

Αφιερωμένο στον φίλο μου Μάνο που έφυγε νωρίς

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά την κα Ελένη Βλαχογιάννη Επίκουρη καθηγήτρια της Σχολής Πολιτικών Μηχανικών, για την εμπιστοσύνη που μου έδειξε με την ανάθεση της παρούσας Διπλωματικής Εργασίας, για την αμέριστη υποστήριξη της καθώς και την εκπληκτική συνεργασία που είχαμε σε όλα τα στάδια της μελέτης. Επιπρόσθετα οφείλω να την ευχαριστήσω για την καθοδήγηση και τις γενικότερες γνώσεις που μου μετέδωσε.

Παράλληλα, θα ήθελα να ευχαριστήσω τα υπόλοιπα μέλη της εξεταστικής επιτροπής: κ. Ι. Γκόλια, Πρύτανη του Εθνικού Μετσόβιου Πολυτεχνείου και Καθηγητή της Σχολής Πολιτικών Μηχανικών ΕΜΠ και κ. Γ. Γιαννή Καθηγητή της Σχολής Πολιτικών Μηχανικών ΕΜΠ, για τις εποικοδομητικές παρατηρήσεις τους αλλά κυρίως για τις γνώσεις που μου προσέφεραν καθ' όλη τη διάρκεια της φοίτησής μου στη σχολή.

Ιδιαίτερα θα ήθελα να ευχαριστήσω τους καθηγητές της Σχολής Πολιτικών Μηχανικών και ιδιαίτερα του Τομέα Μεταφορών και Συγκοινωνιακής Υποδομής για τις γνώσεις που μου μεταλαμπάδευσαν καθ' όλη την διάρκεια της φοίτησης μου στη σχολή αλλά κυρίως την βοήθεια τους στο να αγαπήσω περισσότερο το αντικείμενο που διάλεξα να σπουδάσω.

Ένα μεγάλο ευχαριστώ ανήκει στον κ. Εμμανουήλ Μπαρμπουνάκη, υποψήφιο Διδάκτορα ΕΜΠ, για τις συμβουλές και τις υποδείξεις του κατά την διάρκεια εκπόνησης της παρούσας Διπλωματικής Εργασίας. Χωρίς την βοήθεια του δεν θα τα είχα καταφέρει.

Τέλος θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου για τη στήριξη και τη βοήθεια σε όλα τα χρόνια των σπουδών μου.

.....

Ζυγούρης Μιχαήλ

Διπλωματούχος Πολιτικών Μηχανικών ΕΜΠ.

Copyright © Μιχαήλ Ζυγούρης, 2016

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τη συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τη συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Πρότυπα Μηχανικής Μάθησης για την Αναγνώριση Ακραίας Οδηγικής Συμπεριφοράς

Μιχαήλ Ζυγούρης

Επιβλέπουσα : Ελένη Βλαχογιάννη, Επίκουρη Καθηγήτρια ΕΜΠ

Σύνοψη

Σκοπός της παρούσας Διπλωματικής Εργασίας είναι η πρόβλεψη της ακραίας οδηγικής συμπεριφοράς μέσω της χρήσης αισθητήρων του έξυπνου κινητού. Αξιοποιήθηκε μεγάλος όγκος δεδομένων που καταγράφονταν ανά δευτερόλεπτο και αφορούσαν στην ακριβή θέση του οχήματος, την ταχύτητά του, τα σημεία όπου ο οδηγός πραγματοποιούσε απότομη αλλαγή ταχύτητας ή απότομο ελιγμό. Η αξιολόγηση και έλεγχος των δεδομένων που καταγράφονταν από το έξυπνο κινητό γινόταν από ειδική συσκευή παρακολούθησης των χαρακτηριστικών οδήγησης επί της διαγνωστικής θύρας του αυτοκινήτου. Αναπτύχθηκαν κατηγοριοποιητές για την πρόβλεψη τόσο της απότομης επιτάχυνσης/επιβράδυνσης, όσο και απότομου ελιγμού. Εφαρμόστηκαν ειδικές τεχνικές δειγματοληψίας για ανομοιογενή δείγματα και έγινε αξιολόγηση της επιρροής τους στην ακρίβεια εντοπισμού των συμβάντων. Τα αποτελέσματα κατέδειξαν ως καταλληλότερο κατηγοριοποιητή τον J48- Consolidated και καταλληλότερο ποσοστό συμβάντων στο δείγμα το 3%.

Λέξεις κλειδιά : οδηγική συμπεριφορά, ακραία γεγονότα, έξυπνα κινητά, μηχανική μάθηση, ανομοιογενή δείγματα

Machine Learning Models for Identifying Harsh Driving Behavior

Michail Zygouris

Supervisor : Eleni Vlahogianni

Abstract

The aim of this diploma thesis is to identify harsh driving behavior using data from the Smartphone's sensors. A large dataset was exploited, which included data that were recorded per second and concerned the exact position and speed of vehicle and the time and location where the driver was performing a sudden change of speed or a sudden maneuver. The evaluation of the data that were recorded was succeeded because of a monitoring device ,on board diagnostics, that records all the driving characteristics. Classifiers were developed in order to predict either sudden change of speed, or sudden cornering (left/right). Techniques to cope with inhomogeneous datasets were exploited but also evaluation of their influence on the accuracy of tracking events . The results demonstrated that the J48-consolidated was the most accurate classifier using data that contained 3% of harsh events.

Key words : driving behavior, harsh events, smartphones, machine learning, imhomogeneous samples, sampling

Περίληψη

Αντικείμενο της παρούσας Διπλωματικής Εργασίας είναι η διερεύνηση της κυκλοφοριακής συμπεριφοράς του οδηγού μέσω της χρήσης αισθητήρων του έξυπνου κινητού. Πιο συγκεκριμένα θα εξεταστεί το αν στοιχεία οδήγησης του οδηγού όπως η ταχύτητα, η επιτάχυνση κτλ μπορούν όχι μόνο να βοηθήσουν να καθοριστεί η οδηγική συμπεριφορά αλλά και η πρόβλεψη της. Επιπρόσθετος σκοπός της Διπλωματικής Εργασίας αποτελεί η επιλογή του κατάλληλου κατηγοριοποιητή (classifier) έτσι ώστε να αναπτυχθεί η εφαρμογή. Η σωστή επιλογή κατηγοριοποιητή έχει ως επακόλουθο την πιο ορθή χρήση των μεταβλητών που ορίστηκε με αποτέλεσμα την πιο αξιόπιστη πρόβλεψη της οδηγικής συμπεριφοράς.

Η συλλογή των απαιτούμενων δεδομένων πραγματοποιήθηκε από δύο πηγές. Η πρώτη αφορά στους αισθητήρες που βρίσκονται σε έξυπνα κινητά, όπως επιταχυνσιόμετρο, γυροσκόπιο, και GPS. Η δεύτερη αφορά σε συσκευή στη θύρα OBD-II του οχήματος, η οποία είναι εξοπλισμένη με επιταχυνσιόμετρο και GPS, αλλά συλλέγει και δεδομένα από τον «εγκέφαλο» του οχήματος. Δεδομένου ότι η δυνατότητα αξιοποίησης των στοιχείων αυτών είναι σχετικά πρόσφατη, πρόκειται για ένα θέμα το οποίο δεν έχει μελετηθεί ιδιαίτερος ούτε διεθνώς ούτε στην Ελλάδα, ωστόσο, όπως προκύπτει και από τη βιβλιογραφική ανασκόπηση, τα τελευταία χρόνια χάρη στην εξέλιξη της τεχνολογίας, παρατηρείται μία αυξανόμενη ενασχόληση του επιστημονικού κόσμου γύρω από το συγκεκριμένο ζήτημα.

Τα δεδομένα καταγράφηκαν ανά δευτερόλεπτο στα ταξίδια διαφόρων οδηγών για 2 μήνες (Ιανουάριο και Φεβρουάριο του 2016) χρησιμοποιώντας εναλλασσόμενα τρεις συσκευές smartphone και tablet. Συνολικά χρησιμοποιήθηκαν 24 διαφορετικοί οδηγοί. Ως αποτέλεσμα της προκαταρκτικής ανάλυσης των δεδομένων υπάρχει η τελική αναγνώριση της συγκεκριμένης χρονικής στιγμής για ένδειξη ύπαρξης συμβάντος ή μη- συμβάντος.

Η ανομοιογένεια μεταξύ των δύο κλάδων της πρόβλεψης, συμβάν ή κανονική οδήγηση, αντιμετωπίστηκε με μια στρατηγική δειγματοληψίας για μη ισορροπημένα δείγματα. Μέσω της στρατηγικής αυτής επιτυγχάνεται αύξηση του ποσοστού των συμβάντων λόγω κατάλληλης διαγραφής δεδομένων κανονικής οδήγησης από το δείγμα. Μετά την κατάλληλη επεξεργασία των δεδομένων, αναπτύχθηκαν 4 διαφορετικοί κατηγοριοποιητές :

- ✓ Δέντρο απόφασης : J48- Consolidated
- ✓ Δίκτυο Bayes : Bayes_net
- ✓ Πολυεπίπεδο νευρωνικό δίκτυο: multi-layer perception
- ✓ Λογιστική παλινδρόμηση: logistic

Σκοπός ήταν να προβλεφθούν οι εξαρτημένες μεταβλητές :

- ✓ Απότομη επιτάχυνση/ επιβράδυνση
- ✓ Απότομος ελιγμός

Συνοπτικά, από τις αναλύσεις που διεξήχθησαν προέκυψε ότι τα δεδομένα τα οποία καταγράφονται από τους αισθητήρες των έξυπνων κινητών μπορούν να αποθηκευθούν ηλεκτρονικά σε υπολογιστή και περιέχουν ιδιαίτερες σημαντικές πληροφορίες οι οποίες μετά από κατάλληλη επεξεργασία και ανάπτυξη μαθηματικών μοντέλων, μπορούν να φανούν ιδιαίτερος χρήσιμα στην εξαγωγή συμπερασμάτων για την οδηγική συμπεριφορά του οδηγού.

Βασικός παράγοντας για την ανίχνευση συμβάντος απότομης στροφής είναι η καταγραφή της γωνιακής ταχύτητας μέσω του γυροσκοπίου.

Βασικός παράγοντας για την ανίχνευση συμβάντος απότομης επιτάχυνσης και επιβράδυνσης είναι η καταγραφή της επιτάχυνσης μέσω του επιταχυνσιόμετρου.

Καταλληλότερος κατηγοριοποιητής εμφανίζεται ως ο j48 consolidated όπου με εξαίρεση μια κατηγορία παρουσιάζει την μεγαλύτερη απόδοση.

Αδυναμία στο να προβλέψουν είτε συμβάντα απότομης στροφής, είτε συμβάντα απότομης επιτάχυνσης και επιβράδυνσης έχουν η πολυεπίπεδη ταξινόμηση και η κλασσική λογιστική παλινδρόμηση.

Η ευκρίνεια (Precision) των μοντέλων στην πρόβλεψη συμβάντων απότομης επιτάχυνσης και επιβράδυνσης είναι αρκετά χαμηλή λόγω και της δυσκολίας πρόβλεψης του φαινομένου.

Σε δείγμα που περιέχει συμβάντα σε ποσοστό 3% παρατηρείται η μεγαλύτερη τιμή στους δείκτες μέτρησης Precision και F-Measure.

Περιεχόμενα

Ευχαριστίες.....	3
Σύνοψη	1
Abstract	2
Περίληψη.....	3
1. Εισαγωγή	9
1.1 Έξυπνες Εφαρμογές και Οδηγική Συμπεριφορά.....	9
1.2 Σκοπός της διπλωματικής.....	10
1.3 Διάρθρωση Διπλωματικής Εργασίας.....	10
2. Βιβλιογραφική ανασκόπηση	11
2.1 Γενικά.....	11
2.2 Συναφείς έρευνες.....	11
2.3 Πρότυπα Αξιολόγησης Συμπεριφοράς Οδηγών	14
2.4 Συμπεράσματα Βιβλιογραφίας	15
3. Θεωρητικό Υπόβαθρο	16
3.1 Μεθοδολογική Προσέγγιση.....	16
3.2 Κατηγοριοποίηση	17
3.2.1 Λογιστική Παλινδρόμηση	18
3.2.2 Δέντρα Αποφάσεων	18
3.2.3 Δίκτυα Bayes	20
3.2.4 Πολυεπίπεδα Νευρωνικά Δίκτυα	21
3.3 Αξιολόγηση προτύπων	22
4. Συλλογή και Επεξεργασία Δεδομένων	24
4.1 Εισαγωγή	24
4.2 Το Πείραμα.....	24
4.3 Συλλογή στοιχείων	24
4.4 Προκαταρκτική Επεξεργασία Στοιχείων	27
4.4.1 Διαμόρφωση δεδομένων.....	27
4.4.2 Διαθέσιμη Βάση Δεδομένων	29
4.5 Ανάπτυξη και Αξιολόγηση Κατηγοριοποιητών	29
4.5.1 Προετοιμασία βάσεων δεδομένων	29
4.5.2 Αποτελέσματα μοντέλων.....	32
5. Συμπεράσματα.....	42
5.1 Σύνοψη αποτελεσμάτων	42
5.2 Συνολικά συμπεράσματα.....	43

5.3 Προτάσεις για βελτίωση της κυκλοφοριακής συμπεριφοράς.....	43
5.4 Προτάσεις για περαιτέρω έρευνα	44
Βιβλιογραφία.....	45

Κατάλογος εικόνων

Σχήμα 1: Δομή της εφαρμογής DriveSafe (Πηγή: [1]).....	11
Σχήμα 2: Δομή της εφαρμογής PAYDAYS (Πηγή : [2])	12
Σχήμα 3: Δομή εφαρμογής της μελέτης Engelbrecht (Πηγή: [3])	13
Σχήμα 3: Δομή εφαρμογής μελέτης Predic(Πηγή : [10])	14
Σχήμα 5 : Σχηματική διάταξη μεθοδολογικής προσέγγισης	16
Σχήμα 6: Τυπική μορφή λογιστικής καμπύλης	18
Σχήμα 7: Τυπική μορφή κατηγοριοποιητή J48 CONSOLIDATED	19
Σχήμα 8: Γενική δομή του κατηγοριοποιητή BayesNet	20
Σχήμα 9: Γενική δομή κατηγοριοποιητή MLP [17]	21
Σχήμα 10: Σχηματική απεικόνιση λειτουργίας συστήματος [19]	25
Σχήμα 11: Απλοποιημένη διάταξη χωρητικού επιταχυνσιόμετρου [20]	26
Σχήμα 12: Τυπική διάταξη γυροσκοπίου [21].....	26
Σχήμα 13: α) Δημιουργία εκτελέσιμου αρχείου για την ένωση των csv	28
Σχήμα 13: β) Δημιουργία εκτελέσιμου αρχείου για την ένωση των csv	28
Σχήμα 14: Χρησιμοποίηση φίλτρων για την κατάλληλη χρήση των μεταβλητών	31
Σχήμα 4: Φόρτωση set data	31
Σχήμα 5: Εύρεση ορίου ώστε να μεγιστοποιείται ο όρος F-Measure	32
Διάγραμμα 1: Διάγραμμα μεταβολής ευκρίνειας (Precision) ανάλογα με το ποσοστό των συμβάντων απότομων στρεφουσών κινήσεων	35
Διάγραμμα 2: Διάγραμμα μεταβολής βαθμού σωστών προβλέψεων (TP-Rate) ανάλογα με το ποσοστό των συμβάντων απότομων στρεφουσών κινήσεων	35
Διάγραμμα 3: Διάγραμμα μεταβολής βαθμού εσφαλμένων προβλέψεων (FP-RATE) ανάλογα με το ποσοστό των συμβάντων απότομων στρεφουσών κινήσεων	36
Διάγραμμα 4: Διάγραμμα μεταβολής της ανάκλησης (Recall) ανάλογα με το ποσοστό των συμβάντων απότομων στρεφουσών κινήσεων	36
Διάγραμμα 5: Διάγραμμα μεταβολής ευκρίνειας (Precision) ανάλογα με το ποσοστό των συμβάντων απότομης επιτάχυνσης και επιβράδυνσης.....	39
Διάγραμμα 6: Διάγραμμα μεταβολής του βαθμού σωστών προβλέψεων (TP-Rate) ανάλογα με το ποσοστό των συμβάντων απότομης επιτάχυνσης και επιβράδυνσης	40
Διάγραμμα 7: Διάγραμμα μεταβολής του βαθμού εσφαλμένων προβλέψεων (FP-Rate) ανάλογα με το ποσοστό των συμβάντων απότομης επιτάχυνσης και επιβράδυνσης.....	40
Διάγραμμα 8: Διάγραμμα μεταβολής ανάκλησης (Recall) ανάλογα με το ποσοστό των συμβάντων απότομης επιτάχυνσης και επιβράδυνσης	41

Κατάλογος πινάκων

Πίνακας 1 : Τυπικός πίνακας κατηγοριοποίησης.	22
Πίνακας 2: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 1% συμβάντα απότομων στρεφουσών κινήσεων.	32
Πίνακας 3: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 1,5% συμβάντα απότομων στρεφουσών κινήσεων.	33
Πίνακας 4: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 2% συμβάντα απότομων στρεφουσών κινήσεων.	33
Πίνακας 5 : Αποτελέσματα κατηγοριοποίησης στο δείγμα με 3% συμβάντα απότομων στρεφουσών κινήσεων.	33
Πίνακας 6: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 4% συμβάντα απότομων στρεφουσών κινήσεων.	34
Πίνακας 7: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 5% συμβάντα απότομων στρεφουσών κινήσεων.	34
Πίνακας 8: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 1% συμβάντα απότομων κινήσεων επιτάχυνσης και επιβράδυνσης.	37
Πίνακας 9: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 1,5% συμβάντα κινήσεων επιτάχυνσης και επιβράδυνσης.	37
Πίνακας 10: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 2% συμβάντα απότομων κινήσεων επιτάχυνσης και επιβράδυνσης.	37
Πίνακας 11: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 3% συμβάντα απότομων κινήσεων επιτάχυνσης και επιβράδυνσης.	38
Πίνακας 12: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 4% συμβάντα απότομων κινήσεων επιτάχυνσης και επιβράδυνσης.	38
Πίνακας 13: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 5% συμβάντα απότομων κινήσεων επιτάχυνσης και επιβράδυνσης.	38
Πίνακας 14: Συγκεντρωτικός πίνακας για την πρόβλεψη απότομου ελιγμού.....	42
Πίνακας 15: Συγκεντρωτικός πίνακας για την πρόβλεψη απότομης επιτάχυνσης και επιβράδυνσης	42
Πίνακας 16 : Στατιστικά στοιχεία με διαχωρισμό για το εαν είναι συμβάν απότομης στροφής	47
Πίνακας 17: Στατιστικά στοιχεία με διαχωρισμό για το εάν είναι συμβάν απότομης επιτάχυνσης/ επιβράδυνσης	48

Κεφάλαιο 1

1. Εισαγωγή

1.1 Έξυπνες Εφαρμογές και Οδηγική Συμπεριφορά

Η χρήση έξυπνων τηλεφώνων (Smartphones) βοηθάει στην ανίχνευση της ανθρώπινης συμπεριφοράς. Πρόσφατες έρευνες έχουν κατορθώσει να ανιχνεύσουν αν ο χρήστης περπατάει, τρέχει ή αν επιβαίνει σε κάποιο συγκεκριμένο μέσο μεταφοράς όπως λεωφορείο, ΙΧ και τρένο. Έχουν αναπτυχθεί εφαρμογές που είναι πιο εκλεπτυσμένες όπως το πόσες θερμίδες καίγονται κατά την διάρκεια μίας άσκησης, εφαρμογές που αναγνωρίζουν εάν γελάει ο χρήστης σε μια δημόσια συνάντηση, ακόμα και το πότε ο χρήστης του έξυπνου κινητού κοιμάται πιο ελαφριά ώστε να ξυπνήσει ευκολότερα μέσω ξυπνητηριού.

Η ανίχνευση και πρόβλεψη τέτοιων οδικών συμβάντων θα μπορούσε να γίνει σημαντικό εργαλείο στην διαχείριση της κυκλοφορίας, αλλά και μια πρωτοποριακή λύση για την ασφάλιση των οχημάτων.

