DEVELOPMENT OF A BUS ARRIVAL TIME (TO BUS HALT) PREDICTION MODEL USING MACHINE LEARNING

M P V P Medawatte

228014G

Degree of Master of Science

Department of Civil Engineering

University of Moratuwa

Sri Lanka

November 2025

DEVELOPMENT OF A BUS ARRIVAL TIME (TO BUS HALT) PREDICTION MODEL USING MACHINE LEARNING

M P V P Medawatte

228014G

Thesis submitted in partial fulfillment of the requirements for the degree
Master of Science

Department of Civil Engineering

University of Moratuwa

Sri Lanka

November 2025

DECLARATION

I declare that this is my own work, and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to the University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 05.11.2025

The above candidate has carried out research for the Master's thesis under my supervision. I confirm that the declaration made above by the student is true and correct.

Name and Signature of the supervisors:

Prof. J.M.S.J. Bandara ……………………………. Date: 06. 11. 2025

Dr Loshaka Perera ……………………………. Date: 26/11/2025

Dr Amila Jayasinghe ……………………………. Date: 26/11/2025

Dr Sagara Sumathipala. ……………………………. Date: 27/11/2025

DEDICATION

I dedicate this dissertation to my beloved parents, who have provided me with immense support in everything I do. Your unwavering love, endless support, and boundless encouragement have been the foundation of my journey. I also dedicate this work to my dearest husband, for his immense support, understanding, and love at every step of this academic journey.

M P V P Medawatte

Department of Civil Engineering,

University of Moratuwa.

05.11.2025

ACKNOWLEDGEMENT

I would like to acknowledge many individuals who supported me throughout this study, guiding me and supporting me. Firstly, I am sincerely thankful to the LBS2ITS Project, ERASMUS+ Programme funded by the European Union for partnering with the University of Moratuwa and extending financial support for this research.

I am extremely grateful to the esteemed **Prof. Saman Bandara** for collaborating with me on this study and enriching my research journey immensely.

I would like to express my heartfelt gratitude to my research supervisor **Dr. Loshaka Perera**, Senior Lecturer of the Department of Civil Engineering, University of Moratuwa, who gave me this great opportunity to participate in this study. His invaluable guidance and supervision throughout the entire research study are highly appreciable.

I would also like to thank my research co-supervisor, **Dr. Amila Jayasinghe,** for his provoking discussions that challenged me to think critically and analytically, enhancing the quality of my study. I am sincerely thankful to **Dr. Sagara Sumathipala** of the Department of Computational Mathematics for guiding me through all unfamiliar areas of machine learning and directing my study toward the right path.

I wish to thank the support given by the Department of Civil Engineering, University of Moratuwa and its academic and non-academic staff members. Special mention to **Prof. C. Jayasinghe**, the current Head of Civil Engineering Department in providing all administrative support when required.

I would also like to thank all the lecturers in the Transportation Engineering Group in addition to my supervisor, including **Prof. W.K. Mampearachchi**, **Dr. H.R. Pasindu** and **Dr. G.L.D.I. De Silva,** who helped me both academically and otherwise during my time as a researcher.

I would like to provide a special mention to the postgraduate course coordinator, **Mrs. Melani Jayakody,** for her continuous administrative support and my **fellow research assistants** in the Transportation Engineering Group during the period of 2022/2023 for their constant help and guidance throughout the research period. Finally, I must thank all others whose names are not individually mentioned who had helped me throughout the period of the study.

M P V P Medawatte
Department of Civil Engineering,
University of Moratuwa.
05.11.2025

ABSTRACT

Accurate, real-time bus arrival information is essential for enhancing public transportation efficiency. Although real-time vehicle location data provides valuable information on current bus positions, it alone cannot forecast future arrival times. Prediction models are therefore essential for estimating travel times to upcoming stops by capturing the dynamic relationships among traffic flow, passenger demand, and operational conditions. In Sri Lanka's public transport network, bus arrival times are currently estimated using Google Maps and Google Transit. However, these systems rely solely on static transit data for their predictions, limiting their accuracy in accounting for real-time external conditions. As a result, they do not incorporate dynamic factors that could significantly impact bus travel time predictions. Hence, there is a need to identify unique factors affecting bus arrival times within the Sri Lankan Road system and to develop an accurate travel time prediction model that accounts for these factors. This study focuses on the development of an accurate real-time bus arrival time prediction model using Global Positioning System (GPS) data and machine learning techniques.

