	81.5	81.5

Best F1 Best ACC

Πίνακας 6.2: KNN - SoundScapes

	Bus			BusStop			BusyStreet		
C	PRE	REC	F1	PRE	REC	F1	PRE	REC	F1
1.000	91.0	94.0	92.5	56.8	47.3	51.6	59.2	60.3	60.6
3.000	86.2	96.0	90.9	55.7	65.3	60.1	60.3	62.7	61.4
5.000	91.1	95.3	93.2	60.5	61.3	60.9	60.6	70.7	65.2
7.000	90.7	97.3	93.9	56.0	62.0	58.9	54.9	63.3	58.8
9.0000	90.6	96.7	93.5	60.6	62.7	61.9	54.8	64.7	59.3
11.000	92.5	98.7	95.5	62.5	70.0	66.0	64.7	73.3	68.7
13.000	92.9	96.0	94.4	61.4	57.3	59.3	53.0	71.3	60.8
15.000	94.8	97.3	96.1	60.3	62.7	61.4	55.8	73.3	63.4

	CoastialStreet			OpenMarket			Park		
C	PRE	REC	F1	PRE	REC	F1	PRE	REC	F1
1.000	69.9	80.7	74.9	70.5	65.3	67.8	64.5	60.7	62.5
3.000	71.1	78.7	74.7	67.9	60.7	64.1	72.0	63.3	67.4
5.000	79.3	79.3	79.3	79.8	66.0	72.3	71.2	74.0	72.5
7.000	81.3	72.7	76.8	71.7	69.3	70.5	70.5	62.7	66.0
9.0000	84.7	70.0	76.6	71.7	69.3	72.0	72.5	74.0	69.2
11.000	86.4	72.0	78.5	70.8	64.7	67.6	67.6	71.3	69.7
13.000	80.3	73.3	76.7	64.4	62.7	63.5	63.5	61.3	62.4
15.000	74.5	68.0	71.1	76.3	66.7	71.2	70.3	68.0	69.2

	Pedestrian			Port			QuietStreet		
C	PRE	REC	F1	PRE	REC	F1	PRE	REC	F1
1.000	52.2	72.0	60.5	85.9	73.3	79.1	64.4	58.0	61.1
3.000	58.6	72.7	64.9	84.4	76.0	80.0	81.6	53.3	69.8
5.000	53.0	77.3	62.9	91.1	74.7	82.1	88.3	55.3	64.5
7.000	54.8	76.7	63.9	91.9	82.7	87.0	83.5	50.7	68.0
9.000	57.3	76.0	65.3	91.9	83.3	87.4	82.6	50.7	63.1
11.00	54.9	74.7	63.3	85.4	82.0	83.7	81.2	46.0	58.7
13.00	57.8	71.3	63.9	85.0	75.3	79.9	81.3	58.0	67.7
15.00	58.4	76.7	66.3	93.0	79.3	85.6	82.9	58.0	68.2

	OVERALL	
C	ACC	F1
1.000	67.7	67.6
3.000	69.9	69.8
5.000	72.7	72.9
7.000	70.8	71.0
9.000	71.9	72.0
11.00	72.5	72.4
13.00	69.6	76.8
15.00	72.2	72.5

Best F1 Best ACC

Πίνακας 6.3: SVM - Vehs

	Car			Motor			OVERALL			
C	PRE	REC	F1	PRE	REC	F1	ACC	F1		
0.001	71.0	78.7	74.7	76.1	67.8	71.7	73.3	73.2	Best F1	Best ACC
0.010	72.3	68.4	70.3	70.0	73.7	71.8	71.1	71.1		
0.500	63.6	63.7	63.7	63.6	63.4	63.5	63.6	63.6		
1.000	65.9	62.8	64.3	64.5	67.5	66.0	65.2	65.1		
5.000	69.1	65.6	67.3	67.3	70.6	68.9	68.1	68.1		
10.00	63.9	63.1	63.5	63.6	64.4	64.0	63.7	63.7		

Πίνακας 6.4: KNN - Vehs

	Car			Motor			OVERALL		
C	PRE	REC	F1	PRE	REC	F1	ACC	F1	
1.00	58.4	62.2	60.2	59.5	55.6	57.5	58.9	58.9	
3.00	65.0	73.7	69.1	69.7	60.3	64.7	67.0	66.9	
5.00	65.4	73.1	69.0	69.5	61.2	65.1	67.2	67.1	Best F1
7.00	61.9	70.0	65.7	65.5	56.9	60.9	63.4	63.3	
9.00	63.7	74.1	69.0	69.0	57.8	62.9	65.9	65.7	
11.00	61.3	74.1	67.2	67.5	53.1	59.4	63.7	63.3	
13.00	62.5	78.7	69.7	71.3	52.8	60.7	65.8	65.2	
15.00	64.5	77.2	70.3	70.3	57.5	63.8	67.3	67.0	Best ACC

