



Analysis of Domestic Tourist Forecasts to Bali Based on Destination City Using the LSTM Method

I Putu Bagus Arya Pradnyana^{1,*}, I Putu Oka Wisnawa¹,
and I Nyoman Rai Widartha Kesuma¹

¹ Information Technology Department, Politeknik Negeri Bali, Bali, Indonesia
bagusarya12@pnb.ac.id

Abstract. The recovery of Bali's tourism post-COVID-19 urgently requires spatial forecasting of domestic tourists due to the disparities between regencies/cities. The objective of this research is to predict the monthly number of domestic tourist visits to each regency/municipality in Bali from 2019 to 2025 using the Long Short-Term Memory (LSTM) deep learning method to support data-driven tourism policies. Model evaluation is comprehensively conducted using Mean Absolute Percentage Error (MAPE). The data is sourced from official publications by BPS Bali with a monthly time resolution per district. Based on testing the LSTM model across nine regions in Bali, the model demonstrates a good level of accuracy in five regions, achieving an average MAPE below 20%. Among these, Jembrana exhibits the best performance with a MAPE of 17.25%. Bangli shows a significant mathematical anomaly with a MAPE of 185.61% due to exceptionally low actual baseline values during specific periods. Overall, the developed LSTM model provides valuable spatial insights for tailored tourism mitigation and regional planning in Bali.

Keywords: Domestic tourists, Forecasting, LSTM, MAPE, City

1 Introduction

The tourism industry in Bali Province has long been recognized as the most dynamic engine of economic growth in Indonesia; however, the COVID-19 pandemic that struck in early 2020 exposed the structural fragility of an over-reliance on international markets. As foreign tourist arrivals plummeted to their lowest point in 2021, domestic tourists emerged as the primary frontline in sustaining the survival of local tourism industry stakeholders (Prasetyo et al., 2024). Data indicates that the recovery of the domestic sector progressed at a significantly faster pace; by 2023, the number of domestic tourist arrivals reached 9.87 million, a figure that surprisingly surpassed the government's target of 7 million visits (Jaya & Asmarajaya, 2025). This phenomenon underscores that a profound understanding of domestic tourist movement patterns is no longer merely a statistical supplement, but rather a strategic necessity in formulating regional economic policies.

Despite the positive trend in aggregate visits at the provincial level, a striking disparity remains in the distribution of tourists across regions in Bali (Badung, 2018). Southern areas, such as Badung Regency and Denpasar City, continue to experience high pressure from heavy visitor loads, whereas regencies in the northern and western regions, such as Buleleng and Jembrana, possess unoptimized potential. This imbalance is frequently caused by the lack of predictive systems capable of providing accurate, granular insights at the specific regency/city level (Ardyansyah, 2025). Without a precise forecasting model, local governments often adopt a reactive stance in managing visitor surges during peak holiday seasons, which in turn diminishes the quality of the tourist experience and threatens environmental sustainability (Nurhidayati & Nurchayati, 2025).

In the disciplines of computer science and data science, forecasting tourist arrivals poses unique technical challenges due to the highly volatile nature of the data, strong seasonal patterns, and unexpected external influences. Conventional statistical methods, such as ARIMA or Holt-Winters, often struggle to capture the non-linear relationships and long-term dynamics present in post-pandemic tourism data (Waciko et al., 2025). The emergence of Deep Learning architectures, particularly Long Short-Term Memory (LSTM), offers a viable solution through gating mechanisms that enable the model to selectively retain or discard crucial historical information (Vashellya et al., 2025). The primary research question addressed in this study is how to design and optimize an LSTM model capable of accommodating the characteristics of domestic tourist arrival data across the nine regencies/cities in Bali with a reliable level of accuracy.

2 Methodology

2.1 LSTM

Designed to bypass the limitations of vanishing and exploding gradients in traditional RNNs, the predictive model incorporates an LSTM algorithm equipped with memory cells (Sofia et al., 2025). These memory components utilize an input gate, a forget gate, and an output gate to govern the retention and deletion of information. In terms of network topology, the model integrates an LSTM input layer, a regularization dropout layer to prevent overfitting by temporarily disabling a fraction of neurons, and two consecutive dense layers consisting of a 32-neuron hidden layer alongside a linear single-neuron output layer (Hengyang, 2025).

