



Tourism Marketing Intelligence: Machine Learning for Demand Forecasting Using Google Trends in Indonesia

A Khidhir Akbar Ghofar^{1*}, Muhammad Alif Hidayat², Adib³

^{1,2,3}Program Studi Bisnis Digital, Universitas Teknologi Bandung, Indonesia

E-Mail: ¹khidhirakbarghofar@utb-univ.ac.id, ²muhammadalif@utb-univ.ac.id, ³adib@utb-univ.ac.id

Received Mar 08th 2026; Revised Apr 24th 2026; Accepted May 07th 2026; Available Online Jul 20th 2026

Corresponding Author: A Khidhir Akbar Ghofar

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

Tourism demand forecasting remains challenging due to the dynamic nature of consumer behavior and the lag in traditional economic indicators. This study investigates the predictive power of online search behavior captured through Google Trends in forecasting international tourist arrivals to Indonesia, integrating digital footprints with macroeconomic indicators using machine learning approaches. Monthly data spanning January 2011 to December 2024 (N=168) were collected from multiple sources: tourist arrivals from Statistics Indonesia, Google Trends composite index from 11 destination-related keywords, exchange rates, and consumer price index. Three machine learning algorithms—Random Forest, XGBoost, and LSTM were compared using a temporal train-test split (80:20). Random Forest demonstrated superior performance with a Test MAPE of 14.12% and an R^2 of 0.811, outperforming XGBoost (MAPE 21.41%) and LSTM (MAPE 24.56%). Feature importance analysis revealed that Google Trends emerged as the second most important predictor (12.22% contribution), surpassing all economic indicators combined (7.20%), validating its role as a leading indicator of tourism demand. This study contributes methodologically by providing empirical evidence for the value of integrating digital behavioral signals with traditional predictors in emerging destination contexts, and practically by demonstrating deployment-ready forecasting accuracy for tourism stakeholders in Indonesia.

Keywords: Digital Footprints, Google Trends, Machine Learning, Random Forest, Tourism Demand Forecasting

1. INTRODUCTION

International tourism contributes over 10% of world GDP and employment [1]. For Indonesia, the sector occupies a position of strategic importance: international arrivals grew from approximately 7 million in 2011 to more than 16 million in 2019, before the COVID-19 pandemic brought this trajectory to an abrupt halt. By 2024, tourism's contribution to Indonesia's GDP stood at 5.1%, with projections suggesting it could reach USD 133.3 billion by 2029 [2]. The pandemic caused arrivals to decline by as much as 80%, underscoring the critical need for accurate and responsive demand forecasting systems to support sectoral recovery and long-term planning.

Conventional forecasting approaches classical time series models such as ARIMA and SARIMA, and macroeconomic econometric models have long served as the standard in the forecasting literature [3]. While these methods capture seasonal patterns and long-run trends reasonably well, they share a structural limitation: macroeconomic data are typically available only at low frequencies, quarterly or annually, and are subject to considerable publication lags, making them poorly suited to tracking real-time shifts in consumer behavior [4]. Against this backdrop, the digital transformation of travel planning has created a new class of data. Search engines such as Google now serve as the primary point of contact between prospective travelers and their destinations of interest, generating large-scale digital footprints in the form of online search queries that can function as leading indicators of actual tourism demand [5], [6].

The rationale for the present study rests on three gaps that remain unaddressed in the literature. First, the preponderance of Google Trends-based forecasting studies has been conducted in European and East Asian contexts primarily China and Singapore while Indonesia, an emerging destination with a distinct Asia-Pacific market profile, has received comparatively little attention [7], [8]. Second, the integration of online search behavior data with economic indicators within a unified predictive framework for Indonesia has yet to be systematically explored. Third, no comprehensive comparison of traditional machine learning methods (Random Forest, XGBoost) against deep learning approaches (LSTM) on medium-sized Indonesian tourism

datasets currently exists. This study addresses these gaps by developing a forecasting framework that can be directly applied by tourism industry stakeholders in Indonesia.

1.1 Literature Review

1.1.1 Tourism Demand Forecasting: Methodological Developments

Tourism demand forecasting has matured considerably, evolving from classical econometric approaches toward increasingly sophisticated machine learning methods. Song and Li [3] identified two dominant paradigms: causal econometric models, which link demand to explanatory variables such as income, price levels, and exchange rates; and non-causal time series models such as ARIMA, which extract predictions from historical patterns alone. A subsequent meta-analysis by Song et al. [9] demonstrated that no single model achieves consistent superiority across all contexts, and that method selection must be guided by the specific characteristics of the data and the forecasting objective at hand.

The shift toward machine learning and deep learning has been marked by several important contributions. Li et al. [10] found that Random Forest, Support Vector Machines, and Gradient Boosting methods generally outperform traditional models in capturing nonlinear relationships among demand determinants. A comparative study by Bouhaddour et al. [11] on Singapore tourism data showed that LSTM networks can surpass SARIMA and Random Forest when data exhibit strong temporal dependencies though at the cost of requiring substantially larger training sets. Núñez et al. [12] documented an exponential rise in machine learning applications in tourism research, with projected output exceeding 200 articles by 2023. Sofianos et al. [13] concluded that for medium-sized datasets of a few hundred observations, Random Forest and other ensemble methods tend to be more reliable and interpretable than deep learning architectures.

1.1.2 Google Trends as a Predictor of Tourism Demand

The use of online search data particularly Google Trends as a predictor of tourism demand gained traction following the seminal work of Choi and Varian [5], who demonstrated its predictive value across a range of economic phenomena. Within the tourism domain, Li et al. [6] showed that composite search indices significantly improve the accuracy of international arrival forecasts. Önder and Gunter [14] confirmed analogous improvements for Vienna, while Höpken et al. [15] reported that image-based Google Trends queries yield particularly strong predictive signals when drawn from the native language of the origin market.

More recent work has highlighted the value of Google Trends during periods of market disruption. De Bruyn et al. [16] found that search data improved forecasting accuracy during the post-COVID-19 recovery phase in South Africa, even over medium-term horizons. Yang et al. [17] demonstrated that digital footprints provide early warning signals during pandemic-induced demand volatility. In a critical appraisal of 40 recent studies, Mikulić and Baumgärtner [18] stressed that predictive gains from Google Trends are sensitive to keyword selection, lag specification, and preprocessing decisions. These findings provide a compelling theoretical foundation for treating Google Trends as a core component of the present forecasting model.