Πίνακας 6.5: SVM - Other

	Bird			Dog			Horn		
C	PRE	REC	F1	PRE	REC	F1	PRE	REC	F1
0.001	55.7	68.0	61.3	94.1	64.0	76.2	79.6	78.0	78.8
0.010	67.9	76.0	71.7	85.7	84.0	84.8	93.5	86.0	83.9
0.500	60.6	80.0	69.0	88.6	78.0	83.0	90.7	78.0	83.9
1.000	63.3	76.0	69.1	91.5	86.0	88.7	83.7	82.0	82.8
5.000	55.6	70.0	61.9	81.2	78.0	79.6	91.7	88.0	89.8
10.00	62.1	72.1	66.7	85.1	80.0	82.5	97.6	82.0	89.1

	Speech			OVERALL	
C	PRE	REC	F1	ACC	F1
0.001	67.9	76.0	71.7	71.5	72.0
0.010	67.3	66.0	66.7	78.0	78.2
0.500	70.2	66.0	68.0	75.5	76.0
1.000	75.0	66.0	70.2	77.5	77.7
5.000	63.4	52.0	57.1	72.0	72.1
10.00	66.0	70.0	68.0	76.0	76.6

Best F1 Best ACC

ηχογραφήσεων μέχρι και το τελικό, το οποίο είναι το πρακτικό λειτουργικό κομμάτι (ίδιο μικρόφωνο, ίδιος ψηφιακός μετατροπέας, ίδια μητρική πλακέτα), με σκοπό στο τελικό κομμάτι να υπάρχουν ίδιας φύσης δεδομένα με αυτά που αρχικά μελετήθηκαν σε όλα τα προηγούμενα στάδια.

Αυτό μπορεί να διαπιστωθεί με μια πειραματική διαδικασία, ώστε να παρθεί υπόψιν το κατά πόσο το overfitting επηρεάζει την ταξινόμηση. Για το πείραμα αυτό έγινε αφαίρεση 10 δοκιμαστικών αρχείων από κάθε κατηγορία μέσα από το σύνολο δεδομένων της κατηγορίας Soundscape. Στα αρχεία που παραμένουν χρησιμοποιήθηκε ο αλγόριθμος για την εύρεση των βέλτιστων χαρακτηριστικών. Έγινε διαμόρφωση του προγράμματος που χρησιμοποιείται στην διαδικασία της εκτίμησης (ενότητα 5.2) ώστε να επιλέγει μόνο τα χαρακτηριστικά αυτά κατά την πρόβλεψη (τα οποία υπάρχουν σε μορφή θέσεων στον πίνακα των χαρακτηριστικών και άρα η επιλογή τους προγραμματιστικά είναι απλή). Τέλος, χρησιμοποιούνται τα αρχεία που αρχικά αφαιρέθηκαν, τα οποία σε αυτή τη φάση είναι άγνωστα δεδομένα, μιας και δεν χρησιμοποιήθηκαν στην διαδικασία της επιλογής των χαρακτηριστικών με την μέθοδο της ενσωμάτωσης. Η διαδικασία επιλογής των χαρακτηριστικών επαναλαμβάνεται, αυτήν την φορά όμως χωρίς την αφαίρεση αρχείων από το σύνολο και ελέγχονται τα ίδια 10 από κάθε κατηγορία αρχεία σαν γνωστά δεδομένα σε αυτή τη φάση, έχοντας τα καινούρια χαρακτηριστικά που επιλέχθηκαν. Στην πρώτη περίπτωση επιλέχθηκαν 25 χαρακτηριστικά ενώ στην δεύτερη 35. Τα αποτελέσματα φαίνονται στον πίνακα 6.11.

Όπως φαίνεται λοιπόν, αν και η δεύτερη περίπτωση, με τη πρόβλεψη δηλαδή γνωστών

Πίνακας 6.6: KNN - Other

	Bird			Dog			Horn		
C	PRE	REC	F1	PRE	REC	F1	PRE	REC	F1
1.00	38.3	62.0	47.3	77.1	54.0	63.5	87.0	40.0	54.8
3.00	38.8	80.0	52.3	84.8	56.0	67.5	87.5	14.0	24.1
5.00	34.0	66.0	44.9	78.6	66.0	71.7	100	24.0	38.7
7.00	38.3	72.0	50.0	91.5	86.0	88.7	90.0	18.0	30.0
9.00	32.4	68.0	43.9	94.3	66.0	77.6	100	16.0	27.6
11.00	33.3	72.0	45.6	88.6	62.0	72.9	100	2.0	3.9
13.00	30.4	66.0	42.0	88.9	64.0	74.0	100	8.0	14.8
15.00	37.0	80.0	50.6	86.7	52.0	65.0	100	10.0	18.2

	Speech			OVERALL		
C	PRE	PRE	F1	ACC	F1	
1.00	52.5	64.0	57.7	55.0	55.8	
3.00	60.7	68.0	64.2	54.5	52.0	
5.00	69.4	68.0	68.7	56.0	56.0	
7.00	63.3	62.0	62.6	59.5	57.8	Best F1 Best ACC
9.00	67.7	60.0	60.0	52.5	52.8	
11.00	62.5	70.0	66.0	51.5	47.1	
13.00	47.9	46.0	46.0	46.9	44.5	
15.00	52.6	60.0	56.1	50.0	47.5	

