Abstract No: PS-01

Identification of factors and classifying the accident severity in Colombo-Katunayake expressway, Sri Lanka using multinomial logistic regression

M. A. K. Kushan* and N. V. Chandrasekara

Department of Statistics & Computer Science, Faculty of Science, University of Kelaniya, Sri Lanka
Kalanakushan290@gmail.com*

Accidents are one of the main social problems in the World, which cause damages or injuries unintentionally and unexpectedly. This is a major issue affecting not only in developing countries like Sri Lanka but also in developed countries. Sri Lanka's expressway system was launched in 2011 and currently has three major expressways: Southern Expressway, Colombo-Katunayake Expressway and Outer-Circle Expressway. After the construction of expressways, many people opted for expressways based on time, traffic, ease of driving, etc., rather than ordinary roads. The number of accidents on expressways has been on the rise in recent years compared to the past. At present, the accident rate on the Colombo-Katunayake Expressway, which connects the Sri Lankan capital, Colombo with Bandaranaike International Airport, Katunayake and Negombo, is high compared to the other two expressways, but no research has been done to date regarding this. Therefore, the objective of the study was to identify the factors contributing to accidents on the Colombo-Katunayake Expressway and to develop an appropriate regression model to classify the severity of the accidents. In this study, 704 total accident cases of Colombo-Katunayake expressway were considered during the period from 2013 to 2019. Initially, Pearson Chi-square, Logistic regression and Kruskal–Wallis H tests were used to identify the association between the multinomial response variable (accident severity) and eleven predictor variables identified based on the literature. Finally, from selected predictor variables, seven variables: time category, driver's age category, vehicle type, reason for the accident, number of vehicles involved, cause for accident and rainfall were identified as influencing variables to accident severity under 5% level of significance. Since this is not a time series data, 80% of the data were selected in various ways for model building and the remaining 20% were used to test the performance of the built models. Considering significant variables identified above, Multinomial Logistic Regression (MLR) was trained using the stepwise enter method with different data selections criteria. The Random under-sampling technique was used to overcome the class imbalance problem that persists in the data set considered in the study and after selecting the best model, the adequacy of the model was examined and classified the severity of accidents in Colombo-Katunayake Expressway. The final MLR model predicts accident severity with an overall accuracy of 64.3% and rainfall, cause for accident and time category (it is a categorical variable that divides 24 hours into four equal parts) have been identified as the most influential factors affecting accidents on the Colombo-Katunayake Expressway. Furthermore, the final model depicts, with rainy weather, high speed, sleepiness, technical faults and reckless driving increased the likelihood of an accident on the Colombo-Katunayake Expressway, and [0-6] and [12-18] hours were identified as dangerous time categories. The final model developed by this study can be used to implement safety improvements against traffic accidents in expressways of Sri Lanka. As a future study, machine learning techniques can be employed to identify better models with higher classification accuracy.

Keywords: Classification, Class imbalance problem, Expressway, Random under-sampling, Road accident