



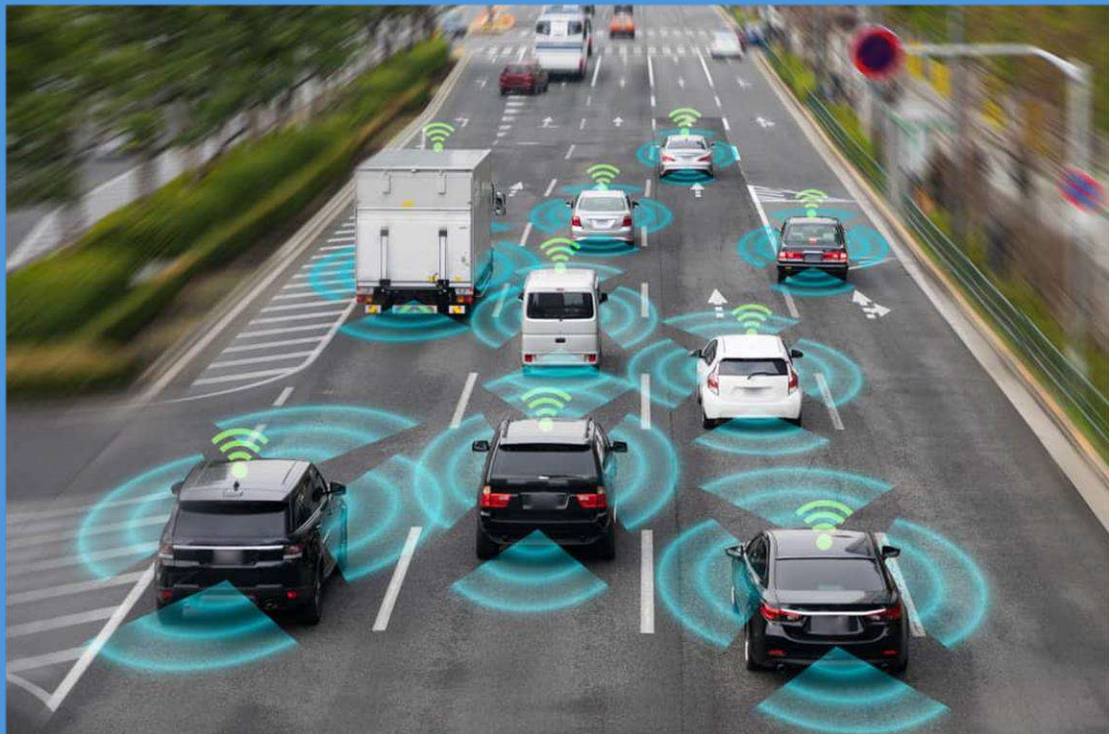
ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ

ΠΡΟΒΛΕΨΗ ΤΗΣ ΟΔΗΓΙΚΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΜΕ ΔΕΔΟΜΕΝΑ ΑΠΟ ΔΙΑΦΟΡΕΤΙΚΕΣ ΧΩΡΕΣ ΚΑΙ ΧΡΗΣΗ ΤΕΧΝΙΚΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Διπλωματική Εργασία



Γιάννης Ρούκος

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

Αθήνα, Μάρτιος 2025

ΕΥΧΑΡΙΣΤΙΕΣ

Με την παρουσίαση της συγκεκριμένης Διπλωματικής Εργασίας, ολοκληρώνεται ο κύκλος των προπτυχιακών μου σπουδών 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 που παρείχε καθ' όλη τη διάρκεια της εκπόνησης της.

Επίσης, θα ήθελα να ευχαριστήσω ιδιαίτερα τον ερευνητή Θοδωρή Γαρεφαλάκη, για την καθοριστική συνεισφορά του στη διεκπεραίωση της παρούσας εργασίας, μέσω των πολύτιμων συμβουλών και υποδείξεων του, καθώς και για το εξαιρετικό κλίμα συνεργασίας και επικοινωνίας που διαμόρφωσε. Τέλος, θα ήθελα να ευχαριστήσω τους φίλους μου και κυρίως τους γονείς μου που με στήριξαν καθ' όλη τη διάρκεια των σπουδών μου.

Αθήνα, Μάρτιος 2025

Γιάννης Ρούκος

Πρόβλεψη της οδηγικής συμπεριφοράς με δεδομένα από διαφορετικές χώρες και χρήση τεχνικών μηχανικής μάθησης

Γιάννης Ρούκος

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

Σύνοψη

Στόχο της παρούσας Διπλωματικής Εργασίας αποτελεί η ανάπτυξη μοντέλων ταξινόμησης για την αναγνώριση της επικίνδυνης συμπεριφοράς του οδηγού και την κατάταξη της σε τρία διαφορετικά επίπεδα ασφαλείας. Για τον σκοπό αυτό, συγκεντρώθηκαν πολύτιμα δεδομένα σχετικά με τη συμπεριφορά των οδηγών μέσω ενός πειράματος οδήγησης που πραγματοποιήθηκε σε πραγματικές συνθήκες στο Βέλγιο και το Ηνωμένο Βασίλειο. Σε πρώτη ανάλυση υπολογίστηκε η σημαντικότητα των μεταβλητών με τον αλγόριθμο “Τυχαίου Δάσους”, με βάση την οποία επιλέχθηκαν εννιά μεταβλητές εισόδου που αξιοποιήθηκαν στη συνέχεια στις περαιτέρω αναλύσεις. Στη συνέχεια, για να αντιμετωπιστεί το πρόβλημα της ανισορροπίας δεδομένων εφαρμόστηκε η μέθοδος υπερδειγματοληψίας SMOTE. Με την ολοκλήρωση των δύο παραπάνω βημάτων, αναπτύχθηκαν τέσσερις αλγόριθμοι ταξινόμησης της οδηγικής συμπεριφοράς, για τους οποίους υπολογίστηκαν οι Μήτρες Σύγχυσης και στη συνέχεια έγινε σύγκριση των μετρικών αξιολόγησης τους. Έπειτα, εξετάστηκαν οι τιμές SHAP για να κατανοηθεί περαιτέρω η επιρροή της κάθε μεταβλητής εισόδου που επιλέχθηκε. Με αυτόν τον τρόπο, υπολογίστηκε η μέση σημαντικότητα SHAP, όπου και επιλέχθηκαν τα μοντέλα CatBoost και LightGBM ως αποτελεσματικότερα. Τέλος, αποτυπώθηκε η σημαντικότητα των μεταβλητών για τα δύο επιλεγμένα μοντέλα μέσω της μεθόδου SHAP για τα τρία επίπεδα ασφαλείας και αναλύθηκε ξεχωριστά η επιρροή της κάθε μεταβλητής στην αλλαγή του επιπέδου ασφαλείας. Η μέση ταχύτητα του οχήματος υπολογίστηκε ως η σημαντικότερη μεταβλητή, ενώ τα απότομα περιστατικά οδήγησης, τόσο της απότομης επιτάχυνσης, όσο και του απότομου φρεναρίσματος, αποδείχτηκε πως επηρεάζουν σημαντικά την κατάταξη της οδηγικής συμπεριφοράς ως επικίνδυνη.

Λέξεις-Κλειδιά: ανάλυση οδηγικής συμπεριφοράς, πρόβλεψη ατυχημάτων, μηχανική μάθηση, βαθιά μάθηση, μοντέλα ταξινόμησης, μέθοδος επαναδειγματοληψίας, τυχαία δάση, ελαφρύς αλγόριθμος βαθιάς ενίσχυσης, αλγόριθμος ενίσχυσης κατηγοριών, πολυεπίπεδο αντιληπτικό, μέθοδος SHAP (SHapley Additive Explanations)

Predicting driver behaviour in a cross-country naturalistic driving study using machine learning techniques

Giannis Roukos

Supervisor: George Yannis, Professor NTUA

Abstract

The aim of this Diploma Thesis is to develop classification models for identifying risky driving behavior and categorizing it into three different safety levels. To achieve this, valuable data on driver behavior were collected through a driving experiment conducted under real-world conditions in Belgium and the United Kingdom. In the initial analysis, the importance of the variables was calculated using the "Random Forest" algorithm, based on which nine input variables were selected for further analysis. Then, to address the issue of data imbalance, the SMOTE oversampling method was applied. Upon completing these two steps, four classification algorithms for driving behavior were developed, for which Confusion Matrices were calculated, followed by a comparison of their evaluation metrics. Subsequently, SHAP values were examined to further understand the influence of each selected input variable. This approach allowed for the calculation of the average SHAP importance, leading to the selection of the CatBoost and LightGBM models as the most effective. Finally, the importance of the variables for the two selected models was visualized using the SHAP method for the three safety levels, and the influence of each variable on changes in safety level was analyzed separately. The average speed of the vehicle was identified as the most significant variable, while sudden driving events, including both harsh acceleration and harsh braking, were found to significantly influence the classification of driving behavior as dangerous.

Keywords: driving behavior analysis, real-time accident prediction, machine learning, deep learning, classification models, resampling method, random forests, light gradient boosting machine (LightGBM), categorical boosting algorithm (CatBoost), multilayer perceptron, SHAP method (SHapley Additive Explanations).

ΠΕΡΙΛΗΨΗ

Στην παρούσα Διπλωματική Εργασία **διερευνάται η επικίνδυνη οδηγική συμπεριφορά μέσω δεδομένων οδήγησης σε πραγματικές συνθήκες**, τα οποία συλλέχθηκαν από το Βέλγιο και το Ηνωμένο Βασίλειο. Πιο συγκεκριμένα, η συμπεριφορά του οδηγού κατατάσσεται σε τρία διαφορετικά επίπεδα επικίνδυνης συμπεριφοράς με βάση τον δείκτη “Ζώνη Ανοχής Ασφαλείας”. Τα δεδομένα προέρχονται από ένα πείραμα φυσική οδήγησης, στο πλαίσιο του ερευνητικού προγράμματος i-DREAMS. Συγκεκριμένα, στη μελέτη συμμετείχαν 42 οδηγοί από το Βέλγιο, δημιουργώντας μια εκτενή βάση δεδομένων με 813 ταξίδια και συνολική διάρκεια 21.412 λεπτών. Αντίστοιχα, για το Ηνωμένο Βασίλειο, το σύνολο δεδομένων περιλάμβανε 54 οδηγούς, καταγράφοντας 3.317 ταξίδια και 58.458 λεπτά οδήγησης.

Βασικό μέρος της μεθοδολογίας αποτελεί η διαδικασία επιλογής χαρακτηριστικών. Στόχος της διαδικασίας είναι η μείωση του αριθμού των μεταβλητών εισόδου με ταυτόχρονη μείωση του υπολογιστικού κόστους του μοντέλου και βελτίωση της προγνωστικής του απόδοσης. Η επιλογή των χαρακτηριστικών γίνεται με γνώμονα την επιρροή της μεταβλητής στην διαδικασία της ταξινόμησης. Σε πρώτη φάση, η σημαντικότητα των μεταβλητών υπολογίζεται **με τον αλγόριθμο “Τυχαίου Δάσους”** (Random Forest Classifier). Αναφέρεται πως, καθώς επιλέγεται να γίνει **ο καθορισμός του επιπέδου ασφαλείας** με την χρήση της μεταβλητής **Headway**, τα διάφορα περιγραφικά στατιστικά στοιχεία του παράγοντα Headway δεν αποτελούν μεταβλητές εισόδου στα μοντέλα.

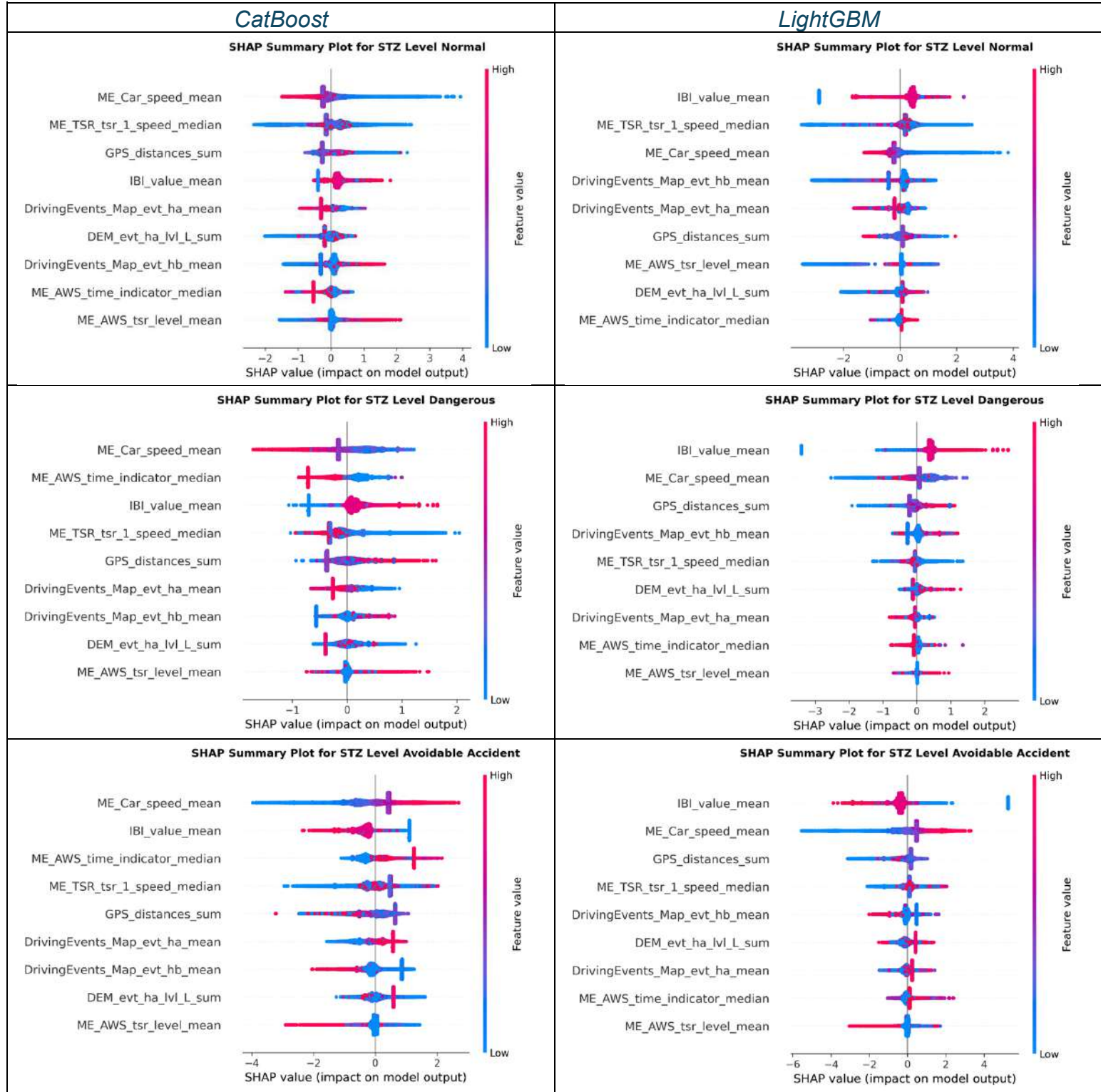
Στη συνέχεια, επιλέγεται ένα σύνολο **εννιά μεταβλητών εισόδου**, κοινό και για τα δύο datasets Βελγίου κι Ηνωμένου Βασιλείου. Για να αντιμετωπιστεί η ανισορροπία των δεδομένων μεταξύ των διαφορετικών κατηγοριών, χρησιμοποιείται η μέθοδος **SMOTE (Synthetic Minority Over-sampling Technique)**. Μετά την εφαρμογή του SMOTE, αναπτύσσονται τέσσερις αλγόριθμοι μηχανικής μάθησης με σκοπό την ταξινόμηση της οδηγικής συμπεριφοράς σε ένα από τα τρία επίπεδα ασφαλείας. Έπειτα, παρουσιάζονται για κάθε μοντέλο ταξινόμησης και για τις δύο χώρες οι μήτρες σύγχυσης για την γραφική αναπαράσταση της επίδοσης τους. Μετά την ολοκλήρωση των παραπάνω, γίνεται η σύγκριση διάφορων μετρικών αξιολόγησης των τεσσάρων μοντέλων.

Στη συνέχεια, για να κατανοήσουμε καλύτερα τα μοντέλα μηχανικής μάθησης, εξετάζονται οι τιμές **SHAP** (SHapley Additive exPlanations). Το **SHAP** χρησιμοποιείται για να ερμηνεύσει την επίδραση κάθε χαρακτηριστικού στο μοντέλο, επιτρέποντας τόσο την ανάλυση σε τοπικό όσο και σε συνολικό επίπεδο. Τα μοντέλα **CatBoost** και **LightGBM** εμφανίζουν τη μεγαλύτερη μέση σημαντικότητα SHAP για τα δεδομένα και των δύο χωρών κι έτσι επιλέγονται για εκ νέου εύρεση της σημαντικότητας των μεταβλητών για κάθε επίπεδο

επικίνδυνης οδηγικής συμπεριφοράς. Στους παρακάτω πίνακες, συγκεντρώνονται τα διαγράμματα που παρουσιάζονται στο κεφάλαιο 5, και δείχνουν την **επιρροή των εννιά επιλεγμένων μεταβλητών** στο κάθε επίπεδο επικινδυνότητας της οδηγικής συμπεριφοράς, με τη **μέθοδο SHAP**.

Ηνωμένο Βασίλειο

Πίνακας 1: Διαγράμματα μεθόδου SHAP για το Ηνωμένο Βασίλειο

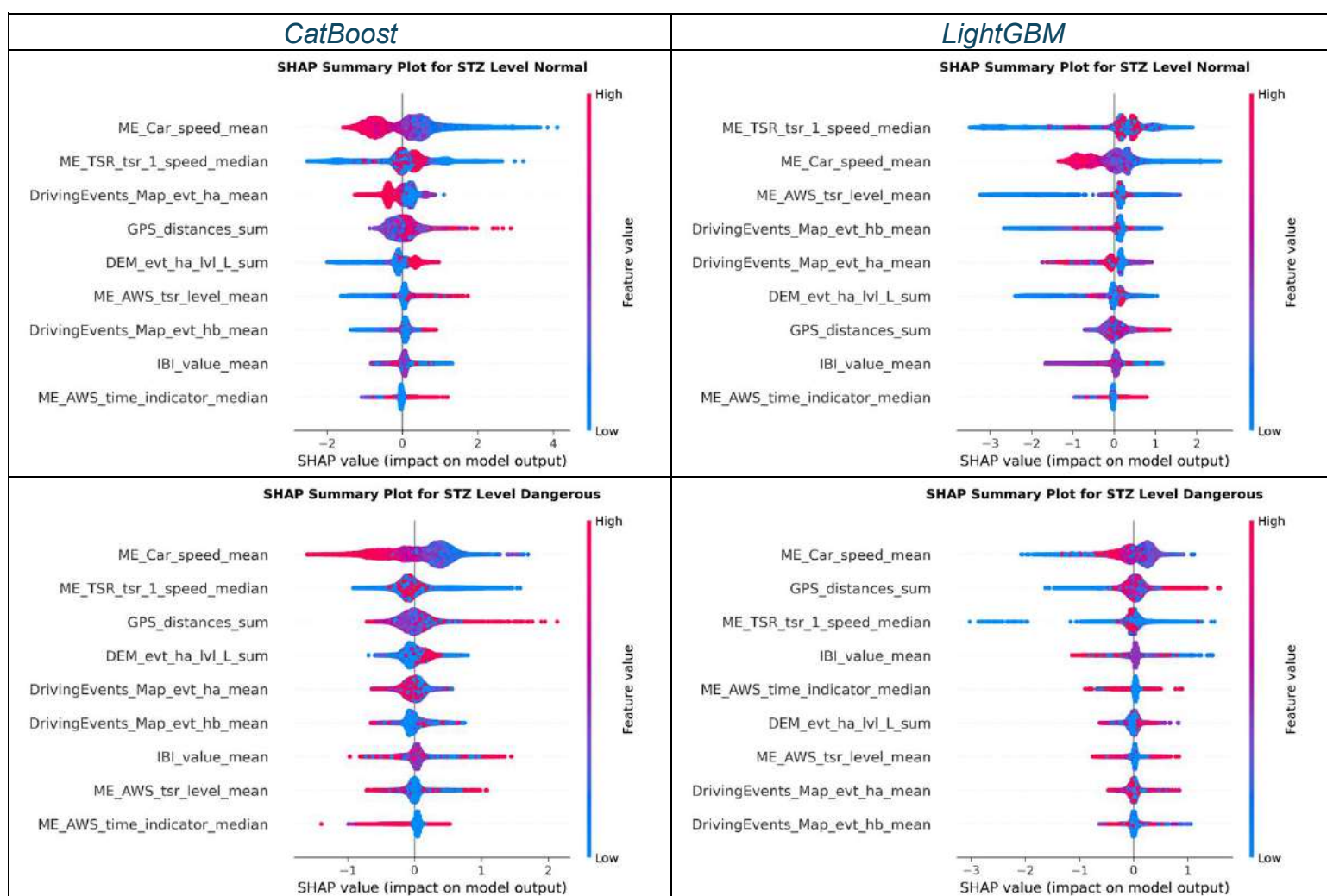


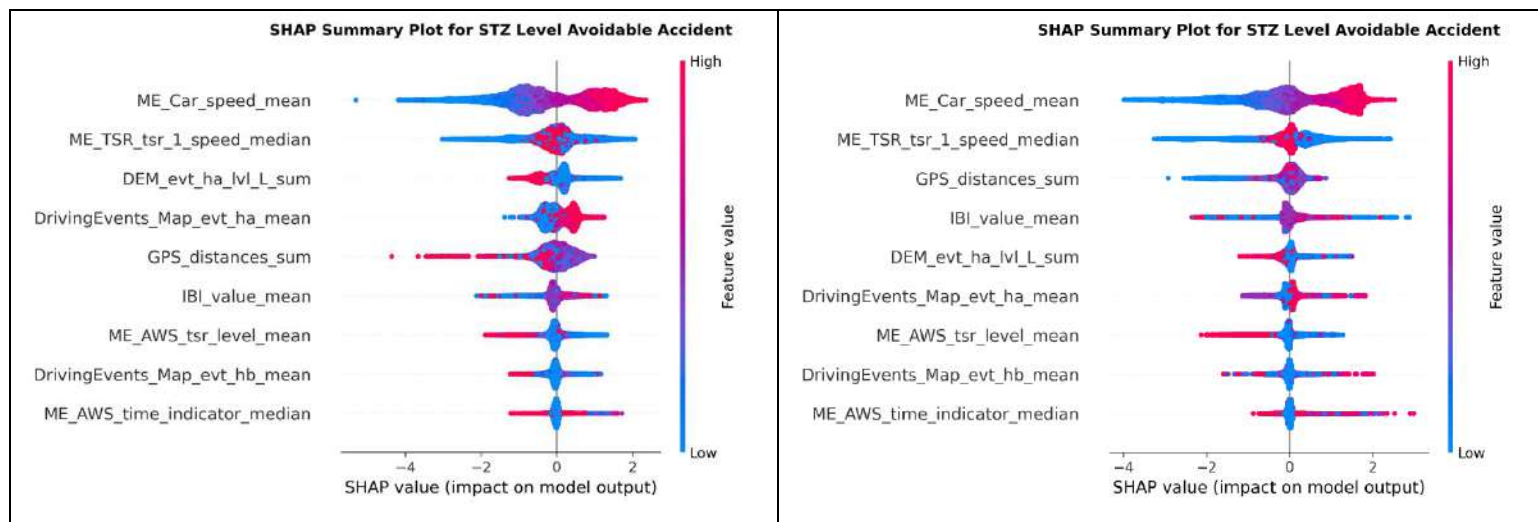
Για τα δεδομένα του **Ηνωμένου Βασιλείου**, προκύπτει για το **CatBoost**, πως η σημαντικότερη μεταβλητή είναι η **μέση ταχύτητα οχήματος (ME_Car_speed_mean)** που ωθεί στην αλλαγή επιπέδου επικίνδυνης οδηγικής συμπεριφοράς. Πιο συγκεκριμένα, στο “Level 0”, καθώς και στο “Level 1”, φαίνεται πως υψηλές τιμές της μεταβλητής (κόκκινο χρώμα) μειώνουν την πιθανότητα να καταταγεί το όχημα στην “STZ Level Normal” και στην “STZ Level Dangerous” αντίστοιχα. Στο “Level 2”, αποδεικνύεται πως υψηλές τιμές της μεταβλητής (κόκκινο χρώμα) ωθούν το όχημα προς την κατάταξη του στο “STZ Level Avoidable Accident”, κάτι αναμενόμενο.

Για το **LightGBM** αντίστοιχα, αναδεικνύεται ως η σημαντικότερη μεταβλητή το **χρονικό διάστημα μεταξύ διαδοχικών καρδιακών παλμών (IBI_value_mean)**, που ωθεί στην αλλαγή επιπέδου επικίνδυνης οδηγικής συμπεριφοράς.

Βέλγιο

Πίνακας 2: Διαγράμματα μεθόδου SHAP για το Βέλγιο





Για τα δεδομένα του **Βελγίου**, για το **CatBoost**, προκύπτει, όπως και στην περίπτωση του Ηνωμένου Βασιλείου, πως η σημαντικότερη μεταβλητή είναι η **μέση ταχύτητα οχήματος (ME_Car_speed_mean)** που ωθεί στην αλλαγή επιπέδου επικίνδυνης οδηγικής συμπεριφοράς. Πιο συγκεκριμένα, στο “Level 0”, καθώς και στο “Level 1”, φαίνεται πως υψηλές τιμές της μεταβλητής (κόκκινο χρώμα) μειώνουν την πιθανότητα να καταταγεί το όχημα στην “STZ Level Normal” και στην “STZ Level Dangerous” αντίστοιχα. Στο “Level 2”, αποδεικνύεται πως υψηλές τιμές της μεταβλητής (κόκκινο χρώμα) ωθούν το όχημα προς την κατάταξη του στο “STZ Level Avoidable Accident”, κάτι αναμενόμενο.

Για το **LightGBM**, αναδεικνύεται επίσης ως η σημαντικότερη μεταβλητή η **μέση ταχύτητα οχήματος (ME_Car_speed_mean)** που ωθεί στην αλλαγή επιπέδου επικίνδυνης οδηγικής συμπεριφοράς. Εξαιρείται από τα παραπάνω το “Level 0”, στο οποίο χρίζεται ως επιδραστικότερη μεταβλητή η **ME_TSR_tsr_1_speed_median**, δηλαδή η **διάμεση τιμή της ταχύτητας που σχετίζεται με το TSR (Traffic Sign Recognition)**, η οποία επηρεάζει το αποτέλεσμα παρόμοια με την ταχύτητα του οχήματος.

Σύμφωνα με τα αποτελέσματα που προκύπτουν από την εφαρμογή της μεθοδολογίας, διατυπώνονται ορισμένα συμπεράσματα που συνδέονται άμεσα με τον στόχο της Διπλωματικής Εργασίας:

- Η εφαρμογή της μεθόδου **SMOTE** πράγματι αντιμετωπίζει το φαινόμενο της ανισορροπίας δεδομένων μεταξύ των διαφορετικών επιπέδων, κάτι που αποδεικνύεται κι από το ικανοποιητικό επίπεδο των μετρικών αξιολόγησης των τεσσάρων διαφορετικών μοντέλων ταξινόμησης
- Από τα τέσσερα μοντέλα ταξινόμησης που αναπτύσσονται, το μοντέλο ταξινόμησης **Multi-Layer Perceptron (MLP)** φαίνεται το πιο αδύναμο μοντέλο για την πρόβλεψη του επιπέδου ασφαλείας, τόσο μέσω των μητρών σύγχυσης όσο και μέσω των μετρικών αξιολόγησης των μοντέλων

- Από τα τέσσερα μοντέλα ταξινόμησης που αναπτύσσονται, τα μοντέλα **CatBoost** και **LightGBM** παρουσιάζουν ικανοποιητικές μετρικές αξιολόγησης καθώς και την υψηλότερη μέση σημαντικότητα SHAP
- Η **μέση ταχύτητα οχήματος** υπολογίζεται ως η σημαντικότερη μεταβλητή από τον υπολογισμό σημαντικότητας μεταβλητών με τον αλγόριθμο “**Τυχαίου Δάσους**”, ενώ το ίδιο αποτέλεσμα εξάγεται κι από τη μέθοδο SHAP για όλες τις περιπτώσεις πλην του μοντέλου **LightGBM**, για το Ηνωμένο Βασίλειο, όπου κατατάσσεται δεύτερη μετά **το χρονικό διάστημα μεταξύ διαδοχικών καρδιακών παλμών**
- **Το χρονικό διάστημα μεταξύ διαδοχικών καρδιακών παλμών** εμφανίζεται ως η πιο σημαντική μεταβλητή του μοντέλου LightGBM για το Ηνωμένο Βασίλειο με τη μέθοδο SHAP, ενώ φαίνεται πως υψηλές τιμές της μεταβλητής, ένδειξη χαλάρωσης-υποαπασχόλησης του οδηγού, ενδέχεται να ωθήσουν την οδηγική συμπεριφορά στην κατάταξη της ως επικίνδυνη και αρκετά σπανιότερα στο επίπεδο αποφυγής ατυχήματος
- Η σχετικά υψηλή σημαντικότητα των μεταβλητών **απότομων επιταχύνσεων** κι **απότομων φρεναρισμάτων** καταδεικνύει πως τα απότομα περιστατικά οδήγησης επηρεάζουν σημαντικά την κατάταξη της οδηγικής συμπεριφοράς ως επικίνδυνη.
- Η υψηλή επιρροή της **συνολικής απόστασης ταξιδιού** υποδεικνύει ότι οι αποστάσεις που καλύπτονται έχουν επίσης σημαντική συνεισφορά στον τρόπο οδήγησης. Οι αποστάσεις που καλύπτει ένας οδηγός επηρεάζουν την επικίνδυνη οδήγηση μέσω της κόπωσης, της εξοικείωσης με τη διαδρομή, της ταχύτητας και της ψυχολογικής του κατάστασης
- Τα αποτελέσματα των αναλύσεων δείχνουν ότι οι δύο χώρες παρουσιάζουν παρόμοια πρότυπα οδήγησης. Η **μέση ταχύτητα του οχήματος** αποτελεί πολύ σημαντική μεταβλητή και για τις δύο χώρες, όπως και η **διάμεση ταχύτητα κατά την εμφάνιση της πρώτης πινακίδας ορίου ταχύτητας** και η **συνολική απόσταση του ταξιδιού**
- Η **διάμεση τιμή της ταχύτητας του οχήματος όταν εντοπίζεται το πρώτο σήμα ορίου ταχύτητας** από το σύστημα αναγνώρισης σημάτων κυκλοφορίας εμφανίζει υψηλή σημαντικότητα τόσο για το Ηνωμένο Βασίλειο, όσο και για το Βέλγιο
- Μεγάλη διαφορά παρατηρείται στη σημαντικότητα **του χρονικού διαστήματος μεταξύ διαδοχικών καρδιακών παλμών**. Πιο συγκεκριμένα, αποδεικνύεται μία από τις πιο σημαντικές μεταβλητές για το Ηνωμένο Βασίλειο σε αντίθεση με το Βέλγιο. Αυτό μας δείχνει ενδεχομένως πως η ψυχική κατάσταση των οδηγών επηρεάζει περισσότερο την οδηγική τους συμπεριφορά στο Ηνωμένο Βασίλειο σε σχέση με το Βέλγιο. Ψυχικές

καταστάσεις όπως στρες, ταχυκαρδία ή αντίστοιχα βραδυκαρδία, χαλάρωση είναι εκείνες που μπορεί να επηρεάσουν τη συμπεριφορά του οδηγού

Περιεχόμενα

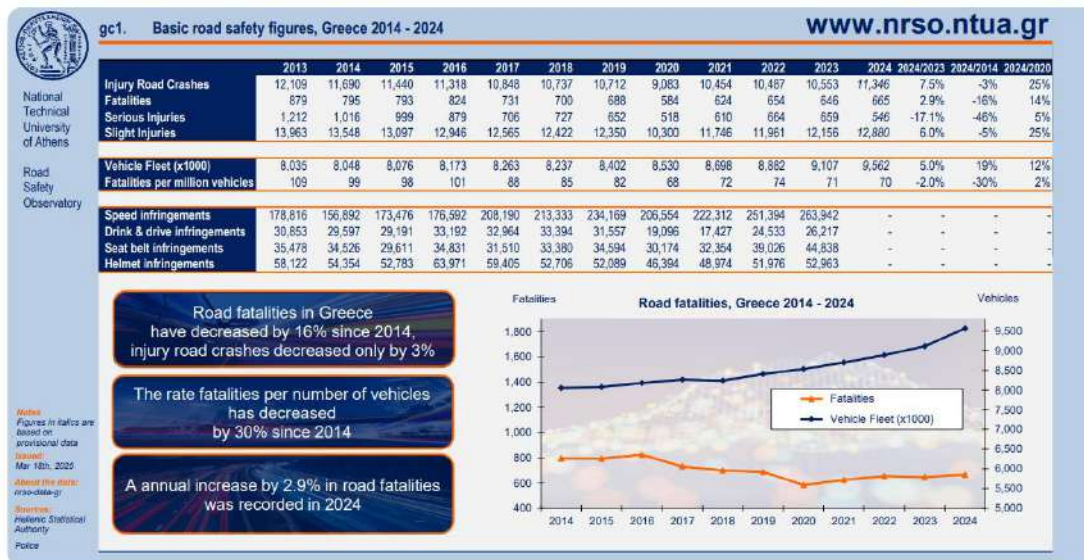
1. ΕΙΣΑΓΩΓΗ.....	1
1.1. Γενική Ανασκόπηση.....	1
1.2. Στόχος Διπλωματικής Εργασίας.....	3
1.3. Μεθοδολογία	4
1.4. Δομή Διπλωματικής εργασίας.....	8
2.ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ	10
2.1. Εισαγωγή	10
2.2. Συναφείς έρευνες και μεθοδολογίες.....	10
2.2.1 Ανάλυση οδηγικής συμπεριφοράς	11
2.2.2 Αλγόριθμοι ταξινόμησης οδηγικής συμπεριφοράς	14
2.3. Πρόβλημα ανισορροπίας δεδομένων σε κάθε τάξη	15
2.4. Αξιολόγηση επίδρασης μεταβλητών στην επικίνδυνη οδήγηση	16
2.5. Σύνοψη.....	17
3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ	19
3.1. Εισαγωγή	19
3.2. Επιλογή χαρακτηριστικών (Feature Selection).....	19
3.3. Μέθοδοι επαναδειγματοληψίας για προβλήματα ανισορροπίας ταξινόμησης.....	19
3.3.1. Τεχνική Συνθετικής Μειονοτικής Υπερδειγματοληψίας (SMOTE).....	20
3.4. Αλγόριθμοι ταξινόμησης	21
3.4.1. Random Forest (Τυχαίο Δάσος).....	22
3.4.2. Light Gradient Boosting Machine (Ελαφρύς Αλγόριθμος Βαθιάς Ενίσχυσης).....	24
3.4.3. Categorical Boosting (Αλγόριθμος Ενίσχυσης Κατηγοριών).....	25
3.4.4. Multi-Layer Perceptron (Πολυεπίπεδο Αντιληπτικό)	27
3.5. Μετρικές αξιολόγησης Μοντέλων Ταξινόμησης.....	28
3.5.1. Μήτρα Σύγχυσης (Confusion Matrix).....	28
3.5.2. Ακρίβεια (Accuracy).....	29
3.5.3. Ακρίβεια Θετικής Πρόβλεψης (Precision).....	30
3.5.4. Ανάκληση (Recall)	30
3.5.5. Ρυθμός Λανθασμένου Συναγερμού (False Alarm Rate).....	30
3.5.6. f1-score	30

3.6. Κατανόηση λειτουργίας μοντέλων μηχανικής μάθησης μέσω της μεθόδου SHAP	31
4. ΣΥΛΛΟΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΣΤΟΙΧΕΙΩΝ	32
4.1. Εισαγωγή	32
4.2. Πείραμα σε πραγματικές συνθήκες οδήγησης	32
4.2.1. Στόχος πειράματος	32
4.2.2. On-Board Diagnostics (OBD II).....	33
4.2.3. Σύστημα Πρόληψης Ατυχημάτων MobilEye	34
4.2.4. Στοιχεία που συλλέχθηκαν από το πείραμα i-Dreams	35
4.3.Επεξεργασία στοιχείων	37
4.4. Περιγραφική Στατιστική Μεταβλητών	41
4.5. Σύνοψη.....	43
5. ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ – ΑΠΟΤΕΛΕΣΜΑΤΑ	44
5.1 Εισαγωγή	44
5.2 Εντοπισμός του Επιπέδου ‘Ζώνης Ανοχής Ασφαλείας’	44
5.2.1 Καθορισμός Επιπέδων Ασφαλείας	45
5.2.2 Επιλογή Χαρακτηριστικών (Feature Selection)	46
5.2.3 Προετοιμασία δεδομένων	51
5.2.4. Αντιμετώπιση άνισης κατανομής δεδομένων	51
5.2.5. Ανάπτυξη μοντέλων ταξινόμησης	54
5.2.6. Σύγκριση Μετρικών αξιολόγησης των μοντέλων.....	61
5.3 Εξήγηση Λειτουργίας Μοντέλων Μηχανικής μάθησης	62
6.ΣΥΜΠΕΡΑΣΜΑΤΑ	74
6.1. Σύνοψη αποτελεσμάτων.....	74
6.2. Σύνοψη συμπερασμάτων	78
6.3. Προτάσεις για αξιοποίηση των αποτελεσμάτων.....	79
6.4. Προτάσεις για μελλοντική έρευνα	80
BIBΛΙΟΓΡΑΦΙΑ	82

1. ΕΙΣΑΓΩΓΗ

1.1. Γενική Ανασκόπηση

Η οδική ασφάλεια αποτελεί ένα κρίσιμο ζήτημα για τις σύγχρονες κοινωνίες, καθώς τα τροχαία ατυχήματα συγκαταλέγονται στις κύριες αιτίες θανάτου και σοβαρού τραυματισμού παγκοσμίως. Η κατανόηση και η αναγνώριση της επικίνδυνης οδηγικής συμπεριφοράς μπορούν να συμβάλουν σημαντικά στη μείωση των ατυχημάτων, αξιοποιώντας τεχνολογίες αιχμής, όπως η μηχανική εκμάθηση.

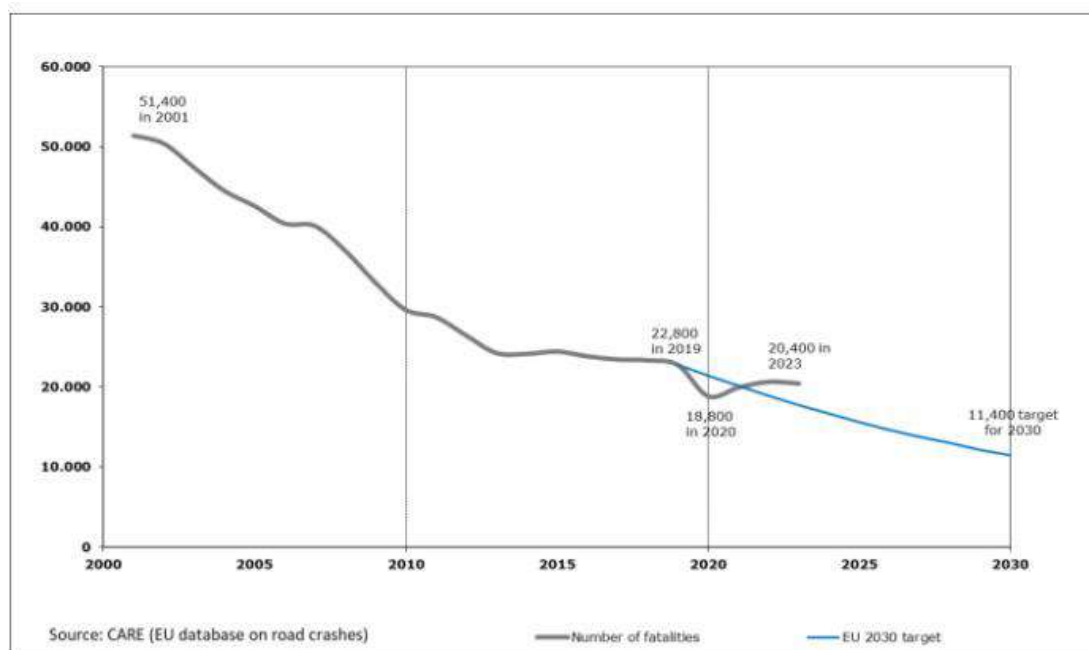


Εικόνα 1.1: Στοιχεία Οδικής Ασφάλειας, Ελλάδα 2014-2024

Πηγή: NTUA Road Safety Observatory (2025)

Όπως αποτυπώνεται στην Εικόνα 1.1, την τελευταία δεκαετία η Ελλάδα παρουσίασε σημαντική βελτίωση της οδικής ασφάλειας, καθώς παρουσίασε μείωση κατά 16% των θανάτων από τροχαία ατυχήματα, ενώ αναλογικά με τον αριθμό των οχημάτων αποτυπώθηκε μείωση 30% του ρυθμού των θανάτων. Το τελευταίο σημαίνει πως σε αυξημένο συνολικό αριθμό οχημάτων κατά 1.514.000, οι απώλειες ζωών από οδικά ατυχήματα εμφανίζονται αναλογικά μειωμένες. Παράλληλα, οι σοβαροί τραυματισμοί από τροχαία ατυχήματα, μειώθηκαν κατά 46% στη δεκαετία 2014-2024.

