Far East Journal of Mathematical Sciences (FJMS)
© 2026 Pushpa Publishing House, Prayagraj, India
https://pphmjopenaccess.com
https://doi.org/10.17654/0972087126033 P-ISSN: 0972-0871
Volume 143, Number 2, February 2026, Pages 563-578 E-ISSN: 2584-1246

EVALUATING PERFORMANCE OF THE SARIMA MODEL IN FORECASTING DAILY KTM RIDERSHIP TRENDS

Suhaila Bahrom[1,2], **Noratikah Abu**[2,*], **Nurul Amira Zainal**[3] **and Nur Haizum Abd Rahman**[2]

[1]Department of Mathematics
Centre for Foundation Studies
International Islamic University Malaysia
26300 Gambang, Pahang, Malaysia

[2]Centre for Mathematical Sciences
Universiti Malaysia Pahang Al-Sultan Abdullah
26300 Kuantan, Pahang, Malaysia

[3]Fakulti Teknologi dan Kejuruteraan Mekanikal
Universiti Teknikal Malaysia Melaka
Hang Tuah Jaya, 76100 Durian Tunggal
Melaka, Malaysia

Received: September 21, 2025; Accepted: November 4, 2025
2020 Mathematics Subject Classification: 62M10, 91B74, 91B84.
Keywords and phrases: KTM Komuter, public transportation, ridership, time series analysis.
[*]Corresponding author

How to cite this article: Suhaila Bahrom, Noratikah Abu, Nurul Amira Zainal and Nur Haizum Abd Rahman, Evaluating performance of the SARIMA model in forecasting daily KTM ridership trends, Far East Journal of Mathematical Sciences (FJMS) 143(2) (2026), 563-578.
https://doi.org/10.17654/0972087126033

Published Online: January 6, 2026

Abstract

Accurate ridership forecasting is crucial for optimizing public transportation operations, including scheduling, capacity planning, and resource allocation. KTM Komuter, one of Malaysia's primary rail services, experiences fluctuations in daily ridership due to factors such as peak-hour demand, weekends, public holidays, and economic conditions. This study aims to evaluate the effectiveness of the Seasonal Autoregressive Integrated Moving Average (SARIMA) model in forecasting the daily KTM Komuter ridership. The dataset obtained from Malaysia's official open data portal, spans from the period of October 1, 2023 to January 31, 2025, captures daily trip counts across KTM Komuter stations. The analysis involved exploratory data analysis, stationarity testing, diagnostic checking, and SARIMA modeling to identify the optimal model. The results indicate that the SARIMA$(0, 0, 2)(0, 1, 2)_{12}$ model successfully captures ridership patterns, achieving a mean absolute percentage error (MAPE) of 9.17%, thereby, demonstrating reliable forecasting accuracy. The findings highlight SARIMA's potential in improving train scheduling, capacity planning, and resource allocation. However, the model's reliance on historical data may limit its adaptability to sudden disruptions, such as service interruptions or external economic shifts. Future research should consider integrating external factors, such as weather conditions and macroeconomic indicators, or exploring advanced machine learning models to enhance predictive accuracy and adaptability.

1. Introduction

Public transportation plays a crucial role in urban life, providing a more sustainable and affordable alternative to private vehicles. Accurate ridership forecasting is essential to optimize train schedules, allocate resources effectively, and enhance the overall passenger experience. In Malaysia, Keretapi Tanah Melayu Berhad (KTMB) manages the Keretapi Tanah Melayu (KTM) Komuter system, which was established in 1995 to offer local rail services in Kuala Lumpur and its surrounding suburban areas within the Klang Valley. The Klang Valley Komuter Line features two main routes, Tanjung Malim to Port Klang and Batu Caves to Pulau Sebang (Tampin). Both routes experience heavy ridership, especially during peak hours. Meanwhile, the Northern Komuter Line, introduced in September 2015, operates between Padang Besar and Butterworth and Padang Rengas to Bukit Mertajam, catering to both daily commuters and long-distance travelers [1]. By connecting residential areas with business hubs and major transit points, KTM Komuter plays a vital role in keeping Malaysia moving forward.

