Peak-Hour Urban Road Traffic Volume Prediction Model Based on Short-Term Traffic Count in Putrajaya

Muhammad Izzuddin Mohd Jamil[1], Basil David Daniel[1]*, Gunawan[2], Muhammad Aliff Iman Shamsudin[3]

[1] Faculty of Civil Engineering and Built Environment,
Universiti Tun Hussein Onn Malaysia, Batu Pahat, 86400, MALAYSIA

[2] Faculty of Engineering,
Universitas Muhammadiyah Surabaya, INDONESIA

[3] IJM Construction Sdn Bhd,
Menara IJM Land, 1, Lebuh Tunku Kudin 3, 11700 Gelugor, Pulau Pinang, MALAYSIA

*Corresponding Author: basil@uthm.edu.my
DOI: https://doi.org/10.30880/ijscet.2024.15.02.017

Abstract

Rapid urbanization and population growth have significantly increased traffic volume, resulting in concern over efficient transportation planning and management. Traffic count data is critical for understanding the present and future transportation demands. However, the traditional data collection process is time-consuming and labour-intensive. Therefore, this study focuses on developing a peak-hour traffic volume prediction model based on short-term traffic counts. This study was conducted at seven arterial roads in Putrajaya for 3 hours during peak times in the morning and evening, with vehicles classified by their type. Single Linear Regression (SLR) and Multiple Linear Regression (MLR) analysis were used to develop the prediction models. The analysis focused on identifying the most relevant predictors for accurate traffic volume prediction. Next, by comparing the predicted and actual traffic volumes, each model was validated using Mean Absolute Percentage Error (MAPE). A lower percentage error shows the model's reliability in accurately anticipating peak hour volume. The result shows that the peak hour traffic volume prediction model based on a 15-minute short-term traffic count using MLR analysis is the best model to use to predict traffic volume in Putrajaya with a MAPE of 12.09%, which suggests its reliability to follow actual traffic volume data. With a low error rate, it demonstrates the model's practical utility for traffic management and transportation planning in Putrajaya, offering more efficient traffic management strategies, ultimately helping to reduce congestion, minimise accidents, and save both time and resources.

1. Introduction

The transport sector is greatly influenced by many external economic and social factors. These factors include the population's living standards, demographics, urban planning, accessibility to transportation infrastructure, and changes in the structure of society [1]. Ali et al., (2018) point out that the transportation industry faces significant challenges, particularly in urban areas, where issues like traffic congestion, high energy consumption, air pollution, excessive use of motor vehicles, and mobility difficulties for disabled individuals and senior citizens

Int. J. of Sustainable Construction Engineering and Tech. Vol. 15 No. 3 (2024) p. 192-208

193

prevail [2]. These challenges have led to growing concerns over traffic volume, especially with the escalation of population and urbanisation.

Calculating traffic flows is a significant task involving vehicle counting, although it can be challenging in congested areas [3]. This process, known as traffic volume study or survey, aims to ascertain traffic volume at specific road segments and times, recognising its fluctuating nature throughout the day, week, and year [4]. As emphasised by Rashid et al., (2017), traffic volume analysis assists in identifying critical flow periods and evaluating the impact of large vehicles or pedestrians. Two techniques are utilised for carrying out traffic volume counts, namely manual and automatic [5].

Traditionally, vehicles have been counted and classified manually on-site when analysing traffic. This is usually done by individual observers who rely on visual examination and observation. The data is then recorded using tally sheets or mechanical counters [6]. As Prudhvi et al. (2019) have noted, automatic methods based on detectors have become increasingly popular. These methods involve the automatic counting of vehicles, with no human involvement. Frequently used sensors include inductive loops and sensors that weigh in motion [4].

However, Perera et al., (2021) highlighted that manual data collection is often time-consuming and leads to human errors [7]. While automated systems are effective, they can be expensive and inaccessible to some organisations or regions [8]. Sensor-based methods are an alternative but costly and may miss slow-moving vehicles [3].

Unmanned Aerial Systems (UAS), specifically small Unmanned Aerial Vehicles (UAVs) or drones, have emerged as a promising traffic management and monitoring technology. UAVs can analyse traffic flow and safety conditions [9]. However, limitations such as restricted battery life and safety concerns exist. Additionally, severe weather conditions like strong winds, rainfall, and fog can hinder UAV operations [10].

In recent years, traffic forecasting has become integral to intelligent transportation systems (ITSs), vital for planning and developing traffic management systems [11]. Traffic volume estimation, achieved through short-term traffic count data, forms the basis for prediction models. Reliable models, as noted by Diao et al. (2019) provide crucial insights for transportation planners and engineers regarding infrastructure investments [12]. The analysis of traffic data and prediction of future traffic volume by using statistical methods, data mining techniques, and artificial intelligence have become increasingly popular [13].

In this study, an approach to developing a prediction model for traffic volume was conducted to reduce the time spent at the observation site while maintaining data accuracy. The study used short-term traffic counts during peak hours on urban roads to create a traffic volume prediction model. Regression analysis was used as a statistical and modelling method to develop the prediction model. Establishing a traffic volume predicted model holds significant importance as it can improve safety measures, mitigate environmental impacts, optimise transportation system planning and management, and provide economic benefits. Thus, this study contributes to developing a prediction model to enhance transportation planning and management and hopefully lead to a more sustainable transportation system.

2. Methodology

The data was collected using various techniques, including manual counting, automatic counting equipment (such as infrared or pneumatic tube sensors), and video cameras. For data collection, this study used manual counting, which involved an observer counting the number of vehicles at a particular location during a given period.

2.1 Study Area

Putrajaya was selected as the study area, considering the arterial roads within the urban city that started and ended at intersections equipped with traffic lights. Putrajaya's status as a planned city allows for controlled studies that can provide valuable insights into urban transportation planning and management practices. As the federal administrative capital of Malaysia, it presents a unique urban environment with planned infrastructure and controlled development. This makes it an ideal location for studying traffic volume patterns, road usage, and traffic management strategies. The figure below shows the traffic count locations in Putrajaya. Figure 1 shows two roads in Putrajaya. The first is in Presint 14, while the second (b) is at Presint 2. Two intersections connect both roads, each with traffic lights, similar to Figure 2, which shows two roads. The first is at Lebuh Perdana Selatan, Presint 16, and the second is at Lebuh Perdana Timur, Presint 1. These intersections also have traffic lights. Next is Figure 3, which showcases two more roads in Putrajaya. The first is at Lebuh Wadi Ehsan, Presint 4, and the second is at Jalan Tun Rusnah, Presint 3. Both roads also are equipped with traffic lights at each intersection. Lastly, Figure 4 shows another road at Presint 14. This intersection is different from the one shown in Figure 1, as it represents a different road in Presint 14.
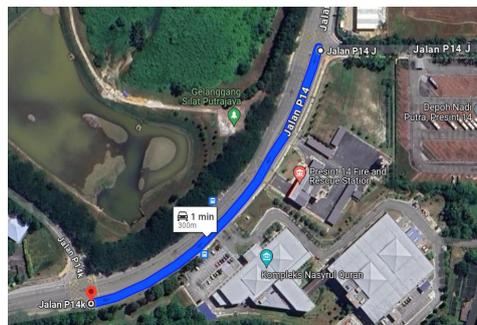
Fig. 1 *(a) Jalan P14, Presint 14, Putrajaya (1st); (b) Lebuh Bestari, Presint 2, 62100 Putrajaya*

Fig. 2 *(a) Lebuh Perdana Selatan, Presint 16, 62150 Putrajaya; (b) Lebuh Perdana Timur, Presint 1, 62000 Putrajaya*

Fig. 3 *(a) Lebuh Wadi Ehsan, Presint 4, 62100 Putrajaya; (b) Jalan Tun Hussein, Presint 3, 62100 Putrajaya*

Fig. 4 *Jalan P14, Presint 14, Putrajaya (2nd)*

2.2 Time for Data Collection

This study collected traffic count data during peak hours (morning and evening), typically the heaviest traffic times of the day. The peak-hour traffic volume in Malaysia was influenced by urbanisation, population growth, and economic development. Based on the Malaysian Public Works Department (JKR) practice, the peak hour is 2 hours. However, for this study, the observation time for data was 3 hours (7.00-10.00 am and 4.00-7.00 pm). This

Int. J. of Sustainable Construction Engineering and Tech. Vol. 15 No. 3 (2024) p. 192-208

195

is because a more extended timeframe can provide a more extensive dataset, potentially improving the quality of the input data. This could result in a more reliable and stable regression prediction model.

