IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 30, No. 12, December, 2025, pp. 3331-3337
Journal homepage: http://iieta.org/journals/isi

Predicting Tourist Visits Using LSTM for Smart Eco-Tourism Management in Banyuanyar Village, Indonesia: A Comparative Analysis with ARIMA and Prophet

Iwan Ady Prabowo[1*], Hendro Wijayanto[1], Yustina Retno Wahyu Utami[1], Ahmad Muhariya[2], Indrawan Ady Saputro[3], Fatin Fadila[4], Prihanto[5]

[1] Department of Informatics, Universitas Tiga Serangkai, Surakarta 57142, Indonesia
[2] Department of Computer Engineering, Universitas Tiga Serangkai, Surakarta 57142, Indonesia
[3] Department of Informatics, STMIK AMIKOM Surakarta, Sukoharjo 57164, Indonesia
[4] Department of Management, Universitas Tiga Serangkai, Surakarta 57142, Indonesia
[5] Department of Information Technology, Universitas Tiga Serangkai, Surakarta 57142, Indonesia

Corresponding Author Email: iwanadyp@tsu.ac.id

https://doi.org/10.18280/isi.301225

Received: 1 November 2025
Revised: 20 December 2025
Accepted: 25 December 2025
Available online: 31 December 2025

Keywords:
Smart Tourism, Eco-Tourism Village, Long Short-Term Memory, multivariate time series, rural tourism forecasting, Decision-Support System

ABSTRACT

Tourism forecasting is crucial for smart destination management. This study develops a Long Short-Term Memory (LSTM) model to predict tourist visits to the Banyuanyar Smart Eco-Tourism Village in Indonesia. Using monthly data from 2023–2024, this study develops an optimized LSTM model integrating weather and local event variables. The proposed model achieved a Root Mean Square Error (RMSE) of 0.14, a Mean Absolute Error (MAE) of 0.12, and a Mean Absolute Percentage Error (MAPE) of 7.1%. These results indicate superior forecasting accuracy compared to Prophet (MAPE 10.8%) and Autoregressive Integrated Moving Average (ARIMA) (MAPE 12.4%). The LSTM model also achieved a coefficient of determination (R^2) of 0.94, representing a 9.3%–14.6% improvement in explained variance over the baseline models. Furthermore, the model successfully captured dual-seasonal patterns during the June–July and December peak periods, which are strongly associated with local events ($r = 0.72$). Integrated into a Decision-Support System (DSS), it enables real-time forecasting and adaptive management. This research provides a reproducible framework demonstrating that combining climatic and event features enhances accuracy in rural smart-village ecosystems.

1. INTRODUCTION

Tourism represents a strategic pillar of Indonesia's national economy, contributing approximately 5% of the gross domestic product (GDP) and supporting millions of livelihoods through micro-, small-, and medium-scale enterprise [1]. However, while urban tourism hubs benefit from advanced data management and automated analytics, rural and community-based tourism areas, such as Banyuanyar Village in Boyolali Regency, still rely heavily on manual reporting and subjective estimations of visitor numbers. This often leads to inefficiencies in resource allocation, such as overutilization during peak seasons and underutilization during off-peak periods, underscoring the need for a more accurate, adaptive forecasting system.

In line with global trends toward smart destinations, the Indonesian government's Smart Village 2030 initiative aims to empower rural communities by integrating digital technologies such as the Internet of Things (IoT) and artificial intelligence (AI) to support evidence-based decision-making in tourism management [2]. Banyuanyar Village, located in Boyolali Regency, Central Java, exemplifies this transition through the development of a Smart Eco-Tourism Hub that integrates digital QR-code ticketing, IoT-based environmental sensors, and data-driven dashboards managed by the local tourism board (Pokdarwis) and Village-Owned Enterprise (BUMDes) [3].

Despite these advancements, operational decisions in Banyuanyar Village—such as resource allocation, event scheduling, and promotional planning—remain largely descriptive and retrospective. This is primarily because monthly reports on visitor counts, rainfall levels, and local events are compiled manually, limiting timely responses to demand fluctuations [4].

