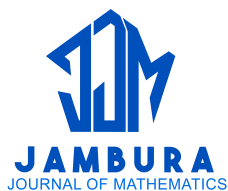


# Comparison of Seasonal ARIMA and Support Vector Machine Forecasting Method for International Arrival in Lombok

Hadyanti Utami MY *et al.*



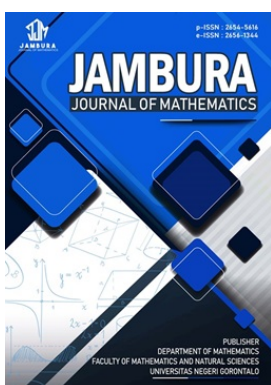
Volume 6, Issue 2, Pages 212–219, August 2024

Received 7 June 2024, Revised 25 July 2024, Accepted 28 July 2024, Published 3 August 2024

To Cite this Article : H. Utami MY, *et al.*, “Comparison of Seasonal ARIMA and Support Vector Machine Forecasting Method for International Arrival in Lombok”, *Jambura J. Math*, vol. 6, no. 2, pp. 212–219, 2024, <https://doi.org/10.37905/jjom.v6i2.26478>

© 2024 by author(s)

## JOURNAL INFO • JAMBURA JOURNAL OF MATHEMATICS

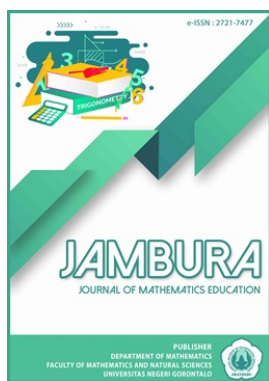


	Homepage	:	<a href="http://ejurnal.ung.ac.id/index.php/jjom/index">http://ejurnal.ung.ac.id/index.php/jjom/index</a>
	Journal Abbreviation	:	Jambura J. Math.
	Frequency	:	Biannual (February and August)
	Publication Language	:	English (preferable), Indonesia
	DOI	:	<a href="https://doi.org/10.37905/jjom">https://doi.org/10.37905/jjom</a>
	Online ISSN	:	2656-1344
	Editor-in-Chief	:	Hasan S. Panigoro
	Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
	Country	:	Indonesia
	OAI Address	:	<a href="http://ejurnal.ung.ac.id/index.php/jjom/oai">http://ejurnal.ung.ac.id/index.php/jjom/oai</a>
	Google Scholar ID	:	iWLjgaUAAAAJ
	Email	:	<a href="mailto:info.jjom@ung.ac.id">info.jjom@ung.ac.id</a>

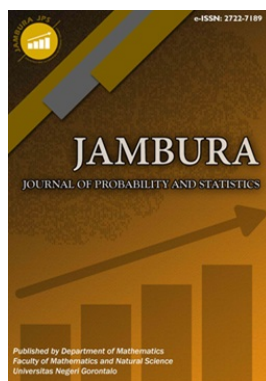
## JAMBURA JOURNAL • FIND OUR OTHER JOURNALS



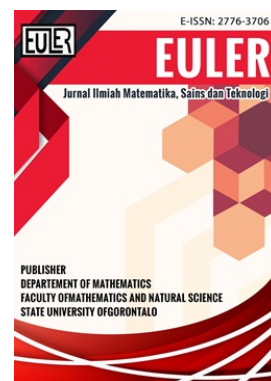
Jambura Journal of Biomathematics



Jambura Journal of Mathematics Education



Jambura Journal of Probability and Statistics



EULER : Jurnal Ilmiah Matematika, Sains, dan Teknologi

# Comparison of Seasonal ARIMA and Support Vector Machine Forecasting Method for International Arrival in Lombok

Hadyanti Utami MY<sup>1,\*</sup>, Silfiana Lis Setyowati<sup>1</sup>, Khairil Anwar Notodiputro<sup>1</sup> ,  
Yenni Angraini<sup>1</sup> , and Laily Nissa Atul Mualifah<sup>1</sup> 

<sup>1</sup>Department of Statistics and Data Science, IPB University, Bogor, Indonesia

## ARTICLE HISTORY

Received 7 June 2024  
Revised 25 July 2024  
Accepted 28 July 2024  
Published 3 August 2024

## KEYWORDS

SARIMA  
Support Vector Machine  
Forecasting  
Lombok International Airport  
Seasonal Patterns  
International Arrivals

**ABSTRACT.** Seasonal Autoregressive Integrated Moving Average is a statistical model designed to analyze and forecast data with that shows seasonal patterns and trends. Support Vector Machine (SVM) is a machine learning-based technique that can be used to forecast time series data. SVM uses the kernel tricks to overcome non-linearity problems, whereas The SARIMA model is well-suited for data that exhibit seasonal fluctuations that repeat over time. Lombok International Airport is the main gateway to West Nusa Tenggara and has become a symbol of tourism growth in the region. Time series analysis is a very useful tool in determining patterns and forecasting the number of international arrivals at Lombok International Airport within a certain period. This study aims to compare the SARIMA model and SVM which can read non-linear patterns in the number of international arrivals at Lombok International Airport. After obtaining the SARIMA and SVM models, the two models are evaluated using test data based on the smallest RMSE value. The SVM model with a linear kernel trick provides the smallest RMSE when compared to SARIMA with SVM RMSE is 238,655. While the best model in Seasonal ARIMA is SARIMA (3,1,0)(1,0,0)<sup>1,2</sup>, the forecasting results show SARIMA is better in the forecasting process for the next 10 months.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. *Editorial of JJoM:* Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B. J. Habibie, Bone Bolango 96554, Indonesia.

