

## FORECASTING ARRIVAL OF FOREIGN TOURISTS USING SEASONAL ARIMA BOX-JENKINS

Nur Ilmayasinta \*

<sup>1</sup> Prodi Pendidikan Matematika, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Islam Lamongan  
Jl. Veteran, No. 53 A, Lamongan, 62213, Indonesia

Corresponding author e-mail: \* [nurilma@unisla.ac.id](mailto:nurilma@unisla.ac.id)

### Abstract

Indonesia's economy is influenced by many factors, including the tourism sector. Through this tourism sector, it is possible for many foreign tourists to visit Indonesia. There are so many foreign tourists who come to Indonesia, forecasting is needed to find out the estimates of foreign tourists in the following months based on existing data. The method that used is the Seasonal Autoregressive Integrated Moving Average (SARIMA) method. The foreign tourist's coming to Indonesia through Soekarno Hatta Airport were taken from the center agency on statistics (BPS) Indonesia. Data on the number of foreign tourists who come to Indonesia through Soekarno Hatta Airport is data with a seasonal pattern. The data used is secondary data obtained from Soekarno Hatta Airport for the period January 2010 to June 2015. In this case it is used to predict the value of the data for the next 6 months using the best model is the ARIMA(0,1,1) (0,1,1)<sup>12</sup>. Forecasting results show the number of each month increases from the previous year. In July it showed the highest yield of 342536, which was 297878 in the previous year. Forecasting results show the number of each month increases from the previous year. In July, the highest yield was 342536, which was 297878 in the previous year.

**Keywords:** Forecasting, ARIMA, Seasonal ARIMA, BOX-JENKINS

### Article info:

Submitted: 09<sup>th</sup> November 2020

Accepted: 09<sup>th</sup> April 2021

### How to cite this article:

N. Ilmayasinta, "PERAMALAN KEDATANGAN WISATAWAN ASING MENGGUNAKAN SEASONAL ARIMA BOX-JENKINS", *BAREKENG: J. Il. Mat. & Ter.*, vol. 15 no. 02, pp. 223-230, Jun. 2021.



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).  
Copyright © 2021 Nur Ilmayasinta

## 1. INTRODUCTION

Indonesia is a rich country. The potential of natural resources is extraordinary, both natural and non-biological natural resources. Indonesia is also the second highest biodiversity country after Brazil. The diversity of flora, fauna and its ecosystem, as well as cultural diversity, are potential attractions for the development of ecotourism in the country. Ecotourism and nature tourism are recognized as being particularly conducive to enriching and enhancing the tourism sector, on the basis that forms of tourism respect the natural heritage and local inhabitants and are in accordance with the carrying capacity of the site.

Forecasting is an activity to predict what will happen in the future. There have been many studies on forecasting, with the ARIMA method and Seasonal ARIMA method, such as: selection of the best models and forecasting the number of visits of foreign tourism (wisman) to Bali in 2014[1]. Model Box-Jenkins in forecasting regional gross domestic products, Bali Province [2]. Use of ARIMA in method predicting inflation movements [3]. Crank-nicolson method for calculating value of stock loan with dividend reinvested before redemption [4]. ARIMA by Box Jenkins Methodology for Estimation and Forecasting Models in Higher Education [5]. Forecasting the Number of Aircraft Passengers at Sultan Iskandar Muda Airport Using the SARIMA Method (Seasonal Autoregressive Integrated Moving Average) [6]. Predictive modeling of pelagic fish catch in malaysia using seasonal ARIMA models [7]. Influence of seasonal factors on time series modeling for river debit forecasting with the SARIMA method [8] and other research on forecasting using the SARIMA method [9]-[13].

In this paper we use data of foreign tourists coming to Indonesia according to the entrance (Soekarno Hatta), because Soekarno Hatta Airport is an airport that is included in the ranks of Indonesian airports with the most visitors because it is located in the capital city. In this study the data used were taken from Central Bureau of Statistics. The Seasonal Autoregressive Integrated Moving Average (SARIMA) method is used for predict the number of foreign tourists coming to Indonesia according to the entrance (Soekarno Hatta), data on numbers passengers at Soekarno Hatta airport show seasonal patterns so SARIMA method is used to predict the number of passengers in the future.