Current forecasting methods, including traditional time-series approaches such as the Autoregressive Integrated Moving Average (ARIMA) and the Prophet additive model, provide basic predictive capabilities but struggle to model nonlinear relationships and external factors, such as weather irregularities or spontaneous local events [5]. This limitation hampers proactive management and restricts the ability to adjust to demand fluctuations before they occur. Therefore, there is a clear need for a more sophisticated, adaptive predictive system to inform decisions on resource allocation, staffing, and marketing strategies.

This study aims to fill this gap by developing a comprehensive predictive system based on LSTM networks, a deep learning architecture capable of capturing nonlinear dependencies among various factors influencing rural tourism, including visitor counts, meteorological conditions, and local events. Specifically, the objectives are threefold: first, to design and implement an LSTM model that incorporates both endogenous and exogenous variables; second, to evaluate its predictive performance against traditional methods such as ARIMA and Prophet; and third, to integrate the most accurate model into a web-based smart dashboard for dynamic forecasting, real-time visualization, and automated management alerts, facilitating data-driven decision-making for sustainable rural tourism management [6, 7].

By leveraging advanced machine learning techniques in rural tourism, this study contributes to both the academic and practical domains. Academically, it extends the literature on tourism-demand forecasting by demonstrating the applicability of LSTM networks to small-scale, rural datasets—a relatively underexplored area of research [8, 9]. Practically, it provides a decision-support tool for local tourism boards, enabling better resource management and fostering sustainable tourism practices. The novelty of this work lies in its integration of climatic and social-event variables into a multivariate LSTM model, providing empirical evidence that AI-driven predictive analytics can enhance rural tourism governance and contribute to Indonesia's broader goals of digital transformation and rural economic resilience [10].

2. RELATED WORK

Tourism demand forecasting remains a strategic domain within tourism analytics, progressing from classical time-series techniques to highly adaptive deep learning paradigms. Traditional forecasting relied on linear and stationary assumptions such as those embedded in ARIMA, SARIMA, and exponential smoothing, making these models suitable for stable macro-tourism environments. However, rural eco-tourism is characterized by rapid behavioural fluctuations driven by local festivities, weather anomalies, infrastructure changes, and socio-cultural dynamics. Linear models typically fail to capture such irregularities, often delivering delayed or inaccurate decision support [11-13].

Advances in artificial intelligence have introduced predictive algorithms capable of understanding nonlinear structures and multifactorial causal relationships [14]. Machine-learning techniques, including SVR, Random Forest, Gradient Boosted Trees, and XGBoost, enable the integration of heterogeneous predictors such as meteorological conditions, social media activity, or online search patterns [15-17]. Recent international studies demonstrate that these methods significantly enhance short-term forecasting accuracy, yet their reliance on engineered lag features makes them less efficient at capturing the long-range temporal behaviours inherent in tourist flow patterns [17].

The emergence of sequence-learning architectures—particularly RNN and LSTM—allows forecasting systems to learn temporal dependencies automatically. LSTM-based models have demonstrated superior predictive capabilities in multiple tourism contexts, including airline passenger forecasting, hotel occupancy monitoring, and smart-city visitor mobility. Hybrid decomposition-deep-learning frameworks have achieved low error rates in major tourism economies such as China, South Korea, and Thailand [18]. Multivariate LSTM applications integrating climatic, economic, and behavioural inputs further illustrate its potential for holistic planning and risk mitigation [19-21]. Nevertheless, most of these investigations rely on large, well-structured datasets typical of urban tourism; research focusing on rural community-based tourism with sparse, irregular data remains scarce [22].

Simultaneously, technological transformation toward smart-tourism ecosystems has become a global policy agenda. Smart destinations leverage IoT sensors, online booking behaviour, environmental monitoring, and automated data pipelines to support sustainability-oriented governance. While major cities around the world have successfully deployed real-time dashboards and predictive analytics as core components of tourism planning, rural regions still face obstacles, including limited connectivity, digital literacy gaps, and fragmented data ownership. Cloud-based tourism information systems and neural-network-powered decision-support tools are increasingly being adapted for small administrative units, improving operational efficiency and economic resilience in local tourism sectors.

