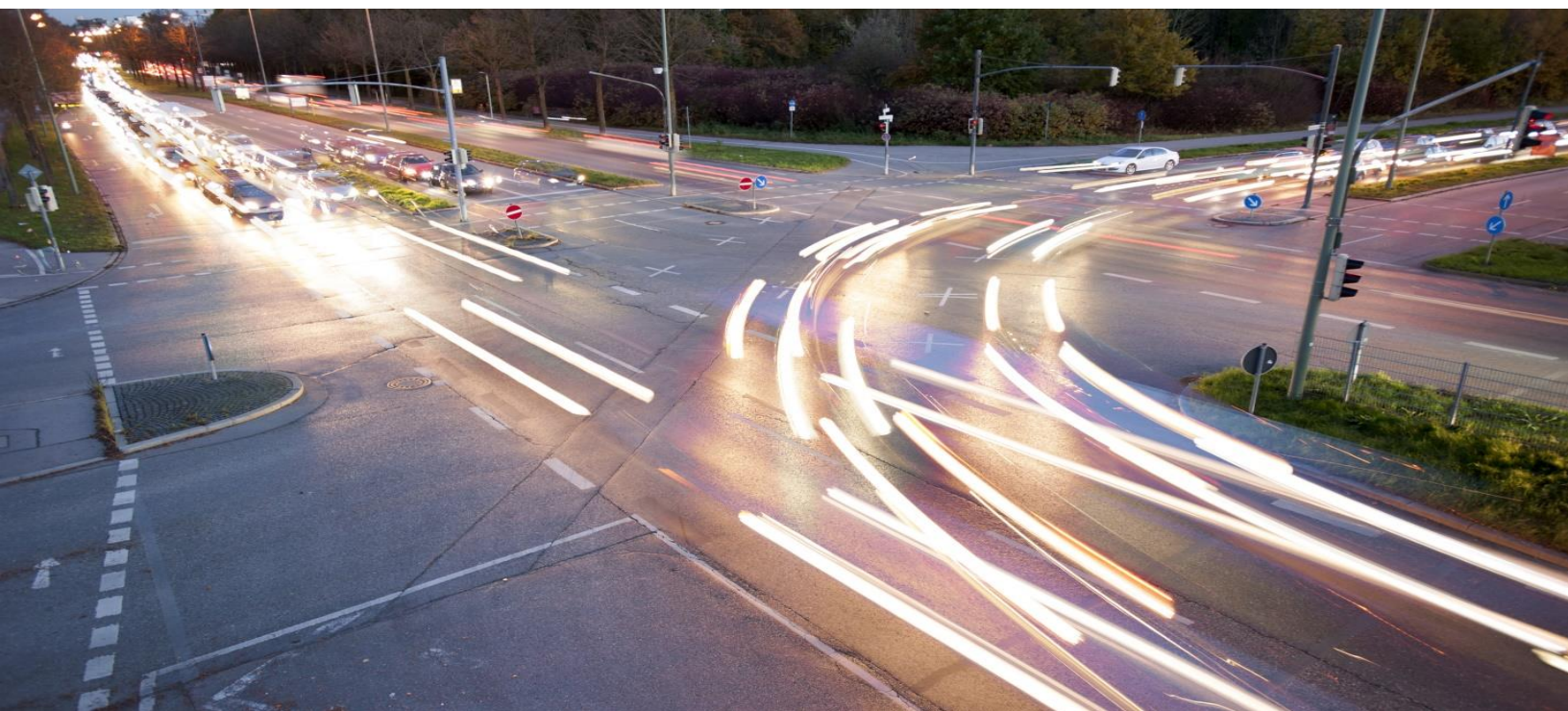




ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ

ΚΡΙΣΙΜΟΙ ΠΑΡΑΓΟΝΤΕΣ ΕΝΤΟΠΙΣΜΟΥ ΟΔΙΚΩΝ ΣΥΜΒΑΝΤΩΝ ΣΕ ΑΣΤΙΚΕΣ ΠΕΡΙΟΧΕΣ

Διπλωματική Εργασία



Μπαρδή Φωτεινή

Επιβλέπων: Γιώργος Γιαννής, Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2020

ΕΥΧΑΡΙΣΤΙΕΣ

Με την παρούσα Διπλωματική Εργασία ολοκληρώνεται ο κύκλος των προπτυχιακών σπουδών μου στη Σχολή Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου.

Θα ήθελα πρωτίστως να ευχαριστήσω θερμά τον κύριο Γιώργο Γιαννή, Καθηγητή του Εθνικού Μετσόβιου Πολυτεχνείου, για την ανάθεση και επίβλεψη της παρούσας Διπλωματικής Εργασίας και για την καθοδήγηση του καθ' όλη τη διάρκεια της εκπόνησης της.

Επίσης, θα ήθελα να ευχαριστήσω εξίσου θερμά τον Δρ. Χρήστο Κατρακάζα για την πολύτιμη βοήθεια και την καθοριστική συμβολή του σε όλα τα στάδια ολοκλήρωσης της εργασίας καθώς και για το εξαιρετικό κλίμα συνεργασίας που διαμόρφωσε.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου για την υποστήριξη που μου προσέφεραν καθ' όλη τη διάρκεια των σπουδών μου.

Αθήνα, Οκτώβριος 2020

Φωτεινή Μπαρδή

Κρίσιμοι παράγοντες εντοπισμού οδικών συμβάντων σε αστικές περιοχές

Μπαρδής Φωτεινή

Επιβλέπων: Γιώργος Γιαννής, Καθηγητής Ε.Μ.Π.

Σύνοψη

Στόχος της παρούσας διπλωματικής εργασίας είναι ο προσδιορισμός των κρίσιμων παραγόντων εντοπισμού οδικών συμβάντων σε αστικές περιοχές. Η συλλογή των στοιχείων έγινε από μία μεγάλη βάση δεδομένων που δημιουργήθηκε μέσω πειράματος σε προσομοιωτή οδήγησης. Για την ανάλυση των δεδομένων αναπτύχθηκαν δύο στατιστικά μοντέλα, συγκεκριμένα ένα μοντέλο διωνυμικής λογιστικής παλινδρόμησης και ένα μοντέλο τυχαίων δασών, με εξαρτημένη μεταβλητή την ύπαρξη ή όχι συμβάντος. Επίσης, πραγματοποιήθηκε ανάλυση παραγόντων σε καταστάσεις που αφορούσαν σε ένα λεπτό πριν το συμβάν, στη διάρκεια του συμβάντος καθώς και στο συνδυασμό αυτών των καταστάσεων με σκοπό την διερεύνηση ύπαρξης κοινών παραγόντων στην ομάδα των εξεταζόμενων μεταβλητών. Σύμφωνα με τα αποτελέσματα, το μοντέλο τυχαίων δασών παρουσίασε τα πιο αξιόπιστα αποτελέσματα στην ικανότητα πρόβλεψης ύπαρξης συμβάντων με χαμηλότερη πιθανότητα λάθος προβλέψεων, συγκριτικά με εκείνο της διωνυμικής λογιστικής παλινδρόμησης. Από την παραγοντική ανάλυση, προέκυψε ότι η κατάσταση ενός λεπτού πριν το συμβάν μπορεί να περιγραφεί μέσω της ταχύτητας, της απόκλισης του οχήματος από το μέσο της οδού και της απόστασης από τη δεξιά οριογραμμή. Ομοίως, η κατάσταση που αφορά σε στοιχεία κατά τη διάρκεια του συμβάντος μπορεί να περιγραφεί καλύτερα μέσω της ταχύτητας και της διαμήκουσ και πλευρικής επιτάχυνσης. Τέλος, γίνονται προτάσεις για αξιοποίηση των αποτελεσμάτων καθώς και για περαιτέρω έρευνα του αντικειμένου.

Λέξεις κλειδιά: εντοπισμός συμβάντων, αστικές περιοχές, οδηγική συμπεριφορά, διωνυμική λογιστική παλινδρόμηση, μοντέλο τυχαίων δασών, ανάλυση παραγόντων

Critical factors for event identification on urban areas

Bardi Fotini

Supervisor: George Yannis, Professor N.T.U.A.

Abstract

The aim of this Diploma Thesis is to determine critical factors for the identification of traffic events in urban areas. Data were obtained from a large dataset created through a driving simulator experiment. For the data analysis, two statistical models were developed, namely a binomial logistic regression and a random forest one, both of which considering the occurrence of an event as the dependent variable. Furthermore, factor analysis was performed with regards to data concerning one minute before the event, the duration of the event as well as the combination of the aforementioned cases, in order to investigate the existence of common factors in the group of independent variables. According to the results, the random forest model proved to produce more reliable results predicting events with a lower false alarm rate, when compared to binomial logistic regression. Moreover, factor analysis demonstrated that data representing one minute before the event can be described by speed, the deviation of the vehicle from the middle of the road and the distance from the right boundary line. Similarly, data during the event can be better described through speed and longitudinal and lateral acceleration. Finally, suggestions are made for the results use as well as for further research in this topic.

Keywords: event identification, urban areas, driving behavior, binomial logistic regression, random forest model, factor analysis

ΠΕΡΙΛΗΨΗ

Στόχος της παρούσας διπλωματικής εργασίας είναι ο **προσδιορισμός κρίσιμων παραγόντων εντοπισμού οδικών συμβάντων σε αστικές περιοχές**. Για το λόγο αυτό αναπτύχθηκαν μοντέλα στατιστικής ανάλυσης και διερευνήθηκε η ικανότητά τους να προβλέπουν την ύπαρξη ενός συμβάντος.

Τα δεδομένα που αναλύθηκαν προέκυψαν από πείραμα σε προσομοιωτή οδήγησης που είχε εκπονηθεί σε παλαιότερη έρευνα. Από αυτά τα δεδομένα διαμορφώθηκαν οι τελικοί πίνακες που περιείχαν στοιχεία που αφορούσαν στην κατάσταση ενός λεπτού πριν το συμβάν αλλά και στη διάρκεια του συμβάντος. Ως **εξαρτημένη** μεταβλητή καθορίστηκε η ύπαρξη ή όχι συμβάντος (Event) και ως **ανεξάρτητες** η ταχύτητα (Speed), η πλευρική (AccLat) και η διαμήκης (AccLon) επιτάχυνση, ο χρόνος μέχρι τη σύγκρουση με το προπορευόμενο όχημα (THead), η απόσταση από τη δεξιά οριογραμμή (DRight), η συνολικά διανυόμενη απόσταση (rdist), η απόκλιση του οχήματος από το μέσο της οδού (rspur), η θέση του τιμονιού (Wheel) και η οδηγική εμπειρία (Driving experience). Οι ανεξάρτητες μεταβλητές προέκυψαν μετά από έλεγχο ότι η συσχέτιση μεταξύ τους δεν είναι υψηλή.

Για τη στατιστική ανάλυση δημιουργήθηκαν το μοντέλο της **διωνυμικής λογιστικής παλινδρόμησης** και το μοντέλο των **τυχαίων δασών** και πραγματοποιήθηκαν δύο παραλλαγές για το καθένα. Η παραλλαγή A περιελάμβανε όλες τις ανεξάρτητες μεταβλητές του τελικού πίνακα δεδομένων ενώ η παραλλαγή B περιελάμβανε τις μεταβλητές που κατατάχθηκαν ως πιο σημαντικές μετά την εφαρμογή της μεθόδου Boruta που εκτίμησε τη σημαντικότητά τους. Πιο συγκεκριμένα:

- **Παραλλαγή A:** : Event (εξαρτημένη) και Speed, AccLat, AccLon, THead, DRight, rdist, rspur, Wheel, Driving experience (ανεξάρτητες)
- **Παραλλαγή B:** Event (εξαρτημένη) και Speed, THead, DRight, rdist, Wheel και Driving experience (ανεξάρτητες)

Στους πίνακες 6.1 – 6.3 παρουσιάζονται τα αποτελέσματα των μοντέλων στατιστικής ανάλυσης και ανάλυσης παραγόντων. Η παραλλαγή B του μοντέλου διωνυμικής λογιστικής παλινδρόμησης δεν έδινε καθόλου ικανοποιητικά αποτελέσματα και δεν κρίθηκε αναγκαία η ανάλυσή του.

ΥΠΟΜΝΗΜΑ	
	Πολύ ικανοποιητικό
	Ικανοποιητικό
	Οριακά ικανοποιητικό
	Καθόλου ικανοποιητικό

Πίνακας 6.1: Μοντέλο διωνυμικής λογιστικής παλινδρόμησης (Παραλλαγή Α)

Μέτρα αξιολόγησης	Ποσοστό (%)	Αξιολόγηση
Ορθότητα (Accuracy)	79.26	
Ανάκληση (Recall)	13.44	
Εξειδικευτικότητα (Specificity)	99.04	
Ακρίβεια (Precision)	80.8	
Μέτρο F (F-measure)	23.05	
Δείκτης λάθος συναγερμού (False alarm rate)	0.96	
Εμβαδόν κάτω από την καμπύλη ROC (AUC)	68.18	

Πίνακας 6.2: Μοντέλο τυχαίων δασών (Παραλλαγή Α)

Μέτρα αξιολόγησης	Ποσοστό (%)	Αξιολόγηση
Ορθότητα (Accuracy)	87.17	
Ανάκληση (Recall)	65.56	
Εξειδικευτικότητα (Specificity)	93.67	
Ακρίβεια (Precision)	75.68	
Μέτρο F (F-measure)	70.26	
Δείκτης λάθος συναγερμού (False alarm rate)	6.33	

Πίνακας 6.3: Μοντέλο τυχαίων δασών (Παραλλαγή Β)

Μέτρα αξιολόγησης	Ποσοστό (%)	Αξιολόγηση
Ορθότητα (Accuracy)	81.19	
Ανάκληση (Recall)	53.51	
Εξειδικευτικότητα (Specificity)	89.51	
Ακρίβεια (Precision)	60.53	
Μέτρο F (F-measure)	56.81	
Δείκτης λάθος συναγερμού (False alarm rate)	10.49	

Επίσης, πραγματοποιήθηκε **παραγοντική ανάλυση** στους πίνακες που αφορούσαν στην κατάσταση ενός λεπτού πριν το συμβάν, στη διάρκεια του συμβάντος και στο σύνολο αυτών για τη διερεύνηση ύπαρξης κοινών παραγόντων (Πίνακας 6.4).

Πίνακας 6.4: Αποτελέσματα παραγοντικής ανάλυσης

Πίνακας	Παράγοντας 1	Παράγοντας 2	Παράγοντας 3
PreEventU2 (1 λεπτό πριν το συμβάν)	DRight	Speed	rspur
DurEventU2 (διάρκεια συμβάντος)	Speed	AccLat AccLon	
EventsU2 (σύνολο των παραπάνω)	AccLat	Speed	DRight

Σύμφωνα με τα αποτελέσματα που προέκυψαν από την εφαρμογή της μεθοδολογίας μπορούν να διατυπωθούν τα **συμπεράσματα** της παρούσας διπλωματικής εργασίας.

- ❖ Οι **μεταβλητές με τη μεγαλύτερη σημαντικότητα** για τον εντοπισμό ύπαρξης συμβάντων σε αστικές οδούς είναι η ταχύτητα, η συνολική διανυόμενη απόσταση, η απόσταση από τη δεξιά οριογραμμή, η θέση του τιμονιού, η οδηγική εμπειρία και ο χρόνος μέχρι τη σύγκρουση με το προπορευόμενο όχημα.
- ❖ Το **μοντέλο διωνυμικής λογιστικής παλινδρόμησης** με όλες τις σημαντικές μεταβλητές (Παραλλαγή Α) κρίθηκε **μη αξιόπιστο** καθώς ενώ παρουσίασε μεγάλη ορθότητα, δηλαδή συνολικά σωστές προβλέψεις, δεν προέβλεπε ικανοποιητικά την ύπαρξη συμβάντος (χαμηλός δείκτης ανάκλησης), που ήταν και ο στόχος της παρούσας διπλωματικής. Το ίδιο μοντέλο για μικρό αριθμό ανεξάρτητων μεταβλητών (Παραλλαγή Β) δεν λειτούργησε καθόλου ικανοποιητικά και δεν κρίθηκε άξιο περαιτέρω ανάλυσης.
- ❖ Το **μοντέλο τυχαίων δασών** παρουσίασε **αξιόπιστα αποτελέσματα** και στις δύο παραλλαγές του. Προέκυψαν υψηλοί δείκτες ορθότητας (σωστή πρόβλεψη ύπαρξης ή όχι συμβάντος) και εξειδικευτικότητας (πρόβλεψη μη ύπαρξης συμβάντος) και στις δύο παραλλαγές. Η **παραλλαγή Α κρίθηκε πιο αποτελεσματική** καθώς είχε αυξημένη πιθανότητα πρόβλεψης ύπαρξης συμβάντος (ανάκληση) και μειωμένη πιθανότητα λάθους πρόβλεψης ύπαρξης συμβάντος (δείκτης λάθος συναγερμού).
- ❖ Το **μοντέλο τυχαίων δασών βρέθηκε να δίνει πιο αξιόπιστα αποτελέσματα στον εντοπισμό συμβάντων σε σχέση με το μοντέλο διωνυμικής λογιστικής παλινδρόμησης** ίσως επειδή δεν θεωρεί ότι το μοντέλο έχει γραμμική σχέση όπως τα μοντέλα παλινδρόμησης και εξετάζει τυχαία δείγματα με στόχο την υψηλότερη ακρίβεια των αποτελεσμάτων.

- ❖ Και στα δύο μοντέλα στατιστικής ανάλυσης, η **παραλλαγή με τον μεγαλύτερο αριθμό μεταβλητών** παρουσίασε πιο ικανοποιητικά αποτελέσματα.
- ❖ Τα αποτελέσματα της παραγοντικής ανάλυσης έδειξαν ότι η κατάσταση που αφορά στο **ένα λεπτό πριν το συμβάν** μπορεί να περιγραφεί μέσω τριών παραγόντων: i) της επιρροής της απόστασης από τη δεξιά οριογραμμή, ii) της ταχύτητας και iii) της απόκλισης του οχήματος από το μέσο της οδού. Η ταχύτητα είναι καθοριστικός παράγοντας για την πιθανότητα εμπλοκής σε ένα συμβάν αλλά και οι αποστάσεις του οχήματος από την δεξιά οριογραμμή και το μέσο της οδού παίζουν σημαντικό ρόλο καθώς σε ένα απροσδόκητο συμβάν το όχημα μπορεί να βρεθεί εκτός πορείας.
- ❖ Όσον αφορά στην κατάσταση **κατά τη διάρκεια του συμβάντος** οι παράγοντες που μπορούν να την περιγράψουν είναι δύο: i) ένας που περιγράφει την επιρροή της ταχύτητας και ii) ένας που περιγράφει την επιρροή της πλευρικής και της διαμήκους επιτάχυνσης. Όπως και η ταχύτητα έτσι και η πλευρική και η διαμήκης επιτάχυνση είναι παράγοντες με καθοριστική σημασία στη δυναμική της κίνησης του οχήματος και συνακόλουθα στην πιθανότητα εμπλοκής σε μη αναμενόμενο συμβάν.
- ❖ Από την παραγοντική ανάλυση επίσης προέκυψε πως τα δεδομένα που αφορούν στην κατάσταση **ενός λεπτού πριν και κατά τη διάρκεια του συμβάντος** μπορούν να περιγραφούν από τρεις παράγοντες: i) την επιρροή της πλευρικής επιτάχυνσης, ii) την επιρροή της ταχύτητας και iii) την επιρροή της απόστασης από τη δεξιά οριογραμμή καθώς αποτελούν το σύνολο των παραπάνω.

ΠΕΡΙΕΧΟΜΕΝΑ

1. ΕΙΣΑΓΩΓΗ.....	1
1.1. Γενική ανασκόπηση.....	1
1.2. Στόχος.....	2
1.3. Μεθοδολογία διπλωματικής εργασίας	3
1.4. Δομή διπλωματικής εργασίας.....	4
 2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ.....	6
2.1. Γενικά.....	6
2.2. Συναφείς έρευνες και μεθοδολογίες	6
2.2.1. Ανάλυση παραγόντων που οδηγούν σε λανθασμένη οδηγική συμπεριφορά..	6
2.2.2. Επίδραση της απόσπασης προσοχής στην οδηγική απόδοση.....	7
2.2.3. Μεταβολή των χαρακτηριστικών του οδηγού στην εμφάνιση ενός συμβάντος	10
2.3. Σύνοψη	11
 3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ.....	16
3.1. Εισαγωγή	16
3.2. Μαθηματικά πρότυπα	16
3.2.1. Μοντέλο διωνυμικής λογιστικής παλινδρόμησης	16
3.2.2. Τυχαία Δάση	17
3.3. Κριτήρια αποδοχής μοντέλων	18
3.3.1. Βασικά κριτήρια ελέγχου λογιστικού μοντέλου	18
3.3.2. Μήτρα σύγχυσης (Confusion Matrix)	18
3.4. Ανάλυση παραγόντων (Factor Analysis).....	21
3.5. Σημαντικότητα μεταβλητών (μέθοδος Boruta).....	22
 4. ΣΥΛΛΟΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΣΤΟΙΧΕΙΩΝ	24
4.1. Εισαγωγή	24
4.2. Συλλογή στοιχείων	24

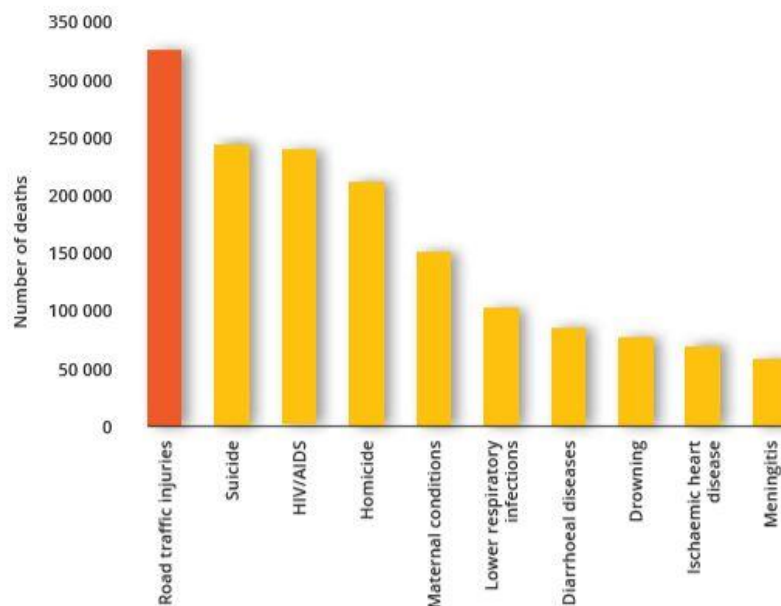
4.3. Βάση δεδομένων.....	28
4.4. Επεξεργασία στοιχείων	28
4.5. Περιγραφική Στατιστική	31
4.6. Συσχέτιση μεταβλητών.....	33
4.7. Πίνακες για τα μοντέλα στατιστικής ανάλυσης	34
4.8. Πίνακες για τη μέθοδο Παραγοντικής Ανάλυσης	35
5. ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ – ΑΠΟΤΕΛΕΣΜΑΤΑ	38
5.1. Εισαγωγή	38
5.2. Μοντέλο διωνυμικής λογιστικής παλινδρόμησης.....	39
5.2.1. Παραλλαγή Α	39
5.2.2. Παραλλαγή Β	43
5.3. Μοντέλο τυχαίων δασών.....	44
5.3.1. Παραλλαγή Α	44
5.3.2. Παραλλαγή Β	46
5.4. Ανάλυση παραγόντων.....	49
5.4.1. Παραγοντική ανάλυση στον πίνακα PreEventU2	49
5.4.2. Παραγοντική ανάλυση στον πίνακα DurEventU2.....	52
5.4.3. Παραγοντική ανάλυση στον πίνακα EventsU2.....	55
5.5. Σύνοψη αποτελεσμάτων	58
6. ΣΥΜΠΕΡΑΣΜΑΤΑ	60
6.1. Σύνοψη αποτελεσμάτων	60
6.2. Συνολικά συμπεράσματα	62
6.3. Προτάσεις για αξιοποίηση των αποτελεσμάτων	63
6.4. Προτάσεις για περαιτέρω έρευνα	64
ΒΙΒΛΙΟΓΡΑΦΙΑ	66
ΠΑΡΑΡΤΗΜΑ	68

1. ΕΙΣΑΓΩΓΗ

1.1. Γενική ανασκόπηση

Η **οδική ασφάλεια** αποτελεί σοβαρό κοινωνικό θέμα καθώς χιλιάδες άνθρωποι χάνουν τη ζωή τους ή τραυματίζονται σοβαρά καθημερινά στην ασφαλτο. Πιο συγκεκριμένα σύμφωνα με έκθεση του Παγκόσμιου Οργανισμού Υγείας ο αριθμός των ανθρώπων που σκοτώνονται σε τροχαία δυστυχήματα ανήλθε μόνο για το 2016 σε 1,35 εκατομμύρια παγκοσμίως (ΠΟΥ, 2018), ενώ τα τροχαία δυστυχήματα αποτελούν σήμερα την πλέον κοινή αιτία θανάτου παιδιών και νέων ηλικίας 5 – 29 ετών σε ολόκληρο τον κόσμο (Γράφημα 1.1). Η Ελλάδα, επίσης, κατέχει την τρίτη υψηλότερη θέση στους δείκτες θνησιμότητας από τροχαία ατυχήματα μεταξύ παιδιών, εφήβων και νέων έως 25 ετών στην Ευρωπαϊκή Ένωση.

Top ten causes of death among people aged 15–29 years, 2012



Γράφημα 1.1: Δέκα κορυφαίες αιτίες θανάτου μεταξύ νέων 15-29 ετών (ΠΟΥ,2014)

Το όχημα, η οδός και το περιβάλλον και οι χρήστες της οδού αποτελούν τους τρεις βασικούς παράγοντες που επιδρούν στην οδική ασφάλεια (Φραντζεσκάκης και Γκόλιας 1994). Ο σημαντικότερος από αυτούς τους παράγοντες για τα οδικά ατυχήματα είναι ο άνθρωπος, είτε ως οδηγός, είτε ως επιβάτης είτε ως πεζός. Ο άνθρωπος μπορεί να προκαλέσει κάποιο συμβάν ή ατύχημα όταν δεν συμμορφώνεται με τους κανόνες οδικής κυκλοφορίας και ασκεί επιθετική οδήγηση (ανάπτυξη υψηλών ταχυτήτων), όταν αποσπά την προσοχή του από το κύριο καθήκον της οδήγησης (χρήση κινητού τηλεφώνου, συνομιλία με συνεπιβάτες) ή όταν διαπράττει κάποιο οδηγικό λάθος (απώλεια ελέγχου οχήματος, αδυναμία τήρησης αποστάσεων).

Οι παραπάνω ανθρώπινες ενέργειες μπορούν να επηρεάσουν άμεσα τα οδηγικά χαρακτηριστικά όπως την ταχύτητα του οδηγού, τη θέση του οχήματος στη λωρίδα, τους χρονικούς διαχωρισμούς των οχημάτων, τις αποστάσεις ασφαλείας από άλλα οχήματα και από τα όρια της οδού, τον χρόνο αντίδρασης του οδηγού και συνεπώς να οδηγήσουν σε αύξηση της πιθανότητας οδικού ατυχήματος.

