



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Πολιτικών Μηχανικών
Τομέας Μεταφορών και Συγκοινωνιακής Υποδομής

**Διερεύνηση της επιρροής της χρήσης κινητού τηλεφώνου
στη συμπεριφορά του οδηγού
με τεχνικές μηχανικής μάθησης
ανισόρροπων δεδομένων**

Κωνσταντίνος-Ειρηναίος Κασελούρης

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

Αθήνα, Μάρτιος 2023



Στόχος Διπλωματικής Εργασίας

- Η διερεύνηση της επιρροής της **χρήσης κινητού τηλεφώνου στη συμπεριφορά του οδηγού**.
- Για την επίτευξη του στόχου πραγματοποιείται:
 - Ανάπτυξη αλγορίθμων για την εύρεση του **βαθμού επιρροής** των οδικών δεδομένων στην επικίνδυνη συμπεριφορά.
 - Ανάπτυξη αλγορίθμων **ταξινόμησης**, για τον εντοπισμό των επιπέδων ασφαλείας.
 - Ανάπτυξη μοντέλων **παλινδρόμησης** για την πρόβλεψη της διάρκειας χρήσης κινητού τηλεφώνου.



Βασικά Στάδια Διπλωματικής Εργασίας

Καθορισμός Στόχου

Βιβλιογραφική Ανασκόπηση

Θεωρητικό υπόβαθρο

Συλλογή Στοιχείων

Περιγραφή και Επεξεργασία
Στοιχείων

Ανάπτυξη Μοντέλων
Μηχανικής Μάθησης

Περιγραφή
αποτελεσμάτων

Συμπεράσματα

Προτάσεις για
Περαιτέρω
Έρευνα



Βιβλιογραφική Ανασκόπηση (1/3)

Συναφείς **έρευνες** και **μεθοδολογίες** :

- Η χρήση κινητού τηλεφώνου αποτελεί **πρωταρχικό παράγοντα** απόσπασης της προσοχής του οδηγού.
- Κρισιμότερη η **πληκτρολόγηση μηνυμάτων** σε σχέση με άλλες δραστηριότητες χρήσης κινητού τηλεφώνου.
- Η χρήση κινητού τηλεφώνου προκαλεί **μείωση** στις **ταχύτητες** διαδρομής και **αύξηση** του **χρόνου αντίδρασης-αντίληψης**.
- Η χρήση κινητού τηλεφώνου μπορεί να αυξήσει την **πλευρική απόκλιση** και τον αριθμό **παρ' ολίγον ατυχημάτων**.



Βιβλιογραφική Ανασκόπηση (2/3)

Συναφείς **έρευνες** και **μεθοδολογίες** :

- Συλλογή στοιχείων από **ερωτηματολόγια**, πείραμα σε **προσομοιωτή** και **πραγματικές** συνθήκες οδήγησης.
- Εφαρμογή μεθοδολογίας πολυεπίπεδης **ταξινόμησης**, **παλινδρόμησης** και **βαθιάς μάθησης** με νευρωνικά δίκτυα.
- Εφαρμογή μεθοδολογιών **επαναδειγματοληψίας**.



Βιβλιογραφική Ανασκόπηση (3/3)

Ελλείψεις και **περιορισμοί** προγενέστερων ερευνών:

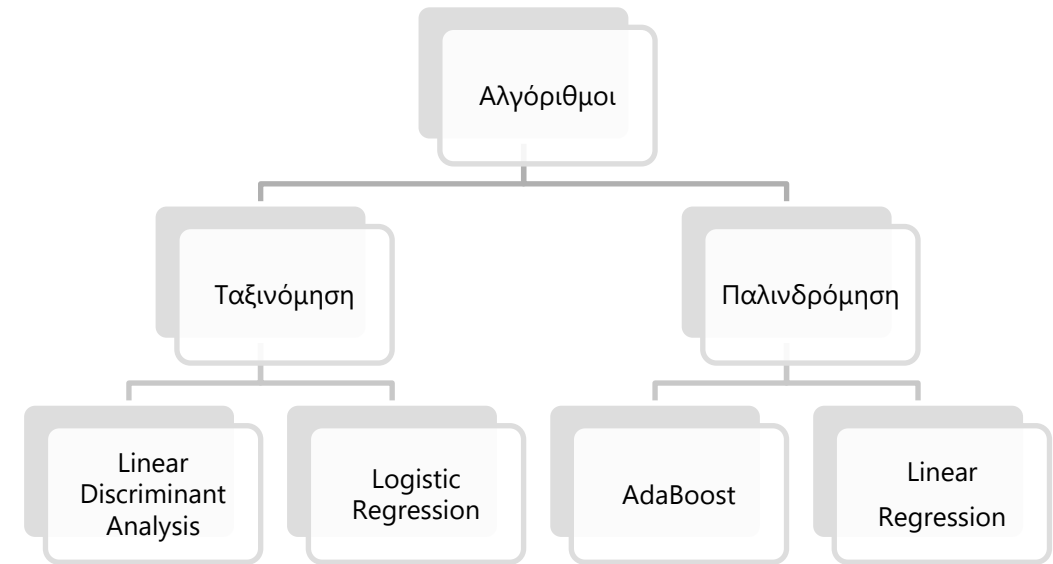
- Εξέταση **περιορισμένου** αριθμού τεχνικών αντιμετώπισης της άνισης κατανομής των δειγμάτων στα επίπεδα επικινδυνότητας.
- **Απουσία** κατάστασης του οδηγού (ψυχολογική ή κόπωση) και κυκλοφοριακών συνθηκών.
- **Αποκλίσεις** προσομοίωσης από πραγματικές συνθήκες οδήγησης.
- **Υποκειμενικότητα** και **ασάφεια** ερωτηματολογίων.
- **Έλλειψη** στοιχείων σχετικά με τον τύπο, την γεωμετρία της οδού και χαρακτηριστικών του οχήματος.