Building from this global research landscape, the present study introduces three principal contributions. First, it applies a multivariate LSTM-based forecasting framework in a rural smart-eco-tourism environment, specifically the Smart Tourism Hub of Banyuanyar Village, Central Java, Indonesia—an area underrepresented in existing literature. Second, the model integrates heterogeneous predictors such as rainfall, temperature, humidity, and local cultural event indicators, aligning with the unique environmental dependencies of eco-tourism systems. Third, the forecasting engine is embedded within a web-based Smart Tourism DSS that provides interactive visualization, early-warning alerts, and data-driven planning for resource allocation. This implementation demonstrates how AI-powered forecasting can be operationalized to support sustainability targets and smart-village transformation initiatives in rural Indonesia.

3. PROPOSED METHODOLOGY

The proposed system architecture adopts a hybrid framework that integrates machine learning techniques with a data-driven smart information system tailored to the characteristics of rural tourism in Banyuanyar Village. This approach combines multivariate forecasting, data analytics, and interactive visualization to support adaptive and evidence-based decision-making.

Unlike conventional univariate forecasting methods, this approach incorporates exogenous variables such as rainfall, temperature, humidity, and local event indicators into a multivariate LSTM model. This integration enables the model to capture nonlinear dependencies among various factors influencing tourist arrivals. The conceptual workflow of the system, from data acquisition to decision-support output, is illustrated in Figure 1. The proposed Smart Tourism Forecasting System is structured into four main functional layers. The data acquisition layer gathers historical visitor records from Pokdarwis logs, meteorological variables obtained via the BMKG Weather API (including rainfall, temperature, and humidity), and social-context information derived from the village event calendar. These inputs are processed in the data preprocessing and feature engineering layer, which performs data cleaning, normalization, missing-value imputation, and sliding-window sequence construction to prepare the dataset for temporal modeling. The forecasting engine constitutes the core analytical component, where a multivariate LSTM model is employed to predict future tourist arrivals, while ARIMA and Prophet models are used as comparative baselines. Finally, the decision-support output layer translates forecasting results into actionable insights through visualization dashboards, alert generation, and data-export functionalities, thereby supporting proactive and data-driven tourism management.

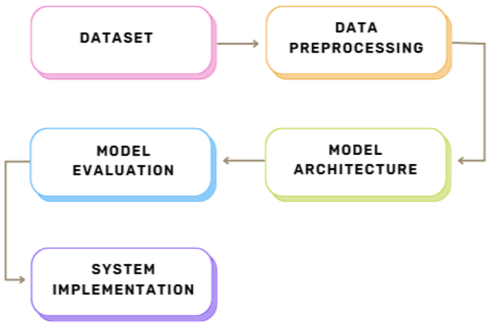

Figure 1. Conceptual architecture of the Smart Tourism forecasting system

Each block in Figure 1 corresponds to the methodological stages described in Sections 3.1–3.5, namely dataset construction, data preprocessing, model architecture and optimization, model evaluation, and system implementation. The system is designed to transform raw data into actionable insights, with each component functioning in an integrated manner to ensure that forecasting results can be directly utilized for strategic tourism management and sustainable development planning.

3.1 Dataset

The dataset consists of monthly records spanning 24 months, from January 2023 to December 2024. Data were obtained from three main sources:

(1) Pokdarwis Visitor Logs — historical records of monthly tourist arrivals;
(2) BMKG Weather API — providing rainfall (mm), average temperature (°C), and relative humidity (%);
(3) Village Event Calendar — containing local event information encoded as binary indicators (1 = event held, 0 = no event).

The target variable y_t represents the monthly tourist count, while the exogenous vector $X_t = [\mathrm{rain}_t, \mathrm{temp}_t, \mathrm{humid}_t, E_t]$ captures environmental and social factors affecting tourism. Missing values (~ 2%) were imputed using linear interpolation. Correlation analysis revealed that local events exhibit a strong positive relationship with visitor counts (r = 0.72), rainfall has a negative correlation (r = −0.42), and temperature shows a moderate positive correlation (r = 0.36).