Ridership patterns in KTM Komuter services show clear differences between weekdays and weekends, indicating varied commuter behaviors. Weekday ridership mainly includes office workers, students, and commuters, peaking during morning and evening rush hours. In contrast, weekend ridership is influenced by leisure activities, shopping, and tourism-related travel. Ridership can experience additional fluctuations due to public holidays, school vacations, and economic conditions. Given these dynamic factors, developing a reliable forecasting model for daily KTM Komuter ridership is essential to mitigate congestion, enhance operational efficiency, and support strategic transportation planning. However, forecasting ridership remains challenging due to the seasonal, non-stationary, and dynamic nature of transit demand [2].

Recent studies have demonstrated the efficacy of autoregressive integrated moving-average (ARIMA) models in forecasting public transportation ridership. For instance, reference [3] employed ARIMA to predict metro passenger traffic in Xi'an, China, highlighting its utility in medium-term forecasting. Similarly, ARIMA was used in Monterey County, California, to forecast ridership impacts of the SURF! Busway and Bus Rapid Transit Project, projecting significant increases in transit usage over two decades [4]. This demonstrated the model's capability not only for long-term forecasts but also for assessing short-term demand. Reference [5] further assessed the forecasting power of ARIMA models for intercity rail demand, finding them effective in predicting daily passenger flows. Furthermore, ARIMA has been applied to post-pandemic transit ridership forecasting, helping public transit agencies adapt to fluctuating demand patterns [6]. A recent study by [7] demonstrated that ARIMA is effective in forecasting short-term passenger flow for urban rail transit, making it a valuable tool for optimizing metro operations and managing congestion. While ARIMA has been proven to be effective, it may not adequately account for seasonal variations in ridership, which are often influenced by weather, holidays, and work schedules.

Given the significant seasonal fluctuations in transit ridership, researchers have increasingly utilized the seasonal ARIMA (SARIMA) model, which is an extension of ARIMA that includes seasonal components. SARIMA has been widely applied to urban rail and bus networks, demonstrating superior performance in capturing periodic ridership patterns. For instance, a study in Dubai utilized the SARIMA model to predict metro ridership, confirming its effectiveness in modeling seasonal variations and long-term trends [8]. Another study used SARIMA to forecast passenger demand for train services between Surabaya and Jakarta, demonstrating its effectiveness in identifying seasonal peaks, especially during holiday periods when ridership demand significantly increases [9].

Furthermore, SARIMA has been applied to forecast passenger demand in Light Rail Transit (LRT) systems, where ridership exhibits strong seasonal variations due to holiday periods and peak-hour travel demand. A study on Palembang's LRT system found that SARIMA effectively predicted ridership trends, supporting mobility planning and optimizing transit operations by enabling authorities to adjust capacity based on demand forecasts [10]. SARIMA has also been used to forecast subway ridership during disruptions such as the COVID-19 pandemic, with a study on New York City's subway showing its effectiveness in capturing shifts in demand, especially when combined with data-driven change-point detection algorithms [11].

In Malaysia, SARIMA has been used to forecast monthly passenger ridership on the Ampang Line LRT, with results confirming its effectiveness in predicting ridership patterns [12]. Similarly, recent research comparing

SARIMA and the prophet method in forecasting public transportation demand found that SARIMA outperformed Prophet by achieving a lower error [13]. Studies in West Sumatra and Java, Indonesia, also confirm SARIMA's effectiveness in forecasting railway passenger demand by accurately capturing seasonal fluctuations and predicting peak ridership during festive seasons [14, 15]. This highlights SARIMA's ability to assist railway operators in anticipating demand surges and optimizing service capacity.

2. Literature Review

Recent advancements have increasingly utilized SARIMA to improve forecasting accuracy in transit systems. For example, reference [16] applied the SARIMA model using multi-source data collected via Internet of Things (IoT) devices and sensor networks to predict short-term urban rail transit passenger flows in Beijing. Their study showed that SARIMA effectively captured the seasonal and periodic variations in passenger traffic, highlighting its strength in modeling the linear and seasonal patterns of urban rail transit data.

Other approaches have explored the integration of decomposition techniques and transfer learning to enhance the accuracy of short-term passenger flow forecasting, particularly during periods of irregular demand fluctuations. For instance, a decomposition-based forecasting model that incorporates transfer learning is proposed to improve railway passenger flow predictions during holidays [17]. The study highlighted that traditional models often struggle to capture seasonal variations and holiday-related demand shifts, resulting in suboptimal forecasting performance. The findings suggest that the proposed model significantly outperformed conventional forecasting methods. Chuwang and Chen [18] developed a time series-based forecasting model to predict daily and weekly passenger demand at urban rail transit stations, emphasizing the importance of structured temporal dependencies in improving prediction accuracy.