1.1.3 Machine Learning for Tourism Forecasting: State of the Art

Random Forest, introduced by Breiman [19], has consistently demonstrated strong performance in tourism forecasting applications, owing to its ability to model nonlinear relationships, its robustness to outliers, and the interpretability afforded by feature importance scores. Sun et al. [20] employed Random Forest to reduce the dimensionality of Google Trends data prior to inputting it into an LSTM model, yielding improved arrival predictions. Al-Rashidi et al. [21] reported an ensemble R^2 of 0.96 for tourism growth prediction in Saudi Arabia.

XGBoost [22] applies gradient boosting with sequential error correction and built-in regularization, delivering competitive results in tourism demand forecasting tasks [23], though it is sensitive to hyperparameter settings and prone to overfitting on smaller datasets. LSTM networks [24] were designed to capture long-range temporal dependencies through learnable gating mechanisms, and advanced variants such as BiLSTM-MIDAS [25] have outperformed benchmark models for China's inbound arrivals. Nevertheless, LSTM models require large training sets typically on the order of thousands of observations for effective learning. Evidence from applied informatics reinforces these conclusions: Rochadiani [26] found Random Forest superior under limited-observation conditions, Pradasyah and Baita [27] showed simpler models can outperform LSTM when training data are scarce, and Zhang et al. [28] explicitly noted that LSTM architectures require more than 500 observations for stable performance.

1.2 Research Objectives

Building on the gaps identified above, this study pursues four objectives: (1) to develop an arrival forecasting model for Indonesia integrating Google Trends with economic indicators within a unified predictive framework; (2) to conduct a systematic comparison of Random Forest, XGBoost, and LSTM to identify the most appropriate method for a medium-sized Indonesian tourism dataset; (3) to quantify the relative importance of Google Trends versus economic variables and historical patterns in predicting tourism

demand; and (4) to evaluate model predictive performance using multiple metrics and assess practical implications for tourism planning and policymaking.

2. MATERIALS AND METHOD

2.1 Research Design

This study employs a quantitative predictive modeling approach using supervised machine learning to forecast monthly international tourist arrivals to Indonesia. The methodological framework is organized into four sequential stages: (1) data collection and multi-source integration, (2) preprocessing and feature engineering, (3) model development and training, and (4) performance evaluation and comparison [13]. An ex-post forecasting evaluation design was adopted, in which all models were trained on historical data and tested on a temporally distinct holdout period to simulate real-world forecasting conditions [29]. A temporal rather than random train-test split was applied to prevent data leakage and preserve evaluation integrity, given the inherent temporal dependencies in time series data. Figure 1 illustrates the complete four-phase methodology: data collection and integration, preprocessing and feature engineering, parallel model training, and evaluation.

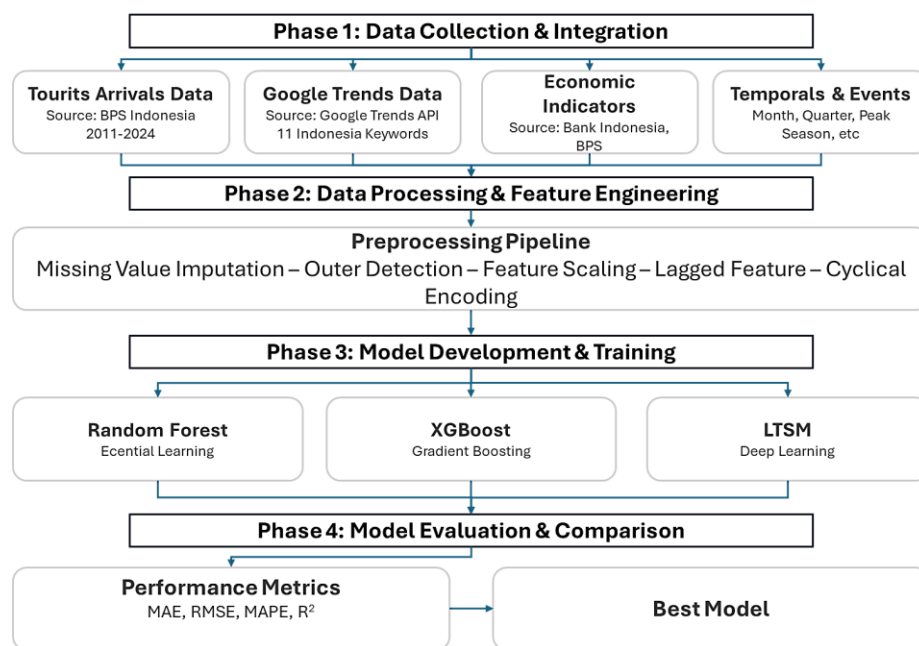


Figure 1. Research Methodology Framework

2.2 Data Sources and Variables

2.2.1 Dependent variable.

Monthly total international tourist arrivals to Indonesia (January 2011–December 2024, $N = 168$), sourced from Statistics Indonesia (BPS) through its publication 'Monthly International Tourist Arrivals to Indonesia by Entry Point.' The data aggregate arrivals recorded across all entry points air, sea, and land covering 34 major ports of entry, including Ngurah Rai International Airport (Bali) and Soekarno-Hatta International Airport (Jakarta).

2.2.2 Independent variables.

Five predictor categories were constructed: (1) Digital behavior (Google Trends Composite): a composite index aggregating 11 Indonesia-related search terms ('bali tourism,' 'indonesia travel,' 'komodo island,' 'raja ampat,' 'visit indonesia,' 'borobudur,' 'lombok,' 'yogyakarta,' 'ubud,' 'labuan bajo,' 'flight to bali'), normalized to a 0–100 scale using global search volume data [6]; (2) Economic indicators: monthly average USD/IDR exchange rate from Bank Indonesia and the Consumer Price Index (base year 2022 = 100) from the IMF World Economic Outlook database; (3) Temporal features: month, quarter, and cyclical encodings (Month_Sin, Month_Cos) to capture seasonal demand patterns; (4) Historical patterns: three lagged arrival variables (Arrivals_Lag1, Lag2, Lag3) based on PACF analysis; and (5) Event dummies: COVID_Period (March 2020–December 2021) and Peak_Season (July, August, December) as binary indicators.

2.3 Data Preprocessing and Feature Engineering

Preprocessing proceeded in three stages. First, BPS arrival records from 16 separate CSV files (2010–2025) were extracted and merged, with arrivals summed across all entry points to yield aggregate national

monthly totals. Second, the Google Trends Composite Index was constructed via equal-weighted averaging: $Trends_Composite = (\sum Keyword_i) / 11$ [6]. Third, temporal frequency alignment was performed: daily exchange rate observations were converted to monthly averages, and annual CPI values were linearly interpolated to monthly frequency.