Η **ταχύτητα** αποτελεί ίσως τον σημαντικότερο παράγοντα κινδύνου που επηρεάζει την πιθανότητα ατυχήματος αλλά και τον δείκτη σοβαρότητας ατυχημάτων καθώς έχει υπολογιστεί ότι αύξηση ενός χιλιομέτρου στη μέση ταχύτητα οδηγεί σε 3% αύξηση της πιθανότητας εμπλοκής σε ατύχημα. Η υπερβολική (μεγαλύτερη του ισχύοντος ορίου) και η μη κατάλληλη (μικρότερη του ισχύοντος ορίου αλλά μη κατάλληλη για τις υπάρχουσες συνθήκες) ταχύτητα συμβάλλουν στην πρόκληση τροχαίων ατυχημάτων και ευθύνονται για ένα υψηλό ποσοστό θανάτων που προκύπτουν από αυτά. Ο έλεγχος της ταχύτητας του οχήματος μπορεί να αποτρέψει τα ατυχήματα και να μειώσει τον αντίκτυπο αυτών όταν συμβαίνουν, μειώνοντας τη σοβαρότητα των τραυματισμών.

Η οδική ασφάλεια στις **αστικές περιοχές** αποτελεί σημαντικό ζήτημα καθώς το μεγαλύτερο ποσοστό οδικών ατυχημάτων με παθόντες γίνονται σε κατοικημένες περιοχές. Αυτό οφείλεται στο γεγονός ότι στο αστικό περιβάλλον κυκλοφορούν πολλοί ευάλωτοι χρήστες της οδού (πεζοί, ποδηλάτες, μοτοσικλετιστές, παιδιά, ηλικιωμένοι) που είναι πιο πιθανό να εμπλακούν σε συμβάν λόγω της αυξημένης κυκλοφορίας.

Τα παραπάνω τονίζουν την ανάγκη προσδιορισμού των κρίσιμων παραγόντων **εντοπισμού οδικών συμβάντων** που συμβαίνουν σε αστικές περιοχές με σκοπό την έγκαιρη πρόβλεψή τους και συνεπώς την προσπάθεια για αποφυγή ή κατάλληλη αντιμετώπιση.

Σε όλα τα παραπάνω θα συμβάλλει η παρούσα διπλωματική εργασία με στόχο τον προσδιορισμό των κρίσιμων παραγόντων εντοπισμού οδικών συμβάντων σε αστικές περιοχές.

1.2. Στόχος

Σύμφωνα με τα παραπάνω η έγκαιρη και αξιόπιστη ανίχνευση απροσδόκητων συμβάντων μπορεί να συμβάλλει θετικά στην οδική ασφάλεια. Συνεπώς, ο στόχος της παρούσας διπλωματικής εργασίας είναι **ο προσδιορισμός των κρίσιμων παραγόντων εντοπισμού οδικών συμβάντων σε αστικές περιοχές**.

Η αξιοποίηση δεδομένων που συλλέχθηκαν μέσω πειράματος προσομοιωτή οδήγησης παλαιότερης έρευνας και αφορούν σε οδηγικά χαρακτηριστικά καθώς και κατάλληλης μεθοδολογίας θα συμβάλλουν στον προσδιορισμό των κρίσιμων παραγόντων εντοπισμού οδικών συμβάντων ειδικά σε αστικές περιοχές.

Ως επιμέρους στόχοι της διπλωματικής εργασίας είναι η εφαρμογή του κατάλληλου θεωρητικού υποβάθρου και η επιλογή αξιόπιστων μεθόδων στατιστικής ανάλυσης που θα οδηγήσουν στην ανάπτυξη μαθηματικών μοντέλων με έγκυρα αποτελέσματα που αφορούν στην κατάσταση ενός λεπτού πριν και κατά τη διάρκεια ενός συμβάντος. Επιπλέον, σημαντικό στόχο αποτελεί η ανάλυση παραγόντων ώστε να πραγματοποιηθεί ομαδοποίηση των ανεξάρτητων μεταβλητών για την κατάσταση που περιγράφει ένα λεπτό πριν το συμβάν, τη διάρκεια του συμβάντος αλλά και τον συνδυασμό αυτών.

Μέσω των αποτελεσμάτων αυτών μπορούν να προκύψουν προτάσεις για τη βελτίωση της οδικής ασφάλειας, καθώς και για περαιτέρω έρευνα του εν λόγω αντικειμένου.

1.3. Μεθοδολογία διπλωματικής εργασίας

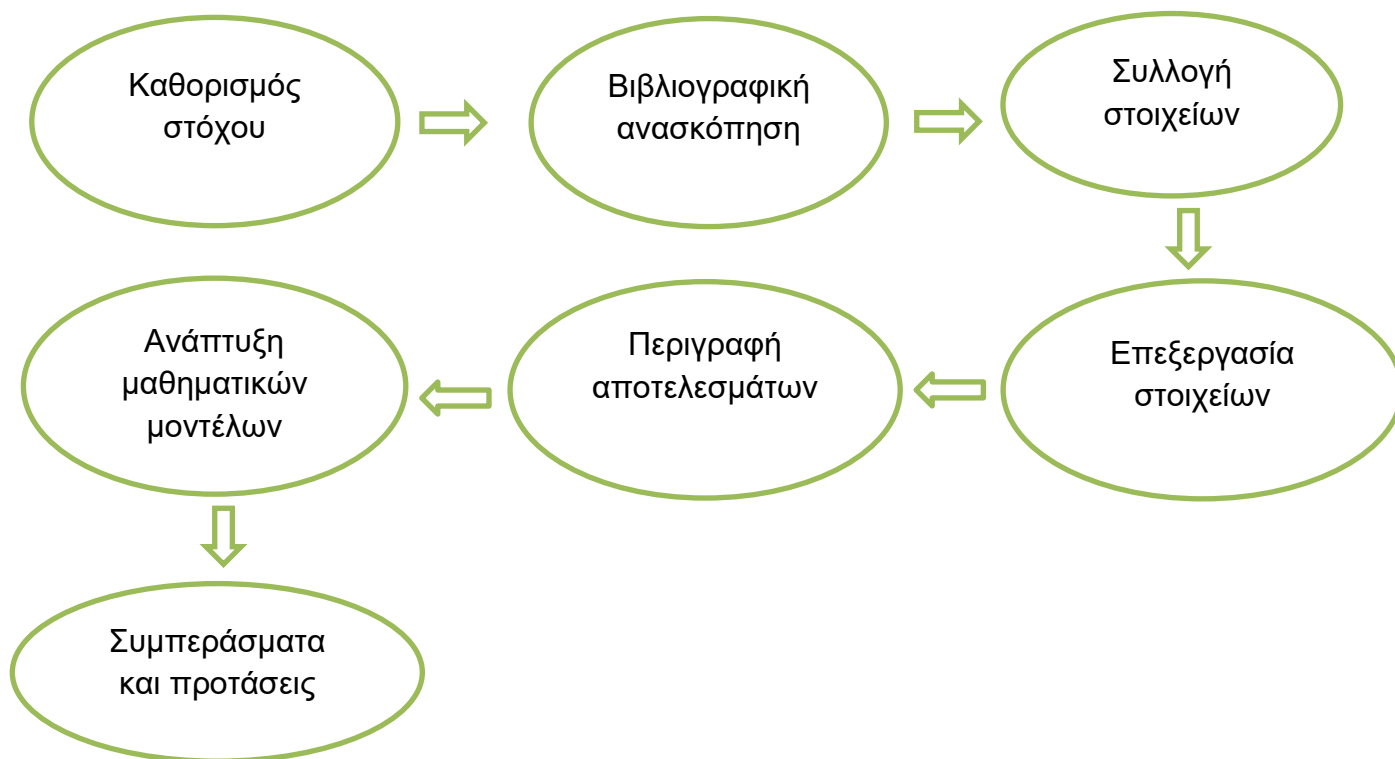
Παρακάτω περιγράφεται η **μεθοδολογία** που ακολουθήθηκε για την επίτευξη του στόχου της διπλωματικής εργασίας.

Αφού γίνει ο καθορισμός του **στόχου** και του θέματος της μελέτης, είναι απαραίτητη η αναζήτηση **συναφών ερευνών** και μεθοδολογιών ανάλυσης από πρόσφατη διεθνή βιβλιογραφία ώστε να εντοπιστούν τα ζητήματα που απαιτούν περαιτέρω έρευνα καθώς και ο τρόπος ανάλυσής τους.

Ακολουθεί η **συλλογή** των στοιχείων και η **επεξεργασία** τους για τη διαμόρφωση της τελικής βάσης δεδομένων. Έπειτα, με ανάπτυξη κατάλληλων μοντέλων **στατιστικής ανάλυσης** προκύπτει σειρά αποτελεσμάτων που αφορούν στην πρόβλεψη ύπαρξης συμβάντων.

Τα παραπάνω βήματα οδηγούν στην επίτευξη του στόχου της παρούσας διπλωματικής εργασίας και στα αντίστοιχα **συμπεράσματα**. Επιπλέον, γίνεται αναφορά σε κάποιες **προτάσεις** για περαιτέρω έρευνα που θα οδηγήσουν στην περαιτέρω και πιο ολοκληρωμένη μελέτη του αντικειμένου.

Τα βήματα της μεθοδολογίας της παρούσας διπλωματικής εργασίας παρουσιάζονται συνοπτικά στο παρακάτω διάγραμμα ροής (Γράφημα 1.1).



Γράφημα 1.1: Βήματα μεθοδολογίας διπλωματικής εργασίας

1.4. Δομή διπλωματικής εργασίας

Παρακάτω παρουσιάζεται η δομή της διπλωματικής εργασίας μέσω της σύνοψης των κεφαλαίων που την απαρτίζουν.

Το παρόν κεφάλαιο αποτελεί το **κεφάλαιο 1**, το οποίο είναι εισαγωγικό και παρουσιάζει την άμεση σύνδεση των κρίσιμων παραγόντων εντοπισμού οδικών συμβάντων σε αστικές περιοχές με την οδική ασφάλεια. Επίσης, παρουσιάζεται ο στόχος της διπλωματικής εργασίας καθώς και η μεθοδολογία που ακολουθήθηκε για την εκπόνησή της.

Το **κεφάλαιο 2** αποτελεί τη βιβλιογραφική ανασκόπηση όπου παρουσιάζονται συναφείς έρευνες και μεθοδολογίες, οι οποίες αποτελούν βάση πάνω στην οποία στηρίζεται ο τελικός καθορισμός του αντικειμένου της εργασίας καθώς και οι παράγοντες που θα εξεταστούν.

Στο **κεφάλαιο 3** παρουσιάζεται το θεωρητικό υπόβαθρο στο οποίο αναλύονται οι μέθοδοι στατιστικής ανάλυσης που θα χρησιμοποιηθούν. Περιγράφονται οι τύποι των μαθηματικών μοντέλων καθώς και οι έλεγχοι για την αποδοχή τους.

Στο **κεφάλαιο 4** παρουσιάζεται η διαδικασία συλλογής των στοιχείων από μεγάλη βάση δεδομένων παλαιάς έρευνας και η επεξεργασία τους για τη διαμόρφωση των τελικών πινάκων προς ανάλυση.

Στο **κεφάλαιο 5** αναλύονται τα βήματα που ακολουθήθηκαν για την ανάπτυξη των μαθηματικών μοντέλων στατιστικής ανάλυσης και πληρούν τα κριτήρια αποδοχής. Χρησιμοποιήθηκαν δύο μοντέλα, το μοντέλο διωνυμικής λογιστικής παλινδρόμησης και το μοντέλο τυχαίων δασών. Επίσης πραγματοποιήθηκε παραγοντική ανάλυση για την αναζήτηση ύπαρξης κοινών παραγόντων σε μια ομάδα μεταβλητών.

Στο **κεφάλαιο 6** παρουσιάζονται τα συμπεράσματα που προκύπτουν από την αξιολόγηση των αποτελεσμάτων και αναφέρονται προτάσεις για την αξιοποίησή τους καθώς και προτάσεις για περαιτέρω έρευνα του αντικειμένου.

Στο **κεφάλαιο 7** παρουσιάζεται η βιβλιογραφία, σε μορφή καταλόγου, η οποία περιλαμβάνει όλες τις πηγές που χρησιμοποιήθηκαν και συνέβαλλαν στην εκπόνηση της διπλωματικής εργασίας.

2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

2.1. Γενικά

Στόχος της βιβλιογραφικής ανασκόπησης είναι ο καθορισμός του αντικειμένου της διπλωματικής εργασίας και η εύρεση της κατάλληλης μεθοδολογίας που θα ακολουθηθεί για την εκπόνησή της. Αποτελεί βάση πάνω στην οποία δομείται η εργασία, καθώς παρουσιάζονται **συναφείς έρευνες και μεθοδολογίες**, ώστε να προκύψουν βασικά συμπεράσματα σύμφωνα με τα οποία θα επιλεγεί η ερευνητική διαδρομή της παρούσας διπλωματικής εργασίας.

Πιο συγκεκριμένα, η αναζήτηση των ερευνών επικεντρώνεται στην ανάλυση των παραγόντων που οδηγούν σε λανθασμένη και μη ασφαλή οδηγική συμπεριφορά, στην επίδραση της απόσπασης προσοχής στην οδηγική απόδοση καθώς και στην μεταβολή των χαρακτηριστικών του οδηγού στην εμφάνιση ενός συμβάντος. Η βιβλιογραφική ανασκόπηση στηρίζεται σε αυτές τις κατευθύνσεις καθώς για την εκπόνηση της παρούσας έρευνας χρειάζονται πληροφορίες για τον τρόπο και τον λόγο που μεταβάλλεται η συμπεριφορά του οδηγού κυρίως όταν συμβαίνουν αναπάντεχα συμβάντα.

Τέλος, εντοπίζονται πιθανά κενά στις υπάρχουσες αναφορές για το διερευνώμενο θέμα και έτσι κρίνεται σκόπιμη η εξέταση του συγκεκριμένου αντικειμένου της διπλωματικής εργασίας με την κατάλληλη μεθοδολογία.

2.2. Συναφείς έρευνες και μεθοδολογίες

2.2.1. Ανάλυση παραγόντων που οδηγούν σε λανθασμένη οδηγική συμπεριφορά

Η μελέτη των Monselise et al. (2019) εξετάζει πολλούς **παράγοντες ατυχημάτων** και στοχεύει στη δημιουργία ενός μοντέλου που μπορεί να προβλέψει την εμφάνιση ενός ατυχήματος δεδομένων των οδικών συνθηκών και της συμπεριφοράς του οδηγού. Η μέθοδος μηχανικής μάθησης που επιλέχθηκε για αυτήν τη μελέτη παράγει ένα μοντέλο που μέσα από δεδομένα 7,707 ταξιδιών επιτρέπει την κατάταξη της σημασίας κάθε παράγοντα που προκαλεί τροχαία ατυχήματα. Πιο σημαντικές μεταβλητές στο μοντέλο πρόβλεψης βρέθηκαν η συμπεριφορά του οδηγού, οι ελιγμοί πριν από το συμβάν και η διάρκεια δευτερευουσών ενεργειών. Πιο συγκεκριμένα, ο πιο σημαντικός παράγοντας που μπορεί να οδηγήσει σε ατύχημα είναι **η συμπεριφορά του οδηγού** δηλαδή σε τι ενέργειες προχώρησε ώστε να προκαλέσει ένα ατύχημα ή μια πρόσκρουση όπως επιθετική οδήγηση, υπνηλία και υπέρβαση του ορίου ταχύτητας. Όσον αφορά του ελιγμούς πριν το συμβάν, η μελέτη επικεντρώθηκε κυρίως σε ατυχήματα σε

διασταυρώσεις και τα αποτελέσματα έδειξαν ότι πιο επικίνδυνες ενέργειες που προκαλούν εμπλοκή σε ατύχημα θεωρούνται η δεξιά στροφή και η είσοδος σε θέση στάθμευσης. Τέλος, περίπου τα δύο τρίτα των οδηγών που συμμετείχαν σε δευτερεύουσες ενέργειες που δεν σχετίζονται με την οδήγηση όπως η ομιλία, το τραγούδι, το φαγητό και το κινητό τηλέφωνο που έχουν μεγάλη διάρκεια, είχαν εμπλοκή σε ατύχημα.

Οι Papantoniou et al. (2019a) σε άρθρο τους διερευνούν διάφορους **παράγοντες** όπως η απόσπαση προσοχής του οδηγού, τα χαρακτηριστικά του οδηγού και το οδικό περιβάλλον, που επηρεάζουν την **εσφαλμένη οδηγική συμπεριφορά**. Πραγματοποίησαν πείραμα σε προσομοιωτή οδήγησης στο οποίο 95 συμμετέχοντες όλων των ηλικιών πήραν μέρος σε 6 δοκιμές σε αστικό και υπεραστικό περιβάλλον με διαφορετικά είδη απόσπασης προσοχής και διάφορες συνθήκες κυκλοφορίας. Τα αποτελέσματα του πολύπλοκου μοντέλου που δημιουργήθηκε από τα δεδομένα του πειράματος αποκαλύπτουν ότι η **επίδραση των χαρακτηριστικών του οδηγού και του τύπου της περιοχής** είναι οι μόνοι στατιστικά σημαντικοί παράγοντες που επηρεάζουν την πιθανότητα οδηγικών σφαλμάτων.

Από τις παραπάνω μελέτες, προκύπτει πως οι κύριοι παράγοντες που μπορούν να προκαλέσουν **λάθη κατά την οδήγηση** είναι η συμπεριφορά του οδηγού, η απόσπαση προσοχής και το περιβάλλον οδήγησης.

2.2.2. Επίδραση της απόσπασης προσοχής στην οδηγική απόδοση

Σε πρόταση μεθοδολογίας ταξινόμησης μέσω μηχανικής μάθησης σε δύο επίπεδα προχώρησαν οι Osman et al. (2019). Σε πρώτο επίπεδο ερευνάται εάν ο οδηγός εμπλέκεται σε κάποια **δευτερεύουσα ενέργεια** ενώ σε δεύτερο επίπεδο εντοπίζεται ο τύπος της ενέργειας αυτής (τηλεφωνική κλήση, γραπτά μηνύματα ή αλληλεπίδραση με συνεπιβάτη). Για την εκπόνηση της μελέτης χρησιμοποιήθηκαν δεδομένα από 373 οδηγούς (95 σε τηλεφωνική κλήση, 96 σε γραπτά μηνύματα, 84 σε συνομιλία με συνεπιβάτη και 98 που δεν εμπλέκονται σε καμία ενέργεια) και εισήχθησαν στα μοντέλα πέντε παράμετροι οδηγικής συμπεριφοράς (ταχύτητα, διαμήκης επιτάχυνση, πλευρική επιτάχυνση, θέση πεντάλ και ρυθμός εκτροπής) μαζί με τις τυπικές αποκλίσεις τους. Τα μοντέλα έδειξαν πως η ακρίβεια εντοπισμού της εμπλοκής σε δευτερεύουσα ενέργεια κυμαινόταν από 66% έως 96%. Για τον προσδιορισμό του τύπου της δευτερεύουσας ενέργειας, η συνολική ακρίβεια κυμαινόταν από 55% έως 79%, με την υψηλότερη ακρίβεια 82,2% να επιτυγχάνεται από τη μέθοδο τυχαίων δασών. Τα ευρήματα της εργασίας δείχνουν πως η εμπλοκή των οδηγών σε δευτερεύουσες ενέργειες (όπως αποστολή γραπτών μηνυμάτων) λειτουργεί ως αντίθετο μέτρο για την πρόληψη ατυχημάτων και εφιστά την προσοχή τους στο κύριο καθήκον της οδήγησης όταν πραγματοποιούνται επικίνδυνες αλλαγές στην οδηγική τους συμπεριφορά.

Επιπρόσθετα οι Choudhary και Velaga (2019a) με μελέτη τους ανέλυσαν και συνέκριναν τις επιπτώσεις της **απόσπασης προσοχής** που προκαλούνται από το φαγητό, το ποτό και τα γραπτά μηνύματα σχετικά με την επικίνδυνη οδηγική

συμπεριφορά σε διασταυρώσεις χωρίς σηματοδότηση. Για το σκοπό αυτό, 89 συμμετέχοντες πλησίασαν τις διασταυρώσεις χωρίς σηματοδότηση από τον δευτερεύων δρόμο και συνάντησαν μια σειρά οχημάτων στον κύριο δρόμο μέσω προσομοιωτή οδήγησης. Τα δεδομένα που συλλέχθηκαν αναλύθηκαν ως προς τη μέση απόσταση αντίδρασης και την πιθανότητα ατυχήματος στις ζώνες προσέγγισης και ολοκλήρωσης αντίστοιχα. Σύμφωνα με τα αποτελέσματα παρατηρήθηκε ότι η απόσταση αντίδρασης αυξήθηκε κατά την κατανάλωση φαγητού και ποτού σε σύγκριση με την αρχική τιμή, ενώ μειώθηκε όταν οι οδηγοί συμμετείχαν στην εργασία γραπτών μηνυμάτων. Η κατανάλωση φαγητού και ποτού δεν επηρέασε αρνητικά την οδηγική απόδοση στη ζώνη προσέγγισης, ωστόσο ο κίνδυνος ατυχήματος στη ζώνη ολοκλήρωσης αυξήθηκε σημαντικά. Απρροσδόκητα, ο κίνδυνος ατυχήματος ήταν χαμηλός (76%) κατά την αποστολή γραπτών μηνυμάτων, πιθανώς επειδή οι οδηγοί αντιλαμβάνονται τη συγκεκριμένη ενέργεια ως πιο επικίνδυνη και προσαρμόζουν αναλόγως τη συμπεριφορά τους.

Η μελέτη των Choudhary και Velaga (2019b) έχει ως στόχο να συγκρίνει τον αντίκτυπο της **συνομιλίας** και των **γραπτών μηνυμάτων** στην οδηγική απόδοση των άπειρων νέων οδηγών με αυτή των επαγγελματιών οδηγών. Για την εκπόνηση των πειραμάτων της έρευνας, που έγιναν με τη βοήθεια προσομοιωτή οδήγησης, χρησιμοποιήθηκε δείγμα 25 νέων οδηγών(κάτω των 25 ετών) και 24 επαγγελματιών οδηγών (άνω των 25 ετών) που εργάζονται ως οδηγοί σε μεταφορική εταιρεία. Η συγκριτική ανάλυση πραγματοποιήθηκε για τον διαμήκη (ταχύτητα και επιτάχυνση) και πλευρικό (θέση στη λωρίδα και κινήσεις διεύθυνσης) έλεγχο του οχήματος. Η σύγκριση καταλήγει στο ότι οι νέοι οδηγοί έχουν χαμηλότερο διαμήκη έλεγχο όταν υπάρχει απόσπαση προσοχής κατά τη διάρκεια της οδήγησης, αλλάζουν πιο συχνά θέση στη λωρίδα όμως καταφέρνουν να διατηρούν λιγότερες αντιστροφές τιμονιού. Το συμπέρασμα της έρευνας είναι πως η χρήση τηλεφώνου έχει μεγαλύτερη επίπτωση στην απόδοση των άπειρων νέων οδηγών χωρίς να υποβαθμίζει ωστόσο την επίδραση της και στους επαγγελματίες οδηγούς δείχνοντας ότι η υψηλότερη εμπειρία δεν μπορεί να εξουδετερώσει τις επιπτώσεις που δημιουργεί η απόσπαση προσοχής στην οδήγηση.

Μια ακόμη έρευνα που μελετά τους νέους οδηγούς είναι αυτή των Yannis et al. (2014) που εξετάζει την επίδραση των **γραπτών μηνυμάτων** στη συμπεριφορά και την ασφάλειά τους σε αστικές και υπεραστικές οδούς. Σε πείραμα που πραγματοποιήθηκε σε προσομοιωτή οδήγησης 34 νέοι οδήγησαν σε διαφορετικά σενάρια οδήγησης(καλές-κακές καιρικές συνθήκες, μέρα-νύχτα). Για την εξαγωγή συμπερασμάτων στη μεταβολή της μέσης ταχύτητας και του μέσου χρόνου αντίδρασης αναπτύχθηκαν μοντέλα κανονικής γραμμικής παλινδρόμησης, ενώ για την πιθανότητα ατυχήματος μοντέλο λογιστικής παλινδρόμησης. Τα αποτελέσματα των μοντέλων δείχνουν ότι η αποστολή γραπτών μηνυμάτων οδηγεί σε σημαντική μείωση της μέσης ταχύτητας και αύξηση του μέσου χρόνου αντίδρασης, άρα και σε αυξημένη πιθανότητα ατυχήματος σε συνδυασμό με την απόσπαση προσοχής. Σε αστικούς δρόμους, η πιθανότητα ατυχήματος είναι 2,9 φορές υψηλότερη όταν γίνεται ανάγνωση SMS και 8,3 φορές όταν γίνεται σύνταξη, σε σύγκριση με απερίσπαστη οδήγηση. Η αντίστοιχη πιθανότητα ατυχήματος είναι 1,4 και 1,5 φορές υψηλότερη σε υπεραστικούς δρόμους.

Επίσης, με σκοπό να αποτυπώσουν τις επιδράσεις του **γραφπού μηνύματος** στη συμπεριφορά των νέων οδηγών οι Dimitriou et al. (2018) πραγματοποίησαν ένα πείραμα σε προσομοιωτή οδήγησης όπου 34 νέοι οδήγησαν σε διαφορετικά περιβάλλοντα και σενάρια οδήγησης όπως αυτοκινητόδρομοι, αστικοί και αγροτικοί δρόμοι, κατά τη διάρκεια της ημέρας και της νύχτας και σε καλές ή άσχημες καιρικές συνθήκες. Με την ανάλυση Corula εξετάστηκε η συμπεριφορά των οδηγών που συμμετείχαν σε σύγκρουση καθώς και των οδηγών που δεν συμμετείχαν σε σύγκρουση και στη συνέχεια ταυτοποιήθηκαν οι μεταβλητές με τις μεγαλύτερες συσχετίσεις. Αποδείχθηκε ότι το ζεύγος μεταβλητών που παρουσίασε τις ισχυρότερες συσχετίσεις στις εξεταζόμενες συνθήκες κυκλοφορίας ήταν η θέση στη λωρίδα και η ταχύτητα. Λαμβάνοντας τα αποτελέσματα της παραπάνω ανάλυσης, εκτελέστηκαν τα μοντέλα Gaussian Mixture τα οποία έδειξαν ότι οι οδηγοί που χρησιμοποιούσαν το κινητό τηλέφωνο και συμμετείχαν σε σύγκρουση παρουσίασαν διαφορετική συμπεριφορά οδήγησης από αυτούς που ήταν μεν απασχολημένοι αλλά δεν ενεπλάκησαν σε σύγκρουση.

Οι Choudhary και Velaga (2018) επιχείρησαν την ποσοτικοποίηση των επιπτώσεων των γραπτών μηνυμάτων και της οδηγικής συμπεριφοράς στον **κίνδυνο ατυχήματος** που σχετίζεται με ένα **ξαφνικό συμβάν** καθώς και τη σύγκριση των επιπτώσεων σε άπειρους και επαγγελματίες οδηγούς που διατρέχουν κίνδυνο κατά τη φάση και μετά την ανίχνευση του γεγονότος. Οι 49 συμμετέχοντες στο πείραμα οδήγησαν σε τρία διαφορετικά σενάρια στον προσομοιωτή οδήγησης: χωρίς απόσπαση προσοχής, με σύνταξη σύντομων και μεγάλων κειμένων ενώ εκτέθηκαν και σε ένα ξαφνικό επικίνδυνο συμβάν κατά τη διάρκεια κάθε οδήγησης. Από τα αποτελέσματα του γενικευμένου γραμμικού μικτού μοντέλου που αναπτύχθηκε, παρατηρήθηκε όπως αναμενόταν ότι και στα δύο σενάρια σύνταξης γραπτών μηνυμάτων ο κίνδυνος ατυχήματος αυξήθηκε, και μάλιστα πολύ περισσότερο στους νέους οδηγούς μετά την ανίχνευση του συμβάντος καθώς ανέπτυσαν υψηλότερες ταχύτητες προσέγγισης από ότι οι επαγγελματίες.

