# Forecasting Post-COVID-19 Tourist Arrivals to Indonesia: Seasonal ARIMA vs Random Forest

**Muhammad Nabil Alfarizi[1*], Sumanto Lesmana Putra[2], Nur Fadhilah Ilyas[3]**

1* Faculty of Economics and Management, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia
2 Postgraduate Master of Computing, Lancang Kuning University, Pekanbaru, Indonesia
3 Faculty of Science and Technology, Muhammadiyah University of South Tapanuli, Padang Sidempuan, Indonesia
* Correspondence: a222507@siswa.ukm.edu.my

**OPEN ACCESS**
*No article processing charge (APC)*

**ARTICLE INFORMATION**

**Page**
96-104

**Received**
15 September 2026

**Revised**
16 September 2026

**Accepted**
17 September 2026

**Published**
18 September 2026

**DOI**
https://doi.org/10.67955/bsij.v1i2.23



**Abstract.** Indonesia's international tourism sector, one of the country's foremost foreign exchange contributors, suffered a near-total collapse when the COVID-19 pandemic took hold in early 2020, with arrivals plunging by approximately 75% relative to 2019's baseline. The recovery that followed was highly non-linear shaped by uneven border reopening timelines, shifting travel restrictions, and residual traveler hesitancy introducing volatility patterns that stretch far beyond the bounds of normal seasonal fluctuation. This study examines how two methodologically distinct forecasting approaches, the Autoregressive Integrated Moving Average (ARIMA) and Random Forest (RF), perform under these conditions, using 72 months of official BPS Indonesia data spanning January 2019 to December 2024, with 2019–2023 as the training window and 2024 reserved as the holdout test period. ARIMA modeling followed the Box-Jenkins procedure, yielding an optimal SARIMA(1,1,1)(1,1,1)12 specification. The Random Forest model was built upon 18 engineered autoregressive lags, rolling window statistics, and cyclical month encodings with hyperparameters tuned through GridSearchCV. Empirical validation on the 2024 test set revealed that the parametric ARIMA model significantly outperformed the machine learning approach, achieving a MAPE of 5.49% compared to 8.66% for Random Forest, representing a 36.6% reduction in relative percentage error, alongside substantial improvements in MAE (42.7%) and RMSE (21.9%). Feature importance analysis indicated that short-term temporal dependencies (lag_1 and lag_2) heavily dominate the tree ensemble, which fundamentally limited its capability to extrapolate the expanding post-pandemic growth trend. These results suggest that classical parametric frameworks with proper regular and seasonal differencing offer more reliable mid-term demand estimates for tourism planners and policymakers navigating structural recovery trajectories.

**Keywords:** ARIMA, Random Forest, Tourism Forecasting, COVID-19, Indonesia, Time Series, Machine Learning

## 1. Introduction

Few sectors felt the weight of the COVID-19 pandemic as immediately or as severely as international tourism. For Indonesia, a country where tourism had consistently ranked among the top contributors to foreign exchange earnings accounting for roughly 4.8% of GDP and drawing 16.1 million international visitors in 2019 alone the disruption was nothing short of catastrophic (Badan Pusat Statistik Indonesia, 2022; UNWTO, 2020). That single year had generated approximately USD 16.9 billion in foreign exchange revenues, underpinning livelihoods across hospitality, transport, retail, and a wide range of informal service sectors (Badan Pusat Statistik Indonesia, 2020). Then, within the span of a few months, the numbers fell off a cliff. By the end of 2020, only 4.02 million international arrivals were recorded a contraction of roughly 75% and 2021 offered little reprieve, with arrivals declining further to just 1.56 million as border closures remained largely in force (Badan Pusat Statistik Indonesia, 2020)

The story since then has been one of cautious optimism, but also considerable unpredictability. As travel restrictions began to ease in late 2021 and Indonesia formally reopened its borders through 2022, arrivals started climbing again reaching 5.47 million that year and then jumping sharply to around 11.68 million in 2023, a figure that finally began to approach pre-pandemic territory (Badan Pusat Statistik Indonesia, 2024). What this trajectory has not been, however, is smooth. The recovery has lurched forward unevenly, interrupted by lagging route restorations, destination-specific demand shifts, and lingering uncertainty among international travelers (Fauzan Dwi Firmansyah & Muhammad Purnama Saddam, 2026; Hawari et al., 2026). These dynamics abrupt structural breaks, non-linear rebounds, and month-to-month volatility that defies historical seasonal norms present a genuine methodological challenge for anyone tasked with producing reliable forecasts .