## 1. Introduction

Time series data is data collected sequentially in a certain time interval and is analyzed to find patterns of past variance that can be used to forecast values for the future [1]. The goal of time series data analysis is to understand the patterns that occur in data measured sequentially over time. After identifying the patterns in the time series data, time series analysis can be used to forecast future values based on historical data. The use of time series analysis is very important because time series data tends to be dependent and often shows correlations between observations. If classical regression is forced on time series data, it will ignore the time dependency between observations and produce an inaccurate model. Lombok International Airport serves as the main gateway to West Nusa Tenggara and has become a symbol of tourism growth in the region. Data shows that since Lombok International Airport began operations, the number of international tourist arrivals has surged, as evidenced by the significant average annual passenger growth. Time series analysis is a very useful tool in determining patterns and forecasting the number of international arrivals at Lombok International Airport over a certain period.

In time series analysis, several methods can be used to forecast future data such as conventional models and machine learning or deep learning. One of the commonly used conventional methods is the SARIMA model. Seasonal Autoregressive Integrated Moving Average is a statistical model designed to an-

alyze and forecast data with that shows seasonal patterns and trends. The SARIMA model is an extension of the Autoregressive Integrated Moving Average model that has seasonal components [2]. The SARIMA model is well-suited for data that exhibit seasonal fluctuations that repeat over time. There are many studies that have conducted forecasting on the number of tourists with the SARIMA model, such as those conducted by Aziza et al. [3] on comparing the SARIMA model with intervention and Prophet to forecast the number of passengers at Soekarno-Hatta International Airport, the results obtained showed that the SARIMA model was better than the Prophet model. In addition, Prianda and Widodo [4] forecasts the number of foreign tourists to Bali by comparing the performance of the SARIMA model and the Extreme Learning Machine. The results given by MAPE on the SARIMA model are smaller than ELM model, this means that SARIMA is better for predicting the number of tourists.

Machine Learning models such as Support Vector Machine (SVM) introduced by Vapnik in 1992 [5] are one of the machine learning-based analysis technique that can be used to forecast time series data. In the context of time series, SVM uses the kernel tricks to overcome non-linearity problems that are often encountered in time series that are difficult to model using traditional statistical methods. The purpose of using this kernel is to transform the data into a high-dimensional space that can be linearly separable [6]. Tourist data tends to fluctuate, this is due to many factors that affect the number of tourist visits in an area that occurs in an extreme manner, such as natural disasters, govern-

\*Corresponding Author.

ment regulations, social stability, riots, and terrorism [7]. Unlike the SARIMA model, the SVM method is more flexible because it does not require the assumption of stationary to produce good predictions. The fluctuating data is done by Saputra et al. [8] forecasting the stock index on the Indonesia Sharia Stock Index (ISSI) which has fluctuating stock data characteristics. The SVR model used can capture non-linear patterns in data with a linear kernel with a parameter optimization process.

This study aims to compare the SARIMA which the conventional model is known to be good at forecasting the seasonal pattern with the SVM which can read non-linear patterns in the number of international arrivals at Lombok International Airport.

## 2. Methods

### 2.1. Data

The data used consists of the Number of Foreign Tourist Visits per month to Indonesia by International Airport Entry Point at Lombok, West Nusa Tenggara. The data is sourced from bps.go.id and kemenparekraf.go.id, collected from January 2014 to February 2024, categorized by month, as shown in Table 1.

Table 1. Lombok foreign tourist visit data

No.	Year	Month	Total
1	2014	January	5105
2	2014	February	4862
3	2014	March	5980
⋮	⋮	⋮	⋮
120	2032	December	7304
121	2024	January	4958
122	2024	February	7748

### 2.2. Stages of Analysis

The stages conducted in the research are visualized with the flowchart in Figure 1. From the dataset, data exploration is carried out to determine the pattern and characteristics of the data through plots. Furthermore, the data is split into training and testing sets to build models and assess the performance of Seasonal ARIMA and SVM models.

- a. In the SARIMA Model, data stationarity is checked using ACF and PACF plots, while checking data stationarity in the mean through the Augmented Dickey-Fuller (ADF) test and stationarity in the variance using Box-Cox, if the value means the data is stationary  $\lambda = 1$  means the data is stationary [4], if the data is not stationary then a differencing process is carried out on the average and variance, if the data is stationary then continue to identify the SARIMA model seen from the ACF and PACF plots. After obtaining a temporary SARIMA model, then estimate the parameters included with the white noise test based on the assumptions of homogeneity of the variance of the residuals, freedom of the residuals, and normality of the residuals are met and compare the smallest AIC (Akaike Information Criterion) value. The SARIMA model is also subjected to a model overfitting process to obtain the best results. Based on these criteria, the best SARIMA model can be selected. The AIC (Akaike Information Criterion) value is obtained based on the following

formulation [9]:

$$AIC = \ln(\hat{\sigma}_a^2) + 2(p + q + 1)/n. \tag{1}$$

- b. Unlike to SARIMA model which is a univariate analysis. In the SVM Model, the first step is to create input and output variables to define. The SVM model is built with training data by tuning the parameters that will be implemented into the SVM kernel function, where the kernels used are: linear, RBF, and sigmoid. The best SVM model was selected based on the smallest RMSE (Root Mean Squared Error).

After obtaining the SARIMA and SVM models, both of model are evaluated using test data based on the smallest RMSE (Root Mean Squared Error), which can then be forecasted for the next 10 periods. The RMSE value calculates the difference between the error in the actual data and forecasting. Therefore, the smaller the resulting RMSE value, the higher the accuracy of the mode [10]. RMSE is formulated as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (A_t - F_t)^2}, \tag{2}$$

where  $A_t$ , is the actual data in period  $t$ ,  $F_t$  is the forecasted value at period  $t$ , and  $n$  is the amount of data.