Σε ακόμα μία έρευνα των Choudhary και Velaga (2017) με αντικείμενο την **απόσπαση προσοχής** αναλύθηκε η επίδραση της συνομιλίας και των γραπτών μηνυμάτων στην οδηγική απόδοση. Πραγματοποιήθηκε πείραμα σε προσομοιωτή οδήγησης στο οποίο 100 οδηγοί τριών διαφορετικών ηλικιακών ομάδων οδήγησαν σε συνθήκες αγροτικής εθνικής οδού και αναλύθηκε η συμπεριφορά τους κυρίως όσον αφορά τη μέση ταχύτητα και τις ικανότητες αποφυγής ατυχημάτων. Τα μοντέλα της γραμμικής παλινδρόμησης έδειξαν ότι οι οδηγοί αντιστάθμισαν σημαντικά την απόσπαση προσοχής από τη χρήση κινητού τηλεφώνου μειώνοντας την ταχύτητά τους και διατηρώντας κατάλληλα επίπεδα ασφαλείας, ενώ σύμφωνα με τα μοντέλα δυαδικής λογιστικής παλινδρόμησης η **πιθανότητα ατυχήματος αυξήθηκε** κατά 3 και 4 φορές αντίστοιχα όταν οι οδηγοί συνομιλούν ή στέλνουν μηνύματα στο τηλέφωνο.

Από τις παραπάνω έρευνες συμπεραίνεται ότι η **απόσπαση προσοχής** του οδηγού μεταβάλλει μη φυσιολογικά τη συμπεριφορά του και έχει μεγαλύτερη επίπτωση στους νέους οδηγούς, οι οποίοι είναι πιο επιρρεπείς στο να αναπτύσσουν υψηλές ταχύτητες και υστερούν εμπειρίας στο χειρισμό του οχήματος, γεγονότα που αυξάνουν την πιθανότητα εμπλοκής σε **ατύχημα**. Επίσης παρατηρείται ότι το τηλέφωνο, ως τύπος

απόσπασης προσοχής, προκαλεί χαμηλότερη πιθανότητα ατυχήματος καθώς οι οδηγοί αντιλαμβάνονται τη συγκεκριμένη ενέργεια ως επικίνδυνη και προσαρμόζουν αναλόγως τη συμπεριφορά τους, σε αντίθεση με την συνομιλία με συνεπιβάτη ή το φαγητό που αποσπούν πιο εύκολα την προσοχή του οδηγού προκαλώντας μεγαλύτερη πιθανότητα λάθους.

2.2.3. Μεταβολή των χαρακτηριστικών του οδηγού στην εμφάνιση ενός συμβάντος

Οι Papantoniou et al. (2019b) σε έρευνα τους στόχευσαν στη διερεύνηση της επίδρασης ενός **απροσδόκητου συμβάντος** στα οδηγικά μέτρα απόδοσης που σχετίζονται με την **ταχύτητα** και πιο συγκεκριμένα στη μέση ταχύτητα και στη μεταβλητότητα της ταχύτητας. Πραγματοποιήθηκε πείραμα προσομοιωτή οδήγησης στο οποίο 95 συμμετέχοντες οδήγησαν υπό διαφορετικούς τύπους απόσπασης προσοχής (χωρίς απόσπαση της προσοχής, συνομιλία με τον επιβάτη, χρήση τηλεφώνου) σε διαφορετικούς τύπους οδών (αστικές / υπεραστικές) και κυκλοφοριακές συνθήκες (υψηλή / χαμηλή). Κάθε συμμετέχων στόχευε να ολοκληρώσει 12 διαφορετικές δοκιμές οδήγησης, ενώ σε κάθε δοκιμή, 2 μη αναμενόμενα περιστατικά είχαν προγραμματιστεί να συμβούν σε σταθερά σημεία κατά μήκος της οδήγησης. Αναπτύχθηκαν δύο μοντέλα γραμμικής παλινδρόμησης προκειμένου να εντοπιστούν διάφορα σύνολα μεταβλητών που συνάδουν με συγκεκριμένα μέτρα απόδοσης όπου η εξαρτημένη μεταβλητή του πρώτου μοντέλου είναι η διαφορά της μέσης ταχύτητας πριν και μετά το συμβάν και η εξαρτημένη μεταβλητή του δεύτερου μοντέλου είναι η διαφορά της τυπικής απόκλισης της ταχύτητας πριν και μετά το συμβάν. Τα αποτελέσματα έδειξαν ότι αφενός, οι οδηγοί μιλώντας στο κινητό παρουσιάζουν πιο προσεκτική συμπεριφορά μετά από ένα απροσδόκητο συμβάν, κάτι που οφείλεται στην αίσθηση απειλής της ασφάλειας που νιώθει ο οδηγός από την παρουσία του κινητού τηλεφώνου στο χέρι του, αφετέρου όταν συνομιλούν με συνεπιβάτη η προσοχή τους αποσπάται πιο εύκολα από το δρόμο και δεν αντιδρούν το ίδιο σωστά.

Την επίδραση πολλών παραμέτρων, όπως η οδηγική απόδοση, οι πηγές απόσπασης προσοχής, τα χαρακτηριστικά του οδηγού, καθώς και το οδικό και κυκλοφοριακό περιβάλλον, στην **πιθανότητα ατυχήματος** όταν συμβαίνουν **απροσδόκητα περιστατικά**, διερεύνησαν οι Papantoniou et al. (2019c) σε μια ακόμα έρευνά τους. Στο πλαίσιο αυτής, πραγματοποιήθηκε ένα πείραμα σε προσομοιωτή οδήγησης στο οποίο οι οδηγοί κλήθηκαν να οδηγήσουν σε διαφορετικούς τύπους απόσπασης της προσοχής και σε διαφορετικές οδικές και κυκλοφοριακές συνθήκες ώστε να αξιολογηθεί η απόδοσή τους. Συμπερασματικά, όπως ήταν αναμενόμενο, βρέθηκε πως η χρήση κινητού τηλεφώνου έχει αρνητική επίδραση στον κίνδυνο ατυχήματος καθώς οι οδηγοί δυσκολεύονται να αντιμετωπίσουν ένα απροσδόκητο συμβάν και ως εκ τούτου είναι πιο πιθανό να διαπράξουν ατύχημα.

Μέσω πειράματος προσομοιωτή οδήγησης οι X. Li et al. (2019) κατέγραψαν και εξέτασαν τις επιδόσεις της διαμήκουσ και πλευρικής αποφυγής σύγκρουσης 45 συμμετεχόντων σε τρία τυπικά σενάρια **σύγκρουσης** που ήταν σύγκρουση ορθής

γωνίας, μετωπική σύγκρουση και σύγκρουση με πεζούς. Από τα πειραματικά αποτελέσματα προέκυψε πως μόνο το φρένο ήταν η πιο κοινή απόκριση μεταξύ των τριών σεναρίων, ακολουθούμενο από φρένο σε συνδυασμό με στροβιλισμό σε σενάρια μετωπικής πρόσκρουσης και σύγκρουσης πεζών. Για το σενάριο σύγκρουσης ορθής γωνίας, στο οποίο ο χρόνος μέχρι τη σύγκρουση ήταν ο μεγαλύτερος, κανένας οδηγός δεν έκανε στραβοτιμονιά και οι οδηγοί με καθυστερημένη αντίδραση έτειναν να αντισταθμίσουν τον κίνδυνο σύγκρουσης λαμβάνοντας μεγαλύτερο μέγιστο ρυθμό επιβράδυνσης. Οι οδηγοί που στράφηκαν προς τη σύγκρουση (κυρίως στα σενάρια μετωπικής σύγκρουσης και σύγκρουσης με πεζούς) είχαν μικρότερο χρόνο αντίδρασης και παρουσιάστηκε καθυστέρηση στη δράση φρένων οδηγώντας τους σε υψηλή πιθανότητα σύγκρουσης.

Οι έρευνες που αναλύθηκαν παραπάνω τονίζουν πως η απόσπαση προσοχής σε συνδυασμό με την εμφάνιση ενός **απροσδόκητου συμβάντος** αυξάνει σε μεγάλο βαθμό τη δημιουργία ατυχήματος.

2.3. Σύνοψη

Στον παρακάτω πίνακα (Πίνακας 2.1) παρουσιάζεται μια σύνοψη των ερευνών που αναφέρθηκαν παραπάνω καθώς και οι ελλείψεις και οι προτάσεις για περαιτέρω έρευνα που παρατηρούνται σε κάθε μία από αυτές.

Πίνακας 2.1: Σύνοψη ερευνών

Έρευνα	Αντικείμενο	Σημαντικότερα αποτελέσματα	Ελλείψεις	Περαιτέρω έρευνα
Osman et al. (2019)	Εντοπισμός δευτερευουσών ενεργειών	αντίθετο μέτρο για την πρόληψη ατυχημάτων οι δευτερεύουσες ενέργειες		Εξέταση διαφορετικών σεναρίων που αφορούν: 1) τύπο οδού 2)γεωμετρικά χαρακτηριστικά 3)χαρακτηριστικά οχήματος

Έρευνα	Αντικείμενο	Σημαντικότερα αποτελέσματα	Ελλείψεις	Περαιτέρω έρευνα
Choudhary και Velaga (2019a)	Φαγητό-ποτό-γραφτά μηνύματα	αύξηση πιθανότητας ατυχήματος για φαγητό-ποτό, χαμηλός κίνδυνος ατυχήματος στην αποστολή γραπτών μηνυμάτων	Κυριαρχία νέων ανδρών στο δείγμα	Ανάλυση πιο περίπλοκων δευτερευουσών ενεργειών
Choudhary και Velaga (2019b)	Χρήση κινητού τηλεφώνου σε νέους και επαγγελματίες οδηγούς	Μεγαλύτερη επίπτωση στους νέους οδηγούς	Κυριαρχία αντρών στο δείγμα	Μελέτη πιο περίπλοκων σεναρίων
Yannis et al. (2014)	Επίδραση γραπτών μηνυμάτων σε νέους οδηγούς	μείωση της μέσης ταχύτητας και αύξηση του μέσου χρόνου αντίδρασης		Σύγκριση διαφορετικών παραγόντων απόσπασης προσοχής σε πιο πολύπλοκα οδικά περιβάλλοντα
Dimitriou et al. (2018)	Επίδραση γραπτών μηνυμάτων σε νέους οδηγούς	μείωση οδηγικής απόδοσης		Μελέτη διαφορετικών παραγόντων απόσπασης προσοχής σε άλλες συνθήκες κυκλοφορίας
Choudhary και Velaga (2018)	Επίδραση γραπτών μηνυμάτων κατά τη διάρκεια απρόσμενου συμβάντος	αύξηση κινδύνου ατυχήματος κυρίως στους νέους οδηγούς	Κυριαρχία αντρών στο δείγμα-μη δυνατή αξιολόγηση του κινδύνου σύγκρουσης σε σχέση με τη συχνότητα των μηνυμάτων	

Έρευνα	Αντικείμενο	Σημαντικότερα αποτελέσματα	Ελλείψεις	Περαιτέρω έρευνα
Choudhary και Velaga (2017)	Ομιλία και αποστολή γραπτών μηνυμάτων μέσω τηλεφώνου	1)μείωση μέσης ταχύτητας 2)αύξηση πιθανότητας ατυχήματος	Μικρό μέγεθος δείγματος ηλικιωμένων και γυναικών-μελέτη μόνο υπεραστικού περιβάλλοντος	Choudhary και Velaga (2017)
Papantoniou et al. (2019b)	Διερεύνηση της επίδρασης ενός απροσδόκητου συμβάντος στα οδηγικά μέτρα απόδοσης που σχετίζονται με την ταχύτητα	πιο προσεκτική συμπεριφορά μετά από ένα απροσδόκητο συμβάν	Ανάλυση της κατάστασης μόνο μετά το συμβάν	
Papantoniou et al. (2019c)	Επίδραση διαφόρων παραμέτρων στην πιθανότητα ατυχήματος όταν συμβαίνουν απροσδόκητα συμβάντα	αύξηση πιθανότητας ατυχήματος	Δεν εξετάζεται ο κίνδυνος ατυχήματος ως μεταβλητή	Μελέτη πολλών άλλων παραμέτρων
X. Li et al. (2019)	Τρόπος αποφυγής συγκρούσεων	1)φρένο η πιο κοινή απόκριση 2)μεγάλος χρόνος αντίδρασης και λάθος ελιγμοί αυξάνουν την πιθανότητα σύγκρουσης		Διερεύνηση του πώς οι στρατηγικές αποφυγής σύγκρουσης ποικίλλουν ανάλογα με τα χαρακτηριστικά των οδηγών

Έρευνα	Αντικείμενο	Σημαντικότεροι παράγοντες	Ελλείψεις	Περαιτέρω έρευνα
Monselise et al. (2019)	Παράγοντες που σχετίζονται με τον κίνδυνο ατυχήματος	η συμπεριφορά του οδηγού	Δεν γίνεται ανίχνευση επικίνδυνων συμπεριφορών σε πραγματικό χρόνο	
Papantoniou et al. (2019a)	Παράγοντες που οδηγούν σε λάθη	1) χαρακτηριστικά οδηγού 2) τύπος περιοχής	Δεν εξετάζεται ο κίνδυνος ατυχήματος ως μεταβλητή	

Από τη διεθνή βιβλιογραφία προκύπτουν οι εξής βασικές παρατηρήσεις:

- Η συμπεριφορά του οδηγού, οι ελιγμοί πριν από το συμβάν και η διάρκεια δευτερευουσών ενεργειών είναι σημαντικοί παράγοντες που μπορούν να οδηγήσουν σε **ατύχημα**.
- Η επίδραση των χαρακτηριστικών του οδηγού και του τύπου της περιοχής είναι οι μόνοι στατιστικά σημαντικοί παράγοντες που επηρεάζουν την πιθανότητα **οδηγικών σφαλμάτων**.
- Ο **κίνδυνος ατυχήματος** είναι χαμηλός κατά την αποστολή γραπτών μηνυμάτων, πιθανώς επειδή οι οδηγοί αντιλαμβάνονται τη συγκεκριμένη ενέργεια ως επικίνδυνη και προσαρμόζουν αναλόγως τη συμπεριφορά τους.
- Οι οδηγοί μιλώντας στο κινητό παρουσιάζουν πιο προσεκτική συμπεριφορά μετά από ένα **απροσδόκητο συμβάν**, κάτι που οφείλεται στην αίσθηση απειλής της ασφάλειας που νιώθει ο οδηγός από την παρουσία του κινητού τηλεφώνου στο χέρι του, αφετέρου όταν συνομιλούν με συνεπιβάτη η προσοχή τους αποσπάται πιο εύκολα από το δρόμο και δεν αντιδρούν το ίδιο σωστά.
- Η χρήση κινητού τηλεφώνου κατά την οδήγηση έχει μεγαλύτερη επίπτωση στην απόδοση των άπειρων νέων οδηγών χωρίς να υποβαθμίζει ωστόσο την επίδραση της και στους επαγγελματίες οδηγούς δείχνοντας ότι η υψηλότερη εμπειρία δεν μπορεί να εξουδετερώσει τις επιπτώσεις που δημιουργεί η **απόσπαση προσοχής** στην οδήγηση.
- Η αποστολή γραπτών μηνυμάτων οδηγεί σε σημαντική μείωση της μέσης ταχύτητας αλλά σε αύξηση του μέσου χρόνου αντίδρασης, άρα και σε αυξημένη **πιθανότητα ατυχήματος** σε συνδυασμό με την απόσπαση προσοχής. Πιο συγκεκριμένα, η πιθανότητα ατυχήματος αυξάνεται κατά 3 και 4 φορές αντίστοιχα όταν οι οδηγοί συνομιλούν ή στέλνουν μηνύματα στο τηλέφωνο.
- Η χρήση κινητού τηλεφώνου έχει αρνητική επίδραση στον κίνδυνο ατυχήματος καθώς οι οδηγοί δυσκολεύονται να αντιμετωπίσουν ένα **απροσδόκητο συμβάν** και ως εκ τούτου είναι πιο πιθανό να διαπράξουν ατύχημα.

Από την παραπάνω σύνοψη παρατηρείται πως έχει γίνει εκτενής έρευνα όσον αφορά το πώς επηρεάζεται η οδηγική συμπεριφορά και η πιθανότητα ατυχήματος από την απόσπαση προσοχής και την συμμετοχή του οδηγού σε δευτερεύουσες ενέργειες. Έχει αναλυθεί η επιρροή των απρόσμενων συμβάντων στα οδηγικά χαρακτηριστικά καθώς και η μεταβολή των χαρακτηριστικών αυτών πριν από ένα ατύχημα με τη χρήση διάφορων μοντέλων στατιστικής ανάλυσης.

Διαπιστώνεται όμως πως δεν υπάρχει επαρκής μελέτη των κρίσιμων παραγόντων που οδηγούν στον εντοπισμό συμβάντων τόσο πριν όσο και κατά τη διάρκεια ενός συμβάντος. Έτσι στην παρούσα διπλωματική εργασία επιχειρείται μέσω ανάλυσης κατάλληλων μεθοδολογιών ο προσδιορισμός των κρίσιμων παραγόντων εντοπισμού οδικών συμβάντων συγκεκριμένα σε αστικές περιοχές.

3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

3.1. Εισαγωγή

Σε αυτό το κεφάλαιο παρουσιάζεται το θεωρητικό υπόβαθρο σύμφωνα με το οποίο πραγματοποιείται η στατιστική ανάλυση της παρούσας εργασίας. Τα μοντέλα που αναπτύχθηκαν για τον εντοπισμό των συμβάντων και την επεξεργασία των μεταβλητών ήταν η διωνυμική λογιστική παλινδρόμηση και τα τυχαία δάση. Επίσης, πραγματοποιήθηκε ανάλυση παραγόντων για την διερεύνηση ύπαρξης κοινών παραγόντων ανάμεσα στις μεταβλητές και με χρήση της μεθόδου Boruta εκτιμήθηκε η σημαντικότητα ορισμένων από αυτές. Οι μεταβλητές που χρησιμοποιήθηκαν στην παρούσα ανάλυση αποτελούν οδηγικά χαρακτηριστικά μιας μεγάλης βάσης δεδομένων που συλλέχθηκαν με τη βοήθεια προσομοιωτή οδήγησης σε προηγούμενη έρευνα. Επίσης, γίνεται αναφορά στα κριτήρια αξιολόγησης και αποδοχής των συγκεκριμένων μοντέλων.

3.2. Μαθηματικά πρότυπα

3.2.1. Μοντέλο διωνυμικής λογιστικής παλινδρόμησης

Η διωνυμική λογιστική παλινδρόμηση (binomial logistic regression) (Cox 1958) είναι ένα στατιστικό μοντέλο που αναζητά τη σχέση μεταξύ μιας διακριτής εξαρτημένης μεταβλητής και μίας ή περισσότερων ανεξάρτητων μεταβλητών. Εξαρτημένη μεταβλητή είναι η μεταβλητή της οποίας γίνεται η πρόβλεψη των τιμών της και ανεξάρτητη η μεταβλητή η οποία είναι δεδομένη και χρησιμοποιείται για την πρόβλεψη της εξαρτημένης.

Η μορφή της εξίσωσης του μοντέλου είναι η εξής:

$$y_i = \text{logit}(P_i) = \ln \frac{P_i}{1-P_i} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni},$$

όπου

n : το πλήθος των ανεξάρτητων μεταβλητών

$\beta_1, \beta_2, \dots, \beta_n$: οι συντελεστές παλινδρόμησης των ανεξάρτητων μεταβλητών $X_1, X_2, \dots, X_n$

β_0 : ο σταθερός όρος της εξίσωσης

P_i : η προβλεπόμενη πιθανότητα, η οποία λαμβάνει τιμές 0 ή 1

Η εξαρτημένη μεταβλητή λαμβάνει την τιμή 1 με πιθανότητα επιτυχίας P και την τιμή 0 με πιθανότητα αποτυχίας $1-P$ και καλείται δυαδική (Binary) ή διωνυμική (Binomial) και το μοντέλο αντίστοιχα. Στη παρούσα εργασία η εξαρτημένη μεταβλητή (Συμβάν)

λαμβάνει την τιμή 1 για ύπαρξη συμβάντος και την τιμή 0 για μη ύπαρξη αντίστοιχα. Η δυαδική μορφή της εξαρτημένης μεταβλητής οδήγησε στην επιλογή του διωνυμικού λογιστικού μοντέλου στην παρούσα ανάλυση καθώς με τη βοήθεια των δεδομένων ανεξάρτητων μεταβλητών θα γίνεται ο εντοπισμός των συμβάντων.

3.2.2. Τυχαία Δάση

Τα τυχαία δάση (random forests Breiman 2001) είναι μια μέθοδος μάθησης για ταξινόμηση, παλινδρόμηση και άλλα καθήκοντα που λειτουργούν κατασκευάζοντας ένα πλήθος δέντρων αποφάσεων κατά το χρόνο εκπαίδευσης και εξάγοντας την τάξη ή τη μέση πρόβλεψη των μεμονωμένων δέντρων.

Ο αλγόριθμος των τυχαίων δασών λειτουργεί ως εξής:

- Επιλογή τυχαίων δειγμάτων από το σύνολο δεδομένων εκπαίδευσης
- Δημιουργία δέντρων αποφάσεων για κάθε δείγμα και λήψη αποτελέσματος πρόβλεψης από κάθε δέντρο
- Εκτέλεση ψηφοφορίας για κάθε προβλεπόμενο αποτέλεσμα
- Επιλογή του αποτελέσματος πρόβλεψης με τις περισσότερες ψήφους ως την τελική πρόβλεψη

Η μέθοδος των τυχαίων δασών επιλέχθηκε να χρησιμοποιηθεί στην παρούσα ανάλυση καθώς παρουσιάζει αρκετά πλεονεκτήματα όπως:

1. Μπορούν να παραλληλιστούν, που σημαίνει ότι υπάρχει η δυνατότητα διαχωρισμού της διαδικασίας σε πολλά μηχανήματα για εκτέλεση οδηγώντας σε μικρότερο χρόνο υπολογισμού.
2. Μπορούν να διαχειριστούν δεδομένα μεγάλων διαστάσεων
3. Διαθέτουν μεθόδους για την εξισορρόπηση σφαλμάτων σε μη ισορροπημένα σύνολα δεδομένων πληθυσμού.
4. Προσφέρουν μια μέθοδο για εργασία με δεδομένα που λείπουν. Οι τιμές που λείπουν αντικαθίστανται από τη μεταβλητή που εμφανίζεται περισσότερο σε έναν συγκεκριμένο κόμβο.
5. Παρέχουν την υψηλότερη ακρίβεια μεταξύ όλων των διαθέσιμων μεθόδων ταξινόμησης.
6. Μπορούν να χρησιμοποιηθούν για την επίλυση τόσο των προβλημάτων ταξινόμησης όσο και των προβλημάτων παλινδρόμησης.
7. Λειτουργούν καλά τόσο με κατηγορίες όσο και με συνεχείς μεταβλητές.
8. Αντιμετωπίζουν αποτελεσματικά τις μη γραμμικές παραμέτρους.
9. Είναι συνήθως ανθεκτικά στα ακραία σημεία και μπορούν να τα χειριστούν αυτόματα.
10. Ο αλγόριθμος τους είναι πολύ σταθερός. Ακόμα και αν εισαχθεί ένα νέο σημείο δεδομένων στο σύνολο δεδομένων, ο συνολικός αλγόριθμος δεν επηρεάζεται πολύ, καθώς τα νέα δεδομένα ενδέχεται να επηρεάσουν ένα δέντρο, αλλά είναι πολύ δύσκολο να επηρεάσει όλα τα δέντρα.

3.3. Κριτήρια αποδοχής μοντέλων

3.3.1. Βασικά κριτήρια ελέγχου λογιστικού μοντέλου

Τα βασικά κριτήρια ελέγχου για την αξιολόγηση και αποδοχή των μοντέλων περιγράφονται παρακάτω. Απαραίτητος είναι ο έλεγχος των συσχετίσεων των μεταβλητών δηλαδή οι ανεξάρτητες μεταβλητές να είναι γραμμικώς ανεξάρτητες μεταξύ τους.

- *Λογική εξήγηση των προσήμων των συντελεστών*

Μετά την δημιουργία της εξίσωσης του μοντέλου εξετάζεται αν το θετικό πρόσημο του β_i υποδηλώνει ότι αύξηση της ανεξάρτητης μεταβλητής οδηγεί σε αύξηση της εξαρτημένης και αντίθετα, το αρνητικό πρόσημο υποδηλώνει ότι μείωση της ανεξάρτητης μεταβλητής οδηγεί σε μείωση της εξαρτημένης. Επίσης, θα πρέπει να δίνεται μια λογική εξήγηση για την τιμή του συντελεστή, αφού αύξηση της ανεξάρτητης μεταβλητής x_i κατά μία μονάδα οδηγεί σε αύξηση της εξαρτημένης μεταβλητής κατά β_i .

- *Στατιστική σημαντικότητα*

Σημαντικός έλεγχος για την αξιολόγηση του λογιστικού μοντέλου είναι ο έλεγχος Wald test (z-test) βάσει του οποίου προσδιορίζεται η στατιστική σημαντικότητα των μεταβλητών ώστε να επιλεγθούν οι κατάλληλες. Ο τύπος είναι ο εξής:

$$z_i = \beta_i / s_{\beta_i},$$

όπου

β_i : οι συντελεστές παλινδρόμησης των ανεξάρτητων μεταβλητών x_i

s_{β_i} : το τυπικό σφάλμα των συντελεστών παλινδρόμησης β_i

3.3.2. Μήτρα σύγχυσης (Confusion Matrix)

Η μήτρα σύγχυσης είναι ένας τρόπος απεικόνισης της απόδοσης ενός μοντέλου πρόβλεψης σε μορφή πίνακα. Κάθε καταχώρηση σε πίνακα σύγχυσης υποδηλώνει τον αριθμό των προβλέψεων που έγιναν από το μοντέλο όπου ταξινομήσε τις τάξεις σωστά ή λανθασμένα. Κάθε σειρά του πίνακα αντιπροσωπεύει τις παρουσίες σε μια πραγματική κλάση, ενώ κάθε στήλη αντιπροσωπεύει τις παρουσίες σε μια προβλεπόμενη κλάση. Ένα πρόβλημα δυαδικής ταξινόμησης έχει μόνο δύο κατηγορίες για ταξινόμηση, κατά προτίμηση θετική και αρνητική τάξη. Στην παρούσα εργασία, ως θετική τάξη ορίστηκε η ύπαρξη συμβάντος και ως αρνητική η μη ύπαρξη, αντίστοιχα. Οι μετρήσεις της μήτρας σύγχυσης κατηγοριοποιούνται ως εξής:

1. True Positive (TP): Αναφέρεται στον αριθμό των προβλέψεων όπου ο ταξινομητής προβλέπει σωστά την ύπαρξη συμβάντος ως ύπαρξη συμβάντος.

2. True Negative (TN): Αναφέρεται στον αριθμό των προβλέψεων όπου ο ταξινομητής προβλέπει σωστά τη μη ύπαρξη συμβάντος ως μη ύπαρξη συμβάντος.
3. False Positive (FP): Αναφέρεται στον αριθμό των προβλέψεων όπου ο ταξινομητής προβλέπει εσφαλμένα τη μη ύπαρξη συμβάντος ως ύπαρξη συμβάντος.
4. False Negative (FN): Αναφέρεται στον αριθμό των προβλέψεων όπου ο ταξινομητής προβλέπει εσφαλμένα την ύπαρξη συμβάντος ως μη ύπαρξη συμβάντος.