Θεωρητικό Υπόβαθρο (1/2)

Ανάπτυξη μοντέλων μηχανικής εκμάθησης

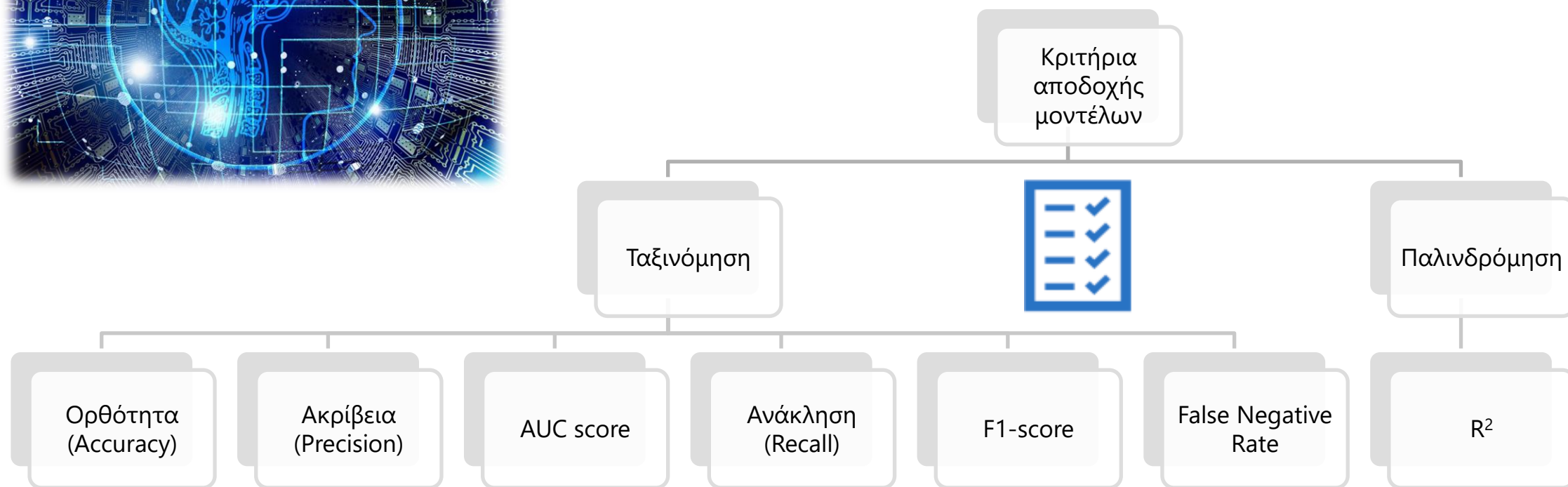
- **Ταξινόμησης:** Για την πρόβλεψη της χρήσης κινητού τηλεφώνου.
- **Παλινδρόμησης:** Για την πρόβλεψη της διάρκειας χρήσης κινητού τηλεφώνου.



Αξιολόγηση σημαντικότητας χαρακτηριστικών:

- Συντελεστές **συσχέτισης** Pearson.
- **Σημαντικότητα** μεταβλητών στις διαδικασίες της ταξινόμησης και παλινδρόμησης.

Θεωρητικό Υπόβαθρο (2/2)



Συλλογή Στοιχείων

- Τα δεδομένα συλλέχθηκαν από τις βάσεις δεδομένων της εταιρείας **OSeven**.
- Δεδομένα **τηλεματικής** μέσω ειδικά σχεδιασμένης εφαρμογής smartphone.
- **356.162** διαφορετικές οδικές διαδρομές.
- Συνολικά **13 από τους 21 δείκτες** αξιοποιήθηκαν για περαιτέρω επεξεργασία.