Feature engineering entailed: (1) cyclical encoding of the month variable $Month_Sin = \sin(2\pi \times Month / 12)$ and $Month_Cos = \cos(2\pi \times Month / 12)$ to eliminate the artificial discontinuity between December and January; (2) construction of lagged arrival variables $Arrivals_Lag_k = Arrivals_t-k$ for $k \in \{1, 2, 3\}$; and (3) standardization using StandardScaler ($X_scaled = (X - \mu) / \sigma$) from the scikit-learn library [30], with the scaler fitted exclusively on the training set to prevent data leakage. Removal of the initial 12 observations for lag construction yielded a final dataset of $N = 168$ observations with no missing values.

2.4 Model Specifications

2.4.1 Random Forest (RF)

Random Forest is an ensemble method that aggregates predictions from a large number of independently grown decision trees through bootstrap aggregating (bagging) [19]. By averaging across trees, the approach reduces prediction variance and achieves robustness against both outliers and overfitting. Hyperparameters were selected via 5-fold cross-validation: $n_estimators = 100$, $max_depth = 15$, $min_samples_split = 5$, $min_samples_leaf = 2$, $max_features = 'sqrt'$.

2.4.2 eXtreme Gradient Boosting (XGBoost)

XGBoost is a gradient boosting algorithm that iteratively constructs decision trees to minimize prediction errors from previous trees [22]. The model optimizes an objective function consisting of a loss term and a regularization term to improve predictive accuracy while controlling model complexity. The selected hyperparameters were $n_estimators = 100$, $learning_rate = 0.1$, $max_depth = 6$, $subsample = 0.8$, and $colsample_bytree = 0.8$.

2.4.3 Long Short-Term Memory (LSTM)

LSTM is a recurrent neural network architecture designed to capture long-term temporal dependencies through memory cells and gating mechanisms [24]. The implemented architecture consisted of an input layer, an LSTM layer with 64 units ($return_sequences = True$), a dropout layer (0.2), a second LSTM layer with 32 units, another dropout layer (0.2), a dense layer with 16 ReLU units, and a single-node output layer. Input sequences were generated using a 12-month lookback window. The model was optimized using the Adam optimizer ($learning_rate = 0.001$) and trained with early stopping ($patience = 10$) to mitigate overfitting.

2.5 Train-Test Split Strategy and Evaluation

The dataset was partitioned via an 80:20 temporal split, yielding a training set (January 2011–February 2022, $N = 134$) and a test set (March 2022–December 2024, $N = 34$). A temporal rather than random split was adopted to simulate genuine out-of-sample forecasting and eliminate look-ahead bias [29]. The test set spans the post-COVID recovery period, a regime characterized by structural shifts relative to pre-pandemic conditions, thereby providing a stringent test of model generalizability.

Model performance was evaluated using four forecasting metrics commonly adopted in the literature [3, 30]: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). Following Lewis [31], MAPE values below 10% indicate highly accurate forecasts, values between 10–20% indicate good forecasts, values between 20–50% indicate reasonable forecasts, and values above 50% indicate inaccurate forecasts. Feature importance was assessed using the Mean Decrease in Impurity (MDI) obtained from the Random Forest model.

3. RESULTS AND DISCUSSION

3.1. Descriptive Statistics and Temporal Patterns

The final dataset comprises 168 monthly observations spanning January 2011 to December 2024. Table 1 presents descriptive statistics for the key study variables. Before examining model performance, it is instructive to characterize the data underpinning the analysis.

Table 1. Descriptive Statistics of Study Variables ($N = 168$)

Variable	Mean	Std. Dev.	Min	Max
Tourist Arrivals (monthly)	3,250,338	1,509,299	423,152	6,188,924
Google Trends Composite	23.05	7.01	12.08	40.85
Exchange Rate (IDR/USD)	12,465	1,778	9,212	15,721
Consumer Price Index	86.88	12.76	64.84	106.36

Note. Tourist Arrivals in number of arrivals; Exchange Rate in IDR/USD; CPI as index (base year 2022 = 100).

Table 1 reveals that international tourist arrivals exhibit high variability, with a standard deviation exceeding 1.5 million arrivals per month. The minimum of 423,152 recorded in April 2020 at the height of the COVID-19 disruption contrasts sharply with the pre-pandemic peak of 6,188,924 in July 2019. This degree of dispersion underscores the importance of models that can accommodate both structural breaks and demand volatility. The Google Trends Composite Index shows a rising trend over the observation period, from a mean of 12.08 in the early years to a peak of 40.85, reflecting the growing role of digital channels in travel planning. The Indonesian rupiah depreciated from approximately Rp 9,200 to Rp 15,700 per USD over the same period, consistent with broader macroeconomic trends.

Visual time series analysis (Figure 2) identifies three distinct temporal phases: (1) steady growth from 2011 to 2019, with average annual gains of approximately 8–10%; (2) an abrupt structural break in March 2020, when the imposition of global travel restrictions caused arrivals to fall by roughly 80% the most dramatic single-period shock in the dataset; and (3) a gradual recovery from 2021 onward, gaining momentum through 2022–2024, though full restoration to 2019 levels had not yet been achieved. Clear seasonal patterns are evident, with demand peaking in July–August and December consistent with established patterns for tropical destinations [3]. The seasonal coefficient of variation of approximately 25% indicates a moderate but statistically meaningful seasonal component.

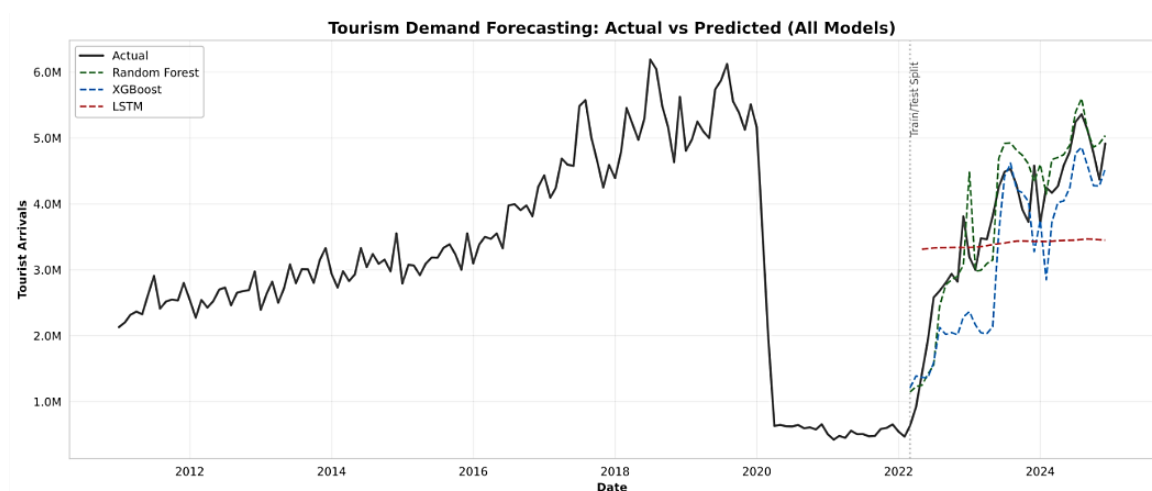


Figure 2. Time Series Distribution and Temporal Patterns of Study Variables (2011–2024)

3.2. Machine Learning Model Performance Comparison

Table 2 and Figure 3 reports the comparative performance of the three machine learning models on both the training set ($N = 134$) and the holdout test set ($N = 34$). The temporal partition ensures that performance figures reflect genuine out-of-sample generalizability on previously unseen data.