Η μήτρα σύγχυσης έχει την μορφή που φαίνεται στον πίνακα 3.1:

Πίνακας 3.1: Παράδειγμα μορφής μήτρας σύγχυσης (Confusion matrix)

Πραγματική κλάση	Προβλεπόμενη κλάση	
	1 (Ύπαρξη συμβάντος)	0 (Μη ύπαρξη συμβάντος)
1 (Ύπαρξη συμβάντος)	True Positive	False Negative
0 (Μη ύπαρξη συμβάντος)	False Positive	True Negative

Μερικά από τα πιο κοινά μέτρα απόδοσης που μπορούν να χρησιμοποιηθούν από τη μήτρα σύγχυσης είναι:

- **Ορθότητα (Accuracy):** Δίνει τη συνολική ακρίβεια του μοντέλου δηλαδή το κλάσμα των συνολικών δειγμάτων που ταξινομούνται σωστά από τον ταξινομητή.

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

Εναλλακτικά της ορθότητας μπορεί να χρησιμοποιηθεί το **ποσοστό εσφαλμένης ταξινόμησης** (Misclassification Rate) που υπολογίζει ποιο μέρος των προβλέψεων είναι λανθασμένο.

$$\text{Misclassification Rate} = 1 - \text{accuracy}$$

- **Ακρίβεια (Precision):** Υπολογίζει ποιο μέρος των προβλέψεων ως θετική τάξη είναι πραγματικά θετικό.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Ανάκληση ή Ευαισθησία (Recall or Sensitivity):** Υπολογίζει ποιο κλάσμα όλων των θετικών δειγμάτων προβλέπονται σωστά ως θετικό από τον ταξινομητή.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **Εξειδικευτικότητα (Specificity):** Υπολογίζει ποιο κλάσμα όλων των αρνητικών δειγμάτων προβλέπονται σωστά ως αρνητικά από τον ταξινομητή.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- **Μέτρο F (F-measure):** Συνδυάζει την ακρίβεια και την ανάκληση σε ένα μόνο μέτρο, μαθηματικά είναι το αρμονικό μέσο ακρίβειας και ανάκλησης.

$$F\text{-measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Δείκτης λάθος συναγερμού (False alarm rate):** Υπολογίζεται ως ο αριθμός των λανθασμένων θετικών προβλέψεων δια του συνολικού αριθμού των αρνητικών.

$$\text{False alarm rate} = \frac{FP}{TN + FP}$$

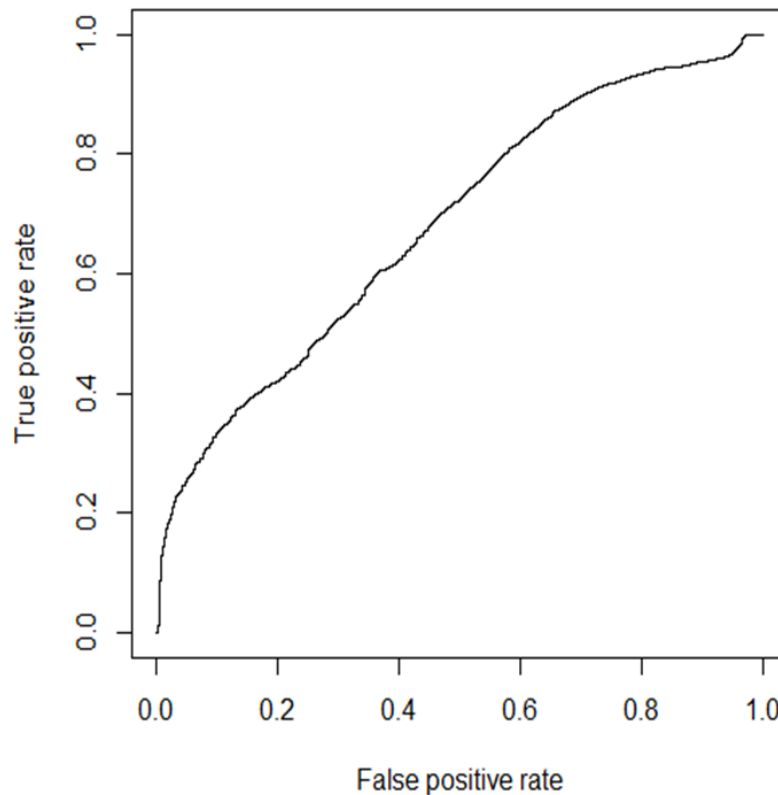
- **Χαρακτηριστική Καμπύλη Λειτουργίας Δέκτη (Receiver Operating Characteristic Curve - ROC Curve):** Η καμπύλη ROC είναι η γραφική παράσταση του αληθινού θετικού ποσοστού έναντι του ψευδούς θετικού ποσοστού για ένα δυαδικό σύστημα ταξινομητή καθώς το όριο της διάκρισής του είναι ποικίλο.

Ο κατακόρυφος άξονας της καμπύλης απεικονίζει το TP (=sensitivity) ενώ ο οριζόντιος το FP ποσοστό (1 - specificity).

Η ορθότητα (accuracy) υπολογίζεται ως το εμβαδόν κάτω από την καμπύλη ROC (Area Under the Curve – AUC)

- AUC = 1: Πρόκειται για ιδανικό μοντέλο
- AUC = 0.5: Πρόκειται για τυχαία πρόβλεψη

Ένα μοντέλο είναι αυστηρά καλύτερο αν έχει μεγαλύτερη περιοχή κάτω από την καμπύλη συνεπώς οι καλοί ταξινομητές βρίσκονται κοντά στην αριστερή πάνω γωνία του διαγράμματος.



Σχήμα 3.1 : Παράδειγμα χαρακτηριστικής καμπύλης ROC

3.4. Ανάλυση παραγόντων (Factor Analysis)

Η ανάλυση παραγόντων (Spearman 1904) είναι ένα γραμμικό στατιστικό μοντέλο που χρησιμοποιείται για να εξηγήσει τη διακύμανση μεταξύ των παρατηρούμενων μεταβλητών και να συμπυκνώσει ένα σύνολό τους σε μη παρατηρούμενες μεταβλητές που ονομάζονται παράγοντες.

Τα στάδια της ανάλυσης παραγόντων είναι τα εξής:

1. Επιλογή ενός σετ μεταβλητών
2. Δημιουργία του πίνακα συσχετίσεων (Correlation matrix)
3. Εξαγωγή παραγόντων
4. Περιστροφή παραγόντων
5. Ερμηνεία νοήματος παραγόντων

Οι μεταβλητές που χρησιμοποιούνται στην ανάλυση παραγόντων πρέπει να καλύπτουν κάποιες βασικές προϋποθέσεις:

- Να είναι συνεχείς
- Να υπάρχει επαρκής συσχέτιση μεταξύ τους ($0.2 < \text{correlation} < 0.8$)
- Το ηλίκο [μέγεθος δείγματος / αριθμός μεταβλητών] πρέπει να ξεπερνά οπωσδήποτε το 5 και στην περίπτωση που οι μεταβλητές αποκλίνουν σοβαρά από την κανονική κατανομή πρέπει να ξεπερνά το 10
- Έλεγχος των μεταβλητών για πιθανή παρουσία ακραίων τιμών
- Ο αριθμός τους να είναι αρκετά μεγάλος ώστε να περιλαμβάνονται τουλάχιστον 3 μεταβλητές σε κάθε παράγοντα
- Το μέγεθος του δείγματος να είναι τουλάχιστον ίσο με το τριπλάσιο του αριθμού των μεταβλητών (δείγμα > 200 θεωρείται ικανοποιητικό)

Παρέχονται δύο δείκτες για τον έλεγχο της ποιότητας των δεδομένων:

- Ο Δείκτης **Keiser-Meyer-Olkin** αξιολογεί την επάρκεια του δείγματος (**KMO > 0.50**)
- Ο Δείκτης **Bartlett's Test of Sphericity** αξιολογεί το κατά πόσο οι συσχετίσεις μεταξύ των μεταβλητών επιτρέπουν την εφαρμογή της ανάλυσης παραγόντων (**p < 0.05**)

Οι πλέον διαδεδομένες μέθοδοι εξαγωγής παραγόντων είναι η **ανάλυση παραγόντων (Factor Analysis)** και η **ανάλυση σε κύριες συνιστώσες (Principal Components Analysis)**. Η ανάλυση παραγόντων είναι καλύτερη μέθοδος όταν σκοπός είναι η κατασκευή παραγόντων καθώς στοχεύει στη μελέτη μόνο του ποσοστού διακύμανσης που έχουν κοινό οι μεταβλητές που μελετώνται. Αντιθέτως, η ανάλυση κυρίων συνιστωσών προτιμάται όταν επιχειρείται η μείωση του αριθμού των μεταβλητών καθώς στόχος της είναι η μελέτη όλης της υπάρχουσας διακύμανσης ώστε να εξαχθεί το μεγαλύτερο ποσοστό της διακύμανσης από τους λιγότερο δυνατούς παράγοντες.

3.5. Σημαντικότητα μεταβλητών (μέθοδος Boruta)

Η μέθοδος **Boruta** (2010) είναι μία μέθοδος που εντοπίζει όλες τις σημαντικές ανεξάρτητες μεταβλητές που μπορεί να υπάρχουν σε ένα σύνολο δεδομένων σε σχέση με μια εξαρτημένη μεταβλητή. Ο αλγόριθμος αυτής της μεθόδου λειτουργεί με την εξής διαδικασία:

1. Πρώτον, προσθέτει τυχαιότητα στο σύνολο δεδομένων δημιουργώντας ανακατεμένα αντίγραφα όλων των χαρακτηριστικών (τα οποία ονομάζονται χαρακτηριστικά σκιάς).
2. Στη συνέχεια, εκπαιδεύει έναν τυχαίο ταξινομητή, όπως των τυχαίων δασών, στο εκτεταμένο σύνολο δεδομένων και εφαρμόζει ένα μέτρο σπουδαιότητας χαρακτηριστικών (η προεπιλογή είναι η μέση μείωση της ακρίβειας) για να

αξιολογήσει τη σημασία κάθε μεταβλητής όπου το υψηλότερο σημαίνει πιο σημαντικό.

3. Σε κάθε επανάληψη, ελέγχει εάν ένα πραγματικό χαρακτηριστικό έχει μεγαλύτερη σημασία από τα καλύτερα χαρακτηριστικά σκιάς του (δηλαδή εάν το χαρακτηριστικό έχει υψηλότερη βαθμολογία από τη μέγιστη βαθμολογία των σκιαγραφικών χαρακτηριστικών του) και αφαιρεί συνεχώς χαρακτηριστικά που θεωρούνται εξαιρετικά ασήμαντα .
4. Τέλος, ο αλγόριθμος σταματά είτε όταν επιβεβαιωθούν ή απορριφθούν όλες οι μεταβλητές είτε φτάσει σε ένα καθορισμένο όριο εκτελέσεων του ταξινομητή.

Σε αυτό το κεφάλαιο παρουσιάστηκε το θεωρητικό υπόβαθρο στο οποίο βασίστηκε η παρούσα εργασία και σε επόμενο θα παρουσιαστεί η εφαρμογή όλων αυτών των μεθοδολογιών με στόχο τον εντοπισμό των συμβάντων.

4. ΣΥΛΛΟΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΣΤΟΙΧΕΙΩΝ

4.1. Εισαγωγή

Ο στόχος της διπλωματικής εργασίας, όπως προέκυψε από τα παραπάνω, είναι ο προσδιορισμός των κρίσιμων παραγόντων εντοπισμού οδικών συμβάντων σε αστικές περιοχές.

Σε αυτό το κεφάλαιο παρουσιάζεται η συλλογή και ο τρόπος επεξεργασίας των στοιχείων που διαμόρφωσαν την τελική βάση δεδομένων σύμφωνα με την οποία έγινε η ανάλυση της παρούσας έρευνας.

4.2. Συλλογή στοιχείων

Η συλλογή των απαραίτητων στοιχείων έγινε από μεγάλη βάση δεδομένων που προέκυψε μετά από πείραμα σε προσομοιωτή οδήγησης για την εκπόνηση μιας διδακτορικής διατριβής.

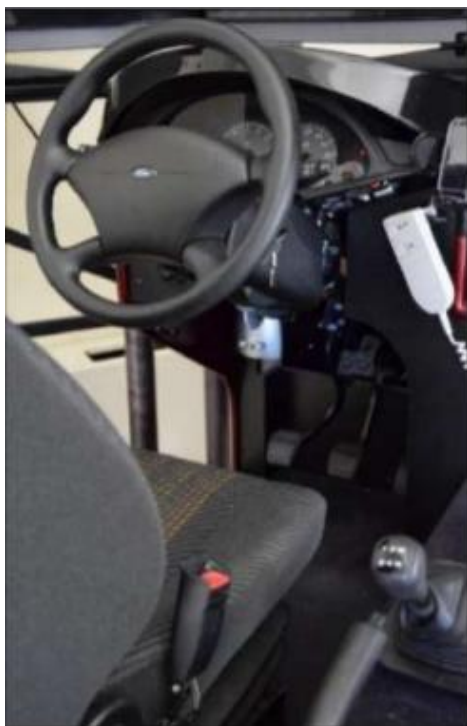
Η πειραματική διαδικασία διεξήχθη στον προσομοιωτή οδήγησης που διαθέτει το Εργαστήριο Κυκλοφοριακής Τεχνικής του Τομέα Μεταφορών και Συγκοινωνιακής Υποδομής της Σχολής Πολιτικών Μηχανικών Ε.Μ.Π., ο οποίος έχει κατασκευαστεί από την εταιρεία Foerst προκειμένου να εξυπηρετήσει ερευνητικούς σκοπούς.

Όπως φαίνεται και στην εικόνα 4.1 ο προσομοιωτής αποτελείται από τρεις οθόνες LCD 40 ιντσών (full HD), συνολική γωνία προβολής 170 μοίρες, θέση οδήγησης και βάση στήριξης. Οι διαστάσεις σε πλήρη ανάπτυξη είναι 230x180 cm με πλάτος βάσης 78 cm.



Εικόνα 4.1: Προσομοιωτής οδήγησης

Στην εικόνα 4.2 απεικονίζεται η θέση οδήγησης του προσομοιωτή, ο οποίος διαθέτει ρυθμιζόμενο κάθισμα οδηγού, τιμόνι διαμέτρου 27 εκατοστών, πεντάλ (γκάζι, φρένο, συμπλέκτης), ταμπλό και δύο εξωτερικούς και έναν κεντρικό καθρέφτη που εμφανίζονται στο πλάι και στην κύρια οθόνη, και εμφανίζονται σε πραγματικό χρόνο αντικείμενα και συμβάντα που συμβαίνουν πίσω από το «όχημα». Τα χειριστήρια που διαθέτει ο οδηγός είναι: μοχλός πέντε ταχυτήτων συν όπισθεν, φλας, υαλοκαθαριστήρες, φώτα, κόρνα, χειρόφρενο και μίζα.



Εικόνα 4.2: Θέση οδήγησης προσομοιωτή

Οι μετρήσεις που γίνονται από τον προσομοιωτή οδήγησης καταγράφονται σε χρονικά διαστήματα των 17-17 ms, δηλαδή περίπου 60 το δευτερόλεπτο. Τα δεδομένα εξάγονται από το πρόγραμμα μέσω αρχείου μορφής κειμένου το οποίο περιλαμβάνει τις τιμές πολλών μεταβλητών, οι οποίες είναι σημαντικές για την ανάλυση και παρουσιάζονται στον Πίνακα 4.1.

Πίνακας 4.1: Μεταβλητές προσομοιωτή οδήγησης

ΜΕΤΑΒΛΗΤΕΣ	ΕΠΕΞΗΓΗΣΗ
Time	πραγματικός χρόνος από την έναρξη της οδήγησης σε ms
x-pos	x θέση του οχήματος σε m
y-pos	y θέση του οχήματος σε m
z-pos	z θέση του οχήματος σε m
road	νούμερο οδού που βρίσκεται το όχημα [int]
richt	κατεύθυνση του οχήματος στην οδό σε [BOOL] (0/1)
rdist	διανυόμενη απόσταση οχήματος από την αρχή της οδήγησης σε m
rspur	απόκλιση οχήματος από το μέσο της οδού σε m
ralpha	κατεύθυνση του οχήματος σε σχέση με την κατεύθυνση της οδού σε μοίρες
Dist	οδηγημένη πορεία από την έναρξη της πορείας σε m
Speed	Πραγματική ταχύτητα σε km/h
Brk	θέση πεντάλ φρένου σε ποσοστό επί τοις εκατό
Acc	θέση πεντάλ γκαζιού σε ποσοστό επί τοις εκατό
Clutch	θέση πεντάλ συμπλέκτη σε ποσοστό επί τοις εκατό
Gear	επιλεγμένη ταχύτητα (0 = αδράνεια, 6 = όπισθεν)
RPM	περιστροφή κινητήρα σε 1/min
HWay	απόσταση από το προπορευόμενο όχημα σε m
Dleft	απόσταση από την αριστερή οριογραμμή σε m
DRight	απόσταση από την δεξιά οριογραμμή σε m
Wheel	θέση τιμονιού σε μοίρες
THead	χρόνος μέχρι τη σύγκρουση με το προπορευόμενο όχημα σε sec
TTL	χρόνος προς τη διασταύρωση έως ότου ξεπεραστεί η γραμμή του ορίου σε sec
TTC	χρόνος μέχρι τη σύγκρουση (όλα τα εμπόδια) σε sec
AccLat	πλευρική επιτάχυνση σε m/s ²
AccLon	διαμήκης επιτάχυνση σε m/s ²
EvVis	συμβάν-ορατό-σημαία/συμβάν-ένδειξη, 0 = όχι συμβάν , 1 = συμβάν
EvDist	συμβάν-απόσταση σε m
Err1No	αριθμός του πιο σημαντικού οδηγικού λάθους από το τελευταίο σύνολο δεδομένων
Err1Val	στιγμή του λάθους, το περιεχόμενο διαφέρει ανάλογα με τον τύπο του λάθους
Err2No	αριθμός του επόμενου οδηγικού λαθους (πιθανόν κενό)
Err2Val	πρόσθετη στιγμή του λάθους 2
Err3No	αριθμός περαιτέρω λάθους οδήγησης (ίσως κενό)
Err3Val	πρόσθετη στιγμή του λάθους 3

Από τις παραπάνω μεταβλητές, αποφασίστηκε πως χρήσιμες για ανάλυση στην παρούσα έρευνα είναι αυτές που παρουσιάζονται στον πίνακα 4.2:

Πίνακας 4.2: Μεταβλητές που χρησιμοποιήθηκαν στην εργασία

ΜΕΤΑΒΛΗΤΕΣ	ΕΠΕΞΗΓΗΣΗ
Time	Πραγματικός χρόνος από την έναρξη της οδήγησης σε ms
rdist	Διανυόμενη απόσταση από την αρχή της οδήγησης σε m
rspur	Απόκλιση οχήματος από το μέσο της οδού σε m
Speed	Πραγματική ταχύτητα σε km/h
HWay	Απόσταση από το προπορευόμενο όχημα σε m
Dleft	Απόσταση από την αριστερή οριογραμμή σε m
DRight	Απόσταση από την δεξιά οριογραμμή σε m
Wheel	Θέση τιμονιού σε μοίρες
THead	Χρόνος μέχρι τη σύγκρουση με το προπορευόμενο όχημα σε sec
TTC	Χρόνος μέχρι τη σύγκρουση (όλα τα εμπόδια) σε sec
AccLat	Πλευρική επιτάχυνση σε m/s^2
AccLon	Διαμήκης επιτάχυνση σε m/s^2

Στον πίνακα 4.3 παρουσιάζονται οι μεταβλητές που αφορούν στα χαρακτηριστικά κάθε οδηγού και συλλέχθηκαν με τη βοήθεια ερωτηματολογίων.

Πίνακας 4.3: Μεταβλητές χαρακτηριστικών οδηγού

VARIABLES	ΜΕΤΑΒΛΗΤΕΣ	ΠΙΘΑΝΕΣ ΤΙΜΕΣ
PersonID	Κωδικός οδηγού	D0i
Age	Ηλικία	Έτη ζωής
AgeGroup	Ηλικιακή ομάδα	Νέος/Μεσήλικας/Ηλικιωμένος
Gender	Φύλο	Άνδρας/Γυναίκα
Education	Εκπαίδευση	Έτη εκπαίδευσης
Driving Experience	Οδηγική εμπειρία	Έτη οδηγικής εμπειρίας
Disease	Ασθένεια	Υγείας

Τέλος, στον πίνακα 4.4 παρουσιάζονται οι μεταβλητές που προέκυψαν από τα διαφορετικά σενάρια που εκτέλεσαν οι οδηγοί. Οι μεταβλητές αυτές αφορούσαν στον κυκλοφοριακό φόρτο (υψηλός/χαμηλός), στην απόσπαση προσοχής (όχι απόσπαση, χρήση κινητού τηλεφώνου, συνομιλία με συνεπιβάτη) καθώς και στην ύπαρξη ή όχι απρόσμενου συμβάντος.

Πίνακας 4.4: Μεταβλητές διαφορετικών σεναρίων

VARIABLES	ΜΕΤΑΒΛΗΤΕΣ	ΠΙΘΑΝΕΣ ΤΙΜΕΣ
Trial	Αριθμός διαδρομής	1,2, ...
Traffic	Κυκλοφοριακός φόρτος	Υψηλός/χαμηλός
Distractor	Απόσπαση	Όχι/κινητό/συνομιλία
Event	Συμβάν (ανάλογα με το είδος)	0 ή αριθμός≠0
State	Συνθήκη	Μηδενική ταχύτητα/όχι συμβάν/συμβάν

4.3. Βάση δεδομένων

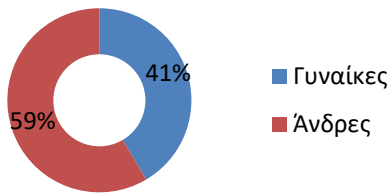
Από τα παραπάνω στοιχεία δημιουργήθηκε η βάση δεδομένων **UrbanControl** που χρησιμοποιήθηκε στην παρούσα έρευνα. Η συγκεκριμένη βάση δεδομένων περιέχει στοιχεία 61 υγιών οδηγών και αφορά μόνο σε αστικές περιοχές. Στον πίνακα 4.5 εμφανίζεται ένα απόσπασμα της βάσης δεδομένων η οποία στο σύνολό της αποτελείται από 1,359,816 σειρές και 27 στήλες (μεταβλητές) δηλαδή από 36,715,032 παρατηρήσεις.

Πίνακας 4.5: Απόσπασμα βάσης δεδομένων UrbanControl (15/1,359,816 σειρές και 10/27 στήλες)

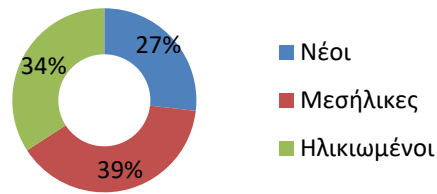
X1	PersonID	Trial	Time	Distractor	Traffic_x	Speed	AccLat	AccLon	Event
55238	D003	1	00:00:00	NO	QL	0	0	0	0
55242	D003	1	00:00:00	NO	QL	0	0	0	0
55254	D003	1	00:00:00	NO	QL	0	0	0	0
55258	D003	1	00:00:00	NO	QL	0	0	0	0
55260	D003	1	00:00:00	NO	QL	0	0	0	0
55282	D003	1	00:00:00	NO	QL	0	0	0	0
55284	D003	1	00:00:00	NO	QL	0	0	0	0
55286	D003	1	00:00:00	NO	QL	0	0	0	0
55288	D003	1	00:00:01	NO	QL	0	0	0	0
55290	D003	1	00:00:01	NO	QL	0	0	0	0
55292	D003	1	00:00:01	NO	QL	0	0	0	0
55294	D003	1	00:00:01	NO	QL	0	0	0	0
55296	D003	1	00:00:01	NO	QL	0	0	0	0
55298	D003	1	00:00:01	NO	QL	0	0	0	0
55300	D003	1	00:00:01	NO	QL	0	0	0	0

4.4. Επεξεργασία στοιχείων

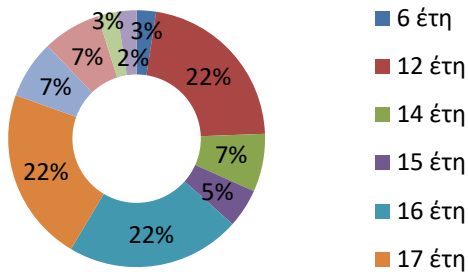
Η παραπάνω βάση δεδομένων με τη βοήθεια της γλώσσας προγραμματισμού R διαχωρίστηκε σε τρεις επιμέρους πίνακες προκειμένου να είναι δυνατός ο εντοπισμός των συμβάντων με βάση τα οδηγικά χαρακτηριστικά. Δημιουργήθηκε πίνακας που περιέχει τα στοιχεία **κατά τη διάρκεια** των συμβάντων, πίνακας που περιέχει στοιχεία **ενός λεπτού πριν** από το συμβάν και ένας **συνολικός** πίνακας που περιέχει το άθροισμα των δύο παραπάνω. Σε αυτούς τους πίνακες παρατηρήθηκε πως στις μεταβλητές AccLon, AccLat και Driving Experience έλειπαν κάποια στοιχεία (τιμές N/A) και έτσι κρίθηκε σωστό να αφαιρεθούν αυτές οι γραμμές. Μετά την αφαίρεση έμειναν τα στοιχεία 41 οδηγών των οποίων τα κοινωνικά χαρακτηριστικά (φύλο, ηλικιακή ομάδα, εκπαίδευση) καθώς και η οδηγική εμπειρία απεικονίζονται στην εικόνα 4.3.

Φύλο

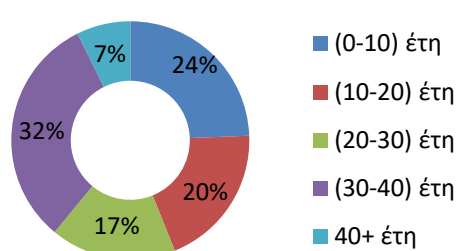
(α)

Ηλικιακή ομάδα

(β)

Εκπαίδευση

(γ)

Οδηγική εμπειρία

(δ)

Εικόνα 4.3: Κατανομή (α) Φύλου, (β) Ηλικιακής ομάδας, (γ) Εκπαίδευσης και (δ) Οδηγικής εμπειρίας συμμετεχόντων

Ο διαχωρισμός των πινάκων έγινε με βάση την μεταβλητή **Event**, η οποία όταν έχει τιμή 0 δείχνει πως δεν υπάρχει συμβάν και όταν έχει τιμή οποιοδήποτε αριθμό διάφορο του 0 δείχνει πως υπάρχει συμβάν.

Για τον διαχωρισμό των πινάκων επίσης, δημιουργήθηκε ο πίνακας **index** ο οποίος καταγράφει τους χρόνους αρχής και τέλους κάθε συμβάντος. Ο πίνακας αυτός αποτελείται από τρεις στήλες, τον κωδικό του οδηγού, τον χρόνο αρχής και τον χρόνο λήξης κάθε συμβάντος και από 656 γραμμές, που αποτελούν τον συνολικό αριθμό των συμβάντων. Στον πίνακα 4.6 παρουσιάζεται ένα απόσπασμα του πίνακα **index** όπου ο χρόνος είναι σε h:m:s:ms.

