International Journal of Applied Mathematics, Sciences, and Technology for National Defense
Volume 4, Issue 2, 169-182
e_ISSN: 2985-9352, p_ISSN: 2986-0776
https://www.journal.foundae.com/index.php/JAS-ND
DOI : https://doi.org/10.58524/app.sci.def.v4i2.1496

Kernel selection in support vector regression for forecasting international tourist arrivals: Implications for economic resilience and national defense

Kris Suryowati*
Universitas AKPRIND Indonesia
INDONESIA

Maria Oktafiana Dedu
Universitas AKPRIND Indonesia
INDONESIA

Rokhana Dwi Bekti
Universitas AKPRIND Indonesia
INDONESIA

Junaidi
Universitas Tadulako
INDONESIA

Article Info

Article history:

Received: May 26, 2026
Revised: July 16, 2026
Accepted: August 27, 2026

Keywords:

BI Rate
Forecasting
Kernel
SVR
Tourist

Abstract

Background: International tourist arrivals are an important component of Indonesia's tourism economy and are influenced by temporal patterns and macroeconomic conditions. Accurate forecasting is therefore important to support tourism planning and economic resilience.
Aims: This study aims to develop and evaluate a Support Vector Regression (SVR) model with different kernel functions for forecasting international tourist arrivals in Indonesia and to compare its performance with the ARIMAX model.
Method: Monthly tourist arrival data from January 2017 to November 2024 were analyzed using SVR with linear, polynomial, radial basis function, and sigmoid kernels. Data were standardized using StandardScaler, while Grid Search was used to optimize the model parameters. Inflation and the BI Rate were incorporated as explanatory variables. The best-performing SVR model was compared with ARIMAX(1,1,1) using MSE, RMSE, MAD, and MAPE.
Result: The SVR with a second-degree Polynomial Kernel achieved the best performance, with an MSE of 3,401,140,779.52, RMSE of 58,319.30, MAD of 44,565.19, and MAPE of 12.94%. The model outperformed ARIMAX(1,1,1), which obtained a MAPE of 16.23%. The results indicate that the polynomial kernel can better capture nonlinear patterns in tourist arrival dynamics.
Conclusion: The SVR Polynomial model provides a useful approach for forecasting international tourist arrivals in Indonesia. Its forecasting capability can support tourism planning, resource allocation, and early identification of changes in tourism demand, contributing to more adaptive and resilient tourism management.

To cite this article: Suyowati, K., Dedu, M. O., Bekti, R. D., & Junaidi. (2026). Kernel selection in support vector regression for forecasting international tourist arrivals: Implications for economic resilience and national defense. *International Journal of Applied Mathematics, Sciences, and Technology for National Defense,* 4(2), 169-182.

***Corresponding Author:**
Kris Suyowati, Universitas AKPRIND Indonesia, Indonesia, Email: suryowati@akprind.ac.id

INTRODUCTION

Indonesia occupies a strategic maritime position in Southeast Asia, situated at the intersection of two continents (Asia and Australia) and two oceans (Indian and Pacific). This geographical advantage has established the country as a hub for economic and cultural exchange and provided a strong foundation for tourism development. The tourism sector is not only a driver of economic growth but also a contributor to national resilience, as foreign exchange earnings and regional development reinforce economic stability, an essential pillar of defense preparedness. Between 2017 and 2019, the number of international tourist arrivals showed an overall upward yet fluctuating trend, peaking in July 2018 with 1,547,232 visits and remaining above 1.5 million per month until the end of 2019. However, in early 2020, the sector experienced a sharp decline due to the COVID-19 pandemic and international travel restrictions, with tourist arrivals dropping to below 200,000 per month by 2021. Signs of recovery began to appear in early 2022, with a significant upward trend throughout 2023, although fluctuations persisted. Based on Ministry of Tourism and Creative Economy in 2024, tourist arrivals increased by 16.9% compared to the previous month but decreased by 18.94% compared to December 2023, totaling 927,746 visits—still far below the July 2018 peak of 1.5 million. This situation underscores the need for effective recovery strategies, including regional collaboration through the ASEAN Tourism Forum (ATF), to strengthen ASEAN's role as a major destination region (Feng et al., 2022; Mishra et al., 2021; Škare et al., 2021).

Macroeconomic factors such as the Bank Indonesia interest rate (BI Rate) and inflation also influence tourism visits. The BI Rate affects credit costs and investments in related sectors, while inflation, measured by the Consumer Price Index (CPI), reflects price stability and affects tourists' purchasing power (Chen & Wang, 2007; Nur Aini, 2022). In this context, forecasting tourist arrivals is crucial for supporting economic policy planning and national tourism marketing strategies. Accurate forecasts help governments and industry stakeholders to prepare infrastructure capacity, promotional strategies, and efficient resource allocation. Moreover, tourist arrival predictions serve as an important indicator for assessing potential foreign exchange revenue and the economic stability of tourism-dependent regions (Nguyen et al., 2013; Suryawan et al., 2025).

Recent studies have shown that applying statistical and machine learning models such as ARIMAX, LSTM, and Support Vector Regression (SVR) can yield high accuracy in forecasting tourist arrivals and tourism-related economic trends across different countries (Akanbi & Bello, 2024; Cankurt & Subaşi, 2016; Guerouate et al., 2023; Sun et al., 2019; Surtiningsih, 2017; Tovmasyan, 2021). Therefore, utilizing adaptive predictive models such as SVR is highly relevant for understanding patterns and uncertainties in Indonesia's post-pandemic tourism data.

SVR is an extension of the Support Vector Machine (SVM) algorithm originally developed for classification tasks (Lalian et al., 2022). The fundamental concept of SVR lies in finding a regression function that minimizes prediction error while maintaining the maximum margin between actual observations and the estimated function. This approach follows the principle of Structural Risk Minimization (SRM), which seeks to balance model complexity and generalization capability for unseen data (Smola & Schölkopf, 2004). In tourism forecasting, SVR has been shown to outperform classical models such as ARIMAX, especially when handling seasonal and highly volatile data (Jouilil & Mentagui, 2023).

This study employs the SVR method, which utilizes various kernel functions to capture data patterns, allowing the model to adapt to the unique characteristics of each dataset. The kernels considered in SVR include the Linear Kernel for linearly separable data, the Polynomial Kernel for nonlinear relationships, and the Radial Basis Function (RBF) Kernel for more complex patterns. Comparing these kernels aims to identify the one most suitable for the data characteristics, as each possesses distinct capabilities in capturing variable relationships. Proper kernel selection is crucial to enhancing model accuracy and efficiency, particularly for datasets with varying levels of complexity.

An important factor influencing SVR performance is hyperparameter tuning, which determines optimal parameters for machine learning algorithms. However, manual hyperparameter optimization can be time-consuming, especially when numerous parameters are involved. One of the most widely used approaches is the grid search method combined with k-fold cross-validation. The main advantage of grid search over other optimization algorithms lies in its systematic iteration structure, which minimizes duplication when parameters are unique (Iriananda et al., 2024). Prediction errors are evaluated using several metrics, including Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD), and Mean Absolute Percentage Error (MAPE).

