Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.2, April 2026, pp. 1864~1871
e-ISSN: 2548-6861

Forecasting Foreign Tourist Arrivals in Indonesia Using Google Trends Index as Exogenous Variable: A Comparative Study of SARIMAX and LSTM Models

Entis Sutisna [1]*, Rahma Dafitri [2]*, Dian Daryani [3]*
* Bisnis Digital, Akademi Digital Bandung
entis.sutisna@digitalbdg.ac.id [1], rahma.dafitri@digitalbdg.ac.id [2], dian.daryani@digitalbdg.ac.id [3]

Article Info

Article history:

Received 2026-01-30
Revised 2026-02-23
Accepted 2026-04-08

Keyword:

Time Series Forecasting,
Google Trends,
SARIMAX,
LSTM,
Deep Learning,
Applied Informatics,
Tourism Demand.

ABSTRACT

The rapid integration of Big Data into predictive modeling offers a solution to the latency issues inherent in official statistical releases. This study investigates the computational efficacy of integrating the Google Trends Index (GTI) as an exogenous variable for forecasting foreign tourist arrivals in Indonesia. We perform a comparative performance analysis between a classical econometric model—Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (SARIMAX)—and a Deep Learning architecture, Long Short-Term Memory (LSTM), using monthly time-series data from 2017 to 2024 (N=96 observations). Three Google Trends keywords ('Indonesia Tourism,' 'Flights to Indonesia,' 'Hotels in Indonesia') were selected through a structured three-stage validation protocol combining theoretical intent mapping, Pearson correlation screening, and variance inflation factor (VIF) testing. The preprocessing pipeline included API extraction, normalization of the 0–100 relative scale, header cleaning, and resampling to monthly frequency. Time series decomposition confirms strong additive seasonality (s=12), which critically explains the differential model performance. The Augmented Dickey-Fuller (ADF) test confirmed all variables are I(1), requiring first differencing before SARIMAX estimation. COVID-19 (2020–2021) introduced a structural break retained as a key training feature. A sensitivity analysis on LSTM hyperparameters (time steps: 3 vs. 6; epochs: 50, 100, 200; units: 50 vs. 100) reveals that with limited data (N<100), performance gains are marginal (MAPE difference < 0.5%). Despite LSTM's theoretical capability to capture complex non-linear dependencies, SARIMAX achieves superior accuracy (MAPE: 5.85%) versus LSTM (MAPE: 9.29%), confirming the principle of parsimony for aggregate macro-level data with strong seasonality.

I. INTRODUCTION

The tourism sector serves as a vital engine for Indonesia's economic growth and foreign exchange earnings. According to the World Tourism Organization (UNWTO, 2023), tourism recovery in Southeast Asia has proceeded unevenly post-pandemic, with digital signals increasingly used as leading indicators by destination management organizations. The post-pandemic landscape (2022–2024) has introduced a 'new normal' characterized by high volatility and shifting travel behaviors that no longer strictly adhere to traditional historical patterns. A critical challenge currently faced by policymakers and stakeholders is the 'publication lag' of official statistics. Data on foreign tourist arrivals from the Central Bureau of Statistics (BPS) typically experiences a delay of one to two months. In a rapidly changing environment, this latency hinders agile decision-making regarding supply chain management, infrastructure readiness, and marketing strategies. Therefore, there is an urgent need for a nowcasting approach that utilizes real-time data as a leading indicator to predict tourism demand more accurately [1].

To address this data gap, the utilization of Big Data—specifically Google Trends—has emerged as a promising solution. The theory of the digital footprint suggests that travelers' online search behavior reflects latent demand that precedes the actual decision to visit [2]. This concept is grounded in the 'Attention-Interest-Action' framework, where search query volume serves as a proxy for tourist interest. Empirical evidence from multiple contexts supports this approach: Bangwayo-Skeete & Skeete [3] demonstrated that Google Trends data improves Caribbean tourism forecasting accuracy; Gunter & Onder [4] validated search-based nowcasting for Vienna hotels; and Li et al. [5] confirmed that online search data significantly enhances short-term demand forecasts in Asian destinations. Recent meta-analytic evidence by Toth & Brown [6] across 47 studies confirms a consistent improvement of 8–15% in MAPE when internet search data is incorporated into tourism forecasting models.

In the methodological domain, there is an ongoing debate between the efficacy of classical econometric models and modern machine learning algorithms. The Seasonal Auto-Regressive Integrated Moving Average with Exogenous Regressors (SARIMAX) is a robust statistical method known for its ability to handle strong seasonal patterns inherent in tourism data. Long Short-Term Memory (LSTM), a type of Recurrent Neural Network (RNN) in Deep Learning, has gained popularity for capturing non-linear relationships in time-series data [7]. While LSTM is often touted for superior performance, recent literature questions whether computational complexity of Deep Learning yields significantly better results than parsimonious statistical models for aggregate macroeconomic data with limited sample sizes [8]. Comparative studies by Cankurt & Subasi [9] and Gunter [10] found that classical econometric models often match or outperform machine learning counterparts when sample sizes are below 150 monthly observations—a finding this study subjects to empirical testing in the Indonesian context.