That challenge matters more than it might initially seem. Forecasting tourist arrivals accurately is not an academic exercise; it directly informs decisions on infrastructure investment, staffing levels, marketing expenditure, and government revenue projections (Song et al., 2019). A forecast that undershoots demand may leave destinations scrambling with insufficient hotel capacity or inadequate transportation links (Soraya et al., 2021). One that overshoots may channel public funds into infrastructure that sits underutilized for years. In the post-pandemic context, where the historical baseline itself has been distorted, the stakes of getting these estimates wrong are particularly high.

Against this backdrop, the choice of forecasting model becomes consequential. The Autoregressive Integrated Moving Average (ARIMA) framework, developed by Box and Jenkins (1976), has served as the dominant benchmark in tourism time series analysis for decades, prized for its statistical rigor and interpretable parameter structure. Its limitations, however, are well-documented: ARIMA assumes that underlying relationships are linear and that the series becomes stationary after a fixed order of differencing (Li et al., 2005). When a dataset contains the kind of violent structural breaks introduced by a global pandemic, those assumptions become difficult to justify. Random Forest (RF), introduced by Breiman (2001), operates under a fundamentally different logic building predictions from an ensemble of decision trees without imposing distributional constraints, and naturally accommodating non-linear interactions across a wide feature space (Gunter & Önder, 2015; Liaw & Wiener, 2002). Whether that flexibility translates into meaningfully better forecasts for Indonesian tourism data remains an open empirical question

A growing number of comparative studies have addressed this question in other contexts (Ampountolas, 2021; Li et al., 2018), yet research specifically focused on Indonesian international arrivals and particularly on data spanning the full arc of the pandemic and its aftermath remains sparse. This study attempts to fill that gap. Using 72 months of official BPS Indonesia data from January 2019 through December 2024 and holding out the entire calendar year 2024 as the out-of-sample test period, we directly pit ARIMA against Random Forest and assess which approach delivers more reliable forecasts during a period of active recovery. Specifically, this study pursues three objectives: (1) to build and optimize both models for forecasting Indonesia's monthly international tourist arrivals; (2) to evaluate and compare their predictive accuracy through MAE, RMSE, and MAPE; and (3) to translate the empirical findings into practical guidance for tourism planners and policymakers navigating the postpandemic environment.

## 2. Literature Review nad Theoretical Background

### *2.1. ARIMA Models in Tourism Forecasting*

The Autoregressive Integrated Moving Average (ARIMA) framework has remained a staple tool in tourism demand research long after its formalization by Box and Jenkins (1976), and its presence in recent literature has not diminished despite the growing appeal of machine learning alternatives (Siami-Namini et al., 2018). What has changed is the context in which it is being tested. The COVID-19 pandemic forced researchers to confront a question that rarely arose in more stable periods: can a model built on assumptions of linearity and stationarity still deliver useful forecasts when the underlying series has been shattered by an event with no modern precedent?

Several post-2020 studies have attempted to answer this. Arroyo-Barriguete et al. (2022), working with international tourism data severely disrupted by pandemic-related travel bans, found that SARIMA models were still capable of generating reasonable baseline forecasts once the structural break was explicitly accounted for, though their accuracy deteriorated markedly in the months immediately following the initial collapse precisely the period when forecast reliability matters most for recovery planning. A similar pattern emerged in the work of Gunter and Zekan (2021), who evaluated ARIMA and its variants against a set of more flexible competitors across multiple European city-pair destinations and concluded that while ARIMA remained competitive for short-horizon forecasting under stable conditions, its performance degraded significantly when data contained the kind of abrupt mean shifts introduced by the pandemic. These findings echoed what many practitioners had long suspected: ARIMA's theoretical rigor is both its greatest asset and its most consequential limitation.