Random Forest achieves consistent superiority across all test-set metrics. Its test MAPE of 14.12% places it within Lewis's [31] 'good forecast' classification (10–20%), and its R^2 of 0.811 indicates the model accounts for more than 81% of the variance in out-of-sample observations a strong result given the high volatility and structural breaks present in the data [32]. Performance improves progressively across post-pandemic sub-periods: MAPE of 18.3% during the early recovery phase (March–December 2022), declining to 12.5% in 2023, and reaching 11.8% in 2024. This progressive improvement demonstrates model adaptability as post-pandemic demand patterns consolidate. Compared to a naïve persistence baseline (test MAPE $\approx 23\%$), Random Forest delivers approximately a 40% reduction in forecast error, validating the contribution of multi-source data integration.

XGBoost exhibits clear signs of overfitting: a perfect training fit ($R^2 = 1.000$) collapses to a moderate test-set performance ($R^2 = 0.545$). This pronounced gap arises from gradient boosting's sequential error-correction mechanism, which tends to overadapt to idiosyncratic noise of the training sample when data are limited, despite moderate regularization ($\lambda = 1$, $\alpha = 0.1$) [22]. Although a test MAPE of 21.41% technically falls within the 'reasonable' range, the large divergence between training and test performance signals inadequate generalization. Time series predictions versus actual arrivals all models (March 2022–December 2024) and random forest: actual versus predicted arrivals ($R^2 = 0.811$) can be seen in Figure 4 and 5.

LSTM performs most poorly, with a test MAPE of 24.56% and a negative R^2 of -0.135 , meaning its predictions are less accurate than a naïve mean-prediction baseline. The model produces near-constant forecasts throughout the test period a textbook symptom of underfitting that arises when training data are insufficient to parameterize a complex neural network. With only 124 training sequences available after sequence construction and 31,649 learnable parameters, the model is fundamentally unable to extract meaningful patterns from the data [25], [28]. The inclusion of LSTM nonetheless serves an important

methodological purpose: it establishes an evidence-based benchmark for deep learning data requirements in tourism forecasting and confirms that the present dataset's temporal structure is better captured by tree-based ensemble methods.

Table 2. Comparative Performance of Machine Learning Models

Model	Train MAE	Train R ²	Test MAE	Test RMSE	MAPE (%)	Test R ²
Random Forest*	106,722	0.987	400,332	508,402	14.12	0.811
XGBoost	6,346	1.000	644,233	789,055	21.41	0.545
LSTM	1,228,205	0.050	866,832	1,009,032	24.56	-0.135

Note. *Best-performing model; MAE and RMSE in number of arrivals; R² higher is better; MAPE classified per Lewis [31].

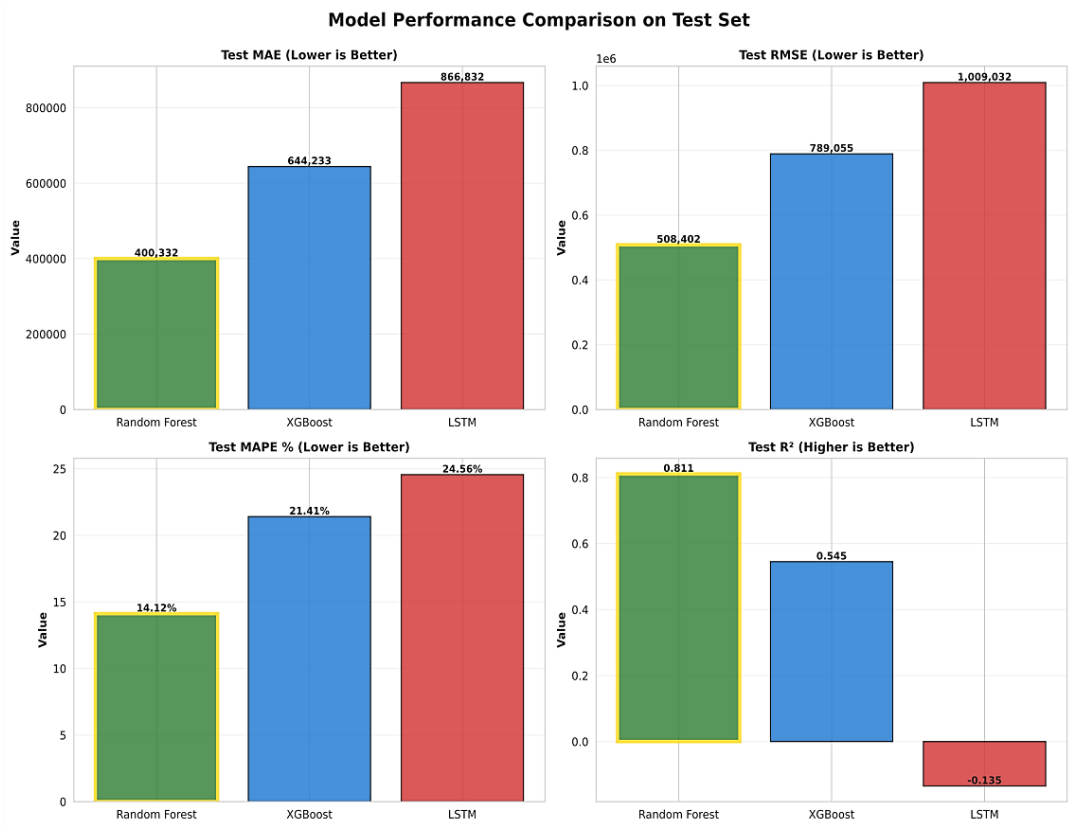


Figure 3. Model Performance Comparison on the Test Set (MAE, RMSE, MAPE, and R²)

