Development of a Method for Bus Arrival Time (to bus halt) Prediction Model Using Machine Learning

M P V P Medawatte[1], H.L.K Perera2, A.B Jayasinghe3, K A S N Sumathipala4

Abstract

The pursuit of innovative methods to enhance various aspects of the transportation field continues to evolve over the years. Intelligent Transportation Systems (ITS) have played a pivotal role in integrating next-generation technologies and advancements into transportation. In recent years, a substantial volume of transportation data has been collected from various sources, including road sensors, probes, GPS, CCTV, and incident reports. Similar to many other industries, transportation has entered the generation of big data. With a rich volume of traffic data, it is challenging to build reliable prediction models based on traditional data analysis techniques. Accurate real-time bus arrival information is a critical component of a public bus transport network, influencing passenger experience, ridership reliability, reduced waiting times, dwell times, and operational efficiency. The Sri Lankan public transport system currently relies on static data to predict bus arrival times, lacking real-time data and factors affecting predictions. This study aims to identify unique factors impacting bus arrival times in Sri Lanka's road system and create precise travel time predictions that consider these factors. Using GPS data and Machine Learning techniques, the study focuses on developing an accurate real-time bus arrival time prediction model. While previous methods like Historical Average Models, Regression, and Kalman Filtering have been used for short-term travel time estimations, Machine Learning techniques show greater accuracy due to their capacity to handle intricate data relationships, large volumes of data, and nonlinear connections among predictors.

GPS data from the Moratuwa to Colombo bus route (100) was collected for the study for 30 days covering weekdays, weekends, and public holidays using five GPS units at various times of the day. This included the data of 330 bus turns giving more than 17500 GPS data points. The route consists of 53 bus stops and data filtering was performed based on GPS locations relative to each bus stop to calculate travel times between each stop separately. To improve accuracy, the total route was segmented, considering the number of bus stops, allowing for more precise travel time predictions by identifying speed changes in each section. The literature review highlighted several key factors influencing bus travel time, encompassing road section, time and day variations, peak/off-peak hours, bus lane availability, distance traveled, traffic flow, weather conditions, signalized intersections, and crossings. Data collection was conducted to encompass all of these factors. Out of the available Machine Learning models, the Support Vector Regression Model, Adaboost Model, K-Nearest Neighbors (KNN) Model, Random Forest Model, XG Boost Model, and Gradient Boosting Regression Model were identified as the models giving the most accurate results for the prepared data set. By using the model performance evaluation parameters like Mean Absolute Error (MAE), R2 Value, and Root Mean Square Error (RMSE), the model prediction accuracies can be compared to select the most explanatory model. Out of the selected Machine Learning models, the Support Vector Regression Model and XG Boost Model were not selected due to their R2 values being less than 0.5 and higher mean absolute errors. In contrast, the Random Forest Model, KNN model, and Gradient Boosting Regression Model yielded lower mean absolute errors and demonstrated

higher R2 values above 0.6. Consequently, these models were chosen for further analysis. The selected factors for prediction can be further narrowed down by performing feature selection before running the above models to further narrow the data collection needed for future studies. The proposed model aims to accurately predict bus arrival times at each stop given the trip's starting time from the origin. Apart from the mentioned models, further exploration of Long Short-Term Memory (LSTM) models using the dataset is suggested. Once finalized, this model can be extended to other routes within Sri Lanka's road network by updating the dataset with specific route characteristics. Future improvements can include incorporating real-time traffic and accident data to enhance prediction accuracies and provide precise arrival times for passengers across the public transport network. This selected travel time prediction model for Sri Lanka's public bus operations holds practical applications, offering fast, advanced, and accurate results.

Authors Details;

1. Graduate Research Assistant, Department of Civil Engineering, University of Moratuwa, Sri Lanka, vidunipramodya@gmail.com
2. Traffic Engineer, City of Rockingham, Australia. perera.loshaka@gmail.com
3. Senior Lecturer, Department of Town & Country Planning, University of Moratuwa, Sri Lanka, amilabj@uom.lk
4. Senior Lecturer, Department of Computational Mathematics, University of Moratuwa, Sri Lanka, sagaras@uom.lk