2.3 Data Analysis

This study utilised Microsoft Excel and SPSS software to analyse the data. Using a regression method, the analysis was conducted to establish the traffic volume during peak hours of 3 hours as the dependent variable. Different trials and combinations were explored to identify the most significant independent variables for analysing SLR and MLR. The independent variables included the number of vehicles in each class, Class 1, Class 2, Class 3, Class 4, Class 5 and Class 6, as shown in Table 1. Other independent variables also included combinations of vehicle classes such as Passenger vehicles (Class 1 + 2), lightweight vehicles (Class 1 + 2 + 6), lorries (Class 3 + 4), and heavyweight vehicles (Class 3 + 4 + 5). A series of analyses were conducted to create an optimal formula model for estimating traffic volume during peak hours.

Table 1 *Vehicle classification*

Class	Type of vehicle
1	Cars & Taxis
2	Small Vans & Utilities
3	Lorries (2-axles)
4	Lorries (3-axles & above)
5	Busses
6	Motorcycles & Scooters
PC	Passenger vehicles
LW	Lightweight vehicles
HW	Heavyweight vehicles
Lorries	Class 3 + Class 4

2.4 Regression Analysis

A statistical evaluation called regression analysis represents the predictor and criterion variables [14]. Equation 1 shows the basic SLR formula, and Equation 2 is the basic MLR formula used for this study.

$$Y' = a + biXi \tag{1}$$

$$Y' = a + b1X1 + b2X + ..bnXn \tag{2}$$

2.4.1 Correlations

Before procced with regression analysis, it is crucial to conduct a correlation analysis to assess the relationship among variables. The degree of relationship between two variables is examine through correlation analysis. The most frequently used methods are Pearson's Product Moment Correlation Coefficient and Spearman's Rank Correlation Coefficient [15]. In this study, the relationship among variables was analysed using Pearson's Product Moment Correlation Coefficient. SPSS was used to analysed the data and obtain the value for the correlation. A common measure that assesses the linear correlation between two variables is the Pearson correlation coefficient (PCC), which ranges from -1 to 1. where 0 represents no correlation, 1 represents a perfect positive correlation, and -1 represents a perfect negative correlation. [16]. The appropriate spectrum for interpreting the correlation coefficient's value is showcased in Figure 5. Regression analysis was conducted after verifying the correlation and ensuring the satisfaction of four key assumptions which is linearity, homoscedasticity, normality, and absence of multicollinearity.
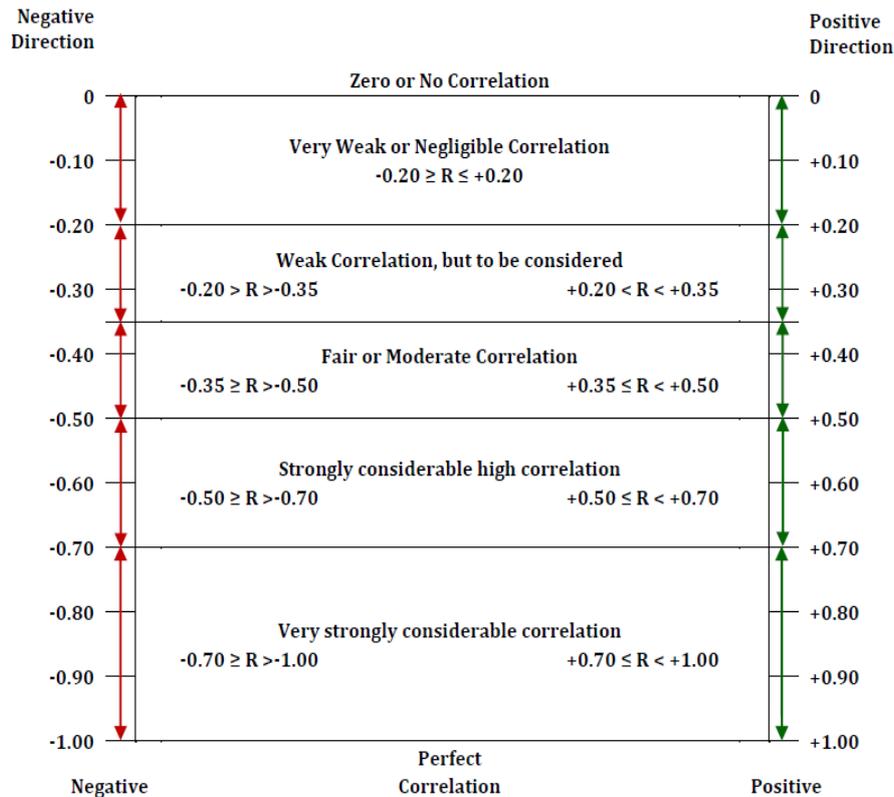
Fig. 5 *Ideal spectrum of interpreting correlation coefficient*

2.4.2 Assumption

The dataset undergoes a thorough examination to verify the assumptions of linearity, homoscedasticity, normality, and multicollinearity before conducting regression analysis. For prediction model to be reliable and valid, particular assumptions are essential.

First, scatterplots were used to represent the relationship between dependent variables and the independent variables before conducting regression analysis to ascertain linearity. Linear regression requires a linear relationship between the independent and dependent variables to ensure the model's reliability. Checking for outliers is crucial since linear regression is sensitive to their effects. Formal tests such as the Durbin-Watson test are conducted to validate the linear relationship.

The following factor to consider is homoscedasticity. This assumption helps assess whether the data demonstrates homoscedasticity by examining the scatter plot for a uniform distribution of residuals along the regression line. Essentially, homoscedasticity refers to the assumption that the variance of the errors in a regression model remains consistent across all levels of the independent variable(s). Residual plots are generated to assess homoscedasticity, wherein a uniform spread of residuals around the horizontal axis indicates constant variance. Verifying this assumption entails examining residual plots, where a uniform spread of residuals around the horizontal axis signifies homoscedasticity.

Moreover, the normality assumption posits that the residuals follow a normal distribution, facilitating unbiased and efficient parameter estimation. This assumption can be assessed through visual inspection via Q-Q or P-P plots and Shapiro-Wilk tests. Deviations from normality may necessitate data transformation techniques to conform to this assumption.