Implemented via deep learning tools like PyTorch or TensorFlow, the fundamental network configuration includes an input layer, one hidden LSTM layer containing optimized units, and a linear dense output layer (Prabowo et al., 2025). To control the movement of data across time steps, each individual LSTM unit features three key regulatory gates:

- a. **Forget Gate(f_t):** Determines which historical information is no longer relevant and should be discarded.
- b. **Input Gate (i_t):** Determines which new information will be stored in the cell state.

- c. **Output Gate (o_t)**: Regulates the value to be propagated to the next hidden state.

The mathematical equations for updating the cell state (C_t) and the hidden state (h_t) are defined as follows (Sugiartawan et al., 2018):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t]) + b_f \quad (1)$$

$$C_t = f_t * C_{t-1} + i_t * \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2)$$

$$h_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) * \tanh(C_t) \quad (3)$$

Execution of the forecasting framework was carried out through the TensorFlow library. The model architecture follows a sequential design, initiating with an LSTM layer that leads into a dropout layer aimed at mitigating overfitting via the random deactivation of hidden nodes. Subsequently, the representation is processed through a 32-unit fully connected dense layer before reaching a single-neuron linear output layer responsible for generating continuous predictive estimations.

Optimal model performance was achieved by configuring the training hyperparameters with a learning rate of 0.001, a batch size of 4, and a total of 100 training epochs. Furthermore, each LSTM node leverages three core gating components—the forget, input, and output gates—to selectively preserve or purge temporal information across sequential steps, thereby mitigating the vanishing gradient challenge inherent in time-series forecasting.

2.2 Normalization

Data normalization involves rescaling attribute values to ensure they are constrained within a specific predetermined boundary. A widely utilized approach is the Min-Max normalization technique, which linearly maps the raw dataset into a range of [0.1, 0.9] (Pradnyana et al., 2025). This method proportionally rescales the values based on the mathematical formulation defined below:

$$X' = 0.8 \times \frac{(X - b)}{(a - b)} + 0.1 \quad (4)$$

In which:

X' : Rescaled data point

X : Raw/unprocessed observation

a : Maximum boundary of the raw feature

b : Minimum boundary of the raw feature

2.3 MAPE

Mean Absolute Percentage Error (MAPE) serves as a widely adopted metric for measuring predictive performance in time-series forecasting. This evaluation index is particularly advantageous when analyzing variables operating on substantial numerical scales (Zhang et al., 2025). By quantifying the relative discrepancy between forecasted outputs and observed ground-truth values, it measures the overall prediction error magnitude (Anisa et al., 2021). The mathematical formulation for MAPE is expressed below:

$$MAPE = \frac{1}{n} \sum \left| \frac{\text{Actual Value} - \text{Forecasted Value}}{\text{Actual Value}} \right| \times 100\% \quad (5)$$

Notes:

n = number of data points

2.4 Collecting Domestic Tourist Data

Domestic tourist visit records were gathered from the official data repository of the Bali Central Statistics Agency (BPS Bali). Spanning from 2019 through 2025, the compiled dataset comprises monthly observation resolution covering all nine regencies and municipalities in Bali Province. The attributes captured include recording year, administrative regency, and monthly visitor arrivals. To rigorously evaluate the predictive capacity of the proposed model, the dataset was chronologically partitioned into an 80% training subset and a 20% testing subset. A representative sample of domestic tourism records for Denpasar City across the specified timeframe is illustrated in Table 1.

Table 1. Denpasar City Tourist Data (2019–2025)

Year	Regency	Number of Tourist
2019	Denpasar	4051988
2020	Denpasar	1090766
2021	Denpasar	1259006
2022	Denpasar	1984425
2023	Denpasar	3526734
2024	Denpasar	4478649
2025	Denpasar	5354919

Based on the observations detailed in Table 1, tourist numbers in Denpasar generally followed a positive annual trend, except during 2020, which experienced a marked contraction. The specific parameter extracted for predictive modeling is the volume of tourist arrivals. This data collection methodology was applied consistently to gather historical records for all other local jurisdictions in Bali spanning from 2019 to 2025. Following aggregation, the raw figures were submitted to the data normalization phase prior to model execution (Hariyanto et al., 2025).