User
Consent



Collect
Raw Data



Data
Processing

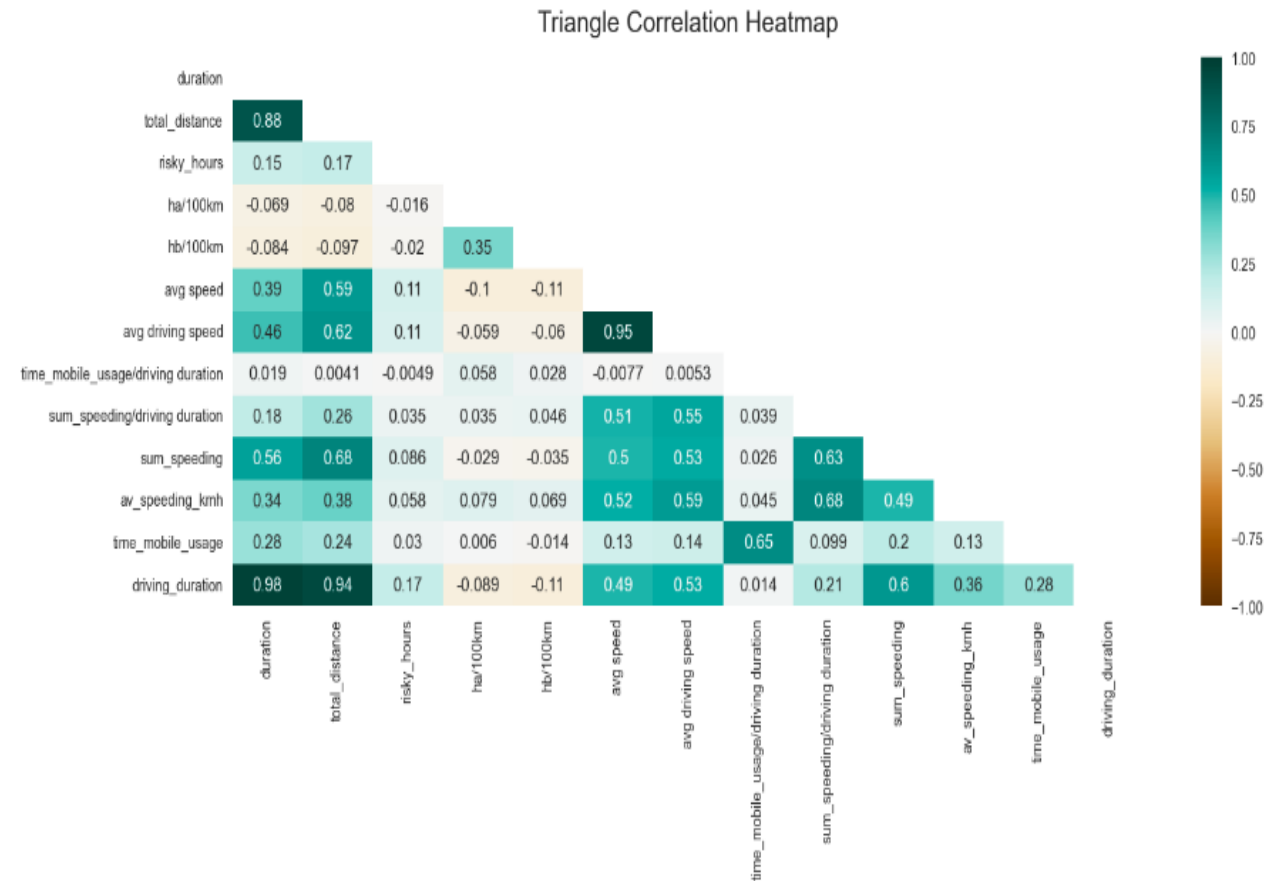


Scores &
Analytics



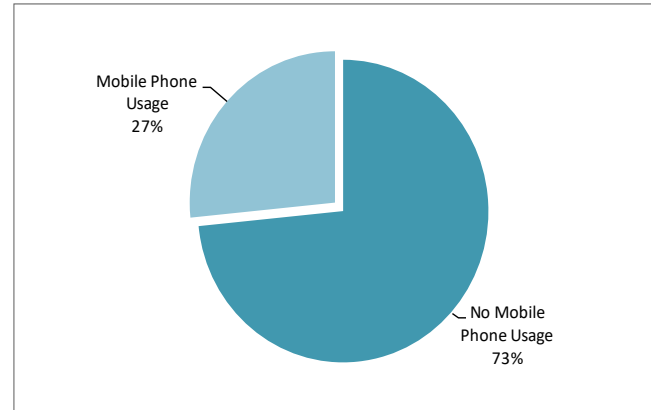
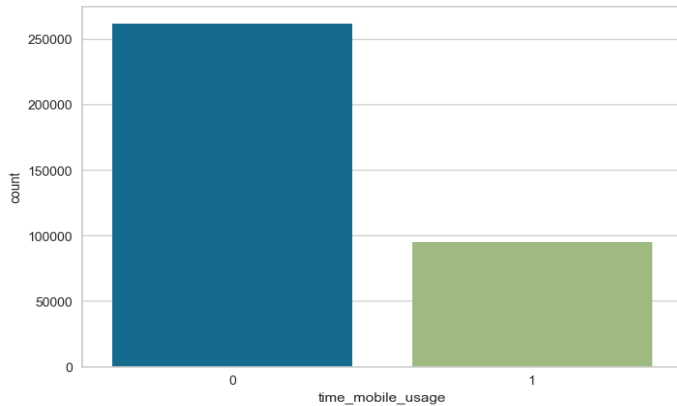
Συσχέτιση Pearson

- **Υψηλή συσχέτιση** μεταξύ των διαφορετικών περιγραφικών στατιστικών της ίδιας μεταβλητής.
- **Θετική συσχέτιση** μεταξύ χρήσης κινητού τηλεφώνου εν ώρα οδήγησης και διάρκειας οδήγησης.
- **Ισχυρή συσχέτιση** μεταξύ συνολικής διανυθείσας απόστασης και συνολικής διάρκειας οδήγησης είτε με είτε χωρίς στάσεις.
- **Μικρή αρνητική συσχέτιση** μεταξύ των ανεξάρτητων μεταβλητών, γεγονός που αποδεικνύει ότι η αύξηση μίας ανεξάρτητης μεταβλητής δεν μειώνει το μέγεθος μίας άλλης.