Αντίστοιχη μείωση παρατηρείται και στην Ευρωπαϊκή Ένωση, όπου τα τελευταία 23 χρόνια ο αριθμός θανάτων από οδικά ατυχήματα εμφανίζεται μειωμένος κατά 60%, ενώ την τελευταία δεκαετία κυμαίνεται σε τιμές που προσεγγίζουν αυτές της Ελλάδας, περί το 20% μειωμένος από το 2013 έως το 2023 όπως φαίνεται στο παρακάτω Διάγραμμα 1.1.



Διάγραμμα 1.1: Διαχρονική Εξέλιξη Νεκρών σε Οδικά Ατυχήματα στην Ε.Ε.

Πηγή: NTUA Road Safety Observatory (2024)

Επιπλέον, σύμφωνα με την Ειδική Έκθεση 04/2024 του Ευρωπαϊκού Ελεγκτικού Συνεδρίου, η Ευρωπαϊκή Επιτροπή έχει θέσει στόχο τη μείωση κατά το ήμισυ των θανάτων και των σοβαρών τραυματισμών από τροχαία ατυχήματα μεταξύ 2020 και 2030. Για την επίτευξη αυτού του στόχου, δίνεται ιδιαίτερη έμφαση στην ενσωμάτωση καινοτόμων τεχνολογιών στην αυτοκινητοβιομηχανία και στην αυτοματοποίηση των μεταφορών. Οι προηγμένες τεχνολογίες υποβοήθησης οδηγού (ADAS), όπως τα συστήματα αυτόματης πέδησης έκτακτης ανάγκης, η υποβοήθηση διατήρησης λωρίδας και τα προσαρμοζόμενα συστήματα ελέγχου ταχύτητας, συμβάλλουν στη μείωση των ατυχημάτων. Επιπλέον, η ανάπτυξη αυτόνομων οχημάτων και η χρήση ευφυών συστημάτων μεταφορών (ITS) βελτιώνουν την οδική ασφάλεια μέσω της μείωσης των ανθρώπινων λαθών. Η ενσωμάτωση αυτών των τεχνολογιών και η εφαρμογή των προαναφερθέντων μέτρων αναμένεται να συμβάλουν σημαντικά στη μείωση των θανατηφόρων τροχαίων ατυχημάτων στην Ε.Ε. κατά 50% έως το 2030.

Σύμφωνα με τα προκαταρκτικά στατιστικά στοιχεία της Ευρωπαϊκής Επιτροπής για το 2024, λιγότεροι άνθρωποι έχασαν τη ζωή τους στους ευρωπαϊκούς δρόμους το 2024, σημειώνοντας μια μικρή μείωση της τάξης του 3% σε σύγκριση με το 2023, που αντιστοιχεί σε 600 λιγότερες απώλειες ανθρωπίνων ζωών. Παρόλο που τα θανατηφόρα τροχαία μειώθηκαν πιο αισθητά το 2024 σε σχέση με το 2023, ο συνολικός ρυθμός βελτίωσης παραμένει πολύ αργός και τα περισσότερα κράτη μέλη δεν βρίσκονται σε καλό δρόμο για να επιτύχουν τον στόχο της ΕΕ για μείωση των θανάτων από τροχαία κατά το ήμισυ έως το 2030.

Επιπρόσθετα, κομβικός είναι ο επιμερισμός των ευθυνών στους κύριους παράγοντες πρόκλησης των οδικών ατυχημάτων. Όπως αναφέρεται στο άρθρο του Kareem (2003), τα αίτια του τροχαίου ατυχήματος είναι ελαττωματικά οχήματα, ανώμαλοι δρόμοι, απρόσεκτη οδήγηση, υπερβολική ταχύτητα, οδήγηση υπό την επήρεια αλκοόλ, ανεπαρκής ύπνος, αλκοόλ, ναρκωτικά και πολλά άλλα. Συνεπώς, συγκεντρωτικά, οι ευθύνες διαμοιράζονται μεταξύ τριών παραγόντων: του ανθρώπου, της οδού και του οχήματος.

Πίνακας 1.1: Συμβολή των τριών κύριων παραγόντων στα οδικά ατυχήματα

Άνθρωπος μόνο	65%
Άνθρωπος και Οδός	24%
Άνθρωπος και Όχημα	4.5%
Άνθρωπος, Οδός και Όχημα	1.25%
Οδός μόνο	2.5%
Οδός και Όχημα	0.25%
Όχημα μόνο	2.5%
ΣΥΝΟΛΟ	100.0%

Πηγή: NTUA Road Safety Observatory

Σύμφωνα με δεδομένα του Τομέα Μεταφορών και Συγκοινωνιακής Υποδομής, της Σχολής Πολιτικών Μηχανικών, ο ανθρώπινος παράγοντας συμμετέχει στο 95% των οδικών ατυχημάτων, το 65% των οποίων οφείλεται αποκλειστικά στον ανθρώπινο παράγοντα. Συνιστά λοιπόν κρίσιμης σημασίας, η διερεύνηση της οδηγικής συμπεριφοράς ανάγοντας τη στη μέτρηση διαφόρων μεταβλητών που την επηρεάζουν.

1.2. Στόχος Διπλωματικής Εργασίας

Στο πλαίσιο αυτής της Διπλωματικής Εργασίας, γίνεται χρήση δεδομένων από το Ηνωμένο Βασίλειο και το Βέλγιο, δύο χώρες με διαφορετικές οδικές υποδομές και πολιτικές οδικής ασφάλειας, προκειμένου **να αναλυθούν οι παράγοντες που επηρεάζουν την επικίνδυνη οδήγηση**. Μέσα από την

εφαρμογή κατάλληλων αλγορίθμων μηχανικής εκμάθησης, επιχειρείται η ανάπτυξη ενός μοντέλου που θα μπορεί να ανιχνεύει την επικίνδυνη οδηγική συμπεριφορά, συμβάλλοντας στη βελτίωση των στρατηγικών πρόληψης και μείωσης των τροχαίων ατυχημάτων.

Αναλυτικότερα, η οδηγική συμπεριφορά θα χωριστεί σε **τρία επίπεδα ασφαλείας** με βάση έναν δείκτη τη **“Ζώνη Ανοχής Ασφαλείας”** ο οποίος σε πολλά σημεία θα αναφέρεται κι ως **“STZ level”**. Αυτά τα επίπεδα είναι: το **κανονικό (normal)**, το **επικίνδυνο (dangerous)** και το επίπεδο **αποφυγής ατυχήματος (avoidable accident)**. Στη συνέχεια, παρατίθενται κάποια βασικά ερωτήματα που τίθενται προς διερεύνηση στη συγκεκριμένη εργασία:

- Ικανότητα αναγνώρισης του επιπέδου ασφαλείας του οδηγού (με βάση τα τρία επίπεδα επικινδυνότητας).
- Σύγκριση της επίδοσης των διαφορετικών μοντέλων ταξινόμησης αλλά και της επιρροής τους στις διαφορετικές συνθήκες πειραμάτων.
- Εντοπισμός των χαρακτηριστικών που επιδρούν στην διαδικασία αναγνώρισης της οδηγικής συμπεριφοράς.

Οι τρόποι αντιμετώπισης των παραπάνω ερωτημάτων συνοψίζονται ως εξής:

- Ανάπτυξη μοντέλων μηχανικής και βαθιάς μάθησης που ταξινομούν τους οδηγούς σε τρία επίπεδα της «Ζώνης Ανοχής Ασφαλείας». Η ταξινόμηση βασίζεται σε χαρακτηριστικά τόσο της οδηγικής συμπεριφοράς όσο και του περιβάλλοντος οδήγησης, με στόχο την ακριβή αξιολόγηση της ασφάλειας κάθε οδηγού.
- Αξιοποίηση της μεθόδου SHAP (SHapley Additive exPlanations) για την αξιολόγηση της συνεισφοράς κάθε χαρακτηριστικού της οδηγικής συμπεριφοράς στο επίπεδο επικινδυνότητας στην τελική πρόβλεψη.

Στόχο της διπλωματικής λοιπόν, συνιστά η εξαγωγή συμπερασμάτων για την **επιρροή των παραγόντων στην κατηγοριοποίηση της επικίνδυνης οδήγησης στα τρία επίπεδα της “Ζώνης Ανοχής Ασφαλείας”**. Επεκτείνοντας, εξετάζοντας τα παραπάνω επιδιώκεται η εξαγωγή συμπερασμάτων για τη χρησιμότητα των συστημάτων υποβοήθησης οδήγησης (ADAS - Advanced Driver Assistance Systems), παράλληλα με τη βελτίωση της οδηγικής αντίληψης και συμπεριφοράς.

1.3. Μεθοδολογία

Το πρώτο βήμα της μελέτης ήταν η οριστικοποίηση του θέματος και ο καθορισμός του στόχου της. Στη συνέχεια, πραγματοποιήθηκε βιβλιογραφική ανασκόπηση, όπου αναζητήθηκαν στη διεθνή βιβλιογραφία σχετικές έρευνες, μελέτες και μεθοδολογίες ανάλυσης, οι οποίες αξιοποιήθηκαν στην παρούσα Διπλωματική Εργασία.

Ακολούθησε η συλλογή και η επεξεργασία των δεδομένων, τα οποία προήλθαν από πείραμα σε πραγματικές συνθήκες οδήγησης, στο πλαίσιο του ερευνητικού έργου i-DREAMS. Τα δεδομένα αφορούσαν τα χαρακτηριστικά οδήγησης 96 οδηγών από το Βέλγιο και το Ηνωμένο Βασίλειο, καθώς και το περιβάλλον οδήγησής τους. Μέσω της κατάλληλης επεξεργασίας, τα δεδομένα προετοιμάστηκαν για την ανάλυσή τους.

Το i-DREAMS ήταν ένα 3ετές έργο που χρηματοδοτήθηκε από το πρόγραμμα έρευνας και καινοτομίας Horizon 2020 της Ευρωπαϊκής Ένωσης, το οποίο παρουσίασε μια πλατφόρμα για τον καθορισμό, την ανάπτυξη, τη δοκιμή και την επικύρωση μιας «Ζώνης Ανοχής Ασφαλείας» για να αποτρέπει τους οδηγούς να πλησιάζουν τα όρια της μη ασφαλούς οδήγησης, μετριάζοντας τους κινδύνους σε πραγματικό χρόνο και μετά το ταξίδι.

Αρκετοί παράγοντες της κατάστασης του οδηγού επηρεάζουν αρνητικά την οδική ασφάλεια, όπως η απόσπαση της προσοχής (εντός του οχήματος ή εξωτερική), η κούραση και η υπνηλία, ανησυχίες για την υγεία (π.χ. ασθένεια, αδυναμία, γνωστική κατάσταση) και ακραία συναισθήματα (π.χ. άγχος, στρες, θυμός). Επιπλέον, οι διαφορές στους κοινωνικο-πολιτιστικούς παράγοντες εξακολουθούν να είναι μεταξύ των βασικών καθοριστικών παραγόντων των οδικών κινδύνων. Ταυτόχρονα, οι τεχνολογικές εξελίξεις καθιστούν εύκολα διαθέσιμα μαζικά και λεπτομερή δεδομένα απόδοσης χειριστή. Για παράδειγμα, μέσω νέων αισθητήρων εντός του οχήματος που καταγράφουν λεπτομερή τρόπο οδήγησης και δεδομένα με βάση τα συμφραζόμενα. Αυτό δημιουργεί νέες ευκαιρίες για τον εντοπισμό και το σχεδιασμό εξατομικευμένων παρεμβάσεων για τον μετριασμό των κινδύνων, την αύξηση της ευαισθητοποίησης και την αναβάθμιση της απόδοσης του οδηγού, συνεχώς και δυναμικά. Η βέλτιστη εκμετάλλευση αυτών των ευκαιριών ήταν η πρόκληση που αντιμετώπισε το i-DREAMS.

Το ακρωνύμιο **"i-DREAMS"** (intelligent Driver and Road Environment Assessment and Monitoring System) σημαίνει ένα έξυπνο σύστημα αξιολόγησης και παρακολούθησης περιβάλλοντος οδηγού και δρόμου. Το ερευνητικό έργο i-DREAMS αποτέλεσε μία κοινοπραξία από 13 συνεργάτες, ερευνητές καθώς και συνεργάτες του κλάδου, από 8 διαφορετικές χώρες.

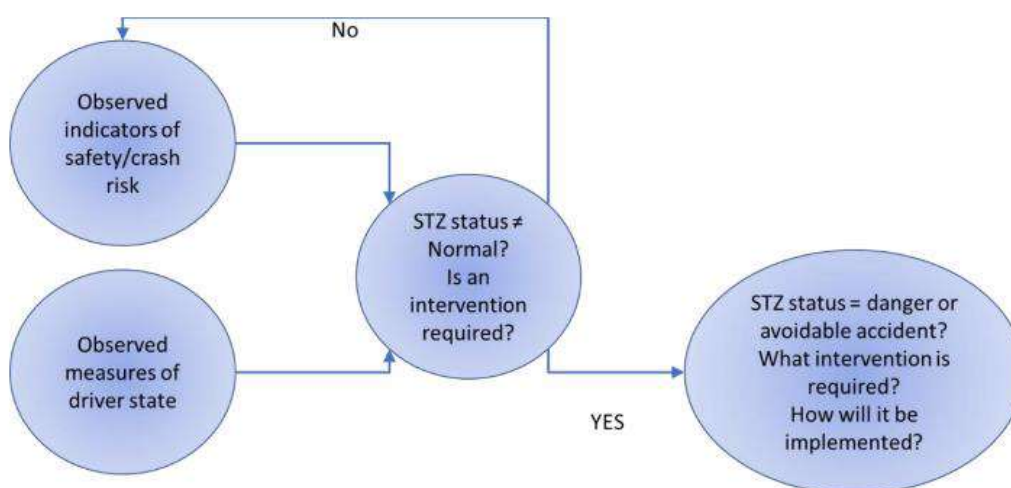
Ο σκοπός του i-DREAMS ήταν η ανάπτυξη ενός πλαισίου που θα καθορίζει, θα δημιουργεί, θα δοκιμάζει και θα επικυρώνει μια «Ζώνη Ανοχής Ασφαλείας» (Safety Tolerance Zone - STZ) με διαφορετικά επίπεδα ασφάλειας. Χρησιμοποιώντας ένα έξυπνο σύστημα παρακολούθησης, το οποίο έλαβε υπόψη το ιστορικό του οδηγού και σχετικούς δείκτες κινδύνου σε πραγματικό χρόνο, αξιολογήθηκαν οι επιδόσεις οδήγησης, η πολυπλοκότητα της εκάστοτε οδηγικής συνθήκης και τα καθήκοντα του οδηγού. Με βάση αυτά τα δεδομένα, προσδιορίστηκε το επίπεδο ασφαλούς οδήγησης του οδηγού και

εφαρμόστηκαν παρεμβάσεις που τον βοήθησαν να παραμείνει εντός ασφαλών ορίων.

Η «Ζώνη Ανοχής Ασφάλειας» περιλαμβάνει τρία επίπεδα:

- 1) Κανονικό – Ασφαλές (Normal)
- 2) Επικίνδυνο (Dangerous)
- 3) Αποφυγής Ατυχήματος (Avoidable Accident)

Στην επόμενη εικόνα, παρουσιάζεται ο διαρκής τρόπος επίδρασης του “STZ level” σε πραγματικό χρόνο οδήγησης στις παρεμβάσεις που απαιτούνται εκάστοτε. Στη συνέχεια, αξιολογείται το πως θα πραγματοποιηθούν εκείνες με κάποιο σύστημα υποβοήθησης οδήγησης (ADAS).



Εικόνα 1.2: Διάγραμμα ροής αξιολόγησης του “STZ level”

Πηγή: i-DREAMS (2023)

Ο πειραματικός σχεδιασμός της μελέτης στο δρόμο i-DREAMS υποδιαιρέθηκε σε τέσσερις διαδοχικές φάσεις:

- Φάση 1: βασική μέτρηση
- Φάση 2: παρέμβαση σε πραγματικό χρόνο
- Φάση 3: παρέμβαση σε πραγματικό χρόνο και ανατροφοδότηση μετά το ταξίδι
- Φάση 4: παρέμβαση σε πραγματικό χρόνο και ανατροφοδότηση μετά το ταξίδι και παιχνιδιοποίηση

Στον Πίνακα 1.2, απεικονίζονται τα χαρακτηριστικά της κάθε φάσης.

Πίνακας 1.2: Περίληψη των διαφορετικών φάσεων του πειράματος *i-Dreams*

Φάση	Παρέμβαση	Περιγραφή	Διάρκεια
Φάση 1 (Βασική γραμμή)	ΟΧΙ	Περίοδος αναφοράς μετά την εγκατάσταση του συστήματος <i>i-DREAMS</i> , προκειμένου να παρακολουθηθεί η οδηγική συμπεριφορά χωρίς παρεμβάσεις.	4 εβδομάδες
Φάση 2	Πραγματικού χρόνου	Περίοδος παρακολούθησης κατά την οποία παρέχονται μόνο ειδοποιήσεις σε πραγματικό χρόνο μέσα στο όχημα, χρησιμοποιώντας προσαρμοστικό ADAS.	4 εβδομάδες
Φάση 3	Πραγματικού χρόνου + Μετά το ταξίδι	Περίοδος παρακολούθησης κατά την οποία, εκτός από τις ειδοποιήσεις σε πραγματικό χρόνο μέσα στο όχημα, οι οδηγοί έλαβαν ανατροφοδότηση για την οδηγική τους απόδοση μέσω της εφαρμογής.	4 εβδομάδες
Φάση 4	Πραγματικού χρόνου + Μετά το ταξίδι + Gamification	Περίοδος παρακολούθησης κατά την οποία οι παρεμβάσεις σε πραγματικό χρόνο στο όχημα ήταν ενεργές μαζί με ανατροφοδότηση, ενώ ταυτόχρονα υπήρχαν ενεργά και στοιχεία παιχνιδοποίησης.	6 εβδομάδες

Πηγή: *i-DREAMS* (2023)

Πιο συγκεκριμένα, έγινε παρέμβαση σε πραγματικό χρόνο κατά τη διάρκεια της οδήγησης και μετά τη διαδρομή για τη βελτίωση της συμπεριφοράς του οδηγού. Στη φάση 4, έγινε και παιχνιδοποίηση με τα δεδομένα που συλλέχθηκαν, δηλαδή ενσωμάτωση στοιχείων παιχνιδιού, όπως συστήματα πόντων, βαθμολογικοί πίνακες, διακριτικά σήματα (badges) ή άλλα στοιχεία που σχετίζονται με παιχνίδια, προκειμένου να αυξηθεί η δέσμευση και τα κίνητρα των οδηγών.

Αφού ολοκληρώθηκε η συλλογή και η επεξεργασία των δεδομένων, ακολούθησε η ανάπτυξη κατάλληλων μοντέλων μηχανικής και βαθιάς μάθησης με στόχο την ταξινόμηση. Η επεξεργασία των δεδομένων, η δημιουργία των μοντέλων και οι σχετικές αναλύσεις πραγματοποιήθηκαν με τη γλώσσα προγραμματισμού Python, αξιοποιώντας βιβλιοθήκες μηχανικής μάθησης όπως οι TensorFlow και scikit-learn, τη βιβλιοθήκη pandas για ανάλυση δεδομένων, καθώς και το πακέτο SHAP (SHapley Additive exPlanations) για την ερμηνεία των μοντέλων. Τέλος, τα αποτελέσματα αξιολογήθηκαν, οδηγώντας στην εξαγωγή χρήσιμων συμπερασμάτων και προτάσεων για μελλοντική έρευνα.

Στο Γράφημα 1.1, παρουσιάζονται τα διαδοχικά στάδια που ακολουθήθηκαν για την εκπόνηση της παρούσας διπλωματικής εργασίας.



Γράφημα 1.1: Διάγραμμα Ροής- Μεθοδολογία Διπλωματικής Εργασίας

1.4. Δομή Διπλωματικής εργασίας

Αυτή η ενότητα παρουσιάζει τη δομή της διπλωματικής εργασίας, συνοψίζοντας το περιεχόμενο κάθε κεφαλαίου.

- **Κεφάλαιο 1:** Εισαγωγή και παρουσίαση του στόχου της εργασίας. Περιγράφεται το γενικό πλαίσιο, εστιάζοντας στον αντίκτυπο των οδικών ατυχημάτων και στη συμβολή των σύγχρονων τεχνολογιών στη μείωση τους, με ιδιαίτερη αναφορά στο ερευνητικό έργο i-DREAMS. Επιπλέον, αναλύονται η μεθοδολογία και η δομή της εργασίας.
- **Κεφάλαιο 2:** Βιβλιογραφική ανασκόπηση σχετικών ερευνών, τόσο ως προς το αντικείμενο της εργασίας όσο και ως προς τις μεθοδολογίες που εφαρμόστηκαν, με πηγές από την διεθνή επιστημονική κοινότητα.
- **Κεφάλαιο 3:** Θεωρητικό υπόβαθρο της έρευνας. Αναλύονται οι τεχνικές επεξεργασίας δεδομένων, η σημασία τους για την ανάπτυξη των μοντέλων, καθώς και οι αλγόριθμοι μηχανικής και βαθιάς μάθησης που χρησιμοποιήθηκαν, μαζί με τις αντίστοιχες μετρικές αξιολόγησής τους.
- **Κεφάλαιο 4:** Παρουσίαση των δεδομένων και της διαδικασίας συλλογής τους μέσω νατουραλιστικών πειραμάτων πραγματικής οδήγησης (i-DREAMS). Εξετάζονται τα βήματα επεξεργασίας των οδηγικών και περιβαλλοντικών χαρακτηριστικών για την περαιτέρω ανάλυση.
- **Κεφάλαιο 5:** Κύρια ενότητα της εργασίας, όπου περιγράφεται λεπτομερώς η μεθοδολογία ανάπτυξης των μοντέλων ταξινόμησης και παλινδρόμησης. Αναλύονται τα στάδια εφαρμογής, η ανάπτυξη των μοντέλων, τα δεδομένα εισόδου και εξόδου, καθώς και τα συνολικά αποτελέσματα της ανάλυσης με συγκρίσεις και μετρικές αξιολόγησης.

- **Κεφάλαιο 6:** Παρουσίαση των συμπερασμάτων της εργασίας με βάση τα αποτελέσματα του προηγούμενου κεφαλαίου. Επιπλέον, προτείνονται κατευθύνσεις για μελλοντική έρευνα, όπως η εφαρμογή διαφορετικών μεθόδων ή η αξιοποίηση νέων δεδομένων.
- **Κεφάλαιο 7:** Βιβλιογραφικές αναφορές που χρησιμοποιήθηκαν κατά την εκπόνηση της διπλωματικής εργασίας.

2.ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

2.1. Εισαγωγή

Στο κεφάλαιο της βιβλιογραφικής ανασκόπησης παρατίθενται σχετικές μεθοδολογίες και μελέτες που συνδέονται με το αντικείμενο της παρούσας διπλωματικής εργασίας. Ειδικότερα, εξετάστηκαν δημοσιευμένες έρευνες στη διεθνή βιβλιογραφία, οι οποίες επικεντρώνονται στην ανάλυση και αναγνώριση της οδηγικής συμπεριφοράς, καθώς και στην πρόβλεψη συγκρούσεων σε πραγματικό χρόνο, αξιοποιώντας τεχνικές μηχανικής μάθησης και βαθιάς μάθησης.

Μέσα από την παρουσίαση και τη σύγκριση αυτών των μελετών, θα καθοριστεί τόσο ο στόχος της παρούσας έρευνας όσο και η κατάλληλη μεθοδολογία για την υλοποίησή του. Ιδιαίτερη έμφαση θα δοθεί στα μοντέλα μηχανικής και βαθιάς μάθησης. Επιπλέον, στις περισσότερες από τις ανασκοπηθείσες έρευνες αναδεικνύεται το πρόβλημα της άνισης κατανομής των δεδομένων μεταξύ των διαφορετικών τάξεων. Για τον λόγο αυτό, θα παρουσιαστούν οι διάφορες τεχνικές επαναδειγματοληψίας δεδομένων που έχουν εφαρμοστεί σε προηγούμενες μελέτες.

2.2. Συναφείς έρευνες και μεθοδολογίες

Οι αριθμοί των αυτοκινήτων ήταν περίπου 800 εκατομμύρια το έτος 2011 και έχουν αυξηθεί σε περίπου ένα δισεκατομμύριο το έτος 2018. Υπολογίζεται ότι ο αριθμός αυτός μπορεί να ξεπεράσει τα 2 δισεκατομμύρια περίπου μέχρι το έτος 2050. Με περισσότερα οχήματα να βρίσκονται σε συμφόρηση στους δρόμους, οι πιθανότητες ατυχημάτων αυξάνονται επίσης αναλογικά. Παρατηρείται διαφορετική συχνότητα ατυχημάτων αναλογικά με τον πληθυσμό στις ηπείρους. Για παράδειγμα, τα θύματα τροχαίων ατυχημάτων ανά 1.000.000 πληθυσμού είναι 10,3 για την Ευρώπη, 18,5 για τη Νοτιοανατολική Ασία και 24,1 για την Αφρική.

Οι αιτίες των ατυχημάτων ταξινομούνται σε τρεις κατηγορίες: κακές καιρικές συνθήκες ή κακές υποδομές (βροχή, λακκούβες στο δρόμο), δυσλειτουργία του οχήματος (κατασκευαστικά ελαττώματα ή φθορά) ή ανθρώπινοι παράγοντες (φυσιολογικοί ή συμπεριφορικοί). Ενώ τα φυσιολογικά λάθη συμβαίνουν λόγω κόπωσης, υπνηλίας, τα λάθη συμπεριφοράς θα μπορούσαν να λάβουν πολλές μορφές, όπως οδήγηση με απόσπαση προσοχής, οδήγηση υπό την επήρεια μέθης, επιθετική οδήγηση, οργή δρόμου, απότομη επιτάχυνση, σκληρό φρενάρισμα και στρίψιμο και υπερβολική ταχύτητα. (Arumugam and Bhargavi, 2019)

Η επιρροή πολλών πτυχών που επηρεάζουν την απόδοση οδήγησης, όπως χαρακτηριστικά οδηγού, η γεωμετρία της οδού και οι καιρικές συνθήκες, διερευνήθηκαν από τους Zolali et al. (2021) χρησιμοποιώντας πειράματα που πραγματοποιήθηκαν σε προσομοιωτή οδήγησης. Όσον αφορά τους

πρακτικούς παράγοντες τροχαίων ατυχημάτων, η ταχύτητα και, ειδικότερα, οι ταχύτητες πάνω από το όριο ταχύτητας είναι ο πιο κρίσιμος παράγοντας που μπορεί να παίξει καθοριστικό ρόλο στη σοβαρότητα των ατυχημάτων. Πολλοί παράγοντες μπορούν να επηρεάσουν την επιλογή ταχύτητας του οδηγού. Περιβαλλοντικοί παράγοντες όπως η γεωμετρία της οδού, οι καιρικές συνθήκες και η χρήση γης ή τα χαρακτηριστικά των οδηγών όπως η ηλικία, το φύλο και η οδηγική εμπειρία επηρεάζουν τη συμπεριφορά επιλογής ταχύτητας των οδηγών.

Επίσης, το αποτέλεσμα του πειράματος προσομοιωτή εικονικής πραγματικότητας δείχνει πως η γρήγορη οδήγηση μπορεί να επηρεάσει την απόδοση του οδηγού. Το αποτέλεσμα δείχνει ότι ο έλεγχος του αυτοκινήτου, η ταχύτητα, οι εντάσεις και η κούραση συνδέονται στενά. (Taheri et al., 2017)

2.2.1 Ανάλυση οδηγικής συμπεριφοράς

Η μελέτη Lyu et al. (2022) στοχεύει να καθορίσει πλαίσιο αναγνώρισης στυλ οδήγησης με βάση τις συνθήκες διαμήκους λειτουργίας οδήγησης (DOC). Η συμβολή αυτής της έρευνας είναι να προτείνει ένα κριτήριο και λύση για τη χρήση δεδομένων διαμήκους συμπεριφοράς οδήγησης για την επισήμανση διαμήκη DOC και προσδιορίζει γρήγορα τα στυλ οδήγησης με βάση αυτά τα μοντέλα DOC και MLC. Αυτή η μελέτη παρέχει μια αναφορά για online αναγνώριση του στυλ οδήγησης σε πραγματικό χρόνο σε οχήματα εξοπλισμένο με ενσωματωμένο εξοπλισμό λήψης δεδομένων, όπως π.χ ADAS.

Επίσης, η μελέτη Wang et al. (2022) πρότεινε ένα Bi-LSTM με μηχανισμό προσοχής για τον εντοπισμό της κατάστασης απόσπασης της προσοχής των οδηγών και αξιολόγησε αυτό το μοντέλο χρησιμοποιώντας φυσιοκρατικά δεδομένα οδήγησης από το SH-NDS. Πέντε μεταβλητές της οδηγικής απόδοσης (ταχύτητα, διαμήκης επιτάχυνση, πλευρική επιτάχυνση, μετατόπιση λωρίδας και ρυθμός τιμονιού) προσδιορίστηκαν ότι συσχετίζονται σε μεγάλο βαθμό με την οδήγηση με απόσπαση προσοχής μέσω τηλεφώνου. (Wang et al., 2022)

Στη συνέχεια, η μελέτη Osman et al. (2019) παρουσιάζει μια μεθοδολογία για τον προσδιορισμό των δευτερευουσών εργασιών ως μία από τις κύριες αιτίες της αποσπασματικής οδηγικής συμπεριφοράς. Η προτεινόμενη μεθοδολογία σε αυτή τη μελέτη χρησιμοποιεί παραμέτρους οδηγικής συμπεριφοράς (ταχύτητα, διαμήκης επιτάχυνση, πλευρική επιτάχυνση, θέση πεντάλ και ρυθμός εκτροπής) καθώς και τις τυπικές αποκλίσεις τους για να προσδιορίσει τους διαφορετικούς τύπους δευτερευουσών εργασιών στις οποίες επιδίδονται οι οδηγοί κατά την οδήγηση. Τρεις δευτερεύουσες εργασίες (που προσδιορίζονται στη βιβλιογραφία ως κρίσιμες για την αλλαγή της συμπεριφοράς στην οδήγηση) εξετάζονται: κλήσεις με κινητό τηλέφωνο, αποστολή μηνυμάτων σε κινητό τηλέφωνο και αλληλεπίδραση με έναν διπλανό επιβάτη.

Επιπρόσθετα, η εργασία των Yi et al. (2019) αντιμετώπισε το πρόβλημα της αναγνώρισης κατάστασης οδήγησης μαθαίνοντας από μη παρεμβατικές, εύκολα προσβάσιμες μετρήσεις σχετικά με το όχημα μέσω smartphone (ταχύτητα μέσω GPS, επιταχύνσεις τριών αξόνων και γωνίες προσανατολισμού μέσω IMU).

Στο άρθρο των Ghandour et al. (2021), οι μέθοδοι ταξινόμησης, μηχανικής μάθησης και τεχνητής νοημοσύνης χρησιμοποιούνται για την ταξινόμηση της συμπεριφοράς του οδηγού. Εξετάστηκαν δύο σύνολα δεδομένων: ανίχνευση λωρίδας και κατάσταση κυκλοφορίας. Αυτά τα δεδομένα συλλέχθηκαν από έξι διαφορετικές συμπεριφορές οδηγών που ταξινομήθηκαν σε τρεις καταστάσεις: φυσιολογική, υπνηλία και επιθετική. Τυχαιοποιήθηκαν για καλύτερη ανάλυση και εφαρμογή των μοντέλων μηχανικής μάθησης. Μια ολοκληρωμένη ανάλυση των χαρακτηριστικών από το σύνολο δεδομένων και η σχέση τους με την ταξινόμηση της ψυχικής κατάστασης του οδηγού είναι μια βασική συμβολή αυτής της εργασίας.

Στην έρευνα των Shangguan et al. (Shangguan et al, 2021) προτείνεται μία μεθοδολογία για την αξιολόγηση και την πρόβλεψη της κατάστασης κινδύνου που βρίσκεται ο οδηγός σε πραγματικό χρόνο. Μέσω της ανάπτυξης αλγορίθμων ομαδοποίησης καθορίζονται 4 στάδια επικινδυνότητας. Επιπλέον για την πρόβλεψη της κατάστασης κινδύνου αναπτύσσονται ορισμένοι αλγόριθμοι ταξινόμησης μηχανικής εκμάθησης. Αναλύοντας την επιρροή των μεταβλητών προκύπτει ότι η διαφορά ταχύτητας, η απόσταση από το προπορευόμενο όχημα, η ταχύτητα και η επιτάχυνση είναι ιδιαίτερα σημαντικές για την πρόβλεψη της κατάστασης επικινδυνότητας του οδηγού.

Η μελέτη των Shi et al. (2018) σχεδιάζει ένα πλαίσιο εξαγωγής και επιλογής χαρακτηριστικών, για την αξιολόγηση της οδήγησης οχήματος και την πρόβλεψη των επιπέδων κινδύνου. Το πλαίσιο ενσωματώνει την επιλογή χαρακτηριστικών που βασίζονται στην μάθηση, την αξιολόγηση κινδύνου χωρίς επίβλεψη και τη μη ισορροπημένη δειγματοληψία δεδομένων. Για την εκτίμηση των δυνατοτήτων κινδύνου των οχημάτων κατά την οδήγηση, προτείνεται η σήμανση δεδομένων χωρίς επίβλεψη. Με βάση τα εξαγόμενα χαρακτηριστικά των δεικτών κινδύνου, τα οχήματα ομαδοποιούνται σε διάφορα επίπεδα επικινδυνότητας με τεχνικές μη επιβλεπόμενης μάθησης. Έπειτα, γίνεται ο καθορισμός των βασικών χαρακτηριστικών σύμφωνα με την κατάταξη της σημασίας των χαρακτηριστικών και την αναδρομική εξάλειψη. Τα επίπεδα κινδύνου των οχημάτων κατά την οδήγηση προβλέπονται με βάση τα βασικά χαρακτηριστικά που έχουν επιλεγεί. Τα ευρήματα δείχνουν ότι αυτή η προσέγγιση είναι αποτελεσματική και αξιόπιστη για τον εντοπισμό σημαντικών χαρακτηριστικών για την αξιολόγηση οδήγησης και για την επίτευξη ακριβούς πρόβλεψης των επιπέδων κινδύνου.

Οι οδηγοί υψηλού κινδύνου είναι πιο πιθανό να εμπλακούν σε τροχαία ατυχήματα και το επίπεδο κινδύνου οδήγησης των οδηγών επηρεάζεται από παράγοντες όπως τα δημογραφικά στοιχεία και χαρακτηριστικά προσωπικότητας. Η μελέτη των Song et al. (2021) με βάση το μοντέλο δοκιμών εξισώσεων (SEM) διερευνά τις σχέσεις μεταξύ των δημογραφικών χαρακτηριστικών των οδηγών (φύλο, ηλικία και αθροιστικά έτη οδήγησης), της αναζήτησης περιπέτειας, της αντίληψης του κινδύνου και της επικίνδυνης οδηγικής συμπεριφοράς του οδηγού. Με την χρήση τεχνικών ομαδοποίησης ανάλογα το φύλο του οδηγού καθορίζονται 3 ομάδες επικινδυνότητας η 'χαμηλή', η 'μεσαία' και η 'υψηλή'. Τέλος, προτείνεται ένα μοντέλο ταξινόμησης του κινδύνου οδήγησης του οδηγού από την ταξινόμηση Τυχαίων Δασών (Random Forests).

Πίνακας 2.1: Ελλείψεις/Προτάσεις για μελλοντική διερεύνηση των ερευνών που παρουσιάστηκαν

Έρευνα	Ελλείψεις	Προτάσεις για μελλοντική διερεύνηση
Lyu et al., 2022	Τα κριτήρια βαθμονόμησης DOCs που αναπτύχθηκαν σε αυτή τη μελέτη ενδέχεται να μην είναι πλήρως εφαρμόσιμα για την αναγνώριση του στυλ οδήγησης σε όλους τους τύπους οδών.	Η μελλοντική έρευνα θα πραγματοποιήσει πιο φυσιοκρατική συλλογή δεδομένων οδήγησης για την επαλήθευση του μοντέλου. Ταυτόχρονα, είναι επίσης απαραίτητο να πραγματοποιηθεί δοκιμή πολλαπλών σεναρίων για να μελετηθεί η δυνατότητα εφαρμογής του μοντέλου σε πολλαπλά σενάρια.
Wang et al., 2022	Θα μπορούσαν να εξεταστούν περισσότεροι τύποι απόσπασης της προσοχής, όπως η οπτική/σωματική και η γνωστική απόσπαση, καθώς δραστηριότητες όπως το φαγητό, το ποτό και η συζήτηση με τους επιβάτες αποσπούν την προσοχή εξίσου με τη χρήση τηλεφώνου.	Με ταχέως αυξανόμενη υπολογιστική ισχύ, ενδέχεται να υπάρχουν άλλα μοντέλα βαθιάς μάθησης που είναι σε θέση να επεξεργάζονται δεδομένα χρονοσειρών πιο αποτελεσματικά προς διερεύνηση.
Osman et al., 2019	Αυτή η μελέτη δεν έλαβε υπόψη την επίδραση του τύπου οδοστρώματος, των γεωμετρικών χαρακτηριστικών και των χαρακτηριστικών του οχήματος στις μεταβλητές της οδηγικής συμπεριφοράς.	Μελλοντική έρευνα θα μελετήσει τον αντίκτυπο του τύπου οδοστρώματος, των γεωμετρικών χαρακτηριστικών και των χαρακτηριστικών του οχήματος στις μεταβλητές της οδηγικής συμπεριφοράς και στη δύναμη προβλεψιμότητας των αναπτυσσόμενων μοντέλων.
Yi et al., 2019	Μπορούν να συμπεριληφθούν περισσότερες σχετικές πληροφορίες, όπως κυκλοφοριακή κατάσταση, καιρικές συνθήκες και συνεχής χρόνος οδήγησης, απαιτώντας μεγαλύτερο όγκο δεδομένων.	Η διαδικτυακή μάθηση μπορεί να ενσωματωθεί στο προτεινόμενο πλαίσιο για μείωση της λανθασμένης ταξινόμησης της νυσταγμένης οδήγησης και ενίσχυση της ασφάλειας.

Έρευνα	Ελλείψεις	Προτάσεις για μελλοντική διερεύνηση
Ghandour et al., 2021	Δεν ελήφθησαν υπόψη παράγοντες όπως το όριο ταχύτητας του δρόμου και ο ψυχικός φόρτος εργασίας. Τα δεδομένα προήλθαν μόνο από αυτοκινητοδρόμους.	Περαιτέρω έρευνα μπορεί να βελτιώσει τα σύνολα δεδομένων, να συνδυάσει διαφορετικά δεδομένα για καλύτερη ταξινόμηση και να προσθέσει σταθμισμένη πιθανότητα στην κατηγοριοποίηση της ψυχικής κατάστασης των οδηγών.
Shangguan et al., 2021	Η μέθοδος πρόβλεψης κινδύνου εστιάζει μόνο στην παρακολούθηση αυτοκινήτου και δεν καλύπτει διαδικασίες αλλαγής λωρίδας ή προσπέρασης.	Η μελλοντική έρευνα θα συλλέξει δεδομένα από αλλαγές λωρίδας και προσπεράσεις μέσω NDS ή πραγματικών δοκιμών, ενώ αλγόριθμοι βαθιάς μάθησης μπορούν να συγκριθούν για καλύτερη πρόβλεψη κινδύνου.
Shi et al., 2018	Η έλλειψη δεδομένων σύγκρουσης καθιστά δύσκολη την επαλήθευση των συνδέσεων μεταξύ υψηλού επιπέδου κινδύνου και πραγματικών ατυχημάτων.	Προτείνεται εις βάθος εξαγωγή χαρακτηριστικών που καλύπτουν ευρύ φάσμα οδηγικής συμπεριφοράς και επικίνδυνων συνθηκών, όπως αλλαγές λωρίδας και συγκρούσεις.
Song et al., 2021	Τα δεδομένα της έρευνας ενδέχεται να είναι μεροληπτικά, καθώς προέρχονται από υποκειμενικά ερωτηματολόγια.	Η αξιολόγηση του επιπέδου κινδύνου των οδηγών μέσω ερωτηματολογίων μπορεί να βοηθήσει στον εντοπισμό υψηλού κινδύνου οδηγών για στοχευμένη εκπαίδευση και βελτίωση της ασφάλειας.

2.2.2 Αλγόριθμοι ταξινόμησης οδηγικής συμπεριφοράς

Στον τομέα της οδικής ασφάλειας, οι αλγόριθμοι μηχανικής εκμάθησης χρησιμοποιούνται όλο και πιο συχνά. Πολλές μελέτες επικεντρώνονται στην ταξινόμηση της οδηγικής συμπεριφοράς και στην ανάπτυξη μοντέλων για την πρόβλεψη συγκρούσεων σε πραγματικό χρόνο, εστιάζοντας στη χρήση διαφορετικών αλγορίθμων ταξινόμησης. Όπως φαίνεται στον Πίνακα 2.2, παρατίθενται οι αλγόριθμοι ταξινόμησης με την καλύτερη απόδοση από τις αντίστοιχες έρευνες.