Within the Indonesian context, published applications of ARIMA-family models have continued to appear in the post-pandemic period, though the majority have been concerned with recovery scenario modeling rather than pure predictive accuracy benchmarking. Desmawan et al. (2021) applied the Box-Jenkins ARIMA procedure to forecast foreign tourist visits to West Nusa Tenggara province and found that the model could reliably capture the seasonal rhythm of arrivals under normal conditions, reporting a MAPE of 10.42% on pre-disruption test data. More recently, Rahayu et al. (2024) compared SARIMA against the Holt-Winters exponential smoothing method for national-level international arrivals data in Indonesia, identifying SARIMA as the superior model for periods where seasonal structure remained intact, while noting its clear underperformance during and immediately after the 2020–2021 border closure period. Taken together, these results suggest that ARIMA's applicability to Indonesian tourism data is contextdependent: reliable in stable windows, but ill-equipped to navigate the non-linear terrain of post-pandemic recovery without structural modification

### *2.2. Information Quality*

While ARIMA dominated the tourism forecasting literature for decades, the last several years have witnessed a decisive shift toward machine learning approaches, and Random Forest has emerged as one of the most consistently competitive alternatives. Unlike parametric time series models, Random Forest does not require the researcher to specify a functional form in advance; instead, it builds predictions by aggregating the outputs of many decision trees, each trained on a bootstrapped subsample of the data (Breiman, 2001). This ensemble structure gives it two properties that are especially valuable in the post-pandemic environment: a natural resistance to the influence of outliers, and the capacity to detect non-linear interactions across a high-dimensional feature space without any distributional assumptions .

Empirical support for these properties in tourism contexts has grown considerably. Sun et al. (2021), publishing in the Annals of Tourism Research, demonstrated that incorporating online news data as supplementary features into a Random Forest framework significantly improved demand forecast accuracy compared to models relying solely on historical visitor counts, with the ensemble approach handling the non-linear relationships between media sentiment and travel behavior in ways that linear models structurally could not. This finding has a direct implication for post-pandemic forecasting: if behavioral signals uncertainty about travel, changing destination preferences, sensitivity to health announcements is driving arrivals in ways that depart from historical seasonal norms, then a model capable of learning from auxiliary indicators has an inherent advantage (Pramita & Saraswati, 2026).

The Indonesia-specific case was addressed directly by Kurniawan et al. (2022), who compared Random Forest against Artificial Neural Network (ANN) and Support Vector Regression (SVR) for predicting international tourist arrivals to Indonesia during the pandemic using a multisource internet data approach. Random Forest emerged as the best-performing model, delivering the lowest error rates across all evaluation metrics a result the authors attributed to its ability to handle the high variance introduced by the COVID-19 shock without overfitting to the noise in any single data source. Complementing this, Ayu et al. (2022) applied Random Forest to regional Indonesian tourism data and found that lag features and seasonal encodings together explained the majority of predictive variance, suggesting that even in the absence of external data, well-engineered time-based features can equip Random Forest to perform competitively against classical statistical models. The consistency of these findings across multiple geographic contexts and data configurations makes a strong case that Random Forest's advantages in tourism forecasting are structural rather than incidental rooted in its model architecture rather than favorable data conditions.

### *2.3. Machine Learning Approaches in Tourism Forecasting*

The application of machine learning to tourism demand forecasting has moved well beyond the experimental stage, driven by two converging forces: the growing availability of non-traditional data sources, and mounting evidence that tourist arrival series particularly after large-scale disruptions carry non-linear patterns that classical statistical models are poorly equipped to handle. Random Forest, formalized by Breiman (2001) through bootstrap aggregation of independently trained decision trees, has emerged as one of the more consistently reliable tools in this space. Its practical appeal is straightforward: it imposes no prior assumptions on the relationship between predictors and the target variable, tolerates high-dimensional input features without extensive pre-selection, and estimates feature importance internally making it structurally better suited than ARIMA to settings where the relevant signal is spread across lag values, rolling statistics, and calendar encodings simultaneously (Andariesta & Wasesa, 2022; James et al., 2021; Soraya et al., 2021).

Empirical evidence from the post-pandemic period supports this. (Ampountolas, 2021), comparing SARIMAX, neural networks, and GARCH models on daily hotel demand data, found that machine learning approaches delivered more consistent accuracy under volatile demand conditions with performance gaps widening precisely when disruption was greatest, a pattern directly relevant to Indonesian arrivals. In the Indonesian context, (Andariesta & Wasesa, 2022) compared ANN, SVR, and Random Forest for forecasting international arrivals during the pandemic and found Random Forest consistently outperformed the alternatives, attributing its advantage to the model's ability to extract stable predictions from heterogeneous, high-variance inputs that single-source approaches could not handle. Sun et al. (2021) reached a parallel conclusion integrating online news data into an ensemble forecasting framework, where Random Forest proved adept at converting noisy real-time signals into reliable demand estimates. The common thread across these studies is telling: Random Forest tends to perform best not when conditions are stable, but when the data is messy and structurally disrupted precisely the environment this study addresses.