Προεπεξεργασία για Ταξινόμηση

➤ Δημιουργία **δυαδικής** μεταβλητής



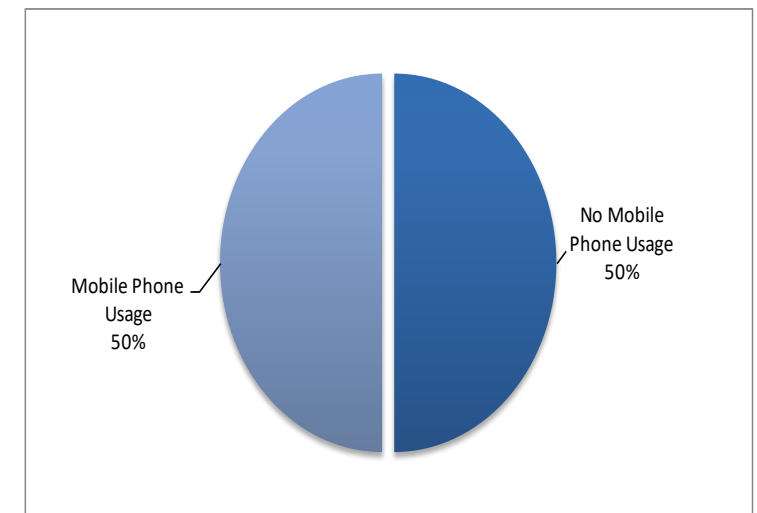
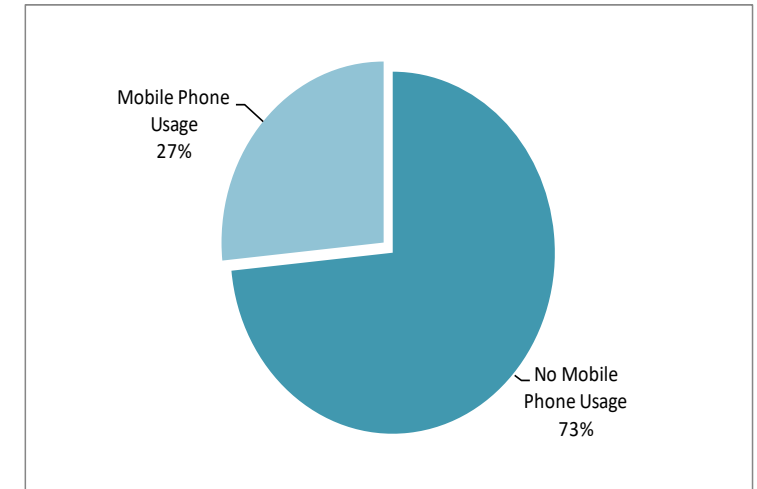
➤ Δύο επίπεδα ταξινόμησης:

- **Επικίνδυνη** (Χρήση κινητού-Τιμή 1)
- **Μη** επικίνδυνη συμπεριφορά (Μη χρήση κινητού-Τιμή 0)



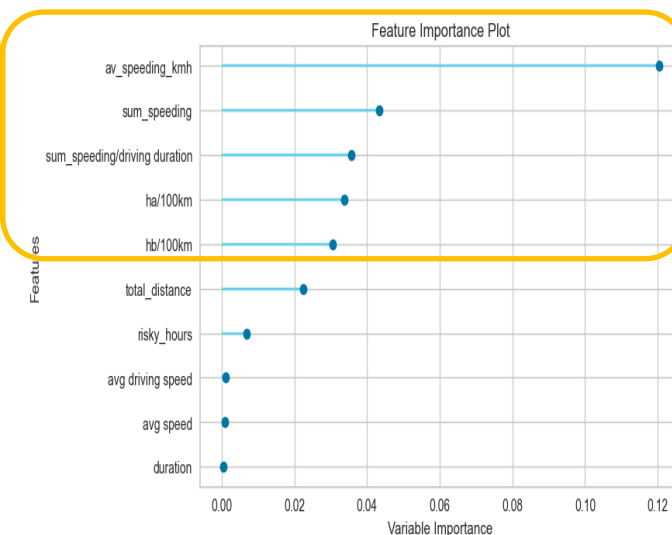
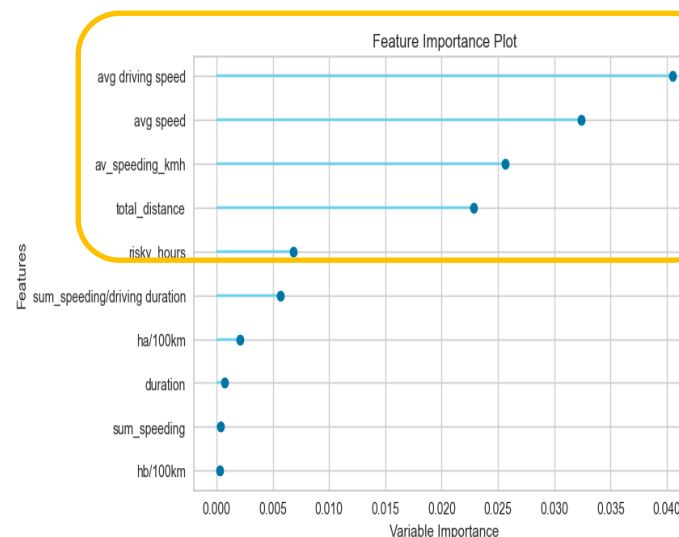
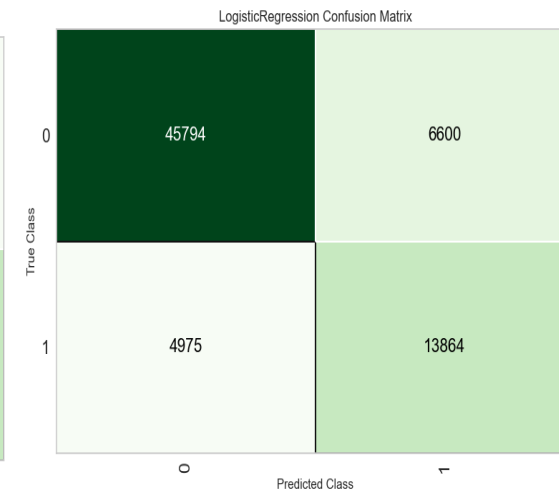
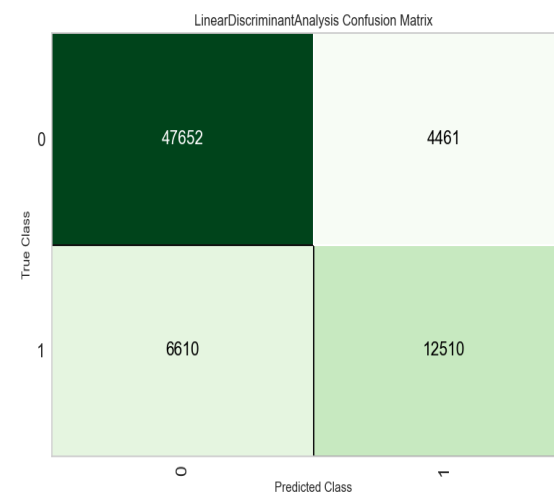
Διαχείριση Μη Ισορροπημένης Μάθησης