Πίνακας 2.2: Έρευνες ανάλυσης οδηγικής συμπεριφοράς κι οι αποτελεσματικότεροι αλγόριθμοι ταξινόμησης

Έρευνα	Σκοπός αλγορίθμων ταξινόμησης	Αλγόριθμοι ταξινόμησης με το υψηλότερο ποσοστό προβλέψεων
Lyu et al., 2022	Αναγνώριση τύπου οδηγικής συμπεριφοράς με βάση 3 κατηγορίες: επιθετικός, μετριοπαθής, συντηρητικός	Multilayer Perceptron (MLP): ποσοστό ορθών προβλέψεων 70%

Έρευνα	Σκοπός αλγορίθμων ταξινόμησης	Αλγόριθμοι ταξινόμησης με το υψηλότερο ποσοστό προβλέψεων
Wang et al., 2022	Εντοπισμός της κατάστασης απόσπασης της προσοχής των οδηγών	Bidirectional Long Short-Term Memory (Bi-LSTM): ποσοστό ορθών προβλέψεων 91.2%
Osman et al., 2019	Αναγνώριση της κατάστασης του οδηγού με βάση 3 κατηγορίες απόσπασης προσοχής	Decision Tree (DT): ποσοστό ορθών προβλέψεων 77.3%
Yi et al., 2019	Αναγνώριση της κατάστασης του οδηγού με βάση 3 κατηγορίες	Random Forest (RF): ποσοστό ορθών προβλέψεων 91.6%
Ghandour et al., 2021	Αναγνώριση της κατάστασης του οδηγού με βάση 3 κατηγορίες: επιθετικός, νυσταγμένος, κανονικός	Gradient Boosting (GB): ποσοστό ορθών προβλέψεων 67%
Shangguan et al., 2021	Μέσω της ανάπτυξης αλγορίθμων ομαδοποίησης καθορίζονται 4 στάδια επικινδυνότητας: ασφαλής, χαμηλού, μεσαίου κι υψηλού κινδύνου αντίστοιχα	Multilayer Perceptron (MLP): ποσοστό ορθών προβλέψεων 89.2%
Shi et al., 2018	Ταξινόμηση της οδηγικής συμπεριφοράς σε 4 επίπεδα	eXtreme Gradient Boosting (XGBoost): ποσοστό ορθών προβλέψεων 89%
Song et al., 2021	Ταξινόμηση επικίνδυνης οδηγικής συμπεριφοράς σε 3 επίπεδα ασφαλείας	Random Forest (RF): ποσοστό ορθών προβλέψεων 90%

2.3. Πρόβλημα ανισορροπίας δεδομένων σε κάθε τάξη

Η ανισορροπία δεδομένων μεταξύ των τάξεων αποτελεί μια σημαντική δυσκολία στα προβλήματα ταξινόμησης, ειδικά όταν εφαρμόζονται αλγόριθμοι μηχανικής μάθησης. Το φαινόμενο αυτό προκύπτει όταν μία ή περισσότερες τάξεις περιλαμβάνουν αισθητά περισσότερα δείγματα από τις υπόλοιπες, γεγονός που μπορεί να προκαλέσει μεροληψία του μοντέλου υπέρ της κυρίαρχης τάξης και να μειώσει την ικανότητά του να προβλέψει σωστά τις λιγότερο συχνές τάξεις.

Σύμφωνα με τους Katrazakas et al. (2020), συμπεραίνεται ότι η ενσωμάτωση της τεχνικής Edited Nearest Neighbors (ENN) με την τεχνική Synthetic Minority Oversampling Technique (SMOTE) αποδίδει βελτιωμένα και ενδιαφέροντα αποτελέσματα όταν εφαρμόζεται σε πρόβλεψη σύγκρουσης σε πραγματικό χρόνο ή για την αξιολόγηση της ασφάλειας δεδομένων σε προσομοιωτή οδήγησης.

Μία από τις πιο διαδεδομένες λύσεις στο πρόβλημα ανισορροπίας δεδομένων, είναι η **υπερδειγματοληψία** (oversampling), όπου δημιουργούνται συνθετικά δείγματα για την ενίσχυση της μειονοτικής τάξης. Η μέθοδος **SMOTE** (Synthetic Minority Over-sampling Technique) είναι χαρακτηριστικό παράδειγμα αυτής της προσέγγισης. Σύμφωνα με μελέτη των Zhu et al. (2022), εξετάστηκαν διάφορες

παραλλαγές του SMOTE σε συνδυασμό με τεχνικές όπως το Adaboost, προκειμένου να βελτιωθεί η απόδοση των μοντέλων σε ανισόρροπα σύνολα δεδομένων.

Η τεχνική dropout (εγκατάλειψη) έχει σημειωθεί με μεγάλη επιτυχία στην εκπαίδευση βαθιών νευρωνικών δικτύων μηδενίζοντας ανεξάρτητα τις εξόδους των νευρώνων τυχαία, αντιμετωπίζοντας το πρόβλημα ανισορροπίας δεδομένων σε κάθε τάξη. Η εγκατάλειψη είναι μια απλή αλλά αποτελεσματική τεχνική για την πρόληψη της υπερπροσαρμογής στην εκπαίδευση των βαθιών νευρωνικών δικτύων.

Σύμφωνα με την έρευνα των Zhe Li et al. (2016) προτείνεται μια αποτελεσματική προσαρμοστική εγκατάλειψη (που ονομάζεται εξελικτική εγκατάλειψη) που υπολογίζει τις πιθανότητες δειγματοληψίας on-the-fly από μια μικρή παρτίδα παραδειγμάτων. Εμπειρικές μελέτες για πολλά σύνολα δεδομένων αναφοράς καταδεικνύουν ότι οι προτεινόμενες εγκαταλείψεις δεν επιτυγχάνουν μόνο πολύ ταχύτερη σύγκλιση αλλά και μικρότερο σφάλμα δοκιμής από την τυπική εγκατάλειψη.

2.4. Αξιολόγηση επίδρασης μεταβλητών στην επικίνδυνη οδήγηση

Η αξιολόγηση της επίδρασης διαφόρων μεταβλητών στην επικίνδυνη οδήγηση αποτελεί κρίσιμο βήμα για την κατανόηση των παραγόντων που συμβάλλουν στην πρόκληση ατυχημάτων. Αυτή η διαδικασία μπορεί να πραγματοποιηθεί μέσω στατιστικών μεθόδων και αλγορίθμων μηχανικής μάθησης, που επιτρέπουν την ανάλυση της σχέσης μεταξύ χαρακτηριστικών όπως η ταχύτητα, η επιτάχυνση, η προσοχή του οδηγού και η κυκλοφοριακή ροή.

Το πειραματικό αποτέλεσμα του LIME (NSC-2021) που δείχνει τη συμβολή των χαρακτηριστικών αυτών είναι εξαιρετικά σημαντικά για τον προσδιορισμό της συμπεριφοράς των οδηγών. Αυτό δίνει τοπική ερμηνευτικότητα, και επίσης παρέχει τρόπο για να καθορίσει ποιες παραλλαγές χαρακτηριστικών θα έχουν τη μεγαλύτερη επίδραση στην πρόβλεψη.

Σύμφωνα με την έρευνα Li et al. (2023), οι τεχνικές τεχνητής νοημοσύνης (AI) έχουν εφαρμοστεί ευρέως στον τομέα των αυτόνομων οχημάτων (AVs). Ωστόσο, οι υπάρχουσες τεχνικές τεχνητής νοημοσύνης, όπως η βαθιά μάθηση και η εκμάθηση συνόλου, έχουν επικριθεί για τη φύση τους με μαύρο κουτί. Το εξηγήσιμο AI είναι μια αποτελεσματική μεθοδολογία για την κατανόηση του μαύρου κουτιού και την οικοδόμηση εμπιστοσύνης του κοινού στα αυτόματα οχήματα. Σε αυτό το άρθρο, προτάθηκε μια εξήγηση πρόσθετου Shapley βάσει της μέγιστης εντροπίας (SHAP) για την εξήγηση της απόφασης αλλαγής λωρίδας (LC). Συγκεκριμένα, κατασκευάστηκε πρώτα ένα μοντέλο απόφασης LC με υψηλή ακρίβεια χρησιμοποιώντας την eXtreme Gradient Boosting.

Στη συνέχεια, για να εξηγηθεί το μοντέλο, προτάθηκε μια τροποποιημένη μέθοδος SHAP με την εισαγωγή μιας μέγιστης βασικής τιμής εντροπίας. Ο πυρήνας αυτής της μεθόδου είναι ο προσδιορισμός της βασικής τιμής του μοντέλου απόφασης LC χρησιμοποιώντας την αρχή της μέγιστης εντροπίας, η οποία παρέχει μια εξήγηση πιο συνεπή με την ανθρώπινη διαίσθηση. Αυτό συμβαίνει επειδή φέρνει δύο ιδιότητες: 1) η μέγιστη εντροπία έχει μια σαφή φυσική σημασία που ποσοτικοποιεί μια απόφαση από το χάος στη βεβαιότητα και 2) το άθροισμα των εξηγήσεων είναι πάντα ισότροπο και θετικό. Επιπλέον,

αναπτύχθηκε εξαντλητική στατιστική ανάλυση και οπτικοποίηση για να παρουσιαστούν διαισθητικές εξηγήσεις του μοντέλου απόφασης LC. Με βάση τα αποτελέσματα της επεξήγησης, αποδόθηκαν οι αιτίες των προβλέψεων με λανθασμένα αποτελέσματα σε ελαττώματα μοντέλου ή σε αραιότητα δειγμάτων, κάτι που παρέχει καθοδήγηση στους χρήστες για βελτιστοποίηση μοντέλων.

Σύμφωνα με την έρευνα Michelaraki et al. (2021) στο πλαίσιο του ερευνητικού έργου i-Dreams, ένας αλγόριθμος σημασίας χαρακτηριστικών που εξήχθη από την Ενίσχυση Ακραίας Κλίσης (XGBoost) χρησιμοποιήθηκε για την αξιολόγηση της σημασίας των μεταβλητών στην πρόβλεψη του επιπέδου “Ζώνης Ανοχής Ασφαλείας”. Για κάθε παράγοντα κινδύνου είναι γνωστό το επίπεδο “Ζώνης Ανοχής Ασφαλείας” και σε κάθε φάση στοχεύεται ένας συγκεκριμένος παράγοντας κινδύνου για την πρόβλεψη του επιπέδου ασφαλείας χωρίς να λαμβάνονται υπόψη άλλες εξίσου σημαντικές μεταβλητές. Επιπρόσθετα, εφαρμόστηκε ένα μοντέλο νευρωνικών δικτύων για την πρόβλεψη δεδομένων σε πραγματικό χρόνο, λαμβάνοντας υπόψη τους πιο σημαντικούς δείκτες κινδύνου.

Επιπλέον, διεξήχθη μια ολοκληρωμένη αξιολόγηση της απόδοσης τριών ταξινομητών μηχανικής μάθησης (δηλαδή Δέντρα απόφασης, Τυχαία Δάση και k-Πλησιότεροι Γείτονες) σε δύο ξεχωριστά σύνολα δεδομένων (δηλαδή σύνολο δεδομένων πειράματος επί δρόμου και προσομοιωτή) για την πρόβλεψη των επιπέδων STZ για πρόοδο. Τα αποτελέσματα έδειξαν ότι το μοντέλο RF ξεπέρασε τα μοντέλα DT και kNN σε όλες τις μετρήσεις, καθιστώντας το το πιο αποτελεσματικό για την πρόβλεψη προόδου με ακρίβεια έως και 90%. Αποκαλύφθηκε επίσης ότι τα Νευρωνικά Δίκτυα απέδειξαν ότι το επίπεδο του STZ μπορεί να προβλεφθεί με εξαιρετική ακρίβεια έως και 89,8%. Επίσης, για την αξιολόγηση της επίδρασης όλων των μεταβλητών στην επικίνδυνη οδήγηση εξετάζεται η χρήση του αλγόριθμου Lime (Local Interpretable Model-Agnostic Explanation) ως υποστηρικτικό εργαλείο για την ανάλυση της επιρροής των ανεξάρτητων μεταβλητών στην πρόβλεψη των μοντέλων βαθιάς εκμάθησης.

2.5. Σύνοψη

Με βάση τη βιβλιογραφική ανασκόπηση που πραγματοποιήθηκε, η παρούσα διπλωματική εργασία στοχεύει να συμβάλει περαιτέρω στη γνώση γύρω από τα **Ευφυή Μεταφορικά Συστήματα (ITS)**, εστιάζοντας στην ανάλυση και τον εντοπισμό της οδηγικής συμπεριφοράς.

Συγκεκριμένα, θα διερευνηθεί η επίδραση διαφόρων οδηγικών χαρακτηριστικών στην αναγνώριση των διαφορετικών επιπέδων οδηγικής συμπεριφοράς, καθώς και η μεταξύ τους συσχέτιση. Επιπλέον, αξιοποιώντας έξυπνα συστήματα και τεχνικές **μηχανικής μάθησης**, θα πραγματοποιηθεί ταξινόμηση της οδηγικής συμπεριφοράς σε τρία επίπεδα, χρησιμοποιώντας τα παρακάτω μοντέλα:

- **CatBoost** (Ενίσχυση Κατηγορικών Δεδομένων)
- **Random Forests** (Τυχαία Δάση)
- **LightGBM** (Ηπια Ενίσχυση Βαθμίδωσης)
- **MLP** (Πολυεπίπεδο Perceptron)

Η ανάλυση αυτή αναμένεται να προσφέρει πολύτιμες πληροφορίες για την κατανόηση της οδηγικής συμπεριφοράς και να συμβάλει στη βελτίωση των Ευφυών Μεταφορικών Συστημάτων.

Η αξιολόγηση της επίδρασης όλων των μεταβλητών στην επικίνδυνη οδήγηση θα πραγματοποιηθεί με τη χρήση του αλγόριθμου Shapley Values με τον οποίο:

- **Υπολογίζουμε την επίδρασή του στη συνολική απόφαση** εξετάζοντας πώς αλλάζει η πρόβλεψη όταν το χαρακτηριστικό περιλαμβάνεται ή αφαιρείται.
- **Λαμβάνουμε υπόψη όλους τους δυνατούς συνδυασμούς χαρακτηριστικών**, ώστε να εξασφαλίσουμε δίκαιη κατανομή της σημασίας κάθε μεταβλητής.
- **Υπολογίζουμε τη μέση συμβολή του χαρακτηριστικού σε όλα τα πιθανά σενάρια**, που είναι η τελική Shapley Value.

3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

3.1. Εισαγωγή

Σε αυτή την ενότητα παρουσιάζεται το θεωρητικό υπόβαθρο που αποτέλεσε τη βάση για την επεξεργασία και ανάλυση των δεδομένων. Αρχικά, αναλύονται οι μέθοδοι επεξεργασίας και επαναδειγματοληψίας, δεδομένου ότι οι κατηγορίες παρουσιάζουν ανισορροπία. Στη συνέχεια, περιγράφονται τα μοντέλα μηχανικής και βαθιάς μάθησης που σχεδιάστηκαν για την ταξινόμηση της οδηγικής συμπεριφοράς στα τρία επίπεδα της Ζώνης Ανοχής Ασφαλείας. Επιπλέον, η μέθοδος SHAP αξιοποιείται ως εργαλείο για την εκτίμηση της επίδρασης διαφόρων παραγόντων στην οδήγηση υπό επικίνδυνες συνθήκες, εξετάζοντας τον ρόλο των ανεξάρτητων μεταβλητών στις προβλέψεις των μοντέλων. Τέλος, δίνεται ιδιαίτερη βαρύτητα στη σημασία των μετρικών αξιολόγησης, των κριτηρίων αποδοχής των μοντέλων και των στατιστικών ελέγχων των αποτελεσμάτων.

3.2. Επιλογή χαρακτηριστικών (Feature Selection)

Η **επιλογή χαρακτηριστικών (feature selection)** είναι μια σημαντική διαδικασία στη μηχανική μάθηση και την ανάλυση δεδομένων, η οποία στοχεύει στην επιλογή των πιο σχετικών μεταβλητών για τη δημιουργία αποδοτικών και ακριβών μοντέλων. Αντί να χρησιμοποιούνται όλα τα διαθέσιμα χαρακτηριστικά, η διαδικασία αυτή επικεντρώνεται στην αναγνώριση εκείνων που επηρεάζουν ουσιαστικά τη μεταβλητή στόχο, απορρίπτοντας παράλληλα τα λιγότερο σημαντικά ή θορυβώδη δεδομένα.

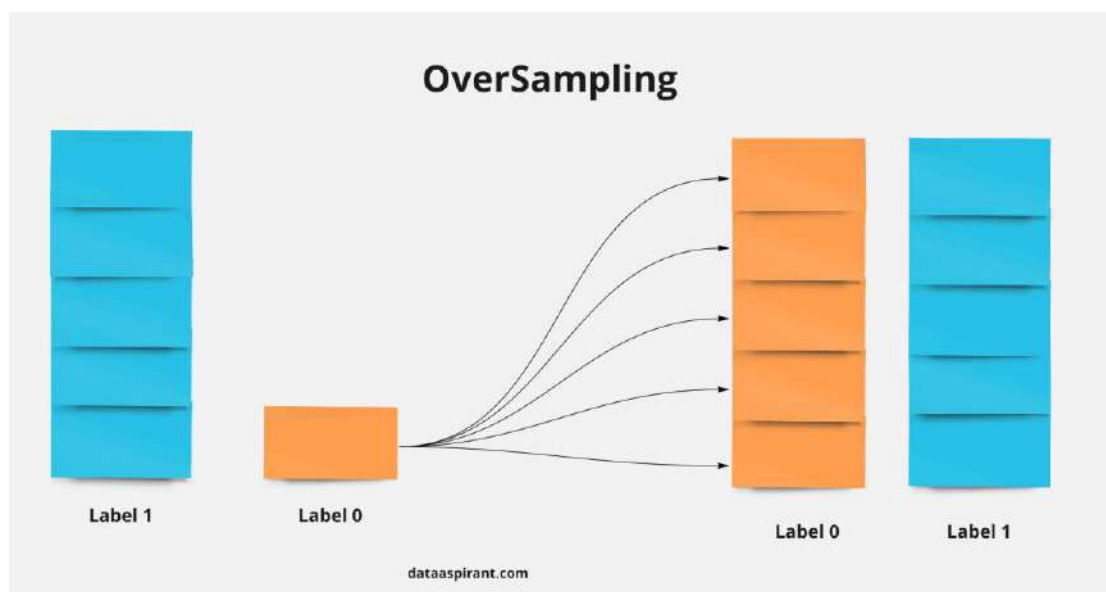
Σε αυτήν την εργασία, χρησιμοποιήθηκε η εκτίμηση της σημαντικότητας των χαρακτηριστικών (feature importance) με χρήση του αλγορίθμου **Random Forest**. Η συγκεκριμένη μέθοδος είναι μια από τις πιο διαδεδομένες τεχνικές για την αναγνώριση των πιο κρίσιμων μεταβλητών σε ένα πρόβλημα μηχανικής μάθησης. Ο αλγόριθμος **Random Forest**, αποτελούμενος από πολλαπλά δέντρα απόφασης, αξιολογεί τη συμβολή κάθε χαρακτηριστικού στην απόδοση του μοντέλου, επιτρέποντας την ταξινόμησή τους ως προς τη σημασία τους και τη δυνητική αφαίρεση όσων έχουν μικρή επιρροή.

3.3. Μέθοδοι επαναδειγματοληψίας για προβλήματα ανισορροπίας ταξινόμησης

Στην περίπτωση που εξετάζεται, η **ανισορροπία των δεδομένων** προκύπτει επειδή οι κλάσεις 1 και 2, που αντιστοιχούν σε επικίνδυνη οδήγηση, έχουν σημαντικά λιγότερα δείγματα σε σχέση με την κλάση 0 που αντιστοιχεί σε ασφαλή τρόπο οδήγησης. Αυτό μπορεί να οδηγήσει σε **μεροληψία του μοντέλου** προς την κυρίαρχη κλάση, με αποτέλεσμα να έχει χαμηλή απόδοση στην πρόβλεψη των υποεκπροσωπούμενων κατηγοριών. Αυτό συμβαίνει διότι,

οι περισσότεροι αλγόριθμοι μηχανικής εκμάθησης που χρησιμοποιούνται για ταξινόμηση, είναι βασισμένοι στην θεώρηση ότι όλες οι κλάσεις έχουν τον ίδιο αριθμό δεδομένων.

Μια κοινή προσέγγιση για την επίλυση του συγκεκριμένου προβλήματος, σύμφωνα με την έρευνα Pérez-Ortiz et al. (2015), είναι η υπερδειγματοληψία της μειονοτικής τάξης μέσω του κυρτού συνδυασμού των προτύπων της. Όπως αποτυπώνεται στο Γράφημα 3.1, σκοπός είναι η υπερδειγματοληψία δεδομένων που ανήκουν στην κλάση μειοψηφίας, στην περίπτωση μας των κλάσεων 1 και 2, που αντιστοιχούν σε επικίνδυνη οδηγική συμπεριφορά.



Γράφημα 3.1: Επαναδειγματοληψία δεδομένων που ανήκουν στην κλάση μειοψηφίας

Πηγή: Dataaspirant

3.3.1. Τεχνική Συνθετικής Μειονοτικής Υπερδειγματοληψίας (SMOTE)

Το **SMOTE** (Τεχνική Συνθετικής Υπερδειγματοληψίας για τη Μειοψηφική Τάξη) είναι μια από τις πιο διαδεδομένες μεθόδους **υπερδειγματοληψίας (oversampling)** για την αντιμετώπιση της ανισορροπίας σε προβλήματα ταξινόμησης. Σε αντίθεση με την απλή τυχαία υπερδειγματοληψία, όπου απλώς αντιγράφονται υπάρχοντα δείγματα, το **SMOTE δημιουργεί νέα, συνθετικά δεδομένα** βασισμένα στις γειτονικές τιμές των υπαρχόντων δειγμάτων της μειοψηφικής τάξης.

Η μέθοδος **SMOTE** εφαρμόζεται σε τρία κύρια στάδια για τη δημιουργία συνθετικών δειγμάτων και την εξισορρόπηση των κλάσεων. Αρχικά, για κάθε δείγμα που ανήκει στη μειοψηφική τάξη, η μέθοδος επιλέγει τυχαία **ένα ή περισσότερα γειτονικά δείγματα**, χρησιμοποιώντας την απόσταση των **k-πλησιέστερων γειτόνων (k-NN)**. Αντί να αντιγράφει υπάρχοντα δείγματα,

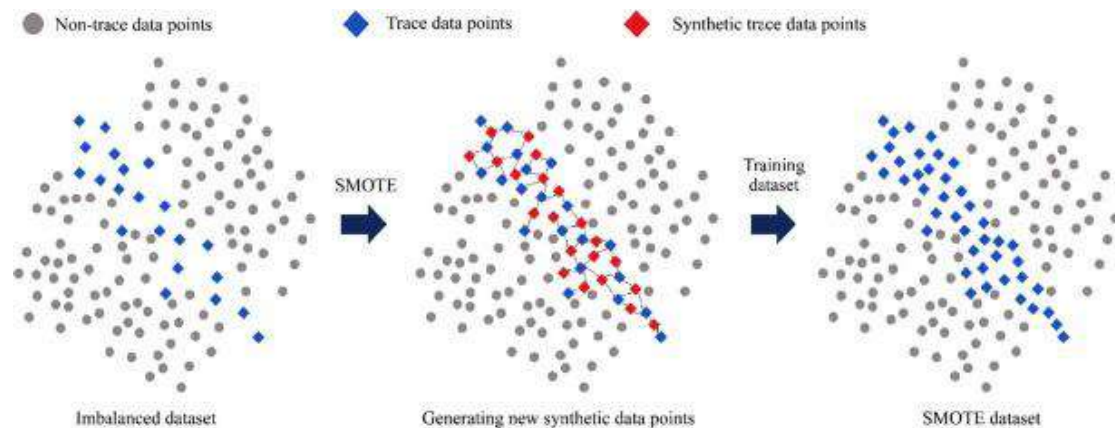
δημιουργεί νέα δεδομένα σε τυχαίες θέσεις μεταξύ του αρχικού δείγματος και των επιλεγμένων γειτόνων.

Η παραγωγή αυτών των νέων σημείων ακολουθεί τον εξής μαθηματικό τύπο:

$$x_{\text{new}} = x_{\text{original}} + \lambda \times (x_{\text{neighbor}} - x_{\text{original}})$$

όπου λ είναι ένας τυχαίος αριθμός στο διάστημα **[0,1]**. Με αυτόν τον τρόπο, το νέο δείγμα τοποθετείται ανάμεσα στο αρχικό και στο γειτονικό σημείο, διατηρώντας τα χαρακτηριστικά της μειοψηφικής τάξης και αποτρέποντας την υπερβολική επανάληψη των ίδιων δεδομένων.

Η διαδικασία αυτή συνεχίζεται έως ότου προστεθεί ο απαιτούμενος αριθμός νέων δειγμάτων, διαμορφώνοντας μια πιο ισορροπημένη κατανομή των κλάσεων στο σύνολο δεδομένων σύμφωνα με την έρευνα Chawla et al. (2002). Στην περίπτωση της εργασίας, επιθυμητός αριθμός για την κλάση 1 τέθηκαν τα 30000 δείγματα, ενώ για την κλάση 2 τα 15000 δείγματα.



Γράφημα 3.2: Διαδικασία που χρησιμοποιείται στο SMOTE: Synthetic Minority Oversampling Technique

Πηγή: ResearchGate (2023)

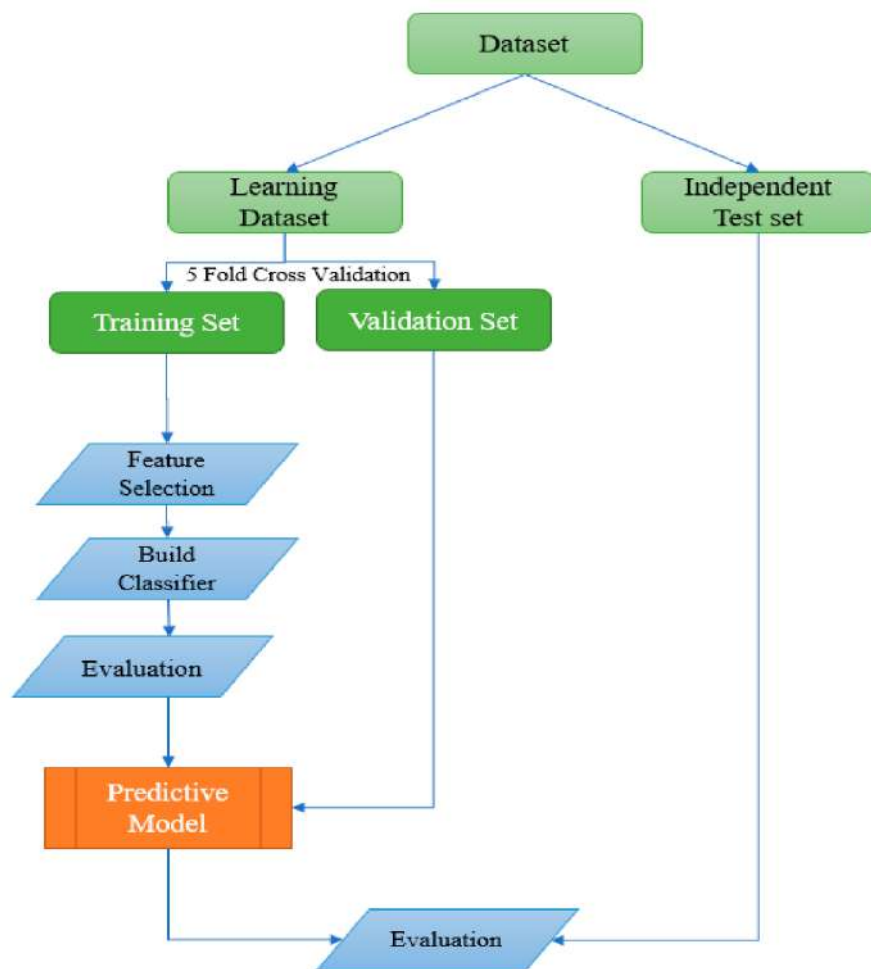
3.4. Αλγόριθμοι ταξινόμησης

Στο πλαίσιο της παρούσας εργασίας, αναπτύχθηκαν τέσσερις αλγόριθμοι μηχανικής μάθησης με στόχο την κατηγοριοποίηση δεδομένων σε πολλαπλές κλάσεις, δηλαδή διενεργήθηκε Πολυταξική Ταξινόμηση (Multiclass Classification).

Για να συμβεί αυτό, πρώτα έγινε διαχωρισμός των δεδομένων σε **σύνολο εκπαίδευσης (training dataset)** και **σύνολο ελέγχου (test dataset)**. Το σύνολο εκπαίδευσης χρησιμοποιείται για την προσαρμογή του μοντέλου στα δεδομένα, ενώ το σύνολο ελέγχου αξιολογεί την απόδοσή του σε μη ορατά δεδομένα, αποτρέποντας έτσι το φαινόμενο της υπερεκπαίδευσης (overfitting). Έτσι, το **80%** των δεδομένων διατέθηκε για εκπαίδευση και το **20%** για έλεγχο.

Η αποτελεσματικότητα των αλγορίθμων ταξινόμησης κρίθηκε μέσω διαφόρων **μετρικών αξιολόγησης**, οι οποίες αποτυπώνουν την ικανότητα του μοντέλου να κατηγοριοποιεί σωστά τα δεδομένα.

Το Γράφημα 3.3 απεικονίζει τη διαδικασία δημιουργίας, αξιολόγησης και εφαρμογής των μοντέλων μηχανικής μάθησης.



Γράφημα 3.3: Διάγραμμα ροής λειτουργίας αλγορίθμων μηχανικής μάθησης

Πηγή: Alruwais N, Zakariah M. *Evaluating Student Knowledge Assessment Using Machine Learning Techniques. Sustainability.* (2023)

3.4.1. Random Forest (Τυχαίο Δάσος)

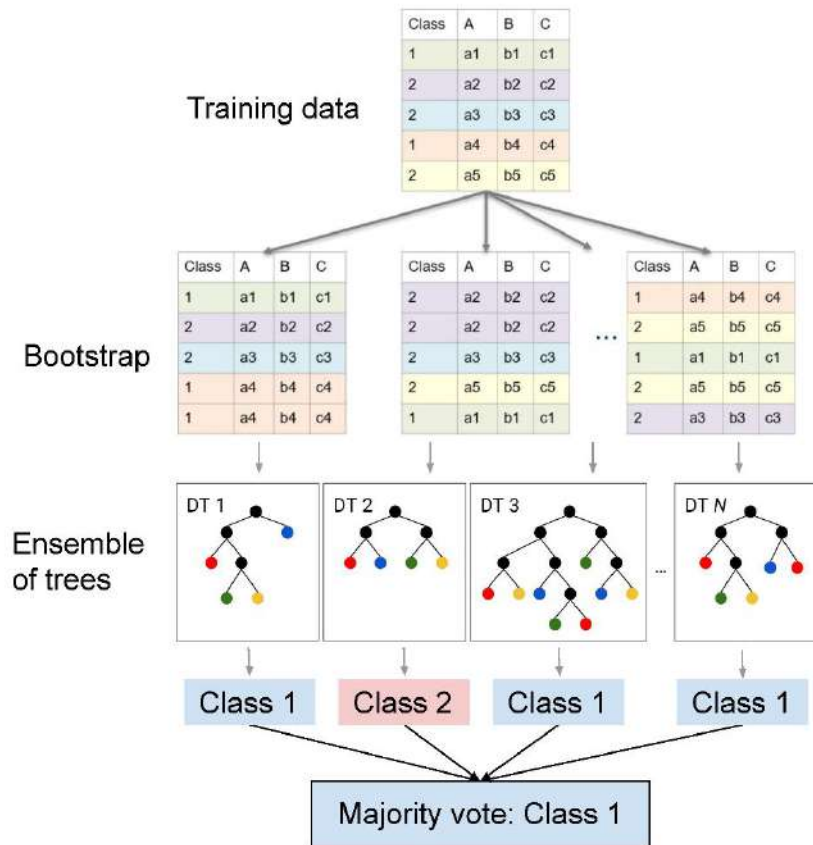
Το **Random Forest** είναι μια μέθοδος μηχανικής μάθησης που χρησιμοποιεί πολλά δέντρα απόφασης για να κάνει πιο αξιόπιστες προβλέψεις. Σύμφωνα με την έρευνα Zakariah et al. (2014), ο αλγόριθμος τυχαίου δάσους χρησιμοποιείται για την ταξινόμηση μεγάλων συνόλων δεδομένων σε διάφορες εφαρμογές. Πιο συγκεκριμένα, αυτό συμβαίνει καθώς αποτελεί μια τεχνική ταξινόμησης τύπου **ensemble** στην **εξόρυξη δεδομένων (data mining)**, όπου ο συνδυασμός πολλών μοντέλων οδηγεί σε καλύτερη ακρίβεια και μειώνει τον

κίνδυνο υπερπροσαρμογής (overfitting). Δημιουργεί πολλαπλά δέντρα, όπου κάθε δέντρο αναπτύσσεται με ένα σύνολο δειγμάτων που επιλέγονται τυχαία. Σε κάθε κόμβο, επιλέγεται ένα υποσύνολο χαρακτηριστικών ώστε να βρεθεί το πιο σημαντικό χαρακτηριστικό για τη διάσπαση του κόμβου σε δύο μέρη, με βάση το κέρδος πληροφορίας (information gain) που προκύπτει.

Μεταξύ άλλων αναφέρεται επίσης στην έρευνα Zakariah et al. (2014), πως ο αλγόριθμος του Τυχαίου Δάσους:

- Διαθέτει απaráμιλλη ακρίβεια σε σύγκριση με τους τρέχοντες αλγορίθμους.
- Εκτελείται αποδοτικά σε μεγάλες βάσεις δεδομένων.
- Μπορεί να διαχειριστεί χιλιάδες μεταβλητές εισόδου χωρίς να απαιτείται η διαγραφή τους.
- Παρέχει εκτιμήσεις για το ποιες μεταβλητές είναι σημαντικές στην ταξινόμηση.
- Δημιουργεί μια εσωτερική, αμερόληπτη εκτίμηση του γενικού σφάλματος καθώς προχωρά η κατασκευή του δάσους.
- Διαθέτει μια αποτελεσματική μέθοδο εκτίμησης των ελλιπών δεδομένων και διατηρεί υψηλή ακρίβεια, ακόμα και όταν λείπει μεγάλο μέρος των δεδομένων.
- Περιλαμβάνει τεχνικές για την εξισορρόπηση του σφάλματος σε σύνολα δεδομένων με ανισομερή κατανομή των κλάσεων.
- Τα παραγόμενα δάση μπορούν να αποθηκευτούν για μελλοντική χρήση σε άλλα δεδομένα.
- Υπολογίζει πρωτότυπα που παρέχουν πληροφορίες για τη σχέση μεταξύ των μεταβλητών και της ταξινόμησης.
- Υπολογίζει εγγύτητες μεταξύ ζευγών περιπτώσεων, οι οποίες μπορούν να χρησιμοποιηθούν για ομαδοποίηση, εντοπισμό ακραίων τιμών ή (με κατάλληλη κλιμάκωση) για την οπτικοποίηση των δεδομένων.
- Οι παραπάνω δυνατότητες μπορούν να επεκταθούν σε μη επισημασμένα δεδομένα, επιτρέποντας μη επιβλεπόμενη ομαδοποίηση, προβολές δεδομένων και ανίχνευση ακραίων τιμών.
- Προσφέρει μια πειραματική μέθοδο για την ανίχνευση αλληλεπιδράσεων μεταξύ μεταβλητών.

Παρακάτω, στο γράφημα 3.4 παρουσιάζεται η διαδικασία ταξινόμησης του Τυχαίου Δάσους.



Γράφημα 3.4: Διαδικασία ταξινόμησης Τυχαίου Δάσους

Πηγή: Introduction to Remote Sensing

3.4.2. Light Gradient Boosting Machine (Ελαφρύς Αλγόριθμος Βαθιάς Ενίσχυσης)

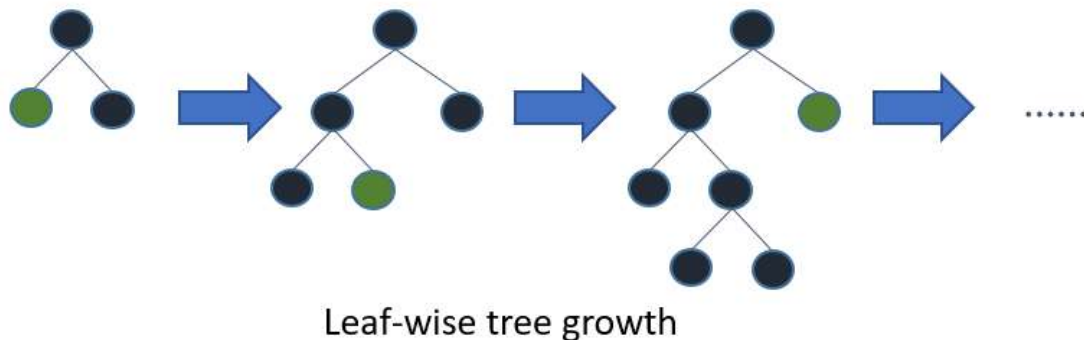
Ο **LightGBM** (Light Gradient Boosting Machine) είναι ένας αλγόριθμος ενισχυτικής μάθησης (boosting) που εφαρμόζεται κυρίως σε επιβλεπόμενα μαθησιακά προβλήματα, όπως η ταξινόμηση και η παλινδρόμηση. Δημιουργήθηκε από τη Microsoft και θεωρείται μία από τις κορυφαίες επιλογές στη μηχανική μάθηση, χάρη στην υψηλή ταχύτητα εκπαίδευσης και την ακρίβειά του. Ο LightGBM χρησιμοποιεί μια μέθοδο ενισχυτικής μάθησης, που έρχεται σε αντίθεση με την παραδοσιακή προσέγγιση επιπέδων (level-wise), με δέντρα αποφάσεων τύπου φύλλων (leaf-wise), η οποία μειώνει την κατανάλωση μνήμης ενώ βελτιώνει την αποδοτικότητα του μοντέλου.

Η θεμελιώδης διαφορά έγκειται στον τρόπο με τον οποίο αναπτύσσεται το δέντρο και επεκτείνονται οι κλάδοι του. Το δέντρο επεκτείνεται επιλέγοντας το φύλλο που προσφέρει το μεγαλύτερο όφελος σε κάθε στάδιο ανάπτυξης. Με άλλα λόγια, σε κάθε βήμα προστίθεται μόνο ένας νέος κόμβος φύλλου. Αυτό έχει ως αποτέλεσμα τη δημιουργία βαθύτερων αλλά πιο στενών δέντρων, τα οποία μπορούν να αποτυπώσουν πιο περίπλοκες σχέσεις μεταξύ των

χαρακτηριστικών. Σε σύγκριση με τη μέθοδο που βασίζεται σε επίπεδα (level-wise), αυτή η προσέγγιση είναι γενικά πιο γρήγορη.

Πιο συγκεκριμένα, σύμφωνα με πειράματα στην έρευνα των Guolin Ke et al. (2017), δημόσια σύνολα δεδομένων δείχνουν ότι ο LightGBM μπορεί να επιταχύνει τη διαδικασία εκπαίδευσης έως και 20 φορές, διατηρώντας σχεδόν την ίδια ακρίβεια. Επίσης, στην παραπάνω έρευνα, παρουσιάζεται μια προσέγγιση που ενσωματώνει δύο καινοτόμες τεχνικές, το Gradient-based One Side Sampling (GOSS) και το Exclusive Feature Bundling (EFB), για να ξεπεράσει τους περιορισμούς της παραδοσιακής μεθόδου ιστογράμματος που χρησιμοποιείται στους αλγορίθμους Gradient Boosting Decision Tree (GBDT).

Ο τρόπος ανάπτυξης του αλγορίθμου ταξινόμησης LightGBM φαίνεται στο Γράφημα 3.5.