3 Result and Discussion

3.1 Result

According to the empirical forecasting outcomes for domestic tourist arrivals across Bali Province, the Mean Absolute Percentage Error (MAPE) exhibited varying degrees of precision among jurisdictions. Jembrana achieved the most superior predictive accuracy with the lowest MAPE of 17.25%, followed closely by Buleleng (19.12%), Tabanan (19.39%), Gianyar (19.60%), Denpasar (19.62%), Karangasem (20.46%), and Klungkung (20.62%), all demonstrating highly favorable model performance. Conversely, higher error margins were observed in Badung (30.23%) and particularly Bangli (185.61%), reflecting diminished forecasting reliability in these specific territories. Across all nine administrative areas, the mean aggregate MAPE stood at 39.10% (see Table 2). Overall, the proposed LSTM framework demonstrated strong predictive capability across the majority of Bali's regions. Graphical comparisons contrasting actual versus predicted tourist trajectories for Jembrana Regency (representing the best performance) and Bangli Regency (representing the worst performance) are illustrated in Figure 1 and Figure 2, respectively.

Table 2. Result MAPE LSTM and SVM

City	LSTM MAPE (%) (This Study)	SVM MAPE (Pradnyana et al., 2025)
Jembrana	17.25	2.245
Tabanan	19.39	0.1375
Badung	30.23	0.072
Gianyar	19.60	0.0804
Klungkung	20.62	0.0413
Bangli	185.61	9.7632
Karangasem	20.46	0.0487
Buleleng	19.12	0.0233
Denpasar	19.62	0.1536