This study analyzes international tourist arrivals in Indonesia from January 2017 to November 2024 and applies SVR with three kernel approaches, Linear, Polynomial, and RBF, to predict future visits (Zhou et al., 2024). Hyperparameter optimization is conducted using grid search with cross-validation, and model performance is evaluated using Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD), and Mean Absolute Percentage Error (MAPE). By comparing kernel performance, the study seeks to identify the most suitable model for Indonesia's tourism data.

Accurate forecasting of tourist arrivals is therefore essential for guiding economic policy, infrastructure planning, and tourism marketing strategies. Forecasts also serve as indicators of potential foreign exchange revenue and economic stability in tourism-dependent regions (Nguyen et al., 2013; Suryawan et al., 2025). Recent studies emphasize the effectiveness of machine learning models, particularly Support Vector Regression (SVR), in forecasting tourism trends. SVR, an extension of Support Vector Machine (SVM), applies the principle of Structural Risk Minimization to balance model complexity and generalization (Lalian et al., 2022; Vapnik, 2000). Compared to classical models such as ARIMAX, SVR has demonstrated superior performance in handling seasonal and volatile data (Amelia et al., 2021; Chen et al., 2022; Jouilil & Mentagui, 2023). Kernel selection (Linear, Polynomial, RBF) and hyperparameter optimization using grid search with cross-validation are critical for enhancing model accuracy, with performance evaluated through metrics such as MSE, RMSE, MAD, and MAPE (Agan Celik et al., 2026; Iriananda et al., 2024). These approaches make SVR highly relevant for analyzing Indonesia's post-pandemic tourism data and supporting recovery strategies. In conclusion, forecasting international tourist arrivals is essential not only for tourism recovery and economic planning but also for reinforcing Indonesia's resilience and defense capacity (Agan Celik et al., 2026). Reliable predictions strengthen economic stability, secure foreign exchange revenues, and enhance Indonesia's strategic position in the ASEAN and Indo-Pacific context, thereby linking tourism forecasting with broader national defense preparedness.

METHOD

Support Vector Regression (SVR)

Support Vector Regression (SVR) is a regression technique designed to overcome the overfitting problem and requires proper parameter selection to achieve optimal performance. Overfitting occurs when a model exhibits very high accuracy on the training data but is fails to generalize to unseen data (Adam et al., 2024; Maulana et al., 2019; Surtiningsih, 2017). The primary goal of SVR is to determine a function $f(x)$ that minimizes the prediction error while maintaining model smoothness, ignoring errors smaller than a predefined threshold ε (epsilon-insensitive loss).

The core concept of SVR involves computing a linear function based on two parameters and which are non-negative Lagrange multipliers. The solution to this optimization problem is generally obtained through quadratic programming, resulting in an optimal approximation surface. When extended to nonlinear spaces, the regression surface is produced through kernel functions, expressed as

$$f(x) = \sum_{j=1}^{n} (a_j^* - a_j)\left(K(x_i, x_j) + \lambda^2\right) \quad (1)$$

where $f(x)$ is predicted output, value $(a_j^* - a_j)$ is Lagrange multipliers, $K(x_i, x_j)$ is kernel function, and λ^2 is scalar (Maulana et al., 2019).

The SVR algorithm can employ different types of kernel functions with its own characteristics and advantages (Adam et al., 2024),

a. Linear Kernel, this is applied when the data are linearly separable in the feature space,

$$K(x_i, x_j) = x_i, x_j^T \tag{2}$$

where i and j is the data $(i, j\text{: } 1, 2, \dots\ n)$, x is the input variable.

b. Polynomial Kernel, this is used when the data cannot be separated linearly. The equation is given by:

$$K(x_i, x_j) = (x_i, x_j^T + 1)^d \tag{3}$$

where d denotes the polynomial degree.

c. Radial Basis Function (RBF) Kernel, one of the most used kernels in SVR, the RBF kernel effectively captures nonlinear relationships and complex patterns,

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \tag{4}$$

where γ Gamma parameter.

Grid Search and k −Fold Cross-Validation

Grid Search is a systematic optimization method used to determine the optimal hyperparameters of the Support Vector Regression (SVR). It defines a search space for each hyperparameter within specified bounds and discretizes this space into candidate values. All possible combinations are evaluated, ensuring that the global optimum within the defined range can be identified. To improve robustness, Grid Search is combined with k −fold cross-validation, which assesses generalization capability by partitioning the dataset into k subsets. The model is trained on $k-1$ folds and tested on the remaining fold, repeated k times. Performance is average across folds. The choice of k influences model performance: smaller k reduces computation time but may yield unstable estimates, while larger k improves reliability at higher computational cost. In this study, $k = 3$ was selected to balance efficiency and accuracy.

The hyperparameter space considered includes cost parameter $C \in \{0.01, 0.1, 1, 10, 100\}$, Epsilon $\varepsilon \in \{0.01, 0.1, 1, 10, 100\}$, Gamma $\gamma \in \{0.01, 0.1, 1, 10, 100\}$ and Polynomial degree $d \in \{0, 1, 2, 3, 4\}$. Each combination was evaluated using cross-validation, and the optimal set was selected based on the highest score, indicating the best generalization capability.

Data and Variables

The dataset used in this study covers the period from January 2017 to November 2024. The cut-off in November was determined by the availability of official publications at the time of data collection, as the most recent consistent data across all variables was only accessible up to that month. In handling the dataset, random splitting was deliberately avoided, since such an approach in time series analysis risks information leakage from future observations into the training set. Instead, the data was partitioned chronologically, with earlier observations allocated for model training and the most recent period reserved for testing and validation. This procedure preserves the temporal structure of the series and ensures that the evaluation reflects realistic forecasting conditions. Dependent variable (Y_t): monthly international tourist arrivals. Independent variables: Lagged arrivals (Y_{t-1}, Y_{t-2}), BI Rate (X_1) and Inflation (X_2)

Justification for variable selection: Lagged tourist arrivals, Y_t: tourism demand is autoregressive, with past arrivals strongly influencing future arrivals due to seasonality and long-term trends. Including lags improves model accuracy by capturing temporal dynamics; BI Rate (X_1) represents monetary policy and liquidity conditions. Lower interest rates stimulate consumption and tourism demand, while higher rates suppress them. Inflation (X_2) directly affects travel affordability, influencing airfare, accommodation, and other costs. Stable inflation enhances destination attractiveness, while high inflation discourages tourism.

These variables are theoretically and empirically relevant to tourism demand. Future studies may enrich the model with additional indicators such as exchange rates, hotel occupancy rates, online search trends (e.g., Google Trends), GDP growth, and global oil prices.