- **Διαχωρισμός των δεδομένων** σε δεδομένα εκπαίδευσης και εξέτασης με αναλογία 80%-20%.
- **Ανισορροπία** δεδομένων εκπαίδευσης της μειονοτικής τάξης.
- Επιλογή μεθόδου **Υπερδειγματοληψίας**, ώστε να μην χαθεί πολύτιμη πληροφορία λόγω υψηλής διακύμανσης.
- Επιλογή μεθόδου υπερδειγματοληψίας **SMOTE**.



Διαδικασία Ταξινόμησης (1/2)

- Ανάπτυξη συνολικά **τεσσάρων (4)** αλγορίθμων **ταξινόμησης**.
- **Εξαρτημένη μεταβλητή** η δυαδικής φύσης μεταβλητή της **χρήσης κινητού τηλεφώνου**.
- Επανάληψη της διαδικασίας για τις **πέντε σημαντικότερες** ανεξάρτητες μεταβλητές.



Διαδικασία Ταξινόμησης (2/2)

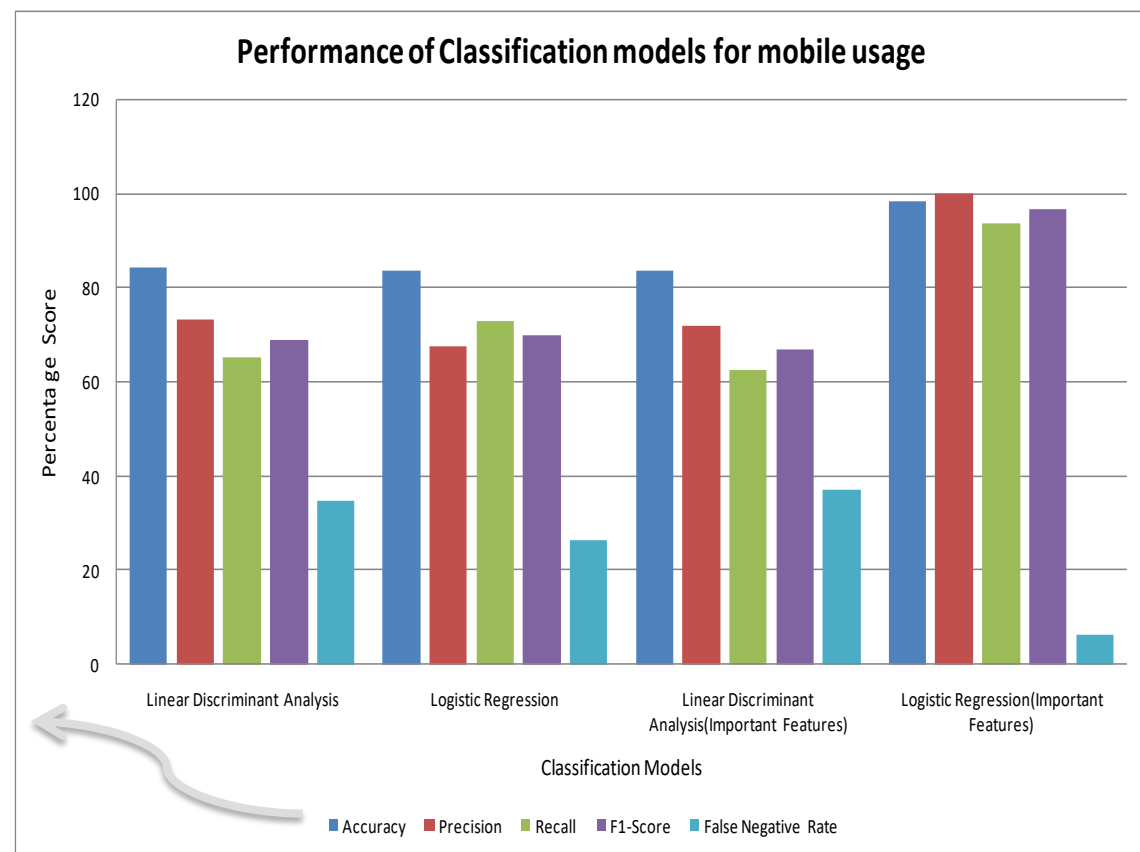
ΤΑΞΙΝΟΜΗΣΗ	Αλγόριθμοι Ταξινόμησης											
	Linear Discriminant Analysis						Logistic Regression					
	Ορθότητα	Ακρίβεια	Ανάκληση	FNR	F1-Score	AUC Score	Ορθότητα	Ακρίβεια	Ανάκληση	FNR	F1-Score	AUC Score
Με όλες τις μεταβλητές	84,4%	73,3%	65,1%	34,6%	68,8%	89,5%	83,4%	67,4%	72,4%	26,4%	70,0%	89,1%
Με τις σημαντικότερες μεταβλητές	83,5%	72,0%	62,4%	37,2%	66,9%	87,7%	98,2%	99,8%	93,5%	6,4%	96,6%	99,9%