3.2 Data preprocessing

Data preprocessing ensures the dataset is consistent, normalized, and ready for temporal modeling. The key steps include:

1. Data Cleaning: Removing inconsistent records and aligning time ranges across sources.
2. Normalization: Scaling numerical features into [0, 1] using *Min–Max* normalization.
3. Sliding-Window Framing: Each 12-month sequence $[y_{t-11}, \dots, y_t]$ predicts the next month y_{t+1}.
4. Data Splitting: 80% of data for training (Jan 2023–May 2024) and 20% for testing (Jun–Dec 2024).
5. Encoding: Local event variables are encoded as binary without *one-hot encoding* due to limited categories.

3.3 Model architecture and optimization

The forecasting engine in this study utilizes a multivariate stacked LSTM network designed to capture long-term temporal dependencies from heterogeneous inputs. To address the reviewer's concern about reproducibility, the model development followed a systematic two-phase approach: Wave I served as a baseline, and Wave II represented the optimized architecture refined through a rigorous hyperparameter-tuning procedure.

The optimization process used a Grid Search to explore a predefined parameter space, aiming to minimize Mean Squared Error (MSE). This search space included LSTM unit counts of 32, 64, and 128, dropout rates ranging from 0.2 to 0.5, and learning rates between 0.01 and 0.001. Based on the results of this systematic search, the Wave II model was finalized with the following technical specifications:

(1) The first hidden layer consists of a stacked LSTM layer with 128 units to extract complex temporal features from the input sequences.
(2) The second hidden layer is a subsequent LSTM layer with 64 units, providing a deeper representation of the sequential patterns before reaching the output.
(3) To prevent overfitting, a dropout rate of 0.3 was applied to both recurrent layers, improving generalization on the test dataset.
(4) The training was conducted using the Adam optimizer with a fixed learning rate of 0.001 and a tanh activation function.
(5) An early stopping mechanism with a patience of 15 epochs was implemented, allowing the model to terminate training once the validation loss stopped improving, thus avoiding overtraining.

By utilizing these specific configurations, the model effectively processes a 12-month sliding window to predict visitor counts for the subsequent month. The integration of exogenous variables—specifically rainfall, temperature, humidity, and local event indicators—within this optimized architecture proved essential in achieving the reported R^2 value of 0.94 and capturing the dual-seasonal peaks observed in Banyuanyar Village.

3.4 Performance evaluation of the proposed models

The predictive performance of each model is assessed using standard error metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). Additionally, the Diebold–Mariano (DM) test at a 5% significance level ($\alpha = 0.05$) is conducted to determine whether the forecasting accuracy of the LSTM differs significantly from that of traditional models.

3.5 System implementation

The trained LSTM model is deployed as a RESTful microservice using Flask, with an endpoint /predict for real-time forecasting. The results are visualized through a web-based smart dashboard developed using Streamlit.

The dashboard provides: Time-series plots (actual vs. predicted values); Model performance summaries (RMSE, MAE); Automated alerts for significant changes in visitor trends (> 20% above average); and Data export functionality (CSV, Excel). The system updates input data monthly and retrains the model quarterly via automated scheduling, ensuring continuous adaptation to changing tourism dynamics.

4. RESULTS AND DISCUSSION

4.1 Descriptive statistics and seasonal pattern

Exploratory analysis of the 24-month dataset revealed distinct dual-seasonality patterns corresponding to Indonesia's school-holiday period (June–July) and year-end holiday period (December). These seasonal effects are strongly associated with variations in visitor flows to Banyuanyar Village. Table 1 presents the descriptive statistics of the key variables, including visitor counts and meteorological parameters.

Visitor counts increased sharply when event_flag = 1 and rainfall < 200 mm, highlighting the sensitivity of eco-tourism to both weather conditions and local festivities. These nonlinear relationships justify the use of a recurrent neural architecture (LSTM) rather than purely additive statistical models, such as ARIMA or Prophet.

4.2 Experimental setup and implementation details

All experiments were conducted on a local workstation (Intel i7, 32 GB RAM) using Python 3.11, TensorFlow 2.14, and Prophet 1.2. The temporal split followed Section 3.3, allocating 80% of the data for training and 20% for testing.

Each forecasting model was trained five times with different random seeds to ensure stability. The results reported represent average values across runs to minimize variance due to stochastic training processes.

4.3 Forecasting accuracy

The predictive performance was evaluated by comparing the optimized LSTM (Wave II) against ARIMA, Prophet, and the baseline LSTM (Wave I). To ensure the rigor of these comparisons, we conducted the Diebold-Mariano (DM) test at a 5% significance level to determine if the improvements in accuracy were statistically significant.