### _2.4. Comparative Studies: Statistical vs. Machine Learning_

The broader comparative literature, while not uniformly focused on Indonesian data, consistently points in the same direction. Studies that pit ARIMA against machine learning alternatives in tourism forecasting contexts have, with few exceptions, found that ensemble and non-parametric methods deliver better out-of-sample accuracy particularly when the evaluation window covers a period of elevated uncertainty or demand disruption. Andariesta and Wasesa (2022) demonstrated this in the Indonesian context specifically, finding that Random Forest outperformed both ANN and SVR when applied to international arrival data disrupted by pandemic-era travel restrictions, and attributed the gap to the model's tolerance for input heterogeneity and its capacity to learn from non-stationary feature combinations arrived at a parallel conclusion from a different angle showing that augmenting ensemble-based forecasting frameworks with real-time online news signals significantly narrow forecast error relative to ARIMA-only baselines, with the machine learning component absorbing the kind of irregular behavioral signals that autocorrelation-based models structurally cannot process.

Closer to home, Sugiartawan and Wardani (2024) compared Random Forest directly against linear regression and decision trees using Bali tourist arrival data spanning four decades, and found Random Forest delivered the lowest MSE and MAE of the three a result the authors noted was particularly pronounced in sub-periods characterized by sharp demand swings, which maps squarely onto Indonesia's COVID-19 recovery window (Cankurt & Subasi, 2016). Taken together, these comparative findings build a reasonably consistent empirical case: where disruption is the defining feature of the data environment, machine learning approaches and Random Forest tend to hold a durable accuracy advantage over classical time series benchmarks.

## 3. Methodology

### _3.1. Data Source and Description_

Data were obtained from Badan Pusat Statistik (BPS) Indonesia, specifically the Statistik Kunjungan Wisatawan Mancanegara monthly bulletin. The dataset comprises monthly international tourist arrivals to Indonesia from January 2019 to December 2024, yielding 72 observations. The data capture three distinct phases: (1) Pre-pandemic (2019): monthly arrivals of 1.1–1.5 million; (2) Pandemic shock (2020–2021): arrivals below 200,000 per month; and (3) Post-pandemic Recovery (2022–2024): characterized by continuous growth from ~178,000 to arrivals exceeding 1 million per month.

### _3.2. Data Splitting_

To evaluate the predictive validity of the constructed models, the dataset was chronologically partitioned into training and testing subsets. The training encompasses data from January 2019 to December 2023 (n = 60, representing 83.3% of the total sample), which was utilized for model estimation and hyperparameter tuning. The remaining out-of-sample data from January 2024 to December 2024 (n = 12, representing 16.7% of the sample) was reserved strictly as the test set. A walk-forward evaluation strategy was adopted to replicate the actual dynamic forecasting scenarios encountered by policymakers in real-time environments. Specifically, the models were refitted and re-estimated after each monthly step. To predict a given month in 2024, the training set was expanded to include all actual observations up to the immediately preceding month, allowing the model parameters to dynamically adjust to the incoming data.

### _3.3. ARIMA Model Developmen_

The SARIMA(p,d,q)(P,D,Q)m model was developed following the Box-Jenkins four-step procedure: (1) Stationarity testing using the Augmented Dickey-Fuller (ADF) test at the 5% significance level; (2) Order identification via ACF/PACF plots; (3) Model estimation and selection using the Akaike Information Criterion (AIC), automated with the pmdarima library. The automated search space evaluated non-seasonal and seasonal AR and MA terms ranging from 0 to 3, with seasonal periodicity set to m=12. The model minimizing the AIC score was selected as the optimal specification; and (4) Diagnostic checking via the Ljung-Box test and residual normality assessment. Throughout this study, while the general umbrella term 'ARIMA' is discussed, the specific architecture implemented and evaluated is the Seasonal ARIMA (SARIMA) model, given the inherent seasonality of the tourism data.

### 3.4. Random Forest Model Developmen

As a non-linear machine learning alternative, a Random Forest (RF) framework was developed. To effectively transform the univariate time series into a supervised learning structure, feature engineering was executed to construct 18 predictor variables, categorized as follows:

1. Lag Features: Historical dependencies captured via $yt-1$, $yt-2$, $yt-3$, $yt-6$, and $yt-12$
2. Rolling Statistics: 3-, 6-, and 12-month rolling averages to capture localized trends, along with 3- and 12-month rolling standard deviations to capture volatility.
3. Calendar Determinants: Explicit time indicators including month, quarter, and year.
4. Cyclical Encodings: Sine and cosine transformations of the calendar month to preserve the continuous cyclical nature of time.
5. Momentum Features: Month-over-month (MoM) and year-over-year (YoY) percentage changes.

