10th Transport Research Arena Conference
Advancing Sustainable and Inclusive Mobility
Dublin, Ireland, April 15-18, 2024

Investigating the Influence of Mobile Phone Use on Driving Behaviour with Machine Learning Analysis

Konstantinos Kaselouris
Transportation Engineer, Researcher

Together with:
Eva Michelaraki, Christos Katrakazas, Marianthi Kallidoni and George Yannis

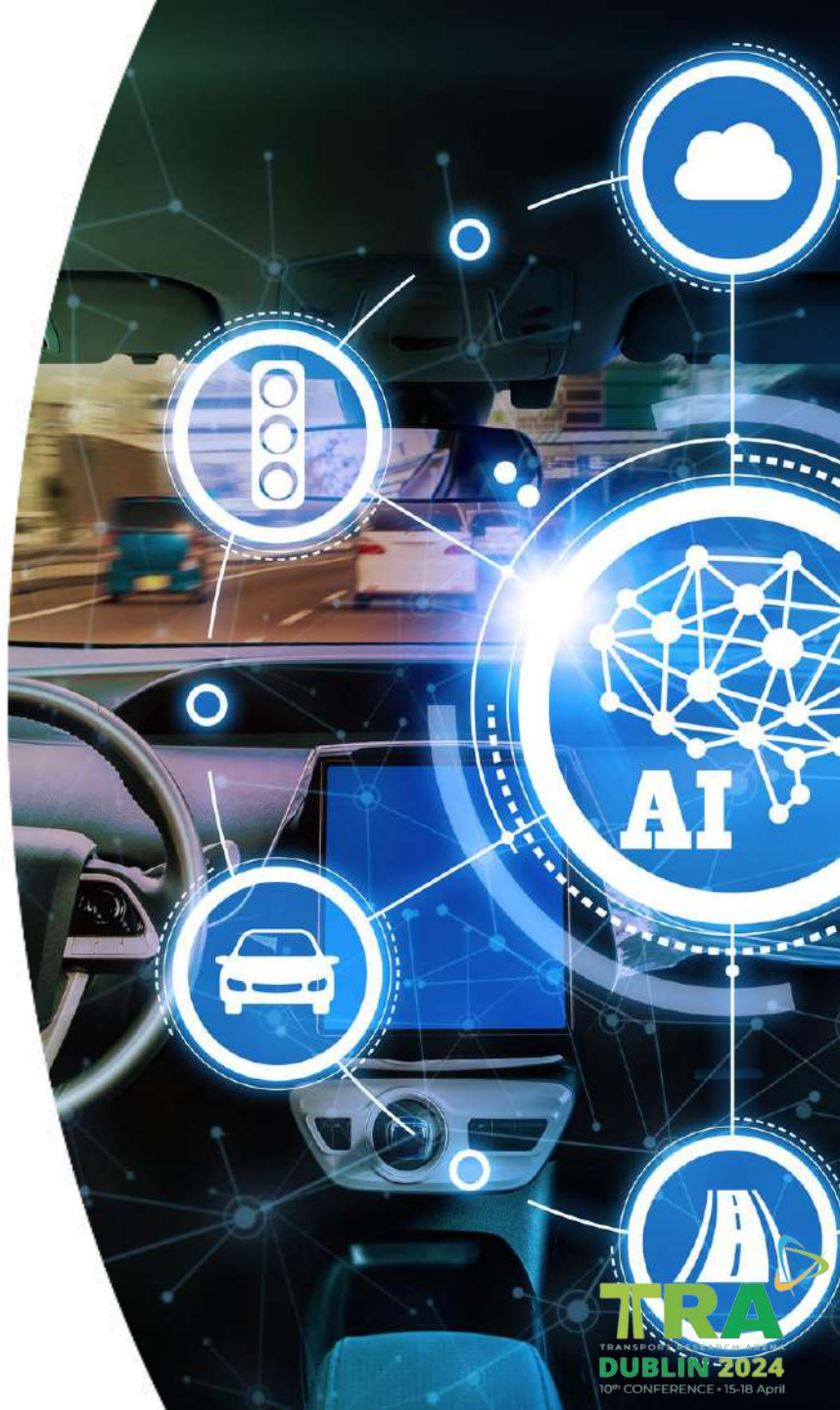
National Technical University of Athens
Department of Transportation Planning and Engineering