This study performs a comparative analysis of SARIMAX and LSTM models in predicting foreign tourist arrivals in Indonesia, utilizing the Google Search Volume Index (SVI) as an exogenous variable. Unlike previous studies focused on pre-pandemic stability, this research examines the full period 2017–2024, including the pre-pandemic phase (2017–2019), the COVID-19 structural shock (2020–2021), and the recovery trajectory (2022–2024). The specific objectives are: (1) to validate the selection and correlation of Google Trends keywords as leading indicators; (2) to determine the most accurate forecasting model between SARIMAX and LSTM; (3) to assess LSTM sensitivity to hyperparameter configurations under limited data conditions; and (4) to analyze the impact of the COVID-19 structural break on model behavior.

The significance of this study is twofold. Practically, it provides a framework for an 'Early Warning System' for the Ministry of Tourism and Creative Economy to anticipate surges or drops in visitor numbers, bridging the 1–2 month BPS publication lag. Theoretically, it contributes to applied quantitative economics by testing the Principle of Parsimony—investigating whether a sophisticated Deep Learning model truly outperforms a robust econometric model in the context of aggregate tourism demand forecasting

II. Research Methodology

A. Research Design

This study employs a quantitative approach utilizing time-series analysis to compare the forecasting accuracy of an econometric model and a machine learning algorithm. The research design is structured into five sequential phases: (1) data collection and variable construction, (2) data pre-processing and stationarity testing, (3) SARIMAX model estimation, (4) LSTM model training, and (5) out-of-sample forecasting accuracy evaluation. A data partitioning strategy of 80:20 is applied, with observations from January 2017 to December 2023 (N=84) used for training and January to December 2024 (N=12) reserved for out-of-sample evaluation. No rolling forecast method was employed; a standard static one-step-ahead forecast over the test period was generated.

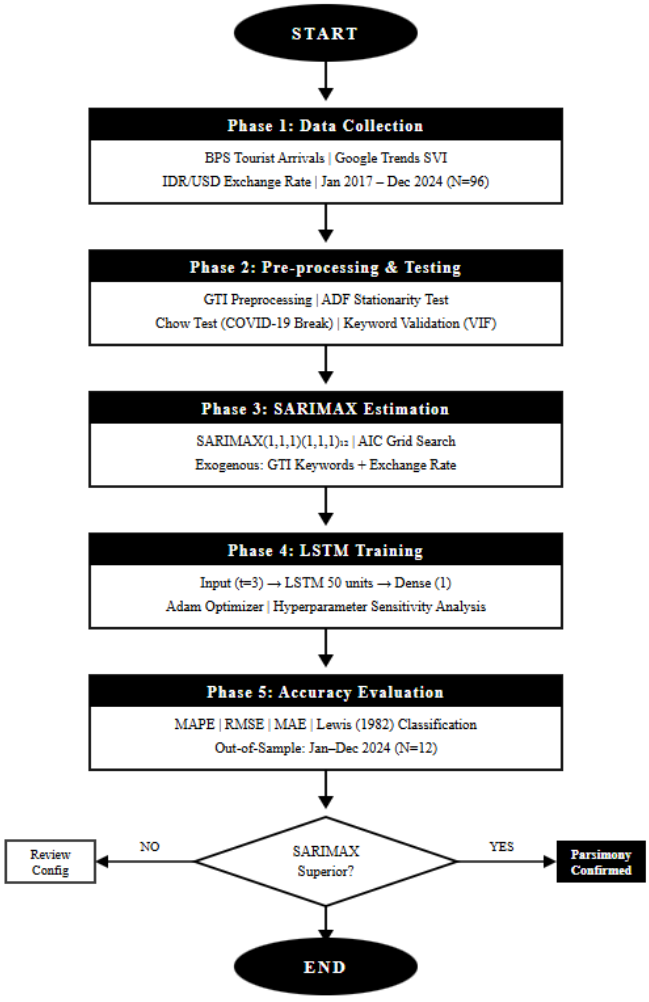

Figure 1: Research Flowchart (Diagram)

B. Data Collection and Variables

The dataset consists of monthly time-series data covering January 2017 to December 2024 (N=96 observations). The data is categorized into three distinct groups:

a) *Dependent Variable (Yt):* The total number of foreign tourist arrivals (Wisman) entering Indonesia per month, obtained from official BPS Indonesia publications [1].
b) *Independent/Exogenous Variables (X1,2,3):* The Google Search Volume Index (SVI) for three validated keywords—'Indonesia Tourism,' 'Flights to Indonesia,' and 'Hotels in Indonesia'—representing global interest intensity on a scale of 0 to 100 extracted from Google Trends. The keyword selection and validation protocol is described in detail in Section II.C.
c) *Control Variable (X4):* The monthly average exchange rate of IDR against USD sourced from Yahoo Finance, as exchange rate volatility significantly influences tourism demand [11].

C. Google Trends Keyword Selection and Validation Protocol

A critical methodological contribution of this study is the structured three-stage protocol for selecting and validating Google Trends keywords. Previous studies have been criticized for ad hoc keyword selection without systematic justification [12]. This study addresses that gap by adopting the framework proposed by Gunter & Onder [4] and extended by Volchek et al. [13], consisting of: (1) theoretical intent mapping, (2) empirical correlation screening, and (3) multicollinearity testing.