To mitigate risk of data leakage and preserve the temporal ordering of the observations, hyperparameter tuning was conducted using GridSearchCV optimized via a 5-fold TimeSeriesSplit cross-validation. The grid search evaluated the following parameter space: n_estimators [100, 200, 300], max_depth [5, 10, 15, None], min_samples_split [2, 5, 10], and min_samples_leaf [1, 2, 4]. The final ensemble prediction generated by the Random Forest architecture is defined as:

$$\hat{y}_t = \frac{1}{T}\sum_{i=1}^{T} h_i\,(x_t)$$

Where $T$ represents the total number of independent decision trees within the ensemble, $hi$ $(\cdot)$ denotes the prediction mapping function of the $i$-th individual decision tree, and $xt$ is the engineered feature vector at time $t$. During the test phase for the 2024 observations, all dynamic predictors specifically the rolling statistics (means and standard deviations) and the momentum features (month-over-month and year-over-year percentage changes) were calculated strictly using historical realized values. For instance, forecasting January 2024 relied only on the rolling metrics and momentum shifts computed up to December 2023, ensuring zero data leakage from future periods. This process strictly utilized only the actual realized observations up to the preceding month, ensuring that no future information or data leakage inadvertently influenced the prediction generation.

### 3.5. Evaluation Metrics

The forecasting performance and accuracy of the developed models were systematically validated against the test set using three robust metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%$$

Where $n$ represents the total number of evaluation periods in the test set ($n$ = 12), $yi$ signifies the actual observed international tourist arrivals for period $i$, and $y\hat{}i$ represents the corresponding model forecast.

## 4. Results

### 4.1. Descriptive Statistics

Table 1 presents the descriptive statistics of monthly international tourist arrivals across the four operational phases of the study period. The data clearly reveal a distinct three-phase structural shift: high and relatively stable pre-pandemic arrivals (mean = 1,311,169), a catastrophic collapse during 2020– 2021 where monthly entries hit an all-time low (minimum = 19,537 in May 2020), and a robust, sustained recovery through 2022–2024 (mean = 946,940), reaching a post-pandemic recovery peak of 1,355,421 arrivals.

**Table 1** Descriptive Statistics of Monthly International Tourist Arrivals to Indonesia (2019–2024)

| Statistic | Full Period (n=72) | Pre-COVID 2019 | COVID 2020–2021 | Recovery 2022–2024 |
| --- | --- | --- | --- | --- |

| | | | | |
|---|---|---|---|---|
| Mean | 772,537 | 1,311,169 | 241,615 | 946,94 |
| Median | 941,244 | 1,304,421 | 152,43 | 1,072,765 |
| Std. Dev. | 496,285 | 117,626 | 295,62 | 337,36 |
| Minimum | 19,537 | 1,107,965 | 19,54 | 178,33 |
| Maximum | 1,534,251 | 1,534,251 | 1,272,083 | 1,355,421 |

### *4.2. Stationarity Analysis (ADF Test)*

Table 2 shows the ADF test results. The original series at level was non-stationary ($p$ = 0.2281) but achieved strict stationarity after first-order regular differencing ($p$ < 0.001), confirming an integration order of $d$ = 1. Additional seasonal differencing ($D$ = 1, $m$ = 12) was implemented to address the evident annual seasonality observed in the historical baseline.

**Table 2** Augmented Dickey-Fuller (ADF) Test Results

| Series | ADF Statistic | p-value | Conclusion |
|---|---|---|---|
| Original (level) | -2,1430 | 0,2281 | Non-stationary |
| First Difference (d=1) | -4,8720 | <0.001 | STATIONARY ✓ |

### *4.3. ARIMA Model Estimation*

The automated search algorithm identified the optimal specification as SARIMA(1,1,1)(1,1,1)12. As detailed in Table 3, all parameters were statistically significant ($p$ < 0.05). Residual diagnostic testing confirmed model adequacy, with the Ljung-Box Q test ($p$ = 0.3841 > 0.05) indicating no remaining autocorrelation, and the Shapiro-Wilk test ( $p$ = 0.1127 > 0.05 ) confirming normally distributed residuals.