Γράφημα 3.5: Τρόπος ανάπτυξης δέντρου αποφάσεων τύπου φύλλων, αλγόριθμος ταξινόμησης LightGBM

Πηγή: LightGBM

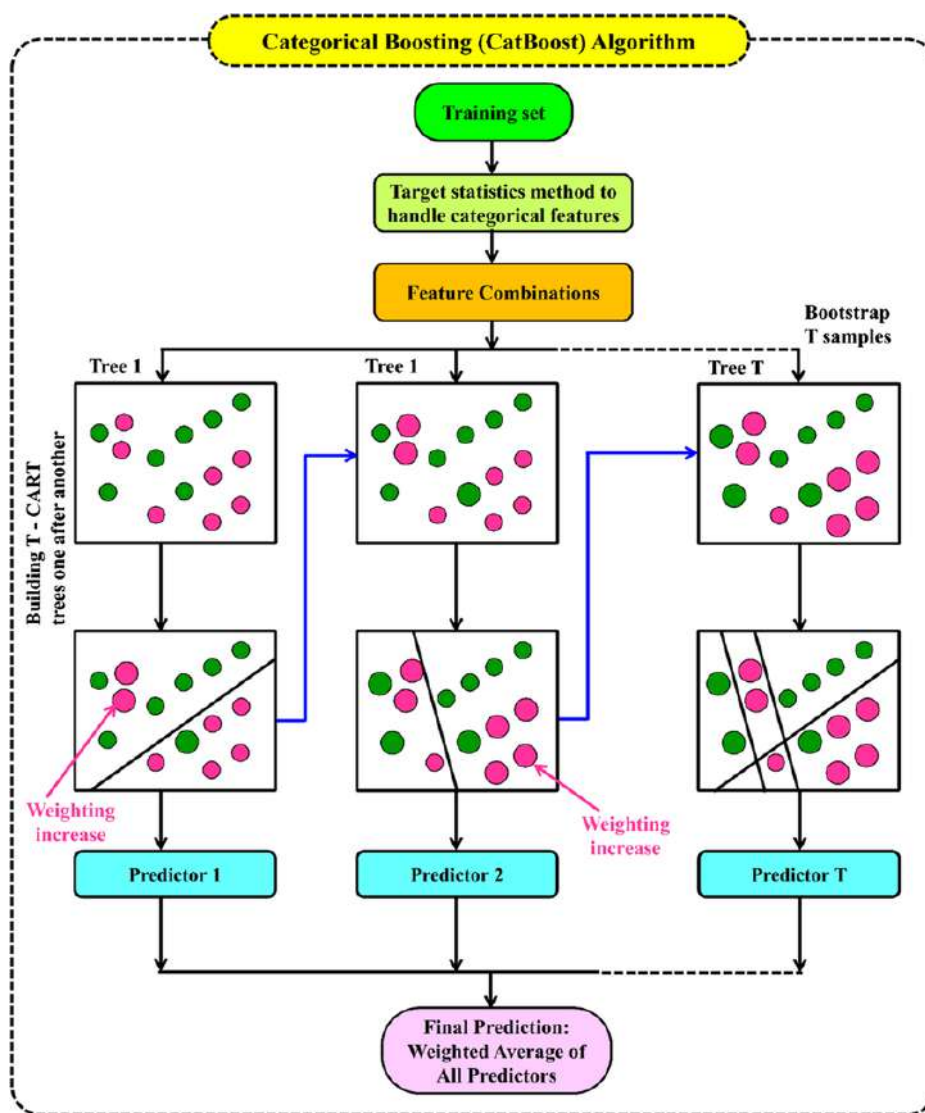
3.4.3. Categorical Boosting (Αλγόριθμος Ενίσχυσης Κατηγοριών)

Ο **CatBoost (Categorical Boosting)** είναι ένας αλγόριθμος ενισχυτικής μάθησης (gradient boosting), που αναπτύχθηκε από τη Yandex και χρησιμοποιείται για προβλήματα ταξινόμησης, παλινδρόμησης και άλλες εφαρμογές μηχανικής μάθησης. Ξεχωρίζει για την αποδοτική διαχείριση κατηγορηματικών δεδομένων, καθώς δεν απαιτεί one-hot encoding (κωδικοποίηση one-hot) ή label encoding (κωδικοποίηση ετικετών), ενώ προσφέρει υψηλή ακρίβεια και γρήγορη εκπαίδευση. Ο CatBoost (Category Boosting) είναι ένας εξειδικευμένος αλγόριθμος δέντρων απόφασης, ο οποίος ανήκει στην οικογένεια των αλγορίθμων gradient boosting. Η βασική του ιδέα είναι να δημιουργήσει ένα αποδοτικό μοντέλο κατασκευάζοντας ένα σύνολο από ασθενείς προβλεπτές δέντρων απόφασης. Αυτά τα δέντρα συνδυάζονται διαδοχικά σε κάθε βήμα, χρησιμοποιώντας τα υπόλοιπα σφάλματα (residuals) του προηγούμενου βήματος. Έτσι, η απόδοση της ταξινόμησης του μοντέλου βελτιώνεται σταδιακά, βήμα προς βήμα. Για δυαδική ταξινόμηση, ο CatBoost χρησιμοποιεί τη συνάρτηση απώλειας log loss, αλλά υπάρχει επίσης η

δυνατότητα χρήσης μιας προσαρμοσμένης (custom) συνάρτησης απώλειας. Η τελική κλάση που αντιστοιχεί σε μια νέα είσοδο καθορίζεται από έναν συνδυασμό των προβλέψεων των επιμέρους δέντρων του συνόλου (ensemble).

Όπως αναφέρεται στην έρευνα Ibrahim et al. (2020), πρόκειται για έναν αλγόριθμο μηχανικής μάθησης που είναι αποδοτικός στην πρόβλεψη κατηγορηματικών χαρακτηριστικών. Στην προαναφερθείσα έρευνα, συγκρίθηκε με άλλους αλγόριθμους ταξινόμησης όπως ο Random Forest και το AdaBoost, και τα πήγε αρκετά καλά, πετυχαίνοντας την υψηλότερη βαθμολογία (ακρίβεια) ενώ σε σύνολο κι οι υπόλοιπες μετρικές αξιολόγησης του μοντέλου ήταν εξίσου καλές με των υπολοίπων.

Στο Γράφημα 3.6 παρουσιάζεται το διάγραμμα ροής της διαδικασίας ανάπτυξης του αλγορίθμου CatBoost.



Γράφημα 3.6: Διάγραμμα ροής ανάπτυξης αλγορίθμου Categorical Boosting

Πηγή: ResearchGate (2023)

3.4.4. Multi-Layer Perceptron (Πολυεπίπεδο Αντιληπτικό)

Τα **πολυεπίπεδα perceptrons (MLPs)** αποτελούν έναν από τους πιο σημαντικούς τύπους νευρωνικών δικτύων, καθώς έχουν εφαρμοστεί με επιτυχία σε πληθώρα περιπτώσεων.

Η δομή ενός MLP περιλαμβάνει:

- Ένα επίπεδο εισόδου (Input Layer) που λαμβάνει τα δεδομένα,
- Ένα ή περισσότερα κρυφά επίπεδα (Hidden Layers), όπου γίνεται η επεξεργασία,
- Ένα επίπεδο εξόδου (Output Layer), το οποίο παράγει το τελικό αποτέλεσμα.

Κάθε επίπεδο αποτελείται από τουλάχιστον έναν νευρώνα.

Η είσοδος περνά μέσα από το δίκτυο ακολουθώντας τη διαδικασία της προωθητικής διάδοσης (forward propagation), καθώς μεταφέρεται από το ένα επίπεδο στο επόμενο.

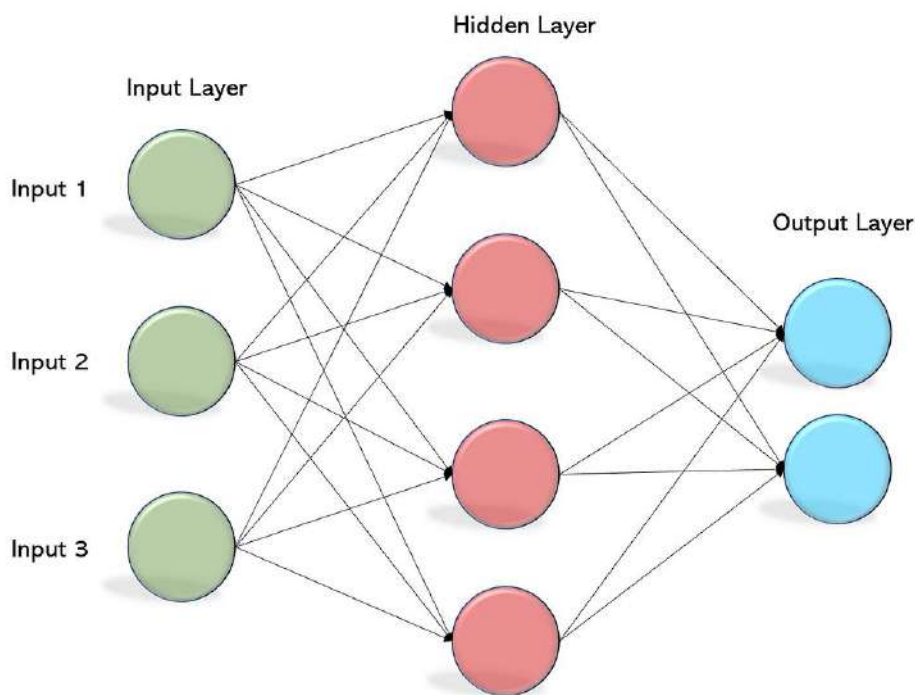
Τα πολυεπίπεδα perceptrons (MLPs) είναι πλήρως συνδεδεμένα δίκτυα, όπου κάθε κόμβος συνδέεται με όλους τους κόμβους του προηγούμενου και του επόμενου επιπέδου.

Στο επίπεδο εισόδου, κάθε κόμβος αντιστοιχεί σε μια ανεξάρτητη μεταβλητή. Οι έξοδοι ενός επιπέδου λειτουργούν ως είσοδοι για το επόμενο, και αυτή η διαδικασία συνεχίζεται διαδοχικά μέχρι να επιτευχθεί η έξοδος του δικτύου.

Για την εκπαίδευσή του, το MLP χρησιμοποιεί την αλγόριθμο back-propagation, μια επιτηρούμενη μέθοδο μάθησης.

Το perceptron υπολογίζει μία έξοδο συνδυάζοντας γραμμικά τις εισόδους του, λαμβάνοντας υπόψη τα αντίστοιχα βάρη, και στη συνέχεια εφαρμόζει (αν χρειάζεται) μια μη γραμμική συνάρτηση ενεργοποίησης.

Στο Γράφημα 3.7 απεικονίζεται η δομή ενός Multi-Layer Perceptron (MLP).



Γράφημα 3.7: Μοντέλο ταξινόμησης Multi-Layer Perceptron (MLP) με τα επίπεδα που αναφέρονται παραπάνω, στο 3.4.4.

Πηγή: Medium (2019)

3.5. Μετρικές αξιολόγησης Μοντέλων Ταξινόμησης

3.5.1. Μήτρα Σύγχυσης (Confusion Matrix)

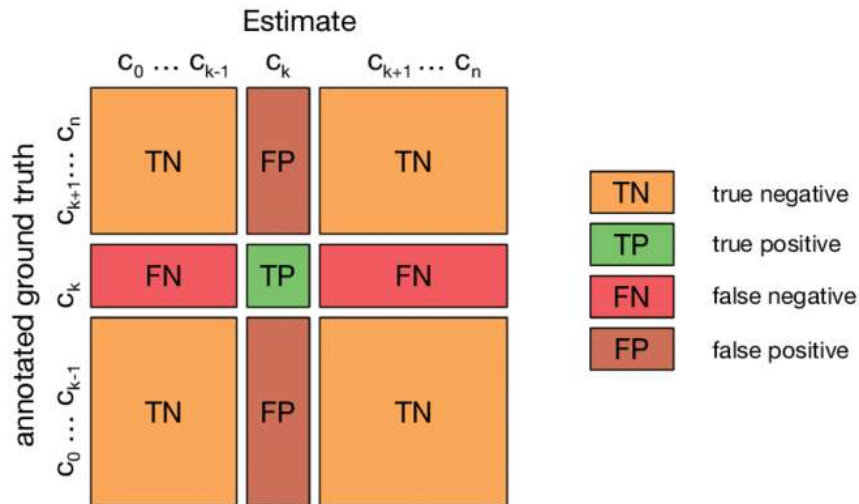
Μία Μήτρα Σύγχυσης είναι ένας πίνακας που χρησιμοποιείται για να περιγράψει την απόδοση ενός μοντέλου ταξινόμησης. Συγκρίνει τις πραγματικές τιμές-στόχους με αυτές που προβλέπει το μοντέλο. Αυτή η σύγκριση πραγματοποιείται για όλες τις κατηγορίες του συνόλου δεδομένων, παρέχοντας μια λεπτομερή ανάλυση της απόδοσης του μοντέλου. Η Μήτρα Σύγχυσης προσφέρει μια πιο λεπτομερή εικόνα της απόδοσης ενός μοντέλου σε σύγκριση με ένα απλό ποσοστό ακρίβειας. Δεν δείχνει μόνο πόσες προβλέψεις ήταν σωστές, αλλά και πού το μοντέλο κάνει λάθη και τι είδους σφάλματα προκύπτουν. Αυτή η πληροφορία είναι ζωτικής σημασίας για τη βελτίωση της απόδοσης του μοντέλου, ιδιαίτερα όταν ορισμένα είδη λαθών έχουν μεγαλύτερο κόστος από άλλα.

Σε ένα πρόβλημα δυαδικής ταξινόμησης, η Μήτρα Σύγχυσης αποτελείται από τέσσερα βασικά στοιχεία:

- **Πραγματικά Θετικά (True Positive (TP)):** Ο αριθμός των περιπτώσεων όπου το μοντέλο προέβλεψε σωστά τη θετική κλάση.

- **Ψευδώς Θετικά (False Positive (FP)):** Ο αριθμός των περιπτώσεων όπου το μοντέλο προέβλεψε εσφαλμένα τη θετική κλάση, ενώ στην πραγματικότητα ήταν αρνητική. Γνωστό και ως Σφάλμα Τύπου I.
- **Ψευδώς Αρνητικά (False Negative (FN)):** Ο αριθμός των περιπτώσεων όπου το μοντέλο προέβλεψε εσφαλμένα την αρνητική κλάση, ενώ στην πραγματικότητα ήταν θετική. Γνωστό και ως Σφάλμα Τύπου II.
- **Πραγματικά Αρνητικά (True Negative (TN)):** Ο αριθμός των περιπτώσεων όπου το μοντέλο προέβλεψε σωστά την αρνητική κλάση.

Το παραπάνω ανάγεται και σε πρόβλημα πολύ-κατηγορικής ταξινόμησης όπως φαίνεται στο Γράφημα 3.8.



Γράφημα 3.8: Μήτρα Σύγχυσης προβλήματος πολύ-κατηγορικής ταξινόμησης

Πηγή: ResearchGate (2016)

3.5.2. Ακρίβεια (Accuracy)

Η ακρίβεια (accuracy) είναι ένας θεμελιώδης δείκτης απόδοσης ενός μοντέλου ταξινόμησης. Εκφράζει το ποσοστό των περιπτώσεων όπου το μοντέλο έκανε σωστή πρόβλεψη, σε σχέση με το σύνολο των περιπτώσεων:

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

Η ακρίβεια είναι καλή μετρική μόνο αν οι κλάσεις είναι ισορροπημένες. Σε ανισορροπημένα δεδομένα, είναι προτιμότερο να εξετάζουμε την ανάκληση, την

ακρίβεια θετικής πρόβλεψης και την f1-score για μια πιο πλήρη εικόνα της απόδοσης του μοντέλου.

3.5.3. Ακρίβεια Θετικής Πρόβλεψης (Precision)

Η ακρίβεια θετικής πρόβλεψης (precision) είναι μια σημαντική μετρική που δείχνει πόσο αξιόπιστες είναι οι θετικές προβλέψεις του μοντέλου.

Ορίζεται ως:

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

Η ακρίβεια θετικής πρόβλεψης είναι ιδιαίτερα σημαντική όταν το κόστος των ψευδώς θετικών (FP) είναι υψηλό.

3.5.4. Ανάκληση (Recall)

Η ανάκληση (recall), γνωστή και ως ευαισθησία (sensitivity) ή ρυθμός αληθώς θετικών (true positive rate - TPR), είναι μια μετρική που δείχνει πόσο καλά το μοντέλο εντοπίζει τα θετικά παραδείγματα.

Ορίζεται ως:

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

3.5.5. Ρυθμός Λανθασμένου Συναγερμού (False Alarm Rate)

Ο ρυθμός ψευδών συναγερμών (False Alarm Rate - FAR), γνωστός και ως ρυθμός ψευδώς θετικών (False Positive Rate - FPR), είναι μια μετρική που δείχνει το ποσοστό των αρνητικών περιπτώσεων που ταξινομήθηκαν λανθασμένα ως θετικές από το μοντέλο.

Ορίζεται ως:

$$FAR = \frac{FP}{FP + TN}$$

3.5.6. f1-score

Το F1-Score είναι μια μετρική που συνδυάζει την Ακρίβεια (Precision) και την Ανάκληση (Recall) σε έναν ενιαίο αριθμό, δίνοντας μια πιο ισορροπημένη εικόνα της απόδοσης ενός μοντέλου ταξινόμησης.

Ορίζεται ως:

$$F1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Ανάκληση}}{\text{Precision} + \text{Ανάκληση}}$$

3.6. Κατανόηση λειτουργίας μοντέλων μηχανικής μάθησης μέσω της μεθόδου SHAP

Το SHAP (SHapley Additive exPlanations) είναι μια τεχνική εμπνευσμένη από τη θεωρία παιγνίων, η οποία χρησιμοποιείται για την ανάλυση και ερμηνεία των προβλέψεων μοντέλων μηχανικής μάθησης. Ο βασικός του στόχος είναι να ενισχύσει τη διαφάνεια αυτών των μοντέλων, επιτρέποντας στους επιστήμονες δεδομένων και σε όσους τα χρησιμοποιούν να κατανοήσουν τους λόγους πίσω από κάθε πρόβλεψη. Η τιμή Shapley αντιπροσωπεύει τη μέση συμβολή μιας μεταβλητής στο τελικό αποτέλεσμα, υπολογίζοντας την επίδρασή της σε όλους τους δυνατούς συνδυασμούς μεταβλητών.

4. ΣΥΛΛΟΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΣΤΟΙΧΕΙΩΝ

4.1. Εισαγωγή

Όπως αναφέρθηκε και στο Κεφάλαιο 1, το i-DREAMS είχε ως στόχο την ανάπτυξη ενός πλαισίου για τον ορισμό, τη δημιουργία, τη δοκιμή και την επικύρωση μιας «**Ζώνης Ανοχής Ασφαλείας**» (**Safety Tolerance Zone - STZ**), η οποία περιλαμβάνει διαφορετικά επίπεδα ασφάλειας. Μέσω ενός έξυπνου συστήματος παρακολούθησης, το οποίο λάμβανε υπόψη το ιστορικό του οδηγού και σχετικούς δείκτες κινδύνου σε πραγματικό χρόνο, αξιολογούνταν η οδηγική του απόδοση, η πολυπλοκότητα των εκάστοτε συνθηκών και τα καθήκοντά του. Με βάση αυτά τα δεδομένα, προσδιοριζόταν το επίπεδο ασφαλούς οδήγησης και εφαρμόζονταν παρεμβάσεις που βοηθούσαν τον οδηγό να διατηρείται εντός ασφαλών ορίων.

Η ανάλυση των δεδομένων της έρευνας αποσκοπεί:

1. Στον εντοπισμό του επιπέδου της «**Ζώνης Ανοχής Ασφαλείας**» στο οποίο βρισκόταν ο οδηγός σε πραγματικό χρόνο, ώστε να ενεργοποιούνται κατάλληλες παρεμβάσεις.
2. Στη διερεύνηση της σχέσης μεταξύ του κινδύνου και των παραγόντων που τον επηρεάζουν άμεσα. Μέσα από αυτή τη διαδικασία, επιδιώχθηκε η βαθύτερη κατανόηση της οδηγικής συμπεριφοράς, με στόχο τη βελτίωση των παρεμβάσεων για μεγαλύτερη οδική ασφάλεια.

4.2. Πείραμα σε πραγματικές συνθήκες οδήγησης

4.2.1. Στόχος πειράματος

Στο πλαίσιο του ερευνητικού προγράμματος i-DREAMS, πραγματοποιήθηκε ένα νατουραλιστικό πείραμα οδήγησης με την συμμετοχή 96 οδηγών από το Βέλγιο και το Ηνωμένο Βασίλειο, και δημιουργήθηκε μία βάση δεδομένων 15.389 διαδρομών και 265.512 λεπτών (καταγεγραμμένων δεδομένων οδήγησης). Οι δοκιμές σε πραγματικές συνθήκες οδήγησης εστιάστηκαν στην παρακολούθηση της οδηγικής συμπεριφοράς και στην επίδραση που ασκούν οι παρεμβάσεις σε πραγματικό χρόνο (δηλαδή προειδοποιήσεις εντός του οχήματος) καθώς και οι παρεμβάσεις μετά τη διαδρομή (δηλαδή ανατροφοδότηση μετά τη διαδρομή και στοιχεία παιχνιδιοποίησης) στην οδηγική συμπεριφορά..

Ο κύριος στόχος του πειράματος ήταν η συλλογή δεδομένων σχετικά με την οδηγική συμπεριφορά και το οδικό περιβάλλον, ώστε να ακολουθήσει η ανάλυσή τους για την επίτευξη των ερευνητικών στόχων.

Όπως φαίνεται στον πίνακα 4.1, και όπως προαναφέρθηκε στο Κεφάλαιο 1, η δοκιμαστική φάση του πειράματος πραγματοποιήθηκε σε τέσσερις φάσεις.

Πίνακας 4.1: Φάσεις πραγμάτωσης του πειράματος οδήγησης στις δύο χώρες

Φάση	Χώρα	Αριθμός οδηγών	Αριθμός διαδρομών	Συνολικός χρόνος (λεπτά)
Φάση 1: Παρακολούθηση	Βέλγιο	38	633	16.393
	Ηνωμένο Βασίλειο	53	3.073	56.853
Φάση 2: Παρεμβάσεις σε πραγματικό χρόνο	Βέλγιο	42	813	21.412
	Ηνωμένο Βασίλειο	54	3.317	58.458
Φάση 3: Παρεμβάσεις σε πραγματικό χρόνο & μετά τη διαδρομή	Βέλγιο	50	990	27.691
	Ηνωμένο Βασίλειο	53	3.417	59.556
Φάση 4: Παρεμβάσεις σε πραγματικό χρόνο, μετά τη διαδρομή & στοιχειοθέτηση παιχνιδιού	Βέλγιο	49	1.222	35.284
	Ηνωμένο Βασίλειο	54	4.594	93.974

Πηγή: i-DREAMS

Καθώς το βασικό αποτέλεσμα του έργου i-DREAMS είναι ένα ολοκληρωμένο σύνολο εργαλείων παρακολούθησης και επικοινωνίας για παρεμβάσεις και υποστήριξη, χρησιμοποιήθηκαν προηγμένες τεχνολογίες και συστήματα προκειμένου να παρακολουθούνται οι δείκτες απόδοσης της οδήγησης. Στη συνέχεια θα αναλυθούν μερικά εργαλεία που συνέβαλαν στην καταγραφή του τρόπου οδήγησης στο έργο i-Dreams.

4.2.2. On-Board Diagnostics (OBD II)

Μια συσκευή OBD-II που υποστηρίζει όλα τα πρωτόκολλα OBD-II ήταν εγκατεστημένη σε κάθε όχημα. Ένα σύγχρονο όχημα υποστηρίζει εκατοντάδες παραμέτρους, οι οποίες καταγράφονται από τη συσκευή OBD-II, η οποία διαθέτει το κατάλληλο Software Development Kit (SDK) για την εξαγωγή των απαραίτητων δεδομένων, καθώς και ένα σύνολο διεπαφών προγραμματισμού εφαρμογών (APIs) για επικοινωνία με συστήματα τρίτων. Αυτή η συσκευή OBD-II ενσωματώνει τεχνολογία GSM/GPRS 2G ή 3G, μέσω της οποίας όλα τα δεδομένα που καταγράφονται από τους αισθητήρες του οχήματος μεταδίδονται

4.2.3. Σύστημα Πρόληψης Ατυχημάτων MobilEye

Κατά τη διάρκεια του πειράματος i-Dreams χρησιμοποιήθηκαν δεδομένα από το σύστημα Mobileye (Mobileye, 2022), μια κάμερα ταμπλό και την πύλη Cardio (CardioID Technologies, 2022), η οποία καταγράφει τη συμπεριφορά οδήγησης (π.χ., ταχύτητα, επιτάχυνση, επιβράδυνση, τιμόνι) μαζί με σήματα GNSS.

[illegible]

Πηγή: OhioLINK (2022)

4.2.4. Στοιχεία που συλλέχθηκαν από το πείραμα i-Dreams

Η θεμελιώδης πρόκληση στο πλαίσιο του έργου i-DREAMS είναι ο τρόπος με τον οποίο οι επεξηγηματικές μεταβλητές (δηλαδή διάφορες μετρικές απόδοσης και δείκτες πολυπλοκότητας του έργου και ικανότητας αντιμετώπισης) συσχετίζονται με την εξαρτημένη μεταβλητή «κίνδυνος», προκειμένου να προβλεφθεί η STZ.

Υπάρχουν τρεις βασικές κατηγορίες μεταβλητών που χρησιμοποιούνται στο i-DREAMS:

- **Διακριτές μεταβλητές:** Μεταβλητές που είναι κατηγορικές (τακτικές ή ονομαστικές) και μπορούν να λάβουν μόνο διακριτές τιμές από το σύνολο των πραγματικών αριθμών. Μερικά παραδείγματα διακριτών μεταβλητών στο i-DREAMS είναι η κόπωση (ναι, όχι), η ώρα της ημέρας (οδήγηση κατά τη διάρκεια της ημέρας ή της νύχτας) και η STZ (κανονική φάση, φάση κινδύνου, φάση αποτρέψιμου ατυχήματος).
- **Συνεχείς μεταβλητές:** Μεταβλητές που μπορούν να λάβουν οποιαδήποτε τιμή από το σύνολο των πραγματικών αριθμών. Ορισμένα παραδείγματα συνεχών μεταβλητών στο i-DREAMS είναι η υπέρβαση ταχύτητας, η απόσταση ασφαλείας και σύνθετες μεταβλητές, όπως μεταβλητές σταθμισμένου αθροίσματος ή σταθμισμένου μέσου όρου.
- **Λανθάνουσες μεταβλητές:** Μεταβλητές που δεν είναι άμεσα παρατηρήσιμες από τον αναλυτή και επομένως δεν είναι γνωστό αν είναι συνεχείς ή διακριτές. Παραδείγματα λανθανουσών μεταβλητών στο i-DREAMS είναι η πολυπλοκότητα του έργου και η ικανότητα αντιμετώπισης, οι οποίες αποτελούν λανθάνουσες επεξηγηματικές μεταβλητές και απαιτούν παρατηρήσιμους δείκτες για τη μέτρησή τους. Ο κίνδυνος αρχικά θεωρείται επίσης στο i-DREAMS ως λανθάνουσα μεταβλητή.

Θα αξιολογηθούν οι επεξηγηματικές μεταβλητές του κινδύνου και οι πιο αξιόπιστοι δείκτες ικανότητας αντιμετώπισης, όπως η μέση ταχύτητα, η απόσταση ασφαλείας, οι παράνομες προσπεράσεις, οι απότομες επιταχύνσεις, τα απότομα φρεναρίσματα, η απόσταση που διανύθηκε, η διάρκεια, οι προειδοποιήσεις μετωπικής σύγκρουσης και οι προειδοποιήσεις σύγκρουσης με πεζούς.

Συγκεκριμένα, οι κύριοι παράγοντες κινδύνου που θα διερευνηθούν στο πλαίσιο του έργου i-DREAMS είναι:

- Υπέρβαση ορίου ταχύτητας
- Απόσταση ασφαλείας
- Προσπέραση
- Κόπωση

- Απότομες επιταχύνσεις
- Απότομα φρεναρίσματα
- Συμβάντα ελέγχου οχήματος (συνδυασμός απότομων επιταχύνσεων, φρεναρισμάτων και στροφών)

Στον πίνακα 4.2 παρουσιάζεται ένας κατάλογος των μεταβλητών που θα εξεταστούν και η περιγραφή τους.

Πίνακας 4.2: Επισκόπηση μεταβλητών που χρησιμοποιούνται στο πείραμα

Μεταβλητή	Περιγραφή	Μονάδες μέτρησης	Τύπος
ME_Car_speed	Ταχύτητα οχήματος	χιλιόμετρα ανά ώρα	αριθμητική
ME_AWS_hw_measurement	Μέτρηση της μεταβλητής headway	Δευτερόλεπτα	αριθμητική
ME_AWS_fcw	Προειδοποίηση μπροστινής σύγκρουσης	-	διακριτή
ME_AWS_pcw	Προειδοποίηση σύγκρουσης πεζών	-	διακριτή
ME_AWS_ldw	Προειδοποίηση αναχώρησης από τη λωρίδα	-	διακριτή
ME_AWS_pedestrian_dz	Πεζός σε επικίνδυνη ζώνη	-	διακριτή
GPS_distances	Συνολική απόσταση του ταξιδιού	χιλιόμετρα	αριθμητική
GPS_spd	Ταχύτητα	χιλιόμετρα ανά ώρα	αριθμητική
ME_TSR_tsr_1_speed	Εμφάνιση κωδικού πινακίδας 1 ταχύτητας	-	διακριτή
ME_AWS_time_indicator	Υποδεικνύει συνθήκες φωτισμού	-	διακριτή
DEM_evt_ha_lvl_M	Απότομα γεγονότα επιτάχυνσης μεσαίου επιπέδου	-	διακριτή
DEM_evt_hb_lvl_M	Απότομο φρενάρισμα μεσαίου επιπέδου	-	διακριτή
DEM_evt_hc_lvl_M	Απότομες στροφές μεσαίου επιπέδου	-	διακριτή
DEM_evt_ha_lvl_L	Απότομα γεγονότα επιτάχυνσης χαμηλού επιπέδου	-	διακριτή
DEM_evt_hb_lvl_L	Απότομο φρενάρισμα χαμηλού επιπέδου	-	διακριτή
DEM_evt_hc_lvl_L	Απότομες στροφές χαμηλού επιπέδου	-	διακριτή
DEM_evt_ha_lvl_H	Απότομα γεγονότα επιτάχυνσης υψηλού επιπέδου	-	διακριτή

Μεταβλητή	Περιγραφή	Μονάδες μέτρησης	Τύπος
DEM_evt_hb_lvl_H	Απότομο φρενάρισμα υψηλού επιπέδου	-	διακριτή
DEM_evt_hc_lvl_H	Απότομες στροφές υψηλού επιπέδου	-	διακριτή
ME_AWS_time_indicator	Υποδεικνύει συνθήκες φωτισμού (ημέρα, σούρουπο, νύχτα)	-	διακριτή
ME_Car_wipers	Υποδεικνύει καιρικές συνθήκες (υαλοκαθαριστήρες on/off)	-	διακριτή
ME_Car_high_beam	Μεγάλη σκάλα (φωτών)	-	διακριτή
IBI_value	Χρονικό διάστημα μεταξύ διαδοχικών καρδιακών παλμών	χιλιοστά του δευτερολέπτου	αριθμητική
ME_LDW_Map_type_L	Προειδοποίηση αναχώρησης από την αριστερή λωρίδα	-	διακριτή
ME_LDW_Map_type_R	Προειδοποίηση αναχώρησης από την δεξιά λωρίδα	-	διακριτή

Έπειτα, παρουσιάζεται η επεξεργασία των στοιχείων, η οποία είναι απαραίτητη για τα επόμενα βήματα της εργασίας.

4.3.Επεξεργασία στοιχείων

Τα δεδομένα που συλλέχθηκαν από τους οδηγούς αποθηκεύτηκαν σε δύο αρχεία .csv, όπου ο τίτλος κάθε αρχείου υποδεικνύει την περιοχή διεξαγωγής του πειράματος (Βέλγιο ή Ηνωμένο Βασίλειο).

Για να διευκολυνθεί η ανάλυση, τα δεδομένα ομαδοποιήθηκαν σε χρονικά διαστήματα των 30 δευτερολέπτων. Σε κάθε τέτοιο διάστημα, υπολογίστηκαν βασικά στατιστικά μεγέθη για κάθε μεταβλητή, όπως:

- Μέση τιμή
- Τυπική απόκλιση
- Ελάχιστη και μέγιστη τιμή
- Διάμεσος

Στον Πίνακα 4.3 παρουσιάζονται οι μεταβλητές που προέκυψαν από αυτή τη διαδικασία.

Πίνακας 4.3: Περιγραφή μεταβλητών μετά την επεξεργασία

Μεταβλητή	Περιγραφή	Μονάδες μέτρησης	Τύπος
Headway	Χρονική απόσταση από το προπορευόμενο όχημα	Δευτερόλεπτα	Αριθμητική
ME_Car_speed_mean	Ταχύτητα οχήματος	Χιλιόμετρα ανά ώρα	Αριθμητική
ME_AWS_hw_measurement_mean	Μέτρηση της μεταβλητής headway	Δευτερόλεπτα	Αριθμητική
ME_AWS_fcw_mean	Προειδοποίηση μπροστινής σύγκρουσης	-	Διακριτή
ME_AWS_pcw_mean	Προειδοποίηση σύγκρουσης πεζών	-	Διακριτή
ME_AWS_pedestrian_dz_mean	Πεζός σε επικίνδυνη ζώνη	-	Διακριτή
GPS_distances_sum	Συνολική απόσταση του ταξιδιού	Χιλιόμετρα	Αριθμητική
GPS_spd_mean	Ταχύτητα	Χιλιόμετρα ανά ώρα	Αριθμητική
ME_TSR_tsr_1_speed_median	Εμφάνιση κωδικού πινακίδας 1 ταχύτητας	-	Διακριτή
ME_AWS_time_indicator_median	Υποδεικνύει συνθήκες φωτισμού	-	Διακριτή
DEM_evt_ha_lvl_M_mean	Δύσκολα γεγονότα επιτάχυνσης μεσαίου επιπέδου	-	Διακριτή
DEM_evt_hb_lvl_M_mean	Δυνατό φρενάρισμα μεσαίου επιπέδου	-	Διακριτή
DEM_evt_hc_lvl_M_mean	Δύσκολες στροφές μεσαίου επιπέδου	-	Διακριτή
DEM_evt_ha_lvl_L_mean	Δύσκολα γεγονότα επιτάχυνσης χαμηλού επιπέδου	-	Διακριτή

Μεταβλητή	Περιγραφή	Μονάδες μέτρησης	Τύπος
DEM_evt_hb_lvl_L_mean	Δυνατό φρενάρισμα χαμηλού επιπέδου	-	Διακριτή
DEM_evt_hc_lvl_L_mean	Δύσκολες στροφές χαμηλού επιπέδου	-	Διακριτή
DEM_evt_ha_lvl_H_mean	Δύσκολα γεγονότα επιτάχυνσης υψηλού επιπέδου	-	Διακριτή
DEM_evt_hb_lvl_H_mean	Δυνατό φρενάρισμα υψηλού επιπέδου	-	Διακριτή
DEM_evt_hc_lvl_H_mean	Δύσκολες στροφές υψηλού επιπέδου	-	Διακριτή
DEM_evt_ha_lvl_M_sum	Δύσκολα γεγονότα επιτάχυνσης μεσαίου επιπέδου	-	Διακριτή
DEM_evt_hb_lvl_M_sum	Δυνατό φρενάρισμα μεσαίου επιπέδου	-	Διακριτή
DEM_evt_hc_lvl_M_sum	Δύσκολες στροφές μεσαίου επιπέδου	-	Διακριτή
DEM_evt_ha_lvl_L_sum	Δύσκολα γεγονότα επιτάχυνσης χαμηλού επιπέδου	-	Διακριτή
DEM_evt_hb_lvl_L_sum	Δυνατό φρενάρισμα χαμηλού επιπέδου	-	Διακριτή
DEM_evt_hc_lvl_L_sum	Δύσκολες στροφές χαμηλού επιπέδου	-	Διακριτή
DEM_evt_ha_lvl_H_sum	Δύσκολα γεγονότα επιτάχυνσης υψηλού επιπέδου	-	Διακριτή
DEM_evt_hb_lvl_H_sum	Δυνατό φρενάρισμα υψηλού επιπέδου	-	Διακριτή

Μεταβλητή	Περιγραφή	Μονάδες μέτρησης	Τύπος
DEM_evt_hc_lvl_H_sum	Δύσκολες στροφές υψηλού επιπέδου	-	Διακριτή
ME_AWS_time_indicator_median	Υποδεικνύει συνθήκες φωτισμού (ημέρα, σούρουπο, νύχτα)	-	Διακριτή
ME_Car_wipers_median	Υποδεικνύει καιρικές συνθήκες (υαλοκαθαριστήρες on/off)	-	Διακριτή
ME_Car_high_beam_median	Μεγάλη σκάλα	-	Διακριτή
IBI_value_mean	Χρονικό διάστημα μεταξύ διαδοχικών καρδιακών παλμών	Χιλιοστά του δευτερολέπτου	Αριθμητική
ME_LDW_Map_type_L_median	Προειδοποίηση αναχώρησης από την αριστερή λωρίδα	-	Διακριτή
ME_LDW_Map_type_R_median	Προειδοποίηση αναχώρησης από τη δεξιά λωρίδα	-	Διακριτή

Στη συνέχεια, αξιολογήθηκαν οι μεταβλητές που σχετίζονται με τον κίνδυνο, καθώς και οι πιο αξιόπιστοι δείκτες που αποτυπώνουν την πολυπλοκότητα της οδήγησης και την ικανότητα των οδηγών να την αντιμετωπίσουν. Ιδιαίτερη έμφαση δόθηκε στον χρόνο διαδρομής, την απόσταση που καλύφθηκε, τις προειδοποιήσεις πρόσθιας σύγκρουσης και τις καιρικές συνθήκες. Επιπλέον, βασικό δείκτη ανάλυσης αποτέλεσε η χρονική απόσταση από το προπορευόμενο όχημα.

Για να ενσωματωθούν τα επίπεδα STZ στη μελέτη, δημιουργήθηκε μία νέα μεταβλητή:

- **STZ_headway**: Αντιπροσωπεύει το επίπεδο STZ για τη χρονική απόσταση από το προπορευόμενο όχημα.

Η εξαρτημένη αυτή μεταβλητή κατηγοριοποιήθηκε σε τρία επίπεδα:

- **0**: Κανονική οδήγηση (Normal)
- **1**: Επικίνδυνη κατάσταση (Dangerous)
- **2**: Φάση αποφυγής ατυχήματος (Avoidable Accident)

Συγκεκριμένα, το επίπεδο 'Κανονικό' πρέπει να είναι η κύρια κατηγορία με το υψηλότερο ποσοστό δειγμάτων, ενώ τα επίπεδα 'Επικίνδυνο' και 'Αποφυγή Ατυχήματος' πρέπει να είναι η μειοψηφική κατηγορία με το χαμηλότερο ποσοστό δειγμάτων. Αναλυτικότερα, από το πείραμα συλλέξαμε τις μεταβλητές iDreams_Headway_Map_level_i ,όπου i=-1,0,1,2,3. Οι παραπάνω μεταβλητές αντιστοιχούν στα επίπεδα παρέμβασης της χρονικής απόστασης από το προπορευόμενο όχημα σε πραγματικό χρόνο. Κάθε επίπεδο παρέμβασης αντιστοιχεί σε ένα από τα τρία επίπεδα της 'Ζώνης Ανοχής Ασφαλείας'. Οι πιθανές τιμές αυτών των μεταβλητών είναι 0 (όταν το επίπεδο παρέμβασης διάφορο του i) και 1 (όταν το επίπεδο παρέμβασης ίσο με i).

Έτσι, για κάθε 30 δευτερόλεπτα προκύπτει το επίπεδο της Ζώνης Ανοχής Ασφαλείας το οποίο χαρακτηρίζεται ως:

- Κανονικό (Normal) όταν η μεταβλητή που επιλέγεται είναι επιπέδου παρέμβασης -1,0,1
- Επικίνδυνο (Dangerous) όταν η μεταβλητή που επιλέγεται είναι επιπέδου παρέμβασης 2
- Αποφυγή Ατυχήματος (Avoidable Accident) όταν η μεταβλητή που επιλέγεται είναι επιπέδου παρέμβασης 3

4.4. Περιγραφική Στατιστική Μεταβλητών

Χρησιμοποιώντας τη βιβλιοθήκη ανάλυσης δεδομένων pandas στο προγραμματιστικό περιβάλλον Python, πραγματοποιήθηκε περιγραφική στατιστική ανάλυση των δεδομένων μετά την επεξεργασία τους. Στους πίνακες 4.4 και 4.5 παρουσιάζονται βασικά περιγραφικά στατιστικά των μεταβλητών που συλλέχθηκαν, όπως η μέση τιμή, η τυπική απόκλιση, οι ελάχιστες και μέγιστες τιμές, καθώς και η διάμεσος για το Βέλγιο και το Ηνωμένο Βασίλειο, αντίστοιχα.