### 2.3. Seasonal Autoregressive Moving Average (SARIMA) Model

The SARIMA model is an extension of the Autoregressive Moving Average (ARIMA) model, which incorporates seasonal patterns in time series data. The ARIMA model ( $p, d, q$ ) introduced by Box and Jenkins in 1994, has an order  $p$  as the operator of AR, an order  $d$  representing differencing, and an order  $q$  as the MA operator. For data with seasonal patterns, it is denoted as  $(p, d, q)(P, D, Q)^s$ , which includes a seasonal factor in the  $t$  time observation written as [11] :

$$\phi_p(B) \Phi_P(B^s) (1 - B)^d (1 - B^s)^D X_t = \theta_q(B) \theta_Q(B^s) e_t \tag{3}$$

with  $X_t$  is the  $t$  time component,  $\phi_p(B)$  is the AR component of order  $p$ ,  $\Phi_P(B^s)$  is the seasonal component of the AR of order  $P$ , while the  $(1 - B)^d$   $d$ -order differencing,  $(1 - B^s)^D$   $D$ -order seasonal element differencing. The MA component is shown at  $\theta_q(B)$  and the seasonal MA is  $\theta_Q(B^s)$ ,  $e_t$  denotes the residual at time  $t$ . In its process, SARIMA assumes that the data is stationary, meaning that the data fluctuates around a constant mean and the variance does not depend on time [12].

### 2.4. Support Vector Machine (SVM) Model

Support Vector Machine, introduced by Vapnik in 1992 can be applied using classification and regression techniques. In the practice of classification, the SVM model works by finding the best hyperplane with the maximum margin. In the case of regression, the SVM hyperplane is built based on clustering with as many points as needed [13]. Margin is the distance of the hyperplane to the closest pattern of each class, while the pattern closest to the maximum margin is called the support vector [5].

It is assumed that the two classes, -1 and +1, are perfectly separated by a  $d$ -dimensional hyperplane, as defined in the equation:

$$\vec{w} \cdot \vec{x} + b = 0. \tag{4}$$

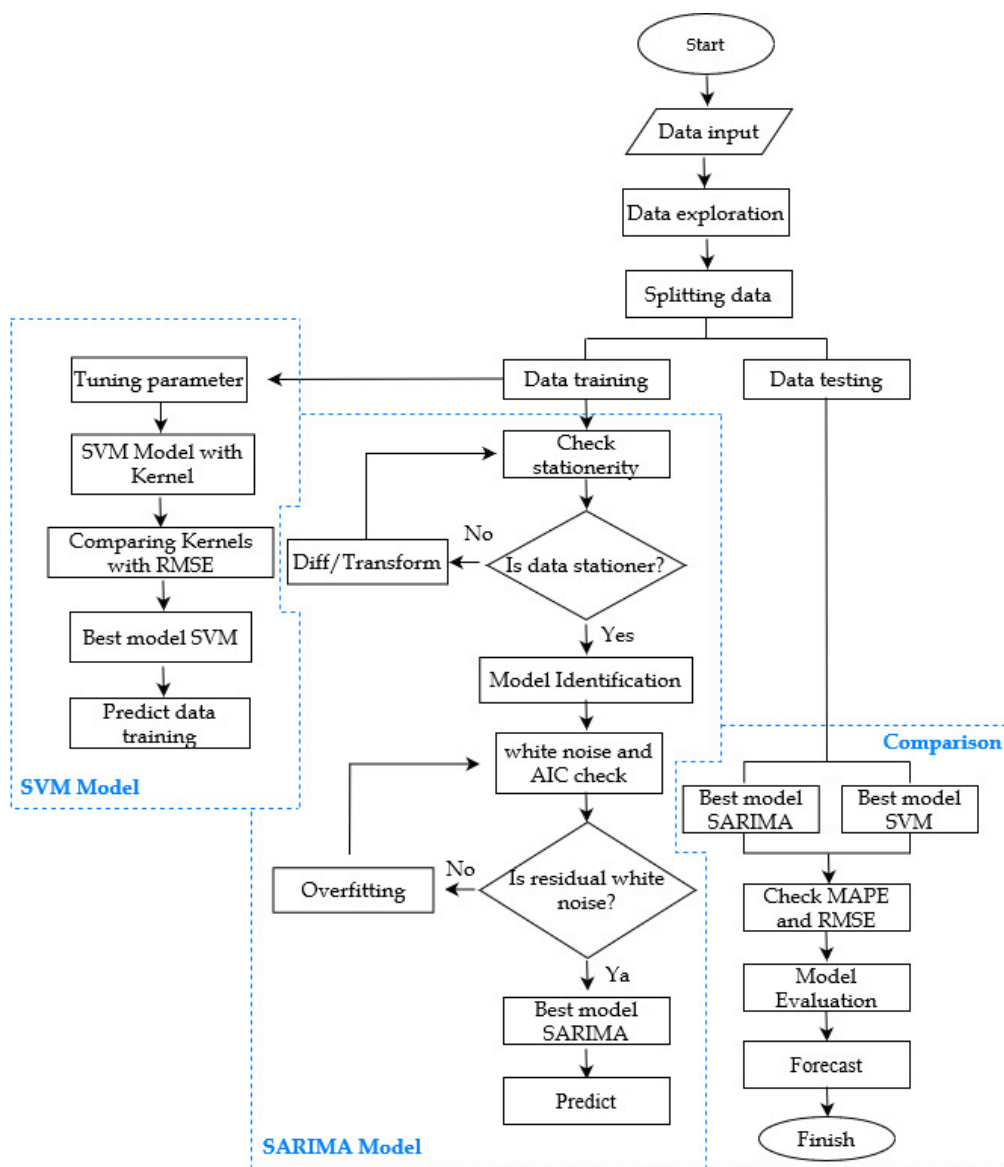


Figure 1. Flowchart of the Research

It is assumed that the two classes, -1 and +1, are perfectly separated by a d-dimensional hyperplane, defined in the equation:

$$\vec{w} \cdot \vec{x} + b \leq -1. \tag{5}$$

If the pattern  $\vec{x}$  belongs to class +1 (negative sample), it can be separated by a hyperplane of dimension  $d$ , with equation :

$$\vec{w} \cdot \vec{x} + b \geq +1, \tag{6}$$

with:  
 $\vec{w}$  = weight vector,  
 $\vec{x}$  = attribute input value,  
 $b$  = bias.

