Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Comparative Performance Analysis of Multilayer Perceptron and Long Short-Term Memory for Daily Demand Forecasting in E-Commerce Delivery Platforms

Ica Unari[1]*, Martanto[2], Raditya Danar[3], Ahmad Rifa'i[4], Ryan Hamongan[5]

[1,4,5] *Informatics Engineering, STMIK IKMI CIREBON*
[2,3] *Informatics Management, STMIK IKMI CIREBON*
icanurani94@gmail.com [1*], martantomusijo@gmail.com [2], radith_danar@yahoo.com [3], a.rifaaii1408@gmail.com [4],
mr.ryansilalahi@gmail.com [5*]

Abstract

This study compares the performance of two deep learning architectures—Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM)—for daily demand forecasting on an e-commerce delivery platform. The dataset consists of 1,827 daily observations from 2020 to 2024 and includes operational, temporal, and behavioral features such as holiday indicators, promotion signals, active customers, and delivery time. Data preprocessing includes cleaning, feature engineering, scaling, and sequence generation using a 30-day sliding window. Both models were trained and evaluated using consistent experimental settings and performance metrics. The results show that the LSTM model achieves better accuracy than the MLP model, with an RMSE of 811.81 compared to 830.15, while the difference in MAE between the two models remains minimal. LSTM demonstrates superior capability in capturing temporal dependencies and reacting to rapid demand fluctuations, whereas both models face challenges when predicting sudden demand spikes. These findings indicate that memory-based models such as LSTM are more effective for highly volatile time-series forecasting in e-commerce operations. However, performance can be further improved with the addition of external variables such as real-time promotions, weather conditions, and multivariate features.

Keywords: *Daily demand forecasting, Deep learning, E-commerce, Long Short-Term Memory, Multilayer Perceptron*

1. Introduction

Recent developments in e-commerce logistics have led to significant changes in demand forecasting dynamics, driven by the rapid adoption of technologies such as parcel lockers, micro-fulfillment centers, shared delivery systems, real-time tracking, and environmentally friendly routing strategies. These innovations have increased operational uncertainty and made demand forecasting more complex and less predictable. In addition, growing difficulties in handling product returns and reverse logistics have further challenged traditional forecasting systems. Researchers argue that conventional mathematical models are no longer adequate and recommend transitioning toward data-driven machine learning approaches, particularly Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM), which can incorporate diverse behavioral and operational signals such as website clicks, return activity, delivery choices, and spatial–temporal constraints [1][2][3][4][5].

Recent studies highlight that advanced machine learning systems—especially deep learning architectures such as LSTM and hybrid neural models—offer substantial improvements in forecasting accuracy within logistics environments. These models effectively capture nonlinear relationships, temporal dependencies, and the influence of external factors including weather, promotions, delivery rerouting, and customer behavior. [6]found that tree-based hybrid models outperform LSTM in retail settings, although LSTM remains superior for sequential input patterns. [7]reported that ML/DL methods can reduce forecasting errors by 15–20% compared to classical time-series models. Similarly, deep neural networks have proven effective for irregular and sparse online sales patterns, reinforcing the relevance of MLP, LSTM, and hybrid models for daily demand forecasting in e-commerce delivery platforms [8][9][10] [11].

In contrast, conventional time-series models such as ARIMA, ETS, and Holt–Winters increasingly show limitations in modern online retail environments. These models rely on assumptions of linearity and stationarity, making them less capable of handling volatile, complex, and rapidly changing demand structures. Studies have demonstrated that such models struggle with structural shifts, nonlinear interactions, and evolving seasonal–trend patterns, which restrict their adaptability in high-variability forecasting scenarios [12][13][14] [15]. As a result, neural network–based approaches have become a preferred alternative.

Neural network models—particularly MLP and LSTM—are increasingly recognized as effective tools for forecasting within complex logistics systems. MLP can model nonlinear interactions across diverse variables, while LSTM excels in capturing long-term sequence dependencies. Empirical research indicates that LSTM often achieves lower forecasting error, with [16] reporting an average reduction of

2.3% in logistic demand prediction tasks compared to simpler models. Additional studies on cross-border e-commerce platforms further highlight the ability of neural networks to enhance inventory management and improve demand accuracy [17][18][19][16].