Introduction

- According to WHO (2023), 1.19 million people lose their lives every year due to road crashes
- Human behaviour plays a crucial role in road safety
- Factors such as speeding, distraction, aggressive or impaired driving and non-compliance with traffic regulations can increase the crash risk
- Mobile phone usage is a main factor of driver distraction

Objectives

- Investigating the influence of mobile phone use on driver behaviour
- Development of algorithms to find the degree of influence of driver data on risky behaviour
  - Classification algorithms to identify safety levels
  - Regression models to predict the duration of mobile phone use



Data Collection

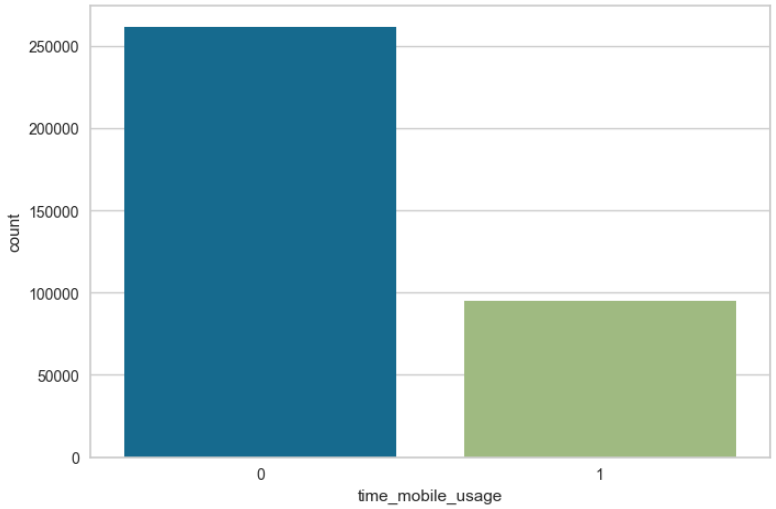
- Data were collected from OSeven database.
- Telematics data via a specially designed smartphone application were used.
- 356,162 different trips were examined.
- A total of 13 variables were exploited for further analysis.

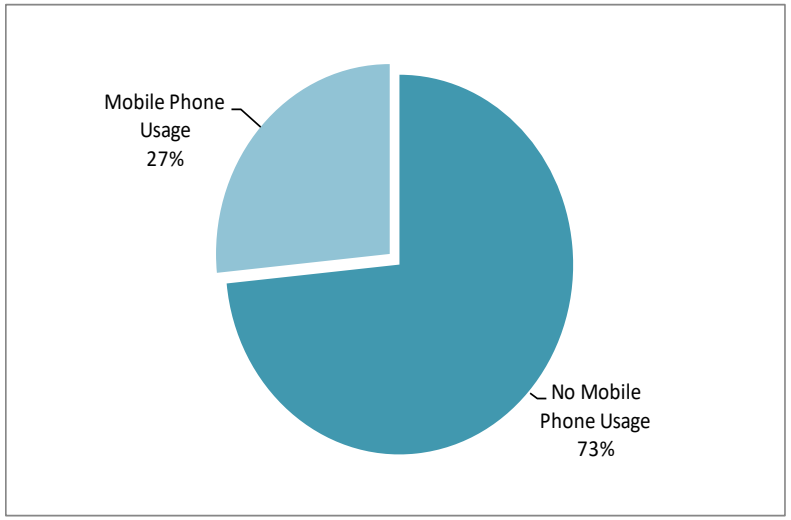
Data Description

- 13 variables: Trip duration, driving duration, total distance, risky hours, ha/100km, hb/100km, avg speed, avg driving speed, time mobile usage/driving duration, sum speeding/driving duration, sum speeding, av speeding kmh, time mobile usage.
- Positive correlation between mobile phone use while driving and driving time.
- Higher correlation between total distance travelled and total driving time either with or without stops.
- Small negative correlation between the 13 independent variables, showing that an increase in one independent variable does not reduce the magnitude of another.