The largest margin value is sought by maximizing the distance value to its closest point, i.e.  $1/\|\vec{w}\|$  at the minimum point.

$$\min \tau(w) = \frac{1}{2} \|\vec{w}\|^2. \tag{7}$$

Suppose a dataset  $(x_1, y_1), \dots, (x_i, y_i)$  where  $x, y \in R$ , where  $x$  is the input variable and  $y$  is the actual (target) value. SVM for regression aims to find a function  $f(x)$  that can predict the target value

with a deviation value of  $\varepsilon$  (epsilon) for each input value, and as much as possible the function  $f(x)$  follows the pattern of fluctuation in the data [14]. By using the Lagrange Multiplier, the optimal condition of the regression function is as follows [15]:

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) (\vec{x}_i \cdot \vec{x}_j) + b, \tag{8}$$

where  $(\vec{x}_i, \vec{x}_j)$  is the pair of two data in training data,  $\alpha_i^*, \alpha_i'$  are Lagrange Multipliers, and  $b$  is coefficient if bias. The estimate value of  $b$  is obtained from  $b = y_i - w \cdot x_i - \varepsilon$  for  $0 \leq \alpha_i \leq C$  and  $b = y_i - w \cdot x_i + \varepsilon$  for  $0 \leq \alpha_i^* \leq C$  [16].

In general, SVM model works in linear separable condition. In non-linear problems, SVM is modified by incorporating a kernel. The kernel functions implemented are linear, Radial Basis Function (RBF), and Sigmoid kernel with the formal as follows [17]:

1. Linear Kernel

$$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i^T \cdot \vec{x}_j. \tag{9}$$

2. Radial Basis Function (RBF) Kernel

$$K(\vec{x}_i, \vec{x}_j) = \exp(-\gamma \|\vec{x}_i - \vec{x}_j\|^2), \tag{10}$$

where  $\gamma$  is the parameter of RBF kernel.

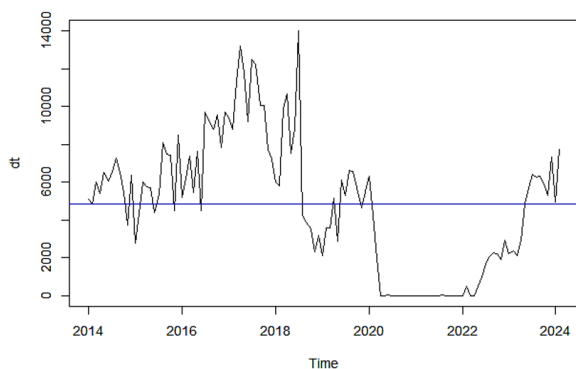
3. Sigmoid Kernel

$$K(\vec{x}_i, \vec{x}_j) = \tanh(\gamma (\vec{x}_i^T \cdot \vec{x}_j) + c). \tag{11}$$

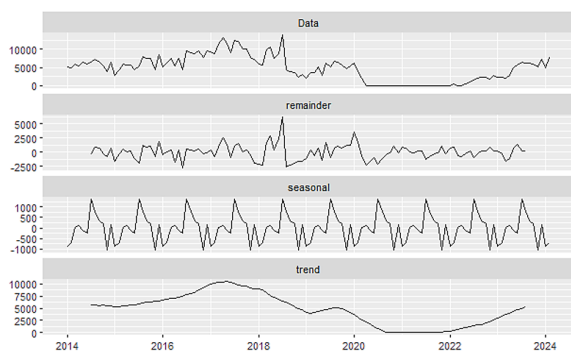
3. Results and Discussion

3.1. Data Exploration

Exploratory data analysis to determine the characteristics of Lombok’s foreign tourist visits can be reviewed with a plot as shown in Figure 2.



(a)



(b)

Figure 2. (a) Data plot and (b) Decomposition plot

Figure 2a illustrates the data does not fluctuate at the average shown on the blue line. The data increased until 2018 and decreased significantly in mid-2018. The COVID-19 pandemic caused no tourist arrivals in 2020-2021, and gradually increased after the pandemic ended. Every year tourist visits increase in certain months, this is also shown in the seasonal Figure 2b which means there is seasonality in the data.

3.2. Data Splitting

The model is formed with training data taken from January 2014 to December 2022 as many as 108 observations. The purpose of partition of the data is to choose the optimal model according to the most significant coefficients, as well as the least value of AIC and satisfaction of the residual assumption. Then, the performance of the model is tested on testing dataset presented by 14 observations for the period from Januari 2023 to

February 2024. This data partitioning ensures an equal number of observations for both the SARIMA and SVM models.