- Αξιοποίηση **μετρικών δεικτών** για την αξιολόγηση της προβλεπτικής ικανότητας.
- Καταλληλότερο μοντέλο αναδείχθηκε το **Linear Discriminant Analysis** με **όλες** τις υπό εξέταση ανεξάρτητες μεταβλητές.

Ακρίβεια: Από τις προβλέψεις που πραγματοποιήθηκαν πόσες ήταν **ορθές**.

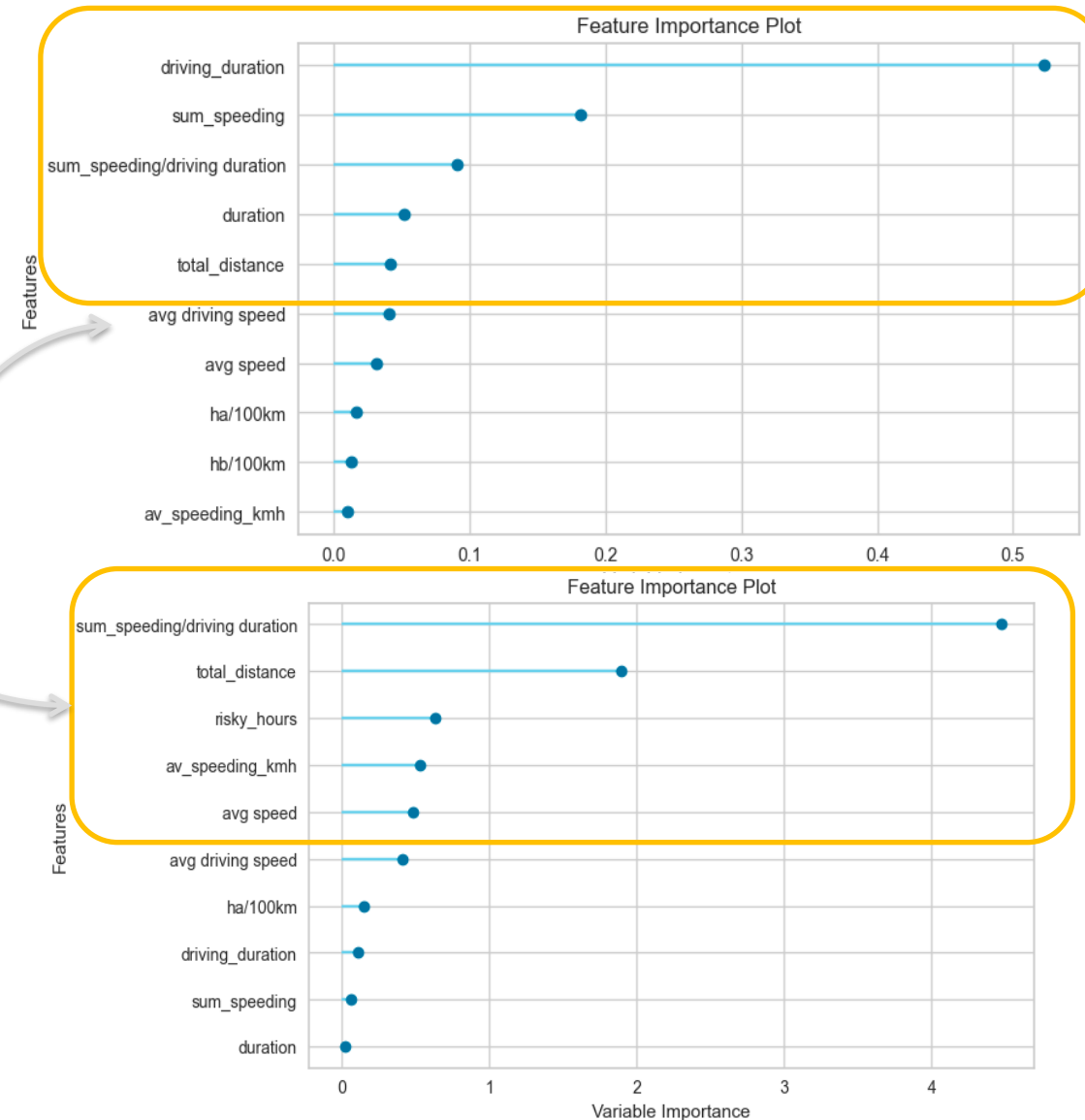
Ανάκληση: Από τα πραγματικά στοιχεία **πόσα** κατάφερε να προβλέψει το μοντέλο.

f1-score: Αρμονικό μέτρο της ακρίβειας και της ανάκλησης.