Research Stages

The stages of this research are described as follows:

1. Preparing the data on the number of international tourist arrivals and their indicators in Indonesia from January 2017 to November 2024
2. Conducting descriptive statistical analysis to observe the characteristics and patterns of the data used.
3. Determining the input variables of the international tourist arrival data in Indonesia for the period January 2017–November 2024 using the Partial Autocorrelation Function (PACF) plot. The significant lags were identified through t-tests in multiple regression analysis.
4. Transforming the input variables into a time series format.
5. Normalizing the data. It was performed to scale the data into a smaller range without altering its original characteristics, following Equation (5)

$$X_{nor} = \frac{\text{minimum range} + (X - X_{\min})(\text{maximum range} - \text{minimum range})}{X_{\max} - X_{\min}} \tag{5}$$

6. The dataset comprised 95 observations, which were split chronologically into training and testing subsets using an 80:20 ratio. Accordingly, 76 observations were allocated for model estimation, while the remaining 19 were reserved for evaluating predictive performance on unseen periods. This proportion provides sufficient data for robust parameter estimation while ensuring a meaningful horizon for out of sample validation, consistent with established forecasting practice (Hyndman & Athanasopoulos, 2021).
7. Implementing the Support Vector Regression (SVR) method, using three kernel functions: Linear Kernel, Polynomial Kernel, and Radial Basis Function (RBF) Kernel.
8. Selecting the optimal model and evaluating its performance based on error metrics, including Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD) and MAPE to determine the model with the smallest overall error.
9. The comparative analysis is conducted by benchmarking the subsequent models with the ARIMAX model (Amelia et al., 2021; Chen et al., 2022).

RESULTS AND DISCUSSION

In the period January 2017–November 2024, the number of foreign tourist visits averaged 258 559/month, with a minimum value of 36 064 (June 2021) and a maximum of 499 810 (December 2018). The BI Rate showed an average of 4.861%/month, recorded a minimum of 3.5% (February 2021) and a maximum of 6.25% (April–August 2024). Meanwhile, inflation averaged 1.578%/month, with a lowest value of 0% (January 2024) and a highest of 5.51% (December 2022).

Figure 1 shows the progress of foreign tourist arrivals in Indonesia, covering January 2017 to early 2025. It clearly illustrates three distinct phases in the tourism sector. During the pre-pandemic period (January 2017–early 2020), the number of foreign tourists was consistently high, fluctuating between 300 000 and 500 000 per month, exhibiting a clear seasonal pattern. This robust trend was violently interrupted in early 2020, plunging the data into the Pandemic Impact Phase (early 2020–early 2022), where visits dropped drastically and remained stagnant below 100 000 per month, representing a significant structural break in the series. Subsequently, the data entered a recovery phase starting in early 2022, demonstrating a steady, sustained upward trend and a gradual re-emergence of monthly fluctuations. While recovery is evident, with monthly figures reaching between 300,000 and 375,000 by early 2025, the volume has not yet fully returned to the peak levels observed prior to 2020. This non-stationarity, dominated by the pandemic shock, necessitates careful handling when applying models like Support Vector Regression (SVR).

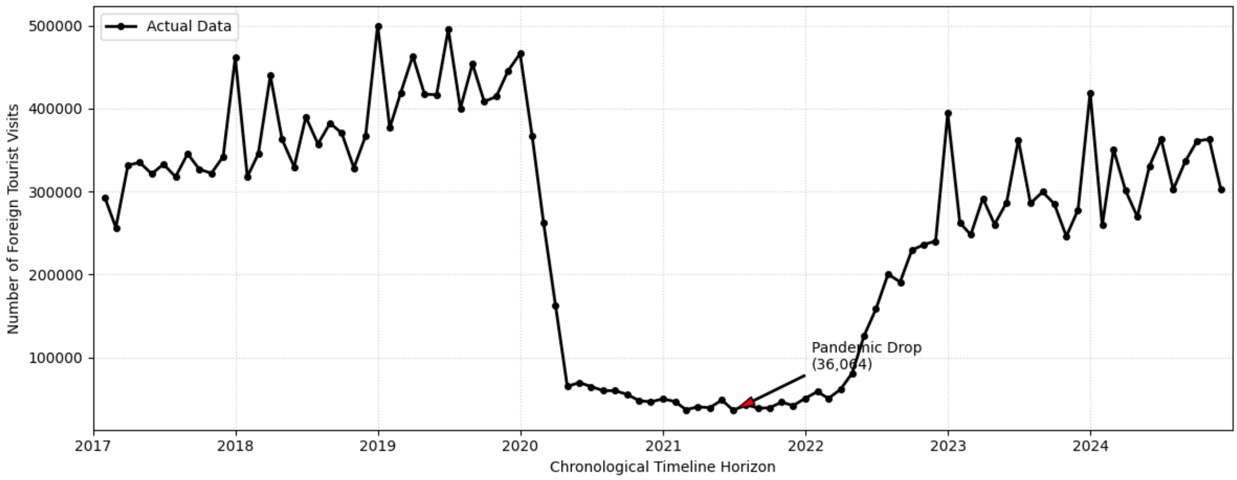

Figure 1. Time series plot of the number of foreign tourist visits.

Based on the visual analysis of BI rate in Figure 2 and inflation in Figure 3, there is an indication of a relationship among foreign tourist arrivals, BI rate, and inflation. A seemingly counter-intuitive pattern emerges where a decrease in the BI Rate (for instance, during August 2017–April 2018 and mid-2019–mid-2022) is associated with a decline in foreign tourist arrivals. Conversely, an increase in the BI Rate (in mid-2018–mid-2019 and mid-2022–2024) is followed by an upward trend in foreign tourist arrivals. Furthermore, the tendency for Indonesian inflation to rise at the end of the year (with a significant peak in December 2023) theoretically reduces domestic purchasing power. However, assuming a stronger foreign exchange rate, the rising inflation can offer a relative advantage to foreign tourists, as the cost of living in Indonesia becomes more affordable (the real price effect)