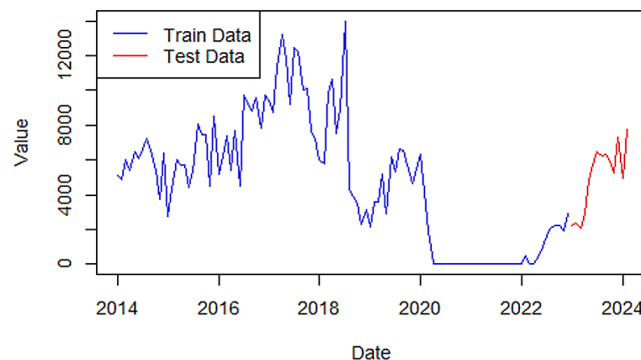
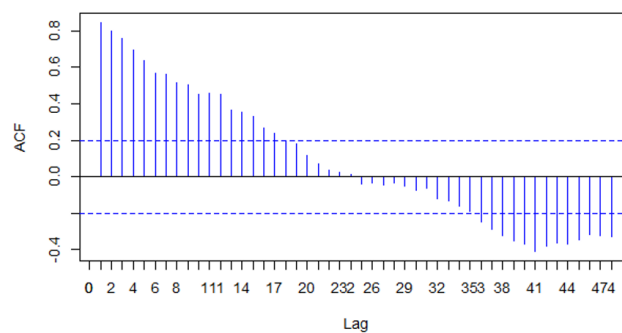


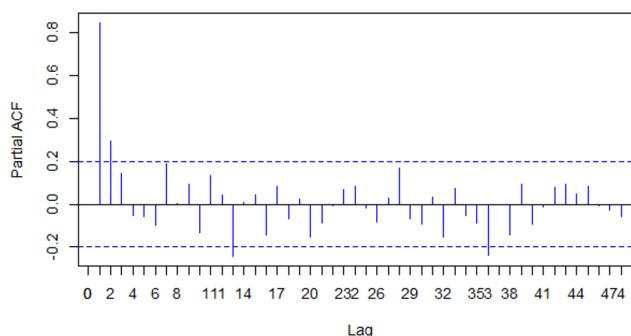
Figure 3. Split of testing and training data

3.3. Seasonal ARIMA Model

Classical model assumptions such as SARIMA require that data be stationary in both mean and variance. Stationarity is checked using ACF and PACF plots to observe the characteristics of lag patterns, along with testing using the ADF and the resulting lambda value.



(a)



(b)

Figure 4. (a) ACF and (b) PACF data

Table 2. Stationary check

ADF Test	Lambda Value
0.4771	0.8481

**Table 3.** SARIMA tentative model parameter estimation

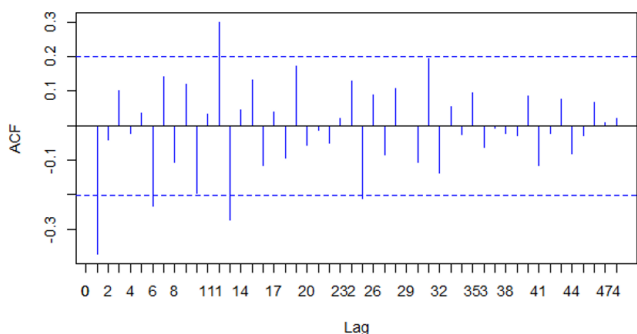
Model	Parameters	Coefficient	P-Value	Parameter Significance	AIC
SARIMA(2,1,1)(1,0,1) <sup>12</sup>	AR(1)	-0.4046	2.87e-05	Sig.	1888.312
	AR(2)	-0.1480	0.1264	No sig.	
	MA(1)	-1.0000	2.2e-16	Sig.	
	SAR(1)	0.3550	0.2407	No sig.	
	SMA(1)	-0.0971	0.7594	No sig.	
SARIMA(0,1,1)(1,0,1) <sup>12</sup>	MA(1)	-0.9999	2e-06	Sig.	1900.352
	SAR(1)	0.4363	0.0694	No sig.	
	SMA(1)	-0.1476	0.5633	No sig.	
SARIMA(2,1,1)(1,0,0) <sup>12</sup>	AR(1)	-0.4072	2.447e-05	Sig.	1886.408
	AR(1)	-0.1470	0.12907	No sig.	
	MA(1)	-0.9999	2.2e-16	Sig.	
	SAR(1)	0.2637	0.00435	Sig.	

Data on foreign tourist arrivals are not stationary in average, illustrated from the ACF plot has a sinusoidal pattern or lag tails off, and the ADF test p-value  $> \alpha (0.05)$ , while the PACF plot is cut off. Stationary data in the variety is determined from the lambda value, the test results show a  $\lambda = 0.8$  or close to 1, according to [18] if lambda is in the rounded value or  $\lambda = 1$  is done by the  $X_t$  value itself, or the data is not transformed which means the data is stationary in variety. The training data was differenced ( $d = 1$ ) in the average, so that the data became stationary with the resulting ADF test, p-value  $(0.01) < \alpha(0.05)$ .

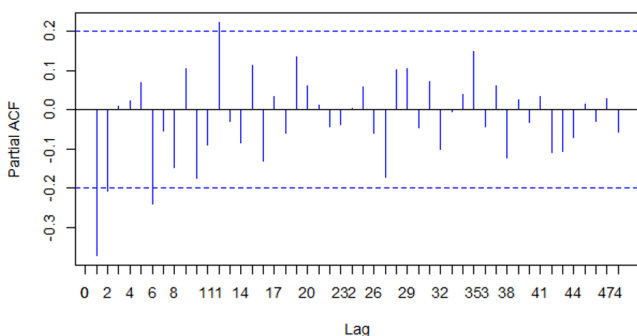
ally, in the seasonal model, the significant pattern at lag 12 onwards indicates stationarity in the data, hence there is no need to differentiate for seasonality. The tentative models obtain are SARIMA(2,1,1)(1,0,1)<sup>12</sup>, SARIMA(0,1,1)(1,0,1)<sup>12</sup>, SARIMA(2,1,0)(1,0,1)<sup>12</sup>, SARIMA(2,1,1)(0,0,1)<sup>12</sup>, and SARIMA(2,1,1)(1,0,0)<sup>12</sup>.

