ENTHUSIASTIC
INTERNATIONAL JOURNAL OF STATISTICS AND DATA SCIENCE
Volume 2, Issue 1, April 2022, pp. 19-25

Forecasting International Tourist Arrivals in Indonesia Using SARIMA Model

Deden Nurhasanah [a,1], Aurielle Maulidya Salsabila [a,2], Mujiati Dwi Kartikasari [a,3,*]

[a] Department of Statistics, Universitas Islam Indonesia, Jalan Kaliurang Km 14.5 Yogyakarta, Indonesia
[1] 19611143@students.uii.ac.id; [2] 19611164@students.uii.ac.id; [3] mujiatikartikasari@uii.ac.id*
* corresponding author

ARTICLE INFO

Article history
Received February 19, 2022
Revised March 20, 2022
Accepted April 27, 2022

Keywords
Tourism
International Tourist
Time Series
Forecasting
SARIMA

ABSTRACT

Tourism is an important sector that significantly contributes to the economy, so the tourism sector is a priority development program. International tourist arrivals indirectly contribute to the country's economic growth. The government has an important task to increase the number of foreign tourist visits. One way to encourage an increase in foreign tourist arrivals is by forecasting. In general, the time series data for the arrival of foreign tourists has a seasonal pattern. The forecasting method that can model seasonal data is SARIMA. This study aims to predict the arrival of foreign tourists in Indonesia using the SARIMA model. Forecasting results show that the appearance of foreign tourists to Indonesia has increased every period.

1. Introduction

Tourism is a variety of travel activities that contain various facilities and services provided by the community, business people, government, and local governments [1] [2] [3]. Tourism is one of the essential industries to increase economic growth. The tourism industry can encourage economic growth. This industry is considered an invisible export industry because of its ability to earn foreign exchange by exporting other commodities. As one of the largest archipelagic countries globally, Indonesia has an excellent opportunity to develop the tourism industry with its potential and natural wealth. The tourism sector is one of the dominant sectors in influencing economic growth. Indonesia's tourism sector has contributed about 4% of the total economy [4].

People who carry out tourism activities are called tourists [1] [2] [3]. Tourists who come from outside Indonesia are called international tourists. In detail, foreign tourists are every visitor who comes to a country outside their place of residence, based on one or more needs without the intention of earning an income, and the duration of their visit is not more than 12 (twelve) months. Increasing the number of foreign tourist visits is one of the essential tasks of the government [5]. It has a linear relationship between the number of foreign tourist arrivals and foreign exchange earnings. With the increase in foreign tourists, foreign exchange earnings also increase.

One way to increase the number of foreign tourist visits is to do forecasting. Previously aware of these conditions, the government could prepare policies regarding the arrival of foreign tourists. The forecasting technique to predict the number of foreign tourist arrivals is the time series method. The

p-ISSN 2798-253X
e-ISSN 2798-3153

time series method is a forecasting technique with time series data and periodic data collection [6]. There are four kinds of time series data patterns, namely horizontal, seasonal, cyclical, and trend [6].

The collection of data on foreign tourist arrivals is carried out periodically, and in general, data on arrivals of foreign tourists will form a seasonal pattern. The method that can model various seasonal data is the SARIMA (Seasonal Autoregressive Integrated Moving Average) model [6]. Many researchers have researched forecasting using SARIMA. Silalahi (2020) [7] forecasted data on poverty using the SARIMA modeling in West Java Province, Andoh et al. (2021) [8] conducted a forecasting analysis of the electricity demand in Ghana using the SARIMA model, Berhane et al. (2018) [9] conducted rainfall modeling and forecasting in Ethiopia using SARIMA, and Urrutia et al. (2017) [10] conducted research in the field of agriculture by testing the quarterly rice and maize production forecasting in the Philippines using the SARIMA model. Based on this description, this study aims to predict the arrival of foreign tourists in Indonesia using SARIMA.

2. Method

There are several stages in this research. Here are the stages.