Recent advancements in deep learning have introduced more complex models, such as Long Short-Term Memory (LSTM) networks, Convolutional

Neural Networks (CNNs), and Deep Belief Networks (DBNs), which have been applied to transit forecasting [19]. These models are capable of capturing nonlinear relationships and external influencing factors such as weather conditions, economic fluctuations, and urban mobility patterns. For instance, Liu and Chen [20] developed an LSTM-based model for short-term metro ridership prediction, demonstrating its ability to learn from historical patterns. Moreover, studies have introduced sequence-to-sequence learning and attention mechanisms to enhance short-term ridership forecasting in metro systems [21]. Despite their strong predictive capabilities, deep learning models require large datasets, significant computational resources, and extensive hyperparameter tuning. This dependence may limit their practicality in certain ridership forecasting applications, especially when data availability is constrained.

Due to the limitations, this study utilized the SARIMA model as the main approach for forecasting. However, its effectiveness in predicting daily ridership trends for KTM Komuter has not been thoroughly investigated. Therefore, this study proposes to forecast daily ridership for KTM Komuter in Malaysia using the SARIMA model and to evaluate its predictive performance by applying standard error metrics, including mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE).

The paper is organized as follows: the first section presents the introduction, including the problem statement, background of the research, and a review of relevant literature. The subsequent sections present the methodology, followed by the findings, and conclude with key insights and directions for future research.

3. Materials and Methods

3.1. Data collection

This study utilized a dataset from Malaysia's official open data portal (https://data.gov.my). The dataset contains daily trip counts, referred to as ridership, for the KTM Komuter service from 1 October 2023 to 31 January

2025. The dataset was divided into a training set (October 1, 2023, to December 13, 2024) and a testing set (December 14, 2024, to January 31, 2025) for model evaluation using Python. The data were pre-processed to check for inconsistencies and outliers, and no missing values were identified.

3.2. Seasonal autoregressive integrated moving average (SARIMA) model

The SARIMA model is an extension of the Autoregressive Integrated Moving Average (ARIMA) model, designed to incorporate both trend and seasonal components in time series forecasting. The SARIMA model effectively captures seasonal effects, long-term trends, periodic changes, and random disturbances in ridership data. The general form of the SARIMA $(p, d, q)(P, D, Q)_{12}$ can be expressed as equation (1):

$$\phi_P(B^s)\varphi_p(B)\nabla_s^D\nabla^d y_t = \Theta_Q(B^s)\theta_q(B^q)a_t, \tag{1}$$

where p, d, q are the non-seasonal components, and P, D, Q are the seasonal components. The p and q parameters refer to the order of the non-seasonal autoregressive (AR) and moving average (MA) terms, respectively, while d represents the degree of differencing required to make the time series stationary. Similarly, P, D, and Q represent the seasonal autoregressive order, seasonal differencing order, and seasonal moving average order, respectively. The parameter s denotes the seasonal period length, which is set based on the observed seasonality in the dataset; meanwhile, a_t represents the random error term.

In time series modelling, SARIMA follows four key steps: model identification, parameter estimation, diagnostic checking, and forecasting. The first step involved checking for stationarity using the Augmented Dickey-Fuller (ADF) test [22]. If the test indicated non-stationarity, then differencing was applied to the time series to remove trends and stabilize the variance.

Next, we focus on analyzing autocorrelation function (ACF) and partial autocorrelation function (PACF) plots to determine the appropriate values for p, q, P, and Q. These parameters were further refined based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values, where the model with the lowest AIC and BIC scores was considered optimal for prediction accuracy.

The third step in SARIMA model development involved diagnostic checking to ensure the adequacy of the fitted model. The Ljung-Box Q test was applied to assess whether the residuals exhibit white noise behavior, confirming the absence of significant autocorrelation.

The final step was forecasting, where the validated SARIMA model was used to predict future values based on historical data. The accuracy of the forecast was assessed using metrics like mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean squared error (RMSE) to measure the difference between predicted and actual values. Prediction intervals were also constructed to account for uncertainty, providing a range within which future observations are expected to fall. Figure 1 summarizes the SARIMA modelling process and evaluates its performance in forecasting daily KTM ridership.