The tentative models formed are estimated parameters on several possible models as in Table 3. The best parameter estimates are shown in the p-value of the model coefficient  $< \alpha(0.05)$  or significant.

In the first model, significant parameters are only found in  $\phi_p(B)$  and  $\theta_q(B)$  components, with the AIC 1888.312, and the second model with the highest AIC shown the significant parameter only in  $\theta_q(B)$ , especially in the third model are only significant in  $\phi_p(B)$ ,  $\theta_q(B)$  and  $\Phi_P(B^s)$ . Overall, the five tentative models resulted in all parameter values being insignificant to the model. However, the third model has the smallest AIC value. Asymptotically, AIC is calculated by minimizing the squared error of the prediction or estimate [19]. For this reason, the SARIMA(2,1,1)(1,0,0)<sup>12</sup> model was selected for further analysis to a diagnostic test of the residuals. Diagnostic tests are important to ensure that the residuals are white noise, which is explained by the assumptions of normality of the distribution and non-autocorrelation of the distribution.



(a)



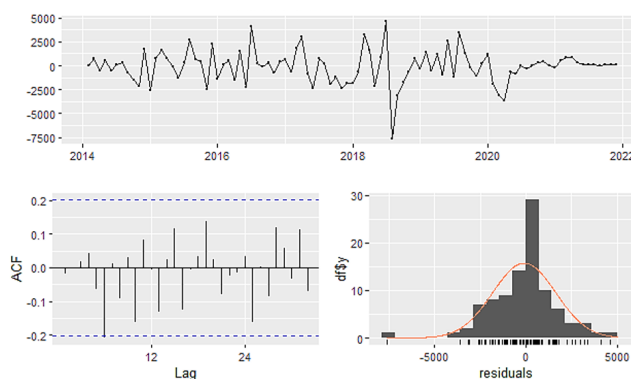
(b)

**Figure 5.** (a) ACF and (b) PACF Plot of differencing results

By differencing once, the ACF plot (Figure 5a) shows significance at lag 1, indicating that the series has a strong correlation at that lag, while the PACF plot (Figure 5b) shows significance at lag 2 in the non-seasonal model. Addition-

**Table 4.** Residuals assumption test

Jarque-Bera Test	Ljung-Box Test (Residuals <sup>2</sup> )
2.2e-16	0.0001359



**Figure 6.** Diagnostic model

**Table 5.** Overfitting model

Model	Parameters	Coefficient	P-Value	Residuals	AIC
SARIMA (0,1,2)(0,0,1) <sup>12</sup>	MA(1)	-1.3987	2.2e-16	No white noise	1886.114
	MA(2)	0.3987	4.311e-05		
	SMA(1)	0.2330	0.004458		
SARIMA (3,1,0)(1,0,0) <sup>12</sup>	AR(1)	-1.0996	2.2e-16	White noise	1909.185
	AR(2)	-0.8098	1.726e-12		
	AR(3)	-0.3652	4.300e-05		
	SAR(1)	0.2872	0.001706		

The autocorrelation plot of the second SARIMA model in Figure 6 shows that there are no significant lags, which means that the assumption of non-autocorrelation is met or indicates that there is no autocorrelation in the model residuals. However, the histogram shows that the residuals do not follow a normal distribution, which is confirmed by the Jarque-Bera Test p-value <  $\alpha$  (0.05). In addition, there is a violation of the assumption of homoscedasticity, which means that the residuals are not white noise. The overfitting process to identify the best model is shown in Table 5.

After overfitting, it can be seen that the SARIMA(3,1,0)(1,0,0)<sup>12</sup> model has a higher AIC value than the first model of the overfitting model, but the residuals exhibit white noise characteristics or no violations of assumptions in residual homoscedasticity, residual independence, and residual normality. Based on these tests, the best model for Seasonal ARIMA is SARIMA(3,1,0)(1,0,0)<sup>12</sup> or  $(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3) (1 - \Phi_1 B^{12}) (1 - B) X_t = e_t$  with an RMSE is 1041.9.

### 3.4. Support Vector Machine Model

Input and output variables of SVM using lag 1, the data is shown in Table 6.

**Table 6.** Value of variables

$X_t$	$Y_t$
5105	4862
4862	5987
5987	5413
⋮	⋮
4958	7748

The output variable ( $Y_t$ ) is the value 1 month after, so each data totals 121 observations. In SVM model, the parameters are selected through a tuning process. This process aims to identify the optimal parameters that will improve the performance of the model on the training data. The initial stage in performing parameter tuning is to determine the range of parameter values of the linear, RBF, and Sigmoid kernels namely  $c, \gamma, \epsilon$ . The range of parameter values used in this study can be seen in Table 7.

**Table 7.** Value of parameters

Parameters	Range Values
$c$ (cost)	$2^2, 2^3, \dots, 2^9$
$\gamma$ (gamma)	$2, 2^2, 2^3$
$\epsilon$ (epsilon)	0.01, 0.1, 0.0005

Each combination of parameters is tested and evaluated to determine which configuration gives the best results. The results of the optimal parameter tuning process are presented in Table 8, which are the best parameters found to be most effective in optimizing SVM model accuracy and predictive power on the trained data.

**Table 8.** Best SVM parameters

Gamma	Cost	Epsilon
2	4	0.01

After obtaining the optimal SVM parameters, the next step was to test these parameters with several SVM kernels: linear, RBF (Radial Basis Function), and sigmoid, to determine the best model configuration. Each kernel was tested on the optimal parameter combination with  $\gamma = 2, c = 4$ , and  $\epsilon = 0.01$ . The performance of each kernel is evaluated with the smallest RMSE value. It is considered the most suitable kernel to minimize RMSE in the SVM model so that it can provide appropriate prediction results.