Preprocessing for Classification

- Creation of a binary variable (dependent variable)

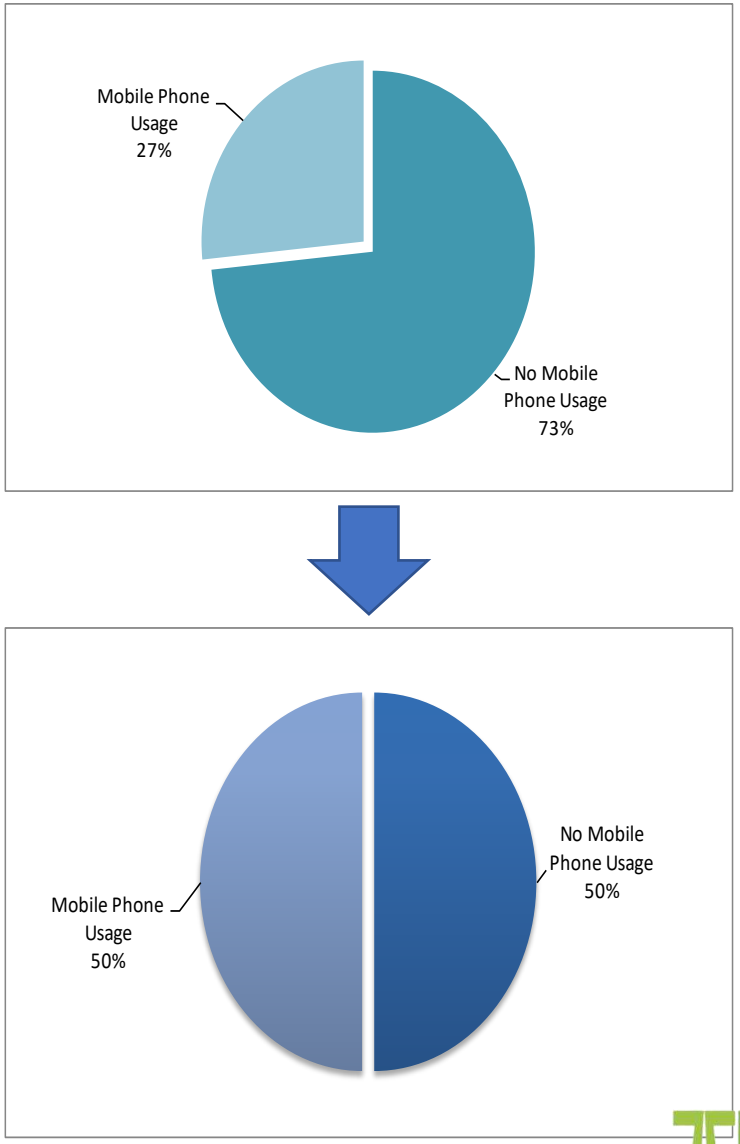




- Two levels of classification:
- Risky behaviour (Mobile use - Value 1)
- Non-risky behaviour (No mobile use - Value 0)

Management of Imbalanced Learning

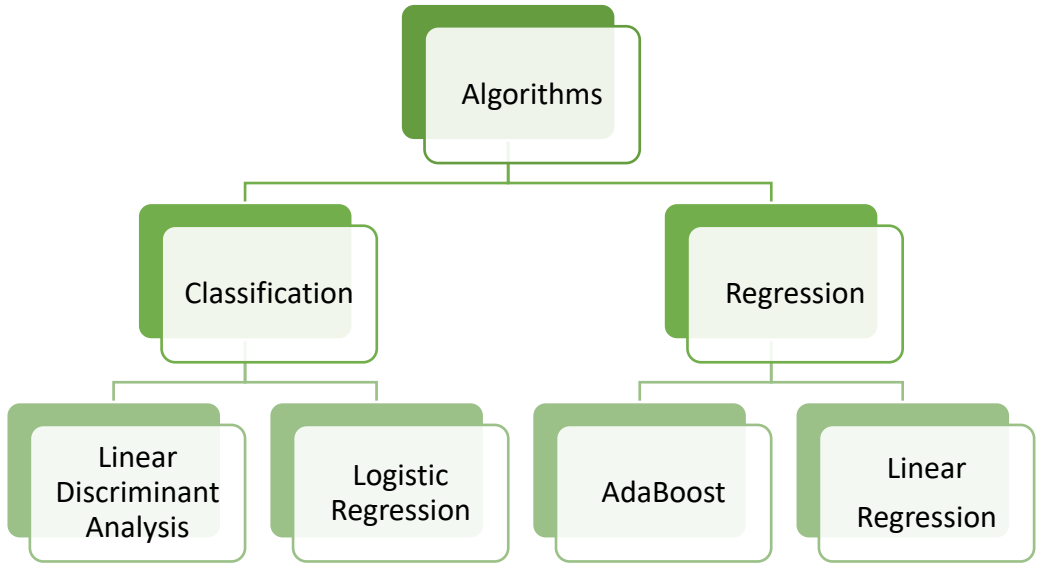
- Separation of data into training and testing data with a ratio of 80%-20%.
- Existence of class imbalance (73% vs 27%).
- Selection of Oversampling method to avoid losing valuable information due to high variance.
- Exploitation of SMOTE oversampling method.