Πίνακας 6.7: Soundscapes Matrix

Πίνακας 6.8: Vehs Matrix

	Car	Motor
Car	239	81
Motor	92	228

Πίνακας 6.9: Other Matrix

	Bird	Dog	Horn	Speech
Bird	38	3	4	5
Dog	4	44	0	2
Horn	7	1	4	2
Speech	9	9	2	3

Πίνακας 6.10: Soundscape Matrix

	Bus	BusStop	BusyStreet	CoastialStreet	OpenMarket	Park	Pedestrian	Port	QuietStreet
Bus	150	0	0	0	0	0	0	0	0
BusStop	1	119	20	8	0	0	0	2	0
CoastialStreet	1	14	119	6	1	0	3	0	6
OpenMarket	1	7	9	130	0	0	3	0	0
Park	0	0	1	2	117	6	16	5	3
Pedestrian	0	1	3	7	16	3	119	0	1
Port	2	1	2	7	2	3	0	126	7
QuietStreet	1	5	2	0	1	16	0	3	122

Πίνακας 6.11: Επιλογή Χαρακτηριστικών: Απόδοση Γνωστών και Αγνώστων δεδομένων

	With Unknown Data			With Known Data		
	Pre	Rec	F1	Pre	Rec	F1
Bus	100	80	88.9	100	90	94.7
BusStop	46.2	60	52.2	46.7	70	56
BusyStreet	61.5	80	69.6	83	100	90.9
CoastialStreet	77.8	70	73.7	100	30	46.2
OpenMarket	90.9	100	95.2	100	100	100
Park	100	90	94.7	75	90	81.8
Pedestrian	90	90	90	88.9	80	84.2
Port	90.9	100	95.2	75	90	81.8
QuietStreet	83.3	50	62.5	100	80	88.9
		ACC	F1		ACC	F1
Overall		80	80.2		81.1	80.5

δεδομένων, αποδίδει λίγο καλύτερα, η διαφορά τους είναι τόσο αμελητέα που δεν παίζει ρόλο το εάν υπάρχει υπερκάλυψη ή όχι, μιας που το κέρδος σε επεξεργαστική ισχύ (άρα και ταχύτητα / φόρτο εργασίας) με τόσο μεγάλη μείωση στον αριθμό των χαρακτηριστικών είναι αρκετά μεγάλο. Το σημαντικό κομμάτι σε αυτή τη διαδικασία είναι η μεγάλη μείωση που έγινε στις διαστάσεις του κάθε μοντέλου (στην περίπτωση των ηχοτοπίων, από 989 σε 35), κρατώντας χονδρικά ίδια απόδοση πρόβλεψης σε παρόμοια πλαίσια.

Με αυτήν λοιπόν την λογική γίνεται χρήση όλου του συνόλου για την επιλογή των χαρακτηριστικών, με το οποίο κερδήθηκαν κατά την διαδικασία αυτή όσο το δυνατό πιο κατατοπισμένα χαρακτηριστικά. Τα αποτελέσματα των μετρήσεων με τους τελικούς ταξινομητές μετά την επιλογή των χαρακτηριστικών φαίνονται στα σχήματα 6.12 6.13 και 6.14

	SVM Soundscapes		
	Pre	Rec	F1
Bus	98	99.3	98.7
BusStop	74.5	76	75.2
BusyStreet	68.3	86.0	76.1
CoastialStreet	81.9	69.3	75.1
OpenMarket	84.2	82	83.1
Park	80.1	75.3	77.7
Pedestrian	83	81.3	82.2
Port	88.2	89.3	88.7
QuietStreet	81.1	77.3	79.2
		ACC	F1
Overall		81.8	81.8

Πίνακας 6.12: Soundscapes

	SVM Others		
	Pre	Rec	F1
Bird	62.7	74	67.96
Dog	88.1	74	80.4
Horn	88	88	88
Speech	79.6	780	78.8
		ACC	F1
Overall		78.5	78.8

Πίνακας 6.13: Others

	SVM Vehs		
	Pre	Rec	F1
Car	61.3	75.6	67.7
Motor	68.2	52.2	59.1
		ACC	F1
Overall		63.9	63.4

Πίνακας 6.14: Vehs

Πίνακας 6.15: Vehs Matrix after feature selection

	Car	Motor
Car	242	78
Motor	153	167

Πίνακας 6.16: Other Matrix after feature selection

	Bird	Dog	Horn	Speech
Bird	73	2	5	6
Dog	10	37	0	3
Horn	4	1	44	1
Speech	8	2	1	39

Πίνακας 6.17: Soundscape Matrix after feature selection

	Bus	BusStop	BusyStreet	CoastialStreet	OpenMarket	Park	Pedestrian	Port	QuietStreet
Bus	149	1	0	0	0	0	0	0	0
BusStop	0	114	20	11	0	1	2	2	0
BusyStreet	1	9	129	8	0	0	1	1	1
CoastialStreet	1	14	21	104	0	4	4	2	0
OpenMarket	0	1	1	1	123	6	8	6	4
Park	0	0	2	0	12	113	3	0	20
Pedestrian	1	2	8	3	10	2	122	2	0
Port	0	6	1	0	1	4	2	134	2
QuietStreet	0	6	7	0	0	11	5	5	116