Despite substantial evidence supporting ML/DL-based forecasting methods, comprehensive comparative analyses between MLP and LSTM—under identical experimental conditions and using real-world daily e-commerce demand data—remain limited. The rising volatility of online purchasing behavior, influenced by temporal patterns, promotional cycles, and sudden external disruptions, makes such comparisons increasingly crucial for both operational planning and supply chain decision-making. Based on these observed gaps, this study conducts a detailed comparative evaluation of MLP and LSTM architectures for daily demand forecasting on an e-commerce delivery platform. The objective is to provide empirical insights into each model's strengths, limitations, and suitability for handling highly dynamic, uncertain, and nonlinear demand patterns commonly found in contemporary online logistics.

2. Research Methodology

2.1. Dataset Description

The dataset used in this study consists of historical daily demand records obtained from an e-commerce delivery platform covering the period from January 2020 to December 2024. The dataset contains a total of **1,827 daily observations** and includes multiple operational, temporal, and behavioral attributes. The primary target variable used for forecasting is **num_orders**, representing the total number of orders placed each day.

The dataset is multivariate and includes the following features:

a) **date**: Daily timestamp of recorded transactions
b) **holiday_flag**: Indicator of national or regional holidays
c) **promotion_flag**: Binary variable representing the presence of promotional campaigns
d) **avg_delivery_time_min**: Average delivery time in minutes
e) **num_active_customers**: Number of active customers per day
f) **avg_order_value**: Average order value in currency units
g) **total_revenue**: Total daily revenue generated
h) **num_orders**: Total number of orders (forecasting target)

These variables capture daily operational fluctuations and customer behavior patterns, making the dataset suitable for evaluating the performance of neural network models in demand forecasting.

Table 1: Dataset Summary

Dataset	Number of Rows	Number of Features	Period Covered	Description
E-commerce Daily Demand	1,827	7 + 1 target	2020–2024	Daily operational and behavioral features for demand forecasting

The multi-year time span and rich feature composition provide sufficient variability for comparing the capabilities of Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) models in forecasting daily demand within e-commerce delivery operations.

2.2. Data Preprocessing

Data preprocessing was conducted to ensure that the e-commerce demand dataset was clean, consistent, and suitable for model training using Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM). Several preprocessing steps were applied, covering data cleaning, feature engineering, normalization, and sequence generation for time-series modeling. First, the dataset was examined for missing values, duplicate entries, and inconsistent data formats. No missing values were found, and all date values were standardized into a uniform datetime format to maintain chronological order. Next, time-based features such as day, month, and weekday were extracted from the date attribute to capture seasonal and periodic patterns. The additional operational attributes—holiday_flag, promotion_flag, avg_delivery_time_min, num_active_customers, avg_order_value, and total_revenue—were retained as numerical inputs without requiring categorical encoding. All numerical features were scaled using the MinMaxScaler method to ensure stable neural-network training and prevent larger-valued variables from dominating the learning process. For the LSTM model, the normalized dataset was further reshaped into a 3-dimensional structure using a 30-day sliding window, where each input sequence contains thirty consecutive daily observations and the target output corresponds to the next day's demand (num_orders). This windowing process enables the LSTM architecture to learn temporal dependencies effectively. Finally, the dataset was divided into training, validation, and testing subsets using chronological splitting to avoid information leakage. The training set includes the earliest portion of the data, while the most recent observations were reserved for testing, ensuring that performance evaluation reflects real-world forecasting conditions.

2.3. Model Architecture

This study employs two deep learning architectures—Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM)—to perform daily demand forecasting on an e-commerce delivery platform. Both models were implemented in Python using the TensorFlow/Keras framework and were trained using consistent experimental settings to enable a fair performance comparison.

2.3.1 Multilayer Perceptron (MLP) Architecture

The MLP model is a feed-forward neural network designed to learn nonlinear relationships across the multivariate input features. The architecture consists of multiple dense layers with ReLU activation functions to enhance learning efficiency and mitigate vanishing-gradient issues. The final output layer uses a linear activation to produce continuous numerical predictions of daily demand. To reduce overfitting, dropout regularization was applied to deactivate random neurons during training. Hyperparameters such as the number of hidden units, batch size, learning rate, and training epochs were tuned experimentally based on validation performance to achieve optimal accuracy.