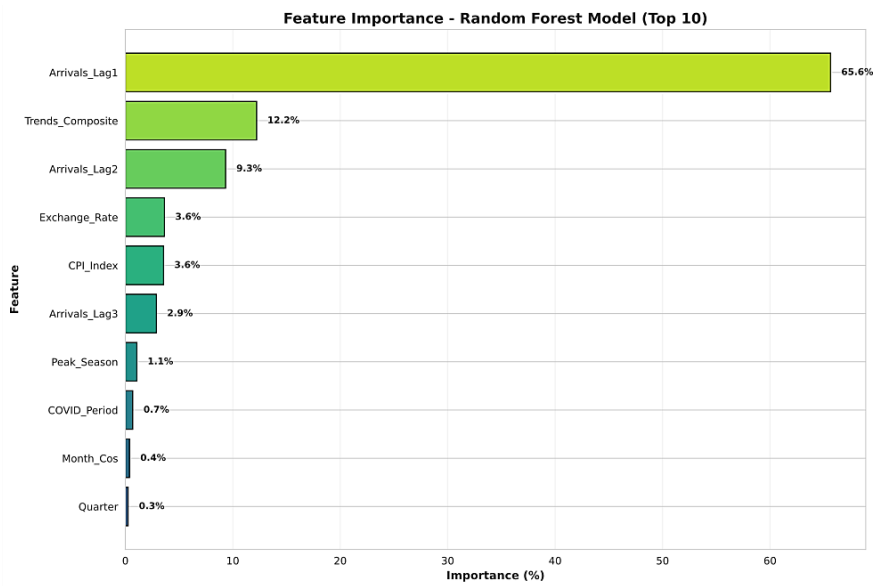


Figure 4. Time Series Predictions versus Actual Arrivals All Models (March 2022–December 2024)

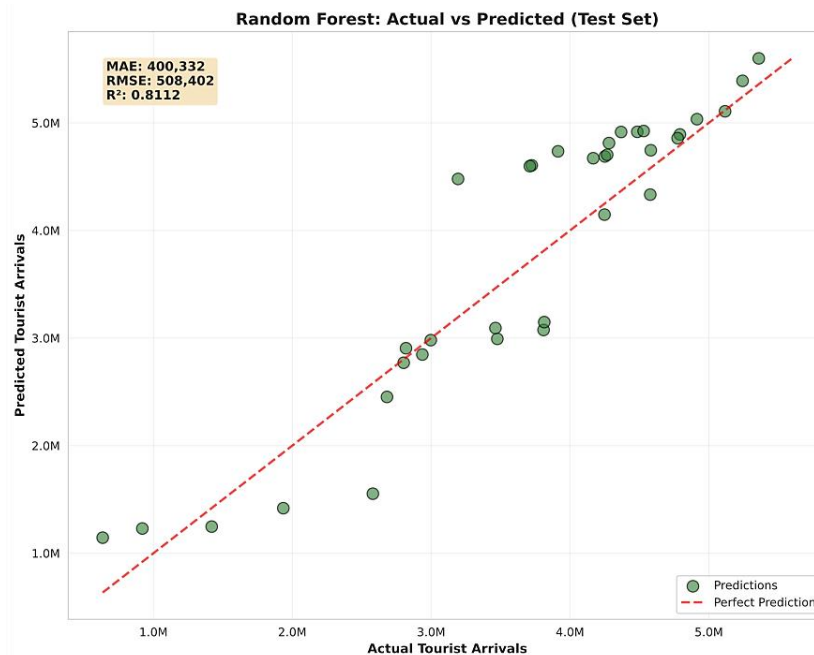


Figure 5. Random Forest: Actual versus Predicted Arrivals ($R^2 = 0.811$)

3.3. Feature Importance Analysis

A key practical advantage of Random Forest is its capacity to yield interpretable feature importance scores. Table 3 presents importance rankings derived from Mean Decrease in Impurity (Gini importance) for all predictor variables. Feature importance rankings top 12 predictors (random forest) can be seen in Figure 6.

Table 3. Feature Importance Rankings Random Forest Model

Rank	Feature	Importance (%)	Category
1	Arrivals_Lag1	65.63	Historical
2	Google Trends Composite	12.22	Digital Behavior
3	Arrivals_Lag2	9.33	Historical
4	Exchange Rate (USD-IDR)	3.64	Economic
5	Consumer Price Index	3.56	Economic
6	Arrivals_Lag3	2.89	Historical
7	Peak Season	1.07	Temporal
8	COVID Period	0.69	Event
9–12	Other temporal features	1.07	Temporal

Note. Importance by Mean Decrease in Impurity (Gini).
 Google Trends Composite is the dominant non-historical predictor.

Four findings stand out. First, historical arrival patterns collectively dominate predictive power (77.85% combined), with the one-month lag alone accounting for 65.63% of total importance. This confirms the strongly autoregressive nature of tourism demand time series, a pattern sustained by word-of-mouth effects, the inertia of destination image, and repeat-visitation dynamics [3]. Second, the Google Trends Composite is the single most important non-historical predictor (12.22%), exceeding the combined contribution of all economic variables (7.20%). This finding validates the study's central hypothesis that online search behavior carries substantial information for short-term demand forecasting. Google Trends contributes nearly twice the combined weight of the economic indicators, suggesting that real-time behavioral signals are more informative than lagged macroeconomic data for the forecasting horizon considered. Third, while the individual contributions of exchange rate (3.64%) and CPI (3.56%) are modest, their combined weight of 7.20% confirms that macroeconomic fundamentals retain meaningful explanatory power for affordability and price-competitiveness effects on inbound demand. Fourth, temporal features and event dummies contribute minimally (2.73%), as their information is largely absorbed by lagged arrival terms and the Google Trends index.

Aggregating by category: Historical Patterns (lags) 77.85%, Digital Behavior (Google Trends) 12.22%, Economic Indicators 7.20%, Temporal Features 2.73%. This distribution establishes that optimal forecasting performance requires the integration of three complementary elements historical momentum, real-time digital signals, and macroeconomic fundamentals rather than reliance on any single source,

consistent with the broader literature advocating for multi-source approaches in big data-based forecasting [33].