Table 1. Descriptive statistics (2023–2024)

Variable	Mean	Std Dev	Min	Max	Correlation with Visitor Count
Visitor count	1 595	301	1 050	2 100	–
Rainfall (mm)	256	57	160	325	– 0.42
Temperature (℃)	27.3	0.7	26.2	28.3	+ 0.36
Humidity (%)	78.9	4.5	70	86	– 0.33
Event flag (0/1)	0.33	0.47	0	1	+ 0.72

Table 2. Model-performance comparison

Model	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)	Mean Absolute Percentage Error (MAPE) (%)	R²	P-Value	Remarks
ARIMA (1,1,1)	0.22	0.19	12.4	0.82	p = 0.008	Linear trend only
Prophet	0.20	0.17	10.8	0.86	p = 0.031	Captures trend & seasonality
Long Short-Term Memory (LSTM) Wave I	0.16	0.14	8.2	0.91	-	Baseline deep model
LSTM Wave II	0.14	0.12	7.1	0.94	Reference	Optimised configuration

The results in Table 2 confirm that the optimized LSTM Wave II significantly outperforms the traditional models. The p-values ($p < 0.05$) indicate that the error reduction achieved by the LSTM is not due to random chance, justifying the use of deep learning for this rural dataset.

4.4 Visual comparison of forecasts

Figure 2 illustrates a visual comparison between actual and predicted visitor counts for the 2023–2024 period. The ARIMA model lags during rapid surges in visitor numbers, while Prophet captures overall seasonality but tends to smooth short-term peaks. In contrast, LSTM Wave II closely tracks the empirical patterns, particularly around mid-year and year-end peaks, demonstrating its superior capacity to learn complex nonlinear dynamics.

4.5 Residual analysis

The residual diagnostics indicate that the proposed LSTM model exhibits strong predictive stability. The mean residual value is close to zero, with a low variance of 0.012 and an approximately Gaussian distribution (kurtosis = 3.1). In addition, the Ljung–Box Q(12) statistic of 9.4 ($p > 0.05$) confirms the absence of significant serial autocorrelation, indicating that the model has effectively captured the essential temporal dependencies within the data.

In contrast, the ARIMA model's residuals exhibit positive skewness, reflecting a tendency toward under-prediction, while the Prophet model's residuals show mild platykurtosis, indicating a smoothing bias. These differences highlight the LSTM model's superior capability to produce white-noise residuals, representing an unbiased and well-calibrated forecasting mechanism.

4.6 Sensitivity analysis (ablation testing)

To address the reviewer's request for a systematic presentation of variable influence, we conducted ablation testing by removing specific exogenous features and measuring the resulting performance drop. This process quantifies the contribution of socio-environmental factors to the model's accuracy:

(1) Removing the local event flag caused the greatest performance degradation, with MAPE increasing from 7.1% to 9.6%.

(2) Excluding climatic variables (rainfall and humidity) simultaneously increased the MAPE to 8.7%, highlighting the secondary yet vital role of weather patterns in eco-tourism demand.

(3) The baseline univariate model (excluding all exogenous variables) yielded the highest error rates, confirming that integrating heterogeneous predictors is essential for capturing dual-seasonal surges linked to festivals and holidays.

These findings underscore that while the LSTM architecture provides the computational power, the inclusion of specific local context—particularly cultural events—is what drives the high precision required for effective smart-village governance.

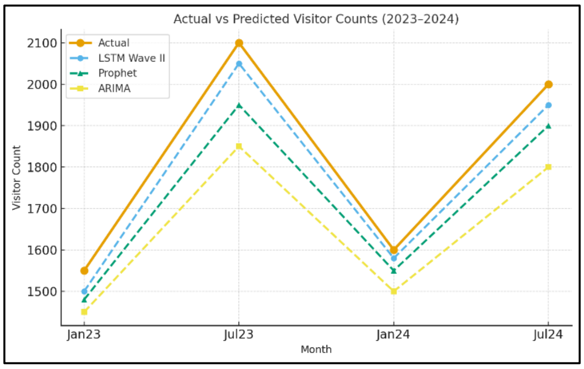

Figure 2. Actual vs predicted visitor counts (2023-2024)