**Table 9.** SVM model selection

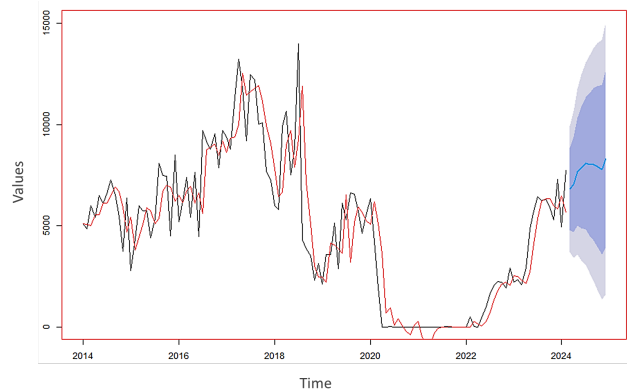
SVM Kernel	RMSE
Linear	238.655
Radial	1023.35
Sigmoid	129052

The smallest error value measured by RMSE is found in the linear kernel, which has an RMSE of 238.655. This shows that the SVM model using the linear kernel is the best among the tested SVM kernels, next the radial basis function (RBF) kernel is higher than the linear kernel and lastly the sigmoid kernel has the highest RMSE among the three, thus making it the least effective SVM model among the three. This comparison shows that the linear kernel in the SVM model can be compared with the best SARIMA model.

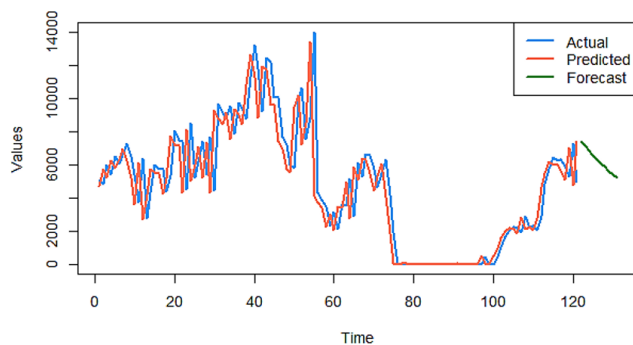
### 3.5. Comparison of SARIMA and SVM Model

Based on the test results using training data with the SARIMA and SVM models, the RMSE value of SVM with a linear kernel is smaller than SARIMA, so it is said that the SVM model is the best in forecasting the number of foreign tourist arrivals through the Lombok arrival gate. The both prediction and forecasting results of the model are shown in Figure 7.

In the SARIMA model (Figure 7a), the black line shows the actual data, while the red line shows the model predictions. The pattern shows that the prediction closely matches the actual data, although it does not fully capture the entire pattern. In con-



(a)



(b)

**Figure 7.** Actual and predicted data of : (a) SARIMA and (b) SVM Models

trast, the SVM model (Figure 7b) produces a prediction pattern that follows the actual data more accurately, this is shown in the smaller error value. The forecasting results of the two models for the next 10 months, the SARIMA model continue the actual data pattern very closely, which indicates consistent forecasting performance. While in the SVM model, the forecasting results show a downward trend every period and do not follow the data pattern. Comparative analysis of forecasting on both models shows that the SARIMA model can provide forecasting results that are closer to historical data patterns.

#### 4. Conclusion

The SARIMA model forecasting results can follow the actual data pattern which means SARIMA is better at forecasting the number of foreign tourists, especially through the Lombok International Airport entrance, and the SVM model can capture nonlinear patterns in the data as indicated by the smallest RMSE. To improve forecasting accuracy, we can propose an intervention process in the SARIMA model which is known to occur in the COVID-19 pandemic in that time range, or can perform a hybrid process so that the model can capture linear and nonlinear patterns in the data.

**Author Contributions.** Hadyanti Utami MY: Software, methodology, formal analysis, data curation, writing—original draft preparation, writing—review and editing, visualization. Silfiana Lis Setyowati: Writing—original draft preparation, writing—review and editing. Khairil Anwar Notodiputro: Conceptualization, methodology, su-

pervision. Yenni Angraini: Conceptualization, methodology, supervision. Laily Nissa Atul Mualifah: Conceptualization, validation, supervision. All authors have read and agreed to the published version of the manuscript.

**Acknowledgement.** Thanks to the lecturers and assistants of the time series analysis course of the Department of Statistics and Data Science, IPB University who have guided and provided advice and solutions when there were obstacles in the process of preparing this article. The authors express their gratitude to the editor and reviewers for their meticulous reading, insightful critiques, and practical recommendations, all of which have greatly enhanced the quality of this work.