Η διαχείριση της κυκλοφορίας μέσω της πρόβλεψης της οδηγικής συμπεριφοράς θα είναι σημαντικό βήμα στο να μειωθούν τα ατυχήματα. Θα μπορούσε με κατάλληλο σύστημα να εμποδίζεται ο χρήστης από το να πατήσει περισσότερο γκάζι, ή θα μπορούσε να ειδοποιείται η τροχαία και ο ίδιος ο χρήστης για την κακή του οδηγική συμπεριφορά. Αυτό θα οδηγούσε στο να υπάρχει πλαφόν στην ταχύτητα ώστε να μην παραβιάζεται το όριο, ανάλογα και με την κατηγορία της οδού.

Ιστορικά οι εταιρίες ασφάλισης αυτοκινήτων βάσιζαν τις χρεώσεις τους σε στατικούς παράγοντες, όπως δημογραφικά και κοινωνικά χαρακτηριστικά δίχως να λαμβάνουν υπ' όψιν τους εξατομικευμένους παράγοντες των χρηστών. Ακόμα και σήμερα η οδική συμπεριφορά δεν ενσωματώνεται στα κόμιστρα καθώς δεν διατίθενται τέτοια δεδομένα. Για αυτόν το λόγο έγινε προσπάθεια από μερικές εταιρίες ώστε να πληρώνουν οι χρήστες ανάλογα με το πόσα χιλιόμετρα διανύουν και τα ατυχήματα στα οποία εμπλέκονται σε διάφορες παραλλαγές όπως η ασφάλιση ανάλογα με τον βαθμό χρησιμοποίησης (Usage Based Insurance -UBI) και η μεγαλύτερη πληρωμή όσο οδηγείς (Pay As You Drive -PAUD).

Τα τρέχοντα προγράμματα όμως έχουν ως στόχευση μόνο τα διανυθέντα χιλιόμετρα και όχι την οδηγική συμπεριφορά. Εξέλιξη των προγραμμάτων πληρωμής ανάλογα με την χρήση είναι η πληρωμή ανάλογα με τον τρόπο οδηγικής συμπεριφοράς (Pay How You Drive -PHUD). Σε αυτή την μέθοδο λαμβάνεται δηλαδή υπ' όψιν και η επικινδυνότητα από τον τρόπο οδήγησης παράγοντας που δεν συνυπολογιζόταν ως τώρα, λόγω έλλειψης δεδομένων.

1.2 Σκοπός της διπλωματικής

Σκοπός της παρούσας Διπλωματικής Εργασίας είναι η διερεύνηση της οδηγικής συμπεριφοράς μέσω της χρήσης αισθητήρων του έξυπνου κινητού. Πιο συγκεκριμένα θα εξεταστεί το αν στοιχεία οδήγησης του οδηγού όπως η ταχύτητα, η επιτάχυνση κτλ μπορούν όχι μόνο να βοηθήσουν να καθοριστεί η οδηγική συμπεριφορά, αλλά και να συμβάλουν στην πρόβλεψη της. Επιπρόσθετος σκοπός της Διπλωματικής Εργασίας αποτελεί η επιλογή του κατάλληλου κατηγοριοποιητή (classifier). Η σωστή επιλογή κατηγοριοποιητή έχει ως επακόλουθο την πιο ορθή χρήση των μεταβλητών που ορίστηκε με αποτέλεσμα την πιο αξιόπιστη πρόβλεψη της οδηγικής συμπεριφοράς. Η διερεύνηση της οδηγικής συμπεριφοράς βασίζεται σε μετρήσεις που συλλέγονται από αισθητήρες τοποθετημένους σε κινητά τηλέφωνα. Το έξυπνο κινητό προτιμήθηκε από τα υπόλοιπα διαγνωστικά μέσα ως ένα μέσο που ήδη κατέχουν οι χρήστες και αρκεί μόνο μια εφάρμογή που εκμεταλλεύεται του υπάρχοντες αισθητήρες του κινητού.

1.3 Διάρθρωση Διπλωματικής Εργασίας

Η διπλωματική εργασία χωρίζεται στα παρακάτω κεφάλαια:

Στο πρώτο κεφάλαιο γίνεται μια εισαγωγή στο αντικείμενο της Διπλωματικής Εργασίας παρουσιάζεται δηλαδή το έξυπνο κινητό ως μέσο ανίχνευσης της οδηγικής συμπεριφοράς και αναλύεται το αντίκτυπο της ανίχνευσης αυτής σε διάφορους τομείς όπως η διαχείριση της κυκλοφορίας, η οδική ασφάλεια και η ασφάλιση των αυτοκινήτων.

Στο δεύτερο κεφάλαιο παρουσιάζονται τα αποτελέσματα της βιβλιογραφικής ανασκόπησης σε έρευνες με συναφές αντικείμενο. Παρατίθενται εργασίες και μελέτες οι οποίες είχαν δημοσιευθεί σε συνέδρια, συγγράμματα και επιστημονικά περιοδικά και πραγματοποιείται σύντομη αναφορά σε κάθε εργασία που περιλαμβάνει το πλαίσιο της έρευνας, τη μεθοδολογία, τα μέσα μέτρησης και τις μεταβλητές που λαμβάνονταν.

Στο τρίτο κεφάλαιο παρουσιάζεται το θεωρητικό υπόβαθρο στο οποίο στηρίχθηκε η ανάλυση των στοιχείων. Αρχικά παρουσιάζονται κάποιες βασικές έννοιες προγραμματισμού και. Έπειτα ακολούθησε η ανάλυση των μοντέλων που χρησιμοποιήθηκαν και οι απαραίτητοι στατιστικοί έλεγχοι στους οποίους υποβάλλονται τα δεδομένα. Το κεφάλαιο ολοκληρώνεται με την στατιστική θεωρία για την ανάλυση των δεδομένων και την τελική επιλογή του κατάλληλου μοντέλου.

Στο τέταρτο κεφάλαιο γίνεται αναλυτική αναφορά των διαδικασιών της συλλογής και επεξεργασίας στοιχείων στα οποία στηρίχθηκε η Διπλωματική Εργασία. Γίνεται εκτενής παρουσίαση στο πείραμα που έγινε έτσι ώστε να δημιουργηθεί η βάση δεδομένων. Στο τέταρτο κεφάλαιο ακόμα περιλαμβάνεται η λεπτομερής περιγραφή της μεθοδολογίας για την εξαγωγή των τελικών αποτελεσμάτων. Γίνεται αναφορά των βημάτων που ακολουθήθηκαν για την εφαρμογή της μεθοδολογίας των διάφορων μοντέλων. Έπειτα παρουσιάζονται τα δεδομένα εισόδου και εξόδου, έμφαση τα δεδομένα εξόδου που αφορούν την ακρίβεια των προβλέψεων και την επιλογή της καλύτερης μεθόδου κατηγοριοποίησης.

Το πέμπτο κεφάλαιο είναι το κεφάλαιο που αναπτύσσονται τα συμπεράσματα από κάθε ανάλυση ξεχωριστά καθώς και προτάσεις για περαιτέρω έρευνα.

Κεφάλαιο 2

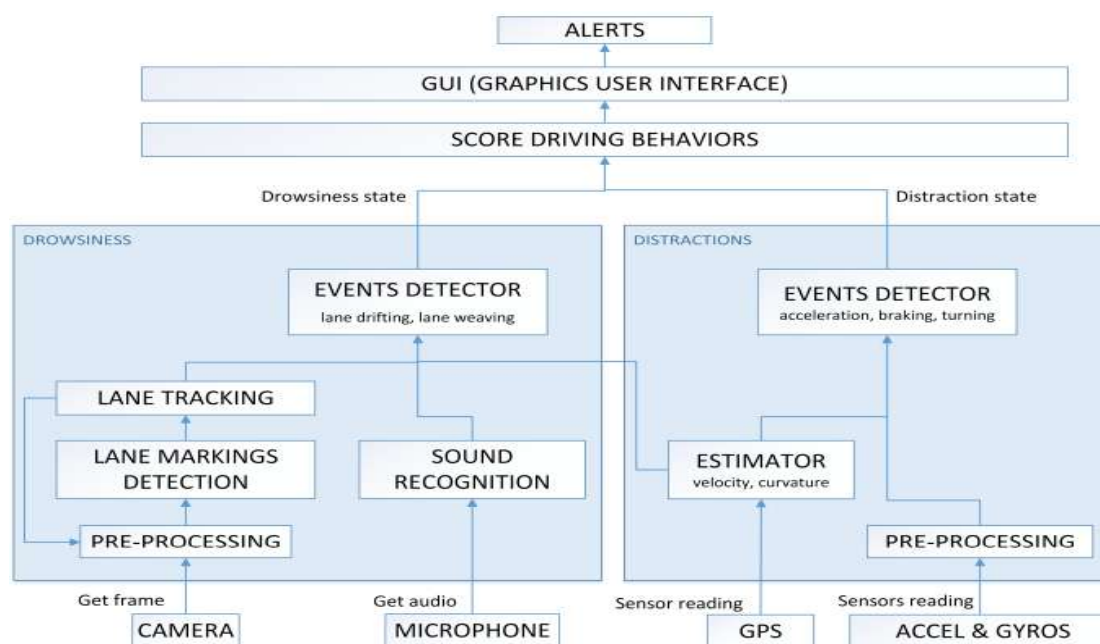
2. Βιβλιογραφική ανασκόπηση

2.1 Γενικά

Η εδραίωση της χρήσης των έξυπνων κινητών στην καθημερινότητα των χρηστών έχει εμπνεύσει αρκετούς ερευνητές τα τελευταία χρόνια να μελετήσουν την οδηγική συμπεριφορά. Το κεφάλαιο αυτό αφορά την βιβλιογραφική ανασκόπηση και περιλαμβάνει έρευνες στον τομέα της διαχείρισης της κυκλοφορίας το αντικείμενο και η μεθοδολογία των οποίων παρουσίαζε συνάφεια με την παρούσα διπλωματική εργασία. Συγκεκριμένα παρουσιάζονται έρευνες που αναφέρονται στην κατανόηση της συμπεριφοράς (ή και γνωρισμάτων της συμπεριφοράς) των οδηγών αυτοκινήτων μέσω χρήσης έξυπνου κινητού. Μέσω της ανασκόπησης των μεθοδολογιών και των ερευνών επιχειρήθηκε ο προσδιορισμός των καταλληλότερων μεθόδων και των κατάλληλων παραμέτρων για την επίτευξη του στόχου της διπλωματικής εργασίας.

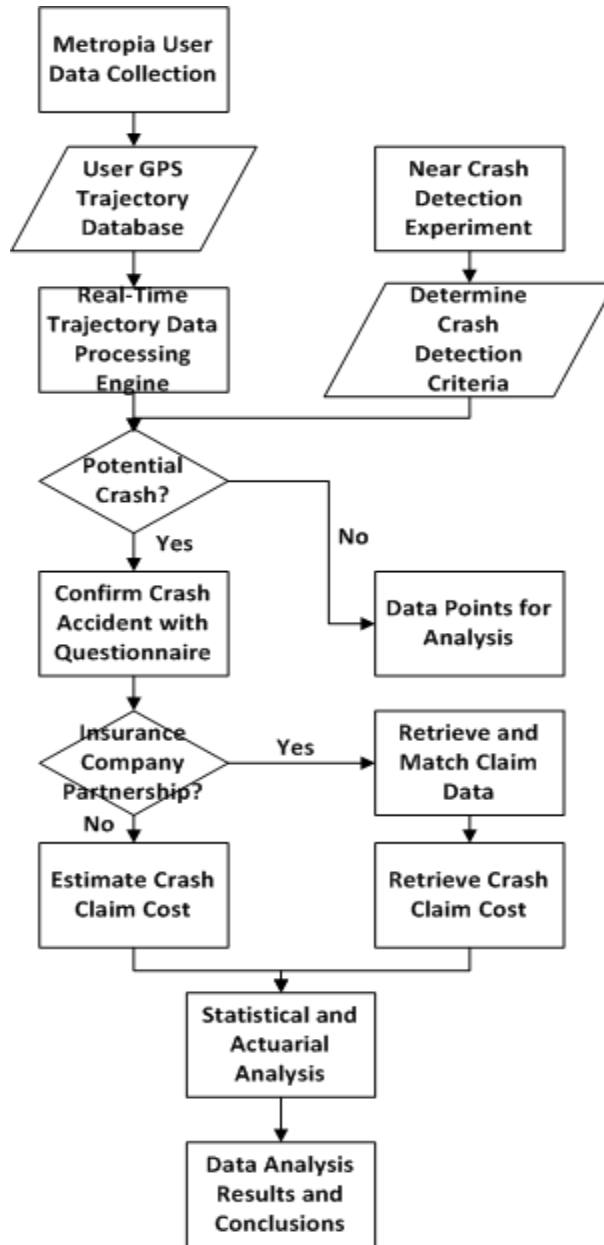
2.2 Συναφείς έρευνες

Η παρακολούθηση της οδικής συμπεριφοράς αποτελεί αντικείμενο πολλών ερευνητών τα τελευταία χρόνια. Μία προσπάθεια έγινε από τους Bergasa et. al. [1] όπου η εφαρμογή DriveSafe που δημιουργήθηκε χρησιμοποιεί τους αδρανειακούς αισθητήρες, την πίσω κάμερα, το μικρόφωνο και το GPS, με σκοπό να βρίσκονται οι μισοκοιμισμένοι και οι οξύθυμοι οδηγοί.



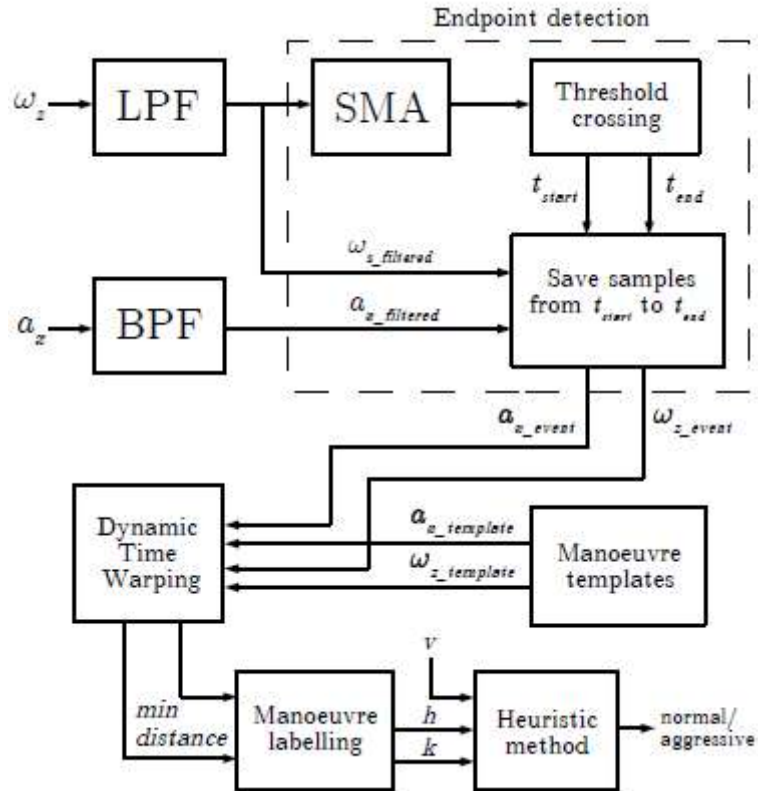
Σχήμα 1: Δομή της εφαρμογής DriveSafe (Πηγή: [1]).

Ακόμα, η Chiou et. al. [2] ανέπτυξαν μια εφαρμογή που δημιουργήθηκε χρησιμοποιεί τα δεδομένα που συλλέγονται από το GPS και αποθηκεύονται σε μια βάση δεδομένων και παρατηρούνται οι επιταχύνσεις, οι επιβραδύνσεις την ώρα ατυχημάτων ή παρ' ολίγον ατυχημάτων.



Σχήμα 2: Δομή της εφαρμογής PAYDAYS (Πηγή : [2])

Επιπρόσθετα η εργασία Engelbrecht et. al. [3] μεταχειρίζεται τους αδρανειακούς αισθητήρες, όπως επιταχυνσιόμετρο (με τιμές a_x , a_y και a_z) και γυροσκόπιο (με τιμές ω_x , ω_y και ω_z), και το GPS έτσι ώστε να βρίσκονται τα καταληκτικά σημεία των οδικών συμβάντων. Ευρετική μέθοδος και δυναμική χρονική στρέβλωση (DTW) χρησιμοποιείται ώστε να διαχωρίζονται οι κανονικές από τις βίαιες κινήσεις.

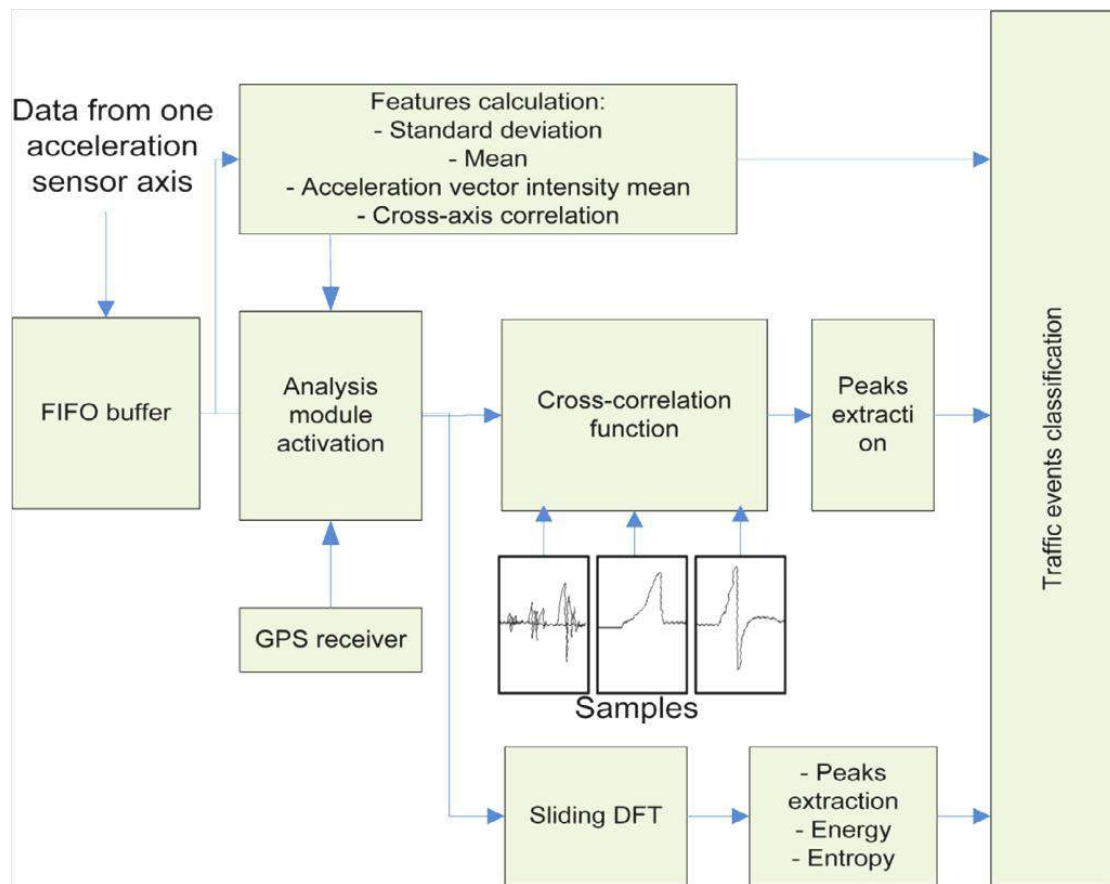


Σχήμα 3: Δομή εφαρμογής της μελέτης Engelbrecht (Πηγή: [3])

Παρόμοιες προσπάθειες έγιναν στις εργασίες Nicholas D. Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T. Campbell et. al. [4], Handel et. al. [5] [6] και Husnjak et. al. [7] όπου χρησιμοποιούνται εφαρμογές ασφαλιστικών εταιρειών με τιμολόγηση ανάλογα με την χρήση των οδηγών UBI (Usage Based Auto Insurance).

Μια διαφορετική προσέγγιση υπήρξε στην εργασία Johnson and Trivedi et. al. [8], [9]. Ναι μεν η εφαρμογή MIROAD (A Mobile-Sensor-Platform for Intelligent Recognition Of Aggressive Driving) που δημιούργησαν χρησιμοποιούσε τους αδρανειακούς αισθητήρες και το GPS αλλά όλη η επεξεργασία με δυναμική χρονική στρέβλωση (DTW) γίνεται εξολοκλήρου από το ίδιο το έξυπνο κινητό.