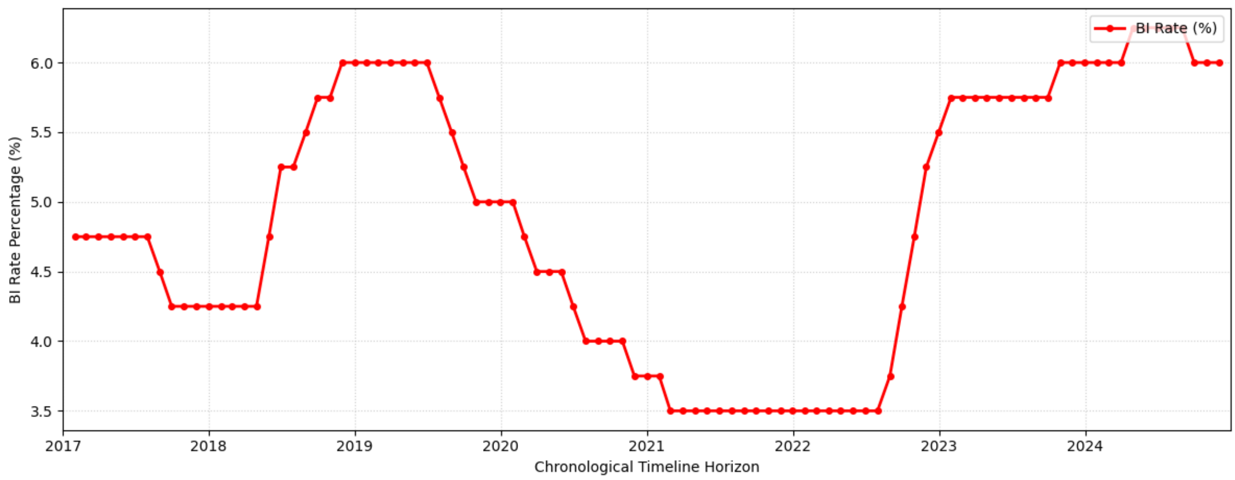

Figure 2. Time series plot of BI rate.

The forecasting process for the number of foreign tourist arrivals using Support Vector Regression (SVR) commences with a crucial step: the determination of input variables for SVR, followed by the model fitting and parameter optimization stages. Input variables for the SVR model were determined through two primary approaches: (1) The selection of BI Rate and Inflation as exogenous economic factors hypothesized to influence foreign tourist arrivals, and (2) The identification of temporal lags using Partial Autocorrelation Function (PACF) analysis. The PACF plot of the tourist arrivals data identified significant lags at specific intervals, which served as the basis for determining the lagged variables included in the SVR model. This approach is analogous to identifying the order of an Autoregressive (AR(p)) model, where the PACF plot helps determine which past periods, Y_{t-k} , significantly influence the current data Y_t.

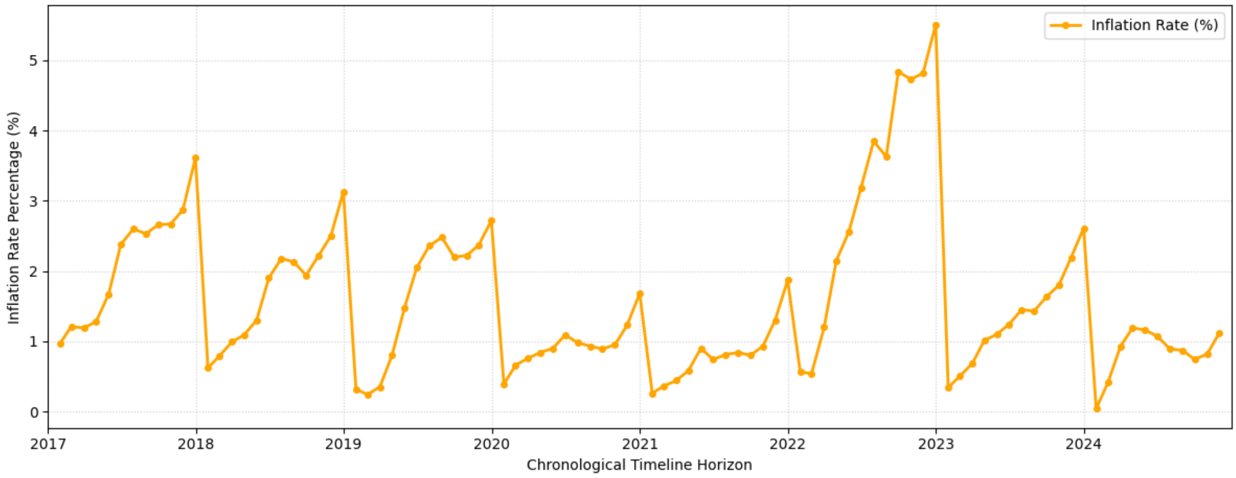

Figure 3. Time series plot of inflation.

The PACF plot in Figure 4 indicated significant cut-offs at lags 1, 2, 4, 7, 13, and 23. A subsequent multiple regression test was conducted at a significance level of α = 5% to validate the influence of these specific lags on the total foreign tourists. The test results showed that lag 1 ($|t_{value}| = 5{,}720$) and lag 2 ($|t_{value}| = 2{,}566$) were statistically significant ($|t_{value}| > t_{\frac{\alpha}{2},89}$), while lags 4, 7, 13, and 23 were not significant ($|t_{value}| < t_{\frac{\alpha}{2},89}$). Based on these findings, the final SVR input variables selected are lag 1, lag 2, the BI Rate, and Inflation.

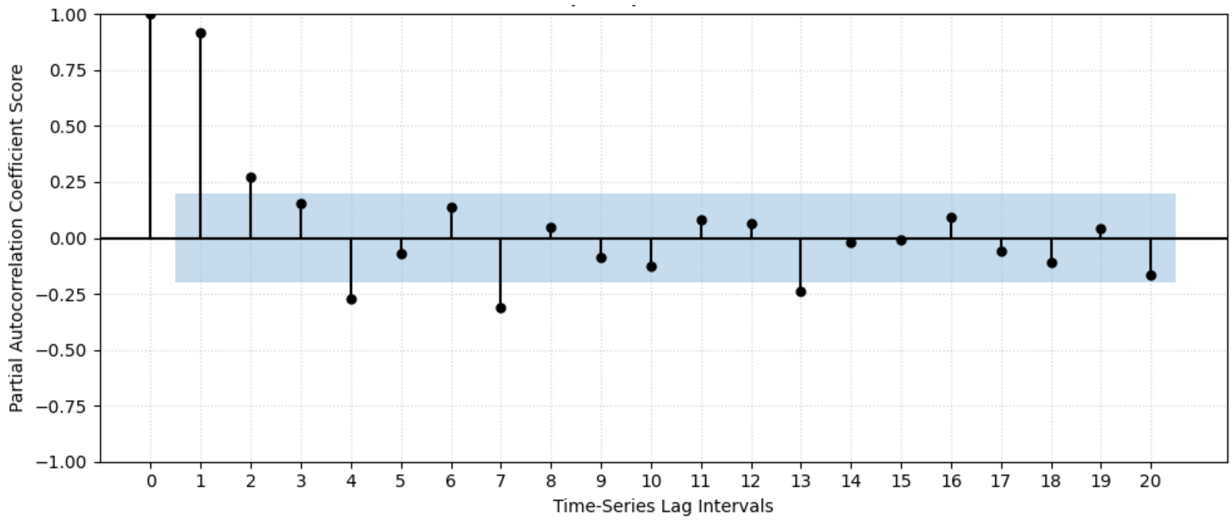

Figure 4. PACF plot.