Τα χαρακτηριστικά που χρησιμοποιούνται για κάθε κατηγορία είναι:

- Soundscape
  - pcm_loudness_sma_quartile1
  - mfcc_sma[2]_kurtosis
  - mfcc_sma[3]_quartile3
  - mfcc_sma[4]_quartile1
  - mfcc_sma[5]_quartile1
  - mfcc_sma[6]_quartile1
  - mfcc_sma[6]_quartile2
  - mfcc_sma[9]_amean
  - mfcc_sma[9]_stddev
  - mfcc_sma[9]_iqr1-2
  - mfcc_sma[11]_amean
  - mfcc_sma[12]_quartile3
  - lspFreq_sma[1]_linregc2
  - lspFreq_sma[1]_quartile3
  - lspFreq_sma[2]_linregerrA
  - lspFreq_sma[5]_stddev
  - lspFreq_sma[6]_linregerrQ
  - lspFreq_sma[6]_iqr2-3
  - lspFreq_sma[7]_quartile1
  - pcm_zcr_sma_linregerrA
  - pcm_zcr_sma_stddev
  - pcm_intensity_sma_de_linregc2
  - pcm_intensity_sma_de_linregerrQ
  - pcm_loudness_sma_de_iqr1-2
  - mfcc_sma_de[2]_linregerrA
  - mfcc_sma_de[2]_linregerrQ
  - mfcc_sma_de[10]_kurtosis
  - mfcc_sma_de[12]_iqr1-2
  - lspFreq_sma_de[1]_quartile1
  - lspFreq_sma_de[4]_skewness
  - lspFreq_sma_de[5]_quartile2
  - lspFreq_sma_de[6]_skewness
  - lspFreq_sma_de[7]_quartile1
  - pcm_zcr_sma_de_linregerrQ
  - voiceProb_sma_de_linregerrQ
- Vehs
  - mfcc_sma[1]_linregc1
  - mfcc_sma[2]_linregc2
  - mfcc_sma[3]_linregerrA
  - mfcc_sma[5]_quartile3
  - mfcc_sma[6]_linregc2
  - mfcc_sma[8]_linregc2
  - mfcc_sma[8]_iqr1-3
  - lspFreq_sma[7]_linregerrQ
  - mfcc_sma_de[2]_linregc1
  - mfcc_sma_de[7]_iqr2-3
  - mfcc_sma_de[12]_stddev
  - lspFreq_sma_de[7]_amean
  - pcm_zcr_sma_de_skewness
  - F0_sma_de_skewness

- Others
  - pcm_loudness_sma_quartile1
  - mfcc_sma[2]_minPos
  - mfcc_sma[3]_min
  - mfcc_sma[5]_quartile3
  - mfcc_sma[8]_stddev
  - mfcc_sma[12]_min
  - lspFreq_sma[1]_iqr2-3
  - lspFreq_sma[3]_amean
  - mfcc_sma_de[8]_max
  - lspFreq_sma_de[4]_max
  - lspFreq_sma_de[4]_skewness
  - F0_sma_de_linregc1
  - F0env_sma_de_iqr1-3

6.2 Reference Data Set

Στο στάδιο αυτό παρουσιάζονται τα αποτελέσματα από την ίδια διαδικασία, αυτήν την φορά με την βάση αναφοράς (βλ. ενότητα 2.2). Λόγω της καλύτερης απόδοσης μέσω SVM στις προηγούμενες βάσεις έγινε δοκιμή μόνο με αυτήν την μέθοδο και στην παρούσα βάση. Αναλυτικότερα, τα αποτελέσματα με όλες τις παραμέτρους (όπως ακριβώς και στις προηγούμενες βάσεις) παρουσιάζονται στον πίνακα 6.18, ενώ ο πίνακας σύγχυσης, μετά από επανάληψη της εκπαίδευσης με τιμή παραμέτρου την βέλτιστη, παρουσιάζεται στο σχήμα 6.19.