Σημαντικό εργαλείο για την διαχείριση της κυκλοφορίας προέρχεται από την εργασία Predic et. al. [10] υπάρχει η εφαρμογή που υπολογίζει μέσο χρόνο ταξιδιού ώστε να υπολογιστούν κρίσιμα συμβάντα και να προωθούνται στους υπόλοιπους χρήστες της εφαρμογής. Ο στόχος της εφαρμογής επιτυγχάνεται μέσω των εργαλείων πλοήγησης των χρηστών. Σε αντίθεση με άλλες έρευνες συναντάμε τους όρους ενέργεια και εντροπία.



Σχήμα 3: Δομή εφαρμογής μελέτης Predic(Πηγή : [10])

Αρωγός στην προσπάθεια αυτή ήταν οι εργασίες για την περιστροφή της θέσης της συσκευής όπως η Ayub et. al. [11], Almazan et. al. [12] και et. al. [13]. Με την μεθοδολογία αυτών των εργασιών, οι ενδείξεις του γυροσκοπίου και του επιταχυνσιόμετρου μπορούν να περιστραφούν ως προς κοινό άξονα.

2.3 Πρότυπα Αξιολόγησης Συμπεριφοράς Οδηγών

Οι διάφορες εργασίες οι οποίες αναλύθηκαν έχουν παρόμοια επιλογή των μεταβλητών όπως οι δείκτες του επιταχυνσιόμετρου, του γυροσκοπίου και του GPS. Παρόλα αυτά υπάρχουν εργασίες όπως Bergasa et. al. [1] όπου εκτός από τα προαναφερθέντα χρησιμοποιούνται το μικρόφωνο και οι κάμερες των κινητών έτσι ώστε να αξιολογείται η ψυχολογική κατάσταση του οδηγού και κατά πόσο είναι νυσταγμένος.

Ενώ οι έρευνες δεν είναι πολλές, λόγω και της πρόσφατης ραγδαίας εξέλιξης των έξυπνων κινητών έχουν δοκιμασθεί διαφορετικές μεθοδολογίες στον τρόπο αξιολόγησης της οδηγικής συμπεριφοράς. Έτσι συναντώνται μεθοδολογίες όπως η χωριστική μέθοδος ,η πολλαπλή-συσχέτιση (cross- correlation) με βιβλιοθήκη συμβάντων και το δέντρο αποφάσεων.

Η ακρίβεια στην αξιολόγηση της συμπεριφοράς των οδηγών είναι αρκετά υψηλή όσον αφορά σε εργασίες που υπήρχε πρόσβαση σε αριθμητικά αποτελέσματα γιατί σε πολλές εργασίες δεν υπήρχε πρόσβαση στους δείκτες μέτρησης της αξιοπιστίας.

Για παράδειγμα η Bergasa et. al. [1] παρουσιάζει δείκτη ακρίβειας 82% και δείκτη ανάκλησης 92%. Τα αποτελέσματα της Engelbrecht et. al είναι ελαφρώς μειωμένα στους ίδιους δείκτες σε σχέση με την Bergasa et. al. αλλά η ακρίβεια (accuracy) του είναι αρκετά υψηλή επιτυγχάνοντας 84.50%. Τελευταία από τις εργασίες που μπορούσε να υπάρχει πρόσβαση ήταν η Trivedi et. al. με ακρίβεια (accuracy) 77%.

2.4 Συμπεράσματα Βιβλιογραφίας

Στο παρόν κεφάλαιο έγινε παρουσίαση και εκτενής αναφορά μερικών από τις έρευνες που σχετίζονται με την εκτίμηση της συμπεριφοράς του οδηγού μέσω έξυπνων κινητών καθώς και τη πιθανή χρήση αυτών στην διαχείριση της κυκλοφορίας και την βελτίωση της οδικής ασφάλειας. Από την ανάλυση των παραπάνω ερευνών και την συνολική εξέτασή τους προκύπτουν τα παρακάτω κύρια σημεία:

- Για τη σωστή αξιολόγηση της οδηγικής συμπεριφοράς είναι απαραίτητη η χρήση των αδρανειακών αισθητήρων (επιταχυνσιόμετρο και γυροσκόπιο).
- Ενώ υπάρχει μεγάλο ενδιαφέρον να αξιολογηθεί η οδηγική συμπεριφορά δεν υπάρχει τόσο μεγάλη ευκολία να προβλεφθεί.
- Για την αποτελεσματική λειτουργία των συστημάτων καταγραφής (OBD, Smartphone) απαιτείται σύνδεση με ηλεκτρονικό υπολογιστή για να αποθηκεύονται και να επεξεργάζονται οι πληροφορίες
- Όταν υπάρχει ειδοποίηση εντός του οχήματος για επικίνδυνη οδήγηση και πιθανή τροχαία παράβαση εμφανίζεται η τάση άμεσης βελτίωσης της οδηγικής συμπεριφοράς
- Η προσπάθεια εκτίμησης της οδηγικής συμπεριφοράς μέσω έξυπνων κινητών είναι πρόσφατη και συνεχώς εξελίσσεται και βελτιώνεται.

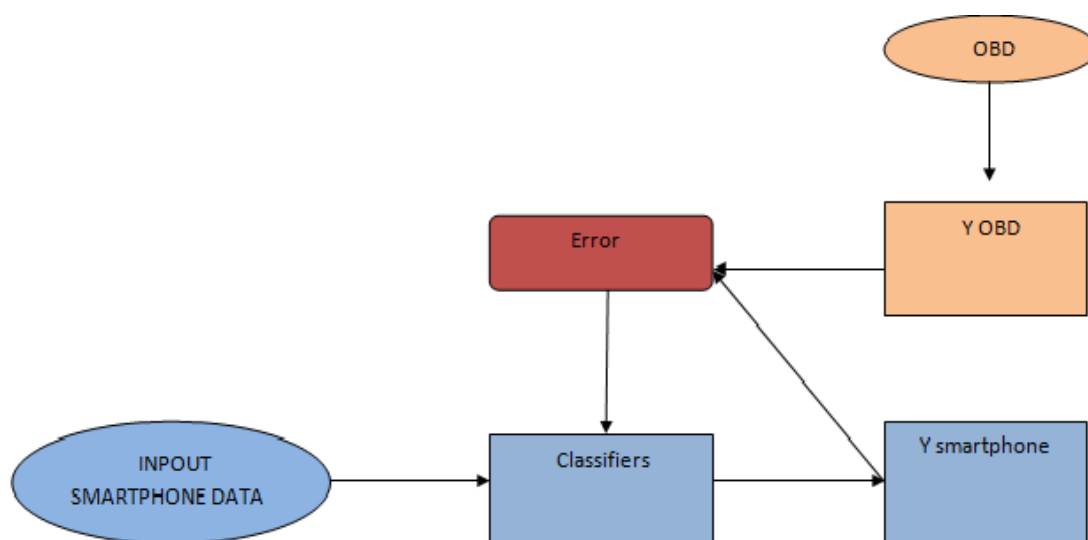
Κεφάλαιο 3

3. Θεωρητικό Υπόβαθρο

3.1 Μεθοδολογική Προσέγγιση

Το πρόβλημα της πρόβλεψης της οδηγικής συμπεριφοράς ενός οδηγού αντιμετωπίζεται ως ένα πρόβλημα κατηγοριοποίησης, κατά το οποίο επιχειρείται η συσχέτιση δεδομένων από επιταχυνσιόμετρο, γυροσκόπιο και GPS ενός έξυπνου κινητού με τα ακραία οδηγικά συμβάντα που έχουν εντοπιστεί από μια συσκευή παρακολούθησης των χαρακτηριστικών του οχήματος που είναι εγκατεστημένη στη θύρα OBD-II.

Όπως φαίνεται στο Σχήμα 5, σκοπός της μεθοδολογίας που ακολουθείται είναι η ελαχιστοποίηση του σφάλματος, μέσω των διάφορων κατηγοριοποιητών, μεταξύ των προβλέψεων που προέρχονται από τα έξυπνα κινητά και της καταγραφής που υπήρχε από το OBD.



Σχήμα 5 : Σχηματική διάταξη μεθοδολογικής προσέγγισης

Το μεγαλύτερο πρόβλημα στην στατιστική ανάλυση των δεδομένων είναι το μικρό δείγμα των οδηγικών συμβάντων (απότομες επιβραδύνσεις, επιταχύνσεις και στροφές) σε σχέση με το δείγμα των μη οδηγικών συμβάντων. Τα δύο υποσύνολα δηλαδή έχουν πολύ μεγάλη ανομοιογένεια και υπερβολικά ασύμμετρες κατανομές. Η μεγάλη αυτή ανομοιογένεια επηρεάζει την αποδοτικότητα και την ακρίβεια των αποτελεσμάτων των αλγορίθμων μηχανικής μάθησης. Για αυτόν τον λόγο επιχειρείται με ειδικό τρόπο η αύξηση του ποσοστιαίου δείγματος των συμβάντων, έτσι ώστε να διερευνηθεί η επιρροή της ομογενοποίησης του δείγματος στην αποδοτικότητα των αλγορίθμων κατηγοριοποίησης.

3.2 Κατηγοριοποίηση

Η κατηγοριοποίηση (classification) είναι μια τεχνική της εξαγωγής δεδομένων, κατά την οποία ένα στοιχείο ανατίθεται σε ένα προκαθορισμένο σύνολο κατηγοριών. Ο όρος κατηγοριοποίηση συναντάται στην βιβλιογραφία και ως ταξινόμηση. Γενικότερα, ο στόχος της διαδικασίας αυτής είναι η ανάπτυξη ενός μοντέλου, το οποίο αργότερα θα μπορεί να χρησιμοποιηθεί για την κατηγοριοποίηση μελλοντικών δεδομένων.

Η κατηγοριοποίηση μπορεί να περιγραφεί ως μία διαδικασία δύο βημάτων:

1. **Εκμάθηση (Learning):** Στο πρώτο βήμα της διαδικασίας δημιουργείται/προσδιορίζεται το μοντέλο με βάση ένα σύνολο προκατηγοριοποιημένων παραδειγμάτων, που ονομάζεται δεδομένα εκπαίδευσης (training data). Τα δεδομένα εκπαίδευσης αναλύονται από έναν αλγόριθμο κατηγοριοποίησης, προκειμένου να σχηματιστεί το μοντέλο. Λόγω του ότι τα δεδομένα εκπαίδευσης ανήκουν σε μία προκαθορισμένη κατηγορία, η οποία είναι γνωστή, η κατηγοριοποίηση αποτελεί μέθοδος εποπτευομένης μάθησης (supervised learning). Το μοντέλο, που λέγεται και αλλιώς κατηγοριοποιητής (classifier), αναπαρίσταται με τη μορφή κανόνων κατηγοριοποίησης (classification rules), δέντρων απόφασης (decision trees) ή μαθηματικών τύπων.
2. **Κατηγοριοποίηση (Classification):** Μετά την δημιουργία του μοντέλου, το επόμενο βήμα είναι η αξιολόγησή του. Για να επιτευχθεί αυτό, χρησιμοποιούμε τα δοκιμαστικά δεδομένα (test data) για να υπολογίσουν την ακρίβεια του μοντέλου. Το μοντέλο κατηγοριοποιεί τα δοκιμαστικά δεδομένα. Έπειτα, η κατηγορία που σχηματίστηκε με βάση τα δοκιμαστικά δεδομένα συγκρίνεται με την πρόβλεψη που έγινε για τα δεδομένα εκπαίδευσης, τα οποία είναι ανεξάρτητα από αυτά της δοκιμής. Η ακρίβεια του μοντέλου υπολογίζεται από το ποσοστό των δειγμάτων δοκιμής που κατηγοριοποιήθηκαν σωστά σε σχέση με το υπό εκπαίδευση μοντέλο.

Παραγωγή κανόνων κατηγοριοποίησης: Η γνώση που αποκτάται κατά την διαδικασία της κατηγοριοποίησης μπορεί να αναπαρασταθεί και με τη χρήση κανόνων. Οι κανόνες κατηγοριοποίησης, σε σχέση με τα δέντρα απόφασης, γίνονται ευκολότερα κατανοητοί όταν το δέντρο που παράχθηκε είναι μεγάλο. Έτσι μπορούμε να μετατρέψουμε ένα δέντρο απόφασης σε ένα σύνολο κανόνων κατηγοριοποίησης. Αυτό μπορεί να επιτευχθεί εάν θεωρήσουμε ότι κάθε κανόνας αντιστοιχεί σε ένα μονοπάτι του δέντρου από τη ρίζα μέχρι ένα κόμβο φύλλο. Άρα κάθε φύλλο παράγει ένα κανόνα. Οι συνθήκες που θα μας οδηγήσουν στο φύλλο(υπόθεση) αποτελούν το αριστερό μέρος του κανόνα, ενώ το φύλλο (αποτέλεσμα) αντιστοιχεί στο δεξιό μέρος του κανόνα. Για παράδειγμα, από την παραπάνω φωτογραφία ισχύει:

Στο κεφάλαιο αυτό παρουσιάζεται η θεωρία στην οποία βασίζεται η στατιστική ανάλυση της Διπλωματικής Εργασίας. Οι μέθοδοι που χρησιμοποιούνται ώστε να ταξινομηθεί το δείγμα μας και να έχουμε ακρίβεια στην πρόβλεψη των συμβάντων είναι α) το δέντρο ταξινόμησης J48 Consolidated β) το Μπεϋζιανό δίκτυο Bayes net γ) Η ταξινόμηση με αντίληψη πολλαπλών επιπέδων (multi-layer perception) και δ) η λογιστική παλινδρόμηση

3.2.1 Λογιστική Παλινδρόμηση

Η λογιστική παλινδρόμηση είναι άλλη μια προσέγγιση σε μαθησιακή συνάρτηση τύπου $f: X \rightarrow Y$ ή $P(Y|X)$ σε περίπτωση όπου το Y έχει διακριτή τιμή και X είναι ένα οποιοδήποτε σύνολο με διακριτές ή συνεχείς μεταβλητές. Στην συγκεκριμένη περίπτωση Y έχει δύο περιπτώσεις.

Η λογιστική παλινδρόμηση υποθέτει για την κατανομή τύπο και εκτιμά από τα δεδομένα την ζητούμενη παράμετρο. Οι τύποι για δυαδική μεταβλητή Y είναι:

$$P(Y = 1|X) = 1/(1 + e^{w_0 + \sum w_i X_i}) \quad (3.1)$$

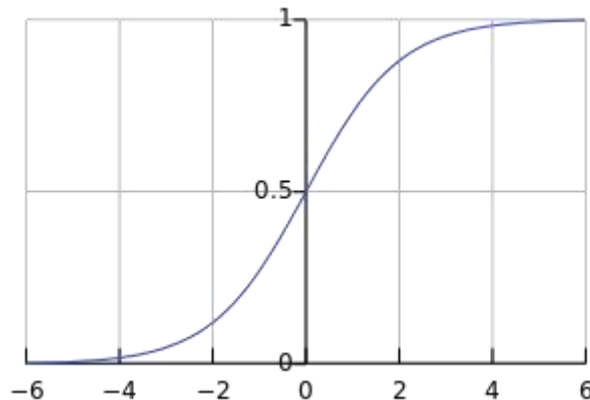
$$P(Y = 0|X) = (e^{w_0 + \sum w_i X_i})/(1 + e^{w_0 + \sum w_i X_i}) \quad (3.2)$$

Προφανώς οι δύο πιθανότητες έχουν άθροισμα ίσο με την μονάδα.

Η ευκολία της μεθόδου έγκειται στο ότι οδηγούμαστε σε απλή γραμμική έκφραση για ταξινόμηση. Για να ταξινομήσουμε οποιοδήποτε δοσμένο X γενικότερα θέλουμε να προσδιορίσουμε την τιμή y_k που μεγιστοποιεί την πιθανότητα $P(Y=y_k|X)$

Αν $P(Y = 0|X) > P(Y = 1|X)$ τότε $0 < w_0 + \sum w_i X_i$ [14]

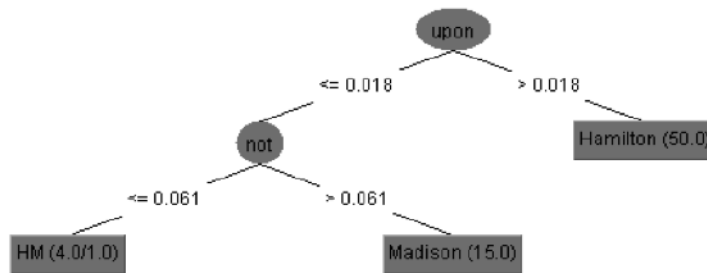
Μια λογιστική συνάρτηση ή λογιστική καμπύλη είναι ένα κοινό "S" σχήμα (Σχήμα 6) (σιγμοειδούς καμπύλης).



Σχήμα 6: Τυπική μορφή λογιστικής καμπύλης

3.2.2 Δέντρα Αποφάσεων

Ένα δέντρο απόφασης είναι ένα μοντέλο πρόβλεψης μηχανικής μάθησης που αποφασίζει για την τιμή στόχο ενός δείγματος με βάση διάφορες τιμές των χαρακτηριστικών των διαθέσιμων δεδομένων. Οι εσωτερικοί κόμβοι του δέντρου απόφασης δηλώνουν τα διαφορετικά χαρακτηριστικά. Τα «κλαδιά» μεταξύ των κόμβων μας δείχνουν τις πιθανές τιμές στα παρατηρούμενα δείγματα των διαφόρων μεταβλητών. Ενώ οι τερματικοί κόμβοι μας δείχνουν την τελική ταξινόμηση της εξαρτημένης μεταβλητής. Η τυπική μορφή του κατηγοριοποιητή παρουσιάζεται στο Σχήμα 7.



Σχήμα 7: Τυπική μορφή κατηγοριοποιητή J48 CONSOLIDATED

Το χαρακτηριστικό που πρόκειται να προβλεφθεί είναι γνωστή ως η εξαρτημένη μεταβλητή, αφού η αξία του εξαρτάται, ή αποφασίζεται από, τις τιμές όλων των άλλων ιδιοτήτων. Τα άλλα χαρακτηριστικά, τα οποία βοηθούν στην πρόβλεψη της τιμής της εξαρτημένης μεταβλητής, είναι γνωστά ως ανεξάρτητες μεταβλητές στο σύνολο δεδομένων.

Το δέντρο απόφασης J48 ακολουθεί τον εξής απλό αλγόριθμο. Για να ταξινομηθεί ένα νέο στοιχείο, πρέπει πρώτα να δημιουργηθεί ένα δέντρο απόφασης με βάση τις τιμές των χαρακτηριστικών των διαθέσιμων δεδομένων εκπαίδευσης. Έτσι, κάθε φορά που συναντά ένα σύνολο στοιχείων (σύνολο εκπαίδευσης) προσδιορίζει το χαρακτηριστικό που κάνει πιο διακριτές τις διάφορες περιπτώσεις. Αυτό το χαρακτηριστικό γνώρισμα που είναι σε θέση να μας πει περισσότερα σχετικά με τις περιπτώσεις των δεδομένων, έτσι ώστε να προκύπτει η καλύτερη ταξινόμηση μέσω υψηλότερου πληροφοριακού κέρδους. Τώρα, μεταξύ των πιθανών τιμών του χαρακτηριστικού αυτού, αν υπάρχει τιμή για την οποία δεν υπάρχει καμία ασάφεια, δηλαδή τιμές για τις οποίες οι περιπτώσεις των δεδομένων που εμπίπτουν στην κατηγορία της, έχουν την ίδια τιμή με τη μεταβλητή-στόχο, τότε τερματίζεται ο εν λόγω κλάδος και εκχωρείται η τιμή στόχος που έχουμε λάβει.

Για τις υπόλοιπες περιπτώσεις αναζητείται η αμέσως επόμενη μεταβλητή με μέγιστο πληροφοριακό κέρδος. Ως εκ τούτου, συνεχίζουμε με αυτόν τον τρόπο μέχρι, είτε να έχουμε σαφή εικόνα για το τι συνδυασμός των χαρακτηριστικών χρειάζεται για να υπάρχει σαφής απόφαση για την εξαρτημένη μεταβλητή είτε να ξεμεινουμε από ανεξάρτητες μεταβλητές. Σε περίπτωση που δεν υπάρχουν άλλες μεταβλητές, ή αν δεν μπορούμε να πάρουμε μια σαφή απόφαση από τις διαθέσιμες πληροφορίες, εκχωρείται στο κλαδί της εξαρτημένης μεταβλητής η τιμή της πλειοψηφίας των στοιχείων στο πλαίσιο του παρόντος κλαδιού.