2.3.2 Long Short-Term Memory (LSTM) Architecture

The LSTM model is a specialized recurrent neural network designed to capture long-term temporal dependencies in sequential data. For this study, the input sequences were constructed using a 30-day sliding window, resulting in a 3-dimensional input structure consisting of samples, time steps, and features. Each LSTM cell contains input, forget, and output gates that regulate information flow and allow the network to retain relevant temporal patterns across time steps. The internal tanh activation is used for state updates, while sigmoid functions control gate operations. A final dense layer with linear activation converts the LSTM output into a single-step demand forecast. Dropout layers were included to prevent overfitting and enhance model generalization.

2.3.3 Training Configuration

Both models were trained using the Adam optimizer, selected for its adaptive learning capability and robustness across noisy and non-stationary datasets. The Mean Squared Error (MSE) loss function was used as the training objective due to its suitability for continuous regression tasks. Training was carried out for several epochs until convergence, with a validation split applied to monitor generalization performance. This setup enables a direct comparison of the MLP's ability to model nonlinear input feature interactions with the LSTM's superiority in capturing sequential patterns in multivariate e-commerce demand data.

2.4. Evaluation Metrics

To evaluate the forecasting performance of the Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) models, this study employs two widely used regression metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). These metrics are widely applied in time-series forecasting because they provide a reliable measurement of how closely model predictions follow actual demand values.

RMSE computes the square root of the average squared differences between the predicted and actual values. Because it penalizes larger errors more heavily, RMSE is particularly useful for evaluating datasets with sudden demand spikes, which frequently occur in e-commerce operations during holidays, promotions, or unexpected external events.

MAE calculates the average of the absolute differences between predicted and true values. Unlike RMSE, MAE treats all errors equally, making it easy to interpret as the typical size of the prediction error in daily operational terms.

Using both RMSE and MAE allows for a more comprehensive evaluation of model performance. RMSE highlights sensitivity to large fluctuations, while MAE provides an intuitive interpretation of average forecasting error. Together, these metrics offer a balanced assessment of how effectively the MLP and LSTM models capture the temporal and operational dynamics of e-commerce daily demand.

3. Results and Discussion

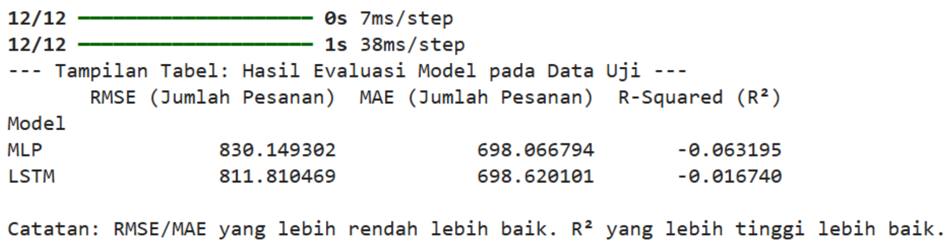
All paragraphs must be justified alignment. With justified alignment, both sides of the paragraph are straight.

3.1. Model Evaluation Results

Both the Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) models were evaluated using the test dataset, with RMSE and MAE serving as the primary accuracy metrics. The results indicate that the LSTM model achieved slightly better predictive performance than the MLP model. Test-set performance:

a) MLP: RMSE = 830.15, MAE = 698.07
b) LSTM: RMSE = 811.81, MAE = 698.62

Although the MAE values of both models are nearly identical, the LSTM achieved a lower RMSE, indicating improved handling of larger errors and more volatile demand fluctuations. Both models produced negative R^2 values, suggesting that the high variability of daily demand—characterized by sudden order spikes, promotions, and holiday effects—poses challenges for simple neural architectures. However, the LSTM's R^2 was closer to zero, demonstrating comparatively better generalization.



```
12/12 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step
12/12 ━━━━━━━━━━━━━━━━━━━━ 1s 38ms/step
--- Tampilan Tabel: Hasil Evaluasi Model pada Data Uji ---
       RMSE (Jumlah Pesanan)  MAE (Jumlah Pesanan)  R-Squared (R²)
Model
MLP               830.149302            698.066794       -0.063195
LSTM              811.810469            698.620101       -0.016740

Catatan: RMSE/MAE yang lebih rendah lebih baik. R² yang lebih tinggi lebih baik.
```