Equation 3 show formula for Variance Inflation Factor (VIF). Multicollinearity occurs when the correlation among independent variables increases the variability of the estimated regression coefficients. This inflation can be quantified using the VIF. VIF is calculated using SPSS to assess multicollinearity among predictor variables. When the VIF value equals 1, it signifies no correlation among the independent variables. The variables are moderately correlated if the VIF value ranges from 1 to 5. However, when the VIF value falls between 5 and 10, it suggests highly correlated variables. If the VIF value exceeds 10, it indicates weak estimated regression coefficients due to multicollinearity [17].

$$VIF = \frac{1}{1 - R^2} = \frac{1}{Tolerance} \tag{3}$$

Int. J. of Sustainable Construction Engineering and Tech. Vol. 15 No. 3 (2024) p. 192-208

197

2.4.3 R-Squared Value (R^2)

After finishing the assumption analysis, R-squared (R^2) was analysed to see the goodness of fit of the regression model to the observed data. According to Kumari & Yadav, (2018), In statistical analysis, the coefficient of determination, or R^2, is essential because it explains how much changes in the dependent variable are influenced by adjustments in the independent variable or variables. [18]. R^2 ranges between 0-1.0. An R^2 value of 0 indicates no link between variables, while a value near or equal to 1 indicates a link between them [19]. The R-squared value can be obtained through regression analysis using Microsoft Excel and SPSS, which then provide information on how well the regression model predicts traffic volume over a 3 hour period and explains the dependent variable's variability (based on vehicle classes).

2.4.4 Hypothesis

After establishing the foundational assumptions and presenting the single and multiple linear regression equation, the next step involved formulating and testing hypotheses to assess the significance of specific coefficients and overall model fit. For a hypothesis test regarding the general coefficient b, the null hypothesis's general form is H_0, and the alternative hypothesis is H_1.

2.4.5 T-Statistics Test

Then, a test using t-statistics was performed to determine if the regression equation, which includes independent variables ($X_1, X_2, ... X_n$), significantly impacts the dependent variable (Y). The t-value was calculated using Equation 4. After that, a critical t-value was obtained from the t-distribution table (right-tailed) based on the degree of freedom and level of confidence (0.05). The p-value for each variable was analysed from the ANOVA table to assess the significance of the independent variables. A variable is deemed acceptable if its p-value exceeds the specified confidence level of 0.05. The alternative hypothesis is accepted when the t-value exceeds the critical t-value and the p-value is below 0.05 and the null hypothesis is rejected.

$$T = \frac{Coefficient\ estimate}{Standard\ error\ of\ the\ coefficient\ estimate} \tag{4}$$

2.4.6 F-Statistic Test

Equation 5 helps determine the F-statistic, while the ANOVA table provides the mean square, residual mean square, and the F-value. The F-distribution table is referenced to find the critical F-value at a significance level 0.05. A variable is considered acceptable when the significance F value falls below 0.05. If the F value exceeds the critical F value and the significance F value is under 0.05, the null hypothesis is rejected in favour of the alternative hypothesis.

$$F = \frac{Mean\ Square}{Residual\ Mean\ Square} \tag{5}$$

2.4.7 Coefficient

The present study analysed the impact of vehicle count and classes on traffic volume in 3 hours. The study employed two different models, SLR and MLR, for 15 minutes and 30 minutes to evaluate the model's coefficient value of slope (b) and y-intercept (a). The slope (b) in operations refers to the variable's rate of increase due to the influence of other variables. On the contrary, intercept (a) represents the variable's increased size due to the impact of different variables [20]. When considering the effects of the dependent variable, the value of the regression coefficient shows the contribution of each independent variable to the dependent variable. [21].

2.5 Validation Using Mean Absolute Percentage Error (MAPE)

The prediction value was validated against the manual count data results to achieve a low degree of percentage error. One statistic for evaluating prediction accuracy is MAPE. The percentage difference between expected and actual values is calculated. After choosing the best model from the SLR and MLR models, both models were tested using short-term data from manual counting. After that, the results from both prediction models were compared to data collected using manual counting using equation 6. Any discrepancies or errors were analysed to identify areas for improvement. The percentage error was analysed using Table 2.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{V\,ob - V\,p}{V\,ob}\right) \text{X } 100 \qquad (6)$$

where:
n= number of data observed (m)
Vob = volume observed (manual counting data)
Vp = volume predicted (prediction model)

Table 2 *Interpretation of typical MAPE values* [22]

MAPE	Interpretation
< 10	Highly accurate forecasting
10 – 20	Good forecasting
20 -50	Reasonable forecasting
> 50	Inaccurate forecasting

3. Results and Discussion

3.1 Correlations

The study employs a Pearson correlation coefficient methodology to analyse the strength and direction of relationships between various variables and total traffic volume over 3-hour periods, utilizing both 15-minute and 30-minute intervals data. Findings are detailed in Table 3 and Table 4, respectively.

For the 15-minute interval data, significant positive correlations were observed with variables X-15 (0.72), C1 (0.70), LW (lightweight vehicles) (0.72), and PC (passenger cars) (0.70), indicating a very strong influence on total traffic volume in the range of +0.70 ≤ R < +1.00. Conversely, variables C2, C5, and C6 exhibited a strong correlation in the +0.50 ≤ R < +0.70 range, highlighting the relevance but comparatively lesser impact of vans, buses, and motorcycles. In contrast, heavyweight vehicles (HW), a composite of lorries and buses (C3+C4+C5), showed a weak correlation (0.33), emphasizing their limited effect on traffic volume. Variables C3, C4, and Lorries (combined C3 and C4) demonstrated very weak correlations, suggesting minimal impact on total traffic volume.

The 30-minute interval data analysis mirrored these findings, with very strong correlations observed for variables X-30, C1 and PC. Other variables like C2, C5, C6, and LW also showed strong correlations, while HW indicated a weak correlation. C3 and C4, along with the lorries combination, exhibited very weak correlations, underlining their minor role in affecting total traffic volume.

The correlation analysis emphasizes significant relationships between various vehicle classifications and total traffic volume. To gain insight into these relationships, additional analysis is necessary. This involves verifying the assumptions of the regression model to ensure the accuracy and reliability of the results. The upcoming section will outline the key assumptions for conducting a robust regression analysis.

Table 3 *Pearson correlation coefficient for 15-minute interval data*

	Total	X-15	C1	C2	C3	C4	C5	C6	PC	LW	HW	Lorries
Total	1.00	0.72	0.70	0.62	0.09	-0.10	0.60	0.64	0.70	0.72	0.33	0.06
X-15	0.72	1.00	0.97	0.93	0.09	-0.13	0.62	0.85	0.98	1.00	0.33	0.06
C1	0.70	0.97	1.00	0.90	0.00	-0.13	0.54	0.71	0.99	0.97	0.22	-0.03
C2	0.62	0.93	0.90	1.00	-0.08	-0.17	0.56	0.66	0.96	0.93	0.16	-0.11
C3	0.09	0.09	0.00	-0.08	1.00	0.24	0.07	0.30	-0.03	0.07	0.89	0.98
C4	-0.10	-0.13	-0.13	-0.17	0.24	1.00	-0.15	-0.08	-0.14	-0.13	0.30	0.43
C5	0.60	0.62	0.54	0.56	0.07	-0.15	1.00	0.61	0.56	0.61	0.49	0.04
C6	0.64	0.85	0.71	0.66	0.30	-0.08	0.61	1.00	0.71	0.84	0.51	0.27
C12	0.69	0.98	0.99	0.96	-0.03	-0.14	0.56	0.71	1.00	0.98	0.21	-0.06
LW	0.72	1.00	0.97	0.93	0.07	-0.13	0.61	0.84	0.98	1.00	0.31	0.04
HW	0.33	0.33	0.22	0.16	0.89	0.30	0.49	0.51	0.21	0.31	1.00	0.89
Lorries	0.06	0.06	-0.03	-0.11	0.98	0.43	0.04	0.27	-0.06	0.04	0.89	1.00