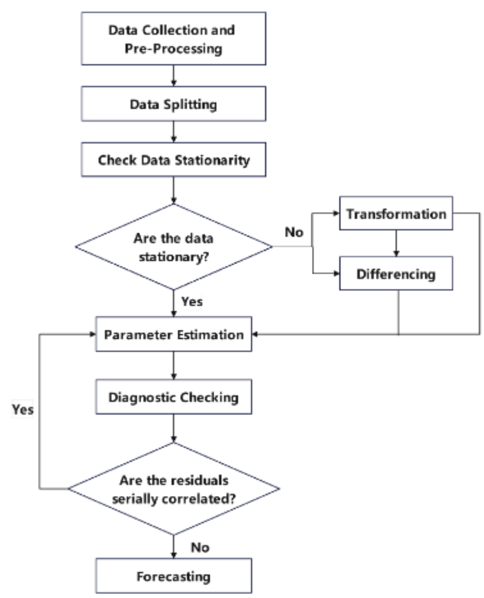

Figure 1. Flowchart of SARIMA modelling.

3.3. Evaluation metrics

Three error metrics; MAPE, MAE and RMSE, were computed to assess the forecasting performance of the SARIMA model. These metrics quantify the deviations between predicted and actual ridership values and are defined in equations (2) to (4):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100, \tag{2}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}| y_t - \hat{y}_t |, \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}, \tag{4}$$

where y_t represents the actual ridership, $\hat{y}_t$ represents the predicted ridership, and n is the number of observations in predicted and actual data by identifying the best model with the smallest error. MAE and RMSE are commonly used scale-dependent metrics and are sensitive to the data's magnitude [23], whereas MAPE is unit-free and widely adopted for comparing performance across multiple time series [24]. Lower values for MAPE, MAE, and RMSE indicate better predictive performance. These metrics provide an objective evaluation of the SARIMA model's ability to accurately capture ridership trends and fluctuations.

4. Results and Discussion

The descriptive statistics of the ridership data, including mean, standard deviation, minimum, and maximum values, are summarized in Table 1. The analysis provides insights into the overall distribution and variability of daily ridership, ensuring a comprehensive understanding of transit demand patterns over time. The results indicate an average daily ridership of 35,288 passengers, with a standard deviation of 6,594, suggesting moderate variability, while the minimum and maximum ridership values are 20,236 and 89,320, respectively.

Table 1. The daily ridership KTM Komuter from 1st October 2023 to 31st January 2025

Mean	Std. dev.	Minimum	Maximum
35288.29	6594.69	20236	89320

The time series plot in Figure 2 illustrates the fluctuations in ridership over this period, with the *x*-axis representing the date and the *y*-axis denoting ridership. The daily ridership exhibits a seasonal trend, with higher passenger volumes on weekdays compared to weekends. Although there is no clear trend of increasing or decreasing ridership over time, the data shows consistent fluctuations, suggesting possible seasonal effects.

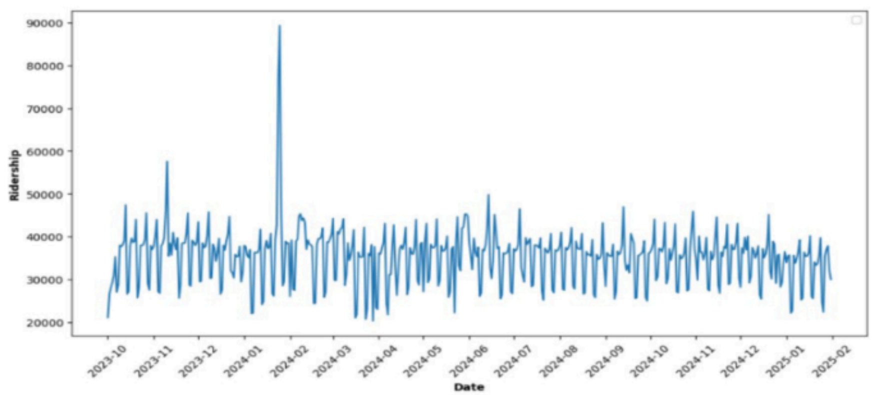

Figure 2. Daily KTM Komuter ridership from 1st October 2023 to 31st January 2025.