The prediction of international tourist arrivals was carried out using the Support Vector Regression (SVR) method with three kernel functions, namely Linear, Polynomial, and Radial Basis Function (RBF). The optimal hyperparameters for each kernel were determined using a grid search approach combined with k-fold cross-validation ($k = 3$), where the data were partitioned into three folds (1/3, 2/3, and 3/3) across multiple iterations. This procedure systematically evaluates various combinations of parameters to identify those that yield the best generalization performance. The cross-validation score obtained in each iteration reflects the model's ability to predict unseen data, with higher values indicating better performance. The overall objective of this process is to ensure that the selected parameters minimize prediction error while maintaining model stability.

For the Linear Kernel, the parameter tuning process involved 25 iterations with variations in the cost parameter (C) and epsilon (ε). The results show that lower parameter values, such as $C = 0.01$ and $\varepsilon = 0.01$, produced relatively low average scores (0.287), indicating limited predictive capability. In contrast, higher parameter values, particularly $C = 0.1$ and $\varepsilon = 0.01$, yielded significantly improved performance, with cross-validation scores of 0.912, 0.691, and 0.846 across the three folds, resulting in an average score of 0.816. This indicates that the model achieves better generalization when a moderate penalty is imposed on the error term. Consequently, the combination $C = 0.1$ and $\varepsilon = 0.01$ was selected as the optimal configuration for the Linear Kernel, with the regularization parameter $\lambda = 0.01$.

For the polynomial kernel, the tuning process was more extensive, involving 125 iterations and incorporating an additional parameter, namely the polynomial degree (d). The results demonstrate that parameter combinations with lower complexity, such as $C = 0.1, d = 1$ and $\varepsilon = 0.01$, produced an average score of 0.817. However, improved performance was observed at higher parameter values, specifically $C = 10, d = 2$ and $\varepsilon = 0.01$, which resulted in cross-validation scores of 0.891, 0.754, and 0.851 with an average score of 0.832. This indicates that a moderate degree of nonlinearity enhances the model's ability to capture underlying patterns in the data. Therefore, this parameter combination was selected as the optimal configuration for the polynomial kernel with $\lambda = 0.01$.

For the RBF Kernel, the parameter tuning involved three key parameters: cost (C), epsilon (ε), and gamma (γ), across 125 iterations. The results reveal that certain parameter combinations, such as $C = 1, \varepsilon = 100$, and $\gamma = 100$, produced negative cross-validation scores, indicating poor model generalization and overfitting or underfitting issues. In contrast, the combination $C = 10, \varepsilon = 0.01$, and $\gamma = 0.01$ yielded significantly better results, with scores of 0.920, 0.699, and 0.826 across the folds, resulting in an average score of 0.815. These findings suggest that appropriate scaling of the gamma parameter is essential for controlling the flexibility of the decision boundary in the feature space. Thus, this configuration was selected as the optimal parameter set for the RBF Kernel, with $\lambda = 0.01$.

Overall, the parameter optimization results demonstrate that the selection of hyperparameters plays a critical role in determining the performance of the SVR model. Each kernel requires a specific combination of parameters to achieve optimal generalization capability. The selected optimal parameters for each kernel were subsequently used during model training and prediction, forming the basis for comparing the predictive performance of the Linear, Polynomial, and RBF kernels. This process ensures that the final SVR models are constructed using the most appropriate parameter configurations, thereby enhancing the reliability and accuracy of the forecasting results. Model for each kernel as below,

a. Linear Kernel with 68 support vector observations:

$$f(x) = \sum_{j=1}^{68} (\alpha_j^* - \alpha_j)\left(K(x_i, x_j) + \lambda^2\right)$$

$$f(x) = \sum_{j=1}^{68} (\alpha_j^* - \alpha_j)\left(\left(\begin{bmatrix} 0.1110 & \cdots & 0.1664 \\ \vdots & \ddots & \vdots \\ 0.1190 & \cdots & 0.1554 \end{bmatrix}\begin{bmatrix} 0.5131 & \cdots & 0.2102 \\ \vdots & \ddots & \vdots \\ 0.1200 & \cdots & 0.0841 \end{bmatrix}^t\right) + 0.01^2 J_{68\times 19}\right) \quad (6)$$

b. Polynomial Kernel with 61 support vector observations:

$$f(x) = \sum_{j=1}^{61} (\alpha_j^* - \alpha_j)\left(K(x_i, x_j) + \lambda^2\right)$$

$$K(x_i, x_j) = (x_i, x_j^T + 1)^2 \quad (7)$$

$$f(x) = \sum_{j=1}^{61} (\alpha_j^* - \alpha_j)\left(\left(\begin{bmatrix} 0.8380 & \cdots & 0.9091 \\ \vdots & \ddots & \vdots \\ 0.1621 & \cdots & 0.4607 \end{bmatrix}\begin{bmatrix} 0.5131 & \cdots & 0.2102 \\ \vdots & \ddots & \vdots \\ 0.1200 & \cdots & 0.0841 \end{bmatrix}^t + 1J_{61\times 19}\right)^2 + 0.01^2 J_{61\times 19}\right)$$

c. RBF Kernel with 69 support vector observations:

$$f(x) = \sum_{j=1}^{69} (\alpha_j^* - \alpha_j)\left(K(x_i, x_j) + \lambda^2\right) \quad (8)$$

$$K(x_i, x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right)$$
$$\|x_i - x_j\|^2 = \left(d(x_i, x_j)\right)^2$$
$$f(x) = \sum_{j=1}^{69}(\alpha_j^* - \alpha_j)\left(\exp\left(-0.01\|x_i - x_j\|^2\right) + 0.01^2\right)$$

The prediction results for all three kernel models are illustrated in Figure 5, which presents the comparison between actual data and predicted values. Visually, all models demonstrate the ability to follow the general pattern of the observed data, including both trend and short-term fluctuations. This indicates that the SVR approach, regardless of the kernel used, can capture the underlying structure of the time series data. However, slight differences in the smoothness and responsiveness of the predicted curves can be observed, reflecting the distinct characteristics of each kernel function in modeling data patterns.