## 2. RESEARCH METHODS

### 2.1 Literature Review

ARIMA is often also called the Box-Jenkins time series method. ARIMA is very good accuracy for short-term forecasting, while for long-term forecasting the accuracy of forecasting is not good. Usually it will tend to be flat (horizontal/constant) for a fairly long period. The Integrated Moving Average Autoregressive Model (ARIMA) is a model that completely ignores independent variables in making forecasting. ARIMA uses past and present values of the dependent variable to produce accurate short-term forecasting. ARIMA is suitable if observations from time series are statistically related to each other (dependent). One of the most popular and frequently used stochastic time series is the Autoregressive Integrated Moving Average (ARIMA) model [14]. The ARIMA model consists of three basic steps, namely the identification phase, the assessment and testing stage, and the diagnostic inspection. Furthermore, the ARIMA model can be used to forecast if the model obtained is adequate

### 2.2 Stationary and Non-stationary

The thing to note is that most periodic series are non-stationary and that the AR and MA aspects of the ARIMA model only concern stationary series. Stationary is a state where mean and variance is constant. Stationarity means there is no growth or decrease in the data. The data must be roughly horizontal along the time axis. In other words, data fluctuations are around a constant average value, independent of time and the variance of these fluctuations remains essentially constant over time. A time series that is not stationary must be converted into stationary data by performing differencing. What is meant by differencing is to count changes or differences in the value of observations. The value of the difference obtained is checked again whether it is stationary or not. If it is not stationary, then do differencing again. If the variance is not stationary, then the logarithmic transformation is performed.

### 2.3 Classification of ARIMA models

The ARIMA model is divided into 3 elements, namely: The Autoregressive (AR) model, Moving Average (MA), and Integrated (I). To determine the best model can be seen based on the Autocorrelation and Partial Autocorrelation plots [15].

**Table 1. ACF and PACF structures in the ARIMA model**

Model	Pola ACF	Pola PACF
AR(p)	Decreases exponentially	Drastically decreases at certain lags
MA(q)	Decreases drastically at a particular lag	Decreasing exponentially
ARMA(p,q)	Decreases exponentially	Decreasing exponentially

These three elements can be modified to form a new model. for example, the autoregressive and moving average (ARMA) models. However, if you want to make it in general form, it becomes ARIMA (p,d,q). p denotes the AR order, d denotes the Integrated order and q represents the moving average order. If the model becomes AR then the model will generally be ARIMA (1,0,0).

### 2.4 Seasonal Autoregressive Integrated Moving Average (SARIMA) Method

The general form of the seasonal moving average process for the period  $S$  level  $Q$  or  $MA(Q)^S$  is defined as follows:

$$X_t = e_t - \Theta_1 e_{t-S} - \Theta_2 e_{t-2S} - \dots - \Theta_Q e_{t-QS} \quad (1)$$

where  $q$  is mutually independent of  $X_{t-1}, X_{t-2}, \dots$  which is normally distributed with mean 0 and variant  $\sigma^2$ . The general form of the seasonal Autoregressive process of the period  $S$  the level of  $P$  or  $AR(P)^S$  is defined as

$$X_t = \Phi_1 X_{t-S} + \Phi_2 X_{t-2S} + \dots + \Phi_P X_{t-PS} + e_t \quad (2)$$

where  $t$  is mutually independent of  $X_{t-1}, X_{t-2}, \dots$  which is normally distributed with mean 0 and variance  $\sigma^2$ . The seasonal data is defined as a pattern that repeats in a fixed time interval. For stationary data, the seasonal factor can be determined by identifying the autocorrelation coefficients at two or three time-lags that are significantly different from zero. Autocorrelations that are significantly different from zero represent a pattern in the data. To recognize seasonal factors, one must look at high autocorrelation. To deal with seasonal, brief general notations are:

$$ARIMA(p, d, q) (P, Q, S)^S \quad (3)$$

With  $(p, d, q)$  = non-seasonal parts of the model

$(P, Q, S)$  = seasonal parts of the model

$S$  = number of periods per season

## 3. RESULTS AND DISCUSSION

Data on the number of foreign tourists coming to Indonesia from 2010-2015 is as follows:

**Table 2. Data of foreign tourists coming to Indonesia**

Month	2010	2011	2012	2013	2014	2015
January	174541	179273	209093	253286	232935	279257
February	147704	191926	207195	225993	241868	276573
March	168205	192579	207907	230957	252210	276573
April	188776	184907	224704	225488	242369	280096
May	190803	203388	209058	220700	247972	286033
June	200566	228045	245652	244080	275667	329654

Month	2010	2011	2012	2013	2014	2015
July	235198	254907	283524	271512	297878	
August	232255	243154	258377	254079	309219	
September	218443	240947	258440	257363	305629	
October	221282	229904	247565	255021	266562	
November	184803	199861	221603	242781	307276	
December	222546	227251	253591	268072	299013	

From the data above we get the time series plot as follows:

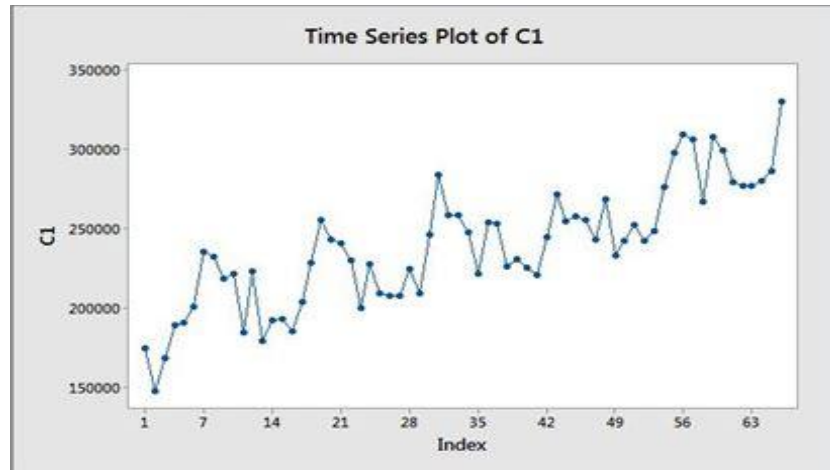


Figure 1. Timeline diagram for initial data

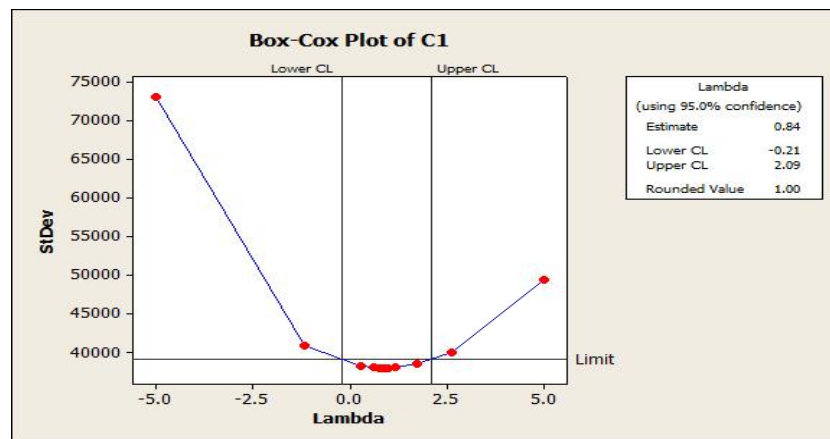


Figure 2. Box-Cox preliminary data

Figure 1 shows that the data on the arrival of foreign tourists to Indonesia through the entrance of Soekarno-Hatta Airport has fluctuated or fluctuated from January 2010 to June 2015. Figure 2 shows that the rounded value on the Box-Cox plot is 1. Lambda value the value is equal to 1 with a confidence interval of 95%. The rounded value of 1 indicates that the data on the arrival of foreign tourists to Indonesia through the entrance to Soekarno-Hatta Airport is stationary in variance.