Based on Figure 2, irregular fluctuations are observed in daily passenger volumes, with some days experiencing sharp increases or decreases. The most significant spike in ridership appears in early January 2024, which may be attributed to a major event, a public holiday, or a temporary service adjustment. Apart from this anomaly, the data maintains a generally stable fluctuation pattern, reinforcing the possibility of underlying seasonal trends.

The data was then split into training and testing based on a ratio of 90:10 for model development and validation, respectively. The training set comprised 440 observations, covering the period from 1st October 2023 to 13th December 2024. Meanwhile, the testing set spanned from 14th December 2024 to 31st January 2025, and included 49 observations. Before modeling, the dataset was examined for stationarity, as non-stationary data can affect the accuracy of forecasting models. The stationarity of the time series data was assessed using the Box-Cox transformation, where the estimated Box-Cox parameter was $\lambda = 0.2202$, with a 95% confidence interval of [0.1654, 0.3427], indicating that the data on daily KTM Komuter ridership was not stationary in variance. As a result, the natural logarithm was used to stabilize the variance and improve model performance. The ADF test on the log-transformed series confirmed stationarity in mean, with a statistic of –2.8684 and a p-value of 0.000. This conclusion is further supported by the ACF and PACF plots in Figure 3. Based on Figure 3, the ACF plot exhibits a rapid decay, while the PACF plot shows a significant cut-off at lower lags, indicating that the data is stationary in both mean and variance. Therefore, no further non-seasonal differencing is required $(d = 0)$. However, seasonal differencing may still be necessary if seasonal patterns are present in the data, requiring an appropriate seasonal differencing order (D) to ensure stationarity in the seasonal component.

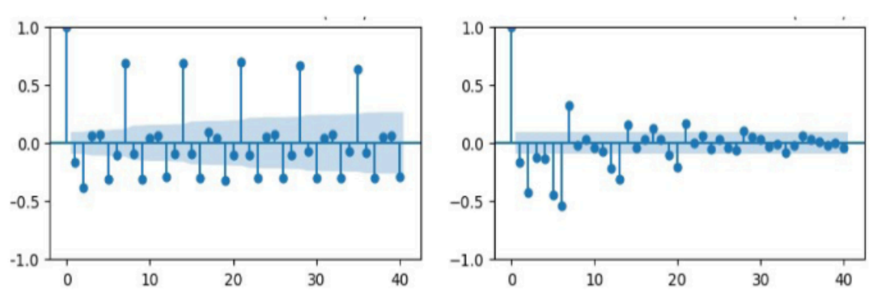

Figure 3. ACF and PACF plots after transformation.

After confirming that the data is stationary, differencing was applied to remove seasonality. Seasonal differencing was performed with the seasonal period set to $s = 7$ accounting for the weekly pattern in the data. The SARIMA model was then specified as $\text{SARIMA}(p, d, q)(P, D, Q)_s$ to serve as the predictive model for daily KTM Komuter ridership. The final model selection was based on the lowest AIC or BIC to optimize forecasting performance. The application of the auto.arima function with seasonal components in Python identified $\text{SARIMA}(0, 0, 2)(0, 1, 2)_7$ as the optimal model for forecasting KTM Komuter ridership. The Ljung-Box test for the $\text{SARIMA}(0, 0, 2)(0, 1, 2)_7$ yielded p-values above 0.05, indicating no evidence of autocorrelation, as shown in Figure 4.

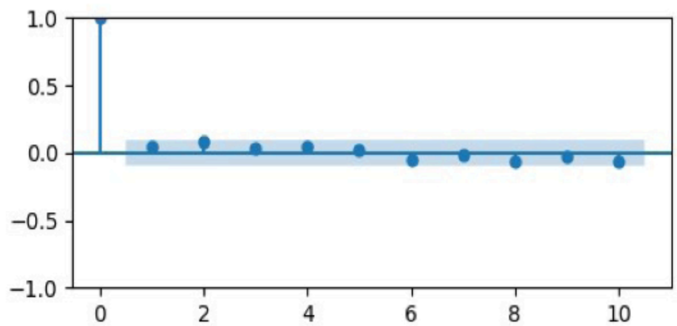

Figure 4. ACF plot of the residuals from $\text{SARIMA}(0, 0, 2)(0, 1, 2)_7$.