1. Creating data plots.
2. Differencing seasonal components.
3. Perform unit root test to detect data stationarity. If the data is not stationary, differencing is performed on the non-seasonal component.
4. Identify models from ACF and PACF plots.
5. Estimating model parameters.
6. Choose the best model based on MAPE.
7. Perform diagnostic tests.
8. Doing forecasting with the best model.

These stages are summarized in the flowchart presented in Figure 1.

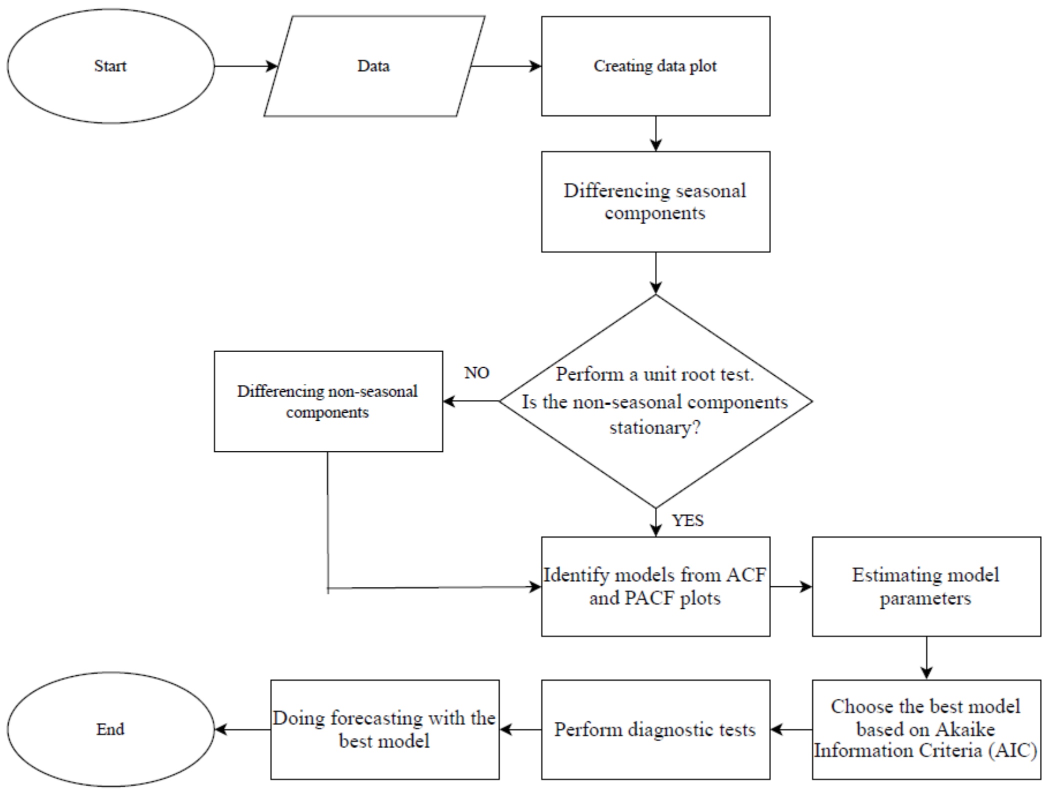

Fig. 1. Research flowchart.

 p-ISSN 2798-253X
e-ISSN 2798-3153

2.1. Seasonal Autoregressive Integrated Moving Average

Seasonal Autoregressive Integrated Moving Average (SARIMA) is a forecasting technique that models various seasonal data [6]. The SARIMA model includes additional seasonal terms in the ARIMA model. The notation of SARIMA model is

$$ARIMA(p,d,q)(P,D,Q)_s$$

where, p is order of the autoregressive part in non-seasonal component, d is degree of first differencing involved in non-seasonal component, q is order of the moving average part in non-seasonal component, P is order of the autoregressive part in seasonal component, D is degree of first differencing involved in seasonal component, Q is order of the moving average part in seasonal component, and s is the number of observations in a year or seasonal period.