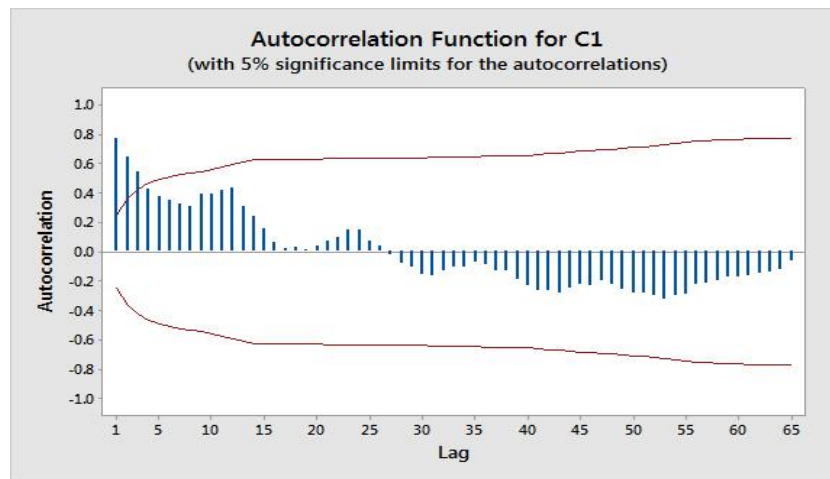


Figure 3. Initial data autocorrelation (ACF) function

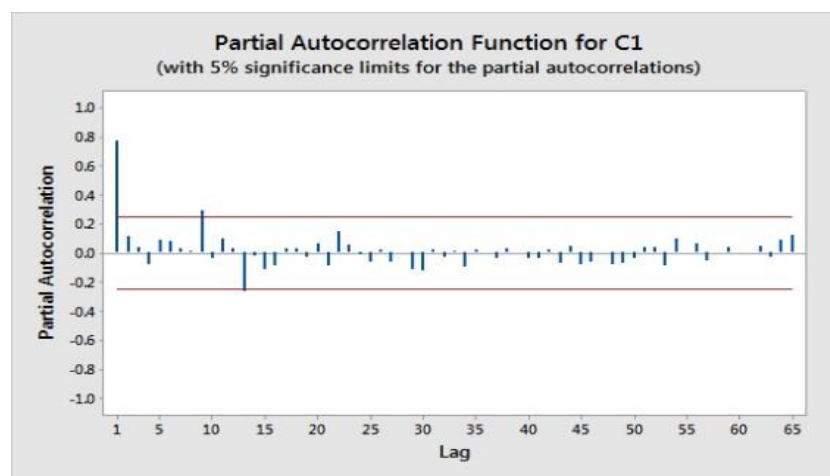


Figure 4. The initial Partial Autocorrelation (PACF) data function

Based on Figure 2, it can be seen that the value of  $\lambda$  is 1 so that the data is stationary for variance. Whereas based on ACF plots and PACF plots it appears that the data is not stationary to the mean. The figure above shows that the value of the autocorrelation function tends to slow down which the value of autocorrelation in a lag is relatively not much different from the previous lag. The value of the partial autocorrelation (PACF) function is truncated after the initial lag. So that this also indicates that the data is not stationary in average. So to get stationary data, the method is to difference the data. The stages of identifying the model for differencing data are as follows:

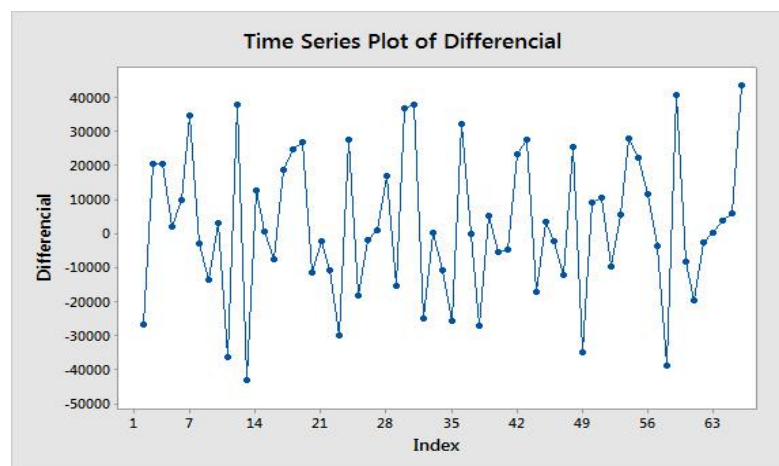


Figure 5. Time series diagram for data after differencing 1 time

The one-time differencing plot above indicates that the data is stationary, so we can use the data to form the seasonal ARIMA model. As for determining the seasonal ARIMA model that is by looking at the Autocorrelation and Partial Autocorrelation functions as follows:

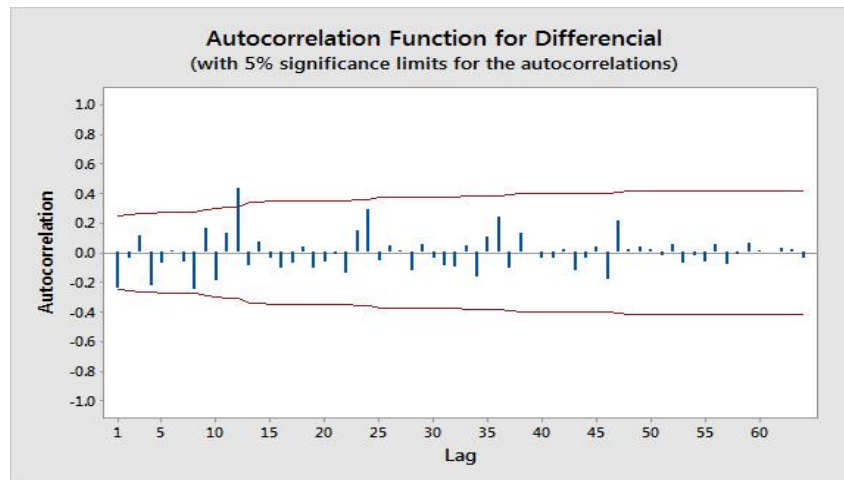


Figure 6. Autocorrelation (ACF) data function after differencing 1 time

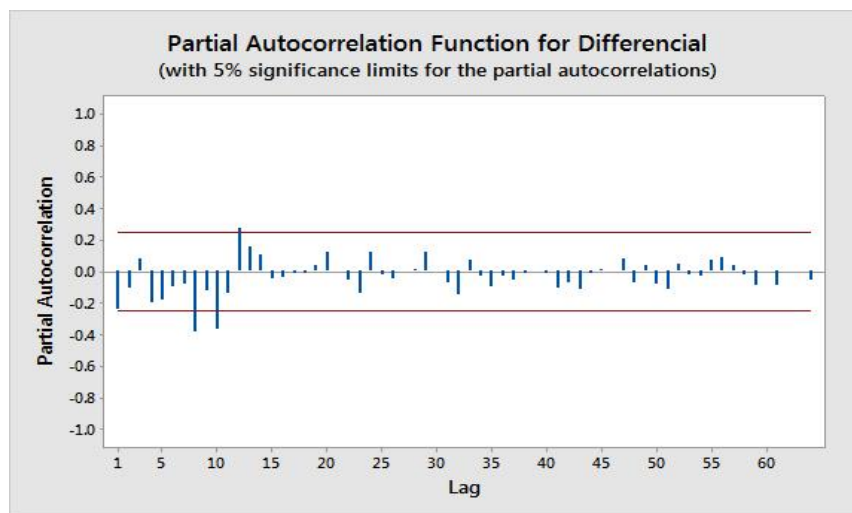


Figure 7. Partial Autocorrelation (PACF) data function after differencing 1 time

Based on the Autocorrelation diagram (ACF), it can be seen that the autocorrelation value is cut off after lag 1 and dies down on PACF so that the first suspected model  $MA(1)$  is  $ARIMA(0,1,1)$ . Whereas, for the second model,  $AR(1)$  or  $ARIMA(1,1,0)$ , which in PACF, the partial autocorrelation value is significant at lag 1. So, the initial estimation of the model that is suitable for the data is  $ARIMA(0,1,1)$  or  $ARIMA(1,1,0)$ , then we choose the best model. From Figure 5 and Figure 6 we can get a seasonal model so that the alleged model is  $(0,1,1)$  with a period of 12. So that the alleged model for seasonality is  $A(0,1,1)(0,1,1)^{12}$ ,  $ARIMA(1,1,0)(0,1,1)^{12}$ .