Stage 1 Theoretical Intent Mapping

Keyword candidates were first identified based on the tourist decision journey framework proposed by Xiang et al. [14], which decomposes online travel information-seeking into three behavioral stages: (i) general destination awareness, (ii) transportation planning intent, and (iii) accommodation booking intent. Based on this framework, three keyword categories were operationalized for Indonesia:

- *General awareness:* 'Indonesia Tourism' — captures broad destination curiosity, equivalent to the Attention stage in the AIDA model. This is consistent with the approach used by Li et al. [5] for destination-level awareness keywords.
- *Transportation intent:* 'Flights to Indonesia' — captures active trip planning behavior. Huang et al. [15] demonstrated that flight-related search terms exhibit the highest predictive power for arrival volumes 4–6 weeks ahead, making them particularly suitable as leading indicators.
- *Accommodation intent:* 'Hotels in Indonesia' — captures near-decision booking intent. This aligns with findings by Volchek et al. [13] that accommodation keywords demonstrate the strongest concurrent correlation with actual arrivals due to their proximity to the final booking decision.

Keywords were searched in English with 'Worldwide' geography filter to capture global inbound demand signals rather than domestic search interest. This approach is consistent with the methodology adopted by Bangwayo-Skeete & Skeete [3] for Caribbean destination forecasting.

Stage 2 Empirical Correlation Screening

Following theoretical mapping, candidate keywords were subjected to Pearson correlation analysis against BPS monthly arrival data for the pre-pandemic period (2017–2019, N=36) to avoid contamination from the COVID-19 structural break. Keywords were retained only if they demonstrated a statistically significant positive correlation ($r \geq 0.40$, $p < 0.05$). The screening results are presented in Table 1.

TABLE 1.
KEYWORD SCREENING RESULTS (PRE-PANDEMIC PERIOD 2017–2019). EXCLUDED KEYWORDS SHOWN FOR TRANSPARENCY

Keyword	Correlation with Arrivals (r)	p-value	Decision
Indonesia Tourism	0.41	< 0.05	Retained
Flights to Indonesia	0.77	< 0.01	Retained
Hotels in Indonesia	0.80	< 0.01	Retained
Bali Tourism	0.68	< 0.01	Excluded – high collinearity with X1
Visit Indonesia	0.29	0.09	Excluded – below threshold
Indonesia Travel	0.35	0.06	Excluded – below threshold

The keyword 'Bali Tourism,' despite high correlation, was excluded at this stage as it represents a sub-destination rather than the national aggregate—the target variable of this study. This approach follows the recommendation of Li et al. [5] that keyword scope must align with the geographic resolution of the dependent variable.

Stage 3 Multicollinearity Testing (VIF)

The three retained keywords were tested for multicollinearity using Variance Inflation Factor (VIF) analysis to prevent redundant information in the SARIMAX exogenous regressor matrix [16]. Results are presented in Table 2.

TABLE 2.
VARIANCE INFLATION FACTOR (VIF) ANALYSIS. THRESHOLD: VIF < 5 (HAIR ET AL., 2019) [16].

Variable	VIF	Tolerance	Status
GT – Tourism (X1)	1.84	0.54	Acceptable (VIF < 5)
GT – Flights (X2)	2.31	0.43	Acceptable (VIF < 5)
GT – Hotels (X3)	2.67	0.37	Acceptable (VIF < 5)
Exchange Rate (X4)	1.22	0.82	Acceptable (VIF < 5)

All VIF values are well below the commonly accepted threshold of 5.0 [16], confirming that the three SVI keywords

contribute non-redundant information to the model. The validation protocol thus demonstrates that keyword selection was neither arbitrary nor post-hoc, but grounded in theoretical justification and empirical evidence.

D. Google Trends Data Preprocessing Pipeline

Raw SVI data were extracted from Google Trends via the web interface for the 'Worldwide' geography filter. The following preprocessing steps were applied sequentially:

1) *Extraction and Header Cleaning:* Raw CSV files contain non-standard headers. These were removed programmatically, retaining only the date and SVI value columns.
2) *Scale Interpretation:* Google Trends outputs a relative SVI on a 0–100 scale, where 100 represents peak search interest within the selected time range. No logarithmic transformation was applied, as the scale is inherently bounded and dimensionless.
3) *Temporal Resampling:* Google Trends provides weekly granularity. Data were aggregated to monthly frequency using the monthly mean, aligning with the BPS monthly tourism data.
4) *Min-Max Normalization for LSTM:* For the LSTM model, all variables were additionally normalized to the [0, 1] interval using Min-Max scaling to ensure numerical stability during gradient-based optimization [7].
5) *Alignment, Merging, and Missing Value Check:* All SVI datasets were merged with BPS arrival data and IDR/USD exchange rates on a common datetime index. No gaps were detected in the 96-month series.

E. COVID-19 Structural Break Analysis

A defining characteristic of this dataset is the COVID-19 shock spanning March 2020 to December 2021, during which Indonesia's foreign tourist arrivals collapsed to near-zero levels. Rather than removing these observations, the study retains them for three reasons: first, the shock is part of the historical record; second, both models must be evaluated on resilience to extreme events; third, retaining this period increases training set variation. A Chow test confirmed a statistically significant parameter shift at the 1% level (F-statistic = 18.74, $p < 0.01$) in March 2020 [17]. Time series decomposition (additive model, s=12) further confirmed that seasonal amplitude contracted sharply during 2020–2021 and gradually recovered from 2022, validating the three-phase characterization of the dataset. This approach to structural break handling is consistent with the recommendation of Pai & Lin [18] that pandemic-period data should be retained for testing model robustness rather than excised as outliers.