Int. J. of Sustainable Construction Engineering and Tech. Vol. 15 No. 3 (2024) p. 192-208

199

Table 4 *Pearson correlation coefficient for 30-minute interval data*

	Total	X-30	C1	C2	C3	C4	C5	C6	PC	LW	HW	Lorries
Total	1.00	0.73	0.71	0.63	0.11	-0.12	0.65	0.65	0.70	0.65	0.37	0.08
X-30	0.73	1.00	0.97	0.93	0.10	-0.15	0.67	0.85	0.98	0.84	0.37	0.06
C1	0.71	0.97	1.00	0.92	0.01	-0.14	0.58	0.73	0.99	0.71	0.26	-0.02
C2	0.63	0.93	0.92	1.00	-0.09	-0.20	0.61	0.67	0.96	0.65	0.18	-0.13
C3	0.11	0.10	0.01	-0.09	1.00	0.26	0.11	0.33	-0.02	0.38	0.88	0.98
C4	-0.12	-0.15	-0.14	-0.20	0.26	1.00	-0.16	-0.09	-0.16	-0.07	0.30	0.44
C5	0.65	0.67	0.58	0.61	0.11	-0.16	1.00	0.67	0.60	0.66	0.54	0.07
C6	0.65	0.85	0.73	0.67	0.33	-0.09	0.67	1.00	0.72	1.00	0.56	0.28
C12	0.70	0.98	0.99	0.96	-0.02	-0.16	0.60	0.72	1.00	0.70	0.24	-0.05
LW	0.65	0.84	0.71	0.65	0.38	-0.07	0.66	1.00	0.70	1.00	0.60	0.34
HW	0.37	0.37	0.26	0.18	0.88	0.30	0.54	0.56	0.24	0.60	1.00	0.88
Lorries	0.08	0.06	-0.02	-0.13	0.98	0.44	0.07	0.28	-0.05	0.34	0.88	1.00

3.2 Assumptions

According to Tranmer et al., (2020), to effectively interpret a model and its limitations, it is crucial to comprehend the assumptions that underlie the model. Additionally, it is important to understand how these assumptions influence the choices regarding data analysis and modelling. Assumptions are made when creating a linear regression model. These assumptions include that the variable is continuous and constant. The relationship between the outcome and independent variables is expected to be linear, and the residuals are expected to be homoscedastic. The residuals are expected to be normally distributed. The model should not have significant multicollinearity, and there should be no external variables with a strong relationship to the dependent variable after controlling for the variables in the model [23]. Therefore, the study conducted a validity check before further evaluation.

The analysis was conducted based on key assumptions on all independent variables to see which variables are significant and compatible. Based on the analysis for single linear regression (SLR), the independent variable (X) that show a good result is traffic volume in the first 15 minutes (X-15) and 30 minutes (X-30). On the other hand, the variables for multiple linear regression (MLR) that were analysed and showed a good result is the combination of vehicles Class 1 and 2 (cars and vans), Class 5 (busses) and Class 6 (motorcycles) for both 15 minutes and 30 minutes model. The result for both SLR and MLR of the particular variable was discussed below.

3.2.1 Linearity

A graph is made to see if the residuals are distributed consistently. It contrasts standardised residual values with those predicted by a conventional normal distribution. By examining the plotted points' alignment with the diagonal line, the distribution can be determined if they are normally distributed or not [24]. Based on the analysis, the relationship between the predictor (Y) and variables (X) in SLR for the 15-minute and 30-minute models is linear as there is no trend, and the spread is random, as shown in Figures 6 (a) and (b). On the other hand, Figures 7 (a) and (b) also show similar results for the MLR 15-minute and 30-minute models as the graph is linear, there is no trend, and the spread is random. Therefore, this study's graph model (SLR 15 and 30 minutes and MLR 15 and 30 minutes) is linear.

3.2.2 Homoscedasticity

Homoscedasticity means equal residual distribution across the regression line, indicating a constant variance for error terms, irrespective of any variable in the model [23]. In other words, if the spread of the residuals in the graph is uniform, then that indicating that the model's accuracy is consistent for all predicted values. Based on Figure 6 and Figure 7, it is assumed that the variance is homogeneous for both SLR and MLR models, as the spread of the residuals is the uniform in any vertical direction. Hence, the scatterplot is homogenous.
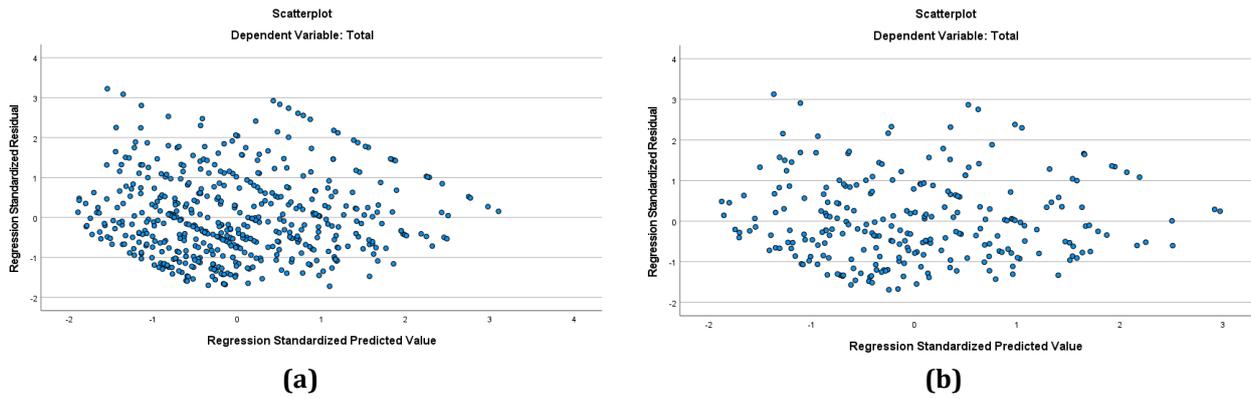
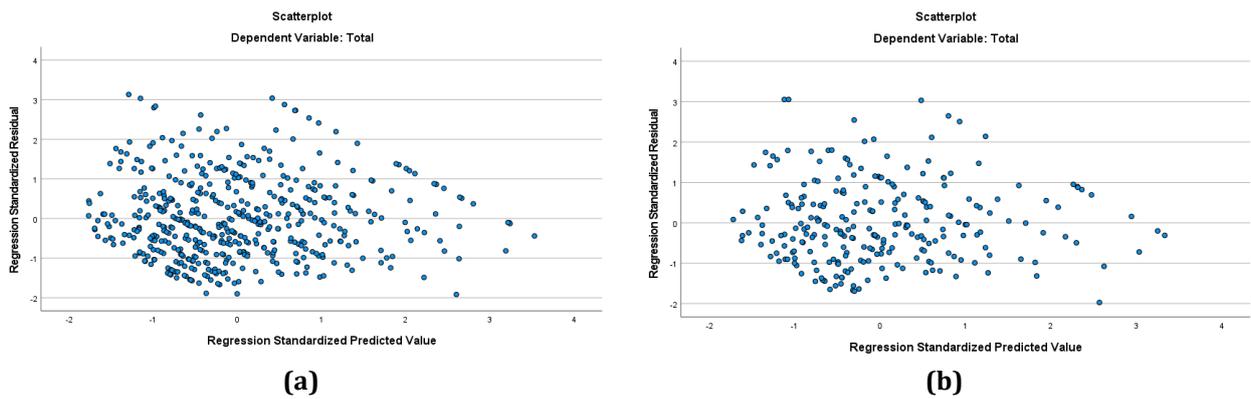**Fig. 6** *SLR Scatter Plot (a) 15 min; (b) 30 min*