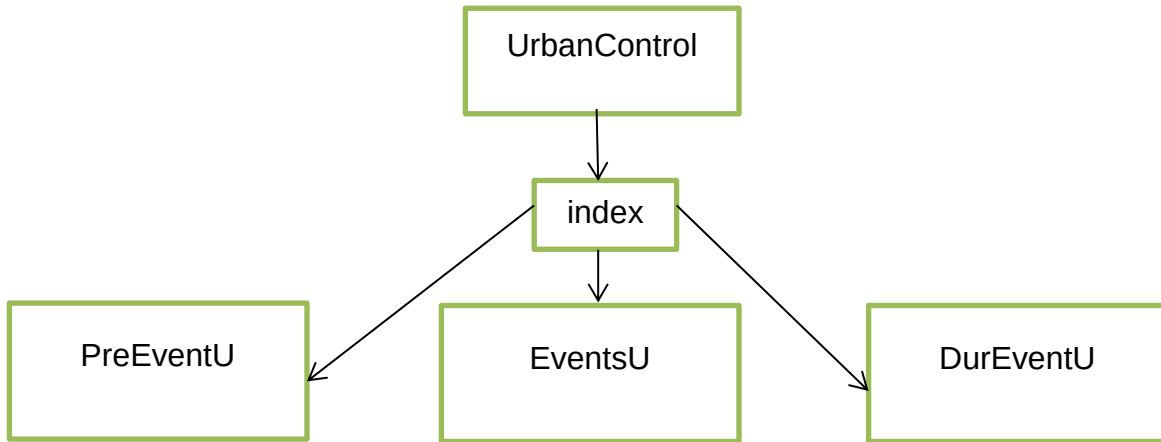
Πίνακας 4.6: Απόσπασμα πίνακα index (15/656 σειρές)

PersonID	Start_Event Time	End_Event Time
D003	00:01:16.049	00:01:38.051
D003	00:02:22.388	00:02:35.055
D003	00:02:36.989	00:02:39.923
D003	00:06:45.941	00:07:13.543
D003	00:07:58.646	00:08:33.581
D003	00:11:03.628	00:11:17.528
D003	00:13:32.971	00:13:46.721
D003	00:17:41.156	00:18:26.644
D003	00:19:15.463	00:19:40.048
D010	00:01:11.465	00:01:29.768
D010	00:02:11.371	00:02:26.906
D010	00:02:29.506	00:02:32.038
D010	00:05:52.237	00:06:13.389
D010	00:06:52.225	00:07:24.093
D010	00:09:51.205	00:10:04.004

Με τη βοήθεια του πίνακα index και κατάλληλου κώδικα δημιουργήθηκαν οι τρεις παρακάτω πίνακες:

- **PreEventU:** Ο πίνακας PreEventU περιέχει τα στοιχεία που αντιστοιχούν σε ένα λεπτό πριν από κάθε συμβάν, δηλαδή αποτελείται από όλες τις στήλες της αρχικής βάσης δεδομένων και από τις γραμμές που αντιστοιχούν στις χρονικές στιγμές μεταξύ της έναρξης ενός συμβάντος μείον 60 δευτερόλεπτα και της έναρξης του συμβάντος, για κάθε ένα από αυτά.
- **DurEventU:** Ο πίνακας DurEventU περιέχει τα στοιχεία που αντιστοιχούν στη διάρκεια κάθε συμβάντος, δηλαδή αποτελείται από όλες τις στήλες της αρχικής βάσης δεδομένων και από τις γραμμές που αντιστοιχούν στις χρονικές στιγμές μεταξύ της έναρξης ενός συμβάντος και της λήξης του συμβάντος, για κάθε ένα από αυτά.
- **EventsU:** Ο πίνακας EventsU περιέχει το άθροισμα των παραπάνω, δηλαδή αποτελείται από όλες τις στήλες της αρχικής βάσης δεδομένων και από τις γραμμές που αντιστοιχούν στις χρονικές στιγμές μεταξύ της έναρξης ενός συμβάντος μείον 60 δευτερόλεπτα και της λήξης του συμβάντος, για κάθε ένα από αυτά.
Στον πίνακα EventsU κάποια στοιχεία επαναλαμβάνοντουσαν καθώς συνέπεφε το ένα λεπτό πριν την έναρξη του συμβάντος με τη διάρκεια του προηγούμενου. Κρίθηκε σωστό αυτές οι τιμές να διαγραφούν.

Η διαδικασία διαχωρισμού των πινάκων φαίνεται στο Γράφημα 4.1.



Γράφημα 4.1: Δημιουργία πινάκων PreEventU, DurEventU και EventsU

Στον πίνακα 4.7 παρουσιάζεται ένα απόσπασμα του πίνακα EventsU. Οι μορφές των πινάκων PreEventU και DurEventU είναι παρόμοιες.

Πίνακας 4.7: Απόσπασμα πίνακα EventsU (15/438,126 σειρές και 10/27 στήλες)

X1	PersonID	Trial	Time	Distractor	Traffic_x	Speed	AccLat	AccLon	Event
56170	D003	1	00:00:16	NO	QL	18.4	0.001	0.384	0
56172	D003	1	00:00:16	NO	QL	18.5	0.002	0.438	0
56174	D003	1	00:00:16	NO	QL	18.5	0.002	0.469	0
56176	D003	1	00:00:16	NO	QL	18.6	0.002	0.495	0
56178	D003	1	00:00:16	NO	QL	18.6	0.002	0.527	0
56180	D003	1	00:00:16	NO	QL	18.7	0.002	0.55	0
56182	D003	1	00:00:16	NO	QL	18.8	0.002	0.571	0
56184	D003	1	00:00:16	NO	QL	18.9	0.003	0.6	0
56186	D003	1	00:00:16	NO	QL	18.9	0.003	0.621	0
56188	D003	1	00:00:16	NO	QL	19	0.003	0.642	0
56190	D003	1	00:00:16	NO	QL	19	0.001	0.273	0
56192	D003	1	00:00:16	NO	QL	19	0	0.002	0
56194	D003	1	00:00:16	NO	QL	19	0	-0.052	0
56196	D003	1	00:00:16	NO	QL	19	0	0.031	0
56198	D003	1	00:00:16	NO	QL	19.1	0.001	0.169	0

4.5. Περιγραφική Στατιστική

Στη συνέχεια, με τη βοήθεια κατάλληλου κώδικα στην R πραγματοποιήθηκε **περιγραφική στατιστική** στους πίνακες PreEventU και DurEventU. Απομονώθηκαν τα αποτελέσματα των συνεχών μεταβλητών που θεωρήθηκαν σημαντικές και άξιες ανάλυσης (Speed, AccLat, AccLon, Hway, Thead, TTC, Dleft, Dright, rdist, rspur, Wheel, Age και Driving Experience). Παρακάτω παρουσιάζονται στους πίνακες 4.8 και

4.9 τα αποτελέσματα της περιγραφικής στατιστικής των πινάκων PreEventU και DurEventU αντίστοιχα.

Πίνακας 4.8: Πίνακας περιγραφικής στατιστικής μεταβλητών PreEventU

	Pre Event (334989)				
	Mean	Standard Deviation	Variance	Max	Min
Speed (km/h)	25.52	15.745	247.900	85.6	0
AccLat (m/s ²)	-0.00901	0.245	0.060	46.079	-63.076
AccLon (m/s ²)	0.05486	1.393	1.941	110.386	-92.203
Hway (m)	3068.4	4550.167	20704020.000	290.7	0
Thead (s)	3931.1	4858.800	23607940.000	9988.8	0.2
TTC (s)	9235.9	2654.460	7046156.000	29.9	0.8
Dleft (m)	1058.94	3074.158	9450447.000	7.72	-3.46
Drigh (m)	1059.07	3074.114	9450178.000	9.55	-0.38
rdist (m)	33.02	27.590	761.214	100	0
rspur (m)	2.758	1.764	3.110	8.7	-8.7
Wheel (degrees)	3.379	25.371	643.665	540	-368
Age (years)	42.6	15.635	244.460	78	22
Driving Experience (years)	20.2	13.200	174.234	46	3

Πίνακας 4.9: Πίνακας περιγραφικής στατιστικής μεταβλητών DurEventU

	During Event (103137)				
	Mean	Standard Deviation	Variance	Max	Min
Speed (km/h)	26.42	16.588	275.166	80.3	0
AccLat (m/s ²)	0.00513	0.250	0.063	54	-7.132
AccLon (m/s ²)	0.0533	2.840	8.064	231.406	-195.662
Hway (m)	3804.8	4798.036	23021150.000	293.2	1.4
Thead (s)	4614.5	4955.611	24558080.000	9972.7	0.3
TTC (s)	9042	2941.578	8652883.000	29.9	0.4
Dleft (m)	640.26	2443.133	5968897.000	7.05	-3.46
Drigh (m)	641.14	2442.905	5967782.000	9.7	-0.06
rdist (m)	31.3	26.108	681.643	100	0
rspur (m)	3.336	1.805	3.257	8.68	-4
Wheel (degrees)	14.56	29.932	895.939	412	-286
Age (years)	43.92	16.204	262.585	78	22
Driving Experience (years)	20.97	13.521	182.824	46	3

Συγκρίνοντας τους δύο παραπάνω πίνακες μπορούν να παρατηρηθούν τα εξής:

- Κατά τη διάρκεια του συμβάντος αυξήθηκαν η τυπική απόκλιση και η διακύμανση της ταχύτητας (Speed) και της διαμήκους επιτάχυνσης (AccLon).
- Αυξήθηκε η μέση τιμή της πλευρικής επιτάχυνσης (AccLat), της απόστασης από το προπορευόμενο όχημα (Hway) και της χρονικής απόστασης από το προπορευόμενο όχημα (Thead).

- Μειώθηκε σημαντικά η διακύμανση της απόστασης τόσο από την δεξιά όσο και από την αριστερή οριογραμμή (Dright, Dleft).

4.6. Συσχέτιση μεταβλητών

Στη συνέχεια, από τον πίνακα EventsU απομονώνονται οι μεταβλητές (στήλες) που κρίθηκαν σημαντικές για ανάλυση (Speed, AccLat, AccLon, Hway, Thead, Dleft, Dright, rdist, rspur, Wheel, Age και Driving Experience) και δημιουργείται ένας πίνακας μόνο με αυτές, στον οποίο προσδιορίστηκε με κατάλληλες εντολές η **συσχέτιση** (correlation) μεταξύ των ανεξάρτητων μεταβλητών. Η μεταβλητή TTC αφαιρέθηκε καθώς έχει παρόμοια ερμηνεία με την μεταβλητή Thead.

Με τη βοήθεια της συσχέτισης βρέθηκαν οι μεταβλητές που δεν είναι υψηλά συσχετισμένες μεταξύ τους (<0.8) και αυτές χρησιμοποιήθηκαν στα μοντέλα στατιστικής ανάλυσης. Στον πίνακα 4.10 παρουσιάζεται η συσχέτιση μεταξύ των ανεξάρτητων μεταβλητών.

Πίνακας 4.10: Πίνακας συσχέτισης (Correlation matrix)

	Speed	AccLat	AccLon	HWay	Thead	DLeft	DRight	rdist	rspur	Wheel	Age	Driving experience
Speed	1	-0.00494	-0.011831	0.0733106	-0.227506	0.0081567	0.0082397	0.2562799	0.0141984	0.0626729	-0.102713077	-0.062015
AccLat	-0.00494	1	0.1210205	-0.03364	-0.026466	-0.080206	-0.080203	0.013857	0.0785	0.0382492	-0.000168036	-0.00369
AccLon	-0.011831	0.1210205	1	0.0033846	-0.001752	0.019228	0.0192369	-0.03049	-0.02535	0.0521512	-0.021469484	-0.020074
HWay	0.0733106	-0.03364	0.0033846	1	0.8260124	0.3082721	0.3084053	-0.109952	-0.002844	-0.014788	0.119708756	0.1091824
Thead	-0.227506	-0.026466	-0.001752	0.8260124	1	0.2494804	0.2495748	-0.199496	0.0031822	-0.025305	0.089078258	0.0676155
DLeft	0.0081567	-0.080206	0.019228	0.3082721	0.2494804	1	0.9999997	-0.267197	-0.05429	0.0301226	0.014788276	0.0139329
DRight	0.0082397	-0.080203	0.0192369	0.3084053	0.2495748	0.9999997	1	-0.267215	-0.05467	0.0301335	0.014731113	0.0139077
rdist	0.2562799	0.013857	-0.03049	-0.109952	-0.199496	-0.267197	-0.267215	1	0.1066635	-0.006929	0.041655942	0.0407544
rspur	0.0141984	0.0785	-0.02535	-0.002844	0.0031822	-0.05429	-0.05467	0.1066635	1	-0.025022	0.06780089	0.036318
Wheel	0.0626729	0.0382492	0.0521512	-0.014788	-0.025305	0.0301226	0.0301335	-0.006929	-0.025022	1	-0.00111387	-0.006512
Age	-0.102713	-0.000168	-0.021469	0.1197088	0.0890783	0.0147883	0.0147311	0.0416559	0.0678009	-0.001114	1	0.866171359
Driving experience	-0.062015	-0.00369	-0.020074	0.1091824	0.0676155	0.0139329	0.0139077	0.0407544	0.036318	-0.006512	0.866171359	1

Στον παραπάνω πίνακα με πράσινο έχουν σημειωθεί οι τιμές που αντιστοιχούν σε μεγάλη συσχέτιση (>0.8) και με κόκκινο οι μεταβλητές που για αυτό το λόγο αποφασίστηκε να αφαιρεθούν.

Όπως φαίνεται στον πίνακα υψηλή συσχέτιση παρουσιάζουν:

- Η απόσταση από το προπορευόμενο όχημα (HWay) με τον χρόνο μέχρι τη σύγκρουση με το προπορευόμενο όχημα (Thead)
- Η απόσταση από την αριστερή οριογραμμή (DLeft) με την απόσταση από την δεξιά οριογραμμή (DRight)
- Η ηλικία (Age) με την οδηγική εμπειρία (Driving experience)

Τελικά, στον πίνακα 4.11 παρουσιάζονται οι μεταβλητές που έμειναν για να χρησιμοποιηθούν στα μοντέλα μετά τη συσχέτιση:

Πίνακας 4.11: Τελικές μεταβλητές

ΤΕΛΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ	
Speed	rdist
AccLat	rspur
AccLon	Wheel
THead	Driving experience
DRight	

4.7. Πίνακες για τα μοντέλα στατιστικής ανάλυσης

Σύμφωνα με όλα τα παραπάνω δημιουργήθηκε ο τελικός πίνακας σύμφωνα με τον οποίο αναπτύχθηκαν τα μοντέλα στατιστικής ανάλυσης (Διωνυμική Λογιστική Παλινδρόμηση, Τυχαία Δάση). Ο πίνακας **Model** προέκυψε απομονώνοντας από τον πίνακα EventsU τις στήλες: Συμβάν (**Event**), Ταχύτητα (**Speed**), Πλευρική επιτάχυνση (**AccLat**), Διαμήκης επιτάχυνση (**AccLon**), Χρόνος μέχρι τη σύγκρουση με το προπορευόμενο όχημα (**THead**), Απόσταση από τη δεξιά οριογραμμή (**DRight**), Συνολική διανυόμενη απόσταση οχήματος (**rdist**), Απόκλιση οχήματος από το μέσο της οδού (**rspur**), Θέση τιμονιού (**Wheel**), Οδηγική εμπειρία (**Driving experience**).

Η στήλη Event (εξαρτημένη μεταβλητή) περιελάμβανε τιμές ή 0 (για μη ύπαρξη συμβάντος) ή κάποιον αριθμό διάφορο του 0 (για ύπαρξη συμβάντος) που συμβόλιζε τη μορφή του αντίστοιχου συμβάντος. Επιλέχθηκε να αντικατασταθούν όλες οι τιμές διάφορες του 0 με την τιμή 1, ώστε να υπάρχουν μόνο οι τιμές 0 και 1 για μη ύπαρξη και ύπαρξη συμβάντος, αντίστοιχα, καθώς αυτό παρουσιάζει ενδιαφέρον στην παρούσα έρευνα.

Στον πίνακα 4.12 παρουσιάζεται ένα απόσπασμα του τελικού πίνακα Model.

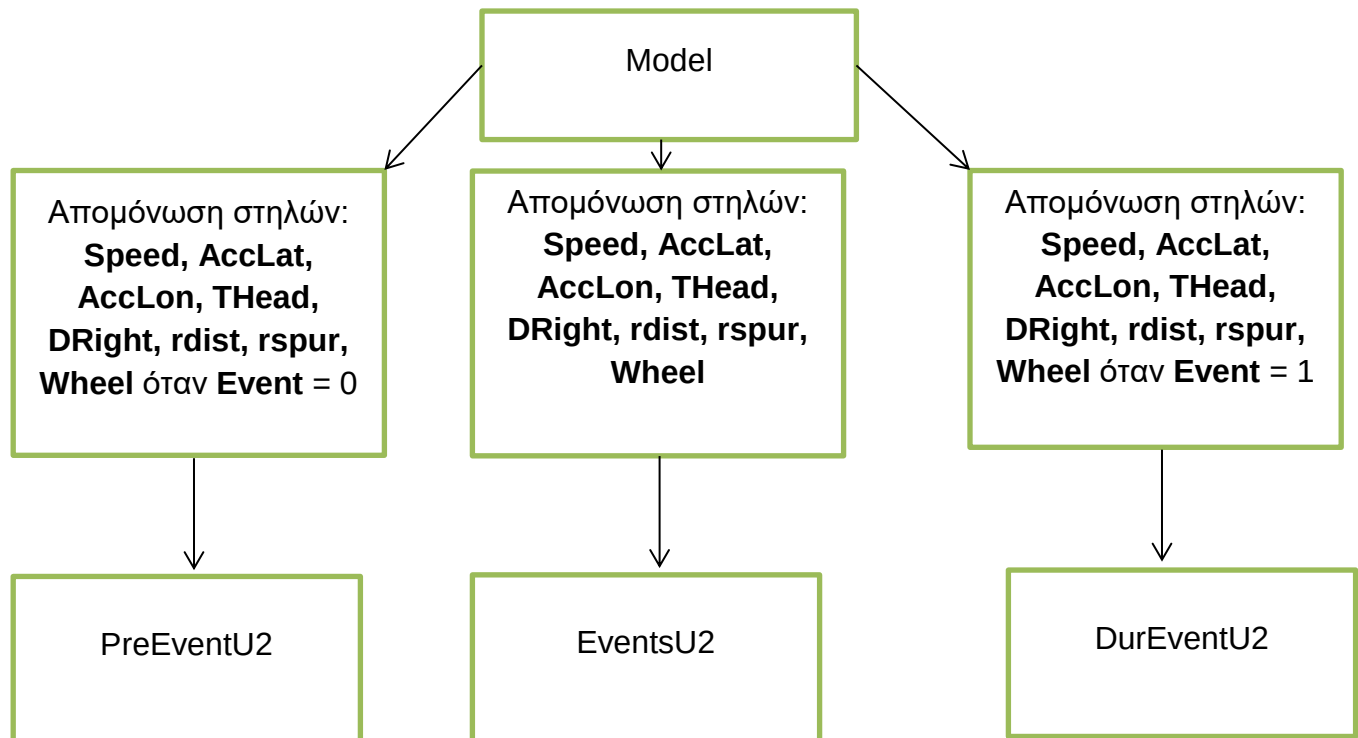
Πίνακας 4.12: Απόσπασμα πίνακα Model (15/438,126 σειρές)

Event	Speed	AccLat	AccLon	THead	DRight	rdist	rspur	Wheel	Driving experience
0	18.4	0.001	0.384	39.4	1.36	30.35	1.84	-2	46
0	18.5	0.002	0.438	39.2	1.36	30.52	1.84	-2	46
0	18.5	0.002	0.469	39.1	1.35	30.69	1.84	-2	46
0	18.6	0.002	0.495	38.9	1.35	30.86	1.84	-2	46
0	18.6	0.002	0.527	38.8	1.35	31.04	1.84	-2	46
0	18.7	0.002	0.55	38.6	1.35	31.21	1.84	-2	46
0	18.8	0.002	0.571	38.4	1.35	31.38	1.84	-2	46
0	18.9	0.003	0.6	38.3	1.35	31.56	1.84	-2	46
0	18.9	0.003	0.621	38.1	1.35	31.73	1.85	-2	46
0	19	0.003	0.642	37.9	1.35	31.91	1.85	-2	46
0	19	0.001	0.273	37.8	1.34	32.09	1.85	-2	46
0	19	0	0.002	37.8	1.34	32.26	1.85	-2	46
0	19	0	-0.052	37.8	1.34	32.44	1.85	-2	46
0	19	0	0.031	37.7	1.34	32.62	1.85	-2	46
0	19.1	0.001	0.169	37.7	1.34	32.79	1.85	-2	46

4.8. Πίνακες για τη μέθοδο Παραγοντικής Ανάλυσης

Η μέθοδος παραγοντικής ανάλυσης (Factor Analysis) πραγματοποιήθηκε στους πίνακες EventsU2, PreEventU2 και DurEventU2. Ο πίνακας EventsU2 δημιουργήθηκε απομονώνοντας από τον πίνακα Model τις στήλες Speed, AccLat, AccLon, THead, DRight, rdist, rspur και Wheel. Ο πίνακας PreEventU2 δημιουργήθηκε απομονώνοντας από τον πίνακα Model τις ίδιες στήλες μόνο όταν Event = 0 δηλαδή τα στοιχεία που αφορούν στη μη ύπαρξη συμβάντος. Ο πίνακας DurEventU2 δημιουργήθηκε απομονώνοντας από τον πίνακα Model τις ίδιες στήλες μόνο όταν Event = 1 δηλαδή τα στοιχεία που αφορούν στην ύπαρξη συμβάντος.

Η παραπάνω διαδικασία δημιουργίας των πινάκων για την παραγοντική ανάλυση μπορεί να περιγραφεί στο Γράφημα 4.2.



Γράφημα 4.2: Δημιουργία πινάκων PreEventU2, DurEventU2, EventsU2

Αποσπάσματα των πινάκων EventsU2, PreEventU2 και DurEventU2 παρουσιάζονται στους πίνακες 4.13, 4.14 και 4.15 αντίστοιχα.

Πίνακας 4.13: Απόσπασμα πίνακα EventsU2 (15/438,126 σειρές)

Speed	AccLat	AccLon	Thead	Drigh	rdist	rspur	Wheel
18.4	0.001	0.384	39.4	1.36	30.35	1.84	-2
18.5	0.002	0.438	39.2	1.36	30.52	1.84	-2
18.5	0.002	0.469	39.1	1.35	30.69	1.84	-2
18.6	0.002	0.495	38.9	1.35	30.86	1.84	-2
18.6	0.002	0.527	38.8	1.35	31.04	1.84	-2
18.7	0.002	0.55	38.6	1.35	31.21	1.84	-2
18.8	0.002	0.571	38.4	1.35	31.38	1.84	-2
18.9	0.003	0.6	38.3	1.35	31.56	1.84	-2
18.9	0.003	0.621	38.1	1.35	31.73	1.85	-2
19	0.003	0.642	37.9	1.35	31.91	1.85	-2
19	0.001	0.273	37.8	1.34	32.09	1.85	-2
19	0	0.002	37.8	1.34	32.26	1.85	-2
19	0	-0.052	37.8	1.34	32.44	1.85	-2
19	0	0.031	37.7	1.34	32.62	1.85	-2
19.1	0.001	0.169	37.7	1.34	32.79	1.85	-2

Πίνακας 4.14: Απόσπασμα πίνακα PreEventU2 (15/334,989 σειρές)

Speed	AccLat	AccLon	Thead	Drigh	rdist	rspur	Wheel
18.4	0.001	0.384	39.4	1.36	30.35	1.84	-2
18.5	0.002	0.438	39.2	1.36	30.52	1.84	-2
18.5	0.002	0.469	39.1	1.35	30.69	1.84	-2
18.6	0.002	0.495	38.9	1.35	30.86	1.84	-2
18.6	0.002	0.527	38.8	1.35	31.04	1.84	-2
18.7	0.002	0.55	38.6	1.35	31.21	1.84	-2
18.8	0.002	0.571	38.4	1.35	31.38	1.84	-2
18.9	0.003	0.6	38.3	1.35	31.56	1.84	-2
18.9	0.003	0.621	38.1	1.35	31.73	1.85	-2
19	0.003	0.642	37.9	1.35	31.91	1.85	-2
19	0.001	0.273	37.8	1.34	32.09	1.85	-2
19	0	0.002	37.8	1.34	32.26	1.85	-2
19	0	-0.052	37.8	1.34	32.44	1.85	-2
19	0	0.031	37.7	1.34	32.62	1.85	-2
19.1	0.001	0.169	37.7	1.34	32.79	1.85	-2

Πίνακας 4.15: Απόσπασμα πίνακα DurEventU2 (15/103,137 σειρές)

Speed	AccLat	AccLon	Thead	Drigh	rdist	rspur	Wheel
23.4	0.039	1.087	9999.9	0.67	25.93	2.48	11
23.6	0.037	1.069	9999.9	0.66	26.26	2.49	11
23.8	0.032	0.955	9999.9	0.66	26.59	2.49	14
23.9	0.028	0.891	9999.9	0.65	26.92	2.5	17
24.1	0.025	0.841	9999.9	0.65	27.25	2.5	18
24.2	0.022	0.809	9999.9	0.65	27.59	2.51	18
24.4	0.021	0.819	9999.9	0.65	27.93	2.51	18
24.5	0.02	0.839	9999.9	0.65	28.27	2.51	18
24.7	0.018	0.849	9999.9	0.65	28.61	2.51	18
24.8	0.016	0.836	9999.9	0.66	28.96	2.51	18
25	0.013	0.766	9999.9	0.66	29.3	2.52	18
25.1	0.011	0.711	9999.9	0.66	29.65	2.51	18
25.2	0.009	0.684	9999.9	0.67	30	2.51	18
25.4	0.008	0.705	9999.9	0.67	30.35	2.51	18
25.5	0.007	0.745	9999.9	0.68	30.71	2.51	18

Σε αυτό το κεφάλαιο περιγράφηκε ο τρόπος συλλογής και επεξεργασίας των στοιχείων που διαμόρφωσαν τους τελικούς πίνακες στους οποίους στο επόμενο κεφάλαιο θα εφαρμοστεί η μεθοδολογία και θα περιγραφούν τα αποτελέσματα που προκύπτουν.

5. ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ – ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1. Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζεται η εφαρμογή της μεθοδολογίας που ακολουθήθηκε στην παρούσα έρευνα καθώς και τα αποτελέσματά της. Όπως αναφέρθηκε και στα παραπάνω κεφάλαια στόχος της εργασίας είναι ο προσδιορισμός των κρίσιμων παραγόντων εντοπισμού οδικών συμβάντων και για την επίτευξή του αναπτύχθηκαν δύο παραλλαγές για κάθε μοντέλο, το μοντέλο διωνυμικής λογιστικής παλινδρόμησης και το μοντέλο τυχαίων δασών. Επίσης, πραγματοποιήθηκε παραγοντική ανάλυση. Όλα τα παραπάνω έγιναν με κατάλληλους κώδικες στο προγραμματιστικό περιβάλλον του RStudio.

Η **παραλλαγή Α** περιελάμβανε όλες τις ανεξάρτητες μεταβλητές που θεωρήθηκαν σημαντικές για ανάλυση μετά τη συσχέτιση, δηλαδή τις: ταχύτητα (**Speed**), πλευρική επιτάχυνση (**AccLat**), διαμήκης επιτάχυνση (**AccLon**), χρόνος μέχρι τη σύγκρουση με το προπορευόμενο όχημα (**THead**), απόσταση από τη δεξιά οριογραμμή (**DRight**), διανυόμενη απόσταση από την αρχή της οδήγησης (**rdist**), απόκλιση οχήματος από το μέσο της οδού (**rspur**), θέση τιμονιού (**Wheel**) και οδηγική εμπειρία (**Driving experience**).