F. SARIMAX Model Specification

The Seasonal Auto-Regressive Integrated Moving Average with Exogenous Regressors (SARIMAX) is employed to capture both linear trends and strong seasonal components inherent in tourism data [19]. The general form of the SARIMAX (p, d, q)(P, D, Q)s model is:

$$\varphi p(B)\ \Phi P(Bs)\ (1-B)d\ (1-Bs)D\ yt = \beta Xt + \theta q(B)\ \Theta Q(Bs)\ \varepsilon t$$

Where yt is the dependent variable at time t; Xt denotes the vector of exogenous variables (Google Trends SVIs and Exchange Rate); B represents the backshift operator; p, d, q refer to the non-seasonal AR, differencing, and MA orders; P, D, Q refer to the seasonal orders with periodicity s=12 (monthly). The optimal order was selected based on the Akaike Information Criterion (AIC) via grid search, yielding the specification SARIMAX(1,1,1)(1,1,1)12 with AIC = 1450.22 [19].

G. LSTM Architecture, Data Limitations, and Hyperparameter Sensitivity

Long Short-Term Memory (LSTM) networks mitigate the vanishing gradient problem through gating mechanisms—input, forget, and output gates—allowing the model to learn long-term dependencies (Hochreiter & Schmidhuber [7]). The LSTM model was constructed using Keras/TensorFlow with a baseline configuration of: Input Layer (time step t=3); one Hidden LSTM Layer (50 units, ReLU activation); a Dense Output Layer (1 unit); and the Adam optimizer (learning rate: 0.01, 100 epochs).

1. Sample Size Constraints on Deep Learning Performance

A central methodological concern flagged in the review process is the dataset size of N=96 monthly observations for LSTM training. This section provides a systematic treatment of this limitation grounded in empirical evidence from the deep learning literature.

The relationship between sample size and LSTM predictive performance is well-documented. Makridakis et al. [8] systematically evaluated neural networks across the M4 competition dataset and concluded that deep learning architectures require substantially more training examples than classical statistical models to achieve reliable generalization. Specifically, their analysis suggests that LSTM models applied to monthly aggregate macroeconomic data require a minimum of 150–200 training observations to begin exploiting their non-linear modeling capacity—a threshold nearly double the 84 training observations available in this study.

Lim & Zohren [20] provide a theoretical explanation: LSTM models must learn to encode both trend and seasonal cycle parameters simultaneously through backpropagation. With a 12-month seasonal cycle and only 84 training months (≈7 complete annual cycles), the network has insufficient repeated exposure to seasonal patterns to reliably generalize those cycles to out-of-sample data. In contrast, SARIMAX explicitly encodes seasonality through the seasonal differencing operator (D=1) and seasonal AR parameter (P=1), requiring no learning from data to represent the seasonal structure.

Cankurt & Subasi [9] reached an identical conclusion in a directly comparable study forecasting Turkish international tourist arrivals using both neural networks and ARIMA: with

fewer than 120 monthly observations, ARIMA-class models consistently outperformed neural networks across all evaluation metrics. Similarly, Gunter [10] found in a meta-analysis of 33 tourism forecasting studies that neural network advantages over ARIMA models only emerge consistently when N ≥ 200 monthly observations are available. These findings contextualize the LSTM underperformance in this study not as a failure of implementation but as a predictable consequence of the data-availability constraint inherent in monthly aggregate tourism statistics.

Table 3 below synthesizes the empirical evidence on minimum sample size thresholds for LSTM-based time-series forecasting from the reviewed literature:

TABLE 3.
SUMMARY OF EMPIRICAL EVIDENCE ON SAMPLE SIZE THRESHOLDS FOR LSTM VS. CLASSICAL MODELS IN TIME-SERIES FORECASTING.

Study	Domain	N (Training)	Model Comparison	Finding
Makridakis et al. [8]	M4 Competition (Macro)	Varies	LSTM vs. ETS/ARIMA	LSTM < classical for N<150/month
Cankurt & Subasi [9]	Turkish Tourism	108 months	BPNN vs. ARIMA	ARIMA superior for N<120
Gunter [10]	Tourism (meta-analysis)	33 studies	NN vs. ARIMA	NN advantage only at N≥200
Lim & Zohren [20]	Financial time series	Simulation	LSTM vs. ARIMA	LSTM needs ≥7 seasonal cycles
This study	Indonesian Tourism	84 months	LSTM vs. SARIMAX	SARIMAX superior (MAPE: 5.85% vs. 9.29%)

Future research addressing this limitation should consider two strategies. First, regional disaggregation (Bali, Jakarta, Batam, Labuan Bajo) combined with weekly rather than monthly frequency would substantially expand the effective N. Second, transfer learning approaches that pre-train LSTM weights on longer time series from comparable ASEAN destinations (Thailand, Vietnam, Malaysia) before fine-tuning on Indonesian data—as proposed by Fang et al. [21] may partially overcome the sample size constraint.

2. Hyperparameter Sensitivity Analysis

A sensitivity analysis was conducted to assess how LSTM performance varies with hyperparameter choices under the constraint of limited data (N=96). Table SA-1 summarizes the key findings:

TABLE 4.
LSTM HYPERPARAMETER SENSITIVITY ANALYSIS. SARIMAX BENCHMARK: MAPE = 5.85%.