Fig. 7 *MLR Scatter Plot (a) 15 min; (b) 30 min*

3.2.3 Normality

A normal probability plot displays sample values about theoretical quantiles. Regression analysis presupposes that the variables follow a normal distribution. It is essential to address skewness and outliers in the data to ensure the validity and reliability of the results. Visual inspection of data plots and using skew and P-P plots are effective methods for assessing data normality to ensure the robustness of the statistical analyses [25]. Based on the analysis, it is assumed that the PP-Plot of the residual data is normally distributed, as illustrated in Figure 8 and Figure 9.
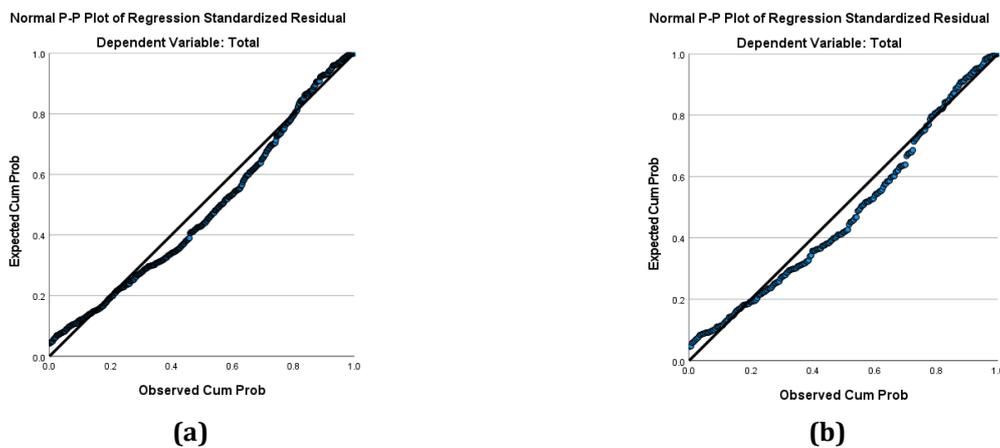

Fig. 8 *SLR Regression line (a) 15 min; (b) 30 min*

Int. J. of Sustainable Construction Engineering and Tech. Vol. 15 No. 3 (2024) p. 192-208
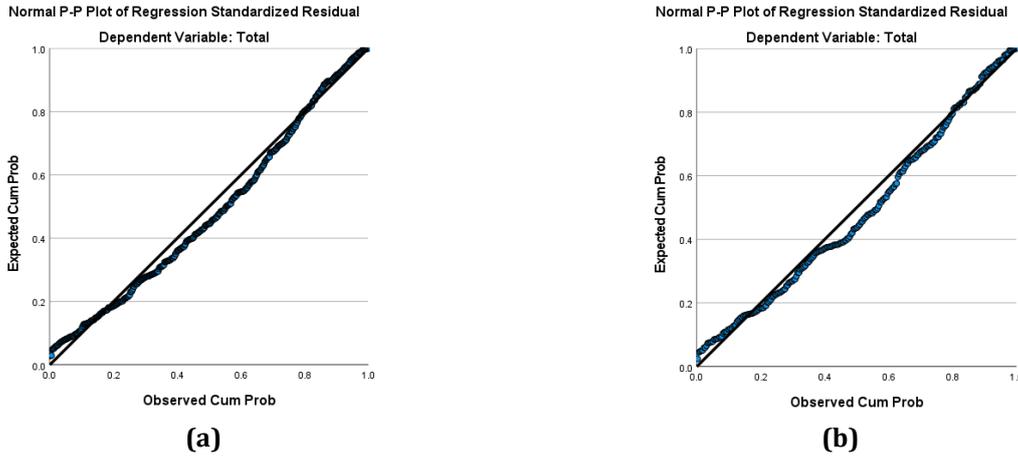
201

Fig. 9 *MLR Regression line (a) 15 min; (b) 30 min*

3.2.4 Multicollinearity

Multicollinearity involves the relationships among variables in MLR analysis. On the other hand, collinearity is defined as the situation where two variables are almost perfect linear combinations of one another. [17]. A linear regression's variance inflation factor (VIF) can be expressed as VIF = 1/T. If the VIF is greater than 5, multicollinearity may present, with VIF values above 10 confirming the same. The variables used in the regression are multicollinear. One of the easiest ways to deal with this issue is by removing independent variables with high VIF values [24]. Based on the analysis of this study, there is no multicollinearity concern as all VIF values are lower than 5.0, as shown in Table 5.

Table 5 *Variance inflation factor (VIF) for MLR model*

Model	Short term	Variable	VIF
MLR	15 min	Class 1+2	2.130
		Class 5	1.676
		Class 6	2.316
	30 min	Class 1+2	2.200
		Class 5	1.913
		Class 6	2.524

3.3 Equation

Based on the analysis from correlation and assumption, the independent variables were chosen and the equations below are the prediction models for SLR and MLR for 15-minute and 30-minute short-term data. The model was chosen based on analysis and evaluation using SPSS and Microsoft Excel software. A detailed explanation was discussed further.

SLR V_{3hr} = 1700.462 + 6.2123($V_{15\ min}$)
SLR V_{3hr} = 1578.0913+3.2591($V_{30\ min}$)

MLR V_{3hr} = 1783.7582 + 4.6724($V_{C1+C2\ 15\ min}$) + 201.996($V_{C5\ 15min}$) + 5.4068($V_{C6\ 15\ min}$)
MLR V_{3hr} = 1728.0320 + 2.3229($V_{C1+C2\ 30\ min}$) + 123.1995($V_{C5\ 30\ min}$) + 2.6000($V_{C6\ 30\ min}$)

3.4 R-squared Value

Table 6 presents the R^2 value for both the SLR for the model using 15- and 30-minute short-term data, while Table 7 presents the R^2 value for both the MLR for the model using 15- and 30-minute short-term data. The R^2 value was calculated using regression analysis in Microsoft Excel. The highest R^2 value for the SLR model is observed to be 0.5177 for a 15-minute duration, with total vehicle as the independent variable. Similarly, for the 30-minute model, the highest R^2 value observed is 0.5416. Moving on to the MLR model, the highest R^2 value for a 15-minute

duration is 0.5863, with C1+C2, C5, and C6 as independent variables. This indicates that the predictors can explain 58.6% of the independent variables. The 30-minute MLR model shows an R^2 value of 0.5798, where C1+C2, C5, and C6 are the independent variables, similar to the previous model.

    After an analysis, it was observed that all the models being considered had a moderate R^2 value. Substantiating this observation, attaining a high R^2 value implies that a significant proportion of the dependent variable can be explained by the independent variables integrated into the model. However, according to Sudin et al., (2021), the R^2 metric should not be a solely reliable indicator of the regression model's adequacy. A model that aligns closely with the data might occasionally exhibit a low R^2 figure, whereas a model with systematic errors could show a high R^2 value. R^2 does not offer insights into the accuracy of biased coefficient prediction and forecasts. It is crucial, instead, to examine the residual plots to evaluate how well the model performs [26]. Therefore, a thorough evaluation is necessary to ensure the model is accurate. This involves testing assumptions such as Normality, Linearity, Homoscedasticity, and Multicollinearity. If these are met, the model is more likely to provide reliable predictions within the range of the observed data.