Σε αυτό το στάδιο υπάρχει η δυνατότητα (και αυτός ήταν ένας από τους λόγους για την επιλογή αυτής της βάσης ως βάση σύγκρισης) για έλεγχο καθαρών ηχητικών γεγονότων και επικαλυπτόμενων (ή όπως ονομάζονται στην σχετική έρευνα για την δημιουργία της βάσης, clean και mix). Αυτό δίνει μια εικόνα για το ποσοστό βελτίωσης που θα είχε η εργασία μας στο ποσοστό επιτυχημένης πρόβλεψης εάν είχε τύχει να ηχογραφηθούν περισσότερα καθαρά γεγονότα κατά την διάρκεια των ηχογραφήσεων. Στα πλαίσια αυτής της εργασίας δεν υπήρχε η δυνατότητα ηχογράφησης των συγκεκριμένων γεγονότων σε καθαρό περιβάλλον και η περίπτωση του να συμβεί αυτό στον εξωτερικό χώρο των ηχογραφήσεων, να μην υπάρχει δηλαδή καθόλου επικάλυψη στα γεγονότα, γίνεται αλλά είναι καθαρά θέμα τύχης. Αυτό ελέγχεται πάλι με πείραμα με την βοήθεια του λογισμικού WEKA. Στον πίνακα 6.20 παρουσιάζονται τα αποτελέσματα τεσσάρων περιπτώσεων. Αρχικά, υπάρχει το ποσοστό επιτυχημένης πρόβλεψης στα καθαρά γεγονότα. Έπειτα, το ποσοστό επιτυχημένης πρόβλεψης στα επικαλυπτόμενα αλλά μέσω της χρήσης του μοντέλου που δημιουργήθηκε από τα καθαρά. Τέλος, υπάρχουν τα ποσοστά επιτυχίας των ίδιων περιπτώσεων αλλά αυτήν την φορά μέσω επιλογής χαρακτηριστικών (μετά από ακολουθία της ίδιας διαδικασίας που περιγράφηκε προηγουμένως για την επιλογή), όπου και για τις δύο αυτές περιπτώσεις η επιλογή των βέλτιστων έγινε βάσει της πρώτης βάσης με τα καθαρά γεγονότα.

Τέλος, στον πίνακα 6.21 παρουσιάζεται ο πίνακας σύγχυσης της τελευταίας διαδικασίας (επικαλυπτόμενα γεγονότα με επιλεγμένα χαρακτηριστικά μέσα από μοντέλο εκπαιδευμένο για καθαρά), ο οποίος έχει και το μεγαλύτερο ενδιαφέρον.

Πίνακας 6.18: SVM - Reference

	Balon			Dientes			Grillo		
C	PRE	REC	F1	PRE	REC	F1	PRE	REC	F1
0.001	100.0	100.0	100.0	97.6	100.0	98.8	100.0	97.5	98.8
0.010	100.0	100.0	100.0	94.7	90.0	92.3	100.0	95.0	97.4
0.500	100.0	100.0	100.0	94.4	85.0	89.5	97.6	100.0	98.8
1.000	100.0	100.0	100.0	92.9	97.5	95.1	100.0	100.0	100.0
5.000	100.0	100.0	100.0	97.2	87.5	92.1	100.0	97.5	98.7
10.00	100.0	100.0	100.0	94.3	82.5	88.0	97.5	97.5	

	Llanto			Llaves			Manos		
C	PRE	REC	F1	PRE	REC	F1	PRE	REC	F1
0.001	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.010	100.0	100.0	100.0	97.6	100.0	98.8	93.0	100.0	96.4
0.500	100.0	100.0	100.0	90.7	97.5	94.0	95.0	9530	95.0
1.000	100.0	100.0	100.0	95.1	97.5	96.3	100.0	92.5	96.1
5.000	100.0	100.0	100.0	95.2	100.0	97.6	93.0	100.0	96.4
10.00	100.0	100.0	100.0	93.0	100.0	96.4	92.9	97.5	95.1

	Tecleado			OVERALL		
C	PRE	REC	F1	ACC	F1	
0.001	100.0	100.0	100.0	99.6	99.6	Best F1 Best ACC
0.010	100.0	100.0	100.0	97.9	97.8	
0.500	100.0	100.0	100.0	96.8	96.7	
1.000	100.0	100.0	100.0	98.2	98.2	
5.000	100.0	100.0	100.0	97.9	97.8	
10.00	100.0	100.0	100.0	96.8	96.7	

- Reference Data Set
  - mfcc_sma[1]_min
  - mfcc_sma[3]_quartile1
  - mfcc_sma[5]_quartile2
  - mfcc_sma[9]_amean
  - lspFreq_sma[6]_min
  - pcm_zcr_sma_linregerrQ
  - voiceProb_sma_linregerrQ
  - mfcc_sma_de[11]_stddev
  - lspFreq_sma_de[2]_linregerrQ
  - lspFreq_sma_de[4]_stddev

Πίνακας 6.19: Reference Matrix

	Balon	Dientes	Grillo	Llanto	Llaves	Manos	Tecleado
Balon	40	0	0	0	0	0	0
Dientes	0	38	1	0	0	1	0
Grillo	0	1	39	0	0	0	0
Llanto	0	0	0	40	0	0	0
Llaves	0	2	0	0	38	0	0
Manos	0	1	0	0	1	38	0
Tecleado	0	0	0	0	0	0	40

Πίνακας 6.20: Ποσοστό Επίδραση Επικάλυψης

	Clean	Mix	Clean - FS	Mix - FS
ACC	97.5	88.5	91.4	81.5
PRE	97.5	89.3	91.4	81.8
REC	97.5	88.6	91.4	81.6
F1	97.5	88.5	91.2	81.3

Πίνακας 6.21: Reference Matrix with feature selection

	Balon	Dientes	Grillo	Llanto	Llaves	Manos	Tecleado
Balon	45	3	3	2	2	0	5
Dientes	0	41	1	6	0	5	7
Grillo	0	0	52	2	0	2	0
Llanto	6	0	0	46	0	0	0
Llaves	0	0	0	0	60	0	0
Manos	0	9	3	0	5	43	4
Tecleado	0	8	0	3	0	0	49