The prediction error indicators were computed and summarized in Table 2. The $\text{MAE} = 2806.50$ represents the average absolute difference between the actual and predicted values for the testing dataset. Meanwhile, MAPE quantifies the mean of the absolute percentage errors in the forecasts. As a relative measure, MAPE expresses forecast errors as a percentage of actual values, making it an intuitive metric for evaluating prediction accuracy. This measure is particularly useful due to its simplicity in interpreting error magnitude. Figure 5 presents a comparison between the actual and forecasted KTM Komuter ridership using the $\text{SARIMA}(0, 0, 2)(0, 1, 2)_7$ with the model and demonstrates strong predictive performance, with only minimal differences between the forecasted and actual ridership data, resulting in a MAPE of 9.17%. Therefore, the $\text{SARIMA}(0, 0, 2)(0, 1, 2)_7$

model has been identified as an optimal approach for forecasting daily KTM Komuter ridership trends, ensuring reliable and accurate predictions.

Table 2. Calculated forecast error indicator

Model	MAPE	MAE	RMSE
SARIMA $(0, 0, 2)(0, 1, 2)_7$	9.17	2806.50	3576.64

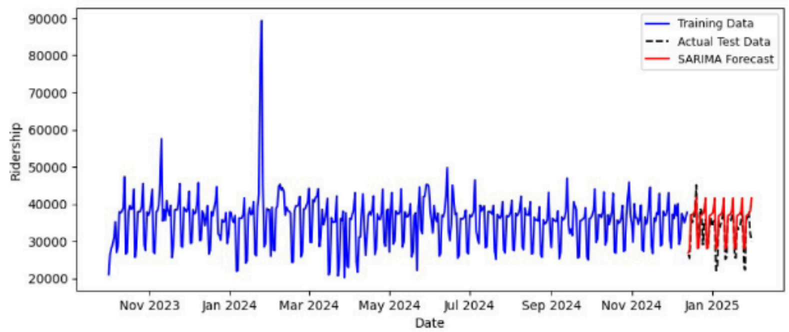

Figure 5. Actual vs forecasting plot of the SARIMA $(0, 0, 2)(0, 1, 2)_7$.

5. Conclusion

The findings of this study address the research objective of evaluating the SARIMA model's suitability for forecasting public transport ridership. The selected SARIMA $(0, 0, 2)(0, 1, 2)_7$ model achieved a low MAPE of 9.17% [25], indicating strong forecasting accuracy. This supports the theoretical basis of time series modeling, particularly the importance of capturing seasonality and stationarity in prediction. The model effectively identifies underlying ridership patterns, enabling KTM Komuter operators to optimize scheduling, adjust service frequency, and allocate resources more efficiently. These forecasts can help improve passenger experience, reduce congestion, and enhance system reliability.

Despite its strengths, the SARIMA model is limited by its reliance on historical data and may not respond well to sudden disruptions such as service failures, economic shifts, or extreme weather. Future research should consider integrating external variables such as fuel prices, road conditions, macroeconomic indicators, or adopting hybrid or machine learning-based models like LSTM to improve adaptability and forecasting performance. Nonetheless, this study demonstrates that SARIMA remains a practical and effective tool for short-term ridership forecasting, offering valuable insights for public transport planning and decision-making.

Acknowledgements

The author sincerely acknowledges the Ministry of Transport Malaysia and Malaysia's official open data portal for providing access to KTM Komuter ridership data and for their invaluable support in facilitating this research. This research was funded by the Research and Innovation Department, Universiti Malaysia Pahang Al-Sultan Abdullah and Geran Padanan Antarabangsa (RDU242705) and Geran Antarabangsa (UIC241506).

References

[1] Keretapi Tanah Melayu Berhad, KTM Komuter, 2025. https://www.ktmb.com.my/komuter.html.

[2] K. Huang, Q. Zhang, C. Zhou, N. Xiong and Y. Qin, An efficient intrusion detection approach for visual sensor networks based on traffic pattern learning, IEEE Trans Syst. Man Cybern. Syst. 47(10) (2017), 2704-2713.

[3] T. Lyu, M. Xu, J. Zhang, Y. Wang, L. Yang and Y. Gao, Influential factor analysis and prediction on initial metro network ridership in Xi'an, China, J. Adv. Transp. 2022 (2022), 1-18.