Fig. 1: Model Evaluation Results

3.2. Training and Validation Behavior

Training curves revealed that the MLP model converged more quickly but exhibited stronger fluctuations in validation loss, indicating a higher susceptibility to overfitting. This behavior suggests that the MLP struggles to capture sequential dependencies inherent in daily demand patterns, especially when sudden external disturbances occur. In contrast, the LSTM model demonstrated stable learning dynamics. The training and validation loss curves remained relatively close throughout the epochs, confirming the LSTM's strength in retaining temporal information and managing long-term dependencies in multivariate time series.

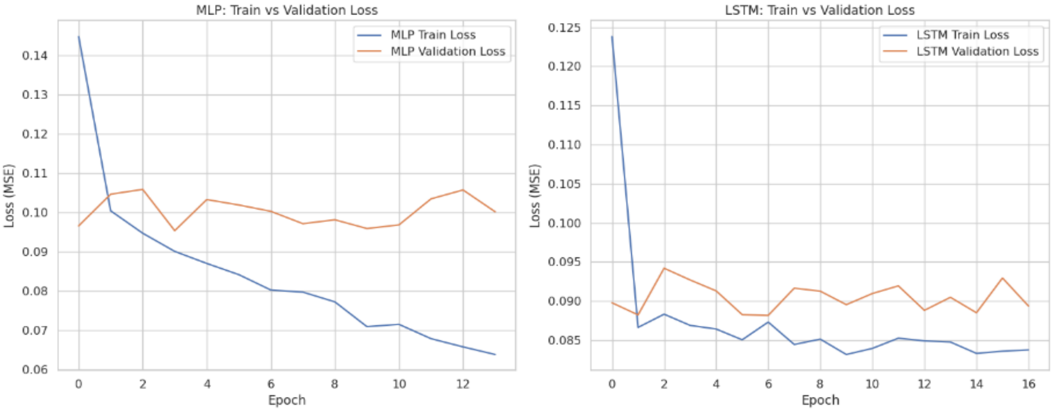


Fig. 2: Training and Validation Loss for MLP and LSTM Models

3.3. Forecast Visualization

A visual comparison between actual demand values and model-generated predictions highlights clear distinctions in forecasting behavior. The MLP tends to produce outputs with higher variance and shows greater deviation during periods of abrupt demand spikes. These deviations indicate that MLP relies more heavily on short-range feature interactions and is less responsive to temporal dynamics. On the other hand, LSTM predictions align more closely with the true demand curve, especially during moderate fluctuations. Although neither model consistently captures extreme spikes, the LSTM provides smoother and more stable predictions, consistent with its ability to learn long-term patterns using gated memory mechanisms.