Διαδικασία Παλινδρόμησης (1/2)

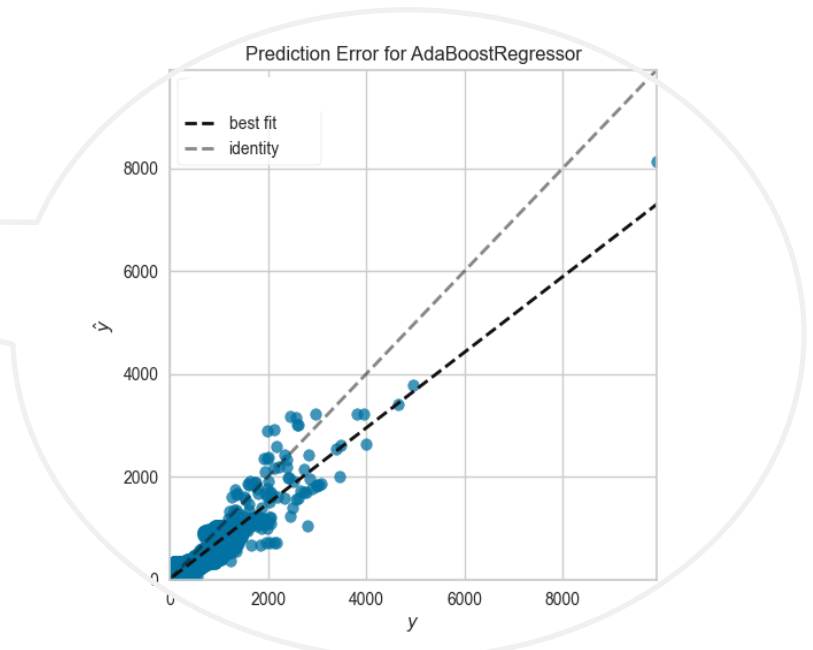
- Ανάπτυξη συνολικά **τεσσάρων (4)** αλγορίθμων **παλινδρόμησης**.
- Η **διάρκεια χρήσης κινητού τηλεφώνου** αποτέλεσε την εξαρτημένη μεταβλητή.
- Διαχωρισμός δεδομένων σε **δεδομένα εκπαίδευσης** και **εξέτασης** σε ποσοστό 80% και 20% αντίστοιχα.
- Επανάληψη της διαδικασίας για τις **πέντε σημαντικότερες** ανεξάρτητες μεταβλητές.



Διαδικασία Παλινδρόμησης (2/2)

- Αξιοποίηση **μετρικών δεικτών** για την αξιολόγηση της προβλεπτικής ικανότητας
- Καταλληλότερο μοντέλο παλινδρόμησης αναδείχθηκε το **AdaBoost Regressor** με όλες τις υπό εξέταση μεταβλητές.
- Μεγαλύτερος συντελεστής προσδιορισμού **R2** και μικρότερα **σφάλματα** για τα μοντέλα με όλες τις ανεξάρτητες μεταβλητές.

ΠΑΛΙΝΔΡΟΜΗΣΗ	Αλγόριθμοι Παλινδρόμησης	
	AdaBoost Regressor	Linear Regression
	R2	R2
Με όλες τις μεταβλητές	0,842	0,497
Με τις σημαντικότερες μεταβλητές	0,840	0,422



Συμπεράσματα (1/2)

- Η **ταχύτητα** αναδεικνύεται ως η βασικότερη παράμετρος επιρροής της **χρήσης** κινητού τηλεφώνου σύμφωνα με τα μοντέλα **ταξινόμησης**.
- Η συνολική **διάρκεια οδήγησης** είτε με είτε χωρίς υπέρβαση του ορίου ταχύτητας αναδεικνύεται ως η βασικότερη παράμετρος επιρροής της **διάρκειας** χρήσης κινητού τηλεφώνου σύμφωνα με τα μοντέλα **παλινδρόμησης**.
- Η **διάρκεια χρήσης κινητού τηλεφώνου** σχετίζεται έμμεσα με την **ταχύτητα**.
- Τα μοντέλα **'Linear Discriminant Analysis'** και **'AdaBoost Regressor'** παρουσίασαν καλύτερη προβλεπτική ικανότητα για τις διαδικασίες της ταξινόμησης και παλινδρόμησης αντίστοιχα.
- Η **συνολική οδηγηθείσα απόσταση** επηρεάζει την χρήση κινητού τηλεφώνου.



Συμπεράσματα (2/2)

- Η **χρήση κινητού τηλεφώνου** φαίνεται να μην επηρεάζει ιδιαίτερα τα απότομα περιστατικά.
- Η οδήγηση κατά τη **διάρκεια της νύχτας** (00:00-05:00) δεν επιδρά αρνητικά στην συμπεριφορά του οδηγού.
- Σημαντικό ποσοστό των οδηγών, της **τάξης του 25%**, χρησιμοποιούσε κινητό τηλέφωνο εν ώρα οδήγησης χωρίς να είναι γνωστή η δραστηριότητα για την οποία γινόταν χρήση κινητού.
- Οι **αλγόριθμοι παλινδρόμησης** με τις σημαντικότερες μεταβλητές παρουσίασαν χαμηλότερο R^2 σε σχέση με τους ίδιους αλγορίθμους λαμβάνοντας υπόψη όλες τις ανεξάρτητες μεταβλητές.
- Η μέθοδος Υπερδειγματοληψίας SMOTE αποδείχθηκε η πιο **αποτελεσματική τεχνική επαναδειγματοληψίας**, λόγω του έντονου φαινομένου υπερπροσαρμογής (Overfitting) χωρίς αυτή.



