The Association of Socio-demographic Characteristics Towards Driver Behaviour and Traffic Fatality in Selangor, Malaysia

Radzuan N.Q.[a], Hassan M.H.A.[a], Abu Kassim K.A.[b], Ab. Rashid A.A.[b],
Mohd Razelan I.S.[c,] Othman N.A.[d]
a Faculty of Mechanical and Automotive Engineering Technology, Universiti Malaysia Pahang, Pekan, Pahang, 26600, Malaysia
b Malaysian Institute of Road Safety Research, Kajang, Selangor, 43000, Malaysia
c Faculty of Civil Engineering Technology, Universiti Malaysia Pahang, Gambang, Pahang, 26300, Malaysia
d Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang, Pekan, Pahang, 26600, Malaysia

ABSTRACT
There are more than 6,000 road traffic fatalities each year in Malaysia since 2004, although the index of road traffic fatality per 100,000 population is decreasing steadily each year. Selangor, one of the states in Malaysia, has recorded the highest road traffic fatalities in 2019. Since traffic accidents and driver's behaviour is correlated, this study aims to understand the association between the socio-demographic characteristics and driver's behaviour, and the correlation between these parameters and road fatality. Road traffic accidents in Selangor in the year 2019 were used as the input data in this study. The selected socio-demographic characteristics are gender, age group, ethnicity, level of education, job specification, and type of driving licence. The statistics show that most of the fatalities involve employed young Malay drivers with more than five years driving experience. A classification model was formulated to classify the driver's behaviour based on the socio-demographic characteristics and vehicle type data. In addition, another classification model was developed, combining the aforementioned parameters as independent variables to classify the occurrence of fatal accidents. The supervised machine learning analysis was conducted using classifiers such as Random Forest, Classification Tree, Neural Network, Support Vector Machine, and Naïve Bayes using Orange data mining software. Random Forest was found to produce the most accurate classification of both driver's behaviour based on the socio-demographic characteristics, and the occurrence of fatal accidents as compared to other classifiers. This study shows that socio-demographic characteristics are indeed associated with driver's behaviour, and both socio-demographic characteristics and driver's behaviour are among the causes of fatal accidents.