Πίνακας 4.4: Περιγραφική στατιστική ανάλυση δεδομένων Βελγίου

Βέλγιο	Μέση Τιμή	Τυπική Απόκλιση	Ελάχιστη Τιμή	Μέγιστη Τιμή	Διάμεσος
DrivingEvents_Map_evt_ha_mean	0.49	0.43	0.00	1.00	0.47
DrivingEvents_Map_evt_hb_mean	0.21	0.36	0.00	1.00	0.00
DrivingEvents_Map_evt_hc_mean	0.27	0.41	0.00	1.00	0.00
DrivingEvents_Map_lvl_H_mean	0.05	0.20	0.00	1.00	0.00
DrivingEvents_Map_lvl_L_mean	0.69	0.42	0.00	1.00	1.00
DrivingEvents_Map_lvl_M_mean	0.26	0.40	0.00	1.00	0.00
Drowsiness_level_median	35.00	0.07	35.00	39.00	35.00
IBI_value_mean	813.85	58.48	375.98	2191.41	812.30
ME_Car_speed_mean	54.31	34.91	0.00	172.43	49.63
ME_Car_wipers_median	0.05	0.22	0.00	1.00	0.00
ME_Car_high_beam_median	0.02	0.15	0.00	1.00	0.00
ME_AWS_tsr_level_mean	0.52	1.26	0.00	7.00	0.00

ME_AWS_fcw_mean	0.00	0.00	0.00	0.13	0.00
ME_AWS_pcw_mean	0.00	0.00	0.00	0.13	0.00
ME_AWS_pedestrian_dz_mean	0.00	0.03	0.00	1.00	0.00
ME_AWS_time_indicator_median	1.30	0.69	1.00	3.00	1.00
ME_TSR_tsr_1_speed_median	121.19	118.59	0.00	254.00	64.00
GPS_spd_mean	50.90	32.32	0.00	158.74	47.17
GPS_distances_sum	431.20	377.00	0.00	42712.52	395.99
DEM_evt_ha_lvl_L_mean	0.31	0.43	0.00	1.00	0.00
DEM_evt_ha_lvl_L_sum	9.35	12.82	0.00	30.00	0.00
DEM_evt_ha_lvl_M_mean	0.10	0.27	0.00	1.00	0.00
DEM_evt_ha_lvl_M_sum	3.01	8.17	0.00	30.00	0.00
DEM_evt_ha_lvl_H_mean	0.04	0.17	0.00	1.00	0.00
DEM_evt_ha_lvl_H_sum	1.09	5.10	0.00	30.00	0.00
DEM_evt_hb_lvl_L_mean	0.19	0.35	0.00	1.00	0.00
DEM_evt_hb_lvl_L_sum	5.65	10.46	0.00	30.00	0.00
DEM_evt_hb_lvl_M_mean	0.02	0.12	0.00	1.00	0.00
DEM_evt_hb_lvl_M_sum	0.58	3.69	0.00	30.00	0.00
DEM_evt_hb_lvl_H_mean	0.00	0.04	0.00	1.00	0.00
DEM_evt_hb_lvl_H_sum	0.06	1.18	0.00	30.00	0.00
DEM_evt_hc_lvl_L_mean	0.12	0.30	0.00	1.00	0.00
DEM_evt_hc_lvl_L_sum	3.61	8.97	0.00	30.00	0.00
DEM_evt_hc_lvl_M_mean	0.14	0.32	0.00	1.00	0.00
DEM_evt_hc_lvl_M_sum	4.20	9.55	0.00	30.00	0.00
DEM_evt_hc_lvl_H_mean	0.01	0.10	0.00	1.00	0.00
DEM_evt_hc_lvl_H_sum	0.41	3.12	0.00	30.00	0.00
ME_LDW_Map_type_L_mean	0.94	0.24	0.00	1.00	1.00
ME_LDW_Map_type_R_mean	0.95	0.23	0.00	1.00	1.00

Πίνακας 4.5: Περιγραφική Στατιστική δεδομένων Ηνωμένου Βασιλείου

Ηνωμένο Βασίλειο	Μέση Τιμή	Τυπική Απόκλιση	Ελάχιστη Τιμή	Μέγιστη Τιμή	Διάμεσος
DrivingEvents_Map_evt_ha_mean	0.48	0.41	0.00	1.00	0.47
DrivingEvents_Map_evt_hb_mean	0.17	0.33	0.00	1.00	0.00
DrivingEvents_Map_evt_hc_mean	0.29	0.42	0.00	1.00	0.00
DrivingEvents_Map_lvl_H_mean	0.03	0.14	0.00	1.00	0.00
DrivingEvents_Map_lvl_L_mean	0.74	0.40	0.00	1.00	1.00
DrivingEvents_Map_lvl_M_mean	0.23	0.39	0.00	1.00	0.00
Drowsiness_level_median	35.00	0.00	35.00	35.00	35.00
IBI_value_mean	802.68	37.70	306.64	1580.08	803.22
ME_Car_speed_mean	44.56	34.19	0.00	173.07	37.80
ME_Car_wipers_median	0.07	0.26	0.00	1.00	0.00
ME_Car_high_beam_median	0.01	0.08	0.00	1.00	0.00
ME_AWS_tsr_level_mean	0.10	0.37	0.00	7.00	0.00
ME_AWS_fcw_mean	0.00	0.00	0.00	0.13	0.00
ME_AWS_pcw_mean	0.00	0.00	0.00	0.07	0.00

ME_AWS_pedestrian_dz_mean	0.00	0.03	0.00	1.00	0.00
ME_AWS_time_indicator_median	1.79	0.95	1.00	3.00	1.00
ME_TSR_tsr_1_speed_median	108.36	110.76	0.00	254.00	64.00
GPS_spd_mean	40.92	32.42	0.00	165.49	34.98
GPS_distances_sum	342.52	270.18	0.00	3017.56	292.89
DEM_evt_ha_lvl_L_mean	0.32	0.42	0.00	1.00	0.03
DEM_evt_ha_lvl_L_sum	9.64	12.63	0.00	30.00	1.00
DEM_evt_ha_lvl_M_mean	0.08	0.24	0.00	1.00	0.00
DEM_evt_ha_lvl_M_sum	2.38	7.31	0.00	30.00	0.00
DEM_evt_ha_lvl_H_mean	0.01	0.11	0.00	1.00	0.00
DEM_evt_ha_lvl_H_sum	0.45	3.21	0.00	30.00	0.00
DEM_evt_hb_lvl_L_mean	0.16	0.32	0.00	1.00	0.00
DEM_evt_hb_lvl_L_sum	4.77	9.48	0.00	30.00	0.00
DEM_evt_hb_lvl_M_mean	0.01	0.09	0.00	1.00	0.00
DEM_evt_hb_lvl_M_sum	0.34	2.70	0.00	30.00	0.00
DEM_evt_hb_lvl_H_mean	0.00	0.03	0.00	1.00	0.00
DEM_evt_hb_lvl_H_sum	0.03	0.79	0.00	30.00	0.00
DEM_evt_hc_lvl_L_mean	0.14	0.32	0.00	1.00	0.00
DEM_evt_hc_lvl_L_sum	4.16	9.56	0.00	30.00	0.00
DEM_evt_hc_lvl_M_mean	0.14	0.32	0.00	1.00	0.00
DEM_evt_hc_lvl_M_sum	4.31	9.67	0.00	30.00	0.00
DEM_evt_hc_lvl_H_mean	0.01	0.09	0.00	1.00	0.00
DEM_evt_hc_lvl_H_sum	0.31	2.69	0.00	30.00	0.00
ME_LDW_Map_type_L_mean	0.95	0.22	0.00	1.00	1.00
ME_LDW_Map_type_R_mean	0.96	0.19	0.00	1.00	1.00

4.5. Σύνοψη

Στο συγκεκριμένο κεφάλαιο, παρατέθηκαν τα συλλεχθέντα στοιχεία από τη βάση δεδομένων του πειράματος i-Dreams στις δύο χώρες που εξετάζονται, το Βέλγιο και το Ηνωμένο Βασίλειο. Στη συνέχεια, παρουσιάστηκαν μερικά από τα βασικά εργαλεία που συνέβαλαν στην καταγραφή του τρόπου οδήγησης στο έργο i-Dreams. Έπειτα, καταγράφηκαν οι μεταβλητές που μετρήθηκαν κατά τη διάρκεια του πειράματος i-Dreams, με τις μεθόδους που αναφέρθηκαν, και τέθηκαν προς επεξεργασία. Τέλος, μέσω της γλώσσας προγραμματισμού Python έγινε περιγραφική στατιστική των δεδομένων για τις δύο χώρες.

Πλέον, τα δεδομένα είναι έτοιμα για περαιτέρω ανάλυση, η οποία πραγματοποιείται στο Κεφάλαιο 5.

5. ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ – ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται αναλυτικά η μεθοδολογία που εφαρμόστηκε καθώς και τα αποτελέσματα που προέκυψαν στο πλαίσιο της μελέτης. Ο στόχος και η κατάλληλη μεθοδολογία για την επίτευξη του προσδιορίστηκαν στη βιβλιογραφική ανασκόπηση. Σύμφωνα με τις μεθοδολογίες παλαιότερων ερευνών, που παρουσιάστηκαν στην βιβλιογραφική ανασκόπηση, για την διερεύνηση της επιρροής των διαφορετικών παραγόντων της οδηγικής συμπεριφοράς θα αναπτυχθούν κατάλληλοι αλγόριθμοι μηχανικής και βαθιάς εκμάθησης για την ταξινόμηση. Αναλυτικότερα θα αξιολογηθεί η σημαντικότητα των μεταβλητών στην ταξινόμηση και η ερμηνευτική τους ικανότητα στην παλινδρόμηση.

Η επικίνδυνη οδήγηση θα εξεταστεί βάσει δύο προσεγγίσεων επομένως, η ανάλυση θα χωριστεί σε δύο μέρη. Στο πρώτο μέρος των αναλύσεων θα αναπτυχθούν τα μοντέλα ταξινόμησης για τον καθορισμό του επιπέδου της 'Ζώνης Ανοχής Ασφαλείας' που βρίσκεται οδηγός για κάθε χρονικό πλαίσιο των 30 δευτερολέπτων. Τα δεδομένα που συλλέχθηκαν από την συμμετοχή 96 οδηγών στο Βέλγιο και την Αγγλία, αποτελούν τις ενδογενείς μεταβλητές, ενώ το επίπεδο της 'Ζώνης Ανοχής Ασφαλείας' αποτελεί την εξωγενή μεταβλητή. Στο δεύτερο μέρος των αναλύσεων θα αναπτυχθεί η μέθοδος SHAP με σκοπό να εξεταστούν και αξιολογηθούν τα χαρακτηριστικά που επιδρούν στην διάρκεια οδήγηση σε κάθε ένα από τα τρία επίπεδα ασφαλείας για κάθε οδηγό. Με την χρήση μετρικών αξιολόγησης θα πραγματοποιηθεί η αξιολόγηση της προγνωστικής ικανότητας των μοντέλων. Με την χρήση της προγραμματιστικής γλώσσας Python θα πραγματοποιηθεί η ανάλυση. Θα αξιοποιηθούν οι παρακάτω ειδικές βιβλιοθήκες και εργαλεία:

- Υπολογισμοί: NumPy
- Ανάλυση και χειρισμός δεδομένων: Pandas
- Χειρισμός ανομοιογένειας δεδομένων: Imbalanced Learn
- Γραφική απεικόνιση: Matplotlib, Seaborn 47
- Μηχανική εκμάθηση: Scikit-Learn
- Βαθιά εκμάθηση: Tensorflow
- Ερμηνεία Μοντέλων: SHAPley values

5.2 Εντοπισμός του Επιπέδου 'Ζώνης Ανοχής Ασφαλείας'

Για τον εντοπισμό της 'Ζώνης Ανοχής Ασφαλείας' εστιάζουμε στον υπολογισμό της επιρροής του κάθε παράγοντα κινδύνου στην αναγνώριση της επικίνδυνης οδηγικής συμπεριφοράς του οδηγού. Η μεθοδολογία που θα ακολουθηθεί περιλαμβάνει την ανάλυση των διαφορετικών παραγόντων κινδύνου με βάση την ανάπτυξη τριών συνδυασμένων αλγόριθμων ταξινόμησης. Τέλος, με βάση

την συνολική επίδοση των μοντέλων ταξινόμησης θα αξιολογήσουμε τους κρίσιμους παράγοντες.

5.2.1 Καθορισμός Επιπέδων Ασφαλείας

Αρχικά, πριν την ανάπτυξη των αλγορίθμων ταξινόμησης και την διερεύνηση της επιρροής των μεταβλητών στην επικίνδυνη οδήγηση, ήταν αναγκαία η κατηγοριοποίηση των δεδομένων οδήγησης σε ένα από **τα τρία επίπεδα της 'Ζώνης Ανοχής Ασφαλείας'**. Σύμφωνα με την εφαρμογή i-DREAMS, καταγράφηκαν ορισμένα επίπεδα παρέμβασης σε πραγματικό χρόνο, τα οποία θεσπίστηκαν στα πλαίσια του έργου με βάση ορισμένες αρχές και όρια από την διεθνή βιβλιογραφία. **Η αντιστοίχιση των επιπέδων** βασίστηκε στην χρονική απόσταση από το προπορευόμενο όχημα (**Headway**) για τις δύο χώρες.

Προκειμένου τα αποτελέσματα να εναρμονίζονται με τα αποτελέσματα της διεθνούς βιβλιογραφίας είναι θεμιτό τα δείγματα της επικίνδυνης οδήγησης να αποτελούν την κλάση μειοψηφίας. Συγκεκριμένα, το επίπεδο 'Κανονικό' πρέπει να είναι η κύρια κατηγορία με το υψηλότερο ποσοστό δειγμάτων, ενώ τα επίπεδα 'Επικίνδυνο' και 'Αποφυγή Ατυχήματος' πρέπει να είναι η μειοψηφική κατηγορία με το χαμηλότερο ποσοστό δειγμάτων. Αναλυτικότερα, από το πείραμα συλλέξαμε τις μεταβλητές iDreams_Headway_Map_level_i, όπου $i = 1, 0, 1, 2, 3$. Οι παραπάνω μεταβλητές αντιστοιχούν στα επίπεδα παρέμβασης της χρονικής απόστασης από το προπορευόμενο όχημα σε πραγματικό χρόνο. Κάθε επίπεδο παρέμβασης αντιστοιχεί σε ένα από τα τρία επίπεδα της 'Ζώνης Ανοχής Ασφαλείας'. Οι πιθανές τιμές αυτών των μεταβλητών είναι 0 (όταν το επίπεδο παρέμβασης διάφορο του i) και 1 (όταν το επίπεδο παρέμβασης ίσο με i).

Πίνακας 5.1: Περιγραφή μεταβλητών i-Dreams

Μεταβλητές i-Dreams	Περιγραφή μεταβλητών
iDreams_Headway_Map_level__-1	Για επίπεδο παρέμβασης σε πραγματικό επίπεδο -1 δεν εντοπίζεται κανένα όχημα και το επίπεδο Ζώνης Ανοχής Ασφαλείας είναι Κανονικό.
iDreams_Headway_Map_level__0	Για επίπεδο παρέμβασης σε πραγματικό χρόνο 0 εντοπίζεται όχημα αλλά με Headway $\geq 2,5$ και το επίπεδο Ζώνης Ανοχής Ασφαλείας είναι Κανονικό.
iDreams_Headway_Map_level__1	Για επίπεδο παρέμβασης σε πραγματικό χρόνο 1 εντοπίζεται όχημα με Headway $< 2,5$ που υπερβαίνει τις οριακές τιμές και το επίπεδο Ζώνης Ανοχής Ασφαλείας είναι Κανονικό.

iDreams_Headway_Map_level__2	Για επίπεδο παρέμβασης σε πραγματικό χρόνο 2 έχουμε το πρώτο προειδοποιητικό στάδιο και επίπεδο Ζώνης Ανοχής Ασφαλείας Επικίνδυνο.
iDreams_Headway_Map_level__3	Για επίπεδο παρέμβασης σε πραγματικό χρόνο 3 έχουμε το δεύτερο προειδοποιητικό στάδιο και το επίπεδο Ζώνης Ανοχής Ασφαλείας Αποφυγής Ατυχήματος.

Για τον καθορισμό των επιπέδων της Ζώνης Ανοχής Ασφαλείας που καταγράφονται, ελέγχθηκαν αρχικά για κάθε 30 δευτερόλεπτα οι παραπάνω μεταβλητές σε κάθε επίπεδο παρέμβασης. Στην συνέχεια, επιλέχθηκε το δυσμενέστερο επίπεδο από άποψη ασφαλείας, δηλαδή επιλέχθηκε η μεταβλητή στην οποία εντοπίζεται μία τιμή και αντιστοιχεί στο δυσμενέστερο επίπεδο ασφαλείας. Επομένως, για κάθε 30 δευτερόλεπτα προκύπτει το επίπεδο της Ζώνης Ανοχής Ασφαλείας το οποίο χαρακτηρίζεται ως:

- Κανονικό (Normal) όταν η μεταβλητή που επιλέγεται είναι επιπέδου παρέμβασης -1,0,1
- Επικίνδυνο (Dangerous) όταν η μεταβλητή που επιλέγεται είναι επιπέδου παρέμβασης 2
- Αποφυγής Ατυχήματος (Avoidable Accident) όταν η μεταβλητή που επιλέγεται είναι επιπέδου παρέμβασης 3

Στον πίνακα 5.2 παρατίθενται τα αποτελέσματα της παραπάνω ανάλυσης και για τα δεδομένα που συλλέχθηκαν στην Αγγλία και το Βέλγιο.

Πίνακας 5.2: Επίπεδο Ζώνης Ανοχής Ασφαλείας στις δύο εξεταζόμενες χώρες

	Normal	Dangerous	Avoidable Accident	Σύνολο
Βέλγιο (Headway)	68462	17165	4787	90414
Αγγλία (Headway)	53942	13960	3365	71267

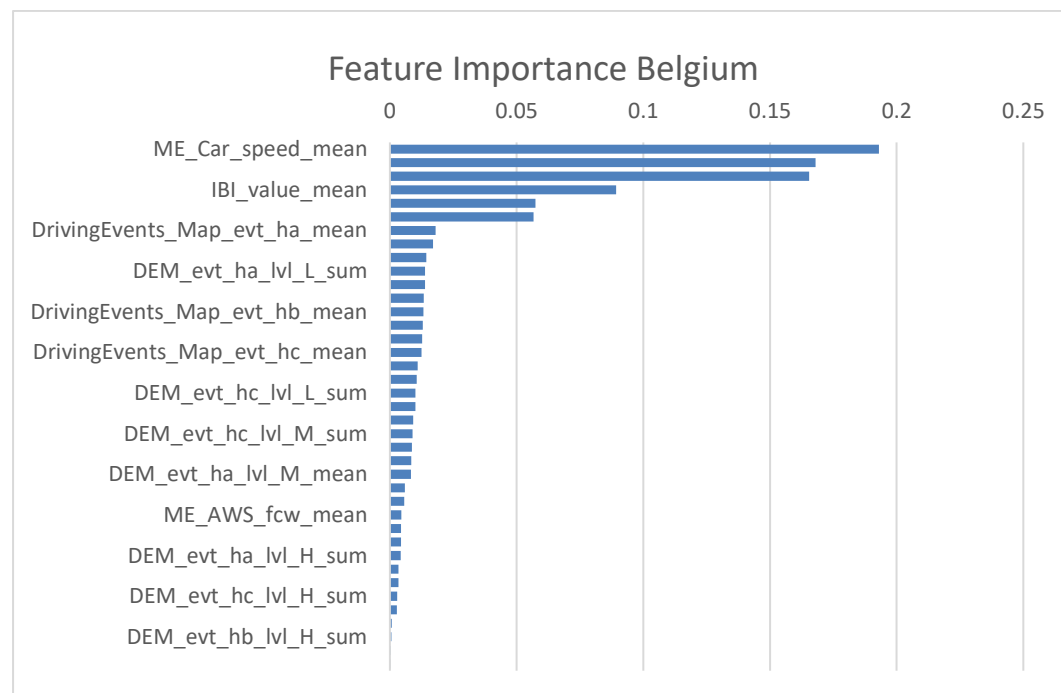
Ο καθορισμός του επιπέδου ασφαλείας θα γίνει με την χρήση της μεταβλητής Headway, δικαιολογημένα καθώς καταγράφονται σημαντικά λιγότερα επίπεδα “Dangerous” και “Avoidable Accident” σε σχέση με τα επίπεδα “Normal”.

5.2.2 Επιλογή Χαρακτηριστικών (Feature Selection)

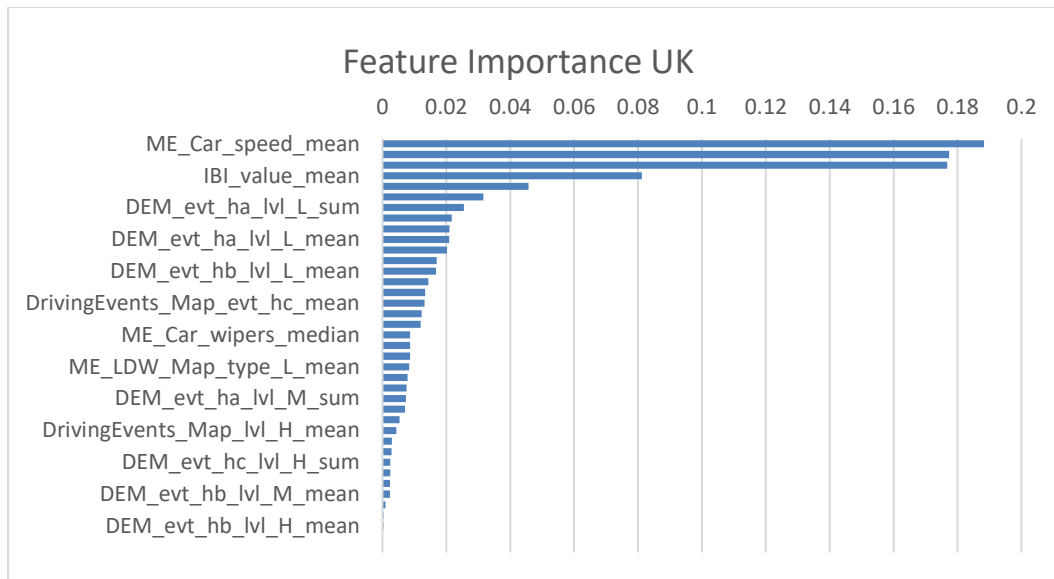
Βασικό κομμάτι της μεθοδολογίας αποτελεί η διαδικασία επιλογής χαρακτηριστικών. Στόχος της διαδικασίας είναι η μείωση του αριθμού των μεταβλητών εισόδου με ταυτόχρονη μείωση του υπολογιστικού κόστους του μοντέλου και βελτίωση της προγνωστικής του απόδοσης. Η επιλογή των χαρακτηριστικών γίνεται με γνώμονα την επιρροή της μεταβλητής στην

διαδικασία της ταξινόμησης. Η μέθοδος αυτή αποτελεί μία αρχική προσέγγιση για την μείωση των μεταβλητών εισόδου και τη βελτίωση των μοντέλων. Διερευνήθηκαν διάφορα σύνολα συνδυάζοντας διαφορετικές μεταβλητές βάσει της επιρροής τους στις προβλέψεις. Για τον εντοπισμό της σημαντικότητας των μεταβλητών στην ταξινόμηση χρησιμοποιήθηκε η τεχνική σημαντικότητας χαρακτηριστικών βάσει την μετάθεση χαρακτηριστικών (feature importance based on feature permutation).

Αρχικά τα δεδομένα χωρίστηκαν σε δύο υποσύνολα: Α. Το πρώτο υποσύνολο περιλαμβάνει όλες τις μεταβλητές εισόδου της μεθόδου δηλαδή τα δεδομένα που συλλέχθηκαν από το πείραμα που διεξήχθη σε πραγματικές συνθήκες οδήγησης. Β. Το δεύτερο υποσύνολο περιλαμβάνει την μεταβλητή εξόδου που αποτελείται από το επίπεδο της 'Ζώνης Ανοχής Ασφαλείας'. Έπειτα χρησιμοποιώντας τα ειδικά εργαλεία της βιβλιοθήκης scikit-learn, αναπτύχθηκε **ο ταξινομητής των 'Τυχαίων Δασών'** (Random Forests Classifier) και υπολογίστηκε η επιρροή της κάθε μεταβλητής (feature importance) με βάση την μετάθεση των χαρακτηριστικών (feature permutation) στη διαδικασία ταξινόμησης. Στα Γραφήματα 5.1 και 5.2 καθώς και στους Πίνακες 5.3 και 5.4 απεικονίζεται η επιρροή της κάθε μεταβλητής στην ταξινόμηση σε κλίμακα τιμών [0,1], για το Βέλγιο και το Ηνωμένο Βασίλειο αντίστοιχα.



Γράφημα 5.1: Σημαντικότητα μεταβλητών για το Βέλγιο



Γράφημα 5.2: Σημαντικότητα μεταβλητών για το Ηνωμένο Βασίλειο

Πίνακας 5.3: Σημαντικότητα μεταβλητών για το Βέλγιο

Feature	Importance
ME_Car_speed_mean	0.19301
GPS_spd_mean	0.168059
GPS_distances_sum	0.165476
IBI_value_mean	0.089276
ME_AWS_tsr_level_mean	0.057424
ME_TSR_tsr_1_speed_median	0.05668
DrivingEvents_Map_evt_ha_mean	0.01806
ME_AWS_time_indicator_median	0.016994
DrivingEvents_Map_lvl_L_mean	0.01443
DEM_evt_ha_lvl_L_sum	0.013916
DEM_evt_ha_lvl_L_mean	0.013914
DrivingEvents_Map_lvl_M_mean	0.013323
DrivingEvents_Map_evt_hb_mean	0.013196
DEM_evt_hb_lvl_L_mean	0.012943
DEM_evt_hb_lvl_L_sum	0.012758
DrivingEvents_Map_evt_hc_mean	0.012452
ME_LDW_Map_type_L_mean	0.010956
ME_LDW_Map_type_R_mean	0.010583
DEM_evt_hc_lvl_L_sum	0.010131
DEM_evt_hc_lvl_L_mean	0.010117
ME_Car_wipers_median	0.009152
DEM_evt_hc_lvl_M_sum	0.008874
DEM_evt_hc_lvl_M_mean	0.008736
DEM_evt_ha_lvl_M_sum	0.008415
DEM_evt_ha_lvl_M_mean	0.008275
DrivingEvents_Map_lvl_H_mean	0.005885
ME_AWS_pedestrian_dz_mean	0.005591

ME_AWS_fcw_mean	0.004468
ME_Car_high_beam_median	0.004398
DEM_evt_ha_lvl_H_mean	0.00438
DEM_evt_ha_lvl_H_sum	0.004301
DEM_evt_hb_lvl_M_mean	0.003326
DEM_evt_hb_lvl_M_sum	0.003312
DEM_evt_hc_lvl_H_sum	0.002814
DEM_evt_hc_lvl_H_mean	0.002757
DEM_evt_hb_lvl_H_mean	0.000723
DEM_evt_hb_lvl_H_sum	0.000644
ME_AWS_pcw_mean	0.00017
Drowsiness_level_median	8.16E-05

Πίνακας 5.4: Σημαντικότητα μεταβλητών για το Ηνωμένο Βασίλειο

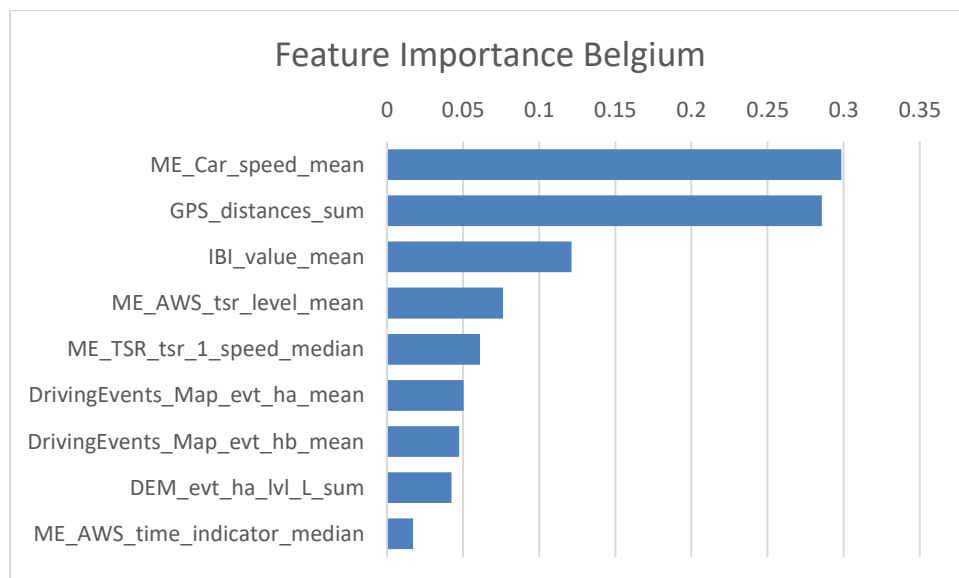
Feature	Importance
ME_Car_speed_mean	0.188244779
GPS_spd_mean	0.177346578
GPS_distances_sum	0.17680185
IBI_value_mean	0.081182426
ME_TSR_tsr_1_speed_median	0.045726241
ME_AWS_tsr_level_mean	0.031564727
DEM_evt_ha_lvl_L_sum	0.025457639
ME_AWS_time_indicator_median	0.021692391
DrivingEvents_Map_evt_ha_mean	0.020964096
DEM_evt_ha_lvl_L_mean	0.020845243
DEM_evt_hb_lvl_L_sum	0.020162531
DrivingEvents_Map_evt_hb_mean	0.016998255
DEM_evt_hb_lvl_L_mean	0.01675055
DrivingEvents_Map_lvl_L_mean	0.01436563
DrivingEvents_Map_lvl_M_mean	0.013361399
DrivingEvents_Map_evt_hc_mean	0.013211932
DEM_evt_hc_lvl_L_mean	0.012239941
DEM_evt_hc_lvl_L_sum	0.011973353
ME_Car_wipers_median	0.008667361
DEM_evt_hc_lvl_M_sum	0.008658806
DEM_evt_hc_lvl_M_mean	0.008631348
ME_LDW_Map_type_L_mean	0.008312557
ME_LDW_Map_type_R_mean	0.007822216
ME_AWS_pedestrian_dz_mean	0.007594819
DEM_evt_ha_lvl_M_sum	0.007369176
DEM_evt_ha_lvl_M_mean	0.007104642
ME_AWS_fcw_mean	0.005356239
DrivingEvents_Map_lvl_H_mean	0.004374043
DEM_evt_ha_lvl_H_mean	0.002928074
DEM_evt_ha_lvl_H_sum	0.002808335

DEM_evt_hc_lvl_H_sum	0.002488669
DEM_evt_hc_lvl_H_mean	0.002413061
DEM_evt_hb_lvl_M_sum	0.00237472
DEM_evt_hb_lvl_M_mean	0.002342834
ME_Car_high_beam_median	0.000991706
DEM_evt_hb_lvl_H_sum	0.000360485
DEM_evt_hb_lvl_H_mean	0.000356545
ME_AWS_pcw_mean	0.000154802
Drowsiness_level_median	0

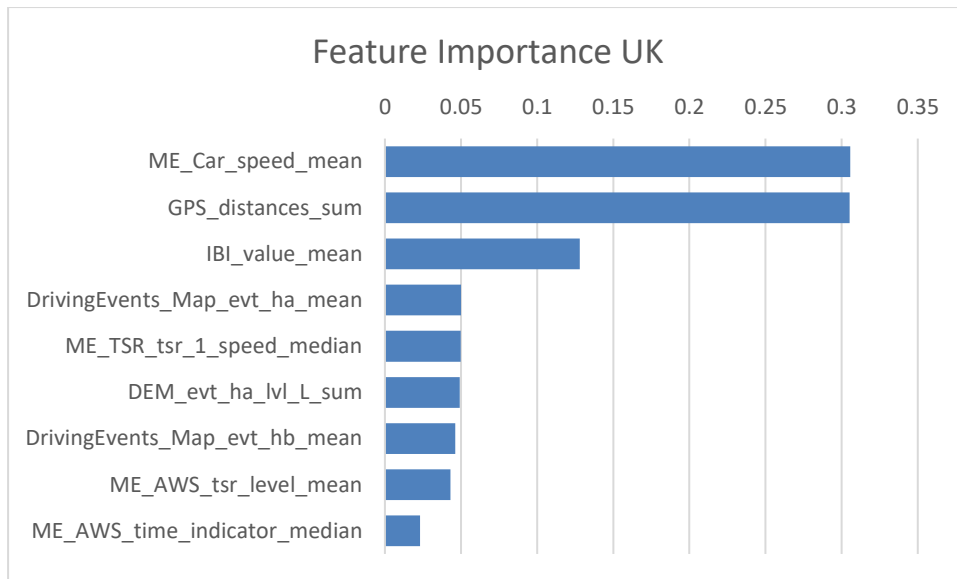
Το σύνολο μεταβλητών εισόδου που επιλέχθηκε και για τα δύο datasets είναι:

1. **ME_Car_speed_mean**
2. **GPS_distances_sum**
3. **IBI_value_mean**
4. **ME_AWS_tsr_level_mean**
5. **ME_TSR_tsr_1_speed_median**
6. **DrivingEvents_Map_evt_ha_mean**
7. **ME_AWS_time_indicator_median**
8. **DEM_evt_ha_lvl_L_sum**
9. **DrivingEvents_Map_evt_hb_mean**

Μετά την επιλογή των μεταβλητών εισόδου, υπολογίστηκε εκ νέου η σημαντικότητα των εννιά επιλεγμένων μεταβλητών με βάση των ταξινομητή “Τυχαίων Δασών” (Random Forests) και αποτυπώνεται στα Γραφήματα 5.3 και 5.4.



Γράφημα 5.3: Σημαντικότητα επιλεγμένων μεταβλητών για το Βέλγιο



Γράφημα 5.4: Σημαντικότητα επιλεγμένων μεταβλητών για το Ηνωμένο Βασίλειο

Αποδείχθηκε λοιπόν, πως η μέση ταχύτητα οχήματος (**ME_Car_speed_mean**) συνιστά τη σημαντικότερη μεταβλητή για τη πρόβλεψη της Ζώνης Ανοχής Ασφάλειας. Σχεδόν εξίσου σημαντική προκύπτει η συνολική απόσταση ταξιδιού (**GPS_distances_sum**), ενώ ακολουθεί το μέσο χρονικό διάστημα μεταξύ διαδοχικών καρδιακών παλμών (**IBI_value_mean**).

5.2.3 Προετοιμασία δεδομένων

Τα δεδομένα από το πείραμα χωρίστηκαν σε δύο ομάδες:

- **Δεδομένα εισόδου**, που περιλαμβάνουν εννιά μεταβλητές.
- **Δεδομένα εξόδου**, που αντιστοιχούν στα τρία επίπεδα της "Ζώνης Ανοχής Ασφαλείας".

Αυτά τα δεδομένα διαχωρίστηκαν σε:

- **Σετ εκπαίδευσης-training dataset (80%)**, που χρησιμοποιείται για να μάθει το μοντέλο να αναγνωρίζει μοτίβα και να προβλέπει το επίπεδο ασφαλείας.
- **Σετ εξέτασης-testing dataset (20%)**, που χρησιμοποιείται για να ελεγχθεί η ακρίβεια του μοντέλου, συγκρίνοντας τις προβλέψεις του με τις πραγματικές τιμές.

Στην ουσία, το μοντέλο μαθαίνει από τα δεδομένα εκπαίδευσης και μετά δοκιμάζεται σε νέα δεδομένα για να αξιολογηθεί η απόδοσή του.

5.2.4. Αντιμετώπιση άνισης κατανομής δεδομένων

Σύμφωνα με τη βιβλιογραφία, οι περισσότερες μελέτες επισημαίνουν το πρόβλημα της ανισορροπίας στα δεδομένα μεταξύ των διαφορετικών κατηγοριών. Συγκεκριμένα, τα δεδομένα που αντιστοιχούν σε επικίνδυνες

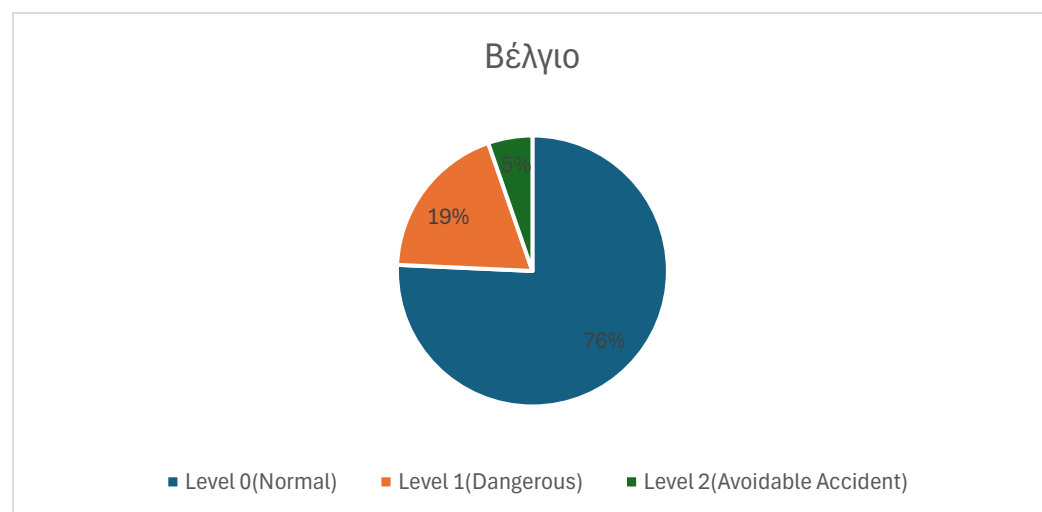
συνθήκες είναι σημαντικά λιγότερα σε σύγκριση με εκείνα που σχετίζονται με ασφαλή οδήγηση.

Επιπλέον, όπως αναφέρθηκε και στο κεφάλαιο 3, τα μοντέλα ταξινόμησης συνήθως υποθέτουν ότι οι κατηγορίες των δεδομένων κατανέμονται ομοιόμορφα. Ωστόσο, αυτή η υπόθεση τα καθιστά ιδιαίτερα επιρρεπή σε σφάλματα όταν τα δεδομένα είναι άνισα κατανεμημένα.

Η ανισορροπία αυτή γίνεται εμφανής κατά την κατηγοριοποίηση των δεδομένων σε επίπεδα ασφαλείας, όπως παρουσιάζεται στους πίνακες για το Βέλγιο και το Ηνωμένο Βασίλειο αντίστοιχα.

Πίνακας 5.5: Ποσοστό δειγμάτων “STZ level” για το Βέλγιο

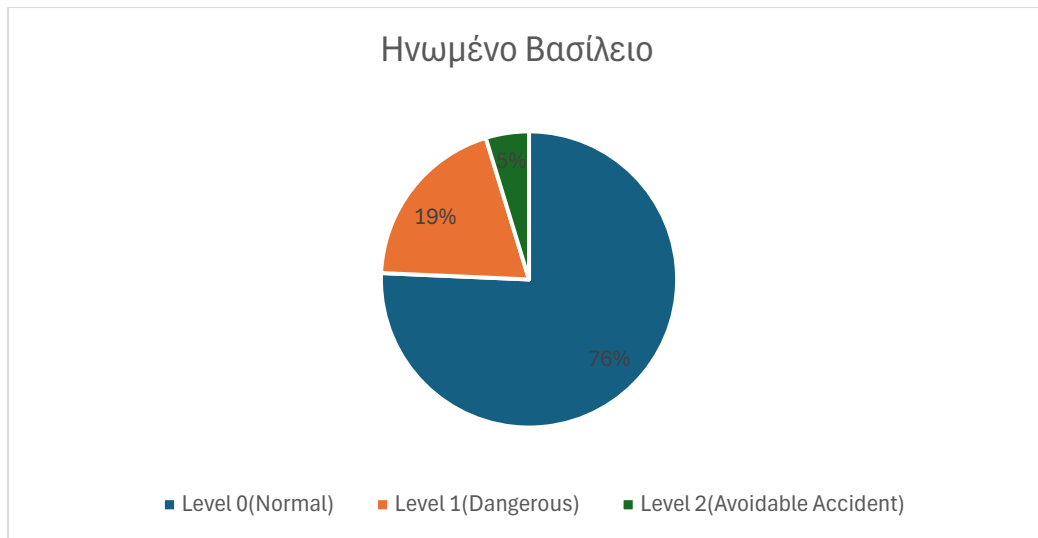
Βέλγιο		
STZ level	Αριθμός δειγμάτων	Ποσοστό δειγμάτων
Level 0 (Normal)	68.462	75,72
Level 1 (Dangerous)	17.165	18,98
Level 2 (Avoidable Accident)	4.787	5,29



Γράφημα 5.5: Ποσοστό δειγμάτων “STZ level” για το Βέλγιο

Πίνακας 5.6: Ποσοστό δειγμάτων “STZ level” για το Ηνωμένο Βασίλειο

Ηνωμένο Βασίλειο		
STZ level	Αριθμός δειγμάτων	Ποσοστό δειγμάτων
Level 0 (Normal)	53942	75.69
Level 1 (Dangerous)	13960	19.59
Level 2 (Avoidable Accident)	3365	4.72



Γράφημα 5.6: Ποσοστό δειγμάτων “STZ level” για το Ηνωμένο Βασίλειο

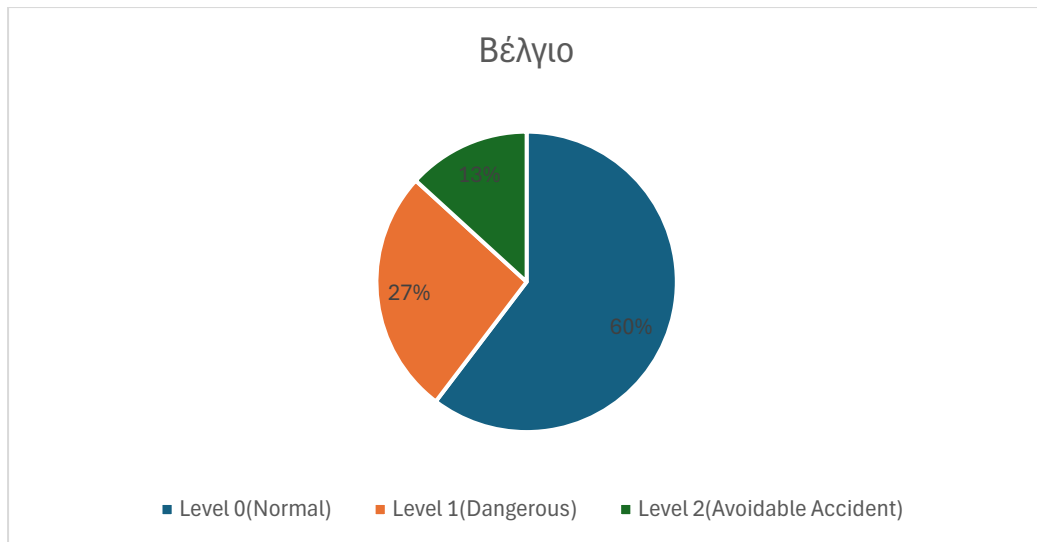
Για να αντιμετωπιστεί η ανισορροπία των δεδομένων μεταξύ των διαφορετικών κατηγοριών, χρησιμοποιήθηκε η μέθοδος **SMOTE (Synthetic Minority Over-sampling Technique)**. Η ανισορροπία αυτή μπορεί να επηρεάσει αρνητικά την απόδοση του μοντέλου, καθώς οι αλγόριθμοι ταξινόμησης τείνουν να δίνουν μεγαλύτερη βαρύτητα στην πλειοψηφούσα κλάση, οδηγώντας σε χαμηλή ακρίβεια πρόβλεψης για τις λιγότερο συχνές κατηγορίες. Το SMOTE αντιμετωπίζει αυτό το ζήτημα δημιουργώντας συνθετικά δείγματα για την υποεκπροσωπούμενη κατηγορία, αντί να αναπαράγει απλώς υπάρχοντα δεδομένα. Με αυτόν τον τρόπο, βελτιώνεται η γενίκευση του μοντέλου και μειώνεται η πιθανότητα υπερπροσαρμογής στα αρχικά δεδομένα εκπαίδευσης, επιτρέποντας πιο αξιόπιστη ταξινόμηση σε πραγματικές συνθήκες.