GPS data from the Moratuwa to Colombo bus route (100) was collected for the study over 30 days, covering weekdays, weekends, and public holidays, using five GPS units at various times of the day. This included data from 335 bus trips, resulting in GPS data points with an accuracy up to 1 sec. 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 by considering the number of bus stops, allowing for more precise travel time predictions by identifying speed changes in each section. Through a review of the literature, key factors affecting bus travel time were identified, including road section, time of day, day of the week, peak/off-peak hours, availability of bus lanes, distance travelled, traffic flow, weather conditions, number of signalized intersections, and number of signalized crossings. Data collection was conducted to encompass all these factors.

After data collection, the most accurate travel time prediction model was identified by evaluating various types of models previously used in research. The available machine learning models, the Support Vector Regression Model 9, Adaboost Model, K-Nearest Neighbors (KNN) Model, Random Forest Model, XG Boost Model, Gradient Boosting Regression Model, and Long Short-Term Memory (LSTM) Model were selected to be trained with the available dataset. Mean Absolute Error (MAE), Mean Square Error (MSE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and R-squared value were selected as model evaluation parameters. Out of the tested models, SVR, RFR, KNN, AdaBoost, XGBoost and GBR performed similarly to each other, showing minute changes with MAE values varying between 17 seconds to 18

seconds and r^2 values greater than 0.6. It was clear that the LSTM model performed prominently better than the other models due to the time series nature of the data set providing the lowest MAE of 3.8 seconds and the highest r^2 value of 0.98. The proposed model aims to generate accurate bus arrival time predictions for each bus stop when the trip's starting time from the origin is known. The model can be further improved in the future by incorporating real-time traffic data and real-time accident data to enhance prediction accuracy, providing accurate bus arrival times for passengers across the public transport network.

Keywords: Intelligent Transport Systems, Bus Travel Time, Machine Learning, Long Short- Term Memory Model

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

GDP - Gross Domestic Product

ITS - Intelligent Transportation Systems

APTS - Advanced Public Transport Systems

GPS - Global Positioning Data

ML - Machine Learning

LBS – Location-Based Services

ITMS - Intelligent Traffic Management Systems

IPIS - Intelligent Passenger Information Systems

MAE - Mean Absolute Error

MSE - Mean Squared Error

MAPE - Mean Absolute Percentage Error

RMSE - Root Mean Square Error

GIS - Geographical Information System

MM - Map Matching

ATM - Automated Teller Machine

TMS - Traffic Management Systems

ATMS - Advanced Traffic Management Systems

FTMS - Freeway Traffic Management Systems

UTC - Urban Traffic Control

ADAS - Advanced Driver Assistance Systems

AVI - Automatic Vehicle indication

IVIS - In-vehicle Information Systems

ATIS - Advanced Traveler Information Systems

TTI - Traffic and Travel Information

PTP - Public Transport Priority

AVM - Automatic Vehicle Monitoring

AVL - Automatic Vehicle Location

GTFS - Google Transit Feed Specifications

APC - Automatic Passenger Counting

WIM - Weight in Motion

ATCS - Adaptive Traffic Control Systems

ANN - Artificial Neutral Network

AI - Artificial Intelligence

SVR - Support Vector Regression

KNN - K- Nearest Neighbours

RFR - Random Forest Regression

GBR - Gradient Boosting Regression

SVM - Support Vector Machines

RBF - Radial Basis Function

SMA - Simple Moving Average

OLS - Ordinary Least Square

FNN - Fully connected Neural Networks

LSTM - Long Short-Term Memory

API - Application Programming Interface

RNN - Recurrent Neural Network

MdAE - Median Absolute Error

PCA - Principle Component Analysis