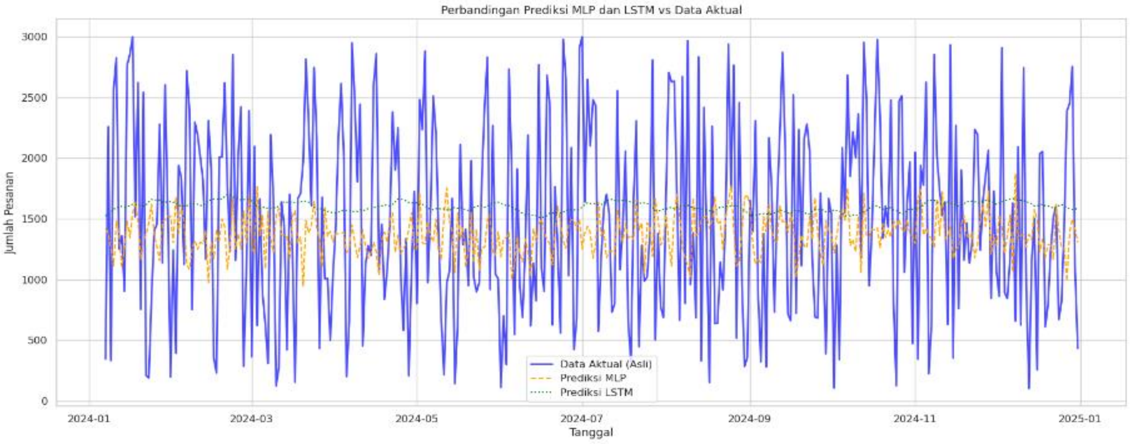


Fig. 3: Actual vs Predicted Values for MLP and LSTM Models

3.4. Interpretation and Comparative Analysis

The results demonstrate several important findings:

1. LSTM outperformed MLP in RMSE and exhibited more stable validation behavior, confirming its superior capability in learning sequential structures in e-commerce daily demand data.
2. Both models struggled with extreme volatility, as shown by negative R^2 scores. This suggests the need for additional external or contextual features, such as promotional intensity, real-time weather conditions, holiday weights, or customer segmentation metrics.
3. Dataset characteristics—such as irregular fluctuations and sudden peaks—indicate that the forecasting task is complex and may require more advanced architectures (e.g., hybrid CNN-LSTM, Transformer-based predictors) to fully capture high-frequency variations.
4. LSTM serves as a stronger baseline model for this type of operational forecasting, while MLP may require significant enhancement or feature engineering to achieve comparable reliability.

Future improvements may include hyperparameter tuning, seasonality decomposition (STL/EEMD), or the integration of real-time behavioral analytics. Such enhancements could substantially improve predictive accuracy and enable more robust deployment in real-world e-commerce logistics systems.

4. Conclusion

This study presents a comparative analysis of Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) models for forecasting daily demand on an e-commerce delivery platform using a multiyear operational dataset. Both models were able to learn general demand patterns; however, the results indicate that the LSTM model provides superior forecasting performance. The LSTM achieved a lower RMSE (811.81) compared to the MLP (830.15), demonstrating better capability in handling nonlinear temporal dependencies and mitigating large prediction errors. The training and validation curves further confirmed that LSTM exhibited more stable learning behavior, while the MLP showed fluctuating validation loss and a higher tendency toward overfitting. Visualization of predicted versus actual demand also revealed that LSTM follows the overall demand trends more closely, particularly during moderate fluctuations. Despite these advantages, both models struggled to accurately predict extreme demand spikes, which is reflected in their negative R^2 scores. This limitation suggests that daily demand forecasting in e-commerce environments is strongly influenced by complex external factors that are not fully represented in the current dataset. Overall, the findings indicate that LSTM is a more reliable baseline model for forecasting daily e-commerce demand. Future research may improve predictive performance by incorporating additional features such as real-time promotions, weather conditions, customer segmentation, and seasonality decomposition, or by adopting more advanced deep learning architectures such as hybrid CNN-LSTM or Transformer-based models.

References

[1] J. A. Cano, A. Londoño-Pineda, and C. Rodas, "Sustainable logistics for e-commerce: A literature review and bibliometric analysis," *Sustainability*, vol. 14, no. 19, p. 12247, 2022, doi: 10.3390/su141912247.

[2] D. Karl, "Forecasting e-commerce consumer returns: A systematic literature review," *Manag. Rev. Q.*, 2024, doi: 10.1007/s11301-024-00436-x.

[3] M. Markowska, J. Marcinkowski, M. Kiba-Janiak, and D. Strahl, "Rural e-customers' preferences for last mile delivery and products purchased via the Internet before and after the COVID-19 pandemic," *J. Theor. Appl. Electron. Commer. Res.*, vol. 18, no. 1, pp. 597–614, 2023, doi: 10.3390/jtaer18010030.

[4] M. Suguna, B. Shah, S. K. Raj, and M. Suresh, "A study on the influential factors of the last-mile delivery projects during COVID-19 era," *Oper. Manag. Res.*, vol. 15, pp. 399–412, 2022, doi: 10.1007/s12063-021-00214-y.