### 3.1. Test parameters for the $ARIMA(0,1,1)(0,1,1)^{12}$

Case Hypothesis:

$$H_0: \hat{\theta} = 0, H_1: \hat{\theta} \neq 0.$$

With a value  $P(0.002) < \alpha(0.05)$ ,  $H_0$  is rejected so we conclude that the  $ARIMA(0,1,1)(0,1,1)^{12}$  model is significantly different from zero.

White noise requirements:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0 \text{ (the remainder meets the white noise requirement)}$$

$$H_1: \text{There is at least one } \rho_j \neq 0, j = 0, 1, 2, \dots, K \text{ (the rest does not meet the white noise requirements).}$$

The P values for lag 12, lag 24, lag 36 and lag 48 are greater than  $\alpha = 0.05$  so we accept  $H_0$  and we conclude that the remainder has met the white noise requirements.

### 3.2. Test parameters for the $ARIMA(1, 1, 0)(0, 1, 1)^{12}$

Case Hypothesis:

$$H_0: \hat{\theta} = 0, H_1: \hat{\theta} \neq 0$$

With a value of  $P(0,000) < \alpha (0.05)$ ,  $H_0$  is rejected so we conclude that the  $ARIMA(1,1,0)(0,1,1)^{12}$  model is significantly different from zero. White noise requirements:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0 \text{ (the remainder meets the white noise requirement)}$$

$H_1$ : There is at least one  $\rho_j \neq 0, j = 0, 1, 2, \dots, K$  (the rest does not meet the white noise requirements).

The P values for *ag* 12, *lag* 24, *lag* 36 and *lag* 48 are greater than  $\alpha = 0.05$  so we accept  $H_0$  and we conclude that the remainder has met the white noise requirements.

### 3.3. Selection of the Best Model

Based on testing the conditions above, it can be explained that the alleged  $ARIMA(0,1,1)(0,1,1)^{12}$  and  $ARIMA(1,1,0)(0,1,1)^{12}$  is appropriate[16]. The next step is to compare the MSE (Mean Square Error) values of the two models. The MSE value for the  $ARIMA(0,1,1)(0,1,1)^{12}$  is 289435993, while the MSE value for the  $ARIMA(1,1,0)(0,1,1)^{12}$  is 401278632. So based on the smallest MSE value, the  $ARIMA(0,1,1)(0,1,1)^{12}$  model is the best model for predicting data values in the next few stages.

### 3.4. Forecasting for the next 6 months

Data values for the next 6 months for the  $ARIMA(0,1,1)(0,1,1)^{12}$ , obtained:

**Table 3. Data of forecasts from period 66**

Period	Forecast	Lower	Upper
67	342536	318756	366316
68	331139	306566	355712
69	329603	304261	354944
70	319896	293810	345983
71	307273	280462	334085
72	330710	303193	358227

Based on the forecasting data above, it can be seen that the number of foreign tourists who come to Indonesia through Soekarno Hatta airport in July to December 2015 will increase when compared to 2014 in months the same one. The highest number of foreign tourists coming to Indonesia through Soekarno Hatta airport is predicted to occur in July 2015, there were 342536 people.

## 4. CONCLUSION

From the results of the above analysis it can be concluded that the best seasonal ARIMA model is  $ARIMA(0,1,1)(0,1,1)^{12}$ . From the ARIMA model the forecast results are obtained for July, August, September, October, November and December are 342536, 331139, 329603, 319896, 307273 and 330710. Forecasting the number of foreign tourists coming to Indonesia through Soekarno Hatta airport for the next 6 months generally increases compared to the previous year in the same month. Only in November 2015 the number of passengers is predicted to experience a slight decrease compared to November 2014. The highest number of passengers is predicted to occur in July 2015 are 342536 people, which in the previous year was 297878.