**Funding.** This research received no external funding.

**Conflict of interest.** The authors declare no conflict of interest related to this article.

#### References

- [1] E. Pujiati, D. Yuniarti, and R. Goejantoro, "Peramalan Dengan Menggunakan Metode Double Exponential Smoothing Dari Brown (Studi Kasus : Indeks Harga Konsumen (IHK) Kota Samarinda)," *J. Eksponensial*, vol. 7, no. 1, 2016. [Online]. Available: <http://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/23>.
- [2] V. P. Ariyanti and Tristyanti Yusnitasari, "Comparison of ARIMA and SARIMA for Forecasting Crude Oil Prices," *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 7, no. 2, pp. 405–413, Mar. 2023, doi: [10.29207/resti.v7i2.4895](https://doi.org/10.29207/resti.v7i2.4895)
- [3] V. N. Aziza, F. H. Moh'd, F. A. Maghfiroh, K. A. Notodiputro, and Y. Angraini, "Performance Comparison of Sarima Intervention and Prophet Models for Forecasting the Number of Airline Passenger At Soekarno-Hatta International Airport," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 4, pp. 2107–2120, 2023, doi:[10.30598/barekengvol17iss4pp2107-2120](https://doi.org/10.30598/barekengvol17iss4pp2107-2120).
- [4] B. G. Prianda and E. Widodo, "Perbandingan Metode Seasonal Arima Dan Extreme Learning Machine Pada Peramalan Jumlah Wisatawan Mancanegara Ke Bali," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 15, no. 4, pp. 639–650, 2021, doi: [10.30598/barekengvol15iss4pp639-650](https://doi.org/10.30598/barekengvol15iss4pp639-650).
- [5] A. S. Nugroho, A. B. Witarto, and D. Handoko, "Support Vector Machine: Teori dan Aplikasinya dalam Bioinformatika," in *Proceeding of Indonesian Scientific Meeting in Central Japan, Gifu - Japan*, 2003.
- [6] N. H. Riyantoni, M. F. Bahreisy, I. Hakim, and D. Rolliawati, "Komparasi Support Vector Machine (Svm) Dan Autoregressive Integrated Moving Average (Arima) Pada Peramalan Hujan Di Daerah Albury, Australia," *J. Sist. Inf. dan Inform.*, vol. 6, no. 1, pp. 59–68, 2023, doi: [10.47080/simika.v6i1.2412](https://doi.org/10.47080/simika.v6i1.2412).
- [7] L. Budiarti, T. Tarno, and B. Warsito, "Analisis Intervensi dan Deteksi Outlier pada Data Wisatawan Domestik (Studi Kasus di Daerah Istimewa Yogyakarta)," *J. Gaussian*, vol. 2, no. 1, pp. 39–48, 2013, doi: [10.14710/j.gauss.2.1.39-48](https://doi.org/10.14710/j.gauss.2.1.39-48).
- [8] G. H. Saputra, A. H. Wigena, and B. Sartono, "Penggunaan Support Vector Regression dalam Pemodelan Indeks Saham Syariah Indonesia dengan Algoritme Grid Search," *Indonesian J. Stat. App.*, vol. 3, no. 2, pp. 148–160, 2019, doi: [10.29244/ijsa.v3i2.172](https://doi.org/10.29244/ijsa.v3i2.172).
- [9] W. A. Woodward, B. P. Sadler, and S. D. Robertson, *Time Series for Data Science: Analysis and Forecasting*, 1st ed. New York: CRC Press, 2022, doi: [10.1201/9781003089070](https://doi.org/10.1201/9781003089070).
- [10] S. Zahara and S. Sugianto, "Peramalan Data Indeks Harga Konsumen Berbasis Time Series," *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 5, pp. 24–30, 2021, doi: [10.29207/resti.v5i1.2562](https://doi.org/10.29207/resti.v5i1.2562).
- [11] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th Ed. Canada: John Wiley & Sons, 2016.
- [12] S. Markidakis, S. C. Wheelwright, and V. E. McGee, *Metode dan Aplikasi Peramalan Jilid 1*, (Terjemahan Ir. Hari Suminto), Edisi Kedua. Jakarta: Binarupa Aksara, 1999.
- [13] N. P. P. Pratama and T. Sukmono, "Forecasting the Amount of Blood Storage Using the Support Vector Machine (Svm) Method," in *Procedia of Engineering and Life Science (SENASAINS 5th)*, Sidoarjo, 2022, doi: [10.21070/pels.v3i0.1359](https://doi.org/10.21070/pels.v3i0.1359).
- [14] L. Bi, O. Tsimhoni, and Y. Liu, "Using the support vector regression approach to model human performance," *IEEE Trans. Syst. Man, Cy-*



- bern. Part A Systems Humans, vol. 41, no. 3, pp. 410–417, 2011, doi: [10.1109/TSMCA.2010.2078501](https://doi.org/10.1109/TSMCA.2010.2078501).
- [15] R. E. Cahyono, J. P. Sugiono, and S. Tjandra, “Analisis Kinerja Metode Support Vector Regression (SVR) dalam Memprediksi Indeks Harga Konsumen,” *JTIM: J. Teknologi Informasi dan Multimedia*, vol. 1, no. 2, pp. 106–116, 2019, doi: [10.35746/jtim.v1i2.22](https://doi.org/10.35746/jtim.v1i2.22).
- [16] D. I. Purnama and O. P. Hendarsin, “Peramalan Jumlah Penumpang Berangkat Melalui Transportasi Udara di Sulawesi Tengah Menggunakan Support Vector Regression (SVR),” *Jambura J. Math.*, vol. 2, no. 2, pp. 49–59, 2020, doi: [10.34312/jjom.v2i2.4458](https://doi.org/10.34312/jjom.v2i2.4458).
- [17] S. Annas, Z. Rais, A. Aswi, I. Indrayasaro, and N. Nurfajriani, *Implementation of Support Vector Regression (SVR) Analysis in Predicting Gold Prices in Indonesia*, vol. 2023, no. Icsmt. Atlantis Press International BV, 2023. doi: [10.2991/978-94-6463-332-0\\_12](https://doi.org/10.2991/978-94-6463-332-0_12).
- [18] M. B. Pamungkas and A. Wibowo, “Aplikasi Metode Arima Box-Jenkins Untuk Meramalkan Kasus DBD Di Provinsi Jawa Timur,” *Indones. J. Public Heal.*, vol. 12, no. 2, pp. 181–194, 2017, doi: [10.20473/ijph.v1i2.2018.181-194](https://doi.org/10.20473/ijph.v1i2.2018.181-194).
- [19] K. R. A. Muslihin and B. N. Ruchjana, “Model Autoregressive Moving Average (ARMA) untuk Peramalan Tingkat Inflasi di Indonesia,” *Limits J. Math. Its Appl.*, vol. 20, no. 2, p. 209, 2023, doi: [10.12962/limits.v20i2.15098](https://doi.org/10.12962/limits.v20i2.15098).