Table 6 *The independent variable's R^2 for the SLR model*

Model	Short term	Variable	R^2
SLR	15 min	Total vehicle	0.5177
		Class 1	0.4911
		Class 2	0.3807
		Class 3	0.0008
		Class 4	0.0011
		Class 5	0.3600
		Class 6	0.4049
	30 min	Total vehicle	0.5416
		Class 1	0.5078
		Class 2	0.3987
		Class 3	0.0121
		Class 4	0.0136
		Class 5	0.4172
		Class 6	0.4282

Table 7 *The independent variable's R^2 for the MLR model*

Model	Short term	Variable	R^2
MLR	15 min	Intercept	0.5863
		Class 1+2	
		Class 5	
		Class 6	
	30 min	Intercept	0.5798
		Class 1+2	
		Class 5	
		Class 6	

3.5 Hypothesis

For a hypothesis testing, it can be expressed as:

- **SLR:**
  H_0: $b_i = 0$ (The slope coefficient (b_i) is equal to zero, and there is no relationship between the independent variable X and the dependent variable Y*)*
  H_1: $b_i \neq 0$ (The slope coefficient (b_i) is not equal to zero, suggesting a linear relationship between the independent variable (X) and the dependent variable (Y).

Int. J. of Sustainable Construction Engineering and Tech. Vol. 15 No. 3 (2024) p. 192-208

203

- **MLR:**
    H_0: b_n = 0 (All of the regression coefficients (b_n) are equal to zero, indicating that none of the independent variables (X_n) has a significant effect on the dependent variable (Y).
    H_1: $b_n \neq 0$ (At least one of the regression coefficients is not equal to zero, suggesting that at least one independent variable significantly affects the dependent variable).

The hypotheses were established based on two scenarios for the SLR. The null hypothesis (H_0: bi = 0) posits that there is no relationship between the independent variable (X) and the dependent variable (Y), indicated by a slope coefficient of zero. In contrast, the alternative hypothesis (H_1: bi≠0) proposes that a direct linear relationship exists between these two variables.

In the context of MLR, the null hypothesis (H_0: bn = 0) asserts that all regression coefficients are zero, suggesting that none of the independent variables (X_n) has a significant impact on the dependent variable (Y). The alternative hypothesis (H_1: bn ≠ 0) indicates that at least one independent variable has a statistically significant effect on the dependent variable.

T-statistic (t-test) is used to analyse the significance of individual regression coefficients. At the same time, the F-statistic (F-test) was used to evaluate the overall significance of the model. The t-test allowed for the assessment of each coefficient's significance in the regression equation, while the F-test was critical in determining the overall relevance of the regression model.

3.5.1 T-statistic

When estimating a parameter, the t-statistic measures the difference between its expected and hypothesised values relative to its standard error. The t-value test determines if a regression equation using independent variables (X1, X2, ... Xn) significantly affects the dependent variable (Y). The influence of the dependent variable is thought to be affected by the independent variables [27]. According to Mishra et al., (2019), in statistical analysis, the t-value is calculated by dividing the difference between sample means by the standard error. The decision to reject the null hypothesis hinges on comparing this t-value to a critical tabulated value based on degrees of freedom and confidence level. The null hypothesis is rejected if the calculated t-value exceeds the tabulated value. This method ensures rigorous and reliable results [28].

Table 8 shows the result of the t-values obtained in ANOVA analysis for the 15-minute and 30-minute SLR and MLR models. In the SLR 15-minute model, each independent variable's t-value exceeds the critical t-value from the T distribution table, which is 1.6484 for 504 observations (Right-tailed, 0.05 significance level). This indicates a positive influence of the vehicle volume in 15 minutes (X-15 min) on the 3-hour traffic volume. The 30-minute model shows a similar result with each independent variable's t-value > t-critical (1.6515) for 252 observations.

In the MLR model, the 15-minute and 30-minute models also show each independent variable's t-value exceeding the critical value of 1.6484 (504 observations) and 1.6515 (252 observations). Notably, the t-values for both SLR and MLR models at 15-minute and 30-minute surpass the critical t-values provided in the t-table, indicating the need for closer examination of the statistical significance of these results.

Table 8 presents models incorporating variables found to be significant across both SLR and MLR analyses. The result reveals that the p-values of these individual variables fall below the predetermined significance value, set at 0.05. This observation aligns with the argument put forth by Alexopoulos (2010), suggesting that when the p-value surpasses 0.05, the linear relationship between variable X and outcome Y is deemed unreliable for predictive purposes, leading to the retention of the null hypothesis [29]. Both the SLR models contain intercept and total vehicle in 15 and 30 minutes, which are statistically significant, implying that the variable significantly affects the dependent variable (3-hour traffic volume) when evaluating the individual coefficients for each independent variable. Conversely, the other variables, specifically Class 2 and Class 3, do not show statistical significance at the selected significance level since their p-values exceed 0.05.

Similarly, with the MLR model, both 15- and 30-minute models have these independent variables, which is the intercept, a combination of vehicle class 1 and class 2 (Class 1+2), Class 5 and Class 6 are deemed to be statistically significant since the variables with p-values less than or equal to 0.05 are considered to considerably influence the dependent variable, 3-hour traffic volume. The variables with p-values higher than 0.05 are insignificant and not present in the table. The variables are Class 2 and Class 3.

Table 8 *The independent variable's t-value and p-value for the SLR and MLR model*

Model	Short term	Variable	T statistic / t-value	P-value
SLR	15 min	Constant	19.44	<.001
		Total vehicle	23.21	<.001
	30 min	Constant	12.88	<.001
		Total vehicle	17.19	<.001
MLR	15 min	Contant	21.16	<.001
		Class 1+2	9.46	<.001
		Class 5	6.58	<.001
		Class 6	4.19	<.001
	30 min	Constant	14.53	<.001
		Class 1+2	6.56	<.001
		Class 5	5.06	<.001
		Class 6	2.66	0.008

3.5.2 F-statistic

According to Cardinal (2004), the value of mean square regression (MSR) and mean square error (MSE) are critical metrics derived from data summaries. The efficacy of a regression model versus a baseline model is evaluated using an F-statistic test, which depends on the degrees of freedom for MSR and MSE to define the F-statistic distribution. By comparing the F-value against the F-distribution table, the test determines a p-value or significance level based on the degrees of freedom, thereby assessing the model's overall significance [30]. This process is crucial in determining whether the independent variables significantly influence the dependent variable. When the probability value (P) falls below the significance value, it indicates that the regression model with variables better fits the data compared to a model without these variables [24].

Table 9 shows the findings regarding the F-values obtained for both the 15-minute and 30-minute SLR and MLR models. F-value was obtained from the ANOVA table in the analysis. For the SLR 15-minute model, the F-value exceeds the critical F-value from the F distribution table (0.05 significance level), 538.82 > 3.86, suggesting a significant departure from the null hypothesis for SLR 15 minutes. For 30 minutes, a similar result was obtained, F-value> F-critical (295.38 > 3.88). Furthermore, the analysis continues to the MLR model, 15 minutes the model, with F-values surpassing the corresponding critical values (208.98 > 2.62). MLR 30-minute model has similar observations as F-value > F-critical (114.09 > 2.64). Notably, the F-values calculated for both SLR and MLR models at the 15-minute and 30-minute surpass the critical values provided in the F-table, prompting a closer examination of the statistical significance of these results.