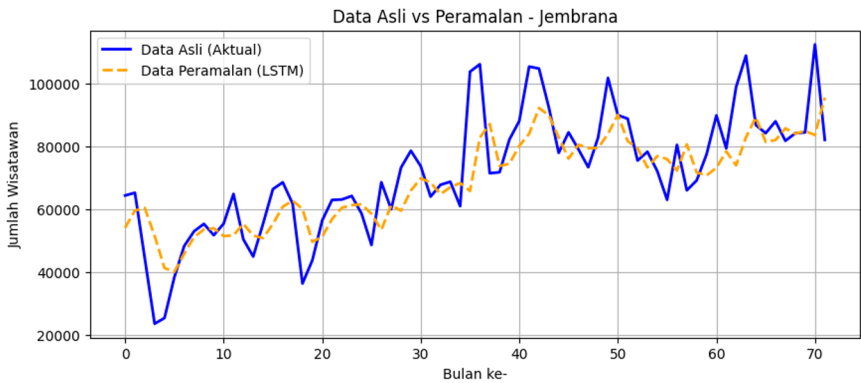


Figure 1. Comparison of Actual vs. Predicted Arrivals of Jembrana City

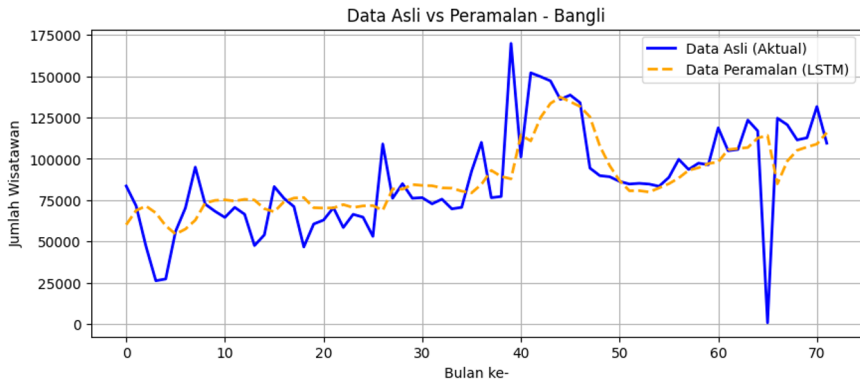


Figure 2. Comparison of Actual vs. Predicted Arrivals of Bangli City

Based on the graphic visualization, the LSTM model for Jembrana City successfully replicated the contour of the actual data line with high precision, capturing both the long-term upward trend and responding effectively to the regular peaks and troughs of seasonal fluctuations from the initial to the final months. Conversely, the graph for Bangli City revealed extreme pattern disruptions; although the LSTM model attempted to follow the general trend trajectory, the prediction line exhibited a response lag during a sudden surge around the 40th month and displayed a substantial deviation gap when the actual data plummeted sharply to near-zero levels at the 65th month.

3.2 Discussion

The empirical findings of this research demonstrate the capability of the Long Short-Term Memory (LSTM) network in generating highly accurate domestic tourist arrival forecasts across municipal jurisdictions in Bali Province. The developed LSTM architecture successfully captured complex temporal dynamics, including seasonal fluctuations and underlying visitation trends, notably within major tourist destinations such as Jembrana, Buleleng, and Tabanan. Unlike traditional statistical forecasting approaches, the LSTM algorithm exhibits superior competency in modeling non-linear patterns and complex time-series behaviors. Nevertheless, the predictive efficacy of this deep learning model remains contingent upon robust data pre-processing and optimal hyperparameter tuning. Overall, these insights underscore the viable integration of machine learning frameworks into regional tourism planning and strategic policy formulation.

The results of this study demonstrate that the Long Short-Term Memory (LSTM) architecture is a robust approach for forecasting domestic tourism arrivals across all nine regencies and cities in Bali. This research significantly extends the previous baseline study by (Pradnyana, 2025), which utilized the Support Vector Machine (SVM) method to forecast nine regions.

A direct comparison of the Mean Absolute Percentage Error (MAPE) reveals a notable disparity in scales between the two methods. While the SVM baseline reported

very low error values (e.g., 0.0233 for Buleleng and 0.0413 for Klungkung), the LSTM model in this study yielded higher MAPE values, such as 19.12% for Buleleng and 20.62% for Klungkung. This difference suggests a variation in the reporting units; the SVM results likely represent decimal error values, whereas the results in this study are expressed as direct percentages. When interpreted using standard forecasting scales, the MAPE values for most regions in this study (ranging from 17.25% to 20.62%) are classified as "Good" accuracy.

One of the most significant findings is the consistent identification of Bangli as a statistical anomaly by both the SVM and LSTM models. The SVM study reported its highest error in Bangli (9.7632), and this study similarly recorded a MAPE of 185.61% for the same region. This consistency across different algorithms suggests that the high error rate is not a failure of the LSTM architecture, but rather a result of the low actual baseline values in Bangli during specific periods, which causes even small absolute deviations to result in disproportionately high percentage errors.

The overall mean MAPE across the nine regions reached 39.10%, a figure heavily influenced by the Bangli anomaly. Excluding such outliers, the LSTM model demonstrated a strong capability in capturing non-linear relationships and seasonal fluctuations that conventional methods often struggle to identify. These findings reinforce the potential of deep learning applications in regional tourism planning, provided that data preprocessing specifically addresses regions with highly volatile or low-volume visitor arrivals.

4 Conclusion

This study has successfully developed an LSTM-based forecasting model to predict domestic tourist arrivals across all nine regencies and cities in Bali for the 2019–2025 period. The results indicate that the LSTM model achieved good prediction accuracy in most regions, with five areas recording MAPE values below 20%. Jembrana emerged as the region with the highest accuracy (MAPE 17.25%), followed by Buleleng and Tabanan.

A comparative analysis with the SVM baseline study shows that while LSTM provides broader spatial coverage (nine regions compared to six), it also confirms a significant mathematical anomaly in Bangli Regency. The exceptionally high MAPE in Bangli (185.61%) is consistent with previous findings and is attributed to the low actual visitor baseline during certain periods. Overall, this research provides a more comprehensive tool for regional tourism planning in Bali, highlighting that while deep learning is effective for non-linear tourism data, specific attention must be given to regions with high data volatility to ensure more targeted policy interventions.