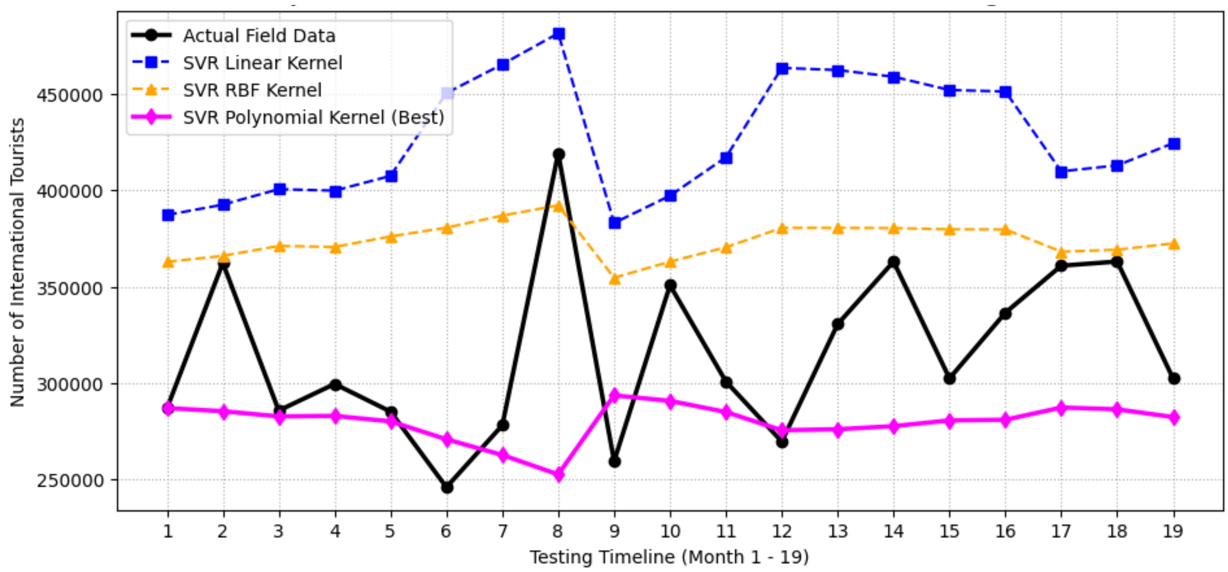

Figure 5. Comparison of evaluation of SVR kernel performances**.**

To evaluate quantitatively model performance, three error metrics were employed: Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD) and Mean Absolute Percentage Error (MAPE). The results of these evaluations for each kernel are summarized in Table 1.

Table 1. Accuracy value of prediction results.

Accuracy	Linear Kernel	Polynomial Kernel	RBF Kernel
MSE	14,764,415,856.32	3,401,140,779.52	7,857,235,590.63
RMSE	121,508.91	58,319.30	88,641.05
MAD	111,351.87	44,565.19	73,609.79
MAPE	37.60%	12.94%	25.63%

The results demonstrate that the polynomial kernel achieved the best predictive performance among the evaluated SVR models. It produced the lowest error values across all evaluation metrics, with an RMSE of 58,319.30, MAD of 44,565.19, and MAPE of 12.94%. These results indicate that the Polynomial Kernel provides more accurate predictions with smaller deviations between predicted and actual tourist arrival values.

The superior performance of the polynomial kernel indicates that the relationship between macroeconomic indicators and tourist arrivals exhibits nonlinear characteristics. By applying polynomial feature transformation, the SVR model can capture more complex interactions between

BI Rate, Inflation, and tourist arrivals. Therefore, the polynomial kernel provides a more suitable representation of the underlying data patterns compared with Linear and RBF kernels.

Based on the evaluation results, the SVR model using the Polynomial Kernel was selected as the optimal configuration for tourist arrival prediction. The obtained MAPE value of 12.94% indicates good forecasting accuracy, demonstrating that the model is capable of capturing the nonlinear relationship between economic variables and tourism demand.

The results of the subsequent models are interpreted in comparison with those obtained from the ARIMAX model. The identification process of the ARIMAX model resulted in the selection of ARIMAX(1,1,1)(BI-Rate,Infl) as the optimal specification based on the lowest Akaike Information Criterion (AIC) value. The model successfully incorporated the dynamic behavior of tourist arrivals through the AR(1) and MA(1) components, while BI Rate and inflation were included to represent the influence of macroeconomic conditions. The diagnostic evaluation indicated that the residuals satisfied the white-noise assumption, confirming that the model was statistically adequate for forecasting purposes. The general formulation of the ARIMAX(1,1,1) model can be represented as follows:

$$(1 - 1 - \varphi_1 B)(1 - B)Y_t = (1 + \theta_1 B)\, e_t + \beta_1 X_{1,t} + \beta_2 X_{2,t} + e_t \tag{9}$$

The ARIMAX(1,1,1) model incorporates both temporal dependencies and external economic factors. The autoregressive component captures the influence of previous tourist arrival patterns, while the moving average component accounts for past forecasting errors. In addition, BI Rate and inflation are included as exogenous variables to explain the contribution of macroeconomic conditions to tourism demand variations.

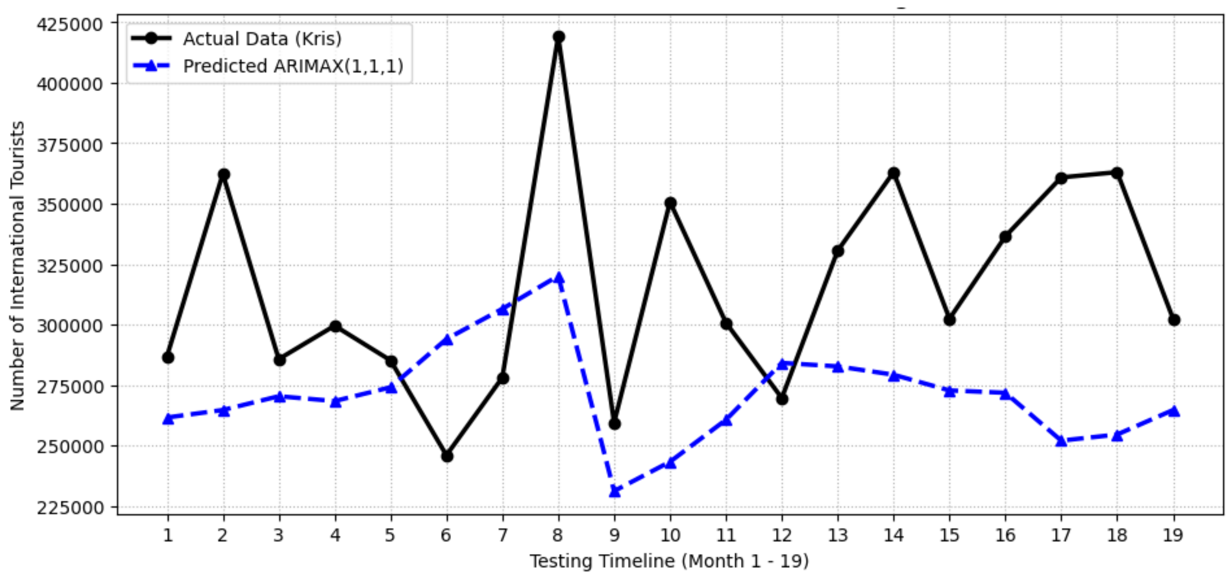

Figure 6. ARIMAX kernel actual vs predicted.

The forecasting performance of the ARIMAX(1,1,1) and SVR-Polynomial models was evaluated using several accuracy metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD), and Mean Absolute Percentage Error (MAPE). The comparison results are presented in Table 2.

Table 2. Forecasting performance comparison between arimax and svr-polynomial.