[5] S. F. Ni, Y. Peng, and Z. J. Liu, "Logistics demand forecast of fresh food e-commerce based on Bi-LSTM model," *J. Comput. Commun.*, vol. 10, pp. 51–65, 2022, doi: 10.4236/jcc.2022.109004.

[6] M. Nasseri, T. Falatouri, P. Brandtner, and F. Darbanian, "Applying Machine Learning in Retail Demand Prediction—A Comparison of Tree-Based Ensembles and Long Short-Term Memory-Based Deep Learning," *Appl. Sci.*, vol. 13, no. 19, p. 11112, 2023, doi: 10.3390/app131911112.

[7] K. Douaioui, R. Oucheikh, O. Benmoussa, and C. Mabrouki, "Machine Learning and Deep Learning Models for Demand Forecasting in Supply Chain Management: A Critical Review," *Appl. Syst. Innov.*, vol. 7, no. 5, p. 93, 2024, doi: 10.3390/asi7050093.

[8] C. Kaoutar Douaioui, Rachid Oucheikh, Othmane Benmoussa, "Model pembelajaran mesin dan pembelajaran mendalam untuk peramalan permintaan dalam manajemen rantai pasok: Tinjauan kritis," *Appl. Syst. Innov.*, vol. 7, no. 5, p. 93, 2024, doi: 10.3390/asi7050093.

[9] Nasseri, "Nasseri, M., Falatouri, T., Brandtner, P., & Darbanian, F. (2023). Penerapan Pembelajaran Mesin dalam Prediksi Permintaan Ritel—Perbandingan antara Ensembel Berbasis Pohon dan Pembelajaran Mendalam Berbasis LSTM. Applied Sciences, 13(19), 11112. https://d".

[10] E. Aldahmani, A. Alzubi, and K. Iyiola, "Demand Forecasting in Supply Chain Using Uni-Regression Deep Approximate Forecasting Model," *Appl. Sci.*, vol. 14, no. 18, p. 8110, 2024, doi: 10.3390/app14188110.

[11] J.-J. Zhou, "An improved machine learning model Shapley value-based to forecast demand for aquatic product supply chain," *Front. Ecol. Evol.*, 2023, doi: 10.3389/fevo.2023.1160684.

[12] R. Baidya and S.-W. Lee, "Addressing the non-stationarity and complexity of time series data for long-term forecasts," *Appl. Sci.*, vol. 14, no. 11, p. 4436, 2024, doi: 10.3390/app14114436.

[13] X. Kong, Z. Chen, W. Liu, K. Ning, and L. Zhang, "Deep learning for time series forecasting: A survey," *Int. J. Mach. Learn. Cybern.*, vol. 16, pp. 5079–5112, 2025, doi: 10.1007/s13042-025-02560-w.

[14] P. Lara-Benítez, M. Carranza-García, and J. C. Riquelme, "An experimental review on deep learning architectures for time series forecasting," 2021. [Online]. Available: https://arxiv.org/abs/2103.12057

[15] W. Li and K. E. Law, "Model pembelajaran mendalam untuk peramalan deret waktu: Sebuah tinjauan," *IEEE Access*, vol. 11, pp. 71234–71244, 2023, doi: 10.1109/ACCESS.2023.3167042.

[16] V. Kuppulakshmi, C. Sugapriya, D. Nagarajan, and A. Kanchana, "Modelling with neural networks and time-series forecasting: inventory control and cost reduction in supply chain process," *SN Oper. Res.*, vol. 6, no. 2, 2025, doi: 10.1007/s43069-025-00466-5.

[17] W. Yang, "A neural network-based model for cross-border e-commerce supply chain demand forecasting and inventory optimization," *Appl. Math. Nonlinear Sci.*, vol. 9, no. 1, 2024, doi: 10.2478/amns-2024-2915.

[18] X. Liu and W. Wang, "Model pembelajaran mendalam untuk peramalan deret waktu: Sebuah survei komprehensif," *Mathematics*, vol. 12, no. 10, p. 1504, 2024, doi: 10.3390/math12101504.

[19] A. Hussain and K. Ahsan, "Hybrid EEMD-LSTM model for accurate e-commerce demand forecasting," *J. Data Digit. Innov.*, vol. 2, no. 2, 2025.