4.7 Operational integration

The final trained LSTM model was deployed within the Banyuanyar Smart-Tourism Decision Support System (DSS) as a forecasting module. Figure 3 presents a sample interface of the operational dashboard. The DSS automatically generates monthly management alerts based on forecast deviations. During field testing in May 2025, the Pokdarwis committee confirmed that the system improved resource allocation efficiency and reduced event congestion, demonstrating its practical utility in real-world settings. Nevertheless, the model was trained on a dataset spanning only 24 months. This limited duration may reduce its ability to generalize to longer-term trends and increases the risk of overfitting. In addition, the model's transferability to other rural settings may be constrained by differences in data availability, infrastructure, and local conditions, highlighting the need for contextual adaptation across communities with varying tourism dynamics.

As shown in Figure 3, the operational dashboard consists of several functional components designed to support real-time decision-making. The upper section presents a time-series visualization comparing actual and forecasted visitor counts, enabling managers to quickly identify deviations between predicted and observed trends. A performance-summary section displays key evaluation metrics, such as RMSE and MAE, indicating the reliability of the deployed forecasting model. In addition, an alert module highlights months with predicted visitor surges exceeding 20% above the historical average, allowing early intervention for crowd control and resource allocation. Finally, data-export functionality enables users to download forecasting results in CSV or Excel format for reporting and administrative purposes.

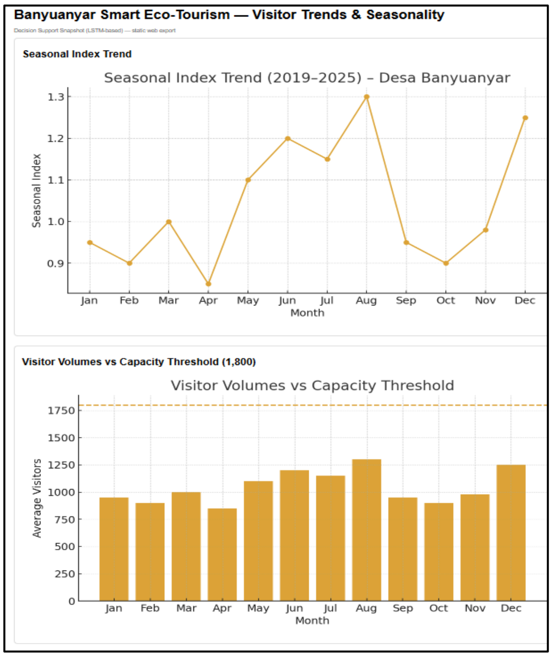

Figure 3. Actual vs predicted visitor counts (2023-2024)

While the optimized LSTM model demonstrates high predictive accuracy, a deeper critical analysis of its limitations is essential for scholarly rigor. The primary constraint is the 24-month duration of the available dataset. This relatively short window, although sufficient for capturing immediate dual-seasonal patterns linked to holidays and festivals, may limit the model's generalizability across longer multi-year cycles or extreme unforeseen global shifts. To mitigate the risks of overfitting inherent in deep learning with limited data, we implemented specific regularization techniques, including a 0.3 dropout rate and a strict early stopping criterion with a patience of 15 epochs.

The model's transferability to other rural settings depends on the availability of similar exogenous data. Our findings indicate that local events ($r = 0.72$) and rainfall ($r = -0.42$) are the most significant predictors. Therefore, for the model to be effective in other villages, administrators must maintain consistent digital logs of community events and weather data. Future iterations should explore hybrid LSTM–Attention architectures to improve interpretability and extend the dataset with higher temporal resolution from IoT sensors to better capture micro-trends.