Table 9 shows that the regression models employed in SLR and MLR are statistically significant. It can be concluded that these regression models offer a better fit for the dataset when contrasted with a model devoid of any independent variables. The significance F values below the predetermined significance value of 0.05 support this conclusion. Since P < α (0.05) hence, this provided evidence against the null hypothesis that none of the independent variables have an effect. Therefore, H_0 is rejected, and the equation can be utilised to create an estimated model of traffic volumes during peak times using short-term traffic counts.

Table 9 *The independent variable's F-value and Significance F for the SLR and MLR model*

Model	Short term	Variable	F-value	Significance F
SLR	15 min	Total vehicle	538.82	1.62 e-63
	30 min	Total vehicle	295.38	3.09 e-44
MLR	15 min	Class 1+2	208.98	7.78 e-88
		Class 5		
		Class 6		
	30 min	Class 1+2	114.09	1.91 e-46
		Class 5		
		Class 6		

Int. J. of Sustainable Construction Engineering and Tech. Vol. 15 No. 3 (2024) p. 192-208

205

3.6 Coefficients for Independent Variables

The coefficient's value determines the direction and size of the change in the predicted Y value for every unit change in a given X variable. Suppose the coefficient for an independent variable in a linear regression model is positive. In that case, this suggests that as the value of the independent variable goes up, so will the predicted outcome of the dependent variable. On the other hand, a negative coefficient implies that an increase in the independent variable's value will lead to a decrease in the predicted outcome of the dependent variable [21].

The study's findings indicated that vehicle count positively influenced traffic volume in both SLR and MLR models. Based on Table 10, the intercept or constant value in the SLR 15-minute model was 1700.46; in the SLR 30-minute model, it was 1578.09. This indicated that if all independent variables were assumed to be zero, the traffic volume would be 1700 and 1578, respectively. The regression coefficient for the variable of total vehicle for the SLR 15-minute model (X-15min) was 6.21, and for the SLR 30-minute model (X-30min), it was 3.26. This showed that a 1% increase in total vehicles in 15 minutes and 30 minutes would result in a 6.21% and 3.26% increase in traffic volume, respectively.

Similarly, the MLR models included the vehicle class and total vehicle count variables, indicating that different vehicle classes impacted traffic volume differently. Based on Table 11, the intercept or constant value for the MLR 15-minute model was 1783.76, and for the MLR 30-minute model, it was 1728.03. The regression coefficient for the variable of vehicle class of Class 1+Class 2 was 4.67 in the MLR 15-minute model and 2.32 in the MLR 30-minute model. This indicated that a 1% increase in total vehicles of Class 1+Class 2 in 15 minutes and 30 minutes would result in a 4.67% and 2.32% increase in traffic volume, respectively. Similarly, for Class 5, the regression coefficient was 201.99 in the MLR 15-minute model and 123.19 in the MLR 30-minute model, and for Class 6, it was 5.54 in the MLR 15-minute model and 2.60 in the MLR 30-minute model. These values showed that both classes positively affected the traffic volume in 3 hours.

Table 10 *The coefficient for SLR*

Model	Short term	Variable	Coefficient	Sign
SLR	15 min	Intercept Total vehicle	1700.46 6.21	Positive
	30 min	Intercept Total vehicle	1578.09 3.26	Positive

Table 11 *The coefficient for MLR*

Model	Short term	Variable	Coefficient	Sign
MLR	15 min	Intercept	1783.76	Positive
		Class 1+2	4.67	Positive
		Class 5	201.99	Positive
		Class 6	5.54	Positive
	30 min	Intercept	1728.03	Positive
		Class 1+2	2.32	Positive
		Class 5	123.19	Positive
		Class 6	2.60	Positive

3.7 Comparison MAPE for SLR MLR Model

Although the models have a high F-statistic value, which indicates good predictive performance, provided the sampling has been carried out correctly. However, more is needed to guarantee that the model with the highest F-statistic will perform the best on newly introduced data [31]. Besides, the predictive MLR model's analysis reveals an R^2 value of 0.6 and 0.57, which indicates a reasonable value for the capability of the dependent variable. While R^2 is a commonly used metric to evaluate a model's performance, it has certain limitations. According to Montaño Moreno et al., (2013), An alternative metric known as the Mean Absolute Percentage Error (MAPE) is used to understand better a model's capability to predict the dependent variable. MAPE has several valuable features, such as being a reliable, unit-free measure that is easy to interpret and clear to present, supports statistical evaluation, and utilises all information related to the error [22]. The concurrent consideration of R-squared and MAPE enhances the overall understanding of the prediction model's quality and performance.

Based on the result of the MAPE for both the SLR and MLR prediction models in Table 12, it can be seen that the MAPE value for the SLR model for 15 minutes is 19.08%, and for 30 minutes, it is 20.29%. On the other hand,

206

Int. J. of Sustainable Construction Engineering and Tech. Vol. 15 No. 3 (2024) p. 192-208

the MLR model for 15 minutes is 12.09%, and for 30 minutes, it is 18.11%. Generally, a prediction or forecast with a MAPE value of less than 10% can be highly accurate. A value of 11% to 20% can be considered good, while a value of 21% to 50% can be considered reasonable. Any 51% or more value can be considered inaccurate [32]. It can be interpreted that the MAPE value from both SLR models is reasonable. Meanwhile, the MLR model is suitable for prediction and forecasting. Based on this, the MLR model results are much better than the SLR model, while the 15-minute MLR model is more accurate than the 30-minute model. Hence, the MLR model for 15-minute short-term traffic volume was chosen to predict the traffic volume during peak hours.

Table 12 *The Mean Absolute Percentage Error (MAPE) from the analysis*

Short term	SLR	MLR
15 min	19.08	12.09
30 min	20.29	18.11

4. Conclusion

This study successfully developed prediction models for peak-hour traffic volume using short-term traffic data. The goal was to avoid the time-consuming and costly labour of the manual data collection process. The method used is statistical regression analysis to predict traffic volume for 3 hours solely based on short-term traffic data. The development process involves the identification of the most relevant predictors or input variables. Other than that, the prediction models were analysed based on pre-determined (0.05) significance levels, and one model was selected for estimating the total traffic volume at peak hours. The study results show that the peak hour road traffic volume prediction model based on a 15-minute short-term traffic count using MLR analysis is the best model to use to predict traffic volume in Putrajaya. This model was released successfully based on the regression analysis results of the parameters determined in advance. The analysis findings also indicate that the model has 12.09% MAPE, which closely approximates actual traffic volume data. The model's precision means that it successfully captures the underlying traffic trends and patterns, which can provide valuable insights for transportation planning and management. Although there is room for improvement, the model's ability to predict traffic volume with a relatively low margin of error demonstrates its practical usefulness in real-world scenarios. The data collected through this model can be utilised to create effective strategies to enhance traffic flow and safety, such as expanding the roadway or introducing a roundabout, which can help to reduce congestion, minimize accidents and save time and money.

Acknowledgement

Conflict of Interest

Authors declare that there is no conflict of interest regarding the publication of the paper.