Κεφάλαιο 7

Επίλογος

7.1 Συμπεράσματα

Μετά από όλα τα πειράματα που πραγματοποιήθηκαν γίνεται κατανοητό πως το αποτελέσματα της ακρίβειας από όλους τους ταξινομητές είναι αρκετά ικανοποιητικό. Όπως φαίνεται στα διαγράμματα σύγχυσης, αρκετές κατηγορίες μπλέκονται μεταξύ τους, πράγμα που οφείλεται στα όμοια μορφολογικά στοιχεία των ηχητικών κατηγοριών (τόσο στα ηχοτοποία, όσο και στα ηχητικά γεγονότα). Αυτό έχει σε μεγάλο βαθμό να κάνει με τα ίδια τα μορφολογικά χαρακτηριστικά της πόλης του Ρεθύμνου όπου έγινε η συλλογή της βάσης των δεδομένων. Το Ρέθυμνο, αν και μικρή πόλη, έχει πολύ μεγάλη ηχητική ποικιλομορφία λόγω του ότι συνδυάζει διαφορετικά ηχητικά στοιχεία, π.χ. σε περιοχές παραλίας/θάλασσας, περιοχές της παλιάς πόλης (όπου επικρατεί περισσότερη ηχητική ηρεμία) και ταυτόχρονα περιέχει ηχητικά στοιχεία πόλεως, όπως μεγάλους δρόμους με αρκετή κίνηση τόσο από πεζούς όσο και από οχήματα.

Επειδή σε αυτήν την εργασία σκοπός ήταν η δημιουργία υλοποίησης ενός αντικειμένου το οποίο να δουλεύει, με τελικό σκοπό την υλοποίηση του κόμβου με το Raspberry Pi και δεν δόθηκε καθόλου σημασία στις κατηγορίες που θα γινόταν ταξινόμηση με αυτόν τον κόμβο. Απλά έγινε προσπάθεια εύρεσης του τι υπάρχει στην πόλη και μπορούσε να χρησιμοποιηθεί σαν κλάση. Τα αποτελέσματά της εργασίας, αν και αρκετά ικανοποιητικά για την διαδικασία, θα μπορούσαν να ήταν ακόμα καλύτερα αν είχε δοθεί λίγη περισσότερη προσοχή στην οργάνωση του συνόλου δεδομένων κατά τα αρχικά στάδια με πιο προσεκτικό σχεδιασμό των κατηγοριών στις οποίες μετέπειτα θα γινόταν η ταξινόμηση. Ακόμα, θα μπορούσε να είχε οργανωθεί η διαδικασία της κατάτμησης αλλιώς και να είχε δοθεί περισσότερη προσοχή στα ίδια τα γεγονότα και στο περιεχόμενό τους, το οποίο πάντα εξαρτάται από το ίδιο το γεγονός. Επιπλέον, όπως αναφέρθηκε, λόγω της φύσης των ηχογραφήσεων που πραγματοποιήθηκαν δεν υπήρχαν *καθαρές* κατηγορίες και πολύ μεγάλο μέρος του υλικού αποτελείται από επικαλυπτόμενα στοιχεία (το ένα μέσα στο άλλο). Αυτό οδήγησε την διαδικασία της εργασίας στο να μην είναι όσο ακριβής όσο θα έπρεπε να είναι σε αυτό το στάδιο.

Όσο αφορά την βάση σύγκρισης, όπως φαίνεται τα αποτελέσματα είναι λίγο καλύτερα από αυτά της βάσης μας. Αυτό συμβαίνει γιατί τα τμήματα ήχου της βάσης αυτής, την οποία και χρησιμοποιήσαμε για την εκπαίδευση, είναι καθαρά, χωρίς δηλαδή να περιέχουν θόρυβο (σε αντίθεση δηλαδή με την δική μας βάση) και όπως φαίνεται και στο πείραμα των επικαλυπτόμενων γεγονότων, σε αυτήν την περίπτωση τα αποτελέσματα πλησιάζουν τα δικά μας. Ταυτόχρονα, η ίδια η ποικιλομορφία των γεγονότων προς αναγνώριση της βάσης (κλάμα, πληκτρολόγηση, ήχος κλειδιών κ.α.) επιτρέπει τον πιο εύκολο διαχωρισμό αυτών των γεγονότων. Παρόλα αυτά σε όλες τις περιπτώσεις τα αποτελέσματα είναι αρκετά ικανοποιητικά για τα πλαίσια της ηχητικής μηχανικής μάθησης.

