



**INFORMATICS
INSTITUTE OF
TECHNOLOGY**

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Driver Turnover Prediction with Machine Learning

A case study of drivers in the ride-hailing industry in Sri Lanka

A Project Proposal by

Ms Manori Priyadarshani

Supervised by

Mr Alroy Mascenghe

Submitted in partial fulfillment of the requirements for the MSc in Advanced Software Engineering degree at the University of Westminster.

October 2021

ABSTRACT

Ride-hailing and online delivery platforms have gained high attention in society because it makes people's life easier. The platform directly depends on individual drivers who have joined the platform willingly with the earning purpose. Most of them tend to churn and do not stay for a long time. The high driver churn makes the platform unstable and pre-identifying churning drivers are crucial for the ride-hailing platforms to take relevant action to make them retain.

This research tries to research, design, and develop an accurate driver churn prediction model in a machine learning approach. The research used a dataset from a leading local ride-hailing company and extracted 18 important driver features to build a binary classification model. Multiple models are built with different machine learning algorithms covering traditional algorithms, ensemble techniques and neural network areas. Models built from algorithms, Naive Bayes, K-Nearest Neighbor, Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, Bagging (Decision Tree), XG Boosting (Decision Tree), ADA Boosting (Decision Tree), Stacking (DT, SVM, RF +LR) and Feed Forward Neural Network are compared to find the best model for this particular research problem. The result recommends ensemble classifiers which have high accuracy compared to other classifiers as the best solution to this problem. The best model was the Random Forest model and the achieved accuracy is 95.7%. All other ensemble learners got accuracies near 94% and feed-forward neural network and decision tree learners also showed an average accuracy, while Naive Bayes and K-Nearest Neighbor showed the lowest accuracy. In summary, the research recommends ensemble machine learning learners to solve the problem of driver churn identification

Key Words: Driver Churn, Ride hailing, Classification, Ensemble, Machine learning