Δοκιμάστηκε το μοντέλο SMOTE όπου για την εξισορρόπηση των δεδομένων διατηρήθηκε ο αριθμός δειγμάτων του “Level 0”, ενώ ο αριθμός δειγμάτων ορίστηκε 30000 και 15000 αντίστοιχα για τα “Level 1” και “Level 2”, όπως παρουσιάζεται στους παρακάτω πίνακες.

Μετά την εφαρμογή της μεθόδου SMOTE:

Πίνακας 5.7: Ποσοστό δειγμάτων “STZ level” για το Βέλγιο με την εφαρμογή SMOTE

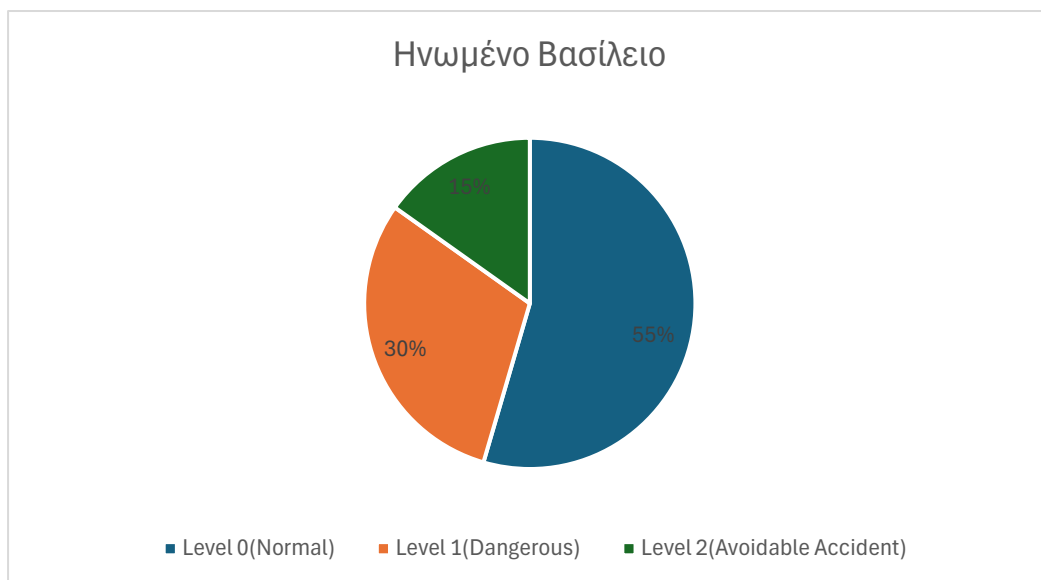
Βέλγιο		
STZ level	Αριθμός δειγμάτων	Ποσοστό δειγμάτων
Level 0 (Normal)	68462	60.34
Level 1 (Dangerous)	30000	26.44
Level 2 (Avoidable Accident)	15000	13.22



Γράφημα 5.7: Ποσοστό δειγμάτων “STZ level” για το Βέλγιο με την εφαρμογή SMOTE

Πίνακας 5.8: Ποσοστό δειγμάτων “STZ level” για το Ηνωμένο Βασίλειο με την εφαρμογή SMOTE

Ηνωμένο Βασίλειο		
STZ level	Αριθμός δειγμάτων	Ποσοστό δειγμάτων
Level 0 (Normal)	53942	54.52
Level 1 (Dangerous)	30000	30.32
Level 2 (Avoidable Accident)	15000	15.16



Γράφημα 5.8: Ποσοστό δειγμάτων “STZ level” για το Ηνωμένο Βασίλειο με την εφαρμογή SMOTE

5.2.5. Ανάπτυξη μοντέλων ταξινόμησης

Όπως αναφέρθηκε προηγουμένως, ορισμένοι αλγόριθμοι ταξινόμησης σχεδιάστηκαν για να αναγνωρίζουν το επίπεδο της "Ζώνης Ανοχής Ασφαλείας"

στο οποίο βρίσκεται ο οδηγός κάθε 30 δευτερόλεπτα. Η επιλογή των τεσσάρων μοντέλων βασίστηκε στη βιβλιογραφική ανασκόπηση. Μεταξύ άλλων, δοκιμάστηκαν και τα μοντέλα Logistic Regression, SVM, XGBoost, τα οποία απορρίφθηκαν διότι δεν εμφάνιζαν επαρκείς τιμές μετρικών αξιολόγησης Ανάκλησης (Recall) και f1-score. Στον πίνακα παρουσιάζονται οι ονομασίες και οι αντίστοιχοι συμβολισμοί των μοντέλων που επιλέχθηκαν τελικά.

Πίνακας 5.9: Μοντέλα Ταξινόμησης Οδηγικής Συμπεριφοράς

Όνομα Μοντέλου (Ελληνικά)	Όνομα Μοντέλου (Αγγλικά)	Συμβολισμός Μοντέλου
Τυχαίο Δάσος	Random Forest	RF
Ελαφρύς Αλγόριθμος Βαθιάς Ενίσχυσης	Light Gradient Boosting Machine	LightGBM
Αλγόριθμος Ενίσχυσης Κατηγοριών	Categorical Boosting	CatBoost
Πολυεπίπεδο Αντιληπτικό	Multi-Layer Perceptron	MLP

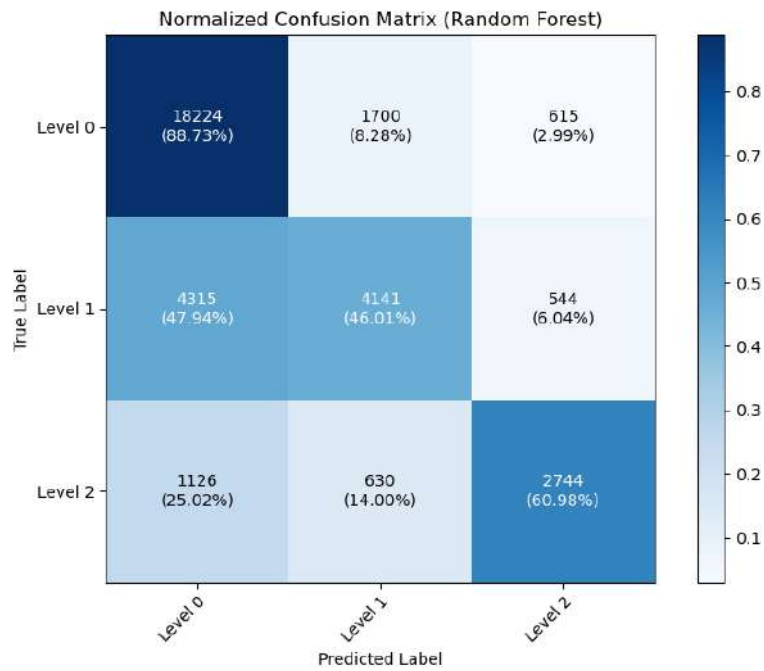
Η εκπαίδευση όλων των μοντέλων έγινε με την τεχνική SMOTE που αναλύθηκε στο 5.2.4., για να μπορέσουν με καλύτερο τρόπο να προβλεφθούν οι 3 κατηγορίες που τέθηκαν.

Παρακάτω παρουσιάζονται για κάθε μοντέλο ταξινόμησης και για τις δύο χώρες οι μήτρες σύγχυσης για την γραφική αναπαράσταση της επίδοσης τους.

Βέλγιο

1. Random Forest

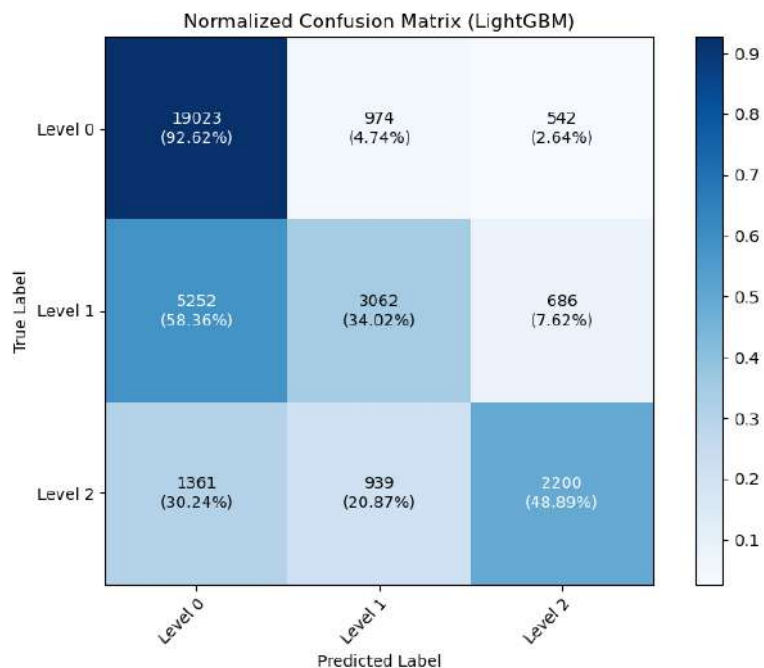
Όπως φαίνεται στο γράφημα ο αλγόριθμος RF έχει την δυνατότητα να προβλέψει το επίπεδο 'Normal' με πολύ καλή ποσότητα πρόβλεψης και τα επίπεδα 'Avoidable Accident' και 'Dangerous' με καλή και μέτρια ευστοχία αντίστοιχα. Επομένως, συνολικά θεωρείται ένα μοντέλο με καλή ικανότητα αναγνώρισης επικίνδυνων συμπεριφορών.



Γράφημα 5.9: Μήτρα σύγχυσης αλγορίθμου Random Forest για το Βέλγιο

2. LightGBM

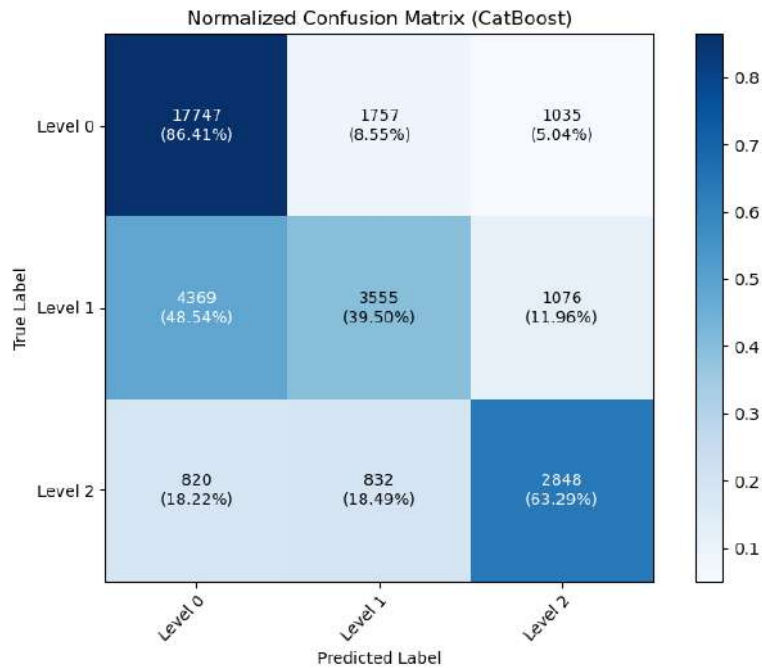
Όπως φαίνεται στο γράφημα ο αλγόριθμος LightGBM έχει την δυνατότητα να προβλέψει το επίπεδο 'Normal' με πολύ καλή ποσότητα πρόβλεψης και τα επίπεδα 'Avoidable Accident' και 'Dangerous' με μέτρια ευστοχία. Επομένως, συνολικά θεωρείται ένα μοντέλο με καλή ικανότητα αναγνώρισης επικίνδυνων συμπεριφορών.



Γράφημα 5.10: Μήτρα σύγχυσης αλγορίθμου LightGBM για το Βέλγιο

3. CatBoost

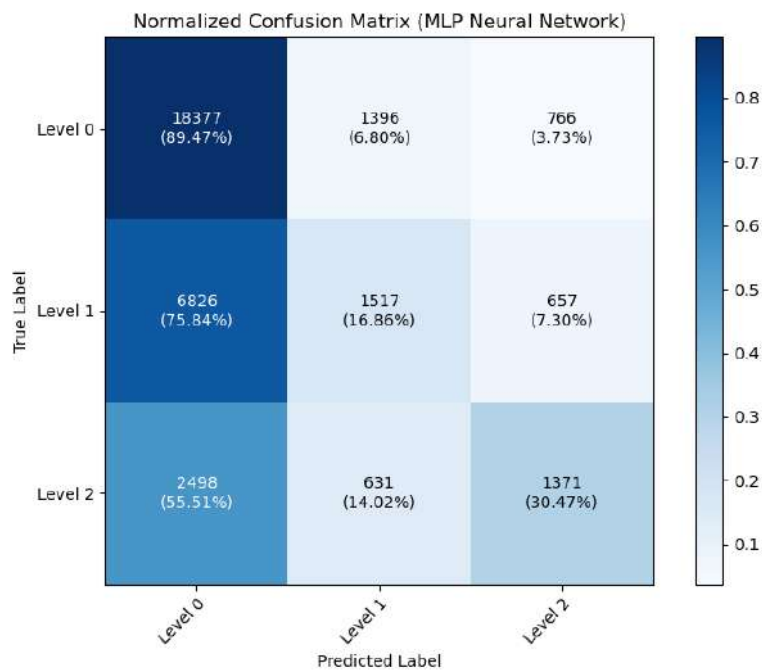
Όπως φαίνεται στο γράφημα ο αλγόριθμος CatBoost έχει την δυνατότητα να προβλέψει το επίπεδο 'Normal' με πολύ καλή ποσότητα πρόβλεψης και τα επίπεδα 'Avoidable Accident' και 'Dangerous' με καλή και μέτρια ευστοχία αντίστοιχα. Επομένως, συνολικά θεωρείται ένα μοντέλο καλή ικανότητα αναγνώρισης επικίνδυνων συμπεριφορών.



Γράφημα 5.11: Μήτρα σύγχυσης αλγορίθμου CatBoost για το Βέλγιο

4. MLP

Όπως φαίνεται στο γράφημα ο αλγόριθμος MLP έχει την δυνατότητα να προβλέψει το επίπεδο 'Normal' με πολύ καλή ποσότητα πρόβλεψης και τα επίπεδα 'Avoidable Accident' και 'Dangerous' με μέτρια και χαμηλή ευστοχία αντίστοιχα. Επομένως, συνολικά θεωρείται ένα μοντέλο με μη ικανοποιητική ικανότητα αναγνώρισης επικίνδυνων συμπεριφορών.

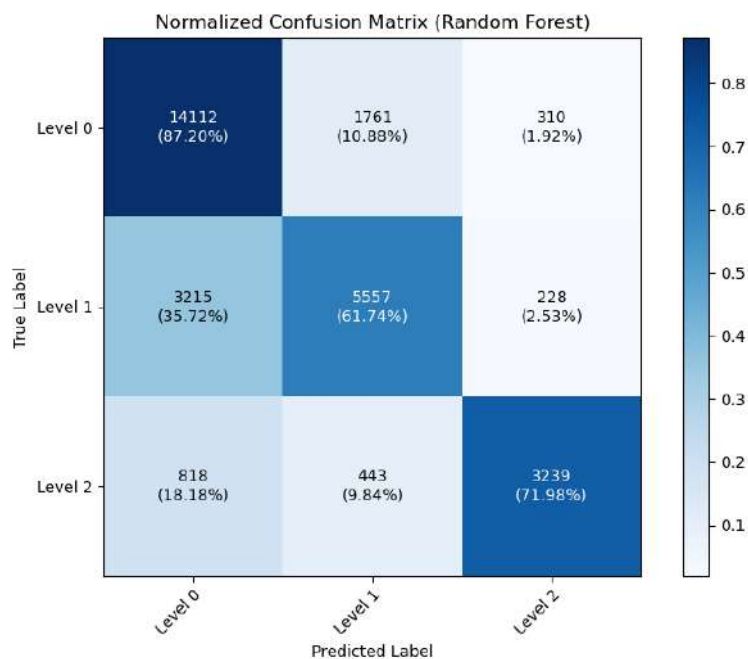


Γράφημα 5.12: Μήτρα σύγχυσης αλγορίθμου MLP για το Βέλγιο

Ηνωμένο Βασίλειο

1. Random Forest

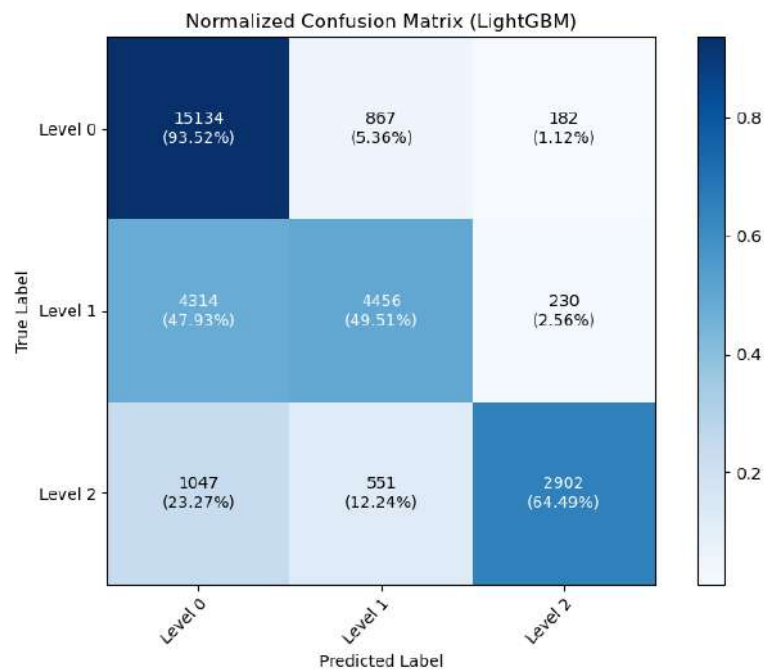
Όπως φαίνεται στο γράφημα ο αλγόριθμος RF έχει την δυνατότητα να προβλέψει το επίπεδο 'Normal' με πολύ καλή ποσότητα πρόβλεψης και τα επίπεδα 'Avoidable Accident' και 'Dangerous' με καλή ευστοχία. Επομένως, συνολικά θεωρείται ένα μοντέλο με καλή ικανότητα αναγνώρισης επικίνδυνων συμπεριφορών.



Γράφημα 5.13: Μήτρα σύγχυσης αλγορίθμου Random Forest για το Ηνωμένο Βασίλειο

2. LightGBM

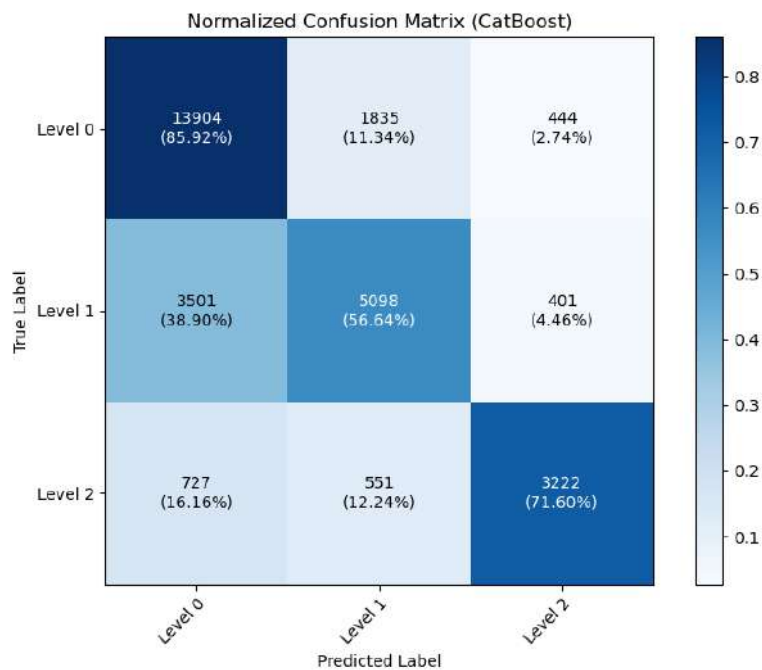
Όπως φαίνεται στο γράφημα ο αλγόριθμος LightGBM έχει την δυνατότητα να προβλέψει το επίπεδο 'Normal' με πολύ καλή ποσότητα πρόβλεψης και τα επίπεδα 'Avoidable Accident' και 'Dangerous' με καλή και μέτρια ευστοχία αντίστοιχα. Επομένως, συνολικά θεωρείται ένα μοντέλο με καλή ικανότητα αναγνώρισης επικίνδυνων συμπεριφορών.



Γράφημα 5.14: Μήτρα σύγχυσης αλγορίθμου LightGBM για το Ηνωμένο Βασίλειο

3. CatBoost

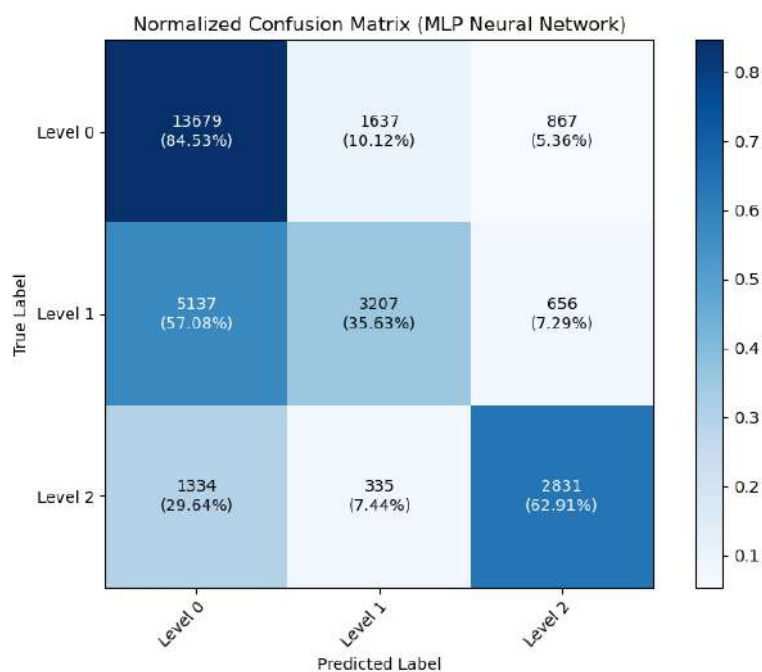
Όπως φαίνεται στο γράφημα ο αλγόριθμος CatBoost έχει την δυνατότητα να προβλέψει το επίπεδο 'Normal' με πολύ καλή ποσότητα πρόβλεψης και τα επίπεδα 'Avoidable Accident' και 'Dangerous' με καλή ευστοχία. Επομένως, συνολικά θεωρείται ένα μοντέλο καλή ικανότητα αναγνώρισης επικίνδυνων συμπεριφορών.



Γράφημα 5.15: Μήτρα σύγχυσης αλγορίθμου CatBoost για το Ηνωμένο Βασίλειο

4. MLP

Όπως φαίνεται στο γράφημα ο αλγόριθμος MLP έχει την δυνατότητα να προβλέψει το επίπεδο 'Normal' με πολύ καλή ποσότητα πρόβλεψης και τα επίπεδα 'Avoidable Accident' και 'Dangerous' με καλή και μέτρια ευστοχία αντίστοιχα. Επομένως, συνολικά θεωρείται ένα μοντέλο με ικανοποιητική ικανότητα αναγνώρισης επικίνδυνων συμπεριφορών.

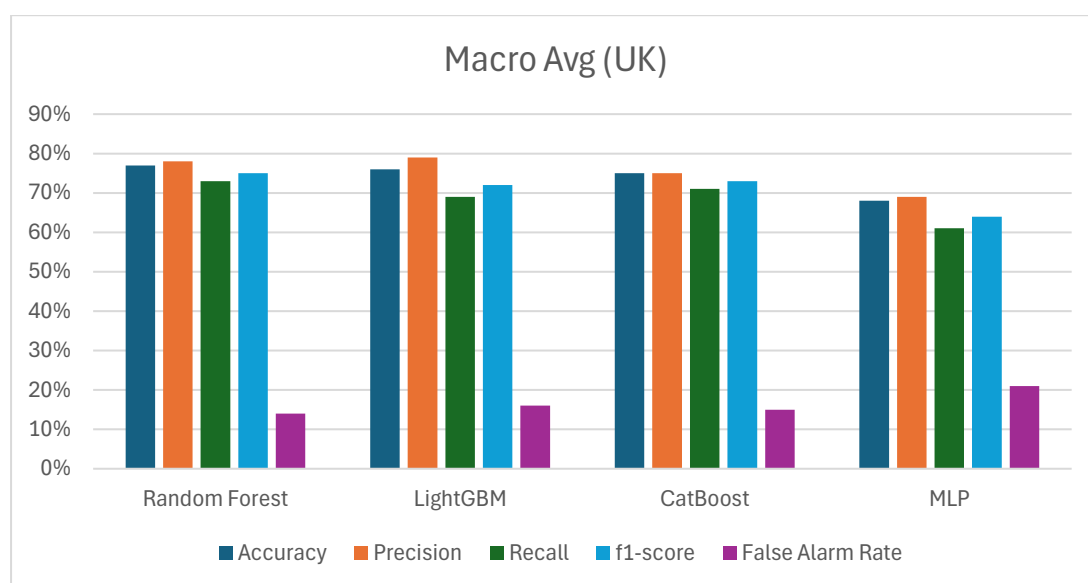


Γράφημα 5.16: Μήτρα σύγχυσης αλγορίθμου MLP για το Ηνωμένο Βασίλειο

5.2.6. Σύγκριση Μετρικών αξιολόγησης των μοντέλων

Οι διάφορες τεχνικές επεξεργασίας δεδομένων, καθώς και η επιλογή των καταλληλότερων παραμέτρων των αλγορίθμων, αποσκοπούσαν στη βελτίωση της προγνωστικής ικανότητας των μοντέλων. Στους πίνακες, καθώς και στα γραφήματα στη συνέχεια, παρουσιάζονται βασικές μετρικές αξιολόγησης των τεσσάρων μοντέλων προς σύγκριση.

Ηνωμένο Βασίλειο

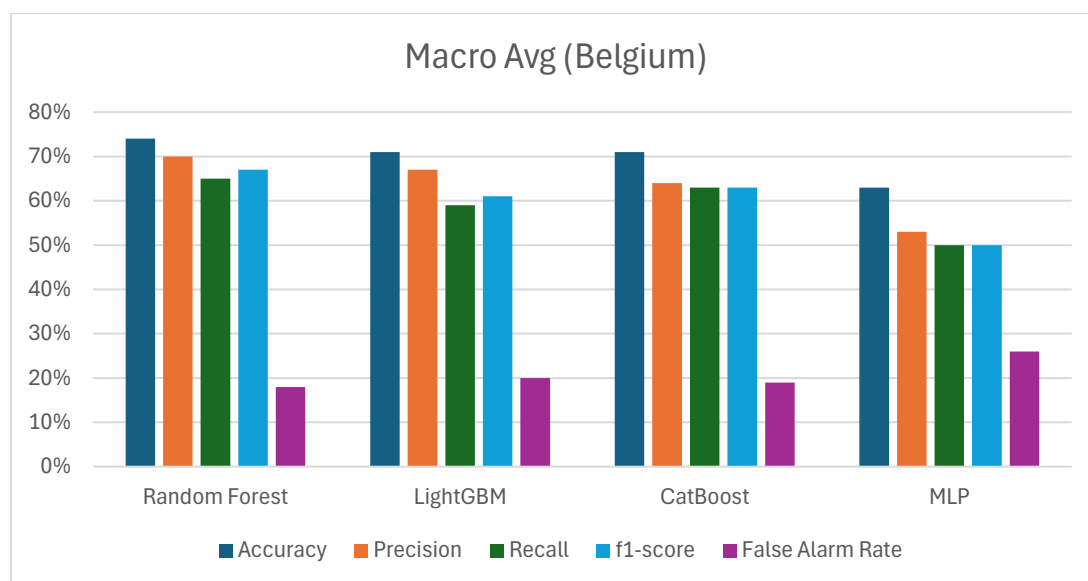


Γράφημα 5.17: Μετρικές αξιολόγησης μοντέλων ταξινόμησης για το Ηνωμένο Βασίλειο

Πίνακας 5.10: Μετρικές αξιολόγησης μοντέλων ταξινόμησης για το Ηνωμένο Βασίλειο

	Accuracy	Precision	Recall	f1-score	False Alarm Rate
Random Forest	77%	78%	73%	75%	14%
LightGBM	76%	79%	69%	72%	16%
CatBoost	75%	75%	71%	73%	15.00%
MLP	68%	69%	61%	64%	21%

Βέλγιο



Γράφημα 5.18: Μετρικές αξιολόγησης μοντέλων ταξινόμησης για το Βέλγιο

Πίνακας 5.11: Μετρικές αξιολόγησης μοντέλων ταξινόμησης για το Βέλγιο

	Accuracy	Precision	Recall	f1-score	False Alarm Rate
Random Forest	74%	70%	65%	67%	18%
LightGBM	71%	67%	59%	61%	20%
CatBoost	71%	64%	63%	63%	19.00%
MLP	63%	53%	50%	50%	26%

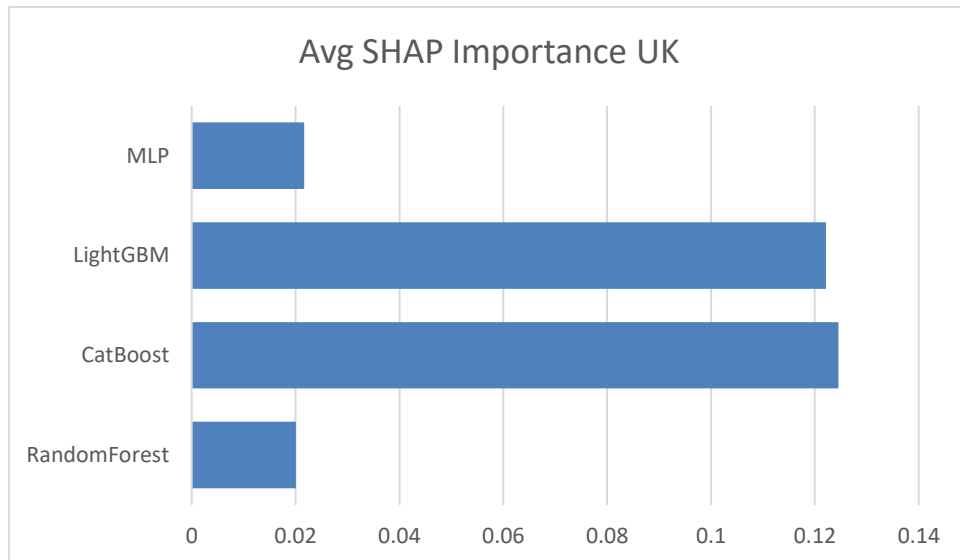
Με βάση τα δείγματα και των δύο χωρών ο **αλγόριθμος ταξινόμησης MLP** παρουσιάζει **τα πιο χαμηλά ποσοστά** Ορθότητας (Accuracy), Ανάκλησης (Recall), Ακρίβειας (Precision) και f1-score ενώ παρουσιάζει και το πιο υψηλό ποσοστό False Alarm Rate. Συνεπώς, κρίνεται ακατάλληλο ως μοντέλο ταξινόμησης.

Τα υπόλοιπα 3 μοντέλα ταξινόμησης, και για τις δύο χώρες, παρουσιάζουν ικανοποιητικά ποσοστά Ορθότητας (Accuracy), Ανάκλησης (Recall), Ακρίβειας (Precision) και f1-score και με βάση αυτά δε μπορεί να ξεχωρίσει κάποιο από τα τρία ως το καταλληλότερο. Για το λόγο αυτόν, στην επόμενη ενότητα, θα εφαρμοστεί η μέθοδος SHAP και μέσω της μέσης σημαντικότητας SHAP θα ξεχωρίσουν οι δύο καλύτεροι αλγόριθμοι ταξινόμησης με βάση τους οποίους θα εφαρμοστεί η μέθοδος SHAP.

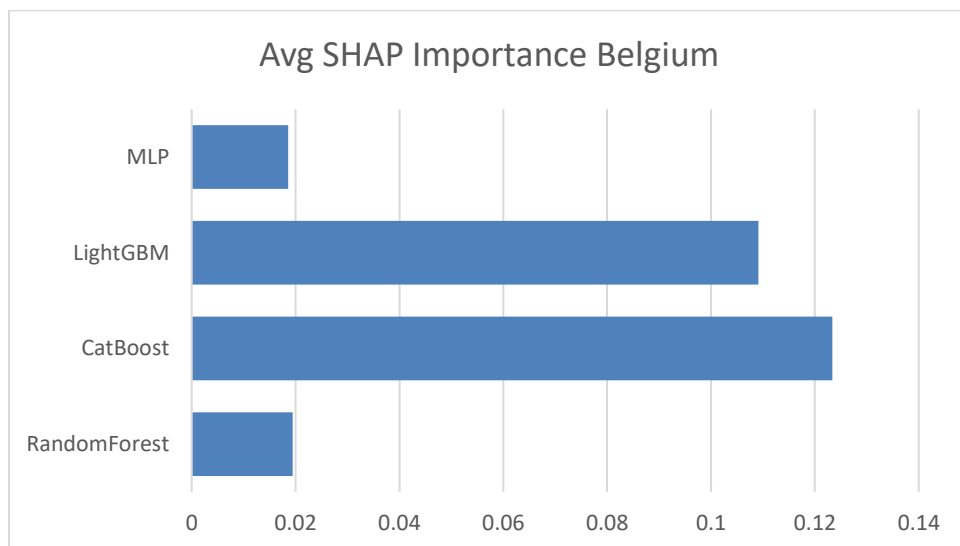
5.3 Εξήγηση Λειτουργίας Μοντέλων Μηχανικής μάθησης

Έπειτα, για να κατανοήσουμε καλύτερα τα μοντέλα μηχανικής μάθησης, αναλύσαμε τις τιμές **SHAP** (SHapley Additive exPlanations). Το SHAP χρησιμοποιείται για να εξηγήσει την επίδραση κάθε χαρακτηριστικού στο

μοντέλο, επιτρέποντας τόσο την τοπική όσο και τη συνολική ανάλυση των δεδομένων. Τα μοντέλα **CatBoost** και **LightGBM** παρουσίασαν την υψηλότερη μέση σημαντικότητα SHAP για τα δεδομένα και των δύο χωρών, όπως φαίνεται και στα παρακάτω γραφήματα.



Γράφημα 5.19: Μέση σημαντικότητα SHAP μοντέλων ταξινόμησης για το Ηνωμένο Βασίλειο

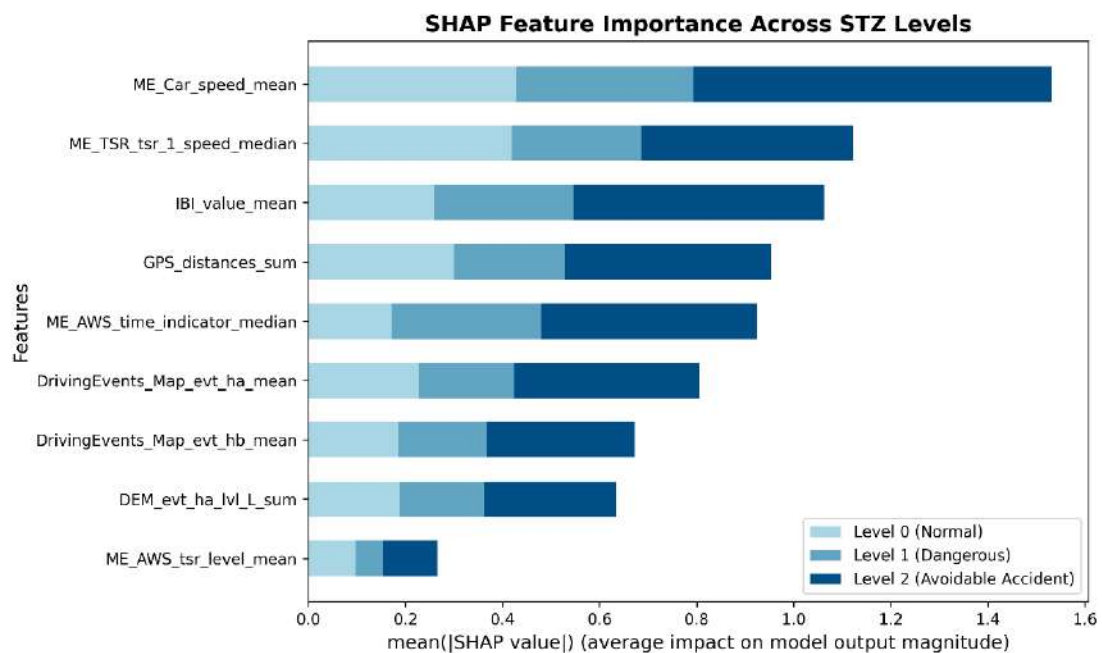


Γράφημα 5.20: Μέση σημαντικότητα SHAP μοντέλων ταξινόμησης για το Βέλγιο

Για τα δύο αυτά μοντέλα δημιουργήθηκαν τα διαγράμματα που ακολουθούν. Τα διαγράμματα δείχνουν τη συμβολή κάθε μεταβλητής στο μοντέλο και τον βαθμό επίδρασης κάθε κατηγορίας στην οποία ανήκει.

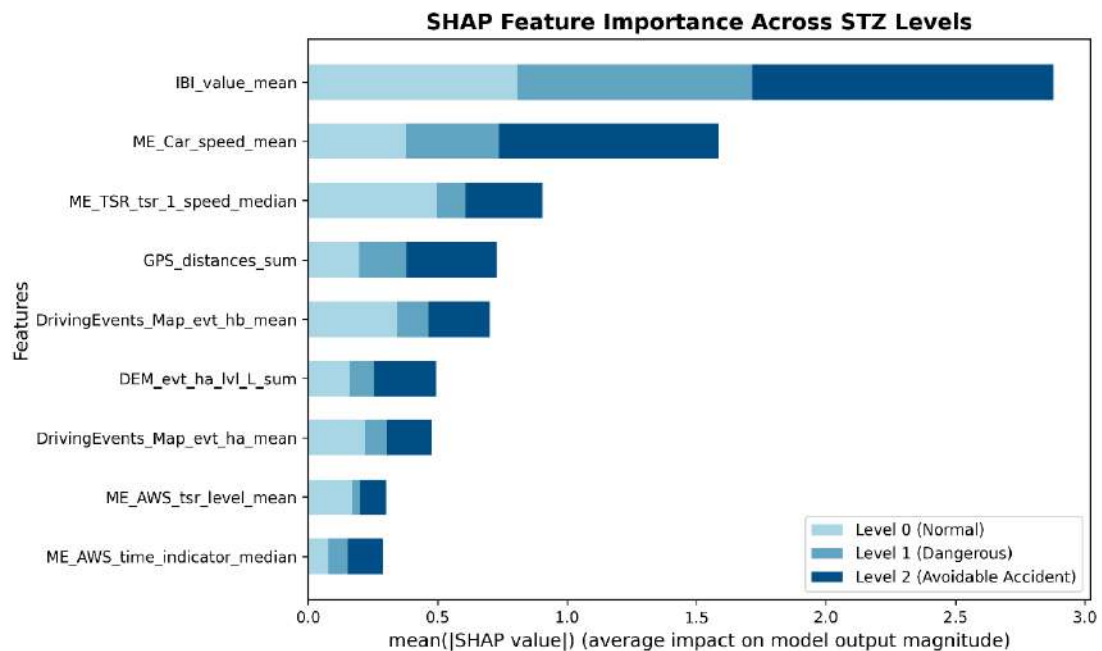
Ηνωμένο Βασίλειο

CatBoost



Γράφημα 5.21: Σημαντικότητα μεταβλητών με CatBoost για κάθε Level, Ηνωμένο Βασίλειο

LightGBM



Γράφημα 5.22: Σημαντικότητα μεταβλητών με LightGBM για κάθε Level, Ηνωμένο Βασίλειο

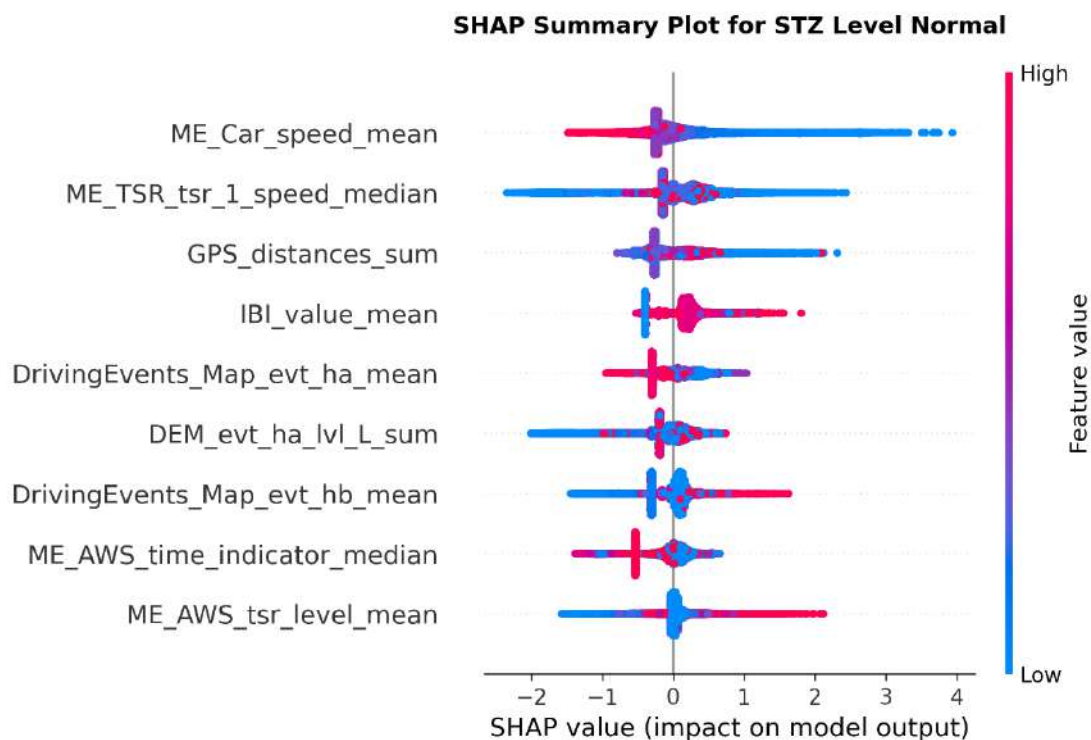
Παρατηρείται πως για τα δεδομένα του **Ηνωμένου Βασιλείου**, για τον αλγόριθμο ταξινόμησης **CatBoost**, σημαντικότερη μεταβλητή αποδεικνύεται η

ME_Car_speed_mean, ιδίως για το Level 2 (Avoidable Accident). Στην περίπτωση χρήσης του αλγορίθμου ταξινόμησης **LightGBM**, προκύπτει ως σημαντικότερη μεταβλητή η **IBI_value_mean** και για τα τρία επίπεδα.