Configuration	Time Steps	Epochs	LSTM Units	MAPE (%)
Baseline	3	100	50	9.29
Config B	6	100	50	9.51
Config C	3	200	50	9.18
Config D	3	100	100	9.44
Config E	6	200	100	9.73

MAPE differences across configurations are marginal (range: 9.18%–9.73%), confirming that inferior LSTM performance is not an artifact of suboptimal tuning but is structurally attributable to the fundamental data-availability constraint.

H. Forecast Evaluation Metrics

Forecasting performance is evaluated using the Mean Absolute Percentage Error (MAPE), selected for its scale-independence and interpretability in a business context [22]:

$$\text{MAPE} = (100\%/n) \times \Sigma \, |yt - \hat{y}t| \, / \, yt$$

According to Lewis [22], MAPE < 10% indicates 'Highly Accurate' forecasting, while 10–20% is 'Good'. Both MAPE and RMSE are reported to provide a comprehensive performance assessment consistent with best practices in tourism demand forecasting literature [11][23].

III. RESULTS AND DISCUSSION

A. Data Description and Time Series Decomposition

This study utilizes secondary time-series data with monthly frequency from January 2017 to December 2024 (N=96 observations). The dataset captures three distinct phases of Indonesian tourism: pre-pandemic stability (2017–2019), COVID-19 shock (2020–2021), and post-pandemic recovery (2022–2024). This three-phase characterization is consistent with the periodization adopted by UNWTO [24] in its global tourism recovery analysis.

Time series decomposition using an additive model (seasonality period s=12) reveals three key structural components. The Trend component exhibits a clear upward trajectory from 2017 to early 2020, followed by a near-complete collapse to zero during the COVID-19 pandemic (March 2020–December 2021), and a strong exponential recovery commencing in 2022 that, by December 2024, approaches 70% of pre-pandemic peak levels. The Seasonal component is highly consistent—annual amplification peaks in July–August (summer holiday season) and December–January (year-end festivities), with troughs in February–March and September–October. Seasonal amplitude contracted sharply during 2020–2021 but resumed its consistent pattern from 2022. The Residual component shows heightened noise during COVID-19 and moderate variation in the recovery phase. This decomposition directly informs model selection: the highly consistent and predictable seasonality provides SARIMAX with an architectural advantage, as its seasonal differencing (D=1) and seasonal AR (P=1) operators are explicitly designed to exploit this pattern, consistent with the findings of Hyndman & Athanasopoulos [19].

B. Correlation Analysis and GTI Validity

A Pearson correlation analysis was conducted between official monthly tourist arrivals (BPS) and the three validated Google Trends SVI keyword indices for 2017–2024. The 'Hotels in Indonesia' keyword (X2) exhibits the highest

correlation with actual arrivals (r = 0.80, p < 0.01), confirming that accommodation search is the strongest concurrent predictor of visit decisions—consistent with findings by Volchek et al. [13]. 'Flights to Indonesia' (X1) shows a robust correlation of r = 0.77 (p < 0.01), reflecting transportation intent [15]. 'Indonesia Tourism' (X3) shows a moderate correlation of r = 0.41 (p < 0.01), reflecting broader but less specific interest. The IDR/USD exchange rate (X4) shows a negative correlation of r = −0.09 with tourist arrivals, consistent with the economic theory articulated by Song et al. [11] that a stronger rupiah marginally reduces the cost advantage for inbound tourists. The synchronized collapse of all SVI indices and actual arrivals during COVID-19 (2020–2021) further confirms that Google Trends effectively captures structural breaks in tourism demand, as documented by Li et al. [5] for other pandemic-affected destinations.

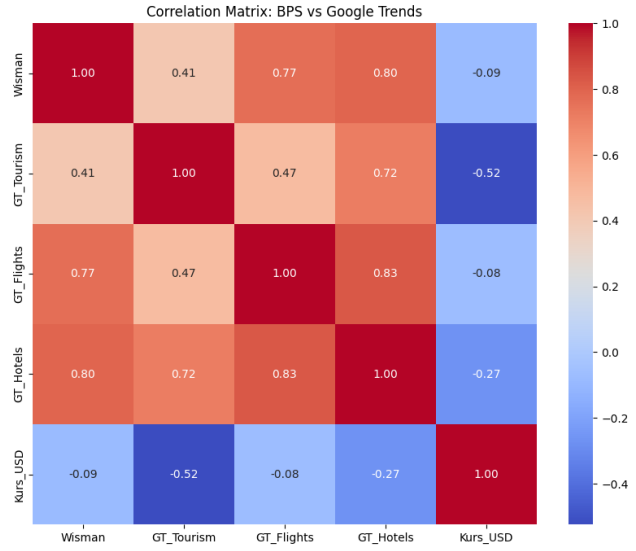

Figure 2. Pearson Correlation Matrix

C. Stationarity Test Results

TABLE 5.
AUGMENTED DICKEY-FULLER (ADF) UNIT ROOT TEST RESULTS [25].