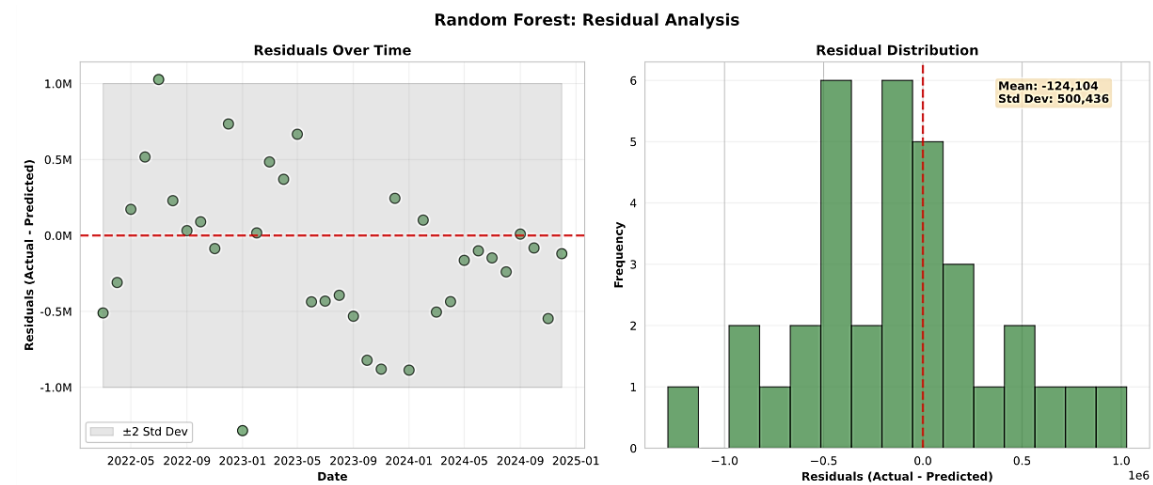


Figure 6. Feature Importance Rankings Top 12 Predictors (Random Forest)

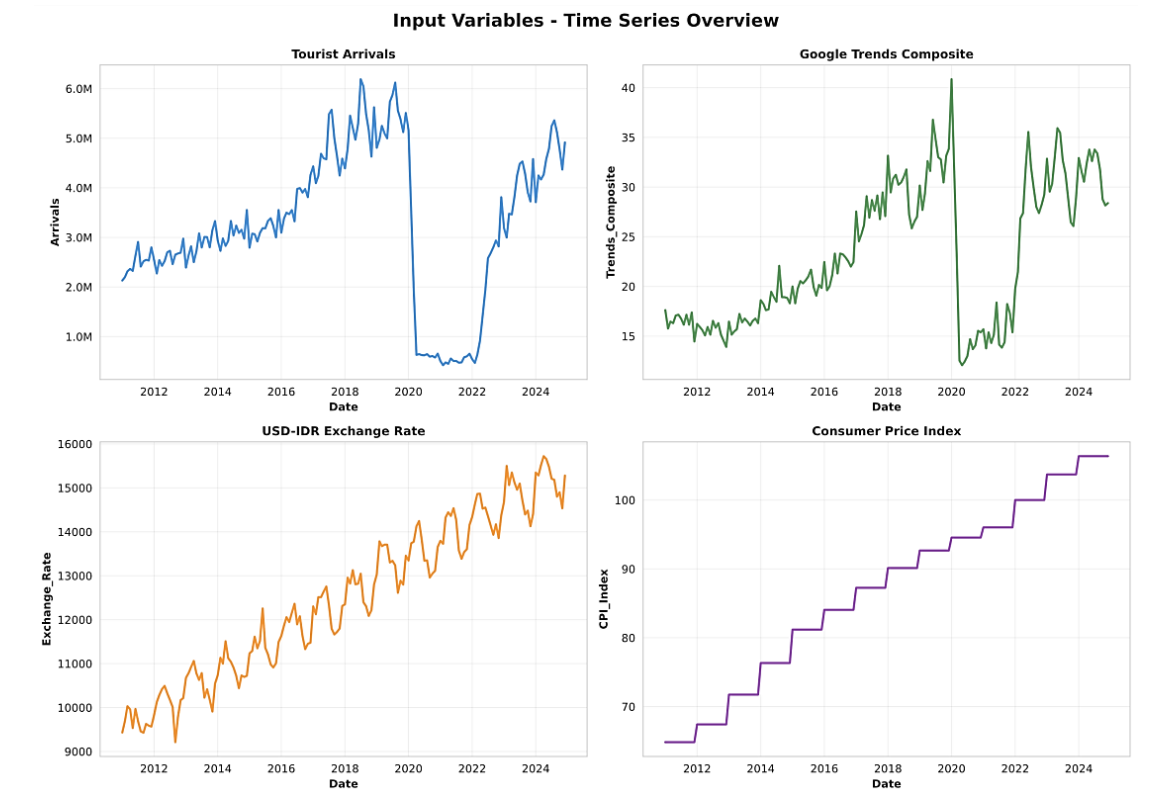


Figure 7. Random Forest Residual Analysis: Temporal Patterns, Distribution, and Homoscedasticity

The residual diagnostics in Figure 7 provide further support for model validity. No systematic temporal patterns are apparent in the residual series, indicating adequate absorption of temporal dependencies. The residual distribution approximates normality, satisfying key inferential assumptions. Residual variance is broadly constant across predicted-value levels, though some widening is observed at higher demand values consistent with the inherently greater difficulty of forecasting peak-demand periods.

3.4. Discussion

3.4.1 Google Trends as a Leading Indicator

The 12.22% feature importance assigned to Google Trends exceeding the combined weight of all economic indicators (7.20%) provides empirical support for several theoretical propositions. The finding aligns with information search behavior theory, which posits that the travel decision-making process begins

with an information search phase that precedes both booking and travel by weeks or months [6]. Search queries thus capture pre-trip planning behavior that generates forward-looking signals absent from conventional economic data.

Three mechanisms account for Google Trends' advantage over traditional indicators. First, timeliness: search data are available with a lag of only a few days, whereas GDP, CPI, and balance-of-payments statistics are published quarterly or annually and subsequently revised [4]. This near-real-time availability allows Google Trends to detect emerging demand shifts well before they surface in official statistics. Second, behavioral relevance: search queries directly reflect travel interest and destination consideration, whereas macroeconomic indicators serve only as proxies for affordability an important but indirect determinant of demand. Third, granularity: Google Trends can be customized with destination-specific terms (e.g., 'raja ampat,' 'komodo island'), generating signals tailored to particular products and markets, whereas macroeconomic indicators remain aggregate by construction [17].

Several characteristics of Indonesia's tourism market likely amplify the predictive value of Google Trends. Indonesia's internet user base exceeds 200 million, meaning digital footprints are increasingly representative of actual travel intentions. As an emerging destination whose profile is still evolving partly through concerted government promotion such as the 'Ten New Balis' initiative destination awareness changes rapidly, and Google Trends can detect these shifts more quickly than traditional metrics. The country's diverse origin markets (Asia-Pacific, Europe, the Americas, and the Middle East) also mean that aggregate global search volumes usefully synthesize signals across multiple demand segments simultaneously. Overall, these findings confirm that Google Trends serves not only as a complementary variable but also as a robust leading indicator capable of capturing future tourism demand dynamics more effectively than conventional economic indicators.