Η **παραλλαγή Β** περιελάμβανε τις εξής ανεξάρτητες μεταβλητές: ταχύτητα (**Speed**), χρόνος μέχρι τη σύγκρουση με το προπορευόμενο όχημα (**THead**), απόσταση από τη δεξιά οριογραμμή (**DRight**), διανυόμενη απόσταση από την αρχή της οδήγησης (**rdist**), θέση τιμονιού (**Wheel**) και οδηγική εμπειρία (**Driving experience**). Ο καθορισμός των μεταβλητών της παραλλαγής Β έγινε μετά την εφαρμογή της μεθόδου Boruta που εκτίμησε την σημαντικότητα των μεταβλητών. Στον πίνακα 5.1 παρουσιάζεται η κατάταξη της σημαντικότητας των ανεξάρτητων μεταβλητών (feature importance). Μετά από δοκιμές αποφασίστηκε να μείνουν όλες οι μεταβλητές που κατατάχθηκαν ως σημαντικές με βάση τη συγκεκριμένη μέθοδο στην παραλλαγή Β.

Πίνακας 5.1: Σημαντικότητα ανεξάρτητων μεταβλητών με βάση τη μέθοδο Boruta

ΜΕΤΑΒΛΗΤΕΣ	ΜΕΣΗ ΣΗΜΑΝΤΙΚΟΤΗΤΑ
Wheel	228.78
rdist	177.71
Speed	173.15
DRight	160.84
Driving experience	143.86
THead	129.70

Και στις δύο παραλλαγές ως εξαρτημένη καθορίστηκε η μεταβλητή **Event** (διακριτή) με τιμές 0 ή 1 ανάλογα με την μη ύπαρξη ή την ύπαρξη συμβάντος αντίστοιχα. Οι υπόλοιπες ανεξάρτητες (συνεχείς) διαφέρουν σε κάθε παραλλαγή.

5.2. Μοντέλο διωνυμικής λογιστικής παλινδρόμησης

5.2.1. Παραλλαγή Α

Για την παραλλαγή Α του μοντέλου διωνυμικής λογιστικής παλινδρόμησης χρησιμοποιήθηκε ο πίνακας **Model** (4,381,260 στοιχεία) με εξαρτημένη τη μεταβλητή **Event** και ανεξάρτητες τις: **Speed**, **AccLat**, **AccLon**, **THead**, **DRight**, **rdist**, **rspur**, **Wheel** και **Driving experience**.

Ο πίνακας Model διαχωρίστηκε σε 2 επιμέρους:

1. Τον πίνακα **training1** που αποτελούσε το 80% του αρχικού πίνακα και δημιουργήθηκε για εκπαίδευση (3,505,000 στοιχεία)
2. Τον πίνακα **test1** που αποτελούσε το 20% του αρχικού πίνακα και δημιουργήθηκε για έλεγχο (876,260 στοιχεία)

Στους πίνακες 5.2 και 5.3 παρουσιάζονται αποσπάσματα των πινάκων training1 και test1 αντίστοιχα.

Πίνακας 5.2: Απόσπασμα πίνακα training1 (15/350,500 σειρές)

Event	Speed	AccLat	AccLon	THead	DRight	rdist	rspur	Wheel	Driving experience
0	18.4	0.001	0.384	39.4	1.36	30.35	1.84	-2	46
0	18.5	0.002	0.438	39.2	1.36	30.52	1.84	-2	46
0	18.5	0.002	0.469	39.1	1.35	30.69	1.84	-2	46
0	18.6	0.002	0.495	38.9	1.35	30.86	1.84	-2	46
0	18.6	0.002	0.527	38.8	1.35	31.04	1.84	-2	46
0	18.7	0.002	0.55	38.6	1.35	31.21	1.84	-2	46
0	18.8	0.002	0.571	38.4	1.35	31.38	1.84	-2	46
0	18.9	0.003	0.6	38.3	1.35	31.56	1.84	-2	46
0	18.9	0.003	0.621	38.1	1.35	31.73	1.85	-2	46
0	19	0.003	0.642	37.9	1.35	31.91	1.85	-2	46
0	19	0.001	0.273	37.8	1.34	32.09	1.85	-2	46
0	19	0	0.002	37.8	1.34	32.26	1.85	-2	46
0	19	0	-0.052	37.8	1.34	32.44	1.85	-2	46
0	19	0	0.031	37.7	1.34	32.62	1.85	-2	46
0	19.1	0.001	0.169	37.7	1.34	32.79	1.85	-2	46

Πίνακας 5.3: Απόσπασμα πίνακα test1 (15/87,626 σειρές)

Event	Speed	AccLat	AccLon	THead	DRight	rdist	rspur	Wheel	Driving experience
0	33.3	0.005	0.362	5.2	0.44	35.32	7.44	0	35
0	33.2	0.005	0.362	5.2	0.43	34.86	7.45	2	35
0	33.1	0.005	0.362	5.2	0.43	34.4	7.45	2	35
0	33.1	0.004	0.361	5.2	0.43	33.94	7.46	2	35
0	33	0.004	0.361	5.2	0.42	33.48	7.46	2	35
0	32.9	0.004	0.361	5.2	0.42	33.02	7.46	2	35
0	32.9	0.004	0.361	5.2	0.41	32.56	7.47	2	35
0	32.8	0.004	0.361	5.2	0.41	32.11	7.47	3	35
0	32.7	0.004	0.361	5.1	0.41	31.65	7.47	4	35
0	32.7	0.004	0.361	5.1	0.41	31.2	7.48	4	35
0	32.6	0.003	0.36	5.1	0.41	30.75	7.48	6	35
0	32.6	0.003	0.36	5.1	0.41	30.29	7.48	8	35
0	32.5	0.003	0.36	5.1	0.41	29.84	7.48	8	35
0	32.4	0.002	0.36	5.1	0.41	29.39	7.48	11	35
0	32.4	0.001	0.36	5.1	0.42	28.94	7.47	12	35

Εκπαίδευση μοντέλου

Η εκπαίδευση του μοντέλου έγινε στον πίνακα training1 και βρέθηκαν οι συντελεστές των μεταβλητών με τη βοήθεια των οποίων δημιουργήθηκε η συνάρτηση χρησιμότητας για την ύπαρξη συμβάντος και έγινε ο έλεγχος ποιότητάς τους όπως παρουσιάζονται στον πίνακα 5.4.

Πίνακας 5.4: Έλεγχος συντελεστών

	Estimate	Std. Error	z value	Pr(> z)
Σταθερός όρος	-2.048	0.01365	-150.018	< 2e-16
Speed	0.0059	0.000267	22.093	< 2e-16
AccLat	0.135	0.02295	5.883	0.00
AccLon	-0.0202	0.002248	-8.985	< 2e-16
THead	0.00002839	0.0000008844	32.104	< 2e-16
DRight	-0.0001003	0.000001779	-56.372	< 2e-16
rdist	-0.005719	0.000166	-34.448	< 2e-16
rspur	0.2018	0.002367	85.245	< 2e-16
Wheel	0.01546	0.0001544	100.103	< 2e-16
`Driving experience`	0.007692	0.000311	24.737	< 2e-16

Η συνάρτηση χρησιμότητας για την ύπαρξη συμβάντος που προκύπτει από τον παραπάνω πίνακα είναι:

$$\begin{aligned} \text{Event} = & - 2.048 + 0.0059 * \text{Speed} + 0.135 * \text{AccLat} - 0.0202 * \text{AccLon} \\ & + 0.00002839 * \text{THead} - 0.0001003 * \text{DRight} - 0.005719 * \text{rdist} + 0.2018 * \text{rspur} \\ & + 0.01546 * \text{Wheel} + 0.007692 * \text{Driving experience} \end{aligned}$$

Τα **πρόσημα των συντελεστών** της συνάρτησης χρησιμότητας έχουν **λογική εξήγηση** καθώς το θετικό πρόσημο που εμφανίζεται στην ταχύτητα (Speed), στην πλευρική επιτάχυνση (AccLat), στην απόκλιση του οχήματος από το μέσο της οδού (rspur) και στη θέση του τιμονιού (Wheel) δείχνει πως όσο αυξάνονται αυτές οι μεταβλητές, αυξάνεται και η πιθανότητα ύπαρξης συμβάντος. Το αρνητικό πρόσημο που εμφανίζεται στην απόσταση από τη δεξιά οριογραμμή (DRight) και στη συνολική διανυόμενη απόσταση (rdist) (ο οδηγός εξοικειώνεται με την διαδρομή) επίσης εξηγείται λογικά καθώς όσο μειώνονται αυτά τα μεγέθη αυξάνεται η πιθανότητα να προκύψει συμβάν. Δεν υπάρχει λογική ερμηνεία για τον χρόνο μέχρι τη σύγκρουση με το προπορευόμενο όχημα (THead), για την οδηγική εμπειρία (Driving experience) και για τη διαμήκη επιτάχυνση (AccLon).

Επίσης, ικανοποιείται ο έλεγχος z-test της **στατιστικής σημαντικότητας** των ανεξάρτητων μεταβλητών καθώς σε όλες τις μεταβλητές ισχύει πως $\text{Pr}(>|z|) < 0.05$.

Έλεγχος μοντέλου

Ο έλεγχος του μοντέλου έγινε στον πίνακα test1 και με τη μέθοδο της μήτρας σύγχυσης (confusion matrix) ελέγχθηκε πόσο καλά προβλέπονται οι τιμές της εξαρτημένης μεταβλητής Event. Ως θετική κλάση ορίστηκε η ύπαρξη συμβάντος που αντιπροσωπεύεται από την τιμή 1. Αντίστοιχα η τιμή 0 αντιπροσωπεύει τη μη ύπαρξη συμβάντος. Στον πίνακα 5.5 παρουσιάζονται τα αποτελέσματα της μήτρας σύγχυσης.

Πίνακας 5.5: Μήτρα σύγχυσης (Παραλλαγή Α)

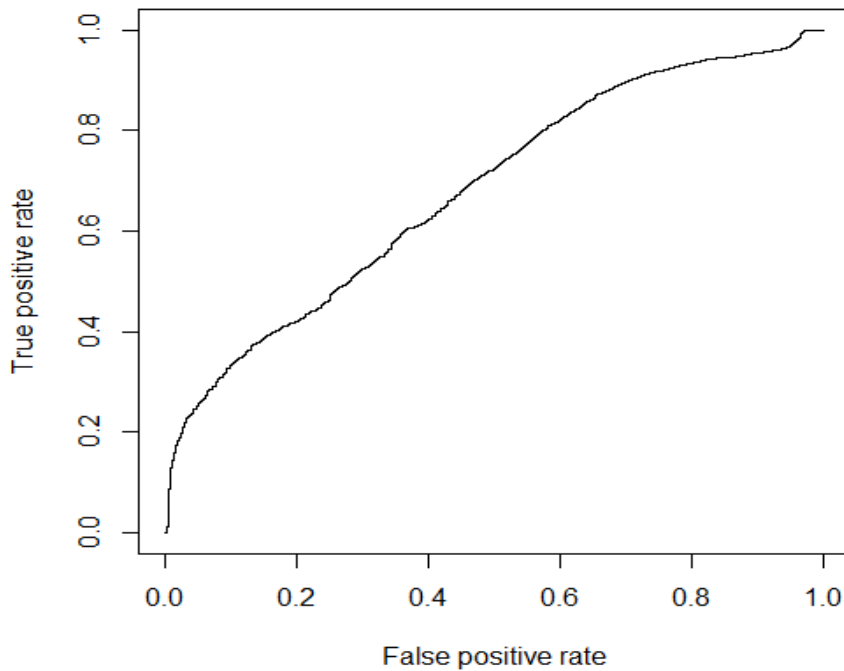
Reference (Πραγματική κλάση)	Prediction (Προβλεπόμενη κλάση)	
	Συμβάν	Όχι συμβάν
Συμβάν	2,723	17,530
Όχι συμβάν	647	66,726

Στον πίνακα 5.6 παρουσιάζονται τα πιο κοινά μέτρα αξιολόγησης (evaluation metrics) του μοντέλου που υπολογίσθηκαν.

Πίνακας 5.6: Μέτρα αξιολόγησης μοντέλου (Παραλλαγή Α)

Accuracy	79.26%
Recall	13.44%
Specificity	99.04%
Precision	80.80%
F-measure	23.05%
False alarm rate	0.96%
AUC	68.18%

Στην εικόνα 5.1 φαίνεται η καμπύλη ROC δηλαδή η γραφική παράσταση του αληθινού θετικού ποσοστού έναντι του ψευδούς θετικού ποσοστού για τον ταξινομητή. Το AUC υπολογίζεται ως το εμβαδόν κάτω από αυτή την καμπύλη.

**Εικόνα 5.1:** Καμπύλη ROC

Αξιολόγηση μοντέλου

Με βάση τον έλεγχο απόδοσης του διωνυμικού λογιστικού μοντέλου (Παραλλαγή Α) γίνεται αξιολόγηση της ικανότητάς του να προβλέψει την ύπαρξη συμβάντος η οποία παρουσιάζεται στον πίνακα 5.7.

Πίνακας 5.7: Αξιολόγηση διωνυμικού λογιστικού μοντέλου (Παραλλαγή Α)

Μέτρα αξιολόγησης	Ποσοστό (%)	Περιγραφή	Αξιολόγηση
Ορθότητα (Accuracy)	79.26	Σωστές προβλέψεις	
Ανάκληση (Recall)	13.44	Σωστές θετικές προβλέψεις	
Εξειδικευτικότητα (Specificity)	99.04	Σωστές αρνητικές προβλέψεις	
Ακρίβεια (Precision)	80.8	Σωστές θετικές προβλέψεις/Συνολικός αριθμός θετικών	
Μέτρο F (F-measure)	23.05	Αρμονικό μέσο ακρίβειας και ανάκλησης	
Δείκτης λάθος συναγερμού (False alarm rate)	0.96	Λανθασμένες θετικές προβλέψεις/Συνολικός αριθμός αρνητικών	
Εμβαδόν κάτω από την καμπύλη ROC (AUC)	68.18	Ορθότητα	

ΥΠΟΜΝΗΜΑ	
	Πολύ ικανοποιητικό
	Ικανοποιητικό
	Οριακά ικανοποιητικό
	Καθόλου ικανοποιητικό

Σύμφωνα με τον παραπάνω πίνακα το μοντέλο διωνυμικής λογιστικής παλινδρόμησης (Παραλλαγή Α) παρουσιάζει πολύ χαμηλό δείκτη ανάκλησης (recall) άρα δεν προβλέπει καθόλου ικανοποιητικά τα θετικά στιγμιότυπα, δηλαδή την ύπαρξη συμβάντος, ενώ όμως έχει την ικανότητα να προβλέψει πολύ καλά τη μη ύπαρξη συμβάντος (υψηλό specificity). Η συνολική ακρίβεια (accuracy) είναι ικανοποιητική δηλαδή το μοντέλο είναι αξιόπιστο για τις σωστές προβλέψεις καθώς και για τον βαθμό πιστότητας της διαδικασίας (precision). Η πιθανότητα λάθους ταξινόμησης των θετικών στιγμιοτύπων (false alarm rate) είναι επίσης αρκετά ικανοποιητική όμως το μέτρο F (F-measure) που εκφράζει τον αρμονικό μέσο της ακρίβειας και της ανάκλησης είναι πολύ χαμηλό. Το εμβαδόν κάτω από την καμπύλη ROC (AUC) είναι οριακά ικανοποιητικό.

Τελικά από τα παραπάνω προκύπτει πως το μοντέλο της διωνυμικής λογιστικής παλινδρόμησης για την Παραλλαγή Α δεν λειτουργεί ικανοποιητικά στον εντοπισμό συμβάντος. Για αυτό τον λόγο γίνεται διερεύνηση και του μοντέλου τυχαίων δασών.

5.2.2. Παραλλαγή Β

Το μοντέλο διωνυμικής λογιστικής παλινδρόμησης για την Παραλλαγή Β δεν θεωρήθηκε άξιο ανάλυσης καθώς δεν προέβλεπε καθόλου ικανοποιητικά την ύπαρξη συμβάντος.

5.3. Μοντέλο τυχαίων δασών

5.3.1. Παραλλαγή A

Για την παραλλαγή A του μοντέλου τυχαίων δασών χρησιμοποιήθηκαν οι πίνακες **training1** (για εκπαίδευση) και **test1** (για έλεγχο), όπως και στο μοντέλο λογιστικής παλινδρόμησης.

Εκπαίδευση μοντέλου

Η εκπαίδευση του μοντέλου πραγματοποιήθηκε στον πίνακα training1 και προέκυψαν τα αποτελέσματα του πίνακα 5.8.

Πίνακας 5.8: Αποτελέσματα εκπαίδευσης (Παραλλαγή A)

Τύπος μεθόδου τυχαίων δασών: Ταξινόμηση
Αριθμός δέντρων: 100
Αριθμός μεταβλητών που δοκιμάστηκαν σε κάθε διαχωρισμό: 3
Εκτίμηση ποσοστού σφάλματος (OOB error rate): 0.48%

Έλεγχος μοντέλου

Ο έλεγχος του μοντέλου πραγματοποιήθηκε στον πίνακα test1 και με τη μέθοδο της μήτρας σύγχυσης (confusion matrix) εξετάστηκε πόσο καλά προβλέπεται η ύπαρξη συμβάντος. Ως θετική κλάση ορίστηκε η ύπαρξη συμβάντος που αντιπροσωπεύεται από την τιμή 1. Αντίστοιχα η τιμή 0 αντιπροσωπεύει τη μη ύπαρξη συμβάντος. Στον πίνακα 5.9 παρουσιάζονται τα αποτελέσματα της μήτρας σύγχυσης.

Πίνακας 5.9: Μήτρα σύγχυσης (Παραλλαγή A)

Reference (Πραγματική κλάση)	Prediction (Προβλεπόμενη κλάση)	
	Συμβάν	Όχι συμβάν
Συμβάν	13,278	6,975
Όχι συμβάν	4,268	63,105

Στον πίνακα 5.10 παρουσιάζονται τα πιο κοινά μέτρα αξιολόγησης (evaluation metrics) του μοντέλου που υπολογίστηκαν.

Πίνακας 5.10: Μέτρα αξιολόγησης μοντέλου (Παραλλαγή Α)

Accuracy	87.17%
Recall	65.56%
Specificity	93.67%
Precision	75.68%
F-measure	70.26%
False alarm rate	6.33%

Αξιολόγηση μοντέλου

Με βάση τον έλεγχο απόδοσης του μοντέλου τυχαίων δασών (Παραλλαγή Α) γίνεται αξιολόγηση της ικανότητάς του να προβλέψει την ύπαρξη συμβάντος η οποία παρουσιάζεται στον πίνακα 5.11.

Πίνακας 5.11: Αξιολόγηση μοντέλου τυχαίων δασών (Παραλλαγή Α)

Μέτρα αξιολόγησης	Ποσοστό (%)	Περιγραφή	Αξιολόγηση
Ορθότητα (Accuracy)	87.17	Σωστές προβλέψεις	
Ανάκληση (Recall)	65.56	Σωστές θετικές προβλέψεις	
Εξειδικευτικότητα (Specificity)	93.67	Σωστές αρνητικές προβλέψεις	
Ακρίβεια (Precision)	75.68	Σωστές θετικές προβλέψεις/Συνολικός αριθμός θετικών	
Μέτρο F (F-measure)	70.26	Αρμονικό μέσο ακρίβειας και ανάκλησης	
Δείκτης λάθος συναγερμού (False alarm rate)	6.33	Λανθασμένες θετικές προβλέψεις/Συνολικός αριθμός αρνητικών	

ΥΠΟΜΝΗΜΑ	
	Πολύ ικανοποιητικό
	Ικανοποιητικό
	Οριακά ικανοποιητικό
	Καθόλου ικανοποιητικό

Σύμφωνα με τον παραπάνω πίνακα το μοντέλο τυχαίων δασών (Παραλλαγή Α) προβλέπει οριακά ικανοποιητικά τα θετικά στιγμιότυπα (recall), δηλαδή την ύπαρξη συμβάντος, και έχει την ικανότητα να προβλέψει πολύ καλά τη μη ύπαρξη συμβάντος (υψηλό specificity). Το μοντέλο είναι επίσης αξιόπιστο για τις σωστές προβλέψεις συνολικά καθώς παρουσιάζει υψηλό δείκτη ορθότητας (accuracy), αλλά και για τον βαθμό πιστότητας της διαδικασίας κατηγοριοποίησης (υψηλό precision). Η πιθανότητα λάθους ταξινόμησης των θετικών στιγμιοτύπων (false alarm rate) είναι ικανοποιητική όπως επίσης και το μέτρο F (F-measure) που εκφράζει τον αρμονικό μέσο της ακρίβειας και της ανάκλησης.

Από τα παραπάνω συμπεραίνεται πως το μοντέλο τυχαίων δασών για την Παραλλαγή Α προβλέπει ικανοποιητικά την ύπαρξη συμβάντος και θεωρείται αξιόπιστο.

5.3.2. Παραλλαγή Β

Για την παραλλαγή Β του μοντέλου τυχαίων δασών δημιουργήθηκε ο πίνακας **Model2** (3,066,882 στοιχεία) ο οποίος προέκυψε από τον πίνακα Model απομονώνοντας τις εξής μεταβλητές: Event (εξαρτημένη), Speed, THead, DRight, rdist, Wheel και Driving experience (ανεξάρτητες).

Ο πίνακας Model2 διαχωρίστηκε σε 2 επιμέρους:

1. Τον πίνακα **training2** που αποτελούσε το 80% του πίνακα Model2 και δημιουργήθηκε για εκπαίδευση (2,453,500 στοιχεία)
2. Τον πίνακα **test2** που αποτελούσε το 20% του πίνακα Model2 και δημιουργήθηκε για έλεγχο (613,382 στοιχεία)

Στους πίνακες 5.12 και 5.13 παρουσιάζονται αποσπάσματα των πινάκων training2 και test2 αντίστοιχα.

Πίνακας 5.12: Απόσπασμα πίνακα training2 (15/350,500 σειρές)

Event	Speed	THead	DRight	rdist	Wheel	Driving experience
0	18.4	39.4	1.36	30.35	-2	46
0	18.5	39.2	1.36	30.52	-2	46
0	18.5	39.1	1.35	30.69	-2	46
0	18.6	38.9	1.35	30.86	-2	46
0	18.6	38.8	1.35	31.04	-2	46
0	18.7	38.6	1.35	31.21	-2	46
0	18.8	38.4	1.35	31.38	-2	46
0	18.9	38.3	1.35	31.56	-2	46
0	18.9	38.1	1.35	31.73	-2	46
0	19	37.9	1.35	31.91	-2	46
0	19	37.8	1.34	32.09	-2	46
0	19	37.8	1.34	32.26	-2	46
0	19	37.8	1.34	32.44	-2	46
0	19	37.7	1.34	32.62	-2	46
0	19.1	37.7	1.34	32.79	-2	46

Πίνακας 5.13: Απόσπασμα πίνακα test2 (15/87,626 σειρές)

Event	Speed	THead	DRight	rdist	Wheel	Driving experience
0	33.3	5.2	0.44	35.32	0	35
0	33.2	5.2	0.43	34.86	2	35
0	33.1	5.2	0.43	34.4	2	35
0	33.1	5.2	0.43	33.94	2	35
0	33	5.2	0.42	33.48	2	35
0	32.9	5.2	0.42	33.02	2	35
0	32.9	5.2	0.41	32.56	2	35
0	32.8	5.2	0.41	32.11	3	35
0	32.7	5.1	0.41	31.65	4	35
0	32.7	5.1	0.41	31.2	4	35
0	32.6	5.1	0.41	30.75	6	35
0	32.6	5.1	0.41	30.29	8	35
0	32.5	5.1	0.41	29.84	8	35
0	32.4	5.1	0.41	29.39	11	35
0	32.4	5.1	0.42	28.94	12	35

Εκπαίδευση μοντέλου

Η εκπαίδευση του μοντέλου πραγματοποιήθηκε στον πίνακα training2 και προέκυψαν τα αποτελέσματα του πίνακα 5.14.

Πίνακας 5.14: Αποτελέσματα εκπαίδευσης (Παραλλαγή Β)

Τύπος μεθόδου τυχαίων δασών: Ταξινόμηση
Αριθμός δέντρων: 100
Αριθμός μεταβλητών που δοκιμάστηκαν σε κάθε διαχωρισμό: 2
Εκτίμηση ποσοστού σφάλματος (OOB error rate): 0.97%

Έλεγχος μοντέλου

Ο έλεγχος του μοντέλου πραγματοποιήθηκε στον πίνακα test2 και με τη μέθοδο της μήτρας σύγχυσης (confusion matrix) εξετάστηκε πόσο καλά προβλέπεται η ύπαρξη συμβάντος. Ως θετική κλάση ορίστηκε η ύπαρξη συμβάντος που αντιπροσωπεύεται από την τιμή 1. Αντίστοιχα η τιμή 0 αντιπροσωπεύει τη μη ύπαρξη συμβάντος. Στον πίνακα 5.15 παρουσιάζονται τα αποτελέσματα της μήτρας σύγχυσης.

Πίνακας 5.15: Μήτρα σύγχυσης (Παραλλαγή Β)

Reference (Πραγματική κλάση)	Prediction (Προβλεπόμενη κλάση)	
	Συμβάν	Όχι συμβάν
Συμβάν	10,838	9,415
Όχι συμβάν	7,067	60,306

Στον πίνακα 5.16 παρουσιάζονται τα πιο κοινά μέτρα αξιολόγησης (evaluation metrics) του μοντέλου που υπολογίσθηκαν.

Πίνακας 5.16: Μέτρα αξιολόγησης μοντέλου (Παραλλαγή Β)

Accuracy	81.19%
Recall	53.51%
Specificity	89.51%
Precision	60.53%
F-measure	56.81%
False alarm rate	10.49%

Αξιολόγηση μοντέλου

Με βάση τον έλεγχο απόδοσης του μοντέλου τυχαίων δασών (Παραλλαγή Β) γίνεται αξιολόγηση της ικανότητάς του να προβλέψει την ύπαρξη συμβάντος η οποία παρουσιάζεται στον πίνακα 5.17.

Πίνακας 5.17: Αξιολόγηση μοντέλου τυχαίων δασών (Παραλλαγή Β)

Μέτρα αξιολόγησης	Ποσοστό (%)	Περιγραφή	Αξιολόγηση
Ορθότητα (Accuracy)	81.19	Σωστές προβλέψεις	
Ανάκληση (Recall)	53.51	Σωστές θετικές προβλέψεις	
Εξειδικευτικότητα (Specificity)	89.51	Σωστές αρνητικές προβλέψεις	
Ακρίβεια (Precision)	60.53	Σωστές θετικές προβλέψεις/Συνολικός αριθμός θετικών	
Μέτρο F (F-measure)	56.81	Αρμονικό μέσο ακρίβειας και ανάκλησης	
Δείκτης λάθος συναγερμού (False alarm rate)	10.49	Λανθασμένες θετικές προβλέψεις/Συνολικός αριθμός αρνητικών	

ΥΠΟΜΝΗΜΑ	
	Πολύ ικανοποιητικό
	Ικανοποιητικό
	Οριακά ικανοποιητικό
	Καθόλου ικανοποιητικό

Σύμφωνα με τον παραπάνω πίνακα το μοντέλο τυχαίων δασών (Παραλλαγή Β) παρουσιάζει οριακά ικανοποιητικό δείκτη ανάκλησης (recall), δηλαδή πρόβλεψη της ύπαρξης συμβάντος, ενώ όμως έχει την ικανότητα να προβλέψει καλά τη μη ύπαρξη συμβάντος (υψηλό specificity). Η συνολική ακρίβεια (accuracy) είναι ικανοποιητική

δηλαδή το μοντέλο είναι αξιόπιστο για τις σωστές προβλέψεις. Η πιθανότητα λάθους ταξινόμησης των θετικών στιγμιοτύπων (false alarm rate) είναι οριακά ικανοποιητική όπως επίσης και το μέτρο F (F-measure) που εκφράζει τον αρμονικό μέσο της ακρίβειας. Ο βαθμός πιστότητας της διαδικασίας ταξινόμησης (precision) είναι επίσης οριακά ικανοποιητικός.