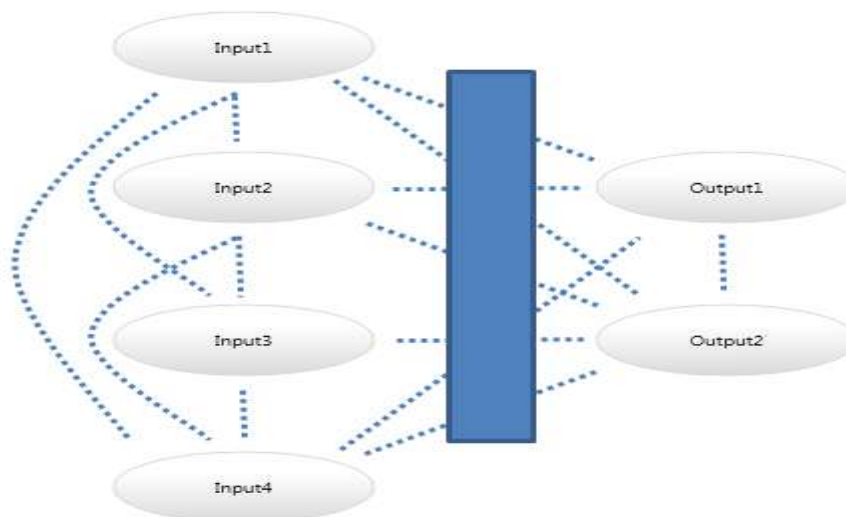
Τώρα που υπάρχει το δέντρο απόφασης, ακολουθείται η σειρά της επιλογής χαρακτηριστικών, όπως έχει ληφθεί για το δέντρο. Με τον έλεγχο όλων των αντίστοιχων χαρακτηριστικών και τις τιμές αυτών με εκείνες που παρατηρήθηκαν στο μοντέλο απόφασης, είναι δυνατή η πρόβλεψη της εξαρτημένης μεταβλητής. [15]

Η διαφοροποίηση του κατηγοριοποιητή J48 CONSOLIDATED από τον J48 έγκειται στο ότι διαθέτει περισσότερες επιλογές για την παραγωγή και προσδιορισμό των δειγμάτων που χρησιμοποιούνται στη διαδικασία σταθεροποίησης (consolidation). Με αυτόν τον τρόπο

εγγυάται το ελάχιστο ποσοστό των παραδειγμάτων του αρχικού δείγματος που περιέχονται ώστε να κατασκευαστούν τα επιμέρους δείγματα. [16]

3.2.3 Δίκτυα Bayes

Ένα Μπεϋζιανό δίκτυο (μοντέλο πεποιθήσεων) είναι ένα πιθανολογικό γραφικό μοντέλο που περιέχει ένα σύνολο τυχαίων μεταβλητών, με τυχηματικές εξαρτήσεις μεταξύ αυτών, μέσω ενός προσανατολισμένου μη κυκλικού γραφήματος. Η τυπική δομή του κατηγοριοποιητή Bayes φαίνεται στην παρακάτω εικόνα (Σχήμα 8).



Σχήμα 8: Γενική δομή του κατηγοριοποιητή BayesNet

Για να κατανοήσουμε το πως λειτουργεί η μέθοδος αυτή καλείται x ο βαθμός της πίστης σε ένα συμβάν. Ενώ η κλασσική πιθανότητα είναι φυσική ιδιότητα των συμβάντων (όπως για παράδειγμα το στρίψιμο ενός κέρματος με σαφή αποτελέσματα το κορόνα ή γράμματα), η μπεϋζιανή πιθανότητα είναι ιδιότητα του ατόμου στο οποίο ανατίθεται η πιθανότητα, δηλαδή είναι ο βαθμός πίστης ενός ατόμου για ένα συγκεκριμένο αποτέλεσμα. Για να ξεχωρίζουμε τους όρους καλείται η κλασσική πιθανότητα ως αληθινή ή φυσική πιθανότητα ενός συγκεκριμένου συμβάντος, ενώ όταν αναφερόμαστε στον βαθμό πεποίθησης ενός ατόμου καλείται μπεϋζιανή ή προσωπική πιθανότητα.

Μια σημαντική διαφορά της φυσικής και της προσωπικής πιθανότητας είναι ότι στην προσωπική πιθανότητα δεν χρειάζονται επαναλαμβανόμενες δοκιμές, χρειάζεται απλά η πεποίθηση για την έκβαση ενός συμβάντος. Σε αντίθεση με την κλασσική πιθανότητα και στατιστική που λαμβάνει για παράδειγμα το πως θα προσγειωθεί το ζάρι.

Συνηθισμένη κριτική στη θεωρία του Bayes για τον ορισμό της προσωπικής πιθανότητας είναι ότι οι πιθανότητες αυτή είναι αυθαίρετη και ότι δεν ικανοποιεί τους κανόνες της κλασσικής πιθανότητας. Επιπλέον υπήρχε η αμφισβήτηση ως προς σε ποιο βαθμό θα μπορούσε να μετρηθεί. Για αυτό τον λόγο έχει μελετηθεί εντατικά από πολλούς μελετητές. Όλοι οι μελετητές που ασχολήθηκαν κατέληξαν στους νόμους των πιθανοτήτων ώστε να μελετήσουν την μέτρηση των πεποιθήσεων. Επίσης θα μπορούσε να ειπωθεί ότι ένας συγκεκριμένος αριθμός στην πιθανότητα x αν και δεν μας οδηγεί με σιγουριά στην τελική

προσωπική πιθανότητα, οδηγεί σε αποφάσεις που βασίζονται σε πιθανότητες, με αυτές τις πιθανότητες να μην είναι ευαίσθητες σε μικρά εύρη πιθανοτήτων.

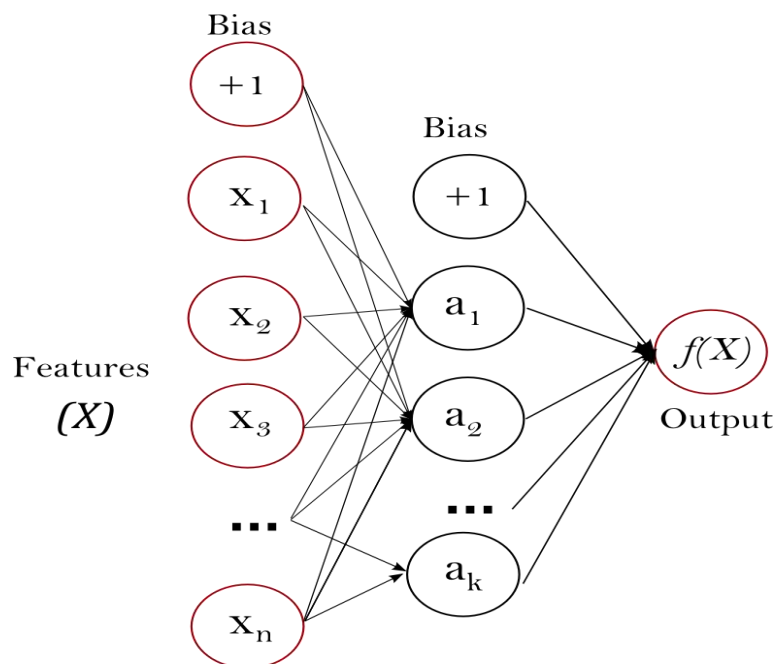
Η διαδικασία της ταξινόμησης με Bayes Net συνοψίζεται στο ότι υπάρχουν ένα σύνολο μεταβλητών $V=\{A_1, A_2, \dots, A_n\}$ και μια σειρά E προσανατολισμένων αλληλεπιδράσεων μεταξύ των μεταβλητών. Τα δύο αυτά σύνολα δημιουργούν ένα προσανατολισμένο ακυκλικό διάγραμμα $G=(V, E)$. Στο μπεϋζιανό δίκτυο οι μεταβλητές που είναι ανεξάρτητες δεν ενώνονται μεταξύ τους. Όσο πιο ανεξάρτητες είναι οι μεταβλητές μεταξύ τους τόσο πιο λίγες οι παράμετροι που απαιτούνται ώστε να δημιουργηθεί η δεσμευμένη πιθανότητα, που είναι το τελικό προϊόν στην μέθοδο αυτή.

$$P(A_1, A_2, \dots, A_n) = \prod_{i=1}^N P\left(\frac{A_i}{P_{Ai}}\right) \quad (3.3)$$

Όπου $P(A_i/P_{Ai})$ είναι η υπό όρους συνεισφορά της μεταβλητής A_i , δεδομένης της P_{Ai} δηλαδή της ύπαρξης των μεταβλητών που την επηρεάζουν. Η δεσμευμένη αυτή πιθανότητα για κάθε μεταβλητή έχει παραμετρικό τύπο από τον οποίο μαθαίνουμε την μέγιστη εκτίμηση πιθανοτήτων.

3.2.4 Πολυεπίπεδα Νευρωνικά Δίκτυα

Η ταξινόμηση με τα πολυεπίπεδα πρόσθιας τροφοδότησης Perceptron (multi-layer perceptron- MLP) γίνεται με εποπτευόμενο μαθησιακό αλγόριθμο, στον οποίο εισάγονται m αριθμός μεταβλητών και y αριθμός μεταβλητών εξόδου. Λαμβάνοντας υπόψη ένα σύνολο χαρακτηριστικών $X = \{x_1, x_2, \dots, x_n\}$ και y στόχο, μπορούν να μάθουν μια μη γραμμική συνάρτηση προσέγγισης είτε για ταξινόμηση είτε για παλινδρόμηση. Είναι διαφορετική από την λογιστική παλινδρόμηση διότι μεταξύ εισόδου και εξόδου μπορεί να δημιουργηθούν ένα ή και περισσότερα μη-γραμμικά στρώματα (κρυμμένα στρώματα). Στο παρακάτω σχήμα (Σχήμα 9) παρουσιάζεται ένα κρυμμένο στρώμα σε πολυεπίπεδη ταξινόμηση.



Σχήμα 9: Γενική δομή κατηγοριοποιητή MLP [17]

Η αριστερή στρώση, γνωστή ως στρώση εισόδου, αποτελείται από ένα σύνολο νευρώνων $\{x_i | x_1, x_2, \dots, x_m\}$ που αντιπροσωπεύει τις δυνατότητες εισόδου. Κάθε νευρώνας στο κρυφό στρώμα μετατρέπει τις τιμές από το προηγούμενο στρώμα με μία σταθμισμένη γραμμική άθροιση $w_1x_1 + w_2x_2 + \dots + w_mx_m$ και ακολουθείται μια μη γραμμική συνάρτηση $g(\cdot) : R \rightarrow R$ όπως η συνάρτηση της υπερβολικής εφαπτομένης. Η στρώση εξόδου λαμβάνει τιμές από την τελευταία κρυμμένη στρώση και τις μετατρέπει σε τιμές εξόδου. [17]

3.3 Αξιολόγηση προτύπων

Τα τελικά αποτελέσματα των μοντέλων με τους διαφορετικούς ταξινομητές στο πρόγραμμα Weka αξιολογούνται με διάφορους δείκτες. Για να οριστούν οι δείκτες πρέπει πρώτα να οριστούν οι ποσότητες TN, FP, FN και TP:

- 1) TN ο κοινός αριθμός των μη συμβάντων που καταγράφηκαν από το OBD και το Smartphone
- 2) FP ο αριθμός των μη συμβάντων σύμφωνα με το OBD που λανθασμένα καταγράφηκε από το Smartphone ως συμβάν
- 3) FN ο αριθμός των συμβάντων σύμφωνα με το OBD που λανθασμένα δεν καταγράφηκε από το Smartphone ως μη συμβάν
- 4) TP ο κοινός αριθμός των συμβάντων που καταγράφηκαν από το OBD και το Smartphone

Η αξιολόγηση ενός κατηγοριοποιητή βασίζεται στον Πίνακα μέσω τεχνικής μείωσης του δείγματος Κατηγοριοποίησης που έχει την παρακάτω μορφή (Πίνακας 1).

Πίνακας 1: Τυπικός πίνακας κατηγοριοποίησης.

		OBD	
		NORMAL	EVENT
smartphone	NORMAL	TN	FN
	EVENT	FP	TP

Οι βασικοί δείκτες αξιολόγησης που θα χρησιμοποιηθούν στην αξιολόγηση των αποτελεσμάτων των κατηγοριοποιητών, που θα αναπτυχθούν στο πλαίσιο της παρούσας διπλωματικής, είναι οι ακόλουθοι:

$$\text{Το ποσοστό σωστών προβλέψεων } TPR = \frac{TP}{TP+FN} \quad (3.4)$$

$$\text{Το ποσοστό εσφαλμένων προβλέψεων } FPR = \frac{FP}{FP+TN} \quad (3.5)$$

$$\text{Ο δείκτης F1 (δείκτης ακρίβειας) } F1 = \frac{2*TP}{2*TP+FP+FN} \quad (3.6)$$

Δείκτης ROC AREA (δείκτης ακρίβειας) $TPR = f(FPR)$ για διάφορα επίπεδα ορίων

$$\text{Ευκρίνεια } PRECISION = \frac{TP}{TP+FP} \quad (3.7)$$

$$\text{Δείκτης ανάκλησης } \mathit{RECALL} = \frac{\mathit{TP}}{\mathit{TP} + \mathit{FN}} \quad (3.8)$$

Κεφάλαιο 4

4. Συλλογή και Επεξεργασία Δεδομένων

4.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζεται η διαδικασία που ακολουθήθηκε για τη συλλογή και επεξεργασία των δεδομένων και σχολιάζεται η ποιότητα και η αξιοπιστία τους. Γίνεται επίσης αναφορά στα προβλήματα που προέκυψαν κατά την επεξεργασία των δεδομένων και παρουσιάζεται ο τρόπος κωδικοποίησής τους και η προκαταρκτική επεξεργασία που υπέστησαν.

4.2 Το Πείραμα

Το πείραμα που πραγματοποιήθηκε προσομοιάζει με μια νατουραλιστική μελέτη οδήγησης με εξαίρεση ότι οι συμμετέχοντες κλήθηκαν να ενεργοποιούν μια εφαρμογή smartphone, η οποία κατέγραφε δεδομένα από αισθητήρες του smartphone, κάθε φορά που εισέρχονταν στο αυτοκίνητο και πριν από την έναρξη του ταξιδιού τους. Τα δεδομένα καταγράφονταν κάθε 1 δευτερόλεπτο και οι οδηγοί έστελναν τα καταγεγραμμένα δεδομένα μία φορά την ημέρα σε ένα διακομιστή. Τρεις οδηγοί για δύο μήνες συμμετείχαν χρησιμοποιώντας εναλλασσόμενα διάφορες συσκευές smartphone και tablet. Οι συμμετέχοντες είχαν τη δυνατότητα να τοποθετήσετε τηλέφωνα ή ταμπλέτες τους σε οποιαδήποτε θέση και είχαν τη δυνατότητα να μιλήσουν, να στείλουν μηνύματα, με τα κινητά τους τηλέφωνα κατά την οδήγηση. Οχήματα εξοπλισμένα με συσκευές OBD-II, η οποία παρείχε τη συνεχή παρακολούθηση των χαρακτηριστικών των οχημάτων και τη συμπεριφορά οδήγησης μέσω βαθμονομημένων οριακών τιμών για τα δεδομένα επιταχυνσιομέτρου.

4.3 Συλλογή στοιχείων

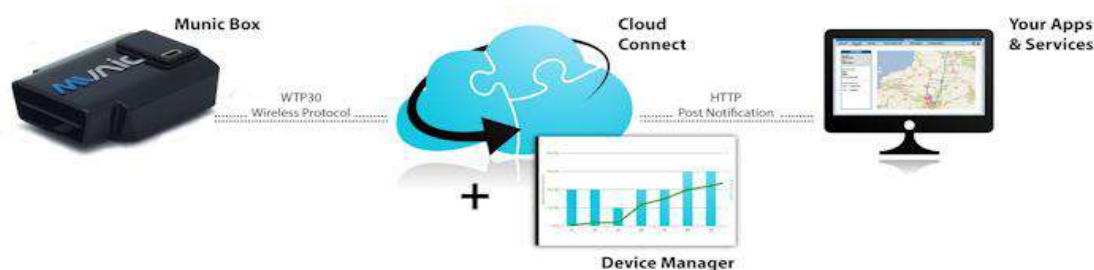
Στην παρούσα Διπλωματική Εργασία, τα δεδομένα προήλθαν από δύο διαφορετικές πηγές:

- Συσκευή στη θύρα OBD-II του οχήματος
- Προσωπική συσκευή Smartphone των συμμετεχόντων στο πείραμα

Το σύστημα OBD αναπτύχθηκε στις Ηνωμένες Πολιτείες Αμερικής από την General Motors το 1981 και είχε ως στόχο την ανίχνευση μηχανικών προβλημάτων στο όχημα, τα οποία προκαλούσαν υψηλά επίπεδα εκπομπής ρύπων, πάνω από τα αποδεκτά όρια.

Το OBD (επί του αυτοκινήτου διάγνωση) είναι ένα μηχάνημα διαστάσεων 27 x 48 x 49.5 cm το οποίο συνδέεται με την κεντρική μονάδα ελέγχου του αυτοκινήτου και βρίσκεται εντός αυτού. Όλα τα δεδομένα που είναι απαραίτητα για την διερεύνηση της κυκλοφοριακής συμπεριφοράς του οδηγού στέλνονται άμεσα στον ηλεκτρονικό υπολογιστή ή το κινητό του χρήστη.

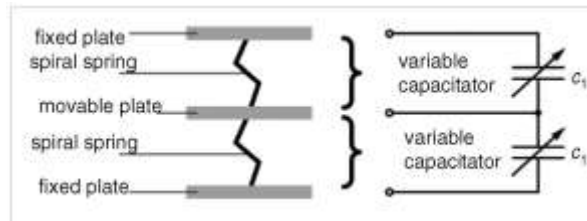
Η σύνδεση μεταξύ της ενσωματωμένης συσκευής στο όχημα και της υπηρεσίας αποθήκευσης δεδομένων (Cloud Connect), όπου αποκτάς πρόσβαση με την δημιουργία λογαριασμού, γίνεται μέσω ενός προηγμένου πρωτοκόλλου το οποίο επιτρέπει σε πραγματικό χρόνο μεταφορά δεδομένων, συμπίεση όλων των αρχείων και βελτιωμένη διαχείριση των δεδομένων παρακολούθησης της οδικής συμπεριφοράς. [18] [19]



Σχήμα 10: Σχηματική απεικόνιση λειτουργίας συστήματος [19]

Το Smartphone (iPhone 6, iPad2, iPhone 4s) διαθέτε την εφαρμογή SensorLog. Το SensorLog είναι μια εφαρμογή διαθέσιμη δωρεάν από το κατάστημα εφαρμογών (App Store) της Apple η οποία περιλαμβάνει την καταγραφή και αποθήκευση των στοιχείων πολλών μετρητών που βρίσκονται ήδη στο κινητό όπως επιταχυνσιόμετρο (accelerometer), πυξίδα, γυροσκόπιο, GPS. Για κάθε διαδρομή του χρήστη, η εφαρμογή τίθεται σε λειτουργία και τα αρχεία αποθηκεύονται στο Smartphone με την μορφή CSV (Comma Separated Values) και παράλληλα στέλνονται μέσω ηλεκτρονικού ταχυδρομείου στον ηλεκτρονικό υπολογιστή.