References

- Anisa, M. P., Irawan, H., & Widiyanesti, S. (2021). Forecasting demand factors of tourist arrivals in Indonesia's tourism industry using recurrent neural network. *IOP Conference Series: Materials Science and Engineering*,

- Ardyansyah, E. (2025). *Penanganan Pungutan Levy terhadap Wisatawan Asing di Ulun Danu Beratan oleh Dinas Pariwisata Provinsi Bali Politeknik Negeri Bali*.
- Badung, B. K. (2018). Kabupaten Badung dalam Angka Badung Regency in Figures 2018. *CV. Bhineka Karya. Badung*.
- Hariyanto, H., Susanto, E., Aji, B., Goeltom, A. D. L., & Widyastuti, S. U. (2025). Strategic roadmap for AI-driven sustainable tourism development in Indonesia: An Expert-Based MCDM Approach. *Jurnal Kepariwisata Indonesia: Jurnal Penelitian dan Pengembangan Kepariwisata Indonesia*, 19(1), 1-14.
- Hengyang, F. (2025). Hybrid LSTM–TCN Architecture for Seasonal Income Forecasting in Rural Tourism Using Macroeconomic and Tourism Indicators. *Informatica*, 49(35).
- Jaya, I. K. N. A., & Asmarajaya, I. K. A. (2025). ANALISIS DATA KUNJUNGAN WISMAN DAN WISNUS KE BALI PADA DINAS PARIWISATA PROVINSI BALI. *Aradhana*, 2(1), 19-24.
- Nurhidayati, H., & Nurchayati, Z. (2025). *Transformasi pariwisata menuju keberlanjutan dan daya saing global*. Deepublish.
- Prabowo, I. A., Wijayanto, H., Utami, Y. R. W., Muhariya, A., Saputro, I. A., & Fadila, F. (2025). Predicting Tourist Visits Using LSTM for Smart Eco-Tourism Management in Banyuanyar Village, Indonesia: A Comparative Analysis with ARIMA and Prophet. *Ingénierie des Systèmes d'Information*, 30(12).
- Pradnyana, I. P. B. A., Wisnawa, I. P. O., & Wiratama, I. K. (2025). Forecasting the Number of Domestic Tourists Based on Destination City Using the Support Vector Machine (SVM) Method in Bali Province. International Conference on Sustainable Green Tourism Applied Science-Social Applied Science 2025 (ICOSTAS-SAS 2025),
- Prasetyo, A. R., Mariel, W. C. F., Ermawan, G. Y., & Pramana, S. (2024). Big data analytics for forecasting tourism recovery in Bali Island using multivariate time series. *Jurnal Kepariwisata Indonesia: Jurnal Penelitian dan Pengembangan Kepariwisata Indonesia*, 18(2), 331-350.
- Sofia, A., Dien, Z., & Erda, G. (2025). Forecasting International Tourist Arrivals to Indonesia Using LSTM: Post-Pandemic Analysis for 2024-2025. *Journal of Mathematics, Computations Statistics*, 8(2), 316-324.
- Sugiartawan, P., Permana, A. A. J., & Prakoso, P. I. (2018). Forecasting Kunjungan Wisatawan Dengan Long Short Term Memory (LSTM). *Jurnal Sistem Informasi dan Komputer Terapan Indonesia*, 1(1), 43-52.
- Vashellya, Z., Nurmawati, E., & Sugiyarto, T. (2025). The Impact of Online Reviews to Predict The Number of International Tourists. *Jurnal Online Informatika*, 10(1), 77-92.
- Waciko, K. J., Susanti, L. A., Muayyad, M., & Fakhrurozi, R. N. J. I. (2025). Forecasting Tourist Arrivals in Bali: A Grid Search-Tuned Comparative Study of Random Forest, XGBoost, and a Hybrid RF-XGBoost Model. *Inferensi*, 8(3), 251-261.
- Zhang, Y., Tan, W. H., & Zeng, Z. (2025). Tourism demand forecasting based on a hybrid Temporal neural network model for sustainable tourism. *Sustainability*, 17(5), 2210.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