Author Contribution

*The authors confirm contribution to the paper as follows: **study conception and design:** Syahrin Azhar, Hilton Ahmad, Sugiman Sugiman; **data collection:** Syahrin Azhar, Subashini Velayutham; **analysis and interpretation of results:** Syahrin Azhar, Sugiman Sugiman, Hilton Ahmad, Demond Daniel Chin; **draft manuscript preparation:** Syahrin Azhar, Subashini Velayutham, Hilton Ahmad. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Paľo, J., Caban, J., Kiktová, M., & Černický. (2019). The comparison of automatic traffic counting and manual traffic counting. *IOP Conference Series: Materials Science and Engineering, 710(1).* https://doi.org/10.1088/1757-899X/710/1/012041

[2] Ali, M., Manogaran, S., Yusof, K. M., & Suhaili, M. R. M. (2018). Analysing vehicular congestion scenario in Kuala Lumpur using open traffic. *Indonesian Journal of Electrical Engineering and Computer Science, 10(3)*, 875–882. https://doi.org/10.11591/ijeecs.v10.i3.pp875-882

Int. J. of Sustainable Construction Engineering and Tech. Vol. 15 No. 3 (2024) p. 192-208

207

[3]   Li, S., Chang, F., Liu, C., & Li, N. (2020). Vehicle counting and traffic flow parameter estimation for dense traffic scenes. *IET Intelligent Transport Systems, 14(12)*, 1517–1523. https://doi.org/10.1049/iet-its.2019.0521

[4]   Prudhvi, M., R., Trinadh, T., A., Sai, C., Reddy, M., Sai, P., U., & Akshith. (2019). Traffic Volume Study using Video Surveillance Technique. *International Journal of Latest Engineering and Management Research (IJLEMR) ||, 04(10)*, 105–118.

[5]   Rashid, M. T., Hossain, A., & Hussain, M. N. (2017). Automated traffic measurement system based on FCD and image processing. In *TENCON 2017-2017 IEEE Region 10 Conference (pp. 1381-1385). IEEE.*

[6]   Zheng, P., & Mike, M. (2012). An Investigation on the Manual Traffic Count Accuracy. *Procedia - Social and Behavioral Sciences, 43*, 226–231. https://doi.org/10.1016/j.sbspro.2012.04.095

[7]   Perera, P., Herath, O. K., Ayesha, B., Thillaiampalam, S., & Kumarage, A. S. (2021). Evaluation of Manual and Video-Based Automated Classified Vehicle Counting Methods for Heterogeneous Traffic Flow. In *Proceedings of the Eastern Asia Society for Transportation Studies (13).*

[8]   Atluri, M., Chowdhury, M., Fries, R., Sarasua, W., Ogle, J., & Kanhere, N. (2007). Development of a Sensor System for Traffic Data Collection. *Journal of Advanced Transportation, 43(1)*, 1–20. https://doi.org/10.1002/atr.5670430102

[9]   Valavanis, K. P., & Vachtsevanos, G. J. (2015). Handbook of unmanned aerial vehicles. *In Handbook of Unmanned Aerial Vehicles*. Springer Netherlands. https://doi.org/10.1007/978-90-481-9707-1

[10]   Ranquist, E., Steiner, M., & Argrow, B. (2017). Exploring the range of weather impacts on UAS operations. In *18th Conference on Aviation, Range and Aerospace Meteorology, Seattle, WA.*

[11]   Lippi, M., Bertini, M., & Frasconi, P. (2013). Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. *IEEE Transactions on Intelligent Transportation Systems, 14(2)*, 871–882. https://doi.org/10.1109/TITS.2013.2247040

[12]   Diao, Z., Zhang, D., Wang, X., Xie, K., He, S., Lu, X., & Li, Y. (2019). A Hybrid Model for Short-Term Traffic Volume Prediction in Massive Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems, 20(3)*, 935–946. https://doi.org/10.1109/TITS.2018.2841800

[13]   Aqib, M., Mehmood, R., Alzahrani, A., Katib, I.A., Albeshri, A.A., & Altowaijri, S.M. (2019). Smarter Traffic Prediction Using Big Data, In-Memory Computing, Deep Learning and GPUs. *Sensors (Basel, Switzerland), 19*. https://doi.org/10.3390/s19092206

[14]   Schneider, A., Hommel, G., & Blettner, M. (2010). Linear regressions analyse, *Deutsches Ärzteblatt International*, *107(44)*, 776–782. https://doi.org/10.3238/arztebl.2010.0776

[15]   Senthilnathan, S. (2019). Usefulness of Correlation Analysis. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3416918

[16]   Schober, P., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia, 126(5)*, 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

[17]   Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, *8(2)*, 39–42. https://doi.org/10.12691/ajams-8-2-1

[18]   Kumari, K., & Yadav, S. (2018). Linear regression analysis study. *Journal of the Practice of Cardiovascular Sciences*, *4(1)*, 33. https://doi.org/10.4103/jpcs.jpcs_8_18

[19]   Manan, W. N. (2011). Accident Prediction Model at Un-Signalized Intersections Using Multiple Regression Method, *Faculty of Civil and Environmental Engineering, University Tun Hussein Onn Malaysia Batu Pahat, Johor*.

[20]   Siswati, B. H., & Corebima, A. D. (2020). The correlation regression equations between metacognitive skills and concept gaining of science and biology of Indonesian students. *Journal of Physics: Conference Series, 1465*, 012043. https://doi.org/10.1088/1742-6596/1465/1/012043

[21]   Roberts, A., & Roberts, J. M. (2020). *Multiple Regression: A Practical Introduction*. SAGE Publications, Incorporated.

[22]   Montaño Moreno, J. J., Palmer Pol, A., Sesé Abad, A., & Cajal Blasco, B. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*, *25(4)*, 500–506. https://doi.org/10.7334/psicothema2013.23

[23]   Tranmer, M., Murphy, J., Elliot, M., & Pampaka, M. (2020). *Multiple Linear Regression (2nd Edition)*.

[24]   Sureiman, O., & Mangera, C. M. (2020). F-test of overall significance in regression analysis simplified. *Journal of the Practice of Cardiovascular Sciences*, *6(2)*, 116. https://doi.org/10.4103/jpcs.jpcs_18_20

[25]   Osborne, J. W., & Waters, E. (2002). Four Assumptions of Multiple Regression That Researchers Should Always Test. *Practical Assessment, Research & Evaluation 8(2)*, 2. https://doi.org/10.7275/r222-hv23

[26]   Sudin, S., Aziz, A. N. A., Saad, F. S. A., Khalid, N. S., & Ibrahim, I. I. (2021). Cycling performance prediction based on cadence analysis by using multiple regression. *Journal of Physics: Conference Series*, *2107(1)*. https://doi.org/10.1088/1742-6596/2107/1/012058

[27] Al-kassab, M. (2022). The Use of One Sample t-Test in the Real Data. *Journal of Advances in Mathematics*, *21*, 134–138. https://doi.org/10.24297/jam.v21i.9279

[28] Mishra, P., Singh, U., Pandey, C. M., Mishra, P., & Pandey, G. (2019). Application of student's t-test, analysis of variance, and covariance. *Annals of Cardiac Anaesthesia*, *22(4)*, 407–411. https://doi.org/10.4103/aca.aca_94_19

[29] Alexopoulos, E. (2010). Introduction to Multivariate Regression Analysis. *Department of Public Health, Medical School, Hippokratia, 14(1)*, 23–28.

[30] Cardinal, R. N., & Aitken, M. R. F. (2005). *ANOVA for the Behavioral Sciences Researcher*. Taylor & Francis. https://doi.org/10.4324/9780203763933

[31] Ivanovski, T., Zhang, G., Jemri´c, T. J., & Guli´cguli´c, M. (2020). Fruit firmness prediction using multiple linear regression. *Communication and Electronic Technology (MIPRO)*, 1306–1311. https://doi.org/10.23919/MIPRO48935.2020.9245213

[32] Omkar, G., & Vasantha Kumar, S. (2018). Finite difference formulation of Lighthill Whitham Richards macroscopic model for traffic flow prediction. *International Journal of Applied Mathematics, 31(5)*, 603–612. https://doi.org/10.12732/ijam.v31i5.6