Ένα από τα σημαντικότερα χαρακτηριστικά των έξυπνων κινητών που χρησιμοποιούμε είναι το **επιταχυνσιόμετρο (accelerometer)**. Επιταχυνσιόμετρα είναι συσκευές που μπορούν να μετρήσουν την επιτάχυνση (το ρυθμό μεταβολής της ταχύτητας), αλλά σε smartphones, είναι σε θέση να ανιχνεύσει τις αλλαγές στον προσανατολισμό ώστε να περιστραφεί η οθόνη για την ευκολία του χρήστη. Παρόλο που το επιταχυνσιόμετρο χρησιμοποιείται κυρίως για παιχνίδια, βίντεο και άλλες ομοειδές δραστηριότητες λίγοι άνθρωποι ξέρουν πώς λειτουργεί πραγματικά το gadget, ή πώς οι μηχανικοί ήταν σε θέση να αξιοποιήσουν ένα τόσο μικρό αλλά σημαντικό κομμάτι της τεχνολογίας, η οποία μπορεί να ανιχνεύσει κίνηση σε τρεις κατευθύνσεις, μέσα σε μια χιλιοστά πάχους. Στις περισσότερες περιπτώσεις, μεταξύ αυτών και στα «έξυπνα» κινητά, η μέτρηση γίνεται με τη χωρητική μέθοδο διότι είναι μια δοκιμασμένη και αξιόπιστη μέθοδος (Handbook of Modern Sensors Physics Designs and Applications). [20]

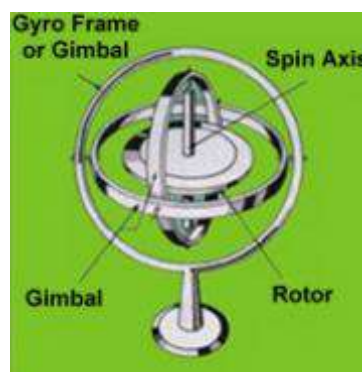


Σχήμα 11: Απλοποιημένη διάταξη χωρητικού επιταχυνσιόμετρου [20]

Τρία φύλλα πυριτίου, τα οποία τοποθετούνται παράλληλα το ένα στο άλλο και συνδέονται με σπειροειδή ελατήρια, σχηματίζουν μια σύνδεση πυκνωτών σε σειρά. Τα δύο εξωτερικά φύλλα είναι σταθερά ενώ το εσωτερικό, το οποίο παίζει το ρόλο της μάζας, έχει τη δυνατότητα να μετακινείται. Οποιαδήποτε μετακίνηση του μεσαίου φύλλου προκαλεί διαφορά στην χωρητικότητα των πυκνωτών η οποία μετρείται και μετατρέπεται σε μετρούμενη επιτάχυνση. Για να κυριολεκτήσουμε, ως εκ τούτου δεν είναι αισθητήρες επιτάχυνσης, αλλά αισθητήρες δύναμης.

Επόμενος μετρητής είναι το **γυροσκόπιο (Gyroscope)**, το οποίο ελέγχει τις αλλαγές στην γωνιακή ταχύτητα και συνεργαζόμενο με το επιταχυνσιόμετρο βελτιώνει την ευαισθησία της οθόνης όταν αλλάζει ο προσανατολισμός. Το γυροσκόπιο βοηθάει πολύ στην απόκριση της συσκευής.

Το γυροσκόπιο εφαρμόζει δύο βασικές αρχές της φυσικής α) την αρχή διατήρησης της στροφορμής και β) την αρχή της αδρανείας με σκοπό να παραμένει σταθερός ο προσανατολισμός της συσκευής. Ο σταθερός προσανατολισμός επιτυγχάνεται μέσω της περιστροφής των μερών της. Πρόκειται για μια διάταξη όμοια με εκείνη που φέρει η σχολική υδρόγειος σφαίρα. Αντί όμως της υδρόγειου φέρεται μια μεταλλική στεφάνη που μπορεί να περιστρέφεται δεξιά ή αριστερά. Αυτή η στεφάνη φέρει δεύτερη εσωτερική που στηρίζεται με συνδέσμους σε οριζόντια διάταξη, ως προς την εξωτερική, δυνάμενη έτσι να περιστρέφεται ελεύθερα με φορά πάνω ή κάτω. Στην εσωτερική αυτή στεφάνη συγκρατείται εσωτερικά σε κάθετη διάταξη σε σχέση με τη προηγούμενη ο "σφόνδυλος" που αποτελεί μια μικρή σφαίρα που περιστρέφεται υπό μορφή σβούρας. Τα σημεία έδρασης της κάθε στεφάνης καθώς και του σφονδύλου εξασφαλίζουν την ελεύθερη περιστροφή όλων των τμημάτων της διάταξης, δηλαδή του γυροσκοπίου σαν σύνολο.



Σχήμα 12: Τυπική διάταξη γυροσκοπίου [21]

Οι τιμές που διαβάζονται από τον μετρητή είναι ουσιαστικά η κυκλική ταχύτητα του έξυπνου κινητού. [21]

Ακόμα καταγράφεται μέσω των αισθητήρων ο προσανατολισμός της συσκευής (device Orientation). Πράγμα πολύ σημαντικό διότι οι τιμές των μεταβλητών οι οποίες καθορίζονται από τους άξονες των τριών διαστάσεων χρειάζονται κατάλληλη προσαρμογή ως προς μία συγκεκριμένη διεύθυνση. Για αυτόν ακριβώς τον λόγο υπάρχουν στην βάση δεδομένων οι στήλες OldAccel, NewAccel, NewRot των διευθύνσεων X, Y και Z.

Επιπρόσθετα μία άλλη εφαρμογή που χρησιμοποιείται είναι το γνωστό σε όλους μας GPS (Global Positioning System) ή αλλιώς Παγκόσμιο Σύστημα Εντοπισμού.

Τα χαρακτηριστικά που χρησιμοποιούμε είναι το γεωγραφικό μήκος και το γεωγραφικό πλάτος ώστε να γνωρίζουμε την ακριβή θέση του οδηγού.

Τέλος είναι σημαντικό να γνωρίζουμε την ακριβή στιγμή της καταγραφής των συμβάντων μέσω μιας συγκεκριμένης σήμανσης χρόνου (Timestamp) που με κατάλληλη μετατροπή μας δείχνει όλες τις χρονικές λεπτομέρειες του συμβάντος (μέρα, ώρα, ημερομηνία κ.τ.λ.).

Οι μετρήσεις γίνονται ταυτόχρονα από το OBD και από το Smartphone και προσπαθούμε να βρούμε αν τα δύο όργανα εντοπίζουν τα ίδια συμβάντα, τα ανώτερα και τα κατώτερα όρια για τις επιταχύνσεις ελέγχου αλλά και την χρονική διαφορά στις διάφορες περιπτώσεις. Σκοπός δηλαδή είναι η ελαχιστοποίηση των σφαλμάτων μεταξύ OBD και Smartphone και η μεγιστοποίηση της αξιοπιστίας των αποτελεσμάτων μας ώστε με ακρίβεια να μπορούμε να εκτιμήσουμε την οδηγική συμπεριφορά, θέμα το οποίο αποτελεί και το αντικείμενο της παρούσας Διπλωματικής Εργασίας.

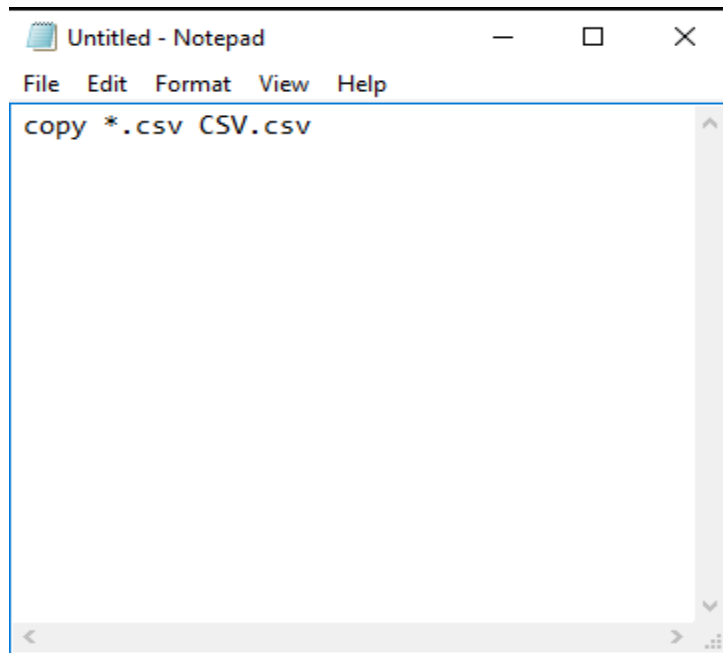
4.4 Προκαταρκτική Επεξεργασία Στοιχείων

4.4.1 Διαμόρφωση δεδομένων

Τα δεδομένα καταγράφονται από τις δύο συσκευές με συχνότητα 1 Hz και στέλνονται για διαμόρφωση στον ηλεκτρονικό υπολογιστή σε μορφή csv.

Πιο απλά, είναι στην ουσία αρχεία που περιέχουν μέσα τους δεδομένα υπό μορφή Excel, δηλαδή περιέχουν πολλαπλά πεδία με μία ή περισσότερες τιμές για κάθε τέτοιο πεδίο, όπως συμβαίνει με έναν πίνακα excel πχ με πολλαπλές στήλες και γραμμές. Το csv είναι τα αρχικά των λέξεων Comma Separated Values που υποδηλώνει στην ουσία μία σειρά μεταβλητών με τιμές μέσα τους, χωρισμένες απλά με ένα κόμμα.

Για να δημιουργηθεί η βάση δεδομένων ενώθηκαν όλα τα πρωτογενή αρχεία σε ένα. Για να γίνει αυτό χρειάστηκε η δημιουργία BAT file.



Σχήμα 13: α) Δημιουργία εκτελέσιμου αρχείου για την ένωση των csv



Σχήμα 13: β) Δημιουργία εκτελέσιμου αρχείου για την ένωση των csv

Η δημιουργία του ήταν απλή και χρειάστηκε η δημιουργία ενός σημειώματος με την χρήση του σημειωματάριου (Notepad) και η αποθήκευση του με την κατάληξη .bat . Το μόνο που έχουμε να κάνουμε για να ενωθούν τα αρχεία με τα ταξίδια είναι να βάλουμε το batch file σε κοινό φάκελο με τα υπόλοιπα csv και να το <<τρέξουμε>>.

Τα συμβάντα διαχωρίζονται σε δυο κατηγορίες:

1. Τα γεγονότα απότομης στροφής (harsh cornering)
2. Τα γεγονότα απότομης επιτάχυνσης και επιβράδυνσης (harsh acceleration)

Για τα συμβάντα απότομης στροφής απαιτείται η χρήση του γυροσκοπίου και λαμβάνεται υπ' όψιν το πόσο γρήγορα μεταβάλλονται οι τιμές της γωνιακής ταχύτητας.

Για τα συμβάντα απότομης στροφής απαιτείται η χρησιμοποίηση του επιταχυσιομέτρου και του Παγκόσμιου Συστήματος Εντοπισμού (GPS) καθώς επίσης χρειάζεται και ο προσανατολισμός των τριών αξόνων του έξυπνου κινητού με βάση τους άξονες του αυτοκινήτου (re-orientation). Τα μεγέθη που αναζητήθηκαν προκειμένου να βρεθούν τα συμβάντα ήταν οι μεγαλύτερες τιμές της μεταβολής του επιταχυνσιομέτρου αλλά και το

πρόσημο των τιμών προκειμένου να αναγνωρίζεται αν είναι λόγω επιβράδυνσης ή επιτάχυνσης.

4.4.2 Διαθέσιμη Βάση Δεδομένων

Η βάση δεδομένων μας που προήλθε από την διεξαγωγή του πειράματος περιέχει 1.484.466 σειρές καταγεγραμμένων δεδομένων εκ των οποίων 2992 2167 είναι τα γεγονότα απότομης στροφής (harsh cornering) με την ένδειξη smartphnepeaksx και 5930 είναι τα γεγονότα απότομης επιτάχυνσης και επιβράδυνσης (harsh acceleration) με την ένδειξη smartphnepeaksy.

Οι μεταβλητές που χρησιμοποιούνται για την ανίχνευση των γεγονότων απότομης στροφής (harsh cornering) είναι οι εξής:

1. Η ταχύτητα του οχήματος
2. Η τιμή της διαφοράς της στιγμιαίας ταχύτητας με την ταχύτητα 2 δευτερολέπτων νωρίτερα
3. Η κυκλική ταχύτητα
4. Η τιμή της απόστασης μεταξύ δύο συνεχόμενων σημείων με ξεχωριστό γεωγραφικό μήκος και πλάτος

Οι μεταβλητές που χρησιμοποιούνται για την ανίχνευση των γεγονότων απότομης επιτάχυνσης και επιβράδυνσης (harsh acceleration) είναι οι εξής:

1. Η ταχύτητα του οχήματος
2. Η ταχύτητα του οχήματος δύο δευτερόλεπτα πριν το παρατηρηθέν γεγονός
3. Η κυκλική ταχύτητα
4. Η τιμή της απόστασης μεταξύ δύο συνεχόμενων σημείων με ξεχωριστό γεωγραφικό μήκος και πλάτος

4.5 Ανάπτυξη και Αξιολόγηση Κατηγοριοποιητών

4.5.1 Προετοιμασία βάσεων δεδομένων

Διαχωρισμός βάσης δεδομένων για εκπαίδευση και γενίκευση

Η βάση δεδομένων σε δύο αρχεία το final csv 1 και το final csv 2. Το μεν final csv 1 έχει συνολικά 1.005.665 σειρές δεδομένων εκ των οποίων 2167 είναι τα γεγονότα απότομης στροφής (harsh cornering) με την ένδειξη smartphnepeaksx και 4565 είναι τα γεγονότα απότομης επιτάχυνσης και επιβράδυνσης (harsh acceleration) με την ένδειξη smartphnepeaksy.

Το αρχείο final csv 2 έχει 478.801 σειρές καταγεγραμμένων δεδομένων εκ των οποίων 825 είναι τα γεγονότα απότομης στροφής (harsh cornering) με την ένδειξη smartphnepeaksx και 1365 είναι τα γεγονότα απότομης επιτάχυνσης και επιβράδυνσης (harsh acceleration) με την ένδειξη smartphnepeaksy.

Το πρώτο αρχείο αποφασίστηκε να είναι το αρχείο εκπαίδευσης των κατηγοριοποιητών (training set), ενώ το 2 θα χρησιμοποιηθεί για τον έλεγχο της κατηγοριοποίησης (test set).

Στο Παράρτημα I παρουσιάζονται τα στατιστικά χαρακτηριστικά του δείγματος της εκπαίδευσης για τις συνεχείς μεταβλητές του δείγματος.

Στρατηγική Δειγματοληψίας για Μη Ισορροπημένα Δείγματα

Ένα τυπικό θέμα σε προβλήματα κατηγοριοποίησης είναι η ανάλυση μη ισορροπημένων δειγμάτων. Στη συγκεκριμένη εφαρμογή, το δείγμα των οδηγικών συμβάντων (απότομες επιβραδύνσεις, επιταχύνσεις και στροφές) σε σχέση με το δείγμα των μη οδηγικών συμβάντων είναι εξαιρετικά μικρό (περίπου 1:500). Τα δύο υποσύνολα δηλαδή έχουν πολύ μεγάλη ανομοιογένεια και υπερβολικά ασύμμετρες κατανομές. Η μεγάλη αυτή ανομοιογένεια επηρεάζει την αποδοτικότητα και την ακρίβεια των αποτελεσμάτων των αλγορίθμων μηχανικής μάθησης. Για αυτόν τον λόγο αποφασίζουμε με ειδικό τρόπο την αύξηση του ποσοστιαίου δείγματος των συμβάντων έτσι ώστε να δούμε κατά πόσο αυξάνεται η αποδοτικότητα των αλγορίθμων.

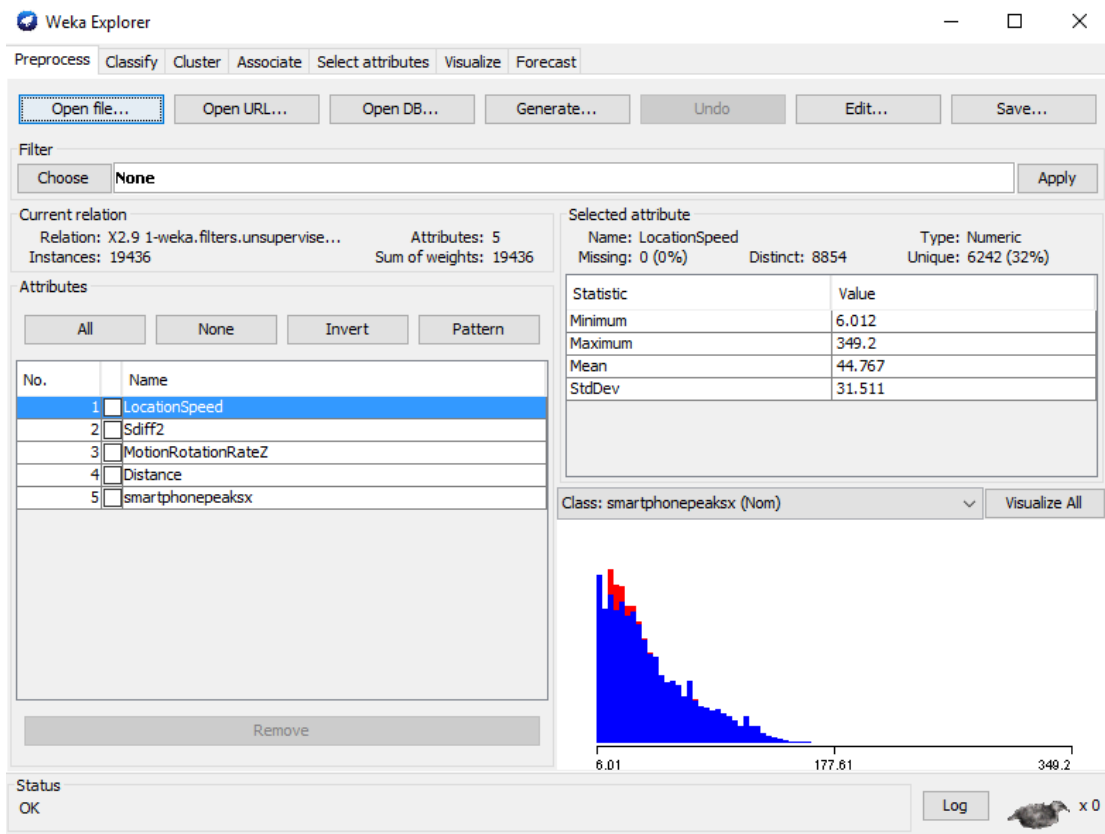
Ο αλγόριθμος που χρησιμοποιήθηκε στην παρούσα διπλωματική έτσι ώστε να επιτευχθεί η ποσοστιαία αύξηση του δείγματος των συμβάντων είναι ο ακόλουθος. Αναπτύχθηκε κώδικας στην R που αφαιρεί τυχαία ορισμένα καταγεγραμμένα μη- συμβάντα του κλάδου 0, που είναι ο κλάδος με την πλειοψηφία των περιπτώσεων, ενώ παράλληλα διατηρεί όλες τις περιπτώσεις του κλάδου 1 (συμβάντα) με σκοπό την δημιουργία μιας πιο ισορροπημένη βάση δεδομένων. Επιτρέπει δύο τρόπους εκτέλεσης της μείωσης του δείγματος (under-sampling) :

- Θέτοντας το ποσοστό των συμβάντων που θέλουμε στον κλάδο 1 μετά το τέλος της διαδικασίας της αλγορίθμου (percPos method)
- Θέτοντας το ποσοστό των μη συμβάντων (percUnder method)

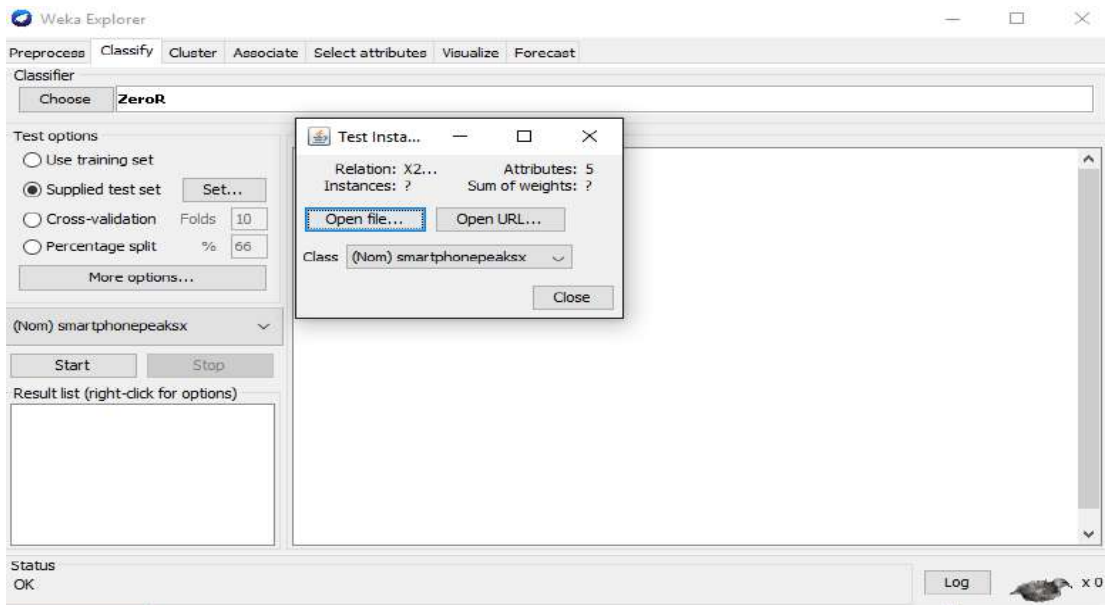
Για την μέθοδο percPos το ποσοστό θα παίρνει τιμές $(N.1/N * 100) \leq \text{ποσοστό} \leq 50$ όπου N.1 ορίζεται ο αριθμός των συμβάντων και N ο αριθμός των συνολικών περιπτώσεων (άθροισμα μη- συμβάντων και συμβάντων). Για την μέθοδο percUnder το ποσοστό θα παίρνει τιμές $(N.1/N.0 * 100) \leq \text{ποσοστό} \leq 100$, όπου N.1 ορίζεται ο αριθμός των συμβάντων και N.0 το σύνολο των μη- συμβάντων.