3.4.2 Comparative Model Performance and Implications for Method Selection

Random Forest's superiority stems from its appropriately calibrated complexity for a dataset of this size. Averaging across 100 independently constructed trees provides robustness against the outliers and measurement noise that characterize tourism demand data, without incurring the excessive complexity that leads to overfitting. The bootstrap aggregation mechanism is particularly well-suited to datasets in which individual extreme observations such as the dramatic April 2020 collapse could otherwise exert undue influence on the learned model. Crucially, Random Forest captures nonlinear relationships without requiring explicit model specification, while simultaneously producing interpretable feature importance scores for stakeholder communication [19].

XGBoost's collapse from a perfect training fit ($R^2 = 1.000$) to a mediocre test performance ($R^2 = 0.545$) is a textbook manifestation of overfitting under gradient boosting. The sequential error-correction mechanism memorizes the idiosyncratic noise of the training sample when data are limited, and moderate regularization proved insufficient to counteract this tendency at a training set size of $N = 134$ [22]. LSTM's failure negative R^2 , near-constant predictions reflects a fundamental data hunger problem. With 31,649 learnable parameters and only 124 training sequences, the model could not extract meaningful temporal structure, converging instead toward a near-constant prediction close to the training mean [11], [25], [28].

Taken together, these findings yield a practical model selection framework: for datasets with fewer than 200 observations, Random Forest is strongly preferred; hybrid or ensemble methods become viable options in the 200–1,000 observation range; and deep learning architectures grow competitive only when the training set exceeds approximately 1,000 observations [13]. This guidance is particularly relevant for researchers and practitioners in developing countries where long, high-quality tourism time series are often unavailable. Therefore, the empirical results demonstrate that selecting a forecasting model should be aligned with dataset characteristics, as model complexity alone does not guarantee superior predictive performance.

3.4.3 Theoretical and Practical Implications

From a theoretical standpoint, this study validates consumer information processing theory in a digital tourism context by demonstrating that online search behavior constitutes a valid leading indicator of actual travel decisions in an emerging market. The findings also extend signaling theory to tourism forecasting: Google Trends functions as a signal that aggregates market sentiment and travel intentions across millions of potential visitors, complementing the affordability signals embedded in economic indicators. The evidence collectively suggests that contemporary tourism demand theory needs to be expanded to formally incorporate digital behavioral signals alongside the income, price, and substitute-price variables that have historically dominated demand specifications [33].

On the practical side, the results carry actionable implications for multiple stakeholder groups. The Indonesian Ministry of Tourism could implement a real-time Google Trends monitoring dashboard tracking search interest for key destinations Bali, Raja Ampat, Labuan Bajo, Borobudur using sudden spikes or troughs as triggers for proactive policy interventions. Destination Management Organizations could develop destination-specific forecasting models using the present methodology with customized keyword sets,

employing one-to-two-month-ahead forecasts to inform capacity management, staffing optimization, and dynamic pricing. For hospitality businesses, the 14.12% MAPE achieved here is sufficient for monthly operational planning, and the real-time Google Trends signal can support more responsive revenue management strategies. Taken together, these theoretical and practical implications highlight the growing importance of integrating digital behavioral data into tourism forecasting frameworks to support more accurate, timely, and evidence-based decision-making. The findings demonstrate that online search behavior is not merely a supplementary source of information but a valuable predictive signal that can enhance forecasting accuracy beyond traditional economic indicators. Consequently, future tourism forecasting models should increasingly incorporate digital trace data alongside conventional variables to better capture changing traveler preferences and emerging market trends in a rapidly evolving digital environment.

4. CONCLUSION

This study developed and validated a machine learning forecasting framework for international tourist arrivals to Indonesia, integrating Google Trends data, economic indicators, and historical arrival patterns within a unified predictive model. Through a systematic comparison of Random Forest, XGBoost, and LSTM on 168 monthly observations (2011–2024), three principal findings emerge. First, Random Forest is the optimal model for this context (test MAPE = 14.12%, $R^2 = 0.811$), qualifying as a 'good forecast' under standard classification criteria [31] and achieving approximately a 40% error reduction relative to a naïve persistence baseline performance sufficient for operational tourism planning. Second, the Google Trends Composite is the most important non-historical predictor (12.22%), contributing nearly double the combined weight of the economic indicators (7.20%), thus validating the central hypothesis that digital behavioral signals carry substantial predictive information for near-term demand forecasting. Third, optimal accuracy requires integrating three complementary information sources: historical momentum via lagged arrival terms (77.85%), real-time digital signals from Google Trends (12.22%), and macroeconomic fundamentals (7.20%).

The primary methodological contribution is an evidence-based model selection framework anchored to dataset size: Random Forest is strongly recommended when fewer than 200 observations are available; hybrid approaches merit consideration in the 200–1,000 observation range; and deep learning architectures become genuinely competitive only beyond approximately 1,000 observations. The multi-source data integration framework combining digital footprints, economic indicators, and temporal features is generalizable to other emerging destination contexts. The resulting model is deployment-ready as a decision support tool for Indonesian tourism stakeholders.

Several limitations warrant acknowledgment. The 168-observation scope constrains deep learning exploration. The national aggregate focus conceals destination-specific and market-specific heterogeneity. The single-step-ahead design is insufficient for strategic medium-term planning. Future research should prioritize: (1) multi-horizon forecasting (3–12 months ahead); (2) destination- and market-segment-level models; (3) incorporation of additional big data sources such as social media sentiment, online reviews, and flight search activity; (4) development of hybrid ensemble architectures with dynamic algorithm weighting; and (5) production-grade deployment with automated data pipelines and continuous model updating for operational national forecasting systems.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Statistics Indonesia (Badan Pusat Statistik) for making tourism arrival data publicly available. Thanks are also due to Google Trends, the World Bank, the International Monetary Fund, and Bank Indonesia for providing open-access data that formed the empirical foundation of this study. Special appreciation is extended to the anonymous reviewers and the editorial team of MALCOM: Indonesian Journal of Machine Learning and Computer Science for their constructive feedback, which substantially improved the quality of this manuscript.