[4] L. Rheinheimer, M. Overmeyer and E. Patel, Ridership Forecast Report: SURF! Busway and Bus Rapid Transit Project, Monterey, California, 2024.

[5] U. Ekici, H. Tüydes-Yaman and N. Şendil, Testing the forecasting power of statistical models for intercity rail passenger flows in turkey, Transportation Research Record: Journal of the Transportation Research Board 2678(11) (2024), 864-880.

[6] A. Hightower, A. Ziedan, J. Guo, X. Zhu and C. Brakewood, A comparison of time series methods for post-COVID transit ridership forecasting, J. Public Trans. 26 (2024), 100097.

[7] S. Li and Z. Wang, Short-term passenger flow prediction of urban rail transit based on ARIMA model, Applied and Computational Engineering 102(1) (2024), 114-121.

[8] S. A. Alblooshi, M. Masmoudi, A. Cheaitou and K. Hamad, Predicting metro ridership in Dubai: analyzing seasonal trends with SARIMA, Holt-Winters, and LSTM, 2024 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD), IEEE, 2024, pp. 1-7.

[9] S. W. Astuti and Jamaludin, Forecasting Surabaya - Jakarta train passengers with SARIMA model, IOP Conf. Ser. Mater. Sci. Eng. 407 (2018), 012105.

[10] A. Saputra, R. Gutriansyah, A. Sanmorino, Z. R. Mair, D. Sartika and S. Puspasari, Prediction passenger numbers in light rail transit using seasonal autoregressive integrated moving average (SARIMA), Electrotechnical Review 1(10) (2024), 45-47.

[11] B. Moghimi, C. Kamga, A. Safikhani, S. Mudigonda and P. Vicuna, Non-stationary time series model for station-based subway ridership during COVID-19 pandemic: case study of New York city, Transportation Research Record: Journal of the Transportation Research Board 2677(4) (2023), 463-477.

[12] M. Shitan, P. K. Karmokar and N. Y. Lerd, Time series modeling and forecasting of ampang line passenger ridership in Malaysia, Pakistan Journal of Statistics 30(3) (2014), 375-386.

[13] G. Küçüker and Ö. Uzun Araz, Forecasting public transport passenger demand using SARIMA and prophet methods, Endüstri Mühendisliği 34(3) (2023), 334-356.

[14] L. Muthoharoh, Predicting the number of train passengers in Java Island using SARIMA model, EKSAKTA: Journal of Sciences and Data Analysis 4 (2023), 40-48.

[15] L. L. Ibrahim and E. Kurniati, Peramalan jumlah penumpang kereta api eksekutif di pulau jawa menggunakan model SARIMA, Jurnal Riset Matematika (2023), 73-82.

[16] W. Li, L. Sui, M. Zhou and H. Dong, Short-term passenger flow forecast for urban rail transit based on multi-source data, EURASIP J. Wirel Commun. Netw. 2021(1) (2021), 9.

[17] K. Wen, G. Zhao, B. He, J. Ma and H. Zhang, A decomposition-based forecasting method with transfer learning for railway short-term passenger flow in holidays, Expert Syst. Appl. 189 (2022), 116102.

[18] D. D. Chuwang and W. Chen, Forecasting daily and weekly passenger demand for urban rail transit stations based on a time series model approach, Forecasting 4(4) (2022), 904-924.

[19] Y. Danfeng and W. Jing, Subway passenger flow forecasting with multi-station and external factors, IEEE Access 7 (2019), 57415-57423.

[20] L. Liu and R.-C. Chen, A novel passenger flow prediction model using deep learning methods, Transp. Res. Part C Emerg. Technol. 84 (2017), 74-91.

[21] L. Tang, Y. Zhao, J. Cabrera, J. Ma and K. L. Tsui, Forecasting short-term passenger flow: an empirical study on Shenzhen metro, IEEE Transactions on Intelligent Transportation Systems 20(10) (2019), 3613-3622.

[22] S. E. Said and D. A. Dickey, Testing for unit roots in autoregressive-moving average models of unknown order, Biometrika 71(3) (1984), 599-607.

[23] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed., OTexts, 2021.

[24] C. Chatfield, A commentary on error measures, Int. J. Forecast. 8 (1992), 100-102.

[25] C. D. Lewis, Industrial and Business Forecasting Methods, Butterworth Scientific, London, Boston, 1982.