Methodology

- Developing machine learning models
  - Classification: For mobile phone usage prediction
  - Regression: For duration of mobile phone use prediction

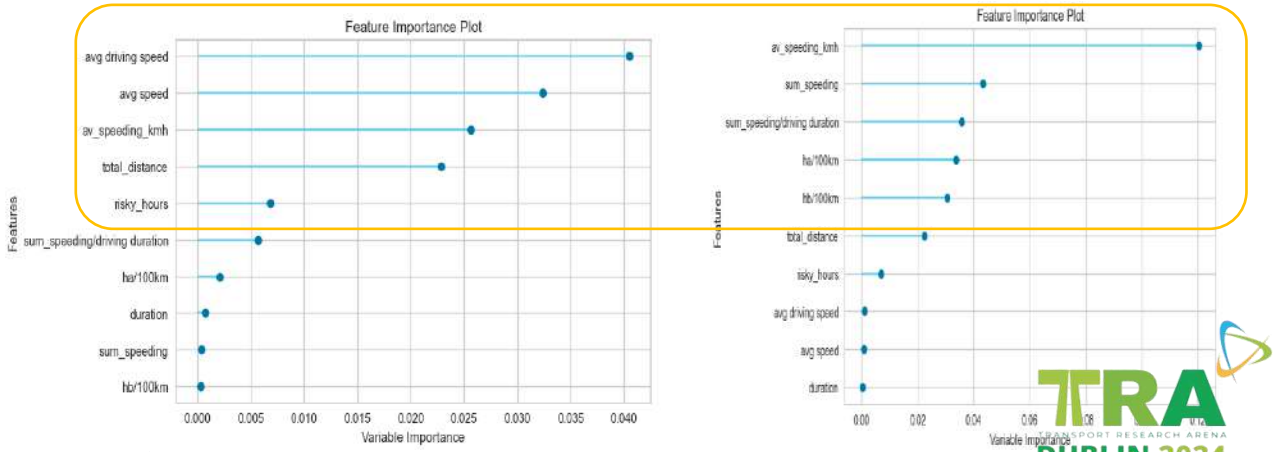
- Assessing the significance of characteristics:
  - Pearson correlation coefficients
  - Significance of variables in classification and regression procedures



Classification Models

- Development of four (4) classification algorithms.
- Binary variable of mobile phone usage was selected as the dependent variable.
- Repetition of the procedure for the five most important independent variables
- Use of metrics to assess predictive ability.
- Linear Discriminant Analysis with all the independent variables under consideration was found to be the most appropriate model.

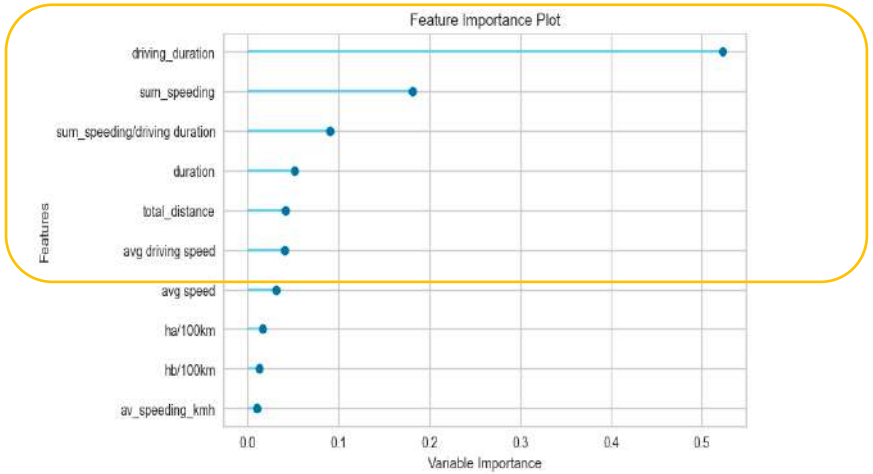
Classification	Linear Discriminant Analysis		Logistic Regression	
	With all variables	With the most significant variables	With all variables	With the most significant variables
Accuracy	84.4%	73.3%	83.4%	98.2%
Precision	73.3%	72.0%	67.4%	99.8%
Recall	65.1%	62.4%	72.4%	93.5%
False Negative Rate (FNR)	34.6%	37.2%	26.4%	6.4%
F1-Score	68.8%	66.9%	70.0%	96.6%
AUC Score	89.5%	87.7%	89.1%	99.9%