## REFERENCES

- [1] I. W. Sukadana and L. G. Meydianawathi, "Pemilihan model terbaik dan peramalan jumlah kunjungan wisatawan mancanegara (wisman) ke Bali Tahun 2014," vol. 20, no. 1, pp. 66–75, 2015.
- [2] I. G. Putu and N. Wirawan, "Model box- jenkins dalam rangka peramalan produk domestik regional bruto provinsi Bali," vol. 19, no. 1, pp. 92–104, 2014.
- [3] Hartati, "Penggunaan metode arima dalam meramal pergerakan inflasi," 2017.
- [4] N. Ilmayasinta, "Metode crank-nicolson untuk menghitung nilai stock loan dengan dividen yang diinvestasikan kembali sebelum pembayaran pinjaman," 2016.

- [5] M. A. Din, "ARIMA by Box Jenkins Methodology for Estimation and Forecasting Models in Higher Education Marilena Aura Din," no. January 2015, 2016.
- [6] F. I. Durrah, T. P. Parhusip, and A. Rusyana, "Peramalan Jumlah Penumpang Pesawat Di Bandara Sultan Iskandar Muda Dengan Metode SARIMA ( Seasonal Autoregressive Integrated Moving Average )," vol. 1, no. 1, pp. 1–11, 2018.
- [7] H. Y. Bako, M. S. Rusiman, I. L. Kane, and H. M. Matias-peralta, "Predictive modeling of pelagic fish catch in malaysia using seasonal ARIMA models," no. November 2016, 2013.
- [8] D. Ruhiat and A. Effendi, "Pengaruh faktor musiman pada pemodelan deret waktu untuk peramalan debit sungai dengan metode sarima," vol. 2, no. 2, pp. 117–128, 2018.
- [9] U. Hanum and D. Murni, "Peramalan Jumlah Pengunjung Objek Wisata Waterboom Kota Sawahlunto Tahun 2019 Menggunakan Metode SARIMA," vol. 2, no. 3, pp. 86–91, 2019.
- [10] F. I. Durrah, Y. Yulia, T. P. Parhusip, and A. Rusyana, "Peramalan Jumlah Penumpang Pesawat Di Bandara Sultan Iskandar Muda Dengan Metode SARIMA (Seasonal Autoregressive Integrated Moving Average)," *J. Data Anal.*, vol. 1, no. 1, pp. 1–11, 2018, doi: 10.24815/jda.v1i1.11847.
- [11] Z. Kafara, F. Y. Rumlawang, and L. J. Sinay, "Peramalan Curah Hujan Dengan Pendekatan Seasonal Autoregressive Integrated Moving Average (Sarima)," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 11, no. 1, pp. 63–74, 2017, doi: 10.30598/barekengvoll1iss1pp63-74.
- [12] E. Zukhronah, S. S. Handajani, U. S. Maret, U. S. Maret, and U. S. Maret, "165 peramalan jumlah barang melalui transportasi kereta api di indonesia menggunakan sarima," pp. 165–172, 2021.
- [13] N. Lestari, Nofinda; Wahyuningsih, "Peramalan Kunjungan Wisata dengan Pendekketan Model SARIMA (Studi kasus : Kusuma Agrowisata)," *J. Sains dan Seni ITS*, vol. 1, no. Vol 1, No 1 (2012): *Jurnal Sains dan Seni ITS (ISSN 2301-928X)*, pp. A29–A33, 2012, [Online]. Available: [http://ejournal.its.ac.id/index.php/sains\\_seni/article/view/1010/675](http://ejournal.its.ac.id/index.php/sains_seni/article/view/1010/675)
- [14] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis, Forecasting and Control*. 2016.
- [15] W. W. S. Wei, "Time Series Analysis : Univariate and Multivariate Methods." 2006.
- [16] J. Sullivan, "A comparison of fuzzy forecasting and Markov modeling," vol. 64, pp. 279–293, 1994.