Από τα παραπάνω συμπεραίνεται πως το μοντέλο τυχαίων δασών για την Παραλλαγή Β προβλέπει οριακά ικανοποιητικά την ύπαρξη συμβάντος και θεωρείται οριακά αξιόπιστο.

5.4. Ανάλυση παραγόντων

5.4.1. Παραγοντική ανάλυση στον πίνακα PreEventU2

Ανάλυση σε κύριες συνιστώσες

Αρχικά δημιουργήθηκε ο πίνακας συσχετίσεων των ανεξάρτητων μεταβλητών του πίνακα PreEventU2 και παρουσιάζεται στον πίνακα 5.18.

Πίνακας 5.18: Πίνακας συσχετίσεων

	Speed	AccLat	AccLon	THead	DRight	rdist	rspur	Wheel
Speed	1.0000							
AccLat	-0.0097	1.0000						
AccLon	-0.0257	-0.0622	1.0000					
THead	-0.2135	-0.0386	-0.0082	1.0000				
DRight	-0.0053	-0.0990	0.0281	0.2957	1.0000			
rdist	0.2571	0.0273	-0.0285	-0.2127	-0.2887	1.0000		
rspur	0.0467	0.1009	-0.0446	0.0277	-0.0362	0.0717	1.0000	
Wheel	0.0078	0.0023	0.0337	-0.0138	0.0798	-0.0246	-0.1079	1.0000

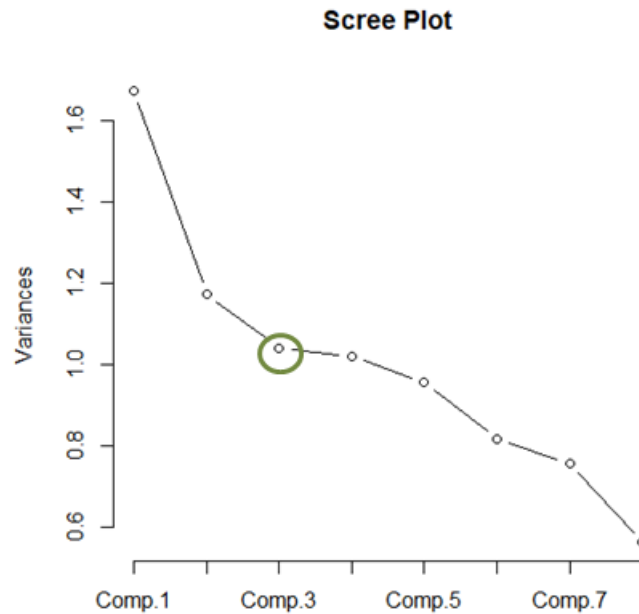
Από τον παραπάνω πίνακα παρατηρείται πως δεν υπάρχει επαρκής συσχέτιση μεταξύ των περισσότερων ανεξάρτητων μεταβλητών ($\text{correlation} < 0.2$).

Στη συνέχεια με τη μέθοδο κυρίων συνιστωσών μελετάται όλη η υπάρχουσα διακύμανση ώστε να εξαχθεί το μεγαλύτερο ποσοστό της διακύμανσης από τους λιγότερο δυνατούς παράγοντες. Στον πίνακα 5.19 παρουσιάζεται η σημαντικότητα των συνιστωσών.

Πίνακας 5.19: Πίνακας σημαντικότητας συνιστωσών

	Συνιστώσα 1	Συνιστώσα 2	Συνιστώσα 3	Συνιστώσα 4	Συνιστώσα 5	Συνιστώσα 6	Συνιστώσα 7	Συνιστώσα 8
Τυπική απόκλιση	1.2947	1.0832	1.0197	1.0100	0.9785	0.9028	0.8698	0.7497
Ποσοστό διακύμανσης	0.2095	0.1467	0.1300	0.1275	0.1197	0.1019	0.0946	0.0703
Αθροιστικό ποσοστό	0.2095	0.3562	0.4861	0.6136	0.7333	0.8352	0.9297	1.0000

Το διάγραμμα της εικόνας 5.1 παρουσιάζει τη διακύμανση ανά συνιστώσα. Με τη βοήθεια αυτής της γραφικής παράστασης προσδιορίζεται ο αριθμός των παραγόντων και εδώ επιλέχθηκαν 3 παράγοντες καθώς μέχρι αυτό το σημείο η διακύμανση είναι πάνω από τη μονάδα(η καμπύλη σε αυτό το σημείο κάνει έναν "αγκώνα").

**Εικόνα 5.1:** Διάγραμμα διακύμανσης – συνιστωσών

Ανάλυση παραγόντων

Έγινε ανάλυση παραγόντων στον πίνακα PreEventU2 για αριθμό παραγόντων ίσο με 3. Στον πίνακα 5.20 φαίνεται η μοναδικότητα των ανεξάρτητων μεταβλητών.

Πίνακας 5.20: Μοναδικότητα μεταβλητών

Speed	AccLat	AccLon	THead	DRight	rdist	rspur	Wheel
0.582	0.977	0.995	0.799	0.005	0.758	0.315	0.977

Πιθανόν οι μεταβλητές AccLat, AccLon και Wheel να μην εκφράζονται μέσω παραγόντων καθώς παρουσιάζουν μεγάλη μοναδικότητα.

Στον πίνακα 5.21 παρουσιάζονται τα αποτελέσματα της ανάλυσης παραγόντων.

Πίνακας 5.21: Ανάλυση παραγόντων

	Παράγοντας 1	Παράγοντας 2	Παράγοντας 3
Speed	0.125	0.634	
AccLat			0.135
AccLon			
THead	0.234	-0.383	
DRight	0.954	-0.191	-0.219
rdist	-0.189	0.440	0.111
rspur	0.153		0.813
Wheel			-0.142
SS loadings	1.048	0.781	0.764
Ποσοστό διακύμανσης	0.131	0.098	0.095
Αθροιστικό ποσοστό	0.131	0.229	0.324

Στον πίνακα 5.22 φαίνονται οι ομάδες παραγόντων των οποίων ο βαθμός ικανότητας να εκφράσουν μια μεταβλητή ξεπερνάει το 50%.

Πίνακας 5.22: Ομάδες παραγόντων

Παράγοντας 1	Παράγοντας 2	Παράγοντας 3
DRight	Speed	rspur

Από τον παραπάνω πίνακα φαίνεται πως ένα σύνολο μεταβλητών που αφορούν την κατάσταση πριν το συμβάν μπορεί να συμπυκνωθεί σε 3 παράγοντες:

1. Τον παράγοντα που εκφράζει την επιρροή της απόστασης από τη δεξιά οριογραμμή (DRight)
2. Τον παράγοντα που εκφράζει την επιρροή της ταχύτητας (Speed)
3. Τον παράγοντα που εκφράζει την επιρροή της απόκλισης του οχήματος από το μέσο της οδού (rspur)

Επίσης, από τον πίνακα 5.21 φαίνεται πως ο παράγοντας 1 που εκφράζει την επιρροή της απόστασης από τη δεξιά οριογραμμή περιγράφει το μεγαλύτερο ποσοστό διακύμανσης.

Έλεγχος ποιότητας δεδομένων

Οι δύο δείκτες που παρέχονται για τον έλεγχο της ποιότητας των δεδομένων είναι ο δείκτης Keiser-Meyer-Olkin (ΚΜΟ) που αξιολογεί την επάρκεια του δείγματος και ο δείκτης Bartlett's Test of Sphericity (p) που αξιολογεί την σφαιρικότητα. Μετά τη διενέργεια των ελέγχων προέκυψαν τα εξής αποτελέσματα:

- ΚΜΟ = 0.5543 > 0.50: ικανοποιείται ο έλεγχος επάρκειας
- $P < 0.05$: ικανοποιείται ο έλεγχος σφαιρικότητας

5.4.2. Παραγοντική ανάλυση στον πίνακα DurEventU2**Ανάλυση σε κύριες συνιστώσες**

Στον πίνακα 5.23 παρουσιάζεται ο πίνακας συσχετίσεων των ανεξάρτητων μεταβλητών του πίνακα DurEventU2.

Πίνακας 5.23: Πίνακας συσχετίσεων

	Speed	AccLat	AccLon	THead	DRight	rdist	rspur	Wheel
Speed	1.0000							
AccLat	0.0070	1.0000						
AccLon	0.0076	0.4257	1.0000					
THead	-0.2776	0.0053	0.0081	1.0000				
DRight	0.0681	-0.0003	0.0073	0.0936	1.0000			
rdist	0.2583	-0.0286	-0.0397	-0.1503	-0.1894	1.0000		
rspur	-0.0963	-0.0021	0.0004	-0.1066	-0.0911	0.2448	1.0000	
Wheel	0.1947	0.0652	0.0837	-0.0975	-0.0888	0.0628	0.1039	1.0000

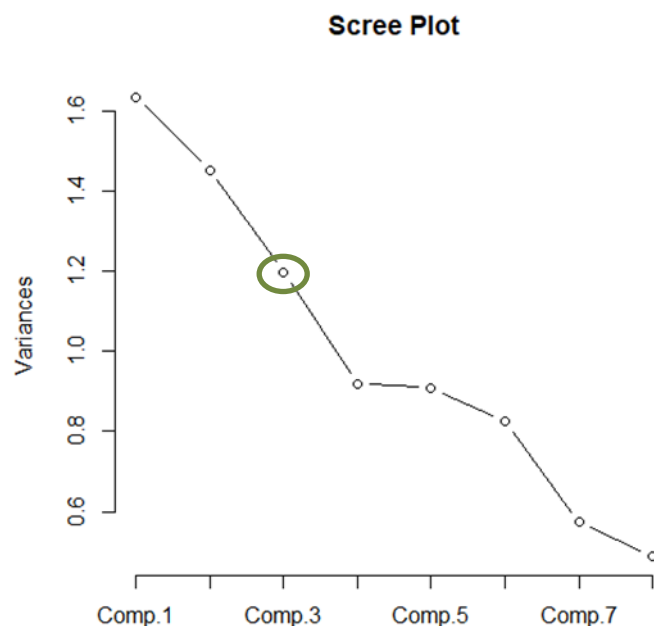
Και σε αυτόν τον πίνακα παρατηρείται πως δεν υπάρχει επαρκής συσχέτιση μεταξύ των περισσότερων ανεξάρτητων μεταβλητών.

Στον πίνακα 5.24 παρουσιάζεται η σημαντικότητα των συνιστωσών.

Πίνακας 5.24: Πίνακας σημαντικότητας συνιστωσών

	Συνιστώσα 1	Συνιστώσα 2	Συνιστώσα 3	Συνιστώσα 4	Συνιστώσα 5	Συνιστώσα 6	Συνιστώσα 7	Συνιστώσα 8
Τυπική απόκλιση	1.2791	1.2055	1.0945	0.9589	0.9529	0.9078	0.7573	0.6983
Ποσοστό διακύμανσης	0.2045	0.1816	0.1498	0.1149	0.1135	0.1030	0.0717	0.0609
Αθροιστικό ποσοστό	0.2045	0.3861	0.5359	0.6508	0.7644	0.8674	0.9391	1.0000

Σύμφωνα με το διάγραμμα της εικόνας 5.2 επιλέχθηκαν 3 παράγοντες καθώς μέχρι εκείνο το σημείο η διακύμανση είναι πάνω από τη μονάδα.



Εικόνα 5.2: Διάγραμμα διακύμανσης – συνιστωσών

Ανάλυση παραγόντων

Έγινε ανάλυση παραγόντων στον πίνακα DurEventU2 για αριθμό παραγόντων ίσο με 3. Στον πίνακα 5.25 φαίνεται η μοναδικότητα των ανεξάρτητων μεταβλητών.

Πίνακας 5.25: Μοναδικότητα μεταβλητών

Speed	AccLat	AccLon	THead	DRight	rdist	rspur	Wheel
0.005	0.661	0.467	0.873	0.888	0.645	0.761	0.925

Πιθανόν οι μεταβλητές THead, DRight και Wheel να μην εκφράζονται μέσω παραγόντων καθώς παρουσιάζουν μεγάλη μοναδικότητα.

Στον πίνακα 5.26 παρουσιάζονται τα αποτελέσματα της ανάλυσης παραγόντων.

Πίνακας 5.26: Ανάλυση παραγόντων

	Παράγοντας 1	Παράγοντας 2	Παράγοντας 3
Speed	0.961		-0.266
AccLat		0.582	
AccLon		0.730	
THead	-0.327		-0.14
DRight			-0.333
rdist	0.39		0.446
rspur			0.488
Wheel	0.232	0.111	
SS loadings	1.239	0.888	0.648
Ποσοστό διακύμανσης	0.155	0.111	0.081
Αθροιστικό ποσοστό	0.155	0.266	0.347

Στον πίνακα 5.27 φαίνονται οι ομάδες παραγόντων των οποίων ο βαθμός ικανότητας να εκφράσουν μια μεταβλητή ξεπερνάει το 50%.

Πίνακας 5.27: Ομάδες παραγόντων

Παράγοντας 1	Παράγοντας 2
Speed	AccLat AccLon

Από τον παραπάνω πίνακα φαίνεται πως ένα σύνολο μεταβλητών που αφορούν την κατάσταση κατά τη διάρκεια του συμβάντος μπορεί να συμπυκνωθεί σε 2 παράγοντες:

1. Τον παράγοντα που εκφράζει την επιρροή της ταχύτητας (Speed)
2. Τον παράγοντα που εκφράζει την επιρροή της πλευρικής επιτάχυνσης (AccLat) και της διαμήκου επιτάχυνσης (AccLon)

Επίσης, από τον πίνακα 5.26 φαίνεται πως ο παράγοντας 1 που εκφράζει την επιρροή ταχύτητας περιγράφει το μεγαλύτερο ποσοστό διακύμανσης.

Έλεγχος ποιότητας δεδομένων

Μετά τη διενέργεια των ελέγχων προέκυψαν τα εξής αποτελέσματα:

- $KMO = 0.4830 < 0.50$: δεν ικανοποιείται ο έλεγχος επάρκειας
- $P < 0.05$: ικανοποιείται ο έλεγχος σφαιρικότητας

5.4.3. Παραγοντική ανάλυση στον πίνακα EventsU2

Ανάλυση σε κύριες συνιστώσες

Στον πίνακα 5.28 παρουσιάζεται ο πίνακας συσχετίσεων των ανεξάρτητων μεταβλητών του πίνακα EventsU2.

Πίνακας 5.28: Πίνακας συσχετίσεων

	Speed	AccLat	AccLon	THead	DRight	rdist	rspur	Wheel
Speed	1.0000							
AccLat	-0.0049	1.0000						
AccLon	-0.0118	0.1210	1.0000					
THead	-0.2275	-0.0265	-0.0018	1.0000				
DRight	0.0082	-0.0802	0.0192	0.2496	1.0000			
rdist	0.2563	0.0139	-0.0305	-0.1995	-0.2672	1.0000		
rspur	0.0142	0.0785	-0.0253	0.0032	-0.0547	0.1067	1.0000	
Wheel	0.0627	0.0382	0.0522	-0.0253	0.0301	-0.0069	-0.0250	1.0000

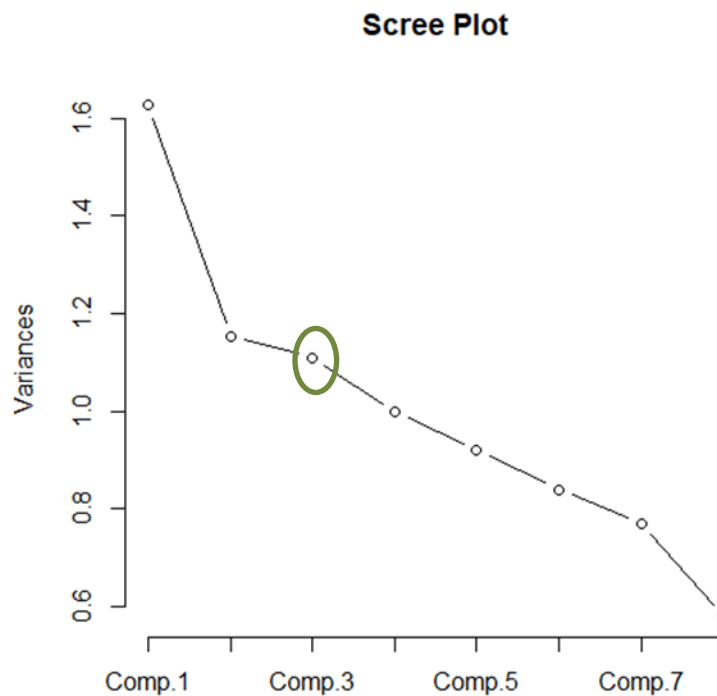
Και σε αυτόν τον πίνακα παρατηρείται πως δεν υπάρχει επαρκής συσχέτιση μεταξύ των περισσότερων ανεξάρτητων μεταβλητών.

Στον πίνακα 5.29 παρουσιάζεται η σημαντικότητα των συνιστωσών.

Πίνακας 5.29: Πίνακας σημαντικότητας συνιστωσών

	Συνιστώσα 1	Συνιστώσα 2	Συνιστώσα 3	Συνιστώσα 4	Συνιστώσα 5	Συνιστώσα 6	Συνιστώσα 7	Συνιστώσα 8
Τυπική απόκλιση	1.2763	1.0738	1.0534	0.9991	0.9590	0.9165	0.8778	0.7615
Ποσοστό διακύμανσης	0.2036	0.1441	0.1387	0.1248	0.1150	0.1050	0.0963	0.0725
Αθροιστικό ποσοστό	0.0725	0.3478	0.4865	0.6112	0.7262	0.8312	0.9275	1.0000

Σύμφωνα με το διάγραμμα της εικόνας 5.3 επιλέχθηκαν 3 παράγοντες καθώς μέχρι εκείνο το σημείο η διακύμανση είναι πάνω από τη μονάδα.



Εικόνα 5.3: Διάγραμμα διακύμανσης – συνιστωσών

Ανάλυση παραγόντων

Έγινε ανάλυση παραγόντων στον πίνακα EventsU2 για αριθμό παραγόντων ίσο με 3. Στον πίνακα 5.30 φαίνεται η μοναδικότητα των ανεξάρτητων μεταβλητών.

Πίνακας 5.30: Μοναδικότητα μεταβλητών

Speed	AccLat	AccLon	THead	DRight	rdist	rspur	Wheel
0.005	0.005	0.983	0.817	0.543	0.770	0.985	0.993

Πιθανόν οι μεταβλητές AccLon, rspur και Wheel να μην εκφράζονται μέσω παραγόντων καθώς παρουσιάζουν μεγάλη μοναδικότητα.

Στον πίνακα 5.31 παρουσιάζονται τα αποτελέσματα της ανάλυσης παραγόντων.

Πίνακας 5.31: Ανάλυση παραγόντων

	Παράγοντας 1	Παράγοντας 2	Παράγοντας 3
Speed	-0.120	0.949	-0.283
AccLat	0.981		-0.166
AccLon	0.127		
THead		-0.111	0.410
DRight		0.203	0.645
rdist		0.124	-0.458
rspur			-0.105
Wheel			
SS loadings	1.006	0.980	0.913
Ποσοστό διακύμανσης	0.126	0.123	0.114
Αθροιστικό ποσοστό	0.126	0.248	0.362

Στον πίνακα 5.31 φαίνονται οι ομάδες παραγόντων των οποίων ο βαθμός ικανότητας να εκφράσουν μια μεταβλητή ξεπερνάει το 50%.

Πίνακας 5.31: Ομάδες παραγόντων

Παράγοντας 1	Παράγοντας 2	Παράγοντας 3
AccLat	Speed	DRight

Από τον παραπάνω πίνακα φαίνεται πως ένα σύνολο μεταβλητών που αφορούν την κατάσταση πριν και κατά τη διάρκεια του συμβάντος μπορεί να συμπυκνωθεί σε 3 παράγοντες:

1. Τον παράγοντα που εκφράζει την επιρροή της πλευρικής επιτάχυνσης (AccLat)
2. Τον παράγοντα που εκφράζει την επιρροή της ταχύτητας (Speed)
3. Τον παράγοντα που εκφράζει την επιρροή της απόστασης από τη δεξιά οριογραμμή (DRight)

Επίσης, από τον πίνακα 5.31 φαίνεται πως ο παράγοντας 1 που εκφράζει την επιρροή της πλευρικής επιτάχυνσης περιγράφει το μεγαλύτερο ποσοστό διακύμανσης.

Έλεγχος ποιότητας δεδομένων

Μετά τη διενέργεια των ελέγχων προέκυψαν τα εξής αποτελέσματα:

- KMO = 0.5470 > 0.50: ικανοποιείται ο έλεγχος επάρκειας
- P < 0.05: ικανοποιείται ο έλεγχος σφαιρικότητας

5.5. Σύνοψη αποτελεσμάτων

Σε αυτό το κεφάλαιο εφαρμόστηκε η μεθοδολογία της στατιστικής ανάλυσης και παρουσιάστηκαν τα αποτελέσματα που προέκυψαν. Με σκοπό τον εντοπισμό συμβάντων αναπτύχθηκαν δύο μοντέλα στατιστικής ανάλυσης, το μοντέλο διωνυμικής λογιστικής παλινδρόμησης και το μοντέλο τυχαίων δασών. Επίσης πραγματοποιήθηκε παραγοντική ανάλυση με σκοπό την ομαδοποίηση των ανεξάρτητων μεταβλητών.

Πραγματοποιήθηκαν δύο παραλλαγές, Α και Β, σε κάθε μοντέλο στατιστικής ανάλυσης. Η παραλλαγή Α περιελάμβανε όλες τις μεταβλητές του τελικού πίνακα δεδομένων ενώ η παραλλαγή Β περιελάμβανε τις μεταβλητές που κατατάχθηκαν ως σημαντικές με βάση τη μέθοδο Boruta.

Σύμφωνα με τα αποτελέσματα προέκυψε πως το **μοντέλο διωνυμικής λογιστικής παλινδρόμησης** στην **παραλλαγή Α** δεν έδινε ικανοποιητικά αποτελέσματα καθώς είχε χαμηλό δείκτη ανάκλησης (recall) άρα δεν ήταν αξιόπιστο στον εντοπισμό της ύπαρξης συμβάντος ενώ προέβλεπε ικανοποιητικά την μη ύπαρξη συμβάντος παρουσιάζοντας υψηλή εξειδικευτικότητα (specificity). Επίσης ο δείκτης λάθος συναγερμού (false alarm rate) ήταν χαμηλός μειώνοντας την πιθανότητα λάθους ταξινόμησης των θετικών στιγμιότυπων δηλαδή της ύπαρξης συμβάντος. Η παραλλαγή Β δεν έδινε καθόλου ικανοποιητικά αποτελέσματα και δεν κρίθηκε σημαντική η ανάλυσή της. Τελικά θεωρείται πως το συγκεκριμένο μοντέλο δεν λειτουργεί σωστά για τον σκοπό της παρούσας εργασίας.

Το **μοντέλο τυχαίων δασών** λειτούργησε καλύτερα κυρίως στην **παραλλαγή Α** όπου παρουσιάζοντας μεγάλη ακρίβεια (accuracy) είχε πολλές σωστές προβλέψεις. Συγκεκριμένα προέβλεπε σε οριακά ικανοποιητικό βαθμό τα θετικά στιγμιότυπα, δηλαδή την ύπαρξη συμβάντος (recall), και σε πολύ ικανοποιητικό βαθμό τα αρνητικά στιγμιότυπα, δηλαδή την μη ύπαρξη συμβάντος (specificity). Επίσης η πιθανότητα λάθος πρόβλεψης ύπαρξης συμβάντος ήταν αρκετά χαμηλή (false alarm rate). Στην **παραλλαγή Β** το μοντέλο τυχαίων δασών θεωρήθηκε οριακά αξιόπιστο καθώς ενώ προέβλεπε πολύ ικανοποιητικά την μη ύπαρξη συμβάντος (specificity), η πρόβλεψη της ύπαρξης συμβάντος ήταν οριακά ικανοποιητική (recall). Οριακά ικανοποιητική ήταν επίσης και η λάθος πρόβλεψη ύπαρξης συμβάντος (false alarm rate).

Συμπερασματικά, η παραλλαγή Α του μοντέλου τυχαίων δασών δίνει τα καλύτερα αποτελέσματα στατιστικής ανάλυσης καθώς παρουσίασε τις πιο αξιόπιστες προβλέψεις.

Στη συνέχεια πραγματοποιήθηκε **παραγοντική ανάλυση** σε στοιχεία που αφορούσαν την κατάσταση ενός λεπτού πριν το συμβάν, σε στοιχεία που αφορούσαν την κατάσταση κατά τη διάρκεια του συμβάντος και στο σύνολό τους. Προέκυψε πως η κατάσταση ενός λεπτού πριν το συμβάν μπορεί να περιγραφεί μέσω της επιρροής της απόστασης από τη δεξιά οριογραμμή, της ταχύτητας και της απόκλισης του οχήματος από το μέσο της οδού. Η κατάσταση κατά τη διάρκεια του συμβάντος μπορεί να εκφραστεί μέσω της επιρροής της ταχύτητας και της διαμήκους επιτάχυνσης ενώ το

σύνολό τους εκφράζεται μέσω της πλευρικής επιτάχυνσης, της ταχύτητας καθώς και της απόστασης από τη δεξιά οριογραμμή.

Στο επόμενο κεφάλαιο θα παρουσιαστούν αναλυτικά τα συμπεράσματα που προκύπτουν από την εφαρμογή της μεθοδολογίας και από τα αποτελέσματα που εξήχθησαν από αυτήν.

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

6.1. Σύνοψη αποτελεσμάτων

Στόχος της παρούσας διπλωματικής εργασίας είναι ο **προσδιορισμός κρίσιμων παραγόντων εντοπισμού οδικών συμβάντων σε αστικές περιοχές**. Για το λόγο αυτό αναπτύχθηκαν μοντέλα στατιστικής ανάλυσης και διερευνήθηκε η ικανότητά τους να προβλέπουν την ύπαρξη ενός συμβάντος.

Τα δεδομένα που αναλύθηκαν προέκυψαν από πείραμα σε προσομοιωτή οδήγησης που είχε εκπονηθεί σε παλαιότερη έρευνα. Από αυτά τα δεδομένα διαμορφώθηκαν οι τελικοί πίνακες που περιείχαν στοιχεία που αφορούσαν στην κατάσταση ενός λεπτού πριν το συμβάν αλλά και στη διάρκεια του συμβάντος. Ως **εξαρτημένη** μεταβλητή καθορίστηκε η ύπαρξη ή όχι συμβάντος (Event) και ως **ανεξάρτητες** η ταχύτητα (Speed), η πλευρική (AccLat) και η διαμήκης (AccLon) επιτάχυνση, ο χρόνος μέχρι τη σύγκρουση με το προπορευόμενο όχημα (THead), η απόσταση από τη δεξιά οριογραμμή (DRight), η συνολικά διανυόμενη απόσταση (rdist), η απόκλιση του οχήματος από το μέσο της οδού (rspur), η θέση του τιμονιού (Wheel) και η οδηγική εμπειρία (Driving experience). Οι ανεξάρτητες μεταβλητές προέκυψαν μετά από έλεγχο ότι η συσχέτιση μεταξύ τους δεν είναι υψηλή.