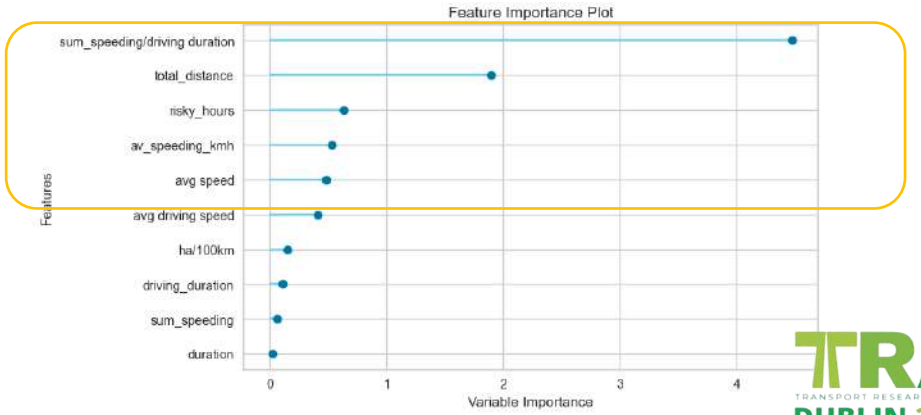


Regression Models

- Development of four (4) regression algorithms.
- The duration of mobile phone use was the dependent variable.
- Split data into training and test data at 80% and 20% respectively.
- Repetition of the procedure for the five most significant independent variables.
- Evaluation of metrics to assess predictive ability.
- The AdaBoost Regressor was found to be the most appropriate regression model with all variables considered.
- Larger coefficient of determination R^2 and smaller errors for models with all independent variables.

Regression	AdaBoost Regressor		Linear Regression	
	With all variables	With the most significant variables	With all variables	With the most significant variables
R^2	0.842	0.840	0.497	0.422





Conclusions (1/2)

- Speed tends to be the key parameter influencing mobile phone use according to the classification models.
- Total driving time either with or without exceeding the speed limit emerges as the most important parameter influencing the duration of mobile phone use according to the regression models.
- The duration of mobile phone use is related indirectly to speed.
- The 'Linear Discriminant Analysis' and 'AdaBoost Regressor' models showed better predictive ability for the classification and regression procedures respectively.
- Total distance driven affects mobile phone use.

Conclusions (2/2)

- Mobile phone use does not seem to have a major impact on harsh driving behaviour events.
- Driving at night (00:00-05:00) does not have a negative effect on driver behaviour.
- A significant proportion of drivers, 25%, used a mobile phone while driving without knowing the activity for which the mobile phone was being used.
- Regression algorithms with the most significant variables showed lower R^2 than the same algorithms considering all independent variables.
- The SMOTE oversampling method proved to be the most effective resampling technique due to the severe overfitting effect without it.

Further Research

- Division in more classification classes for more efficient recognition of driver behaviour according to the duration of mobile phone use
- Utilisation and examination of a larger amount of data, such as demographic, psychographic, road geometry and crash data
- Development of alternative Machine Learning classification and regression models and performance comparison (e.g. Random Forest, Decision Tree, XGBoost, etc.)
- Development of appropriate Deep Learning models, such as Long Short-Term Memory Networks (LSTMs)

10th Transport Research Arena Conference
Advancing Sustainable and Inclusive Mobility
Dublin, Ireland, April 15-18, 2024

Investigating the Influence of Mobile Phone Use on Driving Behaviour with Machine Learning Analysis

Konstantinos Kaselouris
Transportation Engineer, Researcher

Together with:
Eva Michelaraki, Christos Katrakazas, Marianthi Kallidoni and George Yannis

National Technical University of Athens
Department of Transportation Planning and Engineering