Με την παραπάνω διαδικασία δημιουργήθηκαν διάφορα αρχεία που έχουν ως ποσοστά συμβάντων τα 1%, 1.5%, 2%, 3% , 4% και 5%.

Στην συνέχεια στο πρόγραμμα weka ανοίγουμε τα αντίστοιχα αρχεία ως set data και training data όπου με κατάλληλη μορφοποίηση των δεδομένων καθίσταται η μεταβλητή smartphonepeaksx ή smartphonepeaksy από αριθμητική (numeric) σε διακριτή (nominal), καθώς επίσης απαλείφονται όλες οι τιμές των ταχυτήτων οι οποίες είναι μικρότερες των 6 km/h ως ταχύτητα βαδίσματος.



Σχήμα 14: Χρησιμοποίηση φίλτρων για την κατάλληλη χρήση των μεταβλητών

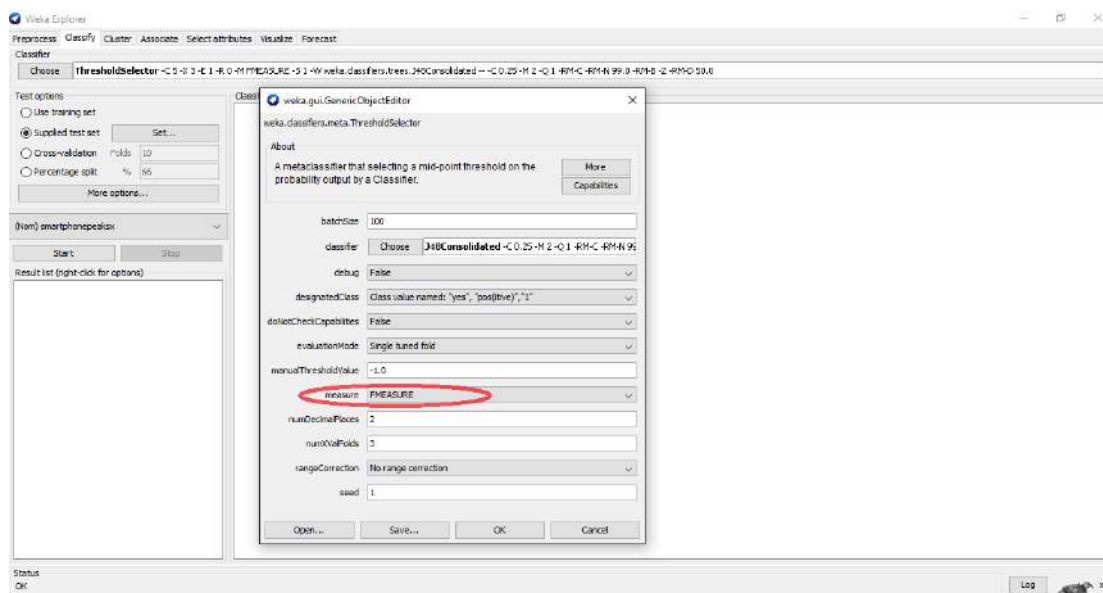


Σχήμα 4: Φόρτωση set data

Στρατηγική Βελτιστοποίησης Κατηγοριοποίησης

Τα δίκτυα που εκπαιδεύονται, παράγουν μια πιθανότητα να συνέβη μια συγκεκριμένη χρονική στιγμή ένα συμβάν, είτε στον εγκάρσιο, είτε στον διαμήκη άξονα του οχήματος.

Στην παρούσα Διπλωματική Εργασία, ο κατηγοριοποιητής, που μαθαίνει μέσω της εκπαίδευσης πάνω σε συγκεκριμένα δεδομένα (Σχήμα 14,15), επιλέγει σταδιακά το συγκεκριμένο όριο με στόχο τη μεγιστοποίηση της τιμής του δείκτη F-Measure (Σχήμα 16) που ορίζεται από τη σχέση 3.6.



Σχήμα 5: Εύρεση ορίου ώστε να μεγιστοποιείται ο όρος F-Measure

4.5.2 Αποτελέσματα μοντέλων

Συμβάντα απότομης στροφής

Στους Πίνακες 2 έως 7 φαίνονται, για τα ποσοστά 1%, 1,5%, 2%, 3%, 4% και 5% συμβάντων απότομης στροφής, τα αποτελέσματα των τεσσάρων διαφορετικών κατηγοριοποιητών³ (το Δέντρο Απόφασης, το Μπεϋζιανό δίκτυο, τα Πολυεπίπεδα Νευρωνικά Δίκτυα, και η Λογιστική Παλινδρόμηση).

Πίνακας 2: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 1% συμβάντα απότομων στρεφουσών κινήσεων.

	j48 consolidated		Bayesian Network		MLP		Logistic Regression	
	Κανονική Οδήγηση	Συμβάν	Κανονική Οδήγηση	Συμβάν	Κανονική Οδήγηση	Συμβάν	Κανονική Οδήγηση	Συμβάν
TP RATE	0.99	0.79	1.00	0.23	1.00	0.00	1.00	0.00
FP RATE	0.21	0.01	0.77	0.00	1.00	0.00	1.00	0.00
Precision	1.00	0.16	1.00	0.28	1.00	0.00	1.00	0.00
Recall	0.99	0.79	1.00	0.23	1.00	0.00	1.00	0.00
F-Measure	0.99	0.27	1.00	0.25	1.00	0.00	1.00	0.00
MCC	0.35	0.35	0.25	0.25	0.00	0.00	0.00	0.00
ROC Area	0.97	0.97	0.99	0.99	0.67	0.67	0.67	0.67
PRC Area	1.00	0.16	1.00	0.23	1.00	0.01	1.00	0.01

Πίνακας 3: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 1,5% συμβάντα απότομων στρεφουσών κινήσεων.

	J48 consolidated		Bayesian Network		MLP		Logistic Regression	
	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν
TP RATE	0.983	0.851	0.991	0.668	1	0	1	0
FP RATE	0.149	0.017	0.332	0.009	1	0	1	0
Precision	1	0.139	0.999	0.191	0.997	0	0.997	0
Recall	0.983	0.851	0.991	0.668	1	0	1	0
F-Measure	0.991	0.238	0.995	0.297	0.998	0	0.998	0
MCC	0.339	0.339	0.354	0.354	0	0	0	0
ROC Area	0.952	0.952	0.99	0.99	0.683	0.683	0.675	0.675
PRC Area	1	0.136	1	0.211	0.999	0.005	0.998	0.005

Πίνακας 4: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 2% συμβάντα απότομων στρεφουσών κινήσεων.

	J48 consolidated		Bayesian Network		MLP		Logistic Regression	
	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν
TP RATE	0.982	0.91	0.994	0.517	1	0	1	0
FP RATE	0.09	0.018	0.483	0.006	1	0	1	0
Precision	1	0.131	0.999	0.216	0.997	0	0.997	0
Recall	0.982	0.91	0.994	0.517	1	0	1	0
F-Measure	0.991	0.229	0.996	0.305	0.998	0	0.998	0
MCC	0.341	0.341	0.332	0.332	0	0	0	0
ROC Area	0.976	0.976	0.987	0.987	0.688	0.688	0.691	0.691
PRC Area	1	0.131	1	0.201	0.999	0.005	0.999	0.005

Πίνακας 5 : Αποτελέσματα κατηγοριοποίησης στο δείγμα με 3% συμβάντα απότομων στρεφουσών κινήσεων.

	J48 consolidated		Bayesian Network		MLP		Logistic Regression	
	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν
TP RATE	0.986	0.509	0.994	0.506	1	0	1	0
FP RATE	0.491	0.014	0.494	0.006	1	0	1	0
Precision	0.978	0.624	0.998	0.22	0.997	0	0.997	0
Recall	0.986	0.509	0.994	0.506	1	0	1	0
F-Measure	0.982	0.561	0.996	0.306	0.998	0	0.998	0
MCC	0.546	0.546	0.33	0.33	0	0	0	0
ROC Area	0.986	0.986	0.983	0.983	0.675	0.675	0.682	0.682
PRC Area	0.999	0.606	1	0.187	0.998	0.005	0.998	0.005

Πίνακας 6: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 4% συμβάντα απότομων στρεφουσών κινήσεων.

	J48 consolidated		Bayesian Network		MLP		Logistic Regression	
	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν
TP RATE	0.978	0.893	0.995	0.417	1	0	0.998	0.004
FP RATE	0.107	0.022	0.583	0.005	1	0	0.996	0.002
Precision	1	0.112	0.998	0.206	0.997	0	0.997	0.008
Recall	0.978	0.893	0.995	0.417	1	0	0.998	0.004
F-Measure	0.989	0.199	0.997	0.276	0.998	0	0.998	0.005
MCC	0.312	0.312	0.29	0.29	0	0	0.003	0.003
ROC Area	0.976	0.976	0.991	0.991	0.67	0.67	0.663	0.663
PRC Area	1	0.119	1	0.176	0.998	0.005	0.998	0.006

Πίνακας 7: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 5% συμβάντα απότομων στρεφουσών κινήσεων.

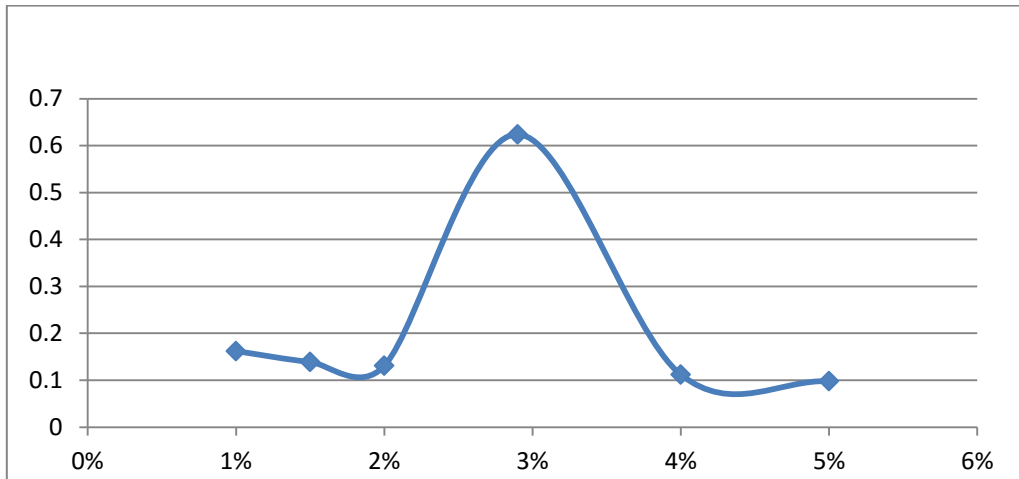
	J48 consolidated		Bayesian Network		MLP		Logistic Regression	
	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν
TP RATE	0.972	0.974	0.988	0.687	1	0	1	0
FP RATE	0.026	0.028	0.313	0.012	1	0	1	0
Precision	1	0.098	0.999	0.149	0.997	0	0.997	0
Recall	0.972	0.974	0.988	0.687	1	0	1	0
F-Measure	0.986	0.177	0.993	0.245	0.998	0	0.998	0
MCC	0.304	0.304	0.316	0.316	0	0	0	0
ROC Area	0.986	0.986	0.99	0.99	0.671	0.671	0.671	0.671
PRC Area	1	0.102	1	0.144	0.999	0.005	0.999	0.005

Τα αποτελέσματα των κατηγοριοποιητών σε όλα τα ποσοστά συμβάντων απότομης στροφής καταδεικνύουν:

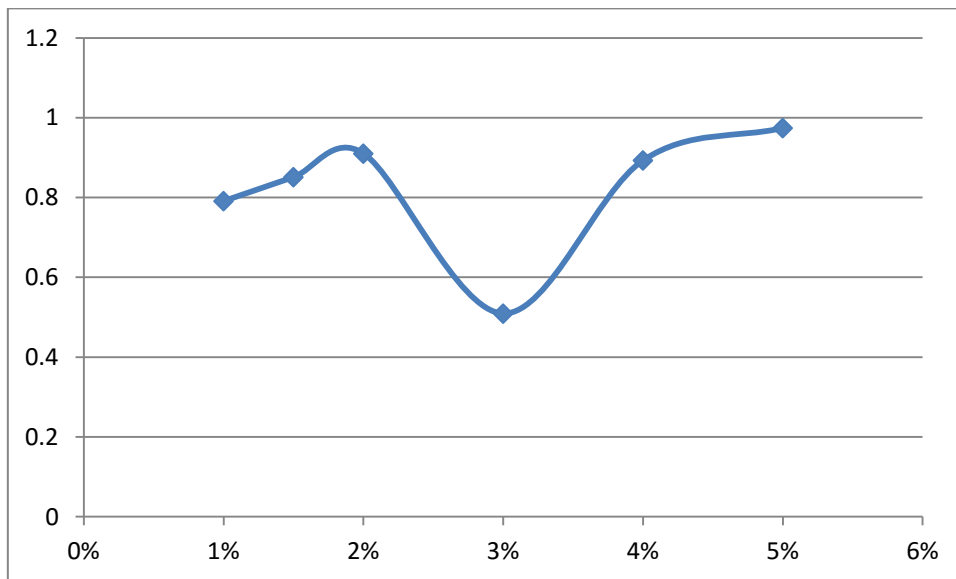
- την αδυναμία της πολυεπίπεδης ταξινόμησης και της λογιστικής παλινδρόμησης να προβλέψει το φαινόμενο
- την αξιόλογη απόδοση στην ευκρίνεια (Precision) του κατηγοριοποιητή bayes_net
- την σαφή υπεροχή του κατηγοριοποιητή J48 consolidated έναντι του bayes_net σε δείκτες όπως ο δείκτης σωστών προβλέψεων (TP-Rate) και η ανάκληση (Recall). Ως επακόλουθο αυτού λαμβάνεται ως καταλληλότερος κατηγοριοποιητής ο J48 consolidated.

Λαμβάνοντας υπ' όψιν τον κατηγοριοποιητή που μπορεί να προβλέψει καλύτερα τα συμβάντα δημιουργήθηκαν διαγράμματα με την εξέλιξη της απόδοσης των μονάδων μέτρησης αξιοπιστίας ανάλογα με το ποσοστό των συμβάντων.

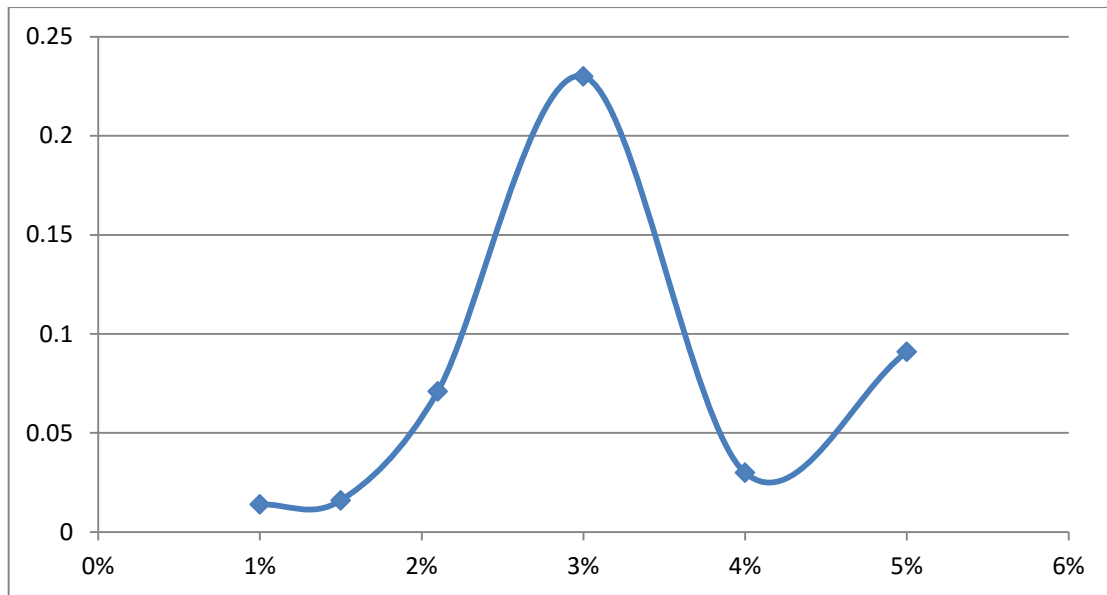
Στα διαγράμματα 1 και 3 φαίνεται η παραβολική πορεία της ακρίβειας (Precision) και του δείκτη ακρίβειας F-Measure. Στο ποσοστό του 3% λαμβάνεται η μέγιστη τιμή στις στους σημαντικότερους δείκτες. Στα Διαγράμματα 2-4 λαμβάνεται η αντίστροφη εικόνα για τους δείκτες σωστής πρόβλεψης (TP-Rate) και ανάκλησης (Recall). Τονίζεται ότι τα διαγράμματα εμπεριέχουν τις αντίστοιχες τιμές από τον καταλληλότερο κατηγοριοποιητή, τον J48-Consolidated.



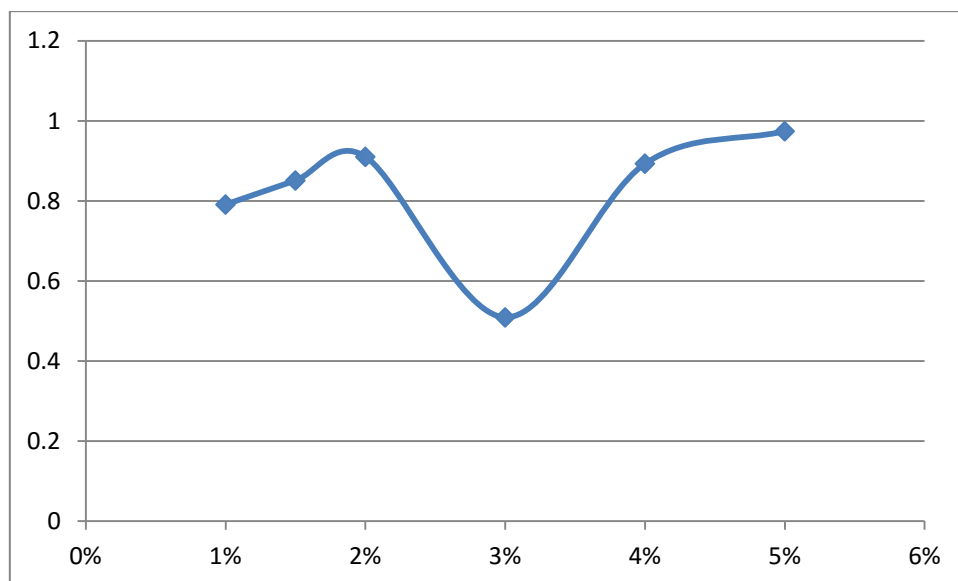
Διάγραμμα 1: Διάγραμμα μεταβολής ευκρίνειας (Precision) ανάλογα με το ποσοστό των συμβάντων απότομων στρεφουσών κινήσεων



Διάγραμμα 2: Διάγραμμα μεταβολής βαθμού σωστών προβλέψεων (TP-Rate) ανάλογα με το ποσοστό των συμβάντων απότομων στρεφουσών κινήσεων



Διάγραμμα 3: Διάγραμμα μεταβολής βαθμού εσφαλμένων προβλέψεων (FP-RATE) ανάλογα με το ποσοστό των συμβάντων απότομων στρεφουσών κινήσεων



Διάγραμμα 4: Διάγραμμα μεταβολής της ανάκλησης (Recall) ανάλογα με το ποσοστό των συμβάντων απότομων στρεφουσών κινήσεων

Συμβάντα απότομης επιτάχυνσης και επιβράδυνσης

Στους Πίνακες 8 έως 13 φαίνονται ,για τα ποσοστά 1%, 1,5%, 2%, 3%, 4% και 5% συμβάντων απότομης επιτάχυνσης και επιβράδυνσης, τα αποτελέσματα των τεσσάρων διαφορετικών κατηγοριοποιητών (το Δέντρο Απόφασης, το Μπεϋζιανό δίκτυο, τα Πολυεπίπεδα Νευρωνικά Δίκτυα , και η Λογιστική Παλινδρόμηση).