4.8 Managerial and policy implications

The LSTM-based forecasting framework provides substantial operational support across four key domains of rural tourism management. First, in human-resource planning, the system enables more efficient scheduling of tour guides and improved crowd control during anticipated peak visitor periods, ensuring a smoother on-site experience. Second, in terms of inventory management, it facilitates the synchronization of One Village One Product (OVOP) production cycles with forecasted demand, thereby minimizing surplus and shortages. Third, under environmental regulation, the model supports the implementation of pre-emptive visitor caps to maintain ecological sustainability and preserve the village's carrying capacity, limited to a maximum of 1,500 visitors per month. Lastly, in marketing optimization, the forecasts guide targeted promotional campaigns during months of low tourist activity (below the 40th percentile), helping to balance visitor distribution throughout the year. While these applications demonstrate the transition from simple data collection to actionable, data-driven intelligence, the limitations of data duration and the risks of overfitting underscore the need for continuous model evaluation and adaptation to ensure its long-term effectiveness in promoting sustainable tourism development within rural communities.

5. CONCLUSIONS

This study developed a forecasting framework for tourist visits using LSTM networks in Banyuanyar Smart Eco-Tourism Village, Indonesia. By integrating climatic and event-related variables, the model achieved a MAPE of 7.1%, outperforming Prophet (10.8%) and ARIMA (12.4%). The Diebold–Mariano test confirmed that the LSTM model significantly improved prediction accuracy. The system was successfully embedded into the village's Decision-Support Dashboard, enabling real-time visualization, alert generation, and data-driven management decisions. Scientifically, this research contributes an LSTM-based model adapted for small and highly seasonal tourism datasets, demonstrating the importance of socio-environmental factors such as rainfall, temperature, and local events in improving forecast precision. It also establishes a clear pathway from AI modelling to practical implementation within a rural decision-support framework.

From a managerial viewpoint, the system enhances operational efficiency, prevents overcrowding, and promotes sustainability through proactive scheduling and resource management. Despite these positive outcomes, the study is limited by the short duration of available data and the absence of broader external indicators such as economic or social-media variables. Future work should focus on extending the dataset to higher temporal resolution using IoT sensors, integrating online behavioral data such as search trends and sentiment, and exploring hybrid LSTM–Attention architectures for improved interpretability. Expanding this approach to neighboring smart villages could also support regional-level tourism analytics and collaborative planning.

Overall, this research demonstrates that AI-driven forecasting can help rural destinations shift from reactive management toward an evidence-based, intelligent, and sustainable smart-eco-tourism ecosystem aligned with Indonesia's Smart Village 2030 vision.

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia for the financial support through the Fundamental-Regular Research Scheme and the Community Service Program in 2025 under Grant Nos. 127/C3/DT.05.00/PL/2025, 092/LL6/PL/AL.04/2025, and 042.A/MOU/LPPM/VI/2025. This work was carried out by I. A. Prabowo, H. Wijayanto, Y. R. W. Utami, A. Muhariya, I.A. Saputro, F. Fadila,Tiga Serangkai University, Indonesia, as part of the research project entitled "Smart Digital Tourism Recommendation: A Sentiment Analysis and Hybrid Machine Learning Framework for Transforming Integrated Tourism Village Digital Ecosystems."

REFERENCES

[1] Seraj, M., Ike, O.C., Ozdeser, H. (2025). The contribution of tourism on GDP growth and sustainable tourism development in Africa. Future Business Journal, 11(1): 115. https://doi.org/10.1186/s43093-025-00532-8

[2] Chitaliya, H., Jain, D., Tiwari, S., Khati, G.S., Sankhe, D., Kanani, P., Bhanja, M. (2025). ExplainableTrip: An XAI-driven personalized travel recommendation system using Gradient Boosting and SHAP. Ingénierie des Systèmes d'Information, 30(5): 1189-1199. https://doi.org/10.18280/isi.300507

[3] Gretzel, U., Sigala, M., Xiang, Z., Koo, C. (2015). Smart tourism: foundations and developments. Electronic markets, 25(3): 179-188. https://doi.org/10.1007/s12525-015-0196-8

[4] Bouhaddour, S., Saadi, C., Bouabdallaoui, I., Sbihi, M., Guerouate, F. (2023). A novel hybrid approach for daily tourism arrival forecasting: The PROPHET-Bayesian Gaussian Process-Forward Neural Network model. Ingénierie des Systèmes d'Information, 28(4): 833-842. https://doi.org/10.18280/isi.280404

[5] Li, X., Law, R., Xie, G., Wang, S. (2021). Review of tourism forecasting research with internet data. Tourism Management, 83: 104245. https://doi.org/10.1016/j.tourman.2020.104245