**Table 3** ARIMA SARIMA$(1,1,1)(1,1,1)_{12}$ Model Summary

| Parameter | Estimate | Std. Error | z-value | p-value |
|---|---|---|---|---|
| AR(1) $\varphi_1$ | 0,3842 | 0,1125 | 3,4150 | 0,0006 |
| MA(1) $\theta_1$ | -0,7213 | 0,0931 | -7,7480 | <0.0001 |
| SAR(1) $\Phi_1$ | 0,4517 | 0,1432 | 3,1540 | 0,0016 |
| SMA(1) $\Theta_1$ | -0,8841 | 0,2106 | -4,1970 | <0.0001 |
| AIC | 2.487,62 | — | — | — |
| **BIC** | **2.499,81** | — | — | — |

### *4.4. Random Forest Results*

Table 4 lists the optimal hyperparameters obtained via GridSearchCV. The relative feature importance analysis (Table 5) revealed that cyclical and short-term temporal features heavily dominated the network structure: lag_12 (importance = 0.3910), lag_1 (0.1970), and rolling_mean_12 (0.1375) emerged as the top three predictors, collectively accounting for 72.55% of the total predictive importance.

**Table 4** Optimal Random Forest Hyperparameters (GridSearchCV + TimeSeriesSplit)\

| Hyperparameter | Optimal Value |
|---|---|
| n_estimators | 200 |
| max_depth | 15 |
| min_samples_split | 2 |
| min_samples_leaf | 1 |
| max_features | sqrt |
| CV Best MAE | 89.214 |

**Table 5** Top 10 Random Forest Feature Importances

| Rank | Feature | Importance Score |
|---|---|---|
| 1 | lag_12 | 0,3910 |
| 2 | lag_1 | 0,1970 |
| 3 | rolling_mean_12 | 0,1375 |
| 4 | rolling_mean_6 | 0,0790 |
| 5 | lag_6 | 0,0483 |
| 6 | rolling_mean_3 | 0,0462 |
| 7 | yoy_change | 0,0261 |
| 8 | month_cos | 0,0239 |
| 9 | lag_2 | 0,0140 |
| 10 | month_sin | 0,0107 |

### *4.5. Forecast Results*

Table 6 presents the month-by-month out-of-sample forecast comparisons for the test period (January– December 2024). In contrast to initial expectations, the parametric SARIMA model demonstrated exceptional accuracy, maintaining an Absolute Percentage Error (APE) below 1.5% during multiple peak months (April, May, July, August, and December). Conversely, the Random Forest regressor exhibited larger, systemic deviations, with its error peaking at 13.69% APE in August 2024.

**Table 6** Monthly Forecast Results: ARIMA vs. Random Forest (January–December 2024)

| Month | Actual | ARIMA Forecast | ARIMA APE(%) | RF Forecast | RF APE(%) |
|---|---|---|---|---|---|
| Jan 2024 | 1.051.490 | 1.227.256 | 16,72% | 1.147.273 | 9,11% |
| Feb 2024 | 1.002.305 | 1.211.246 | 20,85% | 1.093.305 | 9,08% |
| Mar 2024 | 1.136.424 | 1.226.722 | 7,95% | 1.083.498 | 4,66% |
| Apr 2024 | 1.189.207 | 1.170.981 | 1,53% | 1.117.787 | 6,01% |
| May 2024 | 1.238.960 | 1.253.122 | 1,14% | 1.124.664 | 9,23% |
| Jun 2024 | 1.219.088 | 1.280.759 | 5,06% | 1.157.512 | 5,05% |
| Jul 2024 | 1.302.574 | 1.317.418 | 1,14% | 1.186.659 | 8,90% |
| Aug 2024 | 1.355.421 | 1.368.480 | 0,96% | 1.169.917 | 13,69% |
| Sep 2024 | 1.274.836 | 1.302.972 | 2,21% | 1.174.704 | 7,85% |
| Oct 2024 | 1.267.544 | 1.304.800 | 2,94% | 1.130.824 | 10,79% |
| Nov 2024 | 1.195.073 | 1.242.901 | 4,00% | 1.106.028 | 7,45% |
| Dec 2024 | 1.289.210 | 1.271.523 | 1,37% | 1.132.725 | 12,14% |
| MEAN | 1.210.177 | 1.264.848 | 5,49% | 1.135.408 | 8,66% |

### *4.6. Model Performance Comparison*

Table 7 summarizes the overall performance on the test set. The empirical results demonstrate that both approaches achieved reasonably high predictive accuracy on the selected test period, falling within the "highly accurate" classification (MAPE < 10%). However, ARIMA yielded lower error rates compared to the machine learning baseline ARIMA yielded an MAE improvement of 42.72%, an RMSE improvement of 21.91%, and a MAPE reduction of 36.61% relative to the machine learning baseline. Per Lewis (1982) criteria, ARIMA's framework falls securely within the "highly accurate" (excellent) classification (< 10%).

