## CLASSIFYING RIDE FARE PREDICTIONS USING FILTERED ENSAMPLES RANDOM FOREST AND GRADIENT BOOSTING ALGORITHMS

## <u>Uthayasanker. T</u>.

Department of Computer Science and Engineering, Faculty of Engineering, University of Moratuwa, Sri Lanka rtuthaya@cse.mrt.ac.lk

Ride-hailing applications predict the fare for each trip at the beginning of the ride. Accurate prediction fare prediction helps improve the user experience and trip conversion. We aim to build a machine learning model that classify the ride fare estimation. This model helps improve the fare estimate model and feature analysis suggest the key impact factors in ride fare. A selected sample of trips, the fare predictions are labeled as either correct or incorrect. This project aims to develop an accurate and explainable machine learning model to classify these fare predictions. Key features were manually engineered from the provided dataset, including calculated distances, driving times, and various fare-related ratios, to enhance model performance. The dataset was carefully preprocessed to handle missing values and imbalances in the label distribution. Various machine learning models, including Random Forest, XGBoost, and LightGBM, were employed, with XGBoost proving to be the most effective individual model. To further enhance accuracy, a filtering ensemble technique was used to combine these models, achieving an impressive 98.32% accuracy. The explainability of the model was analyzed using SHAP values, which revealed that the most significant factors influencing predictions were ride rate, ride distance, and waiting time. Additionally, other factors such as ride hour, ride day, and ride month also played a role in the model's performance. This approach provides a robust and interpretable solution to the problem of ride fare classification. The study highlights that fare misclassification is analogous to an outlier detection problem, where incorrect fares often exhibit extreme rates in terms of fare per distance or driving time. The findings suggest that further improvements could be made by incorporating additional features related to ride penalties.

*Keywords*: Engineering, Ensemble methods, Feature, Ride fare Classification, Machine learning, XGBoost.