Τα διαγράμματα που ακολουθούν δείχνουν τα χαρακτηριστικά που επηρεάζουν την πρόβλεψη. Όσα αυξάνουν την πρόβλεψη εμφανίζονται με κόκκινο, ενώ όσα τη μειώνουν με μπλε. Τα δεδομένα προέρχονται από την οδηγική συμπεριφορά στο Ηνωμένο Βασίλειο και το Βέλγιο που καταγράφηκε στη βάση δεδομένων i-Dreams.

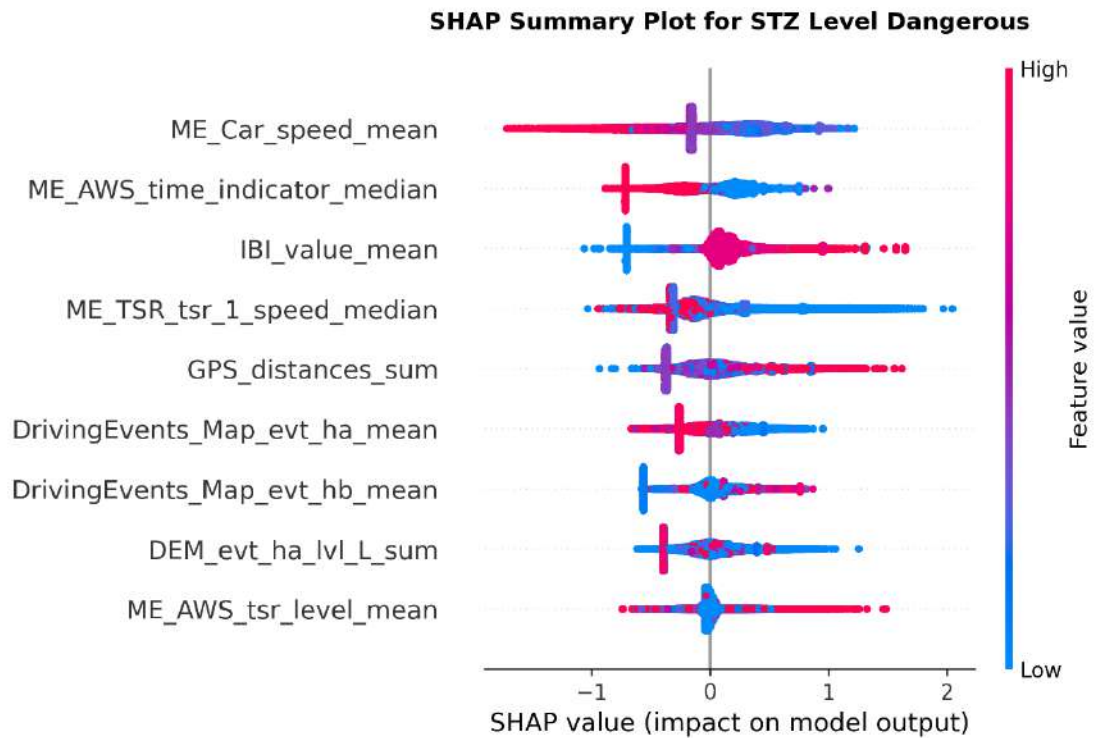
CatBoost

Level 0



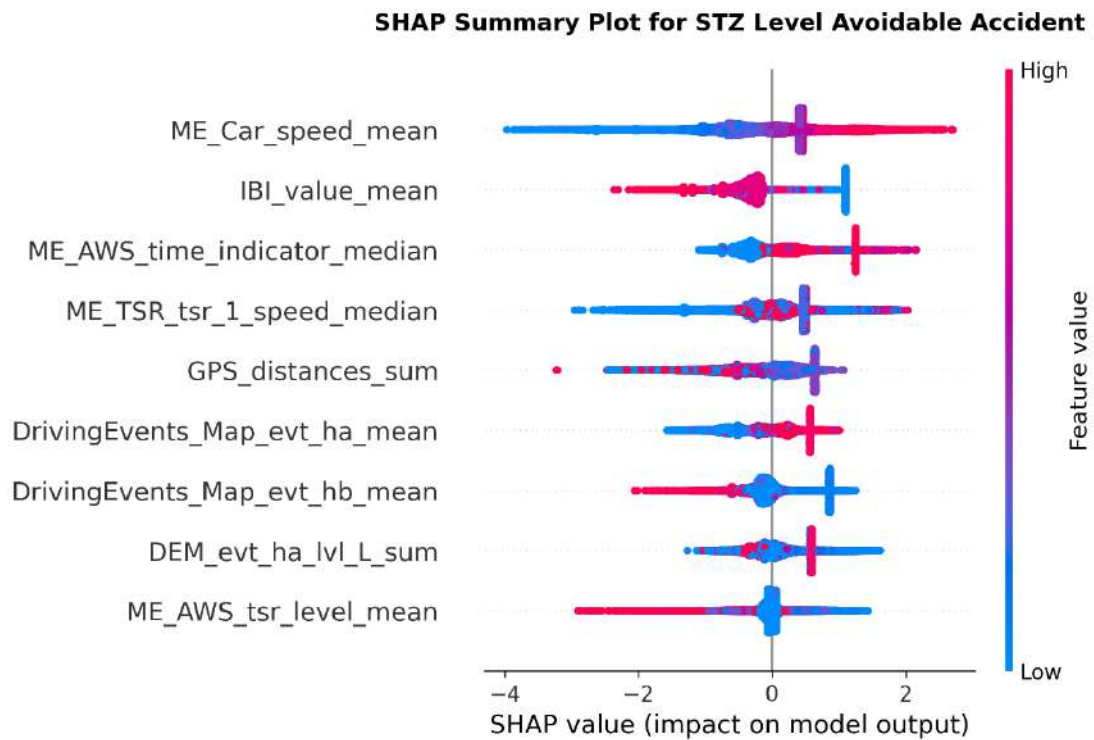
Γράφημα 5.23: Επίδραση μεταβλητών με CatBoost Level Normal, Ηνωμένο Βασίλειο

Level 1



Γράφημα 5.24: Επίδραση μεταβλητών με CatBoost Level Dangerous, Ηνωμένο Βασίλειο

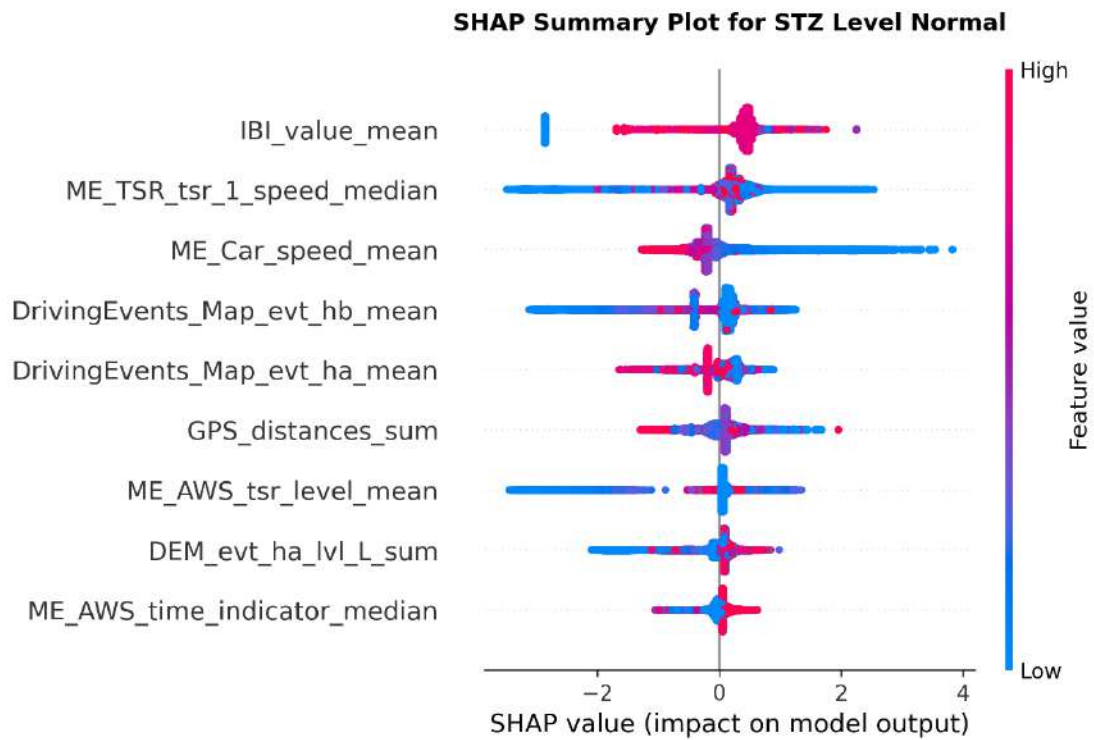
Level 2



Γράφημα 5.25: Επίδραση μεταβλητών με CatBoost Level Avoidable Accident, Ηνωμένο Βασίλειο

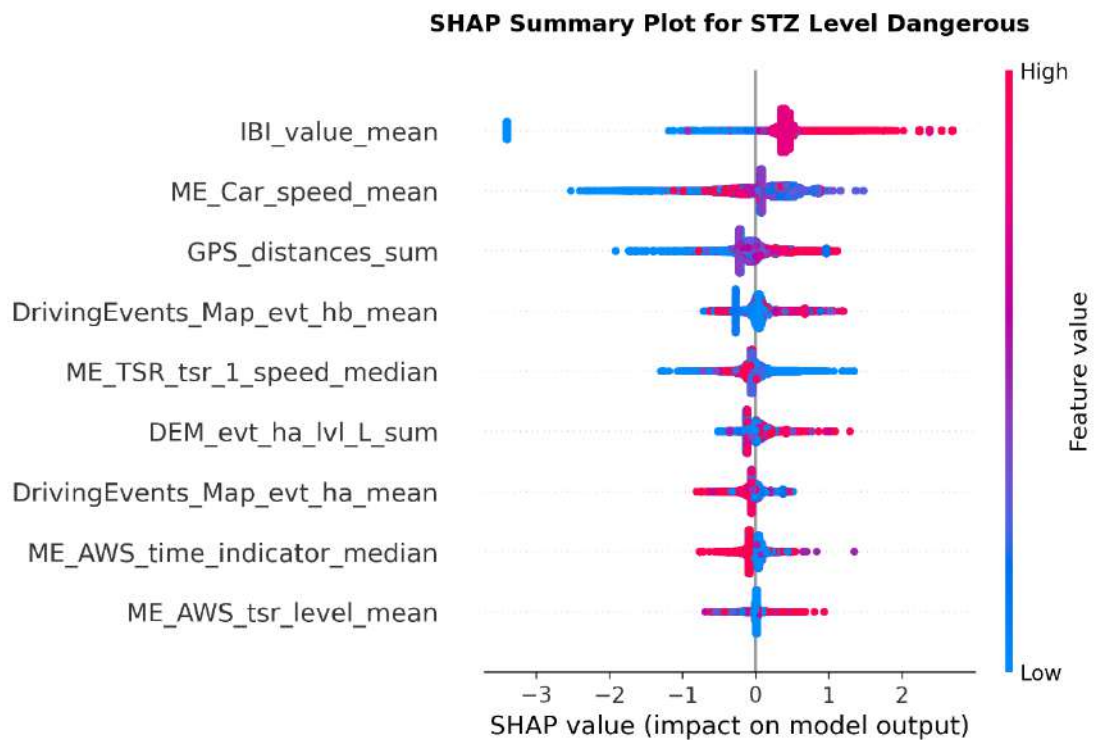
LightGBM

Level 0



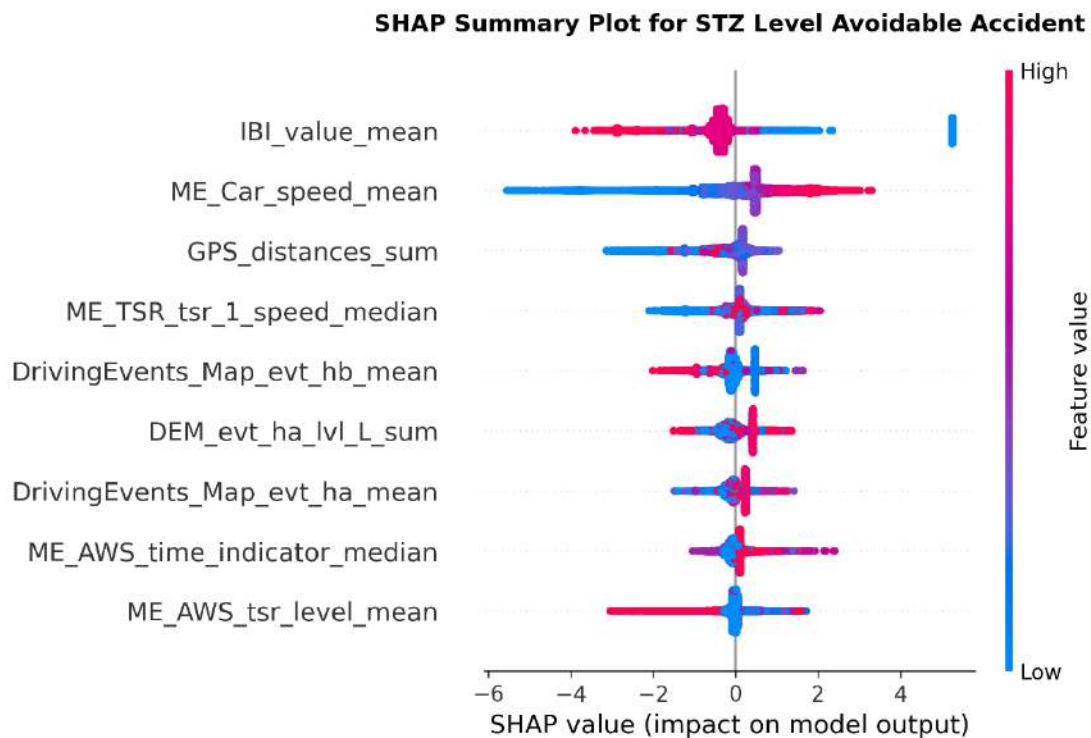
Γράφημα 5.26: Επίδραση μεταβλητών με LightGBM Level Normal, Ηνωμένο Βασίλειο

Level 1



Γράφημα 5.27: Επίδραση μεταβλητών με LightGBM Level Dangerous, Ηνωμένο Βασίλειο

Level 2



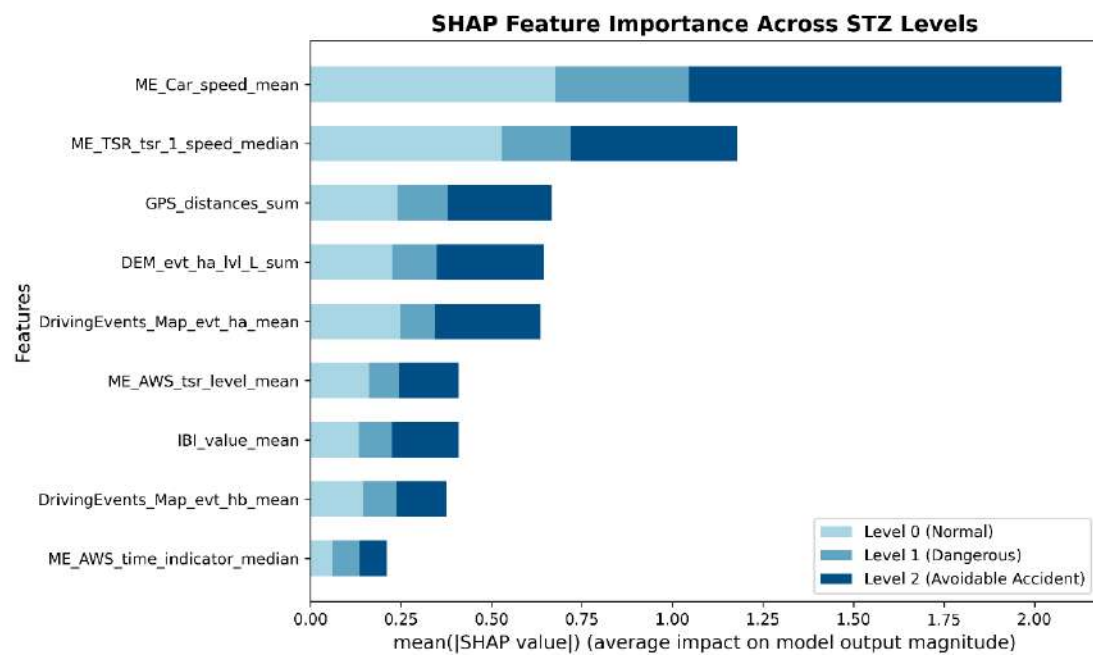
Γράφημα 5.28: Επίδραση μεταβλητών με LightGBM Level Avoidable Accident, Ηνωμένο Βασίλειο

Για τα δεδομένα του **Ηνωμένου Βασιλείου**, προκύπτει για το **CatBoost**, πως η σημαντικότερη μεταβλητή είναι η **ME_Car_speed_mean** που ωθεί στην αλλαγή επιπέδου επικίνδυνης οδηγικής συμπεριφοράς. Πιο συγκεκριμένα, στο “Level 0”, καθώς και στο “Level 1”, φαίνεται πως υψηλές τιμές της μεταβλητής (κόκκινο χρώμα) μειώνουν την πιθανότητα να καταταγεί το όχημα στην “STZ Level Normal” και στην “STZ Level Dangerous” αντίστοιχα. Στο “Level 2”, αποδεικνύεται πως υψηλές τιμές της μεταβλητής (κόκκινο χρώμα) ωθούν το όχημα προς την κατάταξη του στο “STZ Level Avoidable Accident”, κάτι αναμενόμενο.

Για το **LightGBM** αντίστοιχα, αναδεικνύεται ως η σημαντικότερη μεταβλητή η **IBI_value_mean** που ωθεί στην αλλαγή επιπέδου επικίνδυνης οδηγικής συμπεριφοράς.

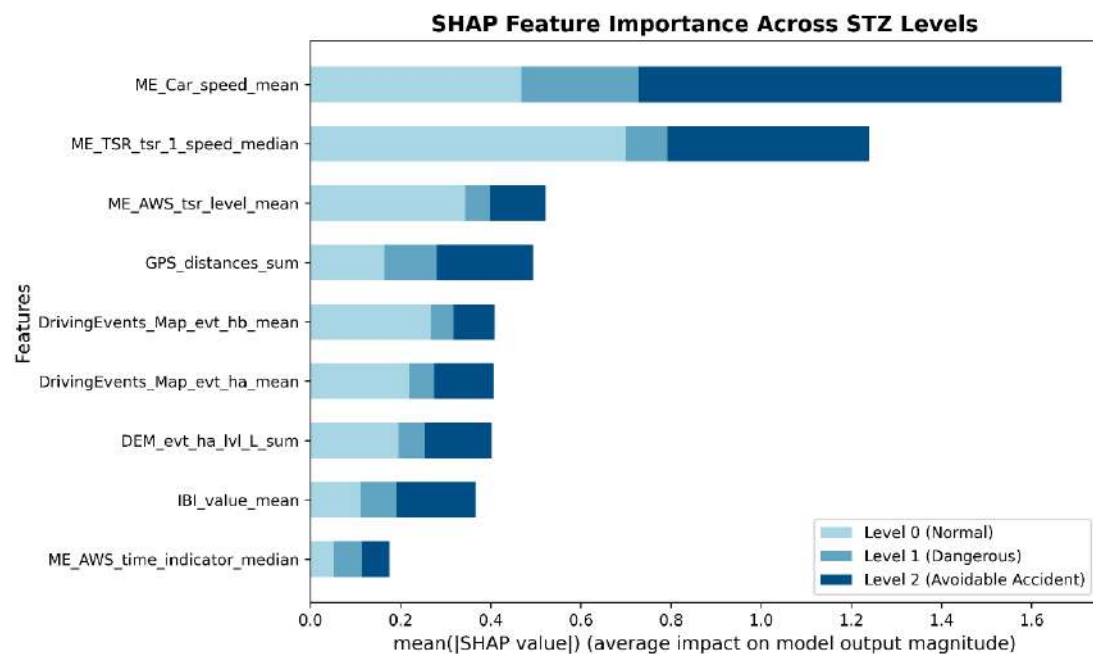
Βέλγιο

CatBoost



Γράφημα 5.29: Σημαντικότητα μεταβλητών με CatBoost για κάθε Level, Βέλγιο

LightGBM



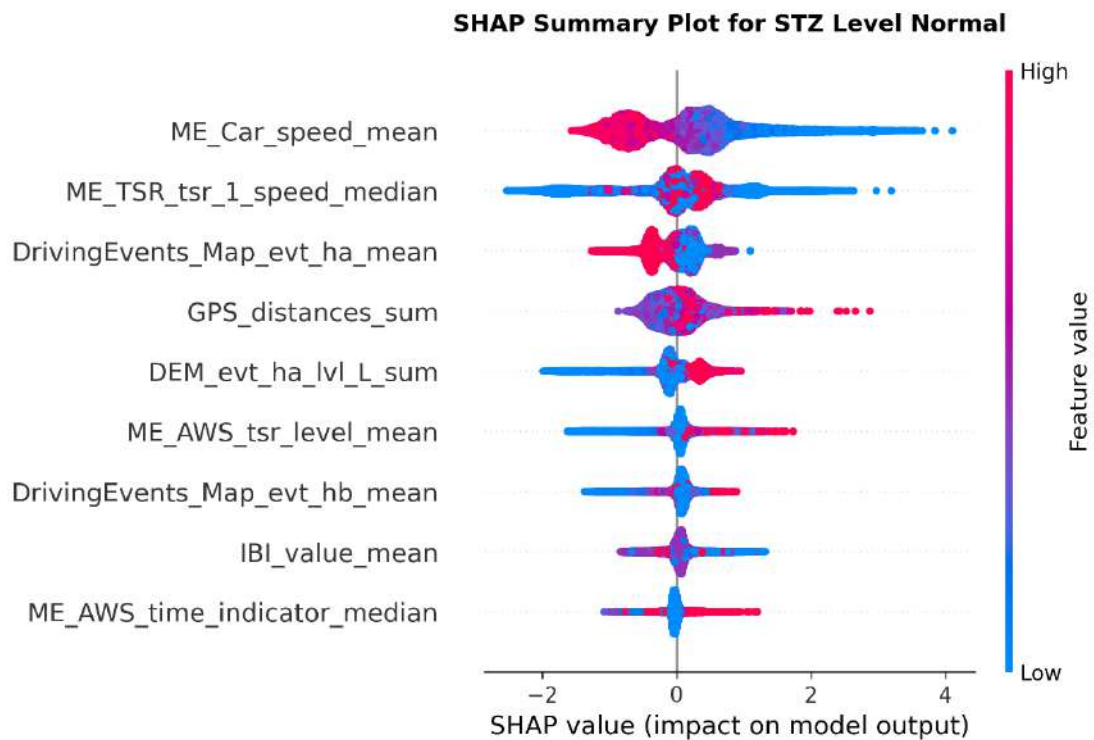
Γράφημα 5.30: Σημαντικότητα μεταβλητών με LightGBM για κάθε Level, Βέλγιο

Προκύπτει πως για τα δεδομένα του **Βελγίου**, για τη χρήση και των δύο αλγορίθμων, σημαντικότερη μεταβλητή αναδεικνύεται η **ME_Car_speed_mean**, κυρίως για το Level 2 (Avoidable Accident).

Τα διαγράμματα που ακολουθούν δείχνουν τα χαρακτηριστικά που επηρεάζουν την πρόβλεψη. Όσα αυξάνουν την πρόβλεψη εμφανίζονται με κόκκινο, ενώ όσα τη μειώνουν με μπλε. Τα δεδομένα προέρχονται από την οδηγική συμπεριφορά στο Ηνωμένο Βασίλειο και το Βέλγιο που καταγράφηκε στη βάση δεδομένων i-Dreams.

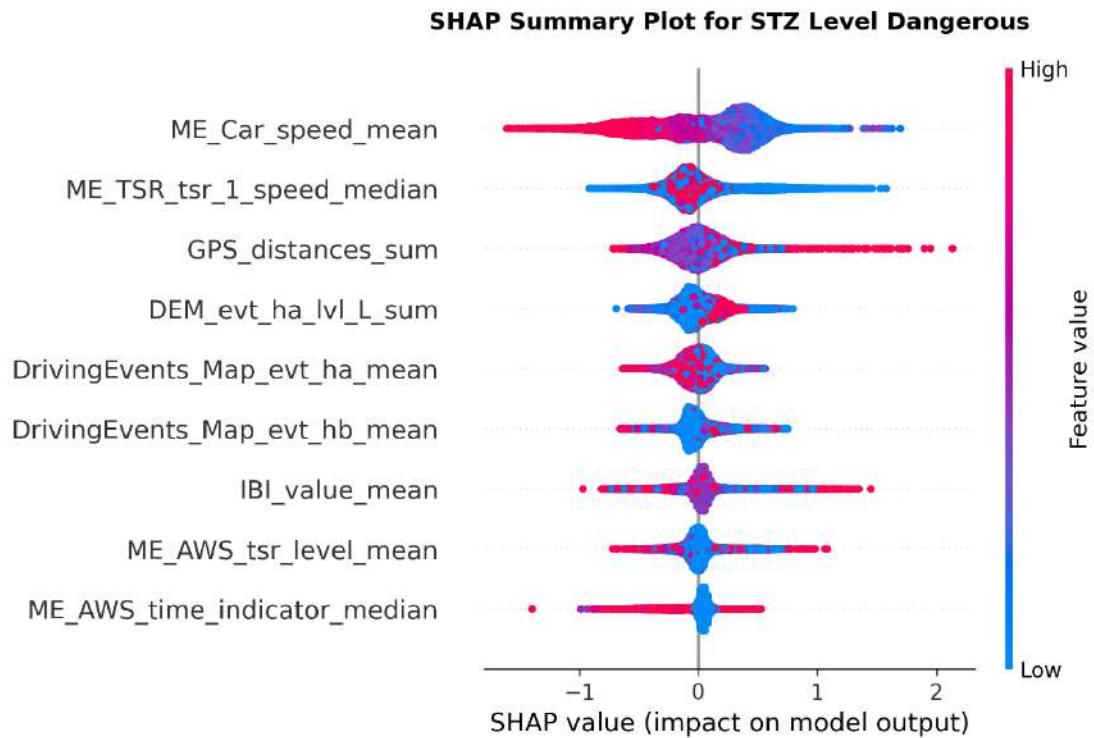
CatBoost

Level 0



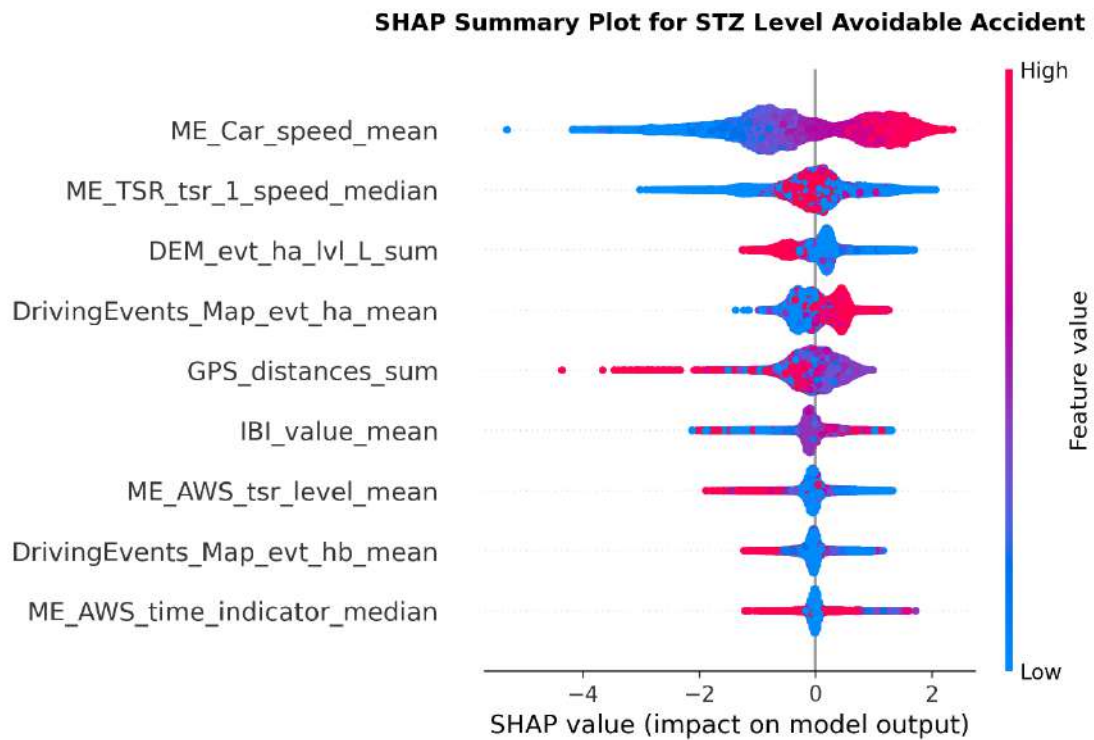
Γράφημα 5.31: Επίδραση μεταβλητών με CatBoost Level Normal, Βέλγιο

Level 1



Γράφημα 5.32: Επίδραση μεταβλητών με CatBoost Level Dangerous, Βέλγιο

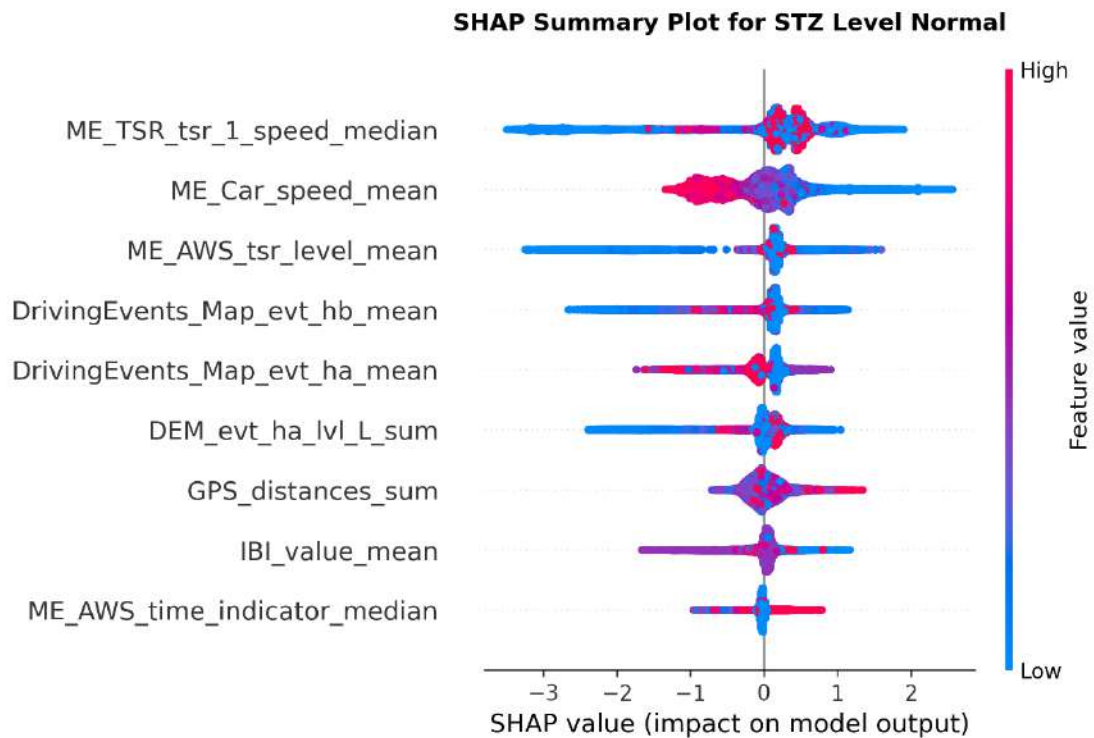
Level 2



Γράφημα 5.33: Επίδραση μεταβλητών με CatBoost Level Avoidable Accident, Βέλγιο

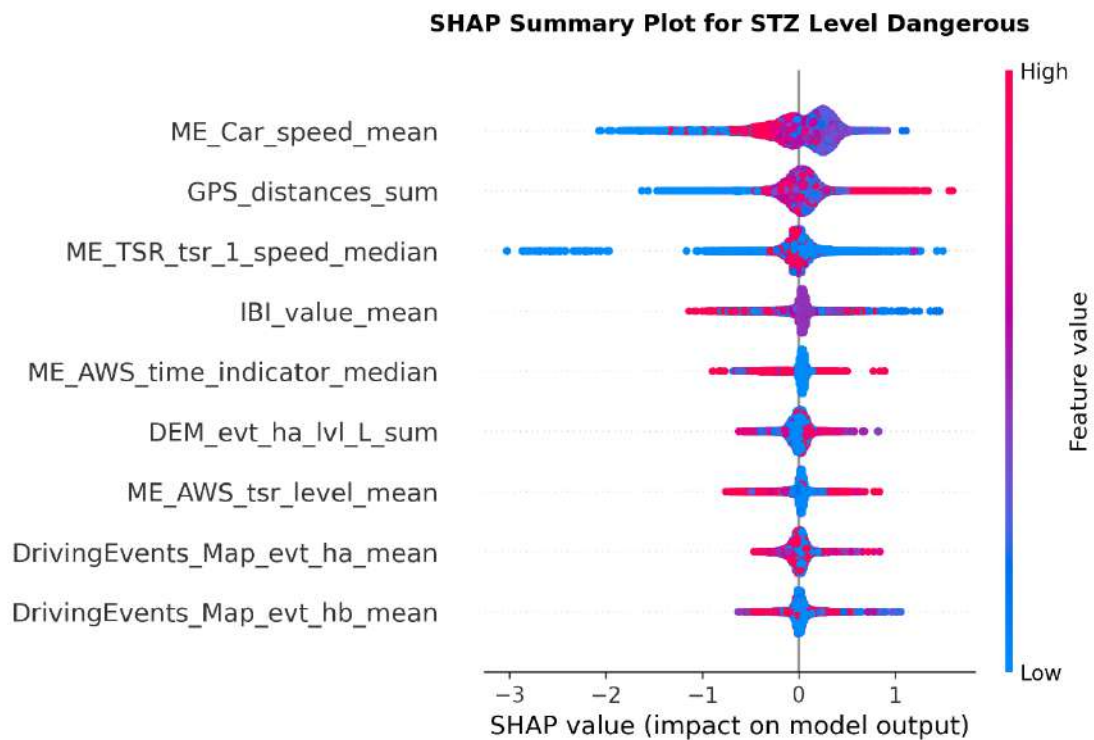
LightGBM

Level 0



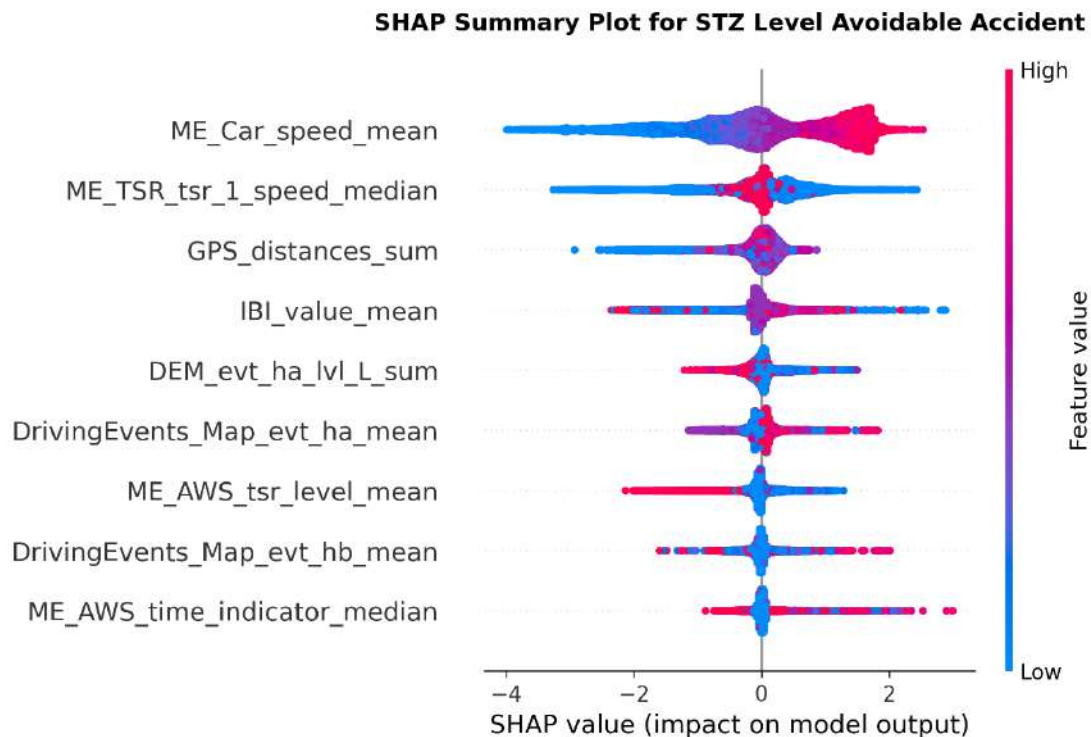
Γράφημα 5.34: Επίδραση μεταβλητών με LightGBM Level Normal, Βέλγιο

Level 1



Γράφημα 5.35: Επίδραση μεταβλητών με LightGBM Level Dangerous, Βέλγιο

Level 2



Γράφημα 5.36: Επίδραση μεταβλητών με LightGBM Level Avoidable Accident, Βέλγιο

Για τα δεδομένα του **Βελγίου**, για το **CatBoost**, προκύπτει, όπως και στην περίπτωση του Ηνωμένου Βασιλείου, πως η σημαντικότερη μεταβλητή είναι η **ME_Car_speed_mean** που ωθεί στην αλλαγή επιπέδου επικίνδυνης οδηγικής συμπεριφοράς. Πιο συγκεκριμένα, στο “Level 0”, καθώς και στο “Level 1”, φαίνεται πως υψηλές τιμές της μεταβλητής (κόκκινο χρώμα) μειώνουν την πιθανότητα να καταταγεί το όχημα στην “STZ Level Normal” και στην “STZ Level Dangerous” αντίστοιχα. Στο “Level 2”, αποδεικνύεται πως υψηλές τιμές της μεταβλητής (κόκκινο χρώμα) ωθούν το όχημα προς την κατάταξη του στο “STZ Level Avoidable Accident”, κάτι αναμενόμενο.

Για το **LightGBM**, αναδεικνύεται επίσης ως η σημαντικότερη μεταβλητή το **ME_Car_speed_mean** που ωθεί στην αλλαγή επιπέδου επικίνδυνης οδηγικής συμπεριφοράς. Εξαιρείται από τα παραπάνω το “Level 0”, στο οποίο χρίζεται ως επιδραστικότερη μεταβλητή το **ME_TSR_tsr_1_speed_median**, η διάμεση τιμή της ταχύτητας που σχετίζεται με το TSR (Traffic Sign Recognition) και η οποία επηρεάζει το αποτέλεσμα παρόμοια με την ταχύτητα του οχήματος.