Εν κατακλείδι, αν και αρκετά σημεία καθ' όλη την διαδικασία θα μπορούσαν να είχαν γίνει αλλιώς, τα αποτελέσματα εξαγωγής των ταξινομητών μας είναι αρκετά ικανοποιητικά, παίρνοντας υπόψιν μάλιστα αρκετά τμήματα ήχου τόσο κατηγοριών ηχοτοπίων όσο και γεγονότων, τα οποία δε μπορεί να τα ταξινομήσει εύκολα κανείς ούτε με το αυτί. Παράλληλα είναι γνωστό πως ούτε ο άνθρωπος έχει 100% επιτυχία στην αναγνώριση ήχων. Η υλοποίηση του κόμβου μας έγινε με οικονομικά υλικά όπως είχε αρχικά σχεδιαστεί και λειτουργεί με πολύ μεγάλη αποδοτικότητα και με χαμηλή καταλάνωση ρεύματος. Παράλληλα, όλα τα στάδια της διαδικασίας είχαν ως σκοπό την μελλοντική βελτίωσή τους και γι αυτό ο αλγόριθμος είναι χωρισμένος σε βήματα (διαδικασίες), εύχρηστος (για χρήση από 3ους) και εύκολα επεκτάσιμος. Με το τέλος της εργασίας αυτής δηλαδή υπάρχει διαθέσιμη για μελλοντική έρευνα η βάση που σχεδιάστηκε και ηχογραφήθηκε καθώς και για οποιαδήποτε άλλη εφαρμογή περιλαμβάνει ηχητική αναγνώριση όπως σε project έξυπνης πόλης κ.α.

Βιβλιογραφία

[1] R. Murray Schafer. *The Soundscape*. Destiny Books, 2018.

[2] R. Murray Schafer, 1969.

[3] Κατερίνα Δ. Σχοινά. Περίπατος σε αστικά ηχοτοπία: μια διδακτική πρόταση, 2015.

[4] Ναπολέων Λαπαθιώτης. *Κάπου περνούσε μια φωνή*. Ερατώ, 2016.

[5] Rodolfo Bonnin. *Machine Learning for Developers*. Packt Publishing, 2018.

[6] Pratap Dangeti. *Statistics for Machine Learning*. Packt Publishing, 2018.

[7] George R. Doddington Daniel Garcia Romero John J. Godfrey Tomi Kinnunen Alvin F. Martin Alan McCree Mark Przybocki Douglas A. Reynolds Craig S. Greenberg, Désiré Bansé. The NIST 2014Speaker Recognition i-Vector Machine Learning Challenge, 2014.

[8] D. A. Reynolds W. M. Campbell, D. E. Sturim. Support Vector Machines using GMM Supervectors for Speaker Verification, 2006.

[9] Martin Woellmer Bjoern Schuller Florian Eyben, Felix Weninger. The Munich Versatile and Fast Open-Source Audio Feature Extractor. `https://www.audeering.com/technology/opensmile/`, 2010.

[10] Tim O'Brien. Learning to understand music from Shazam. `https://blog.shazam.com/learning-to-understand-music-from-shazam-56a60788b62f`, 2017.

[11] P. Cousin C. Pham. Streaming the Sound of Smart Cities: Experimentations on the SmartSantander test-bed, 2013.

[12] Acoustic monitoring in Smart Cities - Research project EarIt. `https://www.idmt.fraunhofer.de/en/Press_and_Media/insight_into_our_research/insight_earit.html`, 2012.

[13] Rosa Ma Alsina-Pagès,Joan Navarro, Francesc Alías, Marcos Hervás. homeSound: Real-Time Audio Event Detection Based on High Performance Computing for Behaviour and Surveillance Remote Monitoring, 2017.

[14] Eric W.M. Yu Cheung-Fat Chan. AN ABNORMAL SOUND DETECTION AND CLASSIFICATION SYSTEM FOR SURVEILLANCE APPLICATIONS , 2010.

[15] James Lyons. Mel Frequency Cepstral Coefficient (MFCC) tutorial. `http://www.practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs`, χ.χ.

[16] HOBNET: Holistic Platform Design for Smart Buildings of the Future Internet. `http://www.cti.gr/el/activities-el/research-projects/item/51-hobnet/51-hobnet`, 2010.

[17] Florian Eyben, Bernd Huber, Erik Marchi, Dagmar Schuller, Björn Schuller. Real-time robust recognition of speakers' emotions and characteristics on mobile platforms, 2015.

[18] Stowell D. , Plumbley MD. Audio-only bird classification using unsupervised feature learning, 2014.

[19] Ellis, Dan and Plumbley, Mark and D. Virtanen, Tuomas. Computational analysis of sound scenes and events, 2018.

[20] Datasets Environmental sounds. `http://www.cs.tut.fi/~heittolt/datasets`, 2016.

[21] Jesús Favela Beltrán-Márquez, Edgar Chávez. allSounds. `"http://sound.natix.org/databases/allSounds.zip"`, 2012.

[22] Jessica Beltrán and Edgar Chávez and Jesús Favela. Scalable identification of mixed environmental sounds, recorded from heterogeneous sources, 2015.

[23] DARES-G1 database of annotated real-world everyday sounds. `https://www.researchgate.net/publication/228726507_DARES-G1_Database_of_Annotated_Real-world_Everyday_Sounds`, 2009.

[24] Annamaria Mesaros, Toni Heittola και Kalle Palomäki. Analysis of acoustic-semantic relationship for diversely annotated real-world audio data, 2013.