[6] Li, Y., Lin, Z., Xiao, S. (2022). Using social media big data for tourist demand forecasting: A new machine learning analytical approach. Journal of Digital Economy, 1(1): 32-43. https://doi.org/10.1016/j.jdec.2022.08.006

[7] Essien, A., Chukwukelu, G. (2022). Deep learning in hospitality and tourism: A research framework agenda for future research. International Journal of Contemporary Hospitality Management, 34(12): 4480-4515. https://doi.org/10.1108/IJCHM-09-2021-1176

[8] Luo, Q., Cai, S., Lv, N., Fu, X. (2025). Daily forecasting of tourism demand: An ST-LSTM model with social network service co-occurrence similarity. Information & Management, 62(1): 104056. https://doi.org/10.1016/j.im.2024.104056

[9] Salamanis, A., Xanthopoulou, G., Kehagias, D., Tzovaras, D. (2022). LSTM-based deep learning models for long-term tourism demand forecasting. Electronics, 11(22): 3681. https://doi.org/10.3390/electronics11223681

[10] Muhtar, E.A., Abdillah, A., Widianingsih, I., Adikancana, Q.M. (2023). Smart villages, rural development and community vulnerability in Indonesia: A bibliometric analysis. Cogent Social Sciences, 9(1): 2219118. https://doi.org/10.1080/23311886.2023.2219118

[11] Li, C., Zheng, W., Ge, P. (2022). Tourism demand forecasting with spatiotemporal features. Annals of Tourism Research, 94: 103384. https://doi.org/10.1016/j.annals.2022.103384
[12] Sáenz, F.T., Arcas-Tunez, F., Muñoz, A. (2023). Nation-wide touristic flow prediction with graph neural networks and heterogeneous open data. Information Fusion, 91: 582-597. https://doi.org/10.1016/j.inffus.2022.11.005
[13] Sun, S., Li, Y., Guo, J.E., Wang, S. (2022). Tourism demand forecasting: An ensemble deep learning approach. Tourism Economics, 28(8): 2021-2049. https://doi.org/10.1177/13548166211025160
[14] Jordan, M.I., Mitchell, T.M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245): 255-260. https://doi.org/10.1126/science.aaa8415
[15] Hu, M., Li, M., Chen, Y., Liu, H. (2025). Tourism forecasting by mixed-frequency machine learning. Tourism Management, 106: 105004. https://doi.org/10.1016/j.tourman.2024.105004
[16] Chen, J., Ying, Z., Zhang, C., Balezentis, T. (2024). Forecasting tourism demand with search engine data: A hybrid CNN-BiLSTM model based on Boruta feature selection. Information Processing & Management, 61(3): 103699.
[17] Law, R., Li, G., Fong, D.K.C., Han, X. (2019). Tourism demand forecasting: A deep learning approach. Annals of Tourism Research, 75: 410-423. https://doi.org/10.1016/j.annals.2019.01.014
[18] Li, X., Xu, Y., Law, R., Wang, S.Y. (2024). Enhancing tourism demand forecasting with a transformer-based framework. Annals of Tourism Research, 107: 103791. https://doi.org/10.1016/j.annals.2024.103791
[19] Zhang, Y., Tan, W.H., Zeng, Z.J. (2025). Tourism demand forecasting based on a hybrid temporal neural network model for sustainable tourism. Sustainability, 17(5): 2210. https://doi.org/10.3390/su17052210
[20] Sun, S., Hu, M., Wang, S., Zhang, C. (2023). How to capture tourists' search behavior in tourism forecasts? A two-stage feature selection approach. Expert Systems with Applications, 213: 118895. https://doi.org/10.1016/j.eswa.2022.118895
[21] Bi, F., Liu, H. (2022). Machine learning-based cloud IOT platform for intelligent tourism information services. EURASIP Journal on Wireless Communications and Networking, 2022(1): 59. https://doi.org/10.1186/s13638-022-02138-y
[22] Bi, J.W., Li, H., Fan, Z.P. (2021). Tourism demand forecasting with time series imaging: A deep learning model. Annals of tourism Research, 90: 103255. https://doi.org/10.1016/j.annals.2021.103255