Πίνακας 8: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 1% συμβάντα απότομων κινήσεων επιτάχυνσης και επιβράδυνσης.

	J48 consolidated		Bayesian Network		MLP		Logistic Regression	
	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν
TP RATE	0.986	0.22	0.975	0.212	0.937	0.357	0.983	0.058
FP RATE	0.78	0.014	0.788	0.025	0.643	0.063	0.942	0.017
Precision	0.995	0.096	0.994	0.056	0.995	0.038	0.993	0.023
Recall	0.986	0.22	0.975	0.212	0.937	0.357	0.983	0.058
F-Measure	0.99	0.133	0.985	0.089	0.965	0.069	0.988	0.033
MCC	0.136	0.136	0.097	0.097	0.099	0.099	0.026	0.026
ROC Area	0.944	0.944	0.899	0.899	0.81	0.81	0.561	0.561
PRC Area	0.999	0.083	0.999	0.045	0.998	0.035	0.994	0.012

Πίνακας 9: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 1,5% συμβάντα κινήσεων επιτάχυνσης και επιβράδυνσης.

	J48 consolidated		Bayesian Network		MLP		Logistic Regression	
	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν
TP RATE	0.984	0.192	1	0	1	0	1	0
FP RATE	0.808	0.016	1	0	1	0	1	0
Precision	0.994	0.08	0.993	0	0.993	0	0.993	0
Recall	0.984	0.192	1	0	1	0	1	0
F-Measure	0.989	0.113	0.996	0	0.996	0	0.996	0
MCC	0.114	0.114	0	0	0	0	0	0
ROC Area	0.776	0.776	0.874	0.874	0.524	0.524	0.561	0.561
PRC Area	0.997	0.038	0.999	0.032	0.991	0.015	0.993	0.011

Πίνακας 10: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 2% συμβάντα απότομων κινήσεων επιτάχυνσης και επιβράδυνσης.

	J48 consolidated		Bayesian Network		MLP		Logistic Regression	
	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν
TP RATE	0.929	0.456	1	0	1	0	1	0
FP RATE	0.544	0.071	1	0	1	0	1	0
Precision	0.997	0.027	0.996	0	0.996	0	0.996	0
Recall	0.929	0.456	1	0	1	0	1	0
F-Measure	0.962	0.051	0.998	0	0.998	0	0.998	0
MCC	0.097	0.097	0	0	0	0	0	0
ROC Area	0.759	0.759	0.847	0.847	0.528	0.528	0.486	0.486
PRC Area	0.998	0.027	0.999	0.014	0.993	0.016	0.994	0.009

Πίνακας 11: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 3% συμβάντα απότομων κινήσεων επιτάχυνσης και επιβράδυνσης.

	J48 consolidated		Bayesian Network		MLP		Logistic Regression	
	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν
TP RATE	0.77	0.701	1	0	1	0	0.999	0
FP RATE	0.299	0.23	1	0	1	0	1	0.001
Precision	0.997	0.021	0.993	0	0.993	0	0.993	0
Recall	0.77	0.701	1	0	1	0	0.999	0
F-Measure	0.869	0.041	0.996	0	0.996	0	0.996	0
MCC	0.093	0.093	0	0	0	0	-0.002	-0.002
ROC Area	0.81	0.81	0.813	0.813	0.481	0.481	0.487	0.487
PRC Area	0.998	0.021	0.997	0.019	0.993	0.007	0.993	0.006

Πίνακας 12: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 4% συμβάντα απότομων κινήσεων επιτάχυνσης και επιβράδυνσης.

	J48 consolidated		Bayesian Network		MLP		Logistic Regression	
	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν
TP RATE	0.97	0.199	1	0	1	0	1	0
FP RATE	0.801	0.03	1	0	1	0	1	0
Precision	0.995	0.042	0.994	0	0.994	0	0.994	0
Recall	0.97	0.199	1	0	1	0	1	0
F-Measure	0.982	0.069	0.997	0	0.997	0	0.997	0
MCC	0.079	0.079	0	0	0	0	0	0
ROC Area	0.777	0.777	0.808	0.808	0.488	0.488	0.506	0.506
PRC Area	0.997	0.025	0.998	0.017	0.993	0.01	0.993	0.012

Πίνακας 13: Αποτελέσματα κατηγοριοποίησης στο δείγμα με 5% συμβάντα απότομων κινήσεων επιτάχυνσης και επιβράδυνσης.

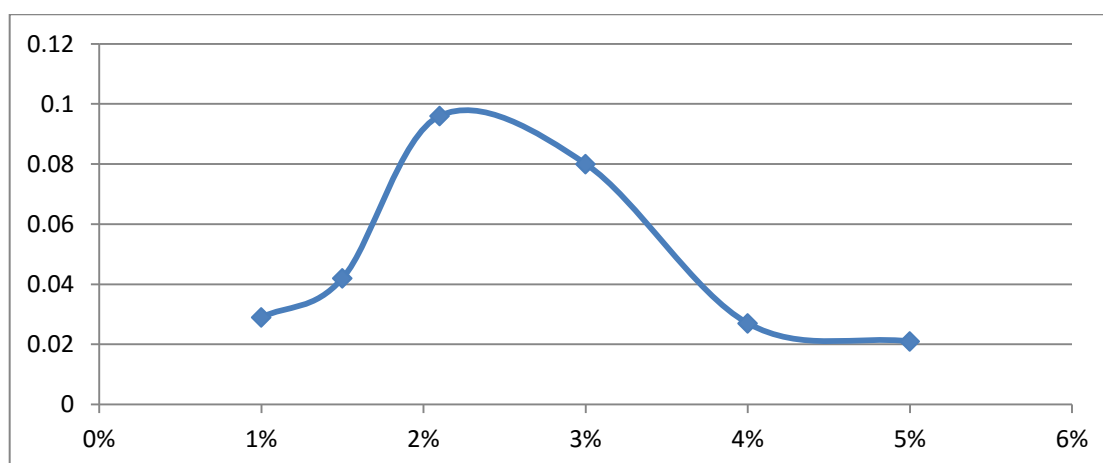
	J48 consolidated		Bayesian Network		MLP		Logistic Regression	
	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν	Κανονική οδήγηση	Συμβάν
TP RATE	0.909	0.391	1	0	1	0	1	0
FP RATE	0.609	0.091	1	0	1	0	1	0
Precision	0.995	0.029	0.993	0	0.993	0	0.993	0
Recall	0.909	0.391	1	0	1	0	1	0
F-Measure	0.951	0.054	0.997	0	0.997	0	0.997	0
MCC	0.086	0.086	0	0	0	0	0	0
ROC Area	0.703	0.703	0.805	0.805	0.526	0.526	0.523	0.523
PRC Area	0.996	0.017	0.997	0.017	0.994	0.009	0.994	0.008

Τα αποτελέσματα των κατηγοριοποιητών σε όλα τα ποσοστά συμβάντων απότομης επιτάχυνσης και επιβράδυνσης καταδεικνύουν:

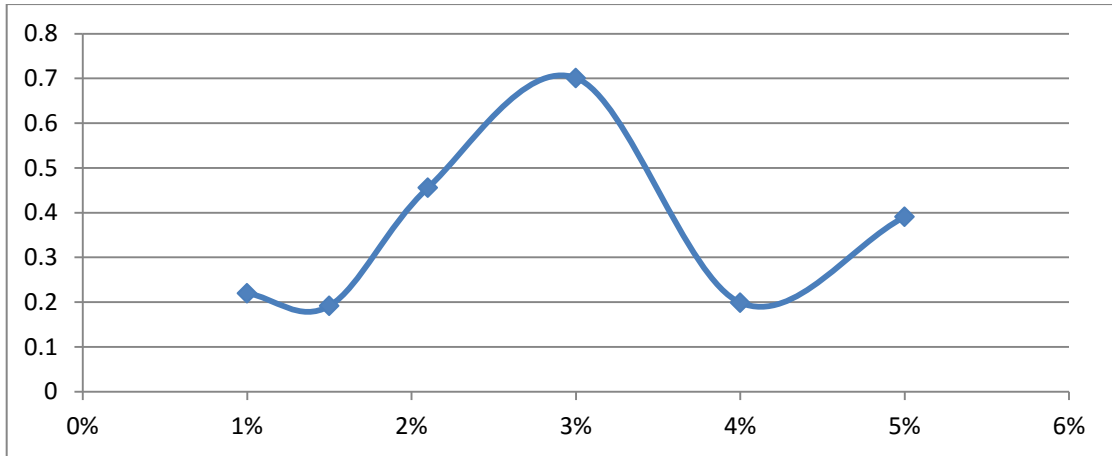
- την αδυναμία της λογιστικής παλινδρόμησης στην πρόβλεψη του φαινομένου
- την αδυναμία της πολυεπίπεδης ταξινόμησης στην πρόβλεψη του φαινομένου, πέρα από μια αξιοπρόσεκτη επίδοση στο 1% των συμβάντων
- την αδυναμία του κατηγοριοποιητή bayes_net στην πρόβλεψη του φαινομένου πέραν του ποσοστού του 1% όπου επιτυγχάνεται εφάμιλλη απόδοση με τον j48 consolidated
- την σαφή υπεροχή του j48 consolidated ελλείψει της καλής απόδοσης των υπολοίπων κατηγοριοποιητών αλλά με εμφανώς μειωμένη απόδοση σε σχέση με την προηγούμενη περίπτωση λόγω και της δυσκολίας του φαινομένου.

Λαμβάνοντας υπ' όψιν τον κατηγοριοποιητή που μπορεί να προβλέψει καλύτερα τα συμβάντα δημιουργήθηκαν διαγράμματα με την εξέλιξη της απόδοσης των μονάδων μέτρησης αξιοπιστίας ανάλογα με το ποσοστό των συμβάντων.

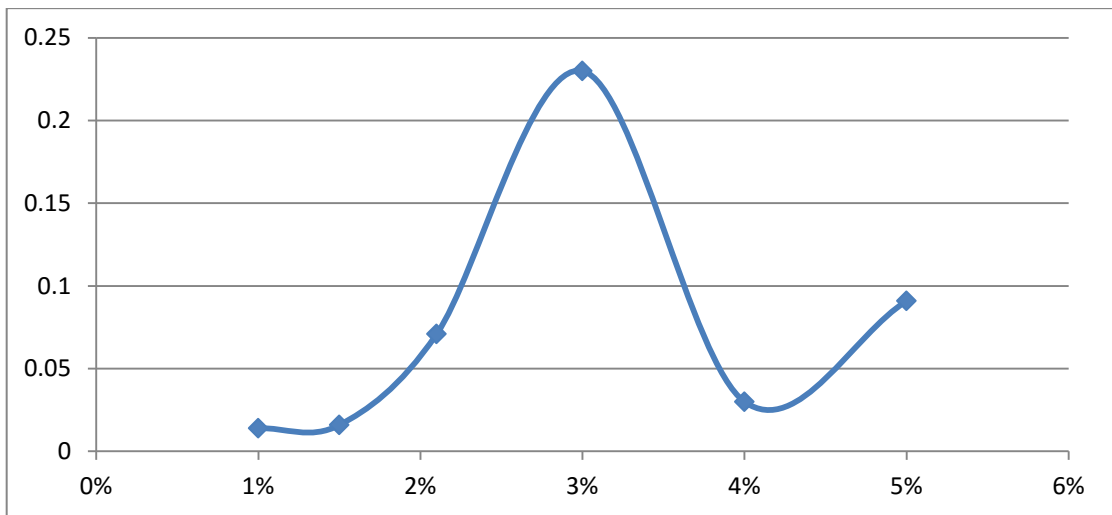
Στα διαγράμματα 1 έως 4 φαίνεται η παραβολική πορεία της ακρίβειας (Precision), του δείκτη ακρίβειας F-Measure, του δείκτη σωστής πρόβλεψης (TP-Rate) και ανάκλησης (Recall). Στο ποσοστό του 3% λαμβάνεται η μέγιστη τιμή στις σε όλους τους δείκτες. Τονίζεται ότι τα διαγράμματα εμπεριέχουν τις αντίστοιχες τιμές από τον καταλληλότερο κατηγοριοποιητή, τον J48-Consolidated.



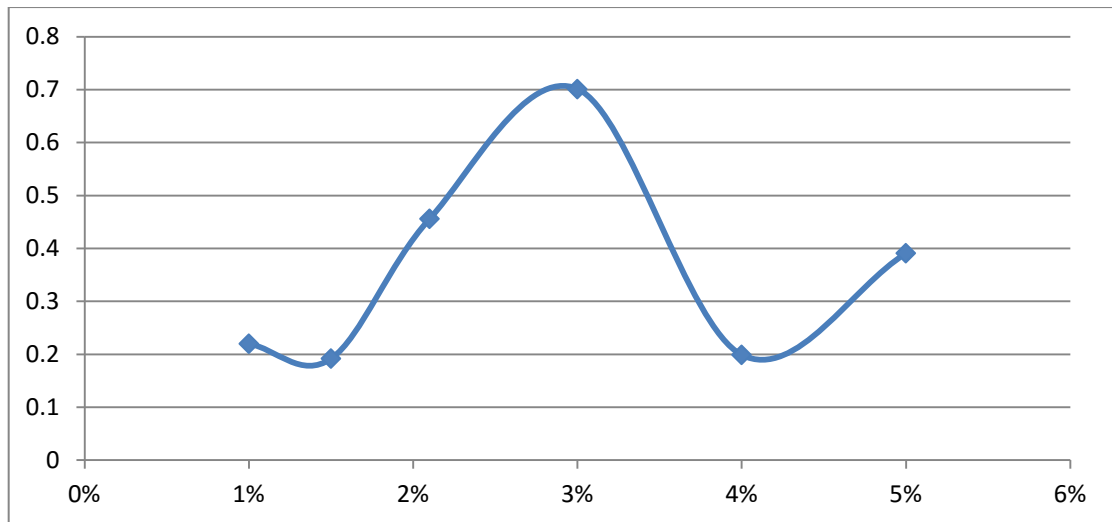
Διάγραμμα 5: Διάγραμμα μεταβολής ευκρίνειας (Precision) ανάλογα με το ποσοστό των συμβάντων απότομης επιτάχυνσης και επιβράδυνσης



Διάγραμμα 6: Διάγραμμα μεταβολής του βαθμού σωστών προβλέψεων (TP-Rate) ανάλογα με το ποσοστό των συμβάντων απότομης επιτάχυνσης και επιβράδυνσης



Διάγραμμα 7: Διάγραμμα μεταβολής του βαθμού εσφαλμένων προβλέψεων (FP-Rate) ανάλογα με το ποσοστό των συμβάντων απότομης επιτάχυνσης και επιβράδυνσης



Διάγραμμα 8: Διάγραμμα μεταβολής ανάκλησης (Recall) ανάλογα με το ποσοστό των συμβάντων απότομης επιτάχυνσης και επιβράδυνσης

Κεφάλαιο 5

5. Συμπεράσματα

5.1 Σύνοψη αποτελεσμάτων

Αντικείμενο της παρούσας Διπλωματικής Εργασίας είναι η διερεύνηση της κυκλοφοριακής συμπεριφοράς του οδηγού μέσω της χρήσης αισθητήρων του έξυπνου κινητού. ότι η δυνατότητα αξιοποίησης των στοιχείων αυτών είναι σχετικά πρόσφατη, πρόκειται για ένα θέμα το οποίο δεν έχει μελετηθεί ιδιαίτερος ούτε διεθνώς ούτε στην Ελλάδα, ωστόσο, όπως προκύπτει και από τη βιβλιογραφική ανασκόπηση, τα τελευταία χρόνια χάρη στην εξέλιξη της τεχνολογίας και συγκεκριμένα εφαρμογών των κινητών τηλεφώνων, παρατηρείται μία αυξανόμενη ενασχόληση του επιστημονικού κόσμου γύρω από το συγκεκριμένο ζήτημα.

Η συλλογή των απαιτούμενων στοιχείων ώστε να γίνει διερεύνηση της κυκλοφοριακής συμπεριφοράς του οδηγού έγινε μέσω κατάλληλα σχεδιασμένης εφαρμογής για την αξιοποίηση των αισθητήρων του έξυπνου κινητού.

Μετά από κατάλληλη επεξεργασία και σειρά δοκιμών αναπτύχθηκαν οι κατηγοριοποιητές j48 consolidated, bayes_net, mlr και logistic που παρείχαν τις καλύτερες δυνατές προβλέψεις για:

1. Συμβάντα απότομης στροφής
2. Συμβάντα απότομης επιτάχυνσης και επιβράδυνσης

Αναπτύχθηκαν τα μοντέλα όλων των κατηγοριοποιητών αλλά στον παρακάτω πίνακα συνοψίζονται τα αποτελέσματα του καλύτερου κατηγοριοποιητή, του δέντρου απόφασης j48 consolidated.

Πίνακας 14: Συγκεντρωτικός πίνακας για την πρόβλεψη απότομου ελιγμού

Απότομη στροφή για την κλάση 1	0.17%	1%	1.50%	2%	3%	4%	5%
TP Rate		0.791	0.851	0.91	0.509	0.893	0.974
FP Rate		0.013	0.017	0.018	0.014	0.022	0.028
Precision		0.162	0.139	0.131	0.624	0.112	0.098
Recall		0.791	0.851	0.91	0.509	0.893	0.974

Πίνακας 15: Συγκεντρωτικός πίνακας για την πρόβλεψη απότομης επιτάχυνσης και επιβράδυνσης

Απότομη επιτάχυνση/επιβράδυνση για την κλάση 1	0.28%	1%	1.50%	2%	3%	4%	5%
TP Rate		0.22	0.192	0.456	0.701	0.199	0.391
FP Rate		0.014	0.016	0.071	0.23	0.03	0.091
Precision		0.096	0.08	0.027	0.021	0.042	0.029
Recall		0.22	0.192	0.456	0.701	0.199	0.391

5.2 Συνολικά συμπεράσματα

Από τα διάφορα στάδια εκπόνησης της παρούσας Διπλωματικής Εργασίας προέκυψε μια σειρά συμπερασμάτων, όπως αυτά συνοψίζονται παρακάτω:

- 1) Τα δεδομένα τα οποία καταγράφονται από τους αισθητήρες των έξυπνων κινητών μπορούν να αποθηκευθούν ηλεκτρονικά σε υπολογιστή και περιέχουν ιδιαίτερες σημαντικές πληροφορίες οι οποίες μετά από κατάλληλη επεξεργασία και ανάπτυξη μαθηματικών μοντέλων, μπορούν να φανούν ιδιαίτερος χρήσιμα στην εξαγωγή συμπερασμάτων για την κυκλοφοριακή συμπεριφορά του οδηγού.
- 2) Βασικός παράγοντας για την ανίχνευση συμβάντος απότομης στροφής (harsh cornering) είναι η καταγραφή της γωνιακής ταχύτητας μέσω του γυροσκοπίου.
- 3) Βασικός παράγοντας για την ανίχνευση συμβάντος απότομης επιτάχυνσης και επιβράδυνσης (harsh acceleration) είναι η καταγραφή της επιτάχυνσης μέσω του επιταχυνσιόμετρου.
- 4) Καταλληλότερος κατηγοριοποιητής εμφανίζεται ως ο j48 consolidated όπου με εξαίρεση μια κατηγορία παρουσιάζει την μεγαλύτερη απόδοση.
- 5) Αδυναμία στο να προβλεψουν είτε συμβάντα απότομης στροφής, είτε συμβάντα απότομης επιτάχυνσης και επιβράδυνσης έχουν η πολυεπίπεδη ταξινόμηση και ο κατηγοριοποιητής logistic.
- 6) Η ευκρίνεια (Precision) των μοντέλων στην πρόβλεψη συμβάντων απότομης επιτάχυνσης και επιβράδυνσης είναι αρκετά χαμηλή λόγω και της δυσκολίας πρόβλεψης του φαινομένου.
- 7) Στο 3% των συμβάντων παρατηρείται η μεγαλύτερη τιμή στους δείκτες μέτρησης Precision και F-Measure.