Model	MSE	RMSE	MAD	MAPE
ARIMAX(1,1,1)	4.109.713.094,16	64.107,04	54.106,17	16.23%
SVR _Polynomial	3.401.140.779,52	58.319,30	44.565,19	12.94%

The SVR-Polynomial model produced lower error values, with improvements of 17.25% in MAE, 9.03% in RMSE, 17.63% in MAD, and 20.33% in MAPE compared with ARIMAX(1,1,1). These results indicate that SVR-Polynomial provides more accurate forecasts and better generalization capability for the tourist arrival dataset.

The superior performance of SVR-Polynomial is attributed to the ability of the polynomial kernel to capture nonlinear relationships between economic variables and tourist arrival patterns. In contrast, ARIMAX relies on linear temporal dependencies through autoregressive and moving average components, which may limit its ability to represent complex nonlinear fluctuations. Therefore, the SVR-Polynomial model is considered more suitable for forecasting tourist arrivals in this study.

To construct the tourism forecast for January–December 2025, external macroeconomic indicators were incorporated as explanatory variables. Monthly inflation rates published by BPS were used as inputs, with values of −0.76% (January), −0.48% (February), 1.65% (March), 1.17% (April), −0.37% (May), 0.19% (June), 0.30% (July), −0.08% (August), 0.21% (September), 0.28% (October), 0.17% (November), and 0.64% (December). In addition, the BI Rate trajectory was included to capture monetary policy effects. The rate began at 5.75% in January 2025, declined gradually, and stabilized at 4.75% from September through December, consistent with Bank Indonesia's policy decision recorded on December 17, 2025. These variables were integrated into the forecasting model to account for both price-level dynamics and interest rate movements, thereby providing a comprehensive representation of external economic conditions influencing tourism demand in Table 3

Table 3. Results of prediction and forecasting of foreign tourist visits.

Month 2025	BI Rate (%)	Inflation (%)	Tourist Forecast (persons)
January	5.75	−0.76	292,122
February	5.75	−0.48	334,163
March	5.75	1.65	314,720
April	5.75	1.17	322,513
May	5.50	−0.37	335,379
June	5.50	0.19	331,066
July	5.25	0.30	335,089
August	5.00	−0.08	346,099
September	4.75	0.21	357,760
October	4.75	0.28	356,469
November	4.75	0.17	353,280
December	4.75	0.64	353,190

Economic and Strategic Implications of Forecasting Results

The comparative evaluation demonstrates that the SVR Polynomial Kernel model provides more accurate forecasts of international tourist arrivals than the ARIMAX(1,1,1) model. The SVR model achieved lower error values, with an MSE of 3,401,140,779.52, RMSE of 58,319.30, MAD of 44,565.19, and MAPE of 12.94%, compared with ARIMAX, which obtained a MAPE of 16.23%. The lower forecasting errors indicate that the nonlinear SVR model is better able to represent the complex patterns in tourism demand observed in the dataset. However, these results should be interpreted within the scope of the models and data considered in this study.

From an economic perspective, more accurate forecasts of international tourist arrivals can support tourism planning and resource allocation. Forecast information can help tourism stakeholders anticipate changes in visitor demand and adjust accommodation capacity, transportation services, tourism infrastructure, workforce requirements, and promotional activities. The inclusion of inflation and the BI Rate as explanatory variables also provides a means of incorporating changes in the domestic economic environment into tourism demand forecasting. Thus, the forecasting model can serve as an analytical tool for supporting adaptive planning rather than relying solely on historical tourism trends.

From a resilience perspective, forecasting can help identify potential declines or increases in international tourist arrivals at an earlier stage. Tourism is sensitive to external shocks such as global economic uncertainty, geopolitical instability, health emergencies, natural disasters, and changes in international travel conditions. A forecasting system that incorporates both historical tourism patterns and macroeconomic indicators can provide an early indication of changes in tourism demand. This information may support the preparation of alternative strategies, such as adjusting tourism promotion, reallocating resources, and preparing contingency measures for periods of declining arrivals. In this context, forecasting contributes not only to tourism growth but also to the ability of the tourism sector to adapt to external disturbances.

From a strategic and national perspective, a stable tourism sector can contribute to broader economic resilience through foreign-exchange generation, employment, business activity, and regional economic development. Therefore, the proposed forecasting approach can be considered a quantitative decision-support tool for monitoring tourism conditions and supporting evidence-based planning. The forecasting results for 2025, which show monthly tourist arrivals ranging from approximately 292,000 to 358,000 visitors, provide an initial reference for anticipating monthly demand and adjusting tourism-related capacity and services accordingly. Nevertheless, the model should be further evaluated using longer forecasting horizons, additional economic and external variables, and alternative machine-learning and time-series models before being applied as a standalone basis for policy decisions.

CONCLUSION

This study examines the characteristics and forecasts the number of international tourist arrivals in Indonesia using the Support Vector Regression (SVR) approach with various kernel functions. Between January 2017 and November 2024, the average monthly number of tourist arrivals was 258 559 visits, with the lowest recorded in June 2021 (36 064 visits) and the highest in December 2018 (499 810 visits). The analysis also reveals that macroeconomic factors, particularly the BI Rate and inflation, are closely linked to fluctuations in tourist arrivals. The average monthly BI Rate was 4.861%, while the average monthly inflation rate stood at 1.578%. This study demonstrates that Support Vector Regression (SVR) with a second-degree Polynomial Kernel is the most effective and dependable model for forecasting international tourist arrivals in the Kris dataset. By applying StandardScaler for data standardization and Grid Search for hyperparameter optimization, the SVR Polynomial model achieved a Mean Absolute Percentage Error (MAPE) of 12.94%, outperforming the classical econometric benchmark ARIMAX(1,1,1), which recorded a higher MAPE of 16.23%.

The novelty of this research lies in the empirical validation of a non-linear machine learning approach that more accurately captures the influence of macroeconomic indicators, specifically the BI Rate and inflation, on tourism demand. The robustness and precision of the SVR Polynomial model highlight its superiority over traditional econometric methods, establishing machine learning as a reliable alternative for tourism forecasting. These findings provide strong evidence that advanced non-linear models can serve as strategic instruments for national tourism policy, enabling more accurate projections and data-driven decision-making. By integrating macroeconomic dynamics into predictive modeling, this study contributes to the methodological advancement of tourism economics and offers practical implications for policymakers seeking to enhance resilience and competitiveness in the tourism sector.

AUTHOR CONTRIBUTIONS

K.S.: Conceptualization of the study, numerical simulation, visualization, and drafting of the original manuscript. M.O.D.: Data curation, methodology development, validation, and the review and editing of the manuscript. R.D.B.: Formal analysis, validation, supervision, and contributed to the review and editing process. J.: Investigation and formal analysis. All authors actively discussed the results and contributed to the preparation and approval of the final manuscript.

CONFLICT OF INTEREST

The authors declare that have no conflict of interest.