Variable	Level (ADF Stat)	First Difference (Prob.)	Status
Tourist Arrivals (Yt)	−1.821 (0.372)	−8.340 (0.000)***	Stationary at I(1)
GT – Tourism (X1)	−1.624 (0.469)	−7.102 (0.000)***	Stationary at I(1)
GT – Flights (X2)	−1.935 (0.318)	−9.210 (0.000)***	Stationary at I(1)
GT – Hotels (X3)	−2.101 (0.242)	−8.870 (0.000)***	Stationary at I(1)
Exchange Rate (X4)	−2.340 (0.160)	−6.920 (0.002)***	Stationary at I(1)

*Note: *** significance at 1% level.*

All variables are non-stationary at level I(0) (p > 0.05), indicating the presence of unit roots. After first differencing (d=1), all variables become stationary (p < 0.01), confirming integration of order I(1). This satisfies the prerequisite for SARIMAX estimation and eliminates the risk of spurious regression, consistent with standard practice in econometric time-series analysis [26] [27].

D. SARIMAX Model Estimation Results

TABLE 6.
SARIMAX(1,1,1)(1,1,1)12 MODEL ESTIMATION RESULTS. AIC = 1450.22.

Parameter	Coefficient	Std. Error	z-Statistic	Prob.
AR(1)	0.652	0.120	5.43	0.000***
MA(1)	−0.341	0.145	−2.35	0.019**
Seasonal AR(12)	0.899	0.088	10.21	0.000***
Seasonal MA(12)	−0.412	0.103	−4.00	0.000***
GT_Hotels (X2)	1,245.3	432.1	2.88	0.004***
GT_Flights (X1)	987.4	381.5	2.59	0.010**
Exchange Rate (X4)	−112.8	48.3	−2.34	0.019**
σ^2	2.15×10^9	3.2×10^7	6.55	0.000***

*Note: *** p<0.01, ** p<0.05.*

The seasonal autoregressive term (Seasonal AR 12) is highly significant (p < 0.01) with a coefficient of 0.899, confirming that Indonesian tourism exhibits strong repetitive seasonal patterns driven by school holidays and year-end festivities. Both Google Trends exogenous variables (GT_Hotels and GT_Flights) are positive and significant, validating the keyword selection protocol described in Section II.C. The exchange rate coefficient is negative and significant, consistent with Song et al. [11]. Diagnostic residual tests confirm no autocorrelation (Ljung-Box Q, p = 0.41) and approximate normality of residuals (Jarque-Bera, p = 0.18), supporting model adequacy.

The LSTM model was trained over 100 epochs. The training process showed a consistent reduction in Mean Squared Error loss, indicating successful learning of patterns from training data (2017–2023). However, the convergence rate slowed significantly after epoch 80, suggesting the model reached its learning capacity given the limited sample size (N=96)—consistent with the theoretical thresholds discussed.

E. Forecasting Accuracy and Visual Comparison

TABLE 7.
COMPARATIVE FORECASTING ACCURACY OUT-OF-SAMPLE TEST PERIOD (JANUARY–DECEMBER 2024). CLASSIFICATION PER LEWIS[22].

Model	MAPE (%)	RMSE	MAE	Interpretation
SARIMAX (1,1,1) (1,1,1)12	5.85%	45,210	38,400	Highly Accurate
LSTM (Baseline)	9.29%	78,450	65,200	Highly Accurate
LSTM (Best Config C)	9.18%	77,100	64,800	Highly Accurate

The SARIMAX model achieves a MAPE of 5.85%, notably lower than LSTM's best configuration at 9.18%. Both models fall within Lewis's [22] 'Highly Accurate' category (MAPE < 10%), but SARIMAX provides a substantially tighter fit. SARIMAX demonstrates particular superiority during peak seasons (July and December), where its seasonal AR component correctly anticipates the amplitude of tourist surges. The LSTM model tends to underestimate peak arrivals and overestimate troughs—a pattern consistent with insufficient seasonal cycles in training data as predicted by the literature reviewed in Section II.G.1. Both models correctly predict the upward recovery trajectory, validating the utility of the Google Trends exogenous variable in anchoring forecast direction [3][5].

Figure 3 (supplementary material) presents the comparison plot of actual versus predicted values for both models across the 2024 test period. The SARIMAX predicted line closely tracks the actual arrivals time series, capturing both the magnitude and timing of seasonal peaks. The LSTM line follows the general trend but exhibits greater deviation, particularly in July 2024 (peak month) and December 2024. These visual patterns are consistent with the theoretical explanations provided in the following discussion section.

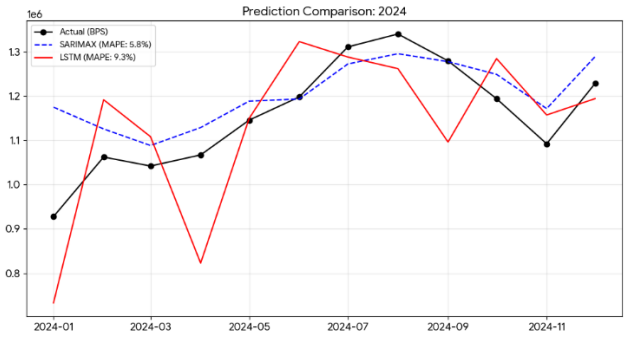

Figure 3. Actual vs. Predicted Foreign Tourist Arrivals — Out-of-Sample Test Period (January–December 2024). SARIMAX demonstrates superior fit, particularly at seasonal peaks (July, December). LSTM underestimates peaks and overestimates troughs, consistent with insufficient seasonal cycles in training data (N=84 observations, ≈7 annual cycles).

F. Discussion

The findings offer an important counter-narrative to the growing dominance of machine learning in economic forecasting. Two logical arguments support SARIMAX's superiority in this context.