6.ΣΥΜΠΕΡΑΣΜΑΤΑ

6.1. Σύνοψη αποτελεσμάτων

Στην παρούσα Διπλωματική Εργασία **διερευνάται η επικίνδυνη οδηγική συμπεριφορά μέσω δεδομένων οδήγησης σε πραγματικές συνθήκες**, τα οποία συλλέχθηκαν από το Βέλγιο και το Ηνωμένο Βασίλειο. Πιο συγκεκριμένα, η συμπεριφορά του οδηγού κατατάσσεται σε τρία διαφορετικά επίπεδα επικίνδυνης συμπεριφοράς με βάση τον δείκτη “Ζώνη Ανοχής Ασφαλείας”. Τα δεδομένα προέρχονται από ένα πείραμα φυσική οδήγησης, στο πλαίσιο του ερευνητικού προγράμματος i-DREAMS. Συγκεκριμένα, στη μελέτη συμμετείχαν 42 οδηγοί από το Βέλγιο, δημιουργώντας μια εκτενή βάση δεδομένων με 813 ταξίδια και συνολική διάρκεια 21.412 λεπτών. Αντίστοιχα, για το Ηνωμένο Βασίλειο, το σύνολο δεδομένων περιλάμβανε 54 οδηγούς, καταγράφοντας 3.317 ταξίδια και 58.458 λεπτά οδήγησης.

Βασικό μέρος της μεθοδολογίας αποτελεί η διαδικασία επιλογής χαρακτηριστικών. Στόχος της διαδικασίας είναι η μείωση του αριθμού των μεταβλητών εισόδου με ταυτόχρονη μείωση του υπολογιστικού κόστους του μοντέλου και βελτίωση της προγνωστικής του απόδοσης. Η επιλογή των χαρακτηριστικών γίνεται με γνώμονα την επιρροή της μεταβλητής στην διαδικασία της ταξινόμησης. Σε πρώτη φάση, η σημαντικότητα των μεταβλητών υπολογίζεται **με τον αλγόριθμο “Τυχαίου Δάσους”** (Random Forest Classifier). Αναφέρεται πως, καθώς επιλέγεται να γίνει **ο καθορισμός του επιπέδου ασφαλείας** με την χρήση της μεταβλητής **Headway**, τα διάφορα περιγραφικά στατιστικά στοιχεία του παράγοντα Headway δεν αποτελούν μεταβλητές εισόδου στα μοντέλα.

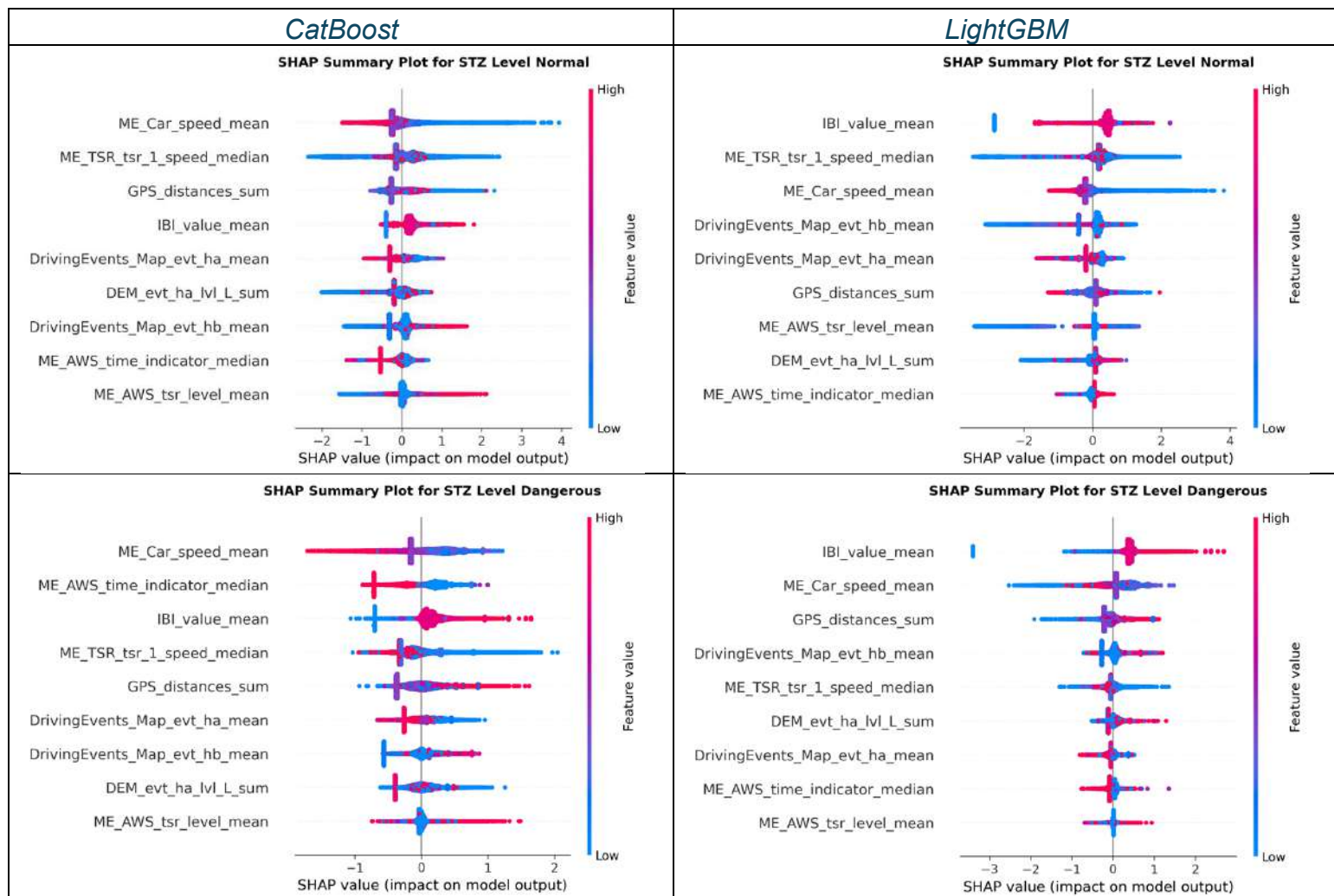
Στη συνέχεια, επιλέγεται ένα σύνολο **εννιά μεταβλητών εισόδου**, κοινό και για τα δύο datasets Βελγίου κι Ηνωμένου Βασιλείου. Για να αντιμετωπιστεί η ανισορροπία των δεδομένων μεταξύ των διαφορετικών κατηγοριών, χρησιμοποιείται η μέθοδος **SMOTE (Synthetic Minority Over-sampling Technique)**. Μετά την εφαρμογή του SMOTE, αναπτύσσονται τέσσερις αλγόριθμοι μηχανικής μάθησης με σκοπό την ταξινόμηση της οδηγικής συμπεριφοράς σε ένα από τα τρία επίπεδα ασφαλείας. Έπειτα, παρουσιάζονται για κάθε μοντέλο ταξινόμησης και για τις δύο χώρες οι μήτρες σύγχυσης για την γραφική αναπαράσταση της επίδοσης τους. Μετά την ολοκλήρωση των παραπάνω, γίνεται η σύγκριση διάφορων μετρικών αξιολόγησης των τεσσάρων μοντέλων.

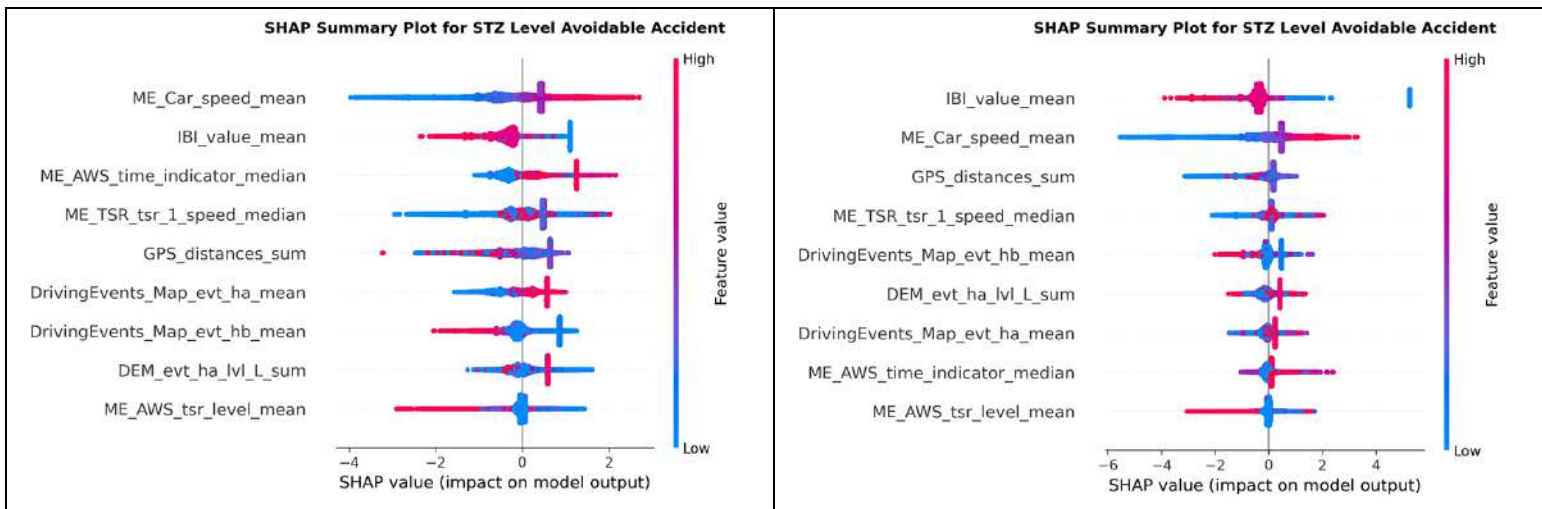
Στη συνέχεια, για να κατανοήσουμε καλύτερα τα μοντέλα μηχανικής μάθησης, εξετάζονται οι τιμές **SHAP** (SHapley Additive exPlanations). Το **SHAP** χρησιμοποιείται για να ερμηνεύσει την επίδραση κάθε χαρακτηριστικού στο μοντέλο, επιτρέποντας τόσο την ανάλυση σε τοπικό όσο και σε συνολικό επίπεδο. Τα μοντέλα **CatBoost** και **LightGBM** εμφανίζουν τη μεγαλύτερη μέση σημαντικότητα SHAP για τα δεδομένα και των δύο χωρών κι έτσι επιλέγονται

για εκ νέου εύρεση της σημαντικότητας των μεταβλητών για κάθε επίπεδο επικίνδυνης οδηγικής συμπεριφοράς. Στους παρακάτω πίνακες, συγκεντρώνονται τα διαγράμματα που παρουσιάζονται στο κεφάλαιο 5, και δείχνουν την **επιρροή των εννιά επιλεγμένων μεταβλητών** στο κάθε επίπεδο επικινδυνότητας της οδηγικής συμπεριφοράς, με τη **μέθοδο SHAP**.

Ηνωμένο Βασίλειο

Πίνακας 6.1: Διαγράμματα μεθόδου SHAP για το Ηνωμένο Βασίλειο



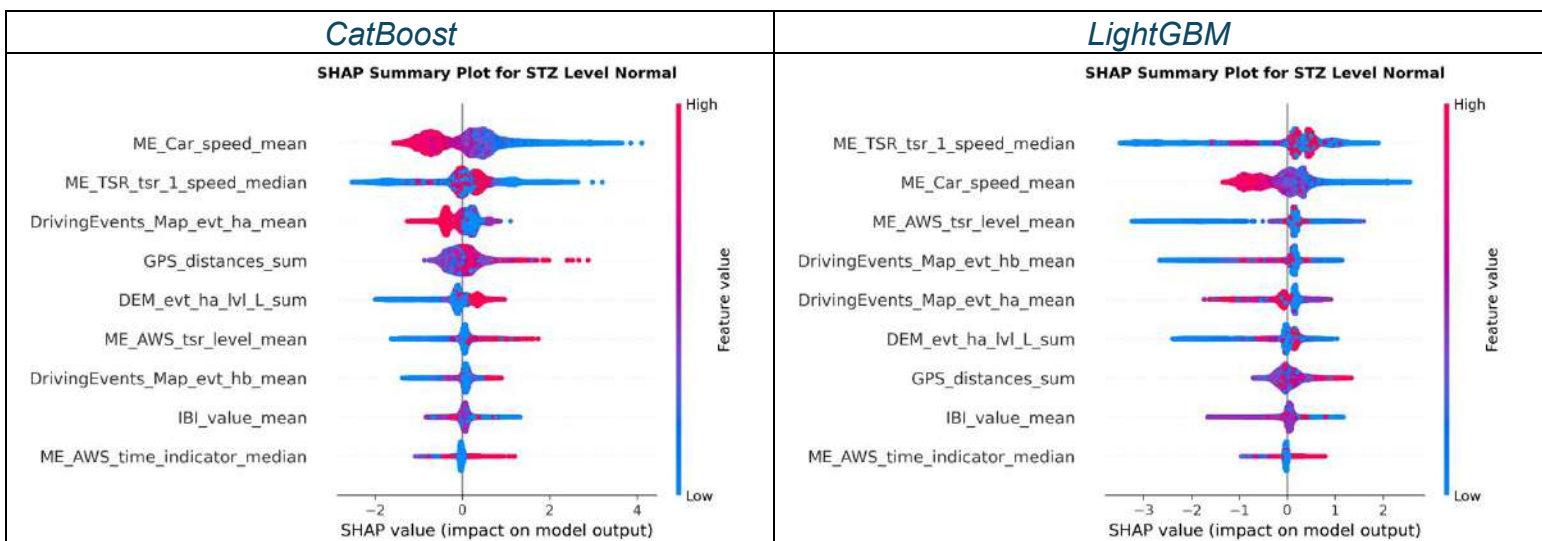


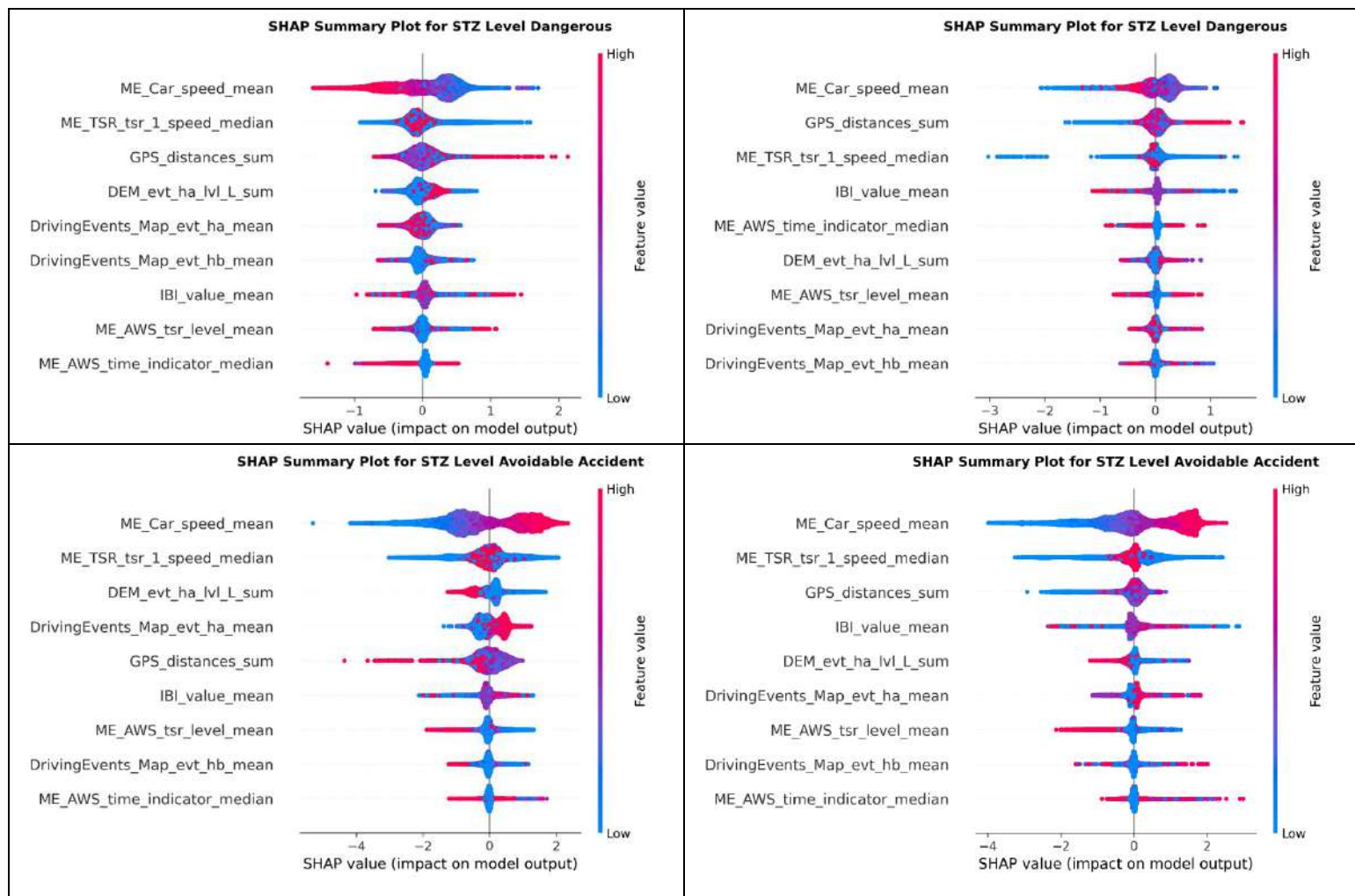
Για τα δεδομένα του **Ηνωμένου Βασιλείου**, προκύπτει για το **CatBoost**, πως η σημαντικότερη μεταβλητή είναι η **ME_Car_speed_mean** που ωθεί στην αλλαγή επιπέδου επικίνδυνης οδηγικής συμπεριφοράς. Πιο συγκεκριμένα, στο “Level 0”, καθώς και στο “Level 1”, φαίνεται πως υψηλές τιμές της μεταβλητής (κόκκινο χρώμα) μειώνουν την πιθανότητα να καταταγεί το όχημα στην “STZ Level Normal” και στην “STZ Level Dangerous” αντίστοιχα. Στο “Level 2”, αποδεικνύεται πως υψηλές τιμές της μεταβλητής (κόκκινο χρώμα) ωθούν το όχημα προς την κατάταξη του στο “STZ Level Avoidable Accident”, κάτι αναμενόμενο.

Για το **LightGBM** αντίστοιχα, αναδεικνύεται ως η σημαντικότερη μεταβλητή η **IBI_value_mean**, που ωθεί στην αλλαγή επιπέδου επικίνδυνης οδηγικής συμπεριφοράς.

Βέλγιο

Πίνακας 6.2: Διαγράμματα μεθόδου SHAP για το Βέλγιο





Για τα δεδομένα του **Βελγίου**, για το **CatBoost**, προκύπτει, όπως και στην περίπτωση του Ηνωμένου Βασιλείου, πως η σημαντικότερη μεταβλητή είναι η **μέση ταχύτητα οχήματος (ME_Car_speed_mean)** που ωθεί στην αλλαγή επιπέδου επικίνδυνης οδηγικής συμπεριφοράς. Πιο συγκεκριμένα, στο “Level 0”, καθώς και στο “Level 1”, φαίνεται πως υψηλές τιμές της μεταβλητής (κόκκινο χρώμα) μειώνουν την πιθανότητα να καταταγεί το όχημα στην “STZ Level Normal” και στην “STZ Level Dangerous” αντίστοιχα. Στο “Level 2”, αποδεικνύεται πως υψηλές τιμές της μεταβλητής (κόκκινο χρώμα) ωθούν το όχημα προς την κατάταξη του στο “STZ Level Avoidable Accident”, κάτι αναμενόμενο.

Για το **LightGBM**, αναδεικνύεται επίσης ως η σημαντικότερη μεταβλητή η **μέση ταχύτητα οχήματος (ME_Car_speed_mean)** που ωθεί στην αλλαγή επιπέδου επικίνδυνης οδηγικής συμπεριφοράς. Εξαιρείται από τα παραπάνω το “Level 0”, στο οποίο κρίζεται ως επιδραστικότερη μεταβλητή η **ME_TSR_tsr_1_speed_median**, δηλαδή η **διάμεση τιμή της ταχύτητας που σχετίζεται με το TSR (Traffic Sign Recognition)** και η οποία επηρεάζει το αποτέλεσμα παρόμοια με την ταχύτητα του οχήματος.

6.2. Σύνοψη συμπερασμάτων

Σύμφωνα με τα αποτελέσματα που προκύπτουν από την εφαρμογή της μεθοδολογίας, διατυπώνονται ορισμένα συμπεράσματα που συνδέονται άμεσα με τον στόχο της Διπλωματικής Εργασίας:

- Η εφαρμογή της μεθόδου **SMOTE** πράγματι αντιμετωπίζει το φαινόμενο της ανισορροπίας δεδομένων μεταξύ των διαφορετικών επιπέδων, κάτι που αποδεικνύεται κι από το ικανοποιητικό επίπεδο των μετρικών αξιολόγησης των τεσσάρων διαφορετικών μοντέλων ταξινόμησης
- Από τα τέσσερα μοντέλα ταξινόμησης που αναπτύσσονται, το μοντέλο ταξινόμησης **Multi-Layer Perceptron** (MLP) φαίνεται το πιο αδύναμο μοντέλο για την πρόβλεψη του επιπέδου ασφαλείας, τόσο μέσω των μητρών σύγχυσης όσο και μέσω των μετρικών αξιολόγησης των μοντέλων
- Από τα τέσσερα μοντέλα ταξινόμησης που αναπτύσσονται, τα μοντέλα **CatBoost** και **LightGBM** παρουσιάζουν ικανοποιητικές μετρικές αξιολόγησης καθώς και την υψηλότερη μέση σημαντικότητα SHAP
- Η **μέση ταχύτητα οχήματος** υπολογίζεται ως η σημαντικότερη μεταβλητή από τον υπολογισμό σημαντικότητας μεταβλητών με τον αλγόριθμο “**Τυχαίου Δάσους**”, ενώ το ίδιο αποτέλεσμα εξάγεται κι από τη μέθοδο SHAP για όλες τις περιπτώσεις πλην του μοντέλου **LightGBM**, για το Ηνωμένο Βασίλειο, όπου κατατάσσεται δεύτερη μετά **το χρονικό διάστημα μεταξύ διαδοχικών καρδιακών παλμών**
- **Το χρονικό διάστημα μεταξύ διαδοχικών καρδιακών παλμών** εμφανίζεται ως η πιο σημαντική μεταβλητή του μοντέλου LightGBM για το Ηνωμένο Βασίλειο με τη μέθοδο SHAP, ενώ φαίνεται πως υψηλές τιμές της μεταβλητής, ένδειξη χαλάρωσης-υποαπασχόλησης του οδηγού, ενδέχεται να ωθήσουν την οδηγική συμπεριφορά στην κατάταξη της ως επικίνδυνη και αρκετά σπανιότερα στο επίπεδο αποφυγής ατυχήματος
- Η σχετικά υψηλή σημαντικότητα των μεταβλητών **απότομων επιταχύνσεων** κι **απότομων φρεναρισμάτων** καταδεικνύει πως τα απότομα περιστατικά οδήγησης επηρεάζουν σημαντικά την κατάταξη της οδηγικής συμπεριφοράς ως επικίνδυνη.
- Η υψηλή επιρροή της **συνολικής απόστασης ταξιδιού** υποδεικνύει ότι οι αποστάσεις που καλύπτονται έχουν επίσης σημαντική συνεισφορά στον τρόπο οδήγησης. Οι αποστάσεις που καλύπτει ένας οδηγός επηρεάζουν την επικίνδυνη οδήγηση μέσω της κόπωσης, της εξοικείωσης με τη διαδρομή, της ταχύτητας και της ψυχολογικής του κατάστασης
- Τα αποτελέσματα των αναλύσεων δείχνουν ότι οι δύο χώρες παρουσιάζουν παρόμοια πρότυπα οδήγησης. Η **μέση ταχύτητα του οχήματος** αποτελεί πολύ σημαντική μεταβλητή και για τις δύο χώρες, όπως και η **διάμεση**

ταχύτητα κατά την εμφάνιση της πρώτης πινακίδας ορίου ταχύτητας και η συνολική απόσταση του ταξιδιού

- Η **διάμεση τιμή της ταχύτητας του οχήματος όταν εντοπίζεται το πρώτο σήμα ορίου ταχύτητας** από το σύστημα αναγνώρισης σημάτων κυκλοφορίας εμφανίζει υψηλή σημαντικότητα τόσο για το Ηνωμένο Βασίλειο, όσο και για το Βέλγιο
- Μεγάλη διαφορά παρατηρείται στη σημαντικότητα **του χρονικού διαστήματος μεταξύ διαδοχικών καρδιακών παλμών**. Πιο συγκεκριμένα, αποδεικνύεται μία από τις πιο σημαντικές μεταβλητές για το Ηνωμένο Βασίλειο σε αντίθεση με το Βέλγιο. Αυτό μας δείχνει ενδεχομένως πως η ψυχική κατάσταση των οδηγών επηρεάζει περισσότερο την οδηγική τους συμπεριφορά στο Ηνωμένο Βασίλειο σε σχέση με το Βέλγιο. Ψυχικές καταστάσεις όπως στρες, ταχυκαρδία ή αντίστοιχα βραδυκαρδία, χαλάρωση είναι εκείνες που μπορεί να επηρεάσουν τη συμπεριφορά του οδηγού

6.3. Προτάσεις για αξιοποίηση των αποτελεσμάτων

Λαμβάνοντας υπόψη τα αποτελέσματα και τα συμπεράσματα που προέκυψαν από την παρούσα διπλωματική εργασία, παρουσιάζεται στη συνέχεια ένα σύνολο προτάσεων αξιοποίησης των ευρημάτων. Οι προτάσεις αυτές αποσκοπούν στη **βελτίωση** της κατανόησης της επίδρασης διαφόρων παραγόντων στην οδική ασφάλεια, καθώς και στην **προώθηση** της έρευνας στον τομέα των **Ευφυών Μεταφορικών Συστημάτων (ITS)**.

- Χρήση **συνδυαστικών μοντέλων ταξινόμησης** για την αξιολόγηση του επιπέδου ασφάλειας των οδηγών σε πραγματικές συνθήκες. Τα αποτελέσματα των τεσσάρων συνδυαστικών αλγορίθμων ταξινόμησης υποδεικνύουν ότι μπορούν να προσφέρουν αξιόπιστες εκτιμήσεις, καθιστώντας τους κατάλληλους για περαιτέρω ανάλυση της οδηγικής συμπεριφοράς.
- **Διερεύνηση των κυριότερων μεταβλητών** που επηρεάζουν την αναγνώριση επικίνδυνης οδηγικής συμπεριφοράς. Η εμβάθυνση στη μελέτη αυτών των παραγόντων μπορεί να συμβάλει ουσιαστικά στη βελτίωση των προηγμένων συστημάτων υποστήριξης οδηγού, ενισχύοντας έτσι τις προσπάθειες τόσο της επιστημονικής κοινότητας όσο και της αυτοκινητοβιομηχανίας.
- **Ανάπτυξη συστήματος ανίχνευσης** της ‘Ζώνης Ανοχής Ασφαλείας’ του οδηγού **σε πραγματικό χρόνο**. Ένα τέτοιο σύστημα θα επιτρέπει στους οδηγούς να παρακολουθούν τη συμπεριφορά τους κατά την οδήγηση και να λαμβάνουν έγκαιρη ενημέρωση για τυχόν μη ασφαλείς πρακτικές, δίνοντάς τους τη δυνατότητα να προσαρμόσουν την οδήγησή τους αναλόγως.

- **Δημιουργία εφαρμογής** για έξυπνα κινητά τηλέφωνα που θα καταγράφει τα δεδομένα οδήγησης και θα προβλέπει τη συνολική διάρκεια παραμονής του οδηγού σε κάθε επίπεδο της 'Ζώνης Ανοχής Ασφαλείας'. Έτσι, ο οδηγός θα έχει μια ολοκληρωμένη εικόνα της οδηγικής του συμπεριφοράς και τη δυνατότητα να τη βελτιώσει μέσω κατάλληλων παρεμβάσεων.
- **Εξοπλισμός των οχημάτων με συστήματα πρόβλεψης ατυχημάτων**, τα οποία, αξιοποιώντας τον αλγόριθμο, θα μπορούν να εντοπίζουν πιθανές επικίνδυνες καταστάσεις και να προειδοποιούν τον οδηγό μέσω ειδοποιήσεων, δίνοντάς του τη δυνατότητα να αντιδράσει έγκαιρα και να αποτρέψει πιθανά ατυχήματα.

6.4. Προτάσεις για μελλοντική έρευνα

Οι σύγχρονες τεχνικές επεξεργασίας και ανάλυσης δεδομένων χρησιμοποιούνται όλο και περισσότερο στον τομέα της οδικής ασφάλειας. Η μελέτη της οδηγικής συμπεριφοράς μέσω μεθόδων μηχανικής μάθησης και βαθιάς εκμάθησης αποτελεί ένα πεδίο έντονου ερευνητικού ενδιαφέροντος. Στην παρούσα μελέτη αναδείχθηκαν ορισμένα ζητήματα, γεγονός που οδήγησε τους ερευνητές στην πρόταση περαιτέρω διερεύνησης επιπρόσθετων παραγόντων και μεθόδων.

Η συγκεκριμένη Διπλωματική Εργασία είχε ως στόχο να καλύψει το κενό που εντοπίστηκε στη βιβλιογραφία, εξετάζοντας διάφορες τεχνικές μηχανικής και βαθιάς μάθησης, ενώ παράλληλα επιδιώκει να αποτελέσει βάση για μελλοντικές συγκριτικές αναλύσεις. Ωστόσο, κατά τη διαδικασία ανάπτυξης της μεθοδολογίας και αξιολόγησης των αποτελεσμάτων, εντοπίστηκαν ορισμένες αδυναμίες, οι οποίες μπορούν να ληφθούν υπόψη σε επόμενες έρευνες.

Ακολουθούν προτάσεις για περαιτέρω μελέτη, οι οποίες μπορούν να συμβάλουν στην εμπάθυνση της έρευνας, τη βελτίωση της κατανόησης του αντικειμένου και την αντιμετώπιση των υφιστάμενων περιορισμών:

- **Αξιοποίηση μεγαλύτερου όγκου δεδομένων** για τη βελτίωση της προγνωστικής ικανότητας των μοντέλων ταξινόμησης. Η αύξηση των δεδομένων μπορεί να μειώσει την πιθανότητα σφάλματος και να ενισχύσει την ακρίβεια των αποτελεσμάτων.
- Ανάπτυξη **εναλλακτικών μεθόδων ανάλυσης της σημαντικότητας των μεταβλητών (feature importance)**. Η εις βάθος διερεύνηση του ρόλου κάθε μεταβλητής θα μπορούσε να συμβάλει στην ακριβέστερη κατανόηση της σχέσης τους με την ταξινόμηση του επιπέδου ασφαλείας των οδηγών.
- Εξέταση **πρόσθετων παραγόντων που επηρεάζουν** την οδηγική συμπεριφορά. Βάσει της παρούσας μελέτης και της ανασκόπησης της σχετικής βιβλιογραφίας, παράγοντες όπως οι καιρικές συνθήκες, τα χαρακτηριστικά της οδού, οι αντιλήψεις των οδηγών για την επικινδυνότητα,

η γεωμετρία των οδών, οι κανόνες κυκλοφορίας και οι πολιτισμικές διαφοροποιήσεις μπορούν να επηρεάσουν την οδική ασφάλεια και χρήζουν περαιτέρω έρευνας.

- Ανάπτυξη εξειδικευμένων μοντέλων ταξινόμησης ακολουθίας (sequence classification) για την **πρόβλεψη του επιπέδου** 'Ζώνης Ανοχής Ασφαλείας' στο οποίο θα βρίσκεται ο οδηγός **στο επόμενο χρονικό διάστημα** των 30 δευτερολέπτων. Ένα τέτοιο μοντέλο θα μπορούσε να παρέχει πιο έγκαιρη και ακριβή πληροφόρηση για την οδηγική συμπεριφορά, συμβάλλοντας στην πρόληψη επικίνδυνων καταστάσεων.

BIBΛΙΟΓΡΑΦΙΑ

1. Kareem A. Review of global menace of road accidents with special reference to malaysia- a social perspective. *Malays J Med Sci.* 2003 Jul;10(2):31-9. PMID: 23386795; PMCID: PMC3561885. <https://pmc.ncbi.nlm.nih.gov/articles/PMC3561885/>
2. Arumugam, S., Bhargavi, R. A survey on driving behavior analysis in usage based insurance using big data. *J Big Data* **6**, 86 (2019). <https://doi.org/10.1186/s40537-019-0249-5>
3. Taheri, S.M., Matsushita, K. and Sasaki, M. (2017) Virtual Reality Driving Simulation for Measuring Driver Behavior and Characteristics. *Journal of Transportation Technologies*, **7**, 123-132. <https://doi.org/10.4236/jtts.2017.72009>
4. Zolali, Mehdi, Mirbaha, Babak, Layegh, Maziyar, Behnood, Hamid Reza, A Behavioral Model of Drivers' Mean Speed Influenced by Weather Conditions, Road Geometry, and Driver Characteristics Using a Driving Simulator Study, *Advances in Civil Engineering*, 2021, 5542905, 18 pages, 2021. <https://doi.org/10.1155/2021/5542905>
5. Diogo Reis, Ricardo Tomás, Margarida C. Coelho, Eloisa Macedo, Correlating driving behavior with safety performance: from the heart to the pedals with a driving simulator, *Transportation Research Procedia*, Volume 69, 2023, Pages 217-224, ISSN 2352-1465, <https://doi.org/10.1016/j.trpro.2023.02.165>
6. Harpreet Singh, Ankit Kathuria, Analyzing driver behavior under naturalistic driving conditions: A review, *Accident Analysis & Prevention*, Volume 150, 2021, 105908, ISSN 0001-4575, <https://doi.org/10.1016/j.aap.2020.105908>
7. Xuesong Wang, Rongjiao Xu, Siyang Zhang, Yifan Zhuang, Yinhai Wang, Driver distraction detection based on vehicle dynamics using naturalistic driving data, *Transportation Research Part C: Emerging Technologies*, Volume 136, 2022, 103561, ISSN 0968-090X, <https://doi.org/10.1016/j.trc.2022.103561>
8. N. Lyu, Y. Wang, C. Wu, L. Peng and A. F. Thomas, "Using naturalistic driving data to identify driving style based on longitudinal driving operation conditions," in *Journal of Intelligent and Connected Vehicles*, vol. 5, no. 1, pp. 17-35, 2022, doi: 10.1108/JICV-07-2021-0008
9. Osama A. Osman, Mustafa Hajji, Sogand Karbalaieali, Sherif Ishak, A hierarchical machine learning classification approach for secondary task identification from observed driving behavior data, *Accident Analysis & Prevention*, Volume 123, 2019, Pages 274-281, ISSN 0001-4575, <https://doi.org/10.1016/j.aap.2018.12.005>
10. Dewei Yi, Jinya Su, Cunjia Liu, Mohammed Quddus, Wen-Hua Chen, A machine learning based personalized system for driving state

- recognition, Transportation Research Part C: Emerging Technologies, Volume 105, 2019, Pages 241-261, ISSN 0968-090X, <https://doi.org/10.1016/j.trc.2019.05.042>
11. Ghandour R, Potams AJ, Boulkaibet I, Neji B, Al Barakeh Z. Driver Behavior Classification System Analysis Using Machine Learning Methods. *Applied Sciences*. 2021; 11(22):10562. <https://doi.org/10.3390/app112210562>
 12. Qiangqiang Shangguan, Ting Fu, Junhua Wang, Tianyang Luo, Shou'en Fang, An integrated methodology for real-time driving risk status prediction using naturalistic driving data, *Accident Analysis & Prevention*, Volume 156, 2021, 106122, ISSN 0001-4575, <https://doi.org/10.1016/j.aap.2021.106122>
 13. Shi, J., Xiao, Y., Tao, L., & Atchley, P. (2017). Factors causing aberrant driving behaviors: A model of problem drivers in China. *Journal of Transportation Safety & Security*, 10(4), 288–302. <https://doi.org/10.1080/19439962.2016.1263706>
 14. 44th National Systems Conference on Systems for Sustainable Healthcare Habitats NSC-2021 May 22-23, 2021. Dayalbagh Educational Institute in collaboration with Systems Society of India, https://www.researchgate.net/profile/Shiwani-4/publication/388176060_Historical_Perspective_of_Artificial_Intelligence_based_expert_system/links/678d39b382501639f5f927fd/Historical-Perspective-of-Artificial-Intelligence-based-expert-system.pdf#page=139
 15. Best Ways To Handle Imbalanced Data In Machine Learning, [Jaiganesh Nagidi](https://dataaspirant.com/handle-imbalanced-data-machine-learning/), <https://dataaspirant.com/handle-imbalanced-data-machine-learning/>
 16. M. Pérez-Ortiz, P. A. Gutiérrez, P. Tino and C. Hervás-Martínez, "Oversampling the Minority Class in the Feature Space," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 9, pp. 1947-1961, Sept. 2016, doi: 10.1109/TNNLS.2015.2461436.
 17. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). **SMOTE: Synthetic Minority Over-sampling Technique**. *Journal of Artificial Intelligence Research*, 16, 321–357. DOI: 10.1613/jair.953
 18. Jiayao Chen, Hongwei Huang, Anthony G. Cohn, Dongming Zhang, Mingliang Zhou, Machine learning-based classification of rock discontinuity trace: SMOTE oversampling integrated with GBT ensemble learning, *International Journal of Mining Science and Technology*, Volume 32, Issue 2, 2022, <https://doi.org/10.1016/j.ijmst.2021.08.004>
 19. Alruwais N, Zakariah M. Evaluating Student Knowledge Assessment Using Machine Learning Techniques. *Sustainability*. 2023; 15(7):6229. <https://doi.org/10.3390/su15076229>
 20. ISSN: 2277-3754 ISO 9001:2008 Certified International Journal of Engineering and Innovative Technology (IJEIT) Volume 4, Issue 3, September 2014,

- https://faculty.ksu.edu.sa/sites/default/files/classification_of_large_dataset_s_using_random.pdf
21. Light GBM Light and Powerful Gradient Boost Algorithm, <https://medium.com/@turkishtechology/light-gbm-light-and-powerful-gradient-boost-algorithm-eaa1e804eca8>
 22. Guolin Ke¹, Qi Meng², Thomas Finley³, Taifeng Wang¹, Wei Chen¹, Weidong Ma¹, Qiwei Ye¹, Tie-Yan Liu¹ ¹Microsoft Research ²Peking University ³Microsoft Redmond, <https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fd848669bdd9eb6b76fa-Abstract.html>
 23. László GÖCS, Zsolt Csaba JOHANYÁK, 1. John von Neumann University, GAMF Faculty of Engineering and Computer Science, Department of Information Technology, Kecskemét, HUNGARY, 2023, <https://annals.fih.upt.ro/pdf-full/2023/ANNALS-2023-3-01.pdf>
 24. Science Direct, <https://www.sciencedirect.com/topics/computer-science/multilayer-perceptron>
 25. What is a Confusion Matrix? Understand the 4 Key Metric of its Interpretation, <https://datasciencedojo.com/blog/confusion-matrix/>
 26. Εικόνα Εξωφύλλου, <https://www.morethanshipping.com/an-overview-of-autonomous-vehicles/>
 27. Εικόνα 1.1, NTUA Road Safety Observatory <https://www.nrso.ntua.gr/slight-increase-in-road-fatalities-in-2024-greece/>
 28. Διάγραμμα 1.1, Road Safety pdf, School of Civil Engineering NTUA 2024, <file:///C:/Users/%CE%93%CE%B9%CE%AC%CE%BD%CE%BD%CE%B7%CF%82/Downloads/DKOA-RoadSafety-Introduction.pdf>
 29. Πίνακας 1.1, Road Safety pdf, School of Civil Engineering NTUA 2024, <file:///C:/Users/%CE%93%CE%B9%CE%AC%CE%BD%CE%BD%CE%B7%CF%82/Downloads/DKOA-RoadSafety-Introduction.pdf>
 30. Εικόνα 1.2, <https://idreamsproject.eu/wp/>
 31. Πίνακας 1.2, <https://idreamsproject.eu/wp/>
 32. Γράφημα 3.1, <https://dataaspirant.com/handle-imbalanced-data-machine-learning/>
 33. Γράφημα 3.2, https://www.researchgate.net/figure/Process-of-Synthetic-Minority-Oversampling-Technique-SMOTE-Algorithm-27_fig1_368993605
 34. Γράφημα 3.3, <https://www.mdpi.com/2071-1050/15/7/6229>
 35. Γράφημα 3.4, classification - Random Forests, https://pages.cms.hu-berlin.de/EOL/geo_rs/S08_Image_classification2.html
 36. Γράφημα 3.5, <https://lightgbm.readthedocs.io/en/latest/Features.html>
 37. Γράφημα 3.6, https://www.researchgate.net/figure/The-flow-diagram-of-the-CatBoost-model_fig3_370695897?_cf_chl_tk=tyjvMqfeM7x4LJ3kLpGVrnUCoaa1x6

[zylExWbiDMS6o-1740249087-1.0.1.1-
kEV.mkSy5qEOIXw2DaOXioGH.eGM89NtLzZfSG603qA](https://zenodo.org/record/1740249/files/1.0.1.1-kEV.mkSy5qEOIXw2DaOXioGH.eGM89NtLzZfSG603qA)

38. Γράφημα 3.7, <https://becominghuman.ai/multi-layer-perceptron-mlp-models-on-real-world-banking-data-f6dd3d7e998f>
39. Γράφημα 3.8, https://www.researchgate.net/figure/Confusion-matrix-for-multi-class-classification-The-confusion-matrix-of-a-classification_fig7_314116591
40. Πίνακας 4.1, <https://idreamsproject.eu/wp/>
41. Εικόνα 4.1, Bartholomew, M. C. (2022). *Track and Screen Evaluation of the Mobileye ADAS Camera System* [Master's thesis, Ohio State University]. OhioLINK Electronic Theses and Dissertations Center. http://rave.ohiolink.edu/etdc/view?acc_num=osu1648740206341633