REFERENCES

Adam, M., Chrisnanto, Y. H., & Umbara, F. R. (2024). Prediksi Pola Keuangan pada Pasar Saham Bursa Efek Indonesia Menggunakan Algoritma Support Vector Machine For Regression (SVR). *Jurnal Edukasi & Penelitian Informatika*, *10*(3), 476–484.

Agan Celik, B., Yulek, M., & Celik, S. (2026). Machine learning assessment of drivers of tourist arrivals in Istanbul: evidence from economic, climatic and behavioral indicators. *Journal of Hospitality and Tourism Technology*, 1–25. https://doi.org/10.1108/JHTT-08-2025-0686

Akanbi, O. B., & Bello, A. O. (2024). Fitting Autoregressive Integrated Moving Average with Exogenous Variables Model with Lognormal Error Term. *Journal of Scientific Research and Reports*, *30*(10), 158–168. https://doi.org/10.9734/jsrr/2024/v30i102442

Amelia, R., Dalimunthe, D. Y., Kustiawan, E., & Sulistiana, I. (2021). ARIMAX model for rainfall forecasting in Pangkalpinang, Indonesia. *IOP Conference Series: Earth and Environmental Science*, *926*(1), 012034. https://doi.org/10.1088/1755-1315/926/1/012034

Cankurt, S., & Subaşi, A. (2016). Tourism demand modelling and forecasting using data mining techniques in multivariate time series: a case study in Turkey. *Turkish Journal Of Electrical Engineering & Computer Sciences*, *24*, 3388–3404. https://doi.org/10.3906/elk-1311-134

Chen, K.-Y., & Wang, C.-H. (2007). Support vector regression with genetic algorithms in forecasting tourism demand. *Tourism Management*, *28*(1), 215–226. https://doi.org/10.1016/j.tourman.2005.12.018

Chen, Y.-P., Liu, L.-F., Che, Y., Huang, J., Li, G.-X., Sang, G.-X., Xuan, Z.-Q., & He, T.-F. (2022). Modeling and Predicting Pulmonary Tuberculosis Incidence and Its Association with Air Pollution and Meteorological Factors Using an ARIMAX Model: An Ecological Study in Ningbo of China. *International Journal of Environmental Research and Public Health*, *19*(9), 5385. https://doi.org/10.3390/ijerph19095385

Feng, S., Yu, Z., Ren, X., & Yu, S. (2022). *The Impact of the COVID-19 Pandemic on the Tourism Industry and What Kind of Future this Industry will be Led to*. https://doi.org/10.2991/aebmr.k.220405.317

Guerouate, F., Sbihi, M., & Laaroussi, H. (2023). Deep learning approach and topic modelling for forecasting tourist arrivals. *International Journal of Electrical and Computer Engineering Systems*, *14*(4), 411–419. https://doi.org/10.32985/ijeces.14.4.5

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice (3rd ed.)* (3rd editiom). OTexts.

Iriananda, S. W., Budiawan, R. W., Rahman, A. Y., & Istiadi, I. (2024). Optimasi Klasifikasi Sentimen Komentar Pengguna Game Bergerak Menggunakan Svm, Grid Search Dan Kombinasi N-Gram. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *11*(4), 743–752. https://doi.org/10.25126/jtiik.1148244

Jouilil, Y., & Mentagui, D. (2023). Comparing The Forecasting Accuracy Metrics of Support Vector Regression and ARIMA Algorithms for Non-Stationary Time Process. *Mathematics and Statistics*, *11*(2), 294–299. https://doi.org/10.13189/ms.2023.110207

Lalian, V. U., Bekti, R. D., Pratiwi, N., Sutanta, E., & Pradnyana, I. W. J. (2022). Analysis of Consumer Satisfaction Levels with GoRide Services Using the Support Vector Machine (SVM) Classification Method. *2022 IEEE 8th Information Technology International Seminar (ITIS)*, 51–56. https://doi.org/10.1109/ITIS57155.2022.10010286

Maulana, N. D., Setiawan, B. D., & Dewi, C. (2019). Implementasi Metode Support Vector Regression (SVR) Dalam Peramalan Penjualan Roti (Studi Kasus: Harum Bakery). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(3), 2986–2995.

Mishra, R. K., Urolagin, S., Jothi, J. A. A., Nawaz, N., & Ramkissoon, H. (2021). Machine Learning based Forecasting Systems for Worldwide International Tourists Arrival. *International Journal of Advanced Computer Science and Applications*, *12*(11). https://doi.org/10.14569/IJACSA.2021.0121107

Nguyen, T.-L., Shu, M.-H., Huang, Y.-F., & Hsu, B.-M. (2013). Accurate forecasting models in predicting the inbound tourism demand in Vietnam. *Journal of Statistics and Management Systems*, *16*(1), 25–43. https://doi.org/10.1080/09720510.2013.777570

Nur Aini, L. (2022). Pengaruh Inflasi, Bank Indonesia Rate Dan Nilai Tukar Rupiah Terhadap Harga Saham Pada Perusahaan Sektor Transportasi Dan Logistik Di Bursa Efek Indonesia Periode 2015-2018. *SIBATIK JOURNAL: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, Dan Pendidikan*, *1*(4), 219–234. https://doi.org/10.54443/sibatik.v1i4.27

Škare, M., Soriano, D. R., & Porada-Rochoń, M. (2021). Impact of COVID-19 on the travel and tourism industry. *Technological Forecasting and Social Change*, *163*, 120469. https://doi.org/10.1016/j.techfore.2020.120469

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Sun, S., Wei, Y., Tsui, K.-L., & Wang, S. (2019). Forecasting tourist arrivals with machine learning and internet search index. *Tourism Management*, *70*, 1–10. https://doi.org/10.1016/j.tourman.2018.07.010

Surtiningsih, L. (2017). *Prediksi jumlah kunjungan wisatawan mancanegara ke Bali menggunakan support vector regression dengan algoritma genetika*. Universitas Brawijaya.

Suryawan, I. W. K., Suhardono, S., Rahman, A., Phan, T. T. T., & Lee, C.-H. (2025). Redesigning Tourism Management Using Hypothetical Scenario Valuation Based on Circular Economy Design Framework. *Circular Economy and Sustainability*, *5*(5), 4027–4053. https://doi.org/10.1007/s43615-025-00581-9

Tovmasyan, G. (2021). Forecasting the number of incoming tourists using Arima model: case study from Armenia. *Marketing and Management of Innovations*, *5*(3), 139–148. https://doi.org/10.21272/mmi.2021.3-12

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. Springer New York. https://doi.org/10.1007/978-1-4757-3264-1

Zhou, Y., Li, J., Ma, G., & Zhu, Y. (2024). Comparison of parameter optimization methods for a runoff forecast model based on a support vector machine. *Physics and Chemistry of the Earth, Parts A/B/C*, *135*, 103653. https://doi.org/10.1016/j.pce.2024.103653