Για τη στατιστική ανάλυση δημιουργήθηκαν το μοντέλο της **διωνυμικής λογιστικής παλινδρόμησης** και το μοντέλο των **τυχαίων δασών** και πραγματοποιήθηκαν δύο παραλλαγές για το καθένα. Η παραλλαγή A περιελάμβανε όλες τις ανεξάρτητες μεταβλητές του τελικού πίνακα δεδομένων ενώ η παραλλαγή B περιελάμβανε τις μεταβλητές που κατατάχθηκαν ως πιο σημαντικές μετά την εφαρμογή της μεθόδου Boruta που εκτίμησε τη σημαντικότητά τους. Πιο συγκεκριμένα:

- **Παραλλαγή A:** : Event (εξαρτημένη) και Speed, AccLat, AccLon, THead, DRight, rdist, rspur, Wheel, Driving experience (ανεξάρτητες)
- **Παραλλαγή B:** Event (εξαρτημένη) και Speed, THead, DRight, rdist, Wheel και Driving experience (ανεξάρτητες)

Στους πίνακες 6.1 – 6.3 παρουσιάζονται τα αποτελέσματα των μοντέλων στατιστικής ανάλυσης και ανάλυσης παραγόντων. Η παραλλαγή B του μοντέλου διωνυμικής λογιστικής παλινδρόμησης δεν έδινε καθόλου ικανοποιητικά αποτελέσματα και δεν κρίθηκε αναγκαία η ανάλυσή του.

ΥΠΟΜΝΗΜΑ	
	Πολύ ικανοποιητικό
	Ικανοποιητικό
	Οριακά ικανοποιητικό
	Καθόλου ικανοποιητικό

Πίνακας 6.1: Μοντέλο διωνυμικής λογιστικής παλινδρόμησης (Παραλλαγή Α)

Μέτρα αξιολόγησης	Ποσοστό (%)	Αξιολόγηση
Ορθότητα (Accuracy)	79.26	
Ανάκληση (Recall)	13.44	
Εξειδικευτικότητα (Specificity)	99.04	
Ακρίβεια (Precision)	80.8	
Μέτρο F (F-measure)	23.05	
Δείκτης λάθος συναγερμού (False alarm rate)	0.96	
Εμβαδόν κάτω από την καμπύλη ROC (AUC)	68.18	

Πίνακας 6.2: Μοντέλο τυχαίων δασών (Παραλλαγή Α)

Μέτρα αξιολόγησης	Ποσοστό (%)	Αξιολόγηση
Ορθότητα (Accuracy)	87.17	
Ανάκληση (Recall)	65.56	
Εξειδικευτικότητα (Specificity)	93.67	
Ακρίβεια (Precision)	75.68	
Μέτρο F (F-measure)	70.26	
Δείκτης λάθος συναγερμού (False alarm rate)	6.33	

Πίνακας 6.3: Μοντέλο τυχαίων δασών (Παραλλαγή Β)

Μέτρα αξιολόγησης	Ποσοστό (%)	Αξιολόγηση
Ορθότητα (Accuracy)	81.19	
Ανάκληση (Recall)	53.51	
Εξειδικευτικότητα (Specificity)	89.51	
Ακρίβεια (Precision)	60.53	
Μέτρο F (F-measure)	56.81	
Δείκτης λάθος συναγερμού (False alarm rate)	10.49	

Επίσης, πραγματοποιήθηκε **παραγοντική ανάλυση** στους πίνακες που αφορούσαν στην κατάσταση ενός λεπτού πριν το συμβάν, στη διάρκεια του συμβάντος και στο σύνολο αυτών για τη διερεύνηση ύπαρξης κοινών παραγόντων (Πίνακας 6.4).

Πίνακας 6.4: Αποτελέσματα παραγοντικής ανάλυσης

Πίνακας	Παράγοντας 1	Παράγοντας 2	Παράγοντας 3
PreEventU2 (1 λεπτό πριν το συμβάν)	DRight	Speed	rspur
DurEventU2 (διάρκεια συμβάντος)	Speed	AccLat AccLon	
EventsU2 (σύνολο των παραπάνω)	AccLat	Speed	DRight

6.2. Συνολικά συμπεράσματα

Σύμφωνα με τα αποτελέσματα που προέκυψαν στο προηγούμενο κεφάλαιο από την εφαρμογή της μεθοδολογίας μπορούν να διατυπωθούν τα **συμπεράσματα** της παρούσας διπλωματικής εργασίας.

- ❖ Οι **μεταβλητές με τη μεγαλύτερη σημαντικότητα** για τον εντοπισμό ύπαρξης συμβάντων σε αστικές οδούς είναι η ταχύτητα, η συνολική διανυόμενη απόσταση, η απόσταση από τη δεξιά οριογραμμή, η θέση του τιμονιού, η οδηγική εμπειρία και ο χρόνος μέχρι τη σύγκρουση με το προπορευόμενο όχημα.
- ❖ Το **μοντέλο διωνυμικής λογιστικής παλινδρόμησης** με όλες τις σημαντικές μεταβλητές (Παραλλαγή Α) κρίθηκε **μη αξιόπιστο** καθώς ενώ παρουσίασε μεγάλη ορθότητα, δηλαδή συνολικά σωστές προβλέψεις, δεν προέβλεπε ικανοποιητικά την ύπαρξη συμβάντος (χαμηλός δείκτης ανάκλησης), που ήταν και ο στόχος της παρούσας διπλωματικής. Το ίδιο μοντέλο για μικρό αριθμό ανεξάρτητων μεταβλητών (Παραλλαγή Β) δεν λειτούργησε καθόλου ικανοποιητικά και δεν κρίθηκε άξιο περαιτέρω ανάλυσης.
- ❖ Το **μοντέλο τυχαίων δασών** παρουσίασε **αξιόπιστα αποτελέσματα** και στις δύο παραλλαγές του. Προέκυψαν υψηλοί δείκτες ορθότητας (σωστή πρόβλεψη ύπαρξης ή όχι συμβάντος) και εξειδικευτικότητας (πρόβλεψη μη ύπαρξης συμβάντος) και στις δύο παραλλαγές. Η **παραλλαγή Α κρίθηκε πιο αποτελεσματική** καθώς είχε αυξημένη πιθανότητα πρόβλεψης ύπαρξης συμβάντος (ανάκληση) και μειωμένη πιθανότητα λάθους πρόβλεψης ύπαρξης συμβάντος (δείκτης λάθος συναγερμού).

- ❖ Το μοντέλο τυχαίων δασών βρέθηκε να δίνει πιο αξιόπιστα αποτελέσματα στον εντοπισμό συμβάντων σε σχέση με το μοντέλο διωνυμικής λογιστικής παλινδρόμησης ίσως επειδή δεν θεωρεί ότι το μοντέλο έχει γραμμική σχέση όπως τα μοντέλα παλινδρόμησης και εξετάζει τυχαία δείγματα με στόχο την υψηλότερη ακρίβεια των αποτελεσμάτων.
- ❖ Και στα δύο μοντέλα στατιστικής ανάλυσης, η **παραλλαγή με τον μεγαλύτερο αριθμό μεταβλητών** παρουσίασε πιο ικανοποιητικά αποτελέσματα.
- ❖ Τα αποτελέσματα της παραγοντικής ανάλυσης έδειξαν ότι η κατάσταση που αφορά στο **ένα λεπτό πριν το συμβάν** μπορεί να περιγραφεί μέσω τριών παραγόντων: i) της επιρροής της απόστασης από τη δεξιά οριογραμμή, ii) της ταχύτητας και iii) της απόκλισης του οχήματος από το μέσο της οδού. Η ταχύτητα είναι καθοριστικός παράγοντας για την πιθανότητα εμπλοκής σε ένα συμβάν αλλά και οι αποστάσεις του οχήματος από την δεξιά οριογραμμή και το μέσο της οδού παίζουν σημαντικό ρόλο καθώς σε ένα απροσδόκητο συμβάν το όχημα μπορεί να βρεθεί εκτός πορείας.
- ❖ Όσον αφορά στην κατάσταση **κατά τη διάρκεια του συμβάντος** οι παράγοντες που μπορούν να την περιγράψουν είναι δύο: i) ένας που περιγράφει την επιρροή της ταχύτητας και ii) ένας που περιγράφει την επιρροή της πλευρικής και της διαμήκους επιτάχυνσης. Όπως και η ταχύτητα έτσι και η πλευρική και η διαμήκης επιτάχυνση είναι παράγοντες με καθοριστική σημασία στη δυναμική της κίνησης του οχήματος και συνακόλουθα στην πιθανότητα εμπλοκής σε μη αναμενόμενο συμβάν.
- ❖ Από την παραγοντική ανάλυση επίσης προέκυψε πως τα δεδομένα που αφορούν στην κατάσταση **ενός λεπτού πριν και κατά τη διάρκεια του συμβάντος** μπορούν να περιγραφούν από τρεις παράγοντες: i) την επιρροή της πλευρικής επιτάχυνσης, ii) την επιρροή της ταχύτητας και iii) την επιρροή της απόστασης από τη δεξιά οριογραμμή καθώς αποτελούν το σύνολο των παραπάνω.

6.3. Προτάσεις για αξιοποίηση των αποτελεσμάτων

Παρακάτω παρατίθενται **προτάσεις** αξιοποίησης των αποτελεσμάτων της παρούσας διπλωματικής εργασίας με κύριο στόχο τη βελτίωση της οδικής ασφάλειας.

- ❖ Δημιουργία **εφαρμογής σε έξυπνο κινητό τηλέφωνο** η οποία εξετάζοντας τη μεταβολή των οδηγικών χαρακτηριστικών θα προειδοποιεί τον οδηγό για πιθανή εμπλοκή σε συμβάν σε πραγματικό χρόνο ή θα τον ενημερώνει για τα συμβάντα που προέκυψαν κατά τη διαδρομή του μετά το πέρας αυτής.

- ❖ Ανάπτυξη εφαρμογής σε έξυπνο κινητό τηλέφωνο η οποία, με τη βοήθεια του αλγορίθμου εντοπισμού συμβάντων, θα έχει τη δυνατότητα να εντοπίζει την πιθανότητα ύπαρξη συμβάντος και να θέτει σε σίγαση τυχόν εισερχόμενα μηνύματα ή ειδοποιήσεις ώστε **να μην αποσπάται η προσοχή του οδηγού** και να είναι σε θέση να διαχειριστεί δύσκολες συνθήκες οδήγησης.
- ❖ Εξοπλισμός των οχημάτων με συστήματα που αξιοποιώντας τον αλγόριθμο θα μπορούν να προβλέψουν την ύπαρξη συμβάντος και θα εμφανίζουν **προειδοποιητικό μήνυμα** στον οδηγό να εκτελέσει κατάλληλες ενέργειες για την αποφυγή του.
- ❖ Αξιοποίηση του αλγορίθμου εντοπισμού συμβάντων με βάση τα οδηγικά χαρακτηριστικά από **Κέντρα Διαχείρισης Κυκλοφορίας** με στόχο τον έγκαιρο εντοπισμό των συμβάντων και την συμβολή για την αντιμετώπισή τους.
- ❖ Δημιουργία **χαρτών** που θα απεικονίζουν την πυκνότητα των συμβάντων στο οδικό δίκτυο.

6.4. Προτάσεις για περαιτέρω έρευνα

Από τη βιβλιογραφική ανασκόπηση καθώς και την παρούσα διπλωματική εργασία γίνονται εμφανείς κάποιες ελλείψεις και απουσίες μελέτης ορισμένων αντικειμένων για τον εντοπισμό των συμβάντων με βάση τα οδηγικά χαρακτηριστικά. Παρακάτω γίνονται κάποιες προτάσεις για περαιτέρω έρευνα που θα συμβάλλουν σε μια ολοκληρωμένη μελέτη του συγκεκριμένου αντικειμένου.

- ❖ Επιλογή **διαφορετικών μεθοδολογικών προσεγγίσεων** για την ανάπτυξη μοντέλων που θα προβλέπουν την ύπαρξη συμβάντων καθώς είναι πιθανό να προκύψουν αξιόπιστα αποτελέσματα.
- ❖ Εξέταση **μεγαλύτερου δείγματος** συμμετεχόντων από εκείνο της παρούσας εργασίας, διότι πιθανόν να προκύπτουν διαφορές στην επιρροή των μεταβλητών στα μοντέλα.
- ❖ Εξέταση **επιπλέον τύπων οδού** εκτός από το αστικό περιβάλλον και σύγκριση των αποτελεσμάτων.
- ❖ Εξέταση **εναλλακτικών σεναρίων οδήγησης με διαφορετικές συνθήκες**, για παράδειγμα με υψηλό ή χαμηλό κυκλοφοριακό φόρτο, με βροχή ή ομίχλη, με απόσπαση προσοχής ή χωρίς.
- ❖ Αξιοποίηση **δεδομένων οδήγησης σε πραγματικές συνθήκες** (naturalistic driving data), αντί των δεδομένων του προσομοιωτή οδήγησης, για πιο αξιόπιστα αποτελέσματα.

- ❖ Μελέτη της μεταβολής των οδηγικών χαρακτηριστικών σε **μεγαλύτερο χρονικό διάστημα του ενός λεπτού πριν** το συμβάν.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Φραντζεσκάκης Ι. Μ., Γκόλιας Ι. Κ., “Οδική Ασφάλεια”, Εκδόσεις Παπασωτηρίου, Αθήνα, 1994
- [2] Osman O., Hajjib M., Karbalaieali S., Ishak S., “A hierarchical machine learning classification approach for secondary task identification from observed driving behavior data”, Accident Analysis and Prevention 123 Louisiana State University LA, United States, 2019
- [3] Choudhary P., Velaga N. R., “Effects of phone use on driving performance: A comparative analysis of young and professional drivers”, Safety Science 111 Indian Institute of Technology Mumbai, India, 2019
- [4] Yannis G., Laiou A., Papantoniou P., Christoforou C., “Impact of texting on young drivers' behavior and safety on urban and rural roads through a simulation experiment”, Journal of Safety Research 49 National Technical University of Athens, Greece, 2014
- [5] Choudhary P., Velaga N. R., “Mobile phone use during driving: Effects on speed and effectiveness of driver compensatory behavior”, Accident Analysis and Prevention 106 Indian Institute of Technology Mumbai, India, 2017
- [6] Dimitriou L., Stylianou K., Yannis G., “Capturing the effects of texting on young drivers behavior based on copula and Gaussian Mixture Models”, Transportation Research Part F 58 University of Cyprus, Cyprus, 2018
- [7] Choudhary P., Velaga N. R., “A comparative analysis of risk associated with eating, drinking and texting during driving at unsignalised intersections”, Transportation Research Part F 63 Indian Institute of Technology Mumbai, India, 2019
- [8] Choudhary P., Velaga N. R., “Effects of texting on accident risk during a sudden hazardous event: Analysis of predetection and postdetection phases”, Traffic Injury Prevention Indian Institute of Technology Mumbai, India, 2018
- [9] Papantoniou P., Pavlou D., Yannis G., Vlachogianni E., “How an unexpected incident affects speed related driving performance measures”, Transportation Research Procedia 41 National Technical University of Athens, Greece, 2018
- [10] Papantoniou P., Antoniou C., Yannis G., Pavlou D., “Which factors affect accident probability at unexpected incidents? A structural equation model approach”, Journal of Transportation Safety & Security National Technical University of Athens, Greece, 2019

- [11] Li X., Rakotonirainy A., Yan X., “How do drivers avoid collisions? A driving simulator-based study”, Journal of Safety Research 70 School of Traffic and Transportation Beijing, China, 2019
- [12] Monselise M., Liang S. O., Yang C. C., “Identifying Important Risk Factors Associated with Vehicle Injuries Using Driving Behavior Data and Predictive Analytics”, 2019
- [13] Papantoniou P., Yannis G., Christofa E., “Which factors lead to driving errors? A structural equation model analysis through a driving simulator experiment”, IATSS Research 43 National Technical University of Athens, Greece, 2019

ΠΑΡΑΡΤΗΜΑ

Κώδικας ανάλυσης στο R-Studio

#Φόρτωση πακέτου για χρήση της εντολής filter

```
install.packages ("dplyr")  
library (dplyr)
```

#Δημιουργία πίνακα index (χρόνοι έναρξης και λήξης κάθε συμβάντος)

```
index<-data.frame()  
k<-0  
for (i in 1:1359815) {  
  if (UrbanControl[i,11]==0 & UrbanControl[i+1,11]!=0) {  
    k<-k+1  
    index[k,1]<-UrbanControl[i,2]  
    index[k,2]<-UrbanControl[i+1,4]  
  }  
  if(UrbanControl[i,11]!=0 & UrbanControl[i+1,11]==0){  
    index[k,3]<-UrbanControl[i,4]  
  }  
}
```

#Δημιουργία πίνακα πριν τα event(PreEventU)

```
PreEventU<-data.frame()  
EVENTS<-data.frame()  
ID<-data.frame()  
x<-unique(index[,1])  
y<-0  
f<-0  
k<-0  
l<-0  
for (y in x) {  
  a<-0  
  b<-0  
  ID<-filter(index,PersonID==y)  
  f<-nrow(ID)  
  k<-k+f  
  l<-k-f+1  
  for(i in l:k) {  
    a<-index[i,2]-60  
    b<-index[i,2]
```

```

EVENTS<-filter(UrbanControl,Time>=a & Time<=b & PersonID==y)
PreEventU<-rbind(PreEventU,EVENTS)
}
}

```

#Δημιουργία πίνακα των events (EventsU)

```

EventsU<-data.frame()
EVENTS<-data.frame()
ID<-data.frame()
x<-unique(index[,1])
y<-0
f<-0
k<-0
l<-0
for (y in x) {
  a<-0
  b<-0
  ID<-filter(index,PersonID==y)
  f<-nrow(ID)
  k<-k+f
  l<-k-f+1
  for(i in l:k) {
    a<-index[i,2]-60
    b<-index[i,3]
    EVENTS<-filter(UrbanControl,Time>=a & Time<=b & PersonID==y)
    EventsU<-rbind(EventsU,EVENTS)
  }
}

```

#Δημιουργία πίνακα κατά τη διάρκεια των event (DurEventU)

```
DureventU<-filter(EventsU,Event!=0)
```

#Drop N/A των στηλών AccLon, AccLat, Driving experience

```

events1<-data.frame()
events2<-data.frame()
events3<-data.frame()
events1<-EventsU[complete.cases(EventsU$AccLon), ]
events2<-events1 [complete.cases(events1$AccLat), ]
events3<-events2[complete.cases(events2$`Driving experience`), ]
EventsU<-events3
preEvent1<-data.frame()
preEvent2<-data.frame()
preEvent3<-data.frame()

```

```

preEvent1<-PreEventU[complete.cases(PreEventU$AccLon), ]
preEvent2<- preEvent1 [complete.cases(preEvent1$AccLat), ]
preEvent3<- preEvent2[complete.cases(preEvent2$`Driving experience`), ]
PreEventU<-preEvent3
durEvent1<-data.frame()
durEvent2<-data.frame()
durEvent3<-data.frame()
durEvent1<-DurEventU[complete.cases(DurEventU$AccLon), ]
durEvent2<- durEvent1 [complete.cases(durEvent1$AccLat), ]
durEvent3<- durEvent2[complete.cases(durEvent2$`Driving experience`), ]
DurEventU<-durEvent3

```

#Αφαίρεση κοινών γραμμών

```

ID<-data.frame()
ID2<-data.frame()
events<-data.frame()
for (y in x) {
  ID<-filter(EventsU,PersonID==y)
  ID2<-ID %>% distinct(Time, .keep_all = TRUE)
  events<-rbind(events,ID2)
}
EventsU<-events
PreEventU<-filter(EventsU,Event==0)

```

#Περιγραφική Στατιστική

```

summary(PreEventU)
sapply(PreEventU,var,na.rm=TRUE)
sapply(PreEventU,sd,na.rm=TRUE)
summary(DurEventU)
sapply(DurEventU,var,na.rm=TRUE)
sapply(DurEventU,sd,na.rm=TRUE)

```

#Δημιουργία πίνακα συσχέτισης (correlation)

```

correlation<-data.frame()
event<-EventsU[,c(8,9,10,13,14,15,16,17,18,19,20,21,25)]
correlation<- cor(event)

```

#Δημιουργία πίνακα μοντέλων στατιστικής ανάλυσης

```

Model<-data.frame()
Model<-EventsU[,c(11,8,9,10,14,17,18,19,20,25)]
for (i in 1:438126) {
  if (Model[i,1]!=0) {

```

```

    Model[i,1]=1}
}

#Μοντέλο Logistic regression (Παραλλαγή A)

#Διαχωρισμός Model σε training (80%) και test (20%)

training1<-Model[1:350500,]
test1<-Model[350501:438126,]

#Μετατροπή της μεταβλητής Event σε διακριτή

training1$Event<-as.factor(training1$Event)
test1$Event<-as.factor(test1$Event)

#Διωνυμική λογιστική παλινδρόμηση

model<-data.frame()
model<-glm(Event~.,data=training1,family=binomial)

#Έλεγχος συντελεστών παλινδρόμησης

summary(model)

#Πρόβλεψη στον πίνακα test

fitted.results <- predict(model,newdata=test1,type='response')
fitted.results <- ifelse(fitted.results > 0.5,1,0)
misClasificError <- mean(fitted.results != test1$Event)
print(paste('Accuracy',1-misClasificError))

#Confusion matrix

install.packages("caret")
library(caret)
pred<-as.data.frame(fitted.results)
pred$fitted.results<-factor(pred$fitted.results)
actual<-test1[,1]
actual$Event<-factor(actual$Event)
confusionMatrix(pred$fitted.results,actual$Event,positive ="1")

#Δημιουργία καμπύλης ROC

install.packages("ROCR")
library(ROCR)
p <- predict(model, newdata=test1, type="response")

```

```
pr <- prediction(p, test1$Event)
prf <- performance(pr, measure = "tpr", x.measure = "fpr")
plot(prf)
```

#Εμβαδόν κάτω από την καμπύλη ROC

```
auc <- performance(pr, measure = "auc")
auc <- auc@y.values\[\[1\]\]
```

#Feature importance

```
install.packages("Boruta")
library(Boruta)
boruta_output <- Boruta(Event ~ ., data=na.omit(training1), doTrace=0)
boruta_signif <- getSelectedAttributes(boruta_output, withTentative = TRUE)
print(boruta_signif)
roughFixMod <- TentativeRoughFix(boruta_output)
boruta_signif <- getSelectedAttributes(roughFixMod)
print(boruta_signif)
imps <- attStats(roughFixMod)
imps2 = imps[imps$decision != 'Rejected', c('meanImp', 'decision')]
head(imps2[order(-imps2$meanImp), ])
plot(boruta_output, cex.axis=.7, las=2, xlab="", main="Variable Importance")
```

#Μοντέλο Random Forest

```
install.packages('randomForest')
library(randomForest)
memory.limit()
memory.limit(100000)
```

#Αλλαγή του ονόματος της στήλης Driving Experience σε Drivingexperience

```
names(training1)[names(training1) == "Driving experience"] <- "Drivingexperience"
names(test1)[names(test1) == "Driving experience"] <- "Drivingexperience"
```

#Τυχαία δάση (Παραλλαγή A)

#Εκπαίδευση

```
set.seed(222)
rf<-randomForest(Event~.,data=training1,ntree=100)
```

#Ελεγχος

```
p1<-predict(rf,test1)
```

```
confusionMatrix(p1,testing$Event,positive = "1")
```

```
#Τυχαία δάση (Παραλλαγή Β)
```

```
Model2<-Model[,c(1,2,5,6,7,9,10)]
```

```
#Διαχωρισμός Model2 σε training (80%) και test (20%)
```

```
training2<-Model2[1:350500,]
```

```
test2<-Model2[350501:438126,]
```

```
#Μετατροπή της μεταβλητής Event σε διακριτή
```

```
training2$Event<-as.factor(training2$Event)
```

```
test2$Event<-as.factor(test2$Event)
```

```
#Εκπαίδευση
```

```
set.seed(222)
```

```
rf<-randomForest(Event~.,data=training2,ntree=100)
```

```
print(rf)
```

```
attributes(rf)
```

```
plot(rf)
```

```
#Έλεγχος
```

```
p2<-predict(rf,test2)
```

```
confusionMatrix(p2,test2$Event,positive = "1")
```

```
#Factor Analysis
```

```
install.packages("REdaS")
```

```
library(REdaS)
```

```
install.packages("grid")
```

```
#Pre-Event
```

```
preEvent <-filter(Model,Event==0)
```

```
PreEventU2<-preEvent[,c(2,3,4,5,6,7,8,9)]
```

```
summary(PreEventU2)
```

```
cor(PreEventU2)
```

```
#Κύριες συνιστώσες για την εκτίμηση του αριθμού των παραγόντων PCA
```

```
pca1<-princomp(PreEventU2,scores=TRUE,cor=TRUE)
```

```
summary(pca1)
```

```
loadings(pca1)
plot(pca1)
screeplot(pca1,type="line",main="Scree Plot")
pca1$scores[1:10,]
```

#Ανάλυση παραγόντων

```
fa1<-factanal(PreEventU2,factors=3,rotation="varimax",scores="regression")
fa1
```

#KMO Test

```
KMOS(PreEventU2)
```

#Bartlett Test

```
bart_spher(PreEventU2,use=c("everything","all.obs","complete.obs","na.or.complete","p
airwise.complete.obs"))
```

#Dur-Event

```
durEvent <-filter(Model,Event==1)
DurEventU2<-durEvent[,c(2,3,4,5,6,7,8,9)]
summary(DurEventU2)
cor(DurEventU2)
```

#Κύριες συνιστώσες για την εκτίμηση του αριθμού των παραγόντων PCA

```
pca2<-princomp(DurEventU2,scores=TRUE,cor=TRUE)
summary(pca2)
loadings(pca2)
plot(pca2)
screeplot(pca2,type="line",main="Scree Plot")
pca2$scores[1:10,]
```

#Ανάλυση παραγόντων

```
fa2<-factanal(DurEventU2,factors=3,rotation="varimax",scores="regression")
fa2
```

#KMO Test

```
KMOS(DurEventU2)
```

#Bartlett Test

```
bart_spher(DurEventU2,use=c("everything","all.obs","complete.obs","na.or.complete","pairwise.complete.obs"))
```

#Events

```
Events <-filter(Model,Event==1)
EventsU2<-Events[,c(2,3,4,5,6,7,8,9)]
summary(EventsU2)
cor(EventsU2)
```

#Κύριες συνιστώσες για την εκτίμηση του αριθμού των παραγόντων PCA

```
pca3<-princomp(EventsU2,scores=TRUE,cor=TRUE)
summary(pca3)
loadings(pca3)
plot(pca3)
screeplot(pca3,type="line",main="Scree Plot")
pca3$scores[1:10,]
```

#Ανάλυση παραγόντων

```
fa3<-factanal(EventsU2,factors=3,rotation="varimax",scores="regression")
fa3
```

#KMO Test

```
KMOS(EventsU2)
```

#Bartlett Test

```
bart_spher(EventsU2,use=c("everything","all.obs","complete.obs","na.or.complete","pairwise.complete.obs"))
```