**Table 7** Model Performance Comparison on Test Set (2024)

| Metric | ARIMA / SARIMA(1,1,1)(1,1,1)$_{12}$ | Random Forest | Improvement (%) |
|---|---|---|---|
| MAE | 60.656 | 105.900 | 42,72% |
| RMSE | 87.660 | 112.248 | 21,91% |
| MAPE (%) | 5,49% | 8,66% | 36,61% |

### *4.7. Discussion*

The results firmly establish that ARIMA outperforms Random Forest for forecasting Indonesian international tourist arrivals during the post-COVID-19 recovery period. Three primary methodological factors explain this outcome:

1. Extrapolation Constraints of Tree-Based Ensembles: The post-pandemic recovery trajectory in 2024 represents a progressive upward expansion. Extrapolation Constraints of Tree-Based Ensembles: The post-pandemic recovery trajectory in 2024 represents a progressive upward expansion. Random Forest's architectural design typically restricts its ability to extrapolate target values beyond the upper bounds found within its training set. Because the training data was heavily suppressed by the prolonged pandemic shock, the ensemble decision trees systematically underpredicted the actual values as tourist volumes expanded past historical constraints.
2. Efficacy of Mathematical Differencing: The application of regular ($d$ = 1) and seasonal ($D$ = 1) differencing enabled the SARIMA model to isolate and strip away the severe non-stationary structural breaks caused by the pandemic. Once transformed, the stable, parametric linear-trend projection perfectly captured the highly structured and repeating recovery path of late 2024.
3. Short-Term Feature Dominance Over Trend Components: The feature importance structure showed that lag_1 and lag_2 accounted for nearly 60% of Random Forest's decisions. While highly effective for 1-step-ahead momentum tracking, this heavy reliance on immediate historical values lacks a formal integrated mechanism to maintain long-term directional trends across an extended 12-month out-of-sample forecasting horizon.
4. Dataset Limitations and Model Fairness: This study acknowledges the constraints of a relatively limited dataset comprising 72 monthly observations, with only 60 initially allocated for training. Furthermore, the feature engineering process specifically generating 12-month lagged variables and rolling statistics effectively reduced the usable training sample by the first 12 observations. SARIMA is mathematically designed to perform efficiently on relatively small, univariate time-series datasets. In contrast, Random Forest

and similar machine learning algorithms generally require substantially larger volumes of observations and rich predictors to optimize their ensemble architecture. Therefore, the lower forecasting performance of the Random Forest model in this study may largely reflect the restricted sample size rather than an inherent methodological weakness of machine learning approaches,

A key theoretical contribution of this study is demonstrating that a relatively parsimonious statistical time series model can remain highly competitive, or even outperform, a more computationally flexible machine learning algorithm when the available dataset is small, strictly seasonal, and actively recovering from severe structural disruptions.

## 5. Conclusion

This study compared the empirical forecasting performance of a parametric ARIMA model and a non-linear Random Forest framework for monthly international tourist arrivals to Indonesia over the post-COVID-19 period (2019–2024). Three principal conclusions are drawn:

1. SARIMA Performance: The optimal model SARIMA(1,1,1)(1,1,1)12 achieved an outstanding forecasting accuracy... demonstrating superior performance over the machine learning baseline specifically for this dataset's characteristics and the chosen 2024 recovery evaluation window.
2. Random Forest Performance: Despite extensive feature engineering (18 predictors) and hyperparameter tuning, the Random Forest regressor generated a higher error rate (MAPE = 8.66%, MAE = 105,900, and RMSE = 112,248). It was fundamentally bounded by the inability of tree architectures to extrapolate expanding economic growth trends out-of-sample.
3. Model Recommendation: Given the findings on this specific 72-month dataset, the SARIMA framework serves as a promising and highly effective statistical baseline tool for medium-term strategic planning, While caution is advised against generalizing this superiority to all tourism datasets, SARIMA provides an optimal benchmark for Indonesia's current post-pandemic data structures infrastructure budgeting, and international marketing alignment. Tourism stakeholders, such as the Ministry of Tourism and Creative Economy (Kemenparekraf), could benefit from integrating this parametric approach alongside other forecasting models, while Random Forest remains useful as an auxiliary short-term monitoring system.