[25] Annamaria Mesaros, Toni Heittola και Kalle Palomäki. Query-by-example retrieval of sound events using an integrated similarity measure of content and label, 2013.

[26] Karol J. Piczak. ESC: Dataset for environmental sound classification, 2015.

[27] Karol J. Piczak. Environmental sound classification with convolutional neural networks, 2015.

[28] Upc-talp database of isolated meeting-room acoustic events. `http://catalog.elra.info/en-us/repository/browse/ELRA-S0268/`, 2008.

[29] Marco Maass Radoslaw Mazur Alfred Mertins Huy Phan, Lars Hertel. Audio phrases for audio event recognition, 2015.

[30] D. Giannoulis and E. Benetos and D. Stowell and M. D. Plumbley. IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events - Public Dataset for Scene Classification Task, 2012.

[31] Janvier Maxime and Xavier Alameda Pineda Laurent Girin and Radu Horaud. Sound Representation and Classification Benchmark for Domestic Robots, 2014.

[32] Pasquale Foggia, Nicolai Petkov, Alessia Saggese, Nicola Strisciuglio and Mario Vento. Reliable detection of audio events in highly noisy environments, 2015.

[33] An open dataset for research on audio field recording archives: freefield1010. Dan Stowell, Mark D. Plumbley, 2013.

[34] Nikolaos Malandrakis, Shiva Sundaram, Alexandros Potamianos. Affective Classification of Generic Audio Clips using Regression Models, 2013.

[35] Samuel Kim, Panayiotis Georgiou, Shrikanth Narayanan. Supervised acoustic topic model with a consequent classifier for unstructured audio classification, 2012.

[36] TUT-SED Synthetic 2016 Synthetic dataset for sound event detection research. `http://www.cs.tut.fi/sgn/arg/taslp2017-crnn-sed/tut-sed-synthetic-2016`, 2016.

[37] TUT Rare sound events, Development dataset. `https://zenodo.org/record/401395#.W9muKlVfi7A`, 2017.

[38] Σπύρος Ι. Λουτρίδης. *Ηλεκτροακουστική & Ηχητικές Εγκαταστάσεις*. Εκδόσεις ίων, 2009.

[39] Stefan Sjogelid. *Raspberry Pi for Secret Agents - Second Edition*. Packt Publishing, 2016.

[40] Paul Boersma, David Weenink. Praat. `http://www.fon.hum.uva.nl/praat/`, 2011.

[41] Tim Mahrt, Hiroshi Seresh. Praatio. `{https://pypi.org/project/praatio/}`, 2014.

[42] Understanding Support Vector Machine algorithm from examples (along with code). `https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/`, 2017.

[43] Support Vector Machines. `http://scikit-learn.org/stable/modules/svm.html`, 2007.

[44] Introduction to k-Nearest Neighbors: Simplified (with implementation in Python). `https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/`, 2018.

[45] Nearest Neighbors. http://scikit-learn.org/stable/modules/neighbors.html, 2007.

[46] The Random Forest Algorithm. https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd, 2018.

[47] Prince Grover. Gradient boosting from scratch. https://medium.com/mlreview/gradient-boosting-from-scratch-1e317ae4587d, 2017.

[48] Pierre Geurts, Damien Ernst, Louis Wehenkel. Extremely randomized trees, 2006.

[49] A beginner's guide to neural networks and deep learning. https://skymind.ai/wiki/neural-network, χ.χ.

[50] Vishal Maini. Machine learning for humans, part 4: Neural networks & deep learning. https://medium.com/machine-learning-for-humans/neural-networks-deep-learning-cdad8aeae49b, 2017.

[51] Andrea Trevino. Introduction to k-means clustering. https://www.datascience.com/blog/k-means-clustering, 2016.

[52] Dan Pelleg, Andrew Moore. X-means: Extending K-means with Efficient Estimation of the Number of Clusters, 2002.

[53] Jason Brownlee. An introduction to feature selection. https://machinelearningmastery.com/an-introduction-to-feature-selection/, 2014.

[54] Saurav Kaushik. Introduction to feature selection methods with an example (or how to select the right variables?). https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables/, 2016.

[55] Harry Theodor Nyquist, 1976.

[56] Gregory Rice Lajos Horvath. An introduction to functional data analysis and a principal component approach for testing the equality of mean curves, 2015.

[57] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort and Vincent Michel. scikit-learn. "http://scikit-learn.org/stable/", 2017.

[58] Davide Albanese, Giuseppe Jurman, Stefano Merler, Roberto Visintainer, Marco Chierici, Lance Hepler. mlpy - Machine Learning Python. http://mlpy.sourceforge.net/, 2012.

[59] Theodoros Giannakopoulos. pyAudioAnalysis. "https://github.com/tyiannak/pyAudioAnalysiss", 2015.

[60] Bob Durrant, Eibe Frank, Lyn Hunt, Geoff Holme,s Mike Mayo, Bernhard Pfahringer Tony Smith Ian Witten . Weka 3: Data Mining Software in Java. `https://www.cs.waikato.ac.nz/ml/weka/`, χ.χ.