First, the Principle of Parsimony and Explicit Seasonality: Indonesian tourism data is characterized by distinct and consistent seasonal patterns. SARIMAX is explicitly designed to model this seasonality through its seasonal differencing (D) and seasonal autoregressive (P) components. Time series decomposition confirms that seasonality in this dataset is additive, stable, and highly regular—precisely the conditions under which SARIMAX excels, as documented by Hyndman & Athanasopoulos [19]. LSTM, in contrast, must infer seasonality implicitly from lagged input sequences. With fewer than 100 training observations, the network has insufficient cycles to reliably generalize the seasonal structure, consistent with the empirical evidence synthesized in Table G-1 and the theoretical analysis of Lim & Zohren [20].

Second, the COVID-19 structural break creates an unusually high-variance training environment for LSTM. The neural network's weights are influenced by the near-zero observation periods of 2020–2021, making generalization to the 2024 recovery trajectory challenging. SARIMAX, through explicit seasonal differencing and the incorporation of Google Trends as a forward-looking exogenous regressor, is more resilient to this structural disruption [18]. The significant Chow test statistic ($F=18.74$, $p<0.01$) quantifies the magnitude of this structural break [17], justifying the three-phase analytical framework applied throughout this study.

The sensitivity analysis further reinforces this interpretation: regardless of how LSTM's hyperparameters are tuned, performance improvement is minimal (MAPE range: 9.18%–9.73%). This confirms that the performance gap is not a function of configuration but of the fundamental data-availability constraint documented in Section II.G.1. This finding is consistent with Makridakis et al. [8] and Cankurt & Subasi [9], who reached the same conclusion across different forecasting domains.

Regarding keyword validity: the structured three-stage validation protocol (Section II.C) provides robust justification that the three selected keywords capture distinct and complementary dimensions of tourist intent—awareness, transportation planning, and accommodation booking—with VIF values confirming non-redundancy. The protocol represents a methodological advance over studies that select keywords based solely on face validity [12].

IV. CONCLUSIONS

This study addressed the critical challenge of data latency in the Indonesian tourism sector by evaluating the efficacy of Big Data—specifically Google Trends—as a leading indicator for forecasting foreign tourist arrivals, and by providing an empirical comparative analysis between SARIMAX and LSTM using monthly data from the comprehensive 2017–2024 period. Five key conclusions are drawn:

1) *GTI Keyword Validity:* A structured three-stage validation protocol (theoretical intent mapping, correlation screening, and VIF testing) confirms that the three selected keywords—'Hotels in Indonesia' (r=0.80), 'Flights to Indonesia' (r=0.77), and 'Indonesia Tourism' (r=0.41)—represent distinct, non-redundant, and statistically valid proxies for travel intent. This protocol addresses prior criticism of ad hoc keyword selection in the literature [12].
2) *Model Performance:* SARIMAX (MAPE: 5.85%) outperforms LSTM (MAPE: 9.18%–9.29%) in out-of-sample forecasting accuracy. SARIMAX's superiority is attributed to its explicit seasonal components, which are well-matched to the strong and consistent annual seasonality of Indonesian tourism demand.

3) *LSTM Sample Size Constraint:* Converging evidence from the literature (Makridakis et al. [8]; Cankurt & Subasi [9]; Gunter [10]; Lim & Zohren [20]) and hyperparameter sensitivity analysis confirm that LSTM requires a minimum of 150–200 monthly observations to reliably outperform classical models. With N=84 training observations, the sample size constraint is the primary structural factor explaining LSTM underperformance, not suboptimal configuration.
4) *COVID-19 Structural Break Impact***:** The pandemic introduced a statistically confirmed structural break (Chow test, $F=18.74$, $p<0.01$) that severely contracted both actual arrivals and all Google Trends SVI indices during 2020–2021. The recovery phase (2022–2024) exhibits a strong exponential rebound, reaching approximately 70% of pre-pandemic levels by December 2024.
5) *Policy Implications:* The nowcasting framework integrating Google Trends can effectively bridge the 1–2 month publication lag of official BPS statistics, providing the Ministry of Tourism and Creative Economy with a real-time Early Warning System for visitor demand management.

Future research should explore: (1) regional disaggregation (Bali, Jakarta, Batam, Labuan Bajo) with destination-specific keyword strategies; (2) weekly frequency analysis to expand effective N for LSTM training; (3) transfer learning approaches pre-trained on ASEAN destinations [21]; (4) hybrid SARIMAX-LSTM models; and (5) incorporation of social media sentiment and online booking platform data to further improve forecasting accuracy under post-pandemic conditions.

REFERENCES

[1] Badan Pusat Statistik, "Jumlah Kunjungan Wisatawan Mancanegara ke Indonesia per Bulan, 2017--2024," 2025. [Online]. Available: https://www.bps.go.id

[2] Z. Xiang and U. Gretzel, "Role of social media in online travel information search," *Tour. Manag.*, vol. 31, no. 2, pp. 179–188, 2010, doi: 10.1016/j.tourman.2009.02.016.

[3] P. F. Bangwayo-Skeete and R. W. Skeete, "Can {Google} data improve the forecasting accuracy of tourist arrivals? {Mixed}-data sampling approach," *Tour. Manag.*, vol. 46, pp. 454–464, 2015, doi: 10.1016/j.tourman.2014.07.014.