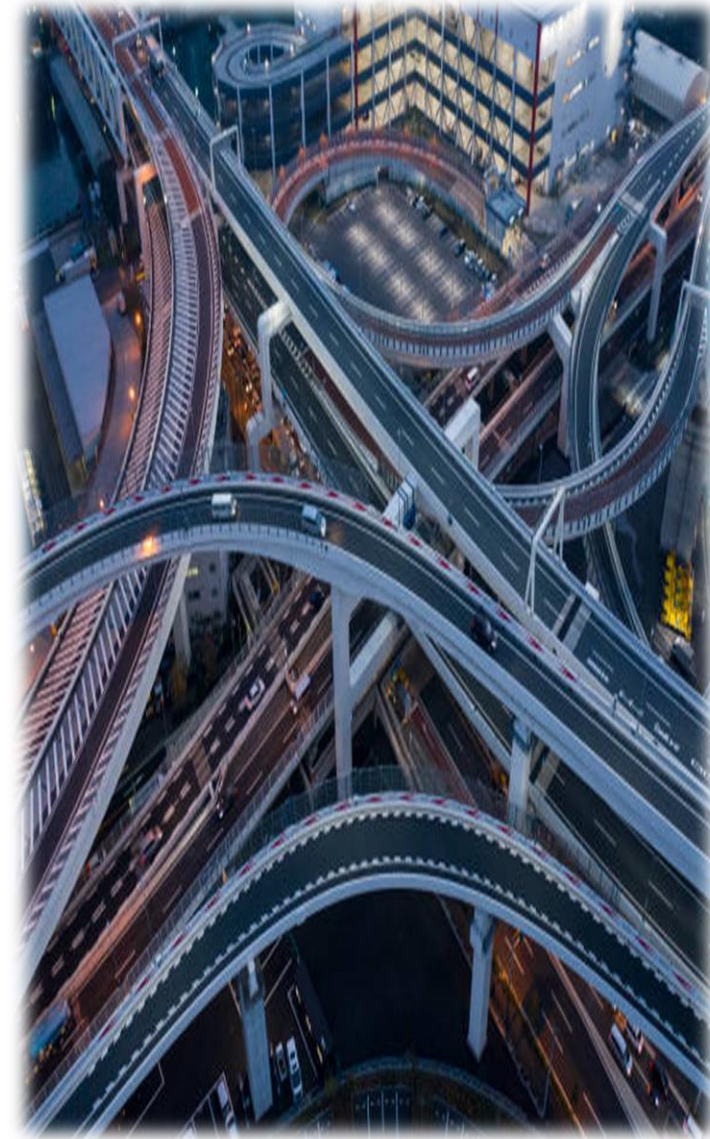
Προτάσεις για Αξιοποίηση των Αποτελεσμάτων

- Αξιοποίηση των επιλεχθέντων μοντέλων ταξινόμησης και πρόβλεψης για την **αναγνώριση** της συμπεριφοράς του οδηγού από δεδομένα **φυσικής** οδήγησης.
- **Σύγκριση** των αποτελεσμάτων με τα αποτελέσματα από τον προσομοιωτή οδήγησης.
- **Αναγγελία** οπτικοακουστικού μηνύματος στο κινητό την ώρα χρήσης του από τον οδηγό με παρότρυνση μη χρήσης του.
- **Περαιτέρω** αξιοποίηση από δημοτικούς, περιφερειακούς και κρατικούς φορείς για τον έμμεσο **εντοπισμό** επικίνδυνων σημείων στο οδικό δίκτυο.



Προτάσεις για Περαιτέρω Έρευνα

- Διάκριση σε περισσότερες κλάσεις ταξινόμησης για **αποδοτικότερη αναγνώριση** της συμπεριφοράς του οδηγού ανάλογα με την διάρκεια χρήσης κινητού τηλεφώνου.
- **Αξιοποίηση και εξέταση** μεγαλύτερου όγκου δεδομένων, όπως δημογραφικών, ψυχογραφικών, δεδομένων γεωμετρίας της οδού και δεδομένων ατυχημάτων.
- Ανάπτυξη **εναλλακτικών μοντέλων** ταξινόμησης και παλινδρόμησης με υψηλές επιδόσεις.
- Ανάπτυξη κατάλληλων μοντέλων **βαθιάς εκμάθησης** (Deep Learning), όπως τα Long Short-Term Memory Networks (LSTMs).





Σας ευχαριστώ πολύ για την προσοχή σας!





Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Πολιτικών Μηχανικών
Τομέας Μεταφορών και Συγκοινωνιακής Υποδομής

**Διερεύνηση της επιρροής της χρήσης κινητού τηλεφώνου
στη συμπεριφορά του οδηγού
με τεχνικές μηχανικής μάθησης
ανισόρροπων δεδομένων**

Κωνσταντίνος-Ειρηναίος Κασελούρης

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

Αθήνα, Μάρτιος 2023