5.3 Προτάσεις για βελτίωση της κυκλοφοριακής συμπεριφοράς

Επακόλουθο των αποτελεσμάτων και των συνολικών συμπερασμάτων που εξήχθησαν κατά την εκπόνηση της παρούσας Διπλωματικής Εργασίας είναι η παράθεση μερικών προτάσεων, οι οποίες θα μπορούσαν να συνεισφέρουν στο να βελτιωθεί η κυκλοφοριακή συμπεριφορά των οδηγών.

- Η παρούσα προσπάθεια θα μπορούσε να διευρυνθεί, χρησιμοποιώντας ολοένα περισσότερους οδηγούς, οι οποίοι θα ενημερώνονται για την συνολική τους οδηγική συμπεριφορά με σκοπό την αυτογνωσία και την προσπάθεια να βελτιωθούν.
- Ο ηλεκτρονικός έλεγχος της ταχύτητας των οχημάτων θα πρέπει να εφαρμοστεί στην πράξη (όχι μόνο στην σήμανση), ιδίως στους δρόμους ταχείας κυκλοφορίας, έτσι ώστε να γνωρίζουν οι ειδικοί φορείς και η τροχαία ποιος υπερβαίνει τα νομοθετημένα όριο και επιβάλλονται οι απαραίτητες κυρώσεις.
- Η εδραίωση και η ανάπτυξη ειδικών προγραμμάτων εκπαίδευσης για μικρούς και μεγάλους σε συνεργασία με την πολιτεία και τους εκπαιδευτικούς φορείς ώστε να γίνει προσπάθεια βελτίωσης της νοοτροπίας των οδηγών.

5.4 Προτάσεις για περαιτέρω έρευνα

Όπως αναφέρθηκε παραπάνω το συγκεκριμένο θέμα είναι αναμφίβολα δύσκολο ως προς την μελέτη του, κυρίως λόγω του μικρού αριθμού συμβάντων κατά την διάρκεια της οδήγησης. Παρ' όλα αυτά η ραγδαία εισβολή των έξυπνων κινητών στην καθημερινότητά μας προσδίδει τεράστιο ενδιαφέρον στο συγκεκριμένο αντικείμενο. Ως εκ τούτου, κρίνεται σημαντικό να επισημανθούν προτάσεις για περαιτέρω έρευνα στο συγκεκριμένο θέμα:

Η διερεύνηση αυτού του θέματος με ένα ακόμα πιο ευρύ δείγμα οδηγών, τέτοιο ώστε να μπορούν να εξαχθούν με μεγαλύτερη ασφάλεια συμπεράσματα μέσω κατηγοριοποίησης των συμμετεχόντων ανάλογα με διαφορετικά στοιχεία των οδηγών όπως η εμπειρία στην οδήγηση, το φύλο, η ηλικία κτλ. Αυτό θα απαιτούσε παράλληλα και την επίτευξη μεγαλύτερης βάσης δεδομένων ή την πρόβλεψη σε πραγματικό χρόνο απ' ευθείας μέσω κινητού τηλεφώνου. Αυτό θα μπορούσε να είναι αντικείμενο περαιτέρω έρευνας.

Στην διπλωματική εργασία λήφθηκε ως δεδομένη η αξιοπιστία του OBD όμως θα μπορούσε να ελεγχθεί περαιτέρω η αξιοπιστία του. Ελέγχοντας και βελτιώνοντας την αξιοπιστία του OBD είναι φανερό πως ενδεχόμενες διαφορές που υπάρχουν μεταξύ της συσκευής OBD και του έξυπνου κινητού θα μπορούσαν να μειωθούν, βελτιώνοντας παράλληλα την απόδοση των κατηγοριοποιητών.

Τέλος, θα ήταν ωφέλιμη η σύνδεση της ψυχολογικής κατάστασης των οδηγών με τον τρόπο οδήγησης τους. Η ανίχνευση της ψυχολογικής κατάστασης θα μπορούσε να γίνει από ειδικά σχεδιασμένες εφαρμογές που κυκλοφορούν σε διαδικτυακά καταστήματα για έξυπνα κινητά. Η έρευνα αυτή θα μπορούσε να δημιουργεί εξατομικευμένο πρότυπο σε περίπτωση ηρεμίας, να ανιχνεύει διαφορές σε σχέση με την οδήγηση σε στιγμές που είναι ο χρήστης συναισθηματικά φορτισμένος, νευριασμένος κτλ. Και αργότερα να ειδοποιείται ο χρήστης για την μεταβολή της οδηγικής του συμπεριφοράς. Εκτός της ψυχολογικής κατάστασης χρήσιμο θα ήταν να συνδεθεί η οδηγική συμπεριφορά με την κατανάλωση αλκοόλ ή την κόπωση. Ήδη αυτοκινητοβιομηχανίες και εταιρίες βαρέων οχημάτων προσπαθούν να αναπτύξουν συστήματα ασφάλειας από κόπωση ([Fatigue Risk Management System \(FRMS\)](#)) [22] τα συστήματα αυτά λειτουργούν παρακολουθώντας μέσω κάμερας τον χρόνο βλεφαρίσματος. Ενώ για την αναγνώριση της κατανάλωσης αλκοόλ κυκλοφορούν ήδη εφαρμογές αλλά δεν έχουν συνδεθεί ποσοτικοποιημένα με την οδηγική συμπεριφορά.

Βιβλιογραφία

- [1] L. M. Bergasa, D. Almería, J. Almazán και R. Arroyo, «DriveSafe: an App for Alerting Inattentive Drivers and Scoring Driving Behaviors,» 2014.
- [2] Y.-C. Chiu, Y.-L. Ma και X. Hu, «Identifying Driving Risk Factors to Support Usage-Based Insurance using Smartphone Personalized Driving Data,» 2014.
- [3] J. Engelbrecht, M. J. Booysen και G.-J. v. Rooyen, «Recognition of driving manoeuvres using,» 2014.
- [4] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury και A. T. Campbell, «A Survey of Mobile Phone Sensing,» 2010.
- [5] H. Peter, S. M. Isaac, W. Johan, B. Farid, W. Richard, O. Jens και O. Martin, «Insurance telematics: opportunities and challenges with the smartphone solution,» 2014.
- [6] H. Peter, O. Jens, O. Martin, S. Isaac και N. Elin, «Smartphone Based Measurement Systems for Road Vehicle Traffic Monitoring and Usage Based Insurance,» 2014.
- [7] H. Siniša, P. Dragan, F. Ivan και M. Marijan, «Telematics System in Usage Based Motor Insurance,» 2014.
- [8] D. A. Johnson και M. M. Trivedi, «Driving Style Recognition Using a Smartphone as a Sensor Platform,» σε *14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2011.
- [9] Trivedi, K. S. Ravi και M. M., «Drive Analysis using Vehicle Dynamics and Vision-based Lane Semantics,» 2014.
- [10] P. Bratislav και S. Dragan, «Enhancing driver situational awareness through crowd intelligence,» 2015.
- [11] A. Shahid, B. Alireza και H. Bahram, «A Sensor Fusion Method for Smart phone Orientation Estimation,» 2012.
- [12] A. Javier, M. B. Luis, Y. Javier, B. Rafael και A. Roberto, «Full auto-calibration of a smartphone on board a vehicle using IMU and GPS embedded sensors,» 2013.
- [13] T. Marco, L. Edward και B. Natalie, «Correcting Smartphone Orientation for Accelerometer-Based Analysis,» 2013.
- [14] «<http://alias-i.com/>,» ALIAS-I, [Ηλεκτρονικό]. Available: <http://alias-i.com/lingpipe/demos/tutorial/logistic-regression/read-me.html>. [Πρόσβαση 2016].
- [15] <http://www.d.umn.edu>, 2016. [Ηλεκτρονικό]. Available: <http://www.d.umn.edu/~padhy005/Chapter5.html>.

- [16] <http://weka.sourceforge.net>, [Ηλεκτρονικό]. Available: <http://weka.sourceforge.net/packageMetaData/J48Consolidated/index.html>. [Πρόσβαση 2016].
- [17] <http://scikit-learn.org>, 2015. [Ηλεκτρονικό]. Available: http://scikit-learn.org/dev/modules/neural_networks_supervised.html. [Πρόσβαση 2016].
- [18] <http://www.obdii.com>, [Ηλεκτρονικό]. Available: <http://www.obdii.com/background.html>. [Πρόσβαση 2016].
- [19] <http://www.munic.io/>, [Ηλεκτρονικό]. Available: <http://www.munic.io/>. [Πρόσβαση 2016].
- [20] <https://physicsandroid.wordpress.com/>, [Ηλεκτρονικό]. Available: <https://physicsandroid.wordpress.com/%CE%B1%CE%B9%CF%83%CE%B8%CE%B7%CF%84%CE%AE%CF%81%CE%B1%CF%82-%CE%B5%CF%80%CE%B9%CF%84%CE%AC%CF%87%CF%85%CE%BD%CF%83%CE%B7%CF%82/>. [Πρόσβαση 2016].
- [21] <http://wikipedia.qwika.com>, [Ηλεκτρονικό]. Available: <http://wikipedia.qwika.com/en2el/Gyroscope>.
- [22] «www.cat.com,» caterpillar inc, [Ηλεκτρονικό]. Available: http://www.cat.com/en_US/support/operations/frms.html. [Πρόσβαση June 2016].
- [23] «2014».
- [24] D. L. Nicholas, M. Emiliano, L. Hong, P. Daniel, C. Tanzeem και T. ., C. Andrew, «A Survey of Mobile Phone Sensing,» 2010.

Παράρτημα Ι

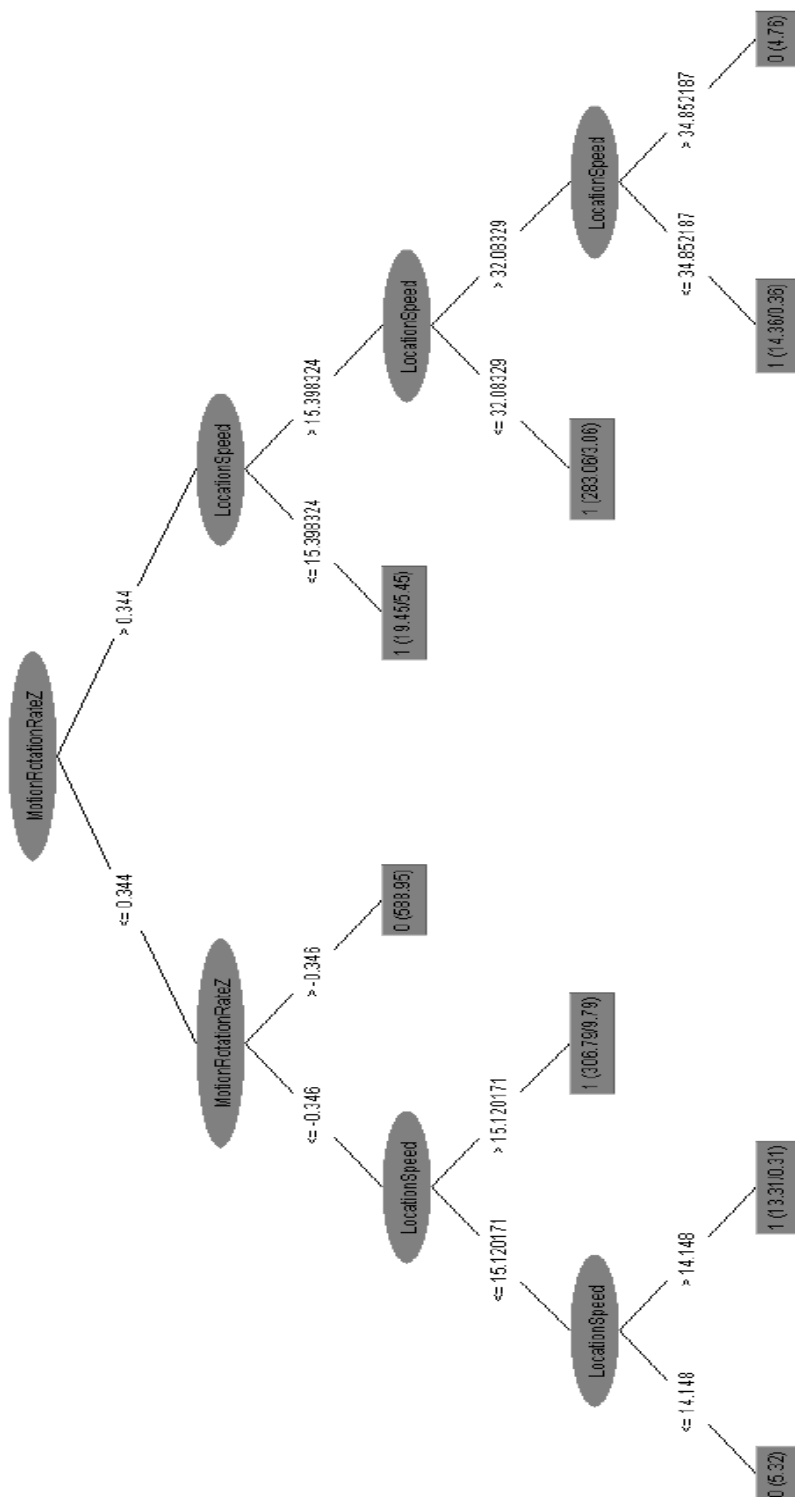
Πίνακας 16 : Στατιστικά στοιχεία με διαχωρισμό για το εάν είναι συμβάν απότομης στροφής

			Row Labels	
			0	1
test set	Μέσος Όρος	LocationSpeed	31.121	25.729
		Sdiff2	0.003	-1.663
		MotionRotationZ	-0.010	-0.042
		Distance	946.234	896.378
	Τυπική Απόκλιση	LocationSpeed	32.840	13.939
		Sdiff2	4.408	7.259
		MotionRotationZ	0.385	1.192
		Distance	20654.788	13900.842
	Διασπορά	LocationSpeed	1078.473	194.298
		Sdiff2	19.432	52.694
		MotionRotationZ	0.148	1.422
		Distance	426620256.360	193233405.102
training set	Μέσος Όρος	LocationSpeed	28.050	25.675
		Sdiff2	0.002	-0.832
		MotionRotationZ	-0.001	-0.005
		Distance	2.473	6.708
	Τυπική Απόκλιση	LocationSpeed	29.831	11.676
		Sdiff2	4.767	7.362
		MotionRotationZ	0.329	0.920
		Distance	14.714	23.214
	Διασπορά	LocationSpeed	889.888	136.324
		Sdiff2	22.727	54.202
		MotionRotationZ	0.108	0.847
		Distance	216.510	538.887

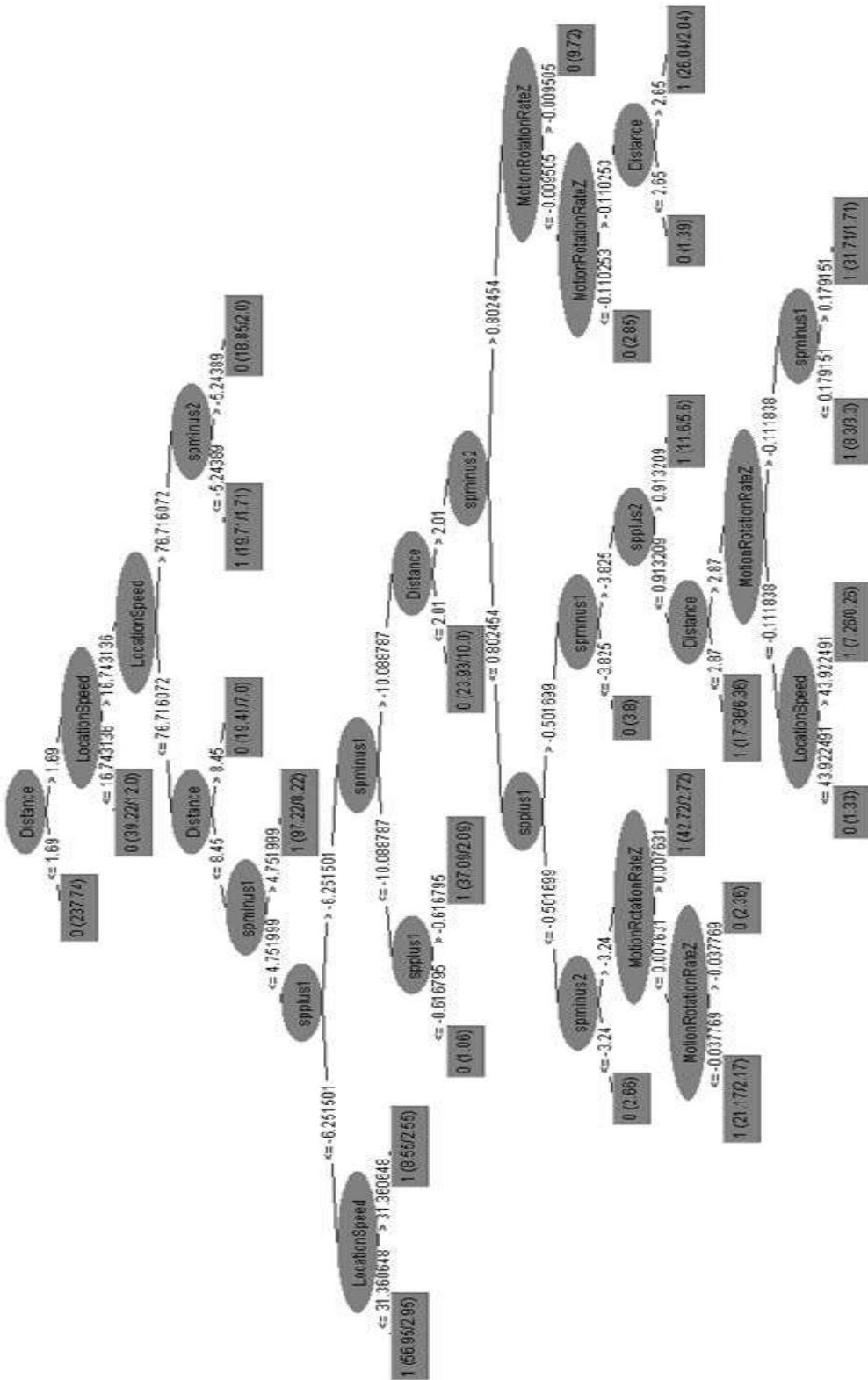
Πίνακας 17: Στατιστικά στοιχεία με διαχωρισμό για το εάν είναι συμβάν απότομης επιτάχυνσης/ επιβράδυνσης

			Row Labels	
			0	1
test set	Μέσος Όρος	LocationSpeed	31.089	39.161
		Sdiff2	-0.001	0.207
		MotionRotationZ	-0.010	0.003
		Distance	947.191	581.346
	Τυπική Απόκλιση	LocationSpeed	32.839	22.772
		Sdiff2	4.387	10.239
		MotionRotationZ	0.388	0.253
		Distance	20669.965	8121.955
	Διασπορά	LocationSpeed	1078.473	194.298
		Sdiff2	19.432	52.694
		MotionRotationZ	0.148	1.422
		Distance	426620256.360	193233405.102
training set	Μέσος Όρος	LocationSpeed	28.005	36.858
		Sdiff2	-0.004	0.803
		MotionRotationZ	-0.001	-0.005
		Distance	2.465	6.267
	Τυπική Απόκλιση	LocationSpeed	29.837	19.314
		Sdiff2	4.733	10.460
		MotionRotationZ	0.331	0.347
		Distance	14.713	19.216
	Διασπορά	LocationSpeed	890.270	373.042
		Sdiff2	22.399	109.410
		MotionRotationZ	0.110	0.120
		Distance	216.484	369.255

Παράρτημα II



Διάγραμμα 9 : Δέντρο απόφασης για 3% των συμβάντων απότομης στροφής



Διάγραμμα 10 : Δέντρο απόφασης για 3% των συμβάντων απότομης επιτάχυνσης/ επιβράδυνσης