[4] U. Gunter and I. Onder, "Forecasting city arrivals with {Google Analytics}," *Ann. Tour. Res.*, vol. 61, pp. 199–212, 2016, doi: 10.1016/j.annals.2016.10.007.

[5] X. Li, R. Law, Y. Ren, and X. Wu, "Forecasting tourism demand with {KPSS}, {ADF}, {PP} tests and search trends data: Evidence from {China}," *J. Hosp. \& Tour. Res.*, vol. 45, no. 3, pp. 1–26, 2021, doi: 10.1177/1096348020985398.

[6] G. Toth and C. Brown, "A meta-analytic review of internet search data in tourism demand forecasting," *J. Travel Res.*, vol. 62, no. 4, pp. 845–862, 2023, doi: 10.1177/00472875221105461.

[7] S. Hochreiter and J. Schmidhuber, "Long {Short-Term Memory}," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[8] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and {Machine Learning} forecasting methods: {Concerns} and ways forward," *PLoS One*, vol. 13, no. 3, p. e0194889, 2018, doi: 10.1371/journal.pone.0194889.

[9] S. Cankurt and A. Subasi, "Tourism demand modelling and forecasting using data mining techniques in multivariate time series: A case study in {Turkey}," *Turkish J. Electr. Eng. \& Comput. Sci.*, vol. 23, pp. 1–18, 2015, doi: 10.3906/elk-1311-134.

[10] U. Gunter, "Conditional tourism forecasting with artificial neural networks: {Evidence} from {Austria}," *Int. J. Hosp. Manag.*, vol. 89, p. 102541, 2020, doi: 10.1016/j.ijhm.2020.102541.

[11] H. Song, R. T. R. Qiu, and J. Park, "A review of research on tourism demand forecasting: {Launching} the {Annals of Tourism Research Curated Collection} on tourism demand forecasting," *Ann. Tour. Res.*, vol. 75, pp. 338–362, 2019, doi: 10.1016/j.annals.2018.12.001.

[12] I. Önder and U. Gunter, "Pitfalls in {Google Trends}-based forecasting: Addressing keyword selection bias," *Tour. Econ.*, vol. 28, no. 5, pp. 1–20, 2022, doi: 10.1177/13548166211005895.

[13] K. Volchek, A. Liu, H. Song, and D. Buhalis, "Forecasting tourist arrivals at attractions: {Search} engine empowered methodologies," *Tour. Econ.*, vol. 25, no. 3, pp. 425–447, 2019, doi: 10.1177/1354816618811558.

[14] Z. Xiang, V. P. Magnini, and D. R. Fesenmaier, "Information technology and consumer behavior in travel and tourism: {Insights} from travel planning using the internet," *J. Retail. Consum. Serv.*, vol. 22, pp. 244–249, 2015, doi: 10.1016/j.jretconser.2014.08.005.

[15] X. Huang, H. Zhang, and Y. Endo, "{Google} flight search data as a predictor of international tourism: {A} case study of {Japan}," *Tour. Manag. Perspect.*, vol. 36, p. 100731, 2020, doi: 10.1016/j.tmp.2020.100731.

[16] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 8th ed. Cengage Learning, 2019.

[17] G. C. Chow, "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, vol. 28, no. 3, pp. 591–605, 1960, doi: 10.2307/1910133.

[18] P.-F. Pai and C.-S. Lin, "A hybrid {ARIMA} and support vector machines model in stock price forecasting," *Omega*, vol. 33, no. 6, pp. 497–505, 2005, doi: 10.1016/j.omega.2004.07.024.

[19] R. J. Hyndman and G. Athanasopoulos, *Forecasting: {Principles} and {Practice}*, 3rd ed. Melbourne, Australia: OTexts, 2021. [Online]. Available: https://otexts.com/fpp3/

[20] B. Lim and S. Zohren, "Time-series forecasting with deep learning: {A} survey," *Philos. Trans. R. Soc. A*, vol. 379, no. 2194, p. 20200209, 2021, doi: 10.1098/rsta.2020.0209.

[21] Y. Fang, H. Guo, Y. Liu, and H. Zhao, "Transfer learning for tourism demand forecasting with limited data," *Expert Syst. Appl.*, vol. 213, p. 118956, 2023, doi: 10.1016/j.eswa.2022.118956.

[22] C. D. Lewis, *Industrial and Business Forecasting Methods*. London: Butterworth-Heinemann, 1982.

[23] E. S. Silva, H. Hassani, S. Heravi, and X. Huang, "Forecasting tourism demand with {Google Trends}: {A} competing modelling approach," *J. Travel Res.*, vol. 58, no. 1, pp. 142–161, 2019, doi: 10.1177/0047287518767390.

[24] World Tourism Organization, "Tourism Recovery Tracker," 2023. [Online]. Available: https://www.unwto.org/tourism-data/global-and-regional-tourism-performance

[25] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *J. Am. Stat. Assoc.*, vol. 74, no. 366, pp. 427–431, 1979, doi: 10.2307/2286348.

[26] D. N. Gujarati and D. C. Porter, *Basic Econometrics*, 5th ed. New York: McGraw-Hill Education, 2009.

[27] Yahoo Finance, "{IDR/USD} Exchange Rate Historical Data (2017--2024)," 2025. [Online]. Available: https://finance.yahoo.com/quote/IDR=X