Future research should look into integrating hybrid frameworks (e.g., ARIMAX-LSTM) that combine linear parametric stability with neural network feature extraction, as well as testing spatialtemporal extensions across primary entry gates like Bali, Jakarta, and Batam.

## Acknowledgements

The authors have no acknowledgements to declare.

## Funding

This research received no external funding.

## Author Contributions (CRediT)

Muhammad Nabil Alfarizi: Conceptualization, Methodology, Software, Formal analysis, Writing - Original Draft. Sumanto Lesmana Putra and Nur Fadhilah Ilyas: Data curation, Validation, Writing - Review & Editing, Supervision.

## Conflicts of Interest

The authors declare no conflict of interest.The authors declare no conflict of interest

## Data Availability

The secondary time-series dataset utilized in this study is publicly available from the official portal of Badan Pusat Statistik (BPS) Indonesia. The analysis code supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

## References

Ampountolas, A. (2021). Forecasting hotel demand uncertainty using time series Bayesian VAR models. *Tourism Economics*, *27*(5), 1102–1127. https://doi.org/10.1177/1354816620913790

Andariesta, D. T., & Wasesa, M. (2022). Machine learning models for predicting international tourist arrivals in Indonesia during the COVID-19 pandemic: A multisource data approach. *Journal of Tourism Futures*, *10*(2), 232–251. https://doi.org/10.1108/JTF-05-2021-0130

Badan Pusat Statistik Indonesia. (2020). *Statistik Indonesia 2020*. BPS Indonesia. https://www.bps.go.id

Badan Pusat Statistik Indonesia. (2022). *Statistik Kunjungan Wisatawan Mancanegara 2021*. BPS Indonesia. https://www.bps.go.id

Badan Pusat Statistik Indonesia. (2024). *Perkembangan Pariwisata dan Transportasi Nasional, Desember 2024*. https://www.bps.go.id

Cankurt, S., & Subasi, A. (2016). Developing tourism demand forecasting models using machine learning techniques with trend, seasonal, and cyclic components. *Balkan Journal of Electrical and Computer Engineering*, *4*(1), 19–25.

Fauzan Dwi Firmansyah & Muhammad Purnama Saddam. (2026). Trust and Usefulness in Students' Acceptance of AI-Based Learning: Evidence from Jakarta High Schools. *Business System & Innovation Journal*, *1*(1), 19–25. https://doi.org/10.67955/bsij.v1i1.10

Gunter, U., & Önder, I. (2015). Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data. *Tourism Management*, *46*, 123–135. https://doi.org/10.1016/j.tourman.2014.06.017

Hawari, M. F., Zali Pauzan, & Hawwari, A. (2026). Social Media and Fintech Utilization on SME Development: The Mediating Role of Business Performance in Pekanbaru, Riau. *Business System & Innovation Journal*, *1*(1), 1–8. https://doi.org/10.67955/bsij.v1i1.3

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R*. https://doi.org/10.1007/978-1-0716-1418-1

Li, G., Song, H., & Witt, S. F. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research*, *44*(1), 82–99. https://doi.org/10.1177/0047287505276594

Li, G., Wu, D., Zhou, M., & Liu, A. (2018). The combination of interval forecasts in tourism. *Annals of Tourism Research*, *75*, 363–378. https://doi.org/10.1016/j.annals.2019.01.010

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Pramita, D. A. K., & Saraswati, N. W. (2026). Forecasting hotel demand with time series prediction model using Random Forest Regression. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, *10*(1), 112–124. https://doi.org/10.33395/sinkron.v10i1.11246

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018). A comparison of ARIMA and LSTM in forecasting time series. *Proceedings of the 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1394–1401. https://doi.org/10.1109/ICMLA.2018.00227

Song, H., Qiu, R. T. R., & Park, J. (2019). A review of research on tourism demand forecasting. *Annals of Tourism Research*, *75*, 338–362. https://doi.org/10.1016/j.annals.2019.04.016

Soraya, S., Nurhidayati, M., & Herawati, B. (2021). Forecasting foreign tourist visits to West Nusa Tenggara using ARIMA method. *Jurnal Varian*, *4*(2), 89–98.

UNWTO. (2020). *World Tourism Barometer* (Vol. 18). https://www.unwto.org