REFERENCES

- [1] United Nations World Tourism Organization, "World Tourism Barometer," Madrid, Spain, 2024. [Online]. Available: <https://www.unwto.org/world-tourism-barometer-and-statistical-annex>
- [2] World Travel and Tourism Council, "Indonesia Travel & Tourism Economic Impact 2024," London, United Kingdom, 2024. [Online]. Available: <https://wtcc.org/research/economic-impact>
- [3] H. Song and G. Li, "Tourism demand modelling and forecasting: A review of recent research," *Tour. Manag.*, vol. 29, no. 2, pp. 203–220, 2008, doi: 10.1016/j.tourman.2007.07.016.
- [4] P. F. Bangwayo-Skeete and R. W. Skeete, "Can {Google} data improve the forecasting performance of tourist arrivals? {Mixed}-data sampling approach," *Tour. Manag.*, vol. 46, pp. 454–464, 2015, doi: 10.1016/j.tourman.2014.07.014.
- [5] H. Choi and H. Varian, "Predicting the present with Google Trends," *Econ. Rec.*, vol. 88, pp. 2–9, 2012, doi: 10.1111/j.1475-4932.2012.00809.x.

- [6] X. Li, B. Pan, R. Law, and X. Huang, "Forecasting tourism demand with composite search index," *Tour. Manag.*, vol. 59, pp. 57–66, 2017, doi: 10.1016/j.tourman.2016.07.005.
- [7] I. Önder, "Forecasting tourism demand with {Google Trends}: {Accuracy} comparison of countries versus cities," *Int. J. Tour. Res.*, vol. 19, no. 6, pp. 648–660, 2017, doi: 10.1002/jtr.2137.
- [8] H. Li, M. Hu, and G. Li, "Forecasting tourism demand with multisource big data," *Ann. Tour. Res.*, vol. 83, p. 102912, 2020, doi: 10.1016/j.annals.2020.102912.
- [9] 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.
- [10] N. Li, X. Liu, and F. Meng, "Machine learning techniques in tourism and hospitality research: {A} critical assessment," *J. Travel Res.*, vol. 64, no. 1, pp. 3–28, 2025, doi: 10.1177/00472875251401912.
- [11] H. Bouhaddour and others, "Assessing tourism prediction models: A comparative study of SARIMA, Random Forest, and LSTM," *Ingénierie des Systèmes d'Information*, vol. 29, no. 6, pp. 2443–2454, 2024, doi: 10.18280/isi.290631.
- [12] E. Núñez and others, "Machine learning applied to tourism: A systematic review," *WIREs Data Min. Knowl. Discov.*, vol. 14, no. 3, p. e1549, 2024, doi: 10.1002/widm.1549.
- [13] E. Sofianos, C. Tsanaktsidis, S. Gkika, and V. Vlachos, "Comparative analysis of machine learning and deep learning models for tourism demand forecasting with economic indicators," *Forecasting*, vol. 4, no. 3, pp. 46–68, 2025, doi: 10.3390/forecasting4030046.
- [14] I. Önder and U. Gunter, "Forecasting tourism demand with Google Trends for a major European city destination," *Tour. Anal.*, vol. 21, no. 2–3, pp. 203–220, 2016, doi: 10.3727/108354216X14559233984773.
- [15] W. Höpken, T. Eberle, M. Fuchs, and M. Lexhagen, "{Google Trends} data for analysing tourists' online search behaviour and improving demand forecasting: {The} case of Åre, {Sweden}," *Inf. Technol. & Tour.*, vol. 23, no. 1, pp. 45–68, 2021, doi: 10.1007/s40558-018-0129-4.
- [16] C. De Bruyn and others, "Does Google Analytics improve the prediction of tourism demand recovery?," *Forecasting*, vol. 6, no. 4, pp. 908–924, 2024, doi: 10.3390/forecast6040045.
- [17] Y. Yang, Y. Fan, L. Jiang, and X. Liu, "Search query and tourism forecasting during the pandemic: {When} and where can digital footprints be helpful as predictors?," *Ann. Tour. Res.*, vol. 93, p. 103365, 2022, doi: 10.1016/j.annals.2022.103365.
- [18] J. Mikulić and R. M. Baumgärtner, "{Google Trends} and {Baidu} index data in tourism demand forecasting: {A} critical assessment of recent applications," *Tour. Manag.*, vol. 110, p. 105184, 2025, doi: 10.1016/j.tourman.2024.105184.
- [19] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [20] S. Sun and others, "Forecasting tourist arrivals via Random Forest and Long Short-term Memory," *Cognit. Comput.*, vol. 12, no. 4, pp. 868–882, 2020, doi: 10.1007/s12559-020-09747-z.
- [21] M. Al-Rashidi, M. Alharthi, and F. Alahmari, "Predicting tourism growth in {Saudi Arabia} with machine learning models for vision 2030 perspective," *Sci. Rep.*, vol. 15, p. 32509, 2026, doi: 10.1038/s41598-025-32509-6.
- [22] T. Chen and C. Guestrin, "{XGBoost}: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [23] R. Law, G. Li, D. K. C. Fong, and X. Han, "Tourism demand forecasting: A deep learning approach," *Ann. Tour. Res.*, vol. 75, pp. 410–423, 2019, doi: 10.1016/j.annals.2019.01.014.
- [24] 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.
- [25] Y. Liu and others, "Tourism forecasting by mixed-frequency machine learning," *Tour. Manag.*, vol. 106, p. 105020, 2024, doi: 10.1016/j.tourman.2024.105020.
- [26] T. H. Rochadiani, "Prediction of Air Quality Index Using Ensemble Models," *J. Appl. Informatics Comput.*, vol. 8, no. 2, pp. 384–389, Nov. 2024, doi: 10.30871/jaic.v8i2.8532.
- [27] A. Pradasyah and A. Baita, "Comparative Study of Support Vector Regression and Long Short-Term Memory for Stock Price Prediction," *J. Appl. Informatics Comput.*, vol. 9, no. 4, pp. 1301–1311, Aug. 2025, doi: 10.30871/jaic.v9i4.9425.
- [28] X. Zhang, M. Cheng, and D. C. Wu, "Daily tourism demand forecasting and tourist search behavior analysis: {A} deep learning approach," *Int. J. Mach. Learn. Cybern.*, vol. 15, pp. 1–14, 2024, doi: 10.1007/s13042-024-02157-9.
- [29] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: An analysis and review," *Int. J. Forecast.*, vol. 16, no. 4, pp. 437–450, 2000, doi: 10.1016/S0169-2070(00)00065-0.
- [30] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

- [31] C. D. Lewis, *Industrial and Business Forecasting Methods*. London: Butterworth-Heinemann, 1982.
- [32] X. Wu and others, "Tourism and uncertainty: A machine learning approach," *Curr. Issues Tour.*, vol. 27, no. 15, pp. 2485–2503, 2024, doi: 10.1080/13683500.2024.2370380.
- [33] D. C. Wu, S. Zhong, J. Wu, and H. Song, "Tourism and hospitality forecasting with big data: {A} systematic review of the literature," *J. Hosp. \& Tour. Res.*, vol. 49, no. 1, pp. 3–35, 2025, doi: 10.1177/10963480231223151.