General formulation of $ARIMA(p,d,q)(P,D,Q)_s$ is describe as follows [6]

$$\phi_p(\mathbf{B})\Phi_P(B)^s(1-B)^d(1-B^s)^D y_t=\theta_q(B)\Theta_Q(B^{\boldsymbol{s}})e_t \tag{1}$$

where,

$$\phi_P(B)=\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_2 B^p\right), \tag{2}$$

$$\theta_P(B)=\left(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_2 B^q\right), \tag{3}$$

$$\Phi_P(B^s)=\left(1-\Phi_1 B-\Phi_2 B^{2s}-\cdots-\Phi_2 B^{Ps}\right), \tag{4}$$

$$\Theta_P(B^s)=\left(1-\Theta_1 B-\Theta_2 B^2-\cdots-\Theta_2 B^{Qs}\right). \tag{5}$$

The notation of B is the backshift operator which is a notation when working with time series lags, ϕ_p is parameter of autoregressive of order p in non-seasonal component, θ_q is parameter of moving average of order p in non-seasonal component, Φ_P is parameter of autoregressive of order P in seasonal component, Θ_Q is parameter of moving average of order Q in seasonal component, e_t is a white noise process with zero mean and constant variance, $t=1,2,\cdots,n$, and n is the number of observations.

In determining the values of p and q as well as P and Q, it can be helped by observing the pattern of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) with reference to the Table 1.

Table 1. Order Identification of ACF and PACF Plots

Order	Description	Pattern of ACF and PACF Plots
AR(p)	An autoregressive model of order p	There is a significant spike in lag p on the PACF plot (if the line slowly decays to 0 exponentially, then AR(0)).
MA(q)	A moving average model of order q	There is a significant spike in lag q on the ACF plot.
SAR(P)	A seasonal autoregressive model of order P	There is a significant spike in lag Ps on the PACF plot, with s being a seasonal period.
SMA(Q)	A seasonal moving average model of order Q	There is a significant spike in lag Qs on the PACF plot, with s being a seasonal period.

2.2. Model Evaluation

To find out the forecasting model is good, it is necessary to evaluate the forecasting model. Model evaluation can be calculated from the resulting error. One measure of error that can be used to evaluate the forecasting model is the Mean Absolute Percentage Error (MAPE). The MAPE formulation is as follows [6],

$$MAPE=\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t-\hat{y}_t}{y_t}\right|\times 100\%. \tag{6}$$

p-ISSN 2798-253X
e-ISSN 2798-3153

3. Results and Discussion

The data for this research is data on the number of foreign tourist visits to Indonesia from January 2015 to December 2019 based on the Central Statistics Agency web page [11]. The data plot is shown in Figure 2.

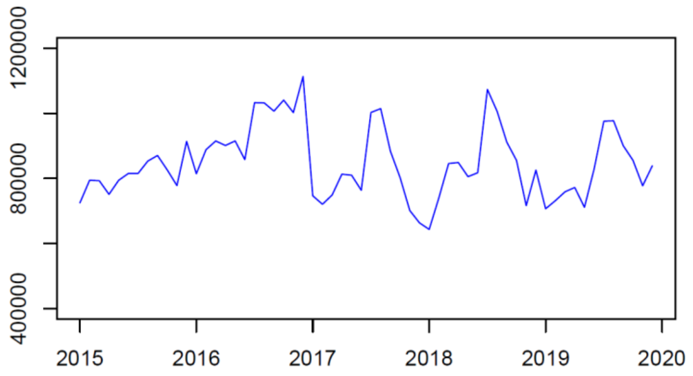

Fig. 2. Plot of the data.

Figure 2 shows the highest number of foreign tourist arrivals occurred in the middle of the year, and the lowest number of foreign tourist arrivals occurred at the beginning and end of the year. The data shows a seasonal component. To ensure the data has a seasonal pattern, there are repeated fluctuations in a certain period, by drawing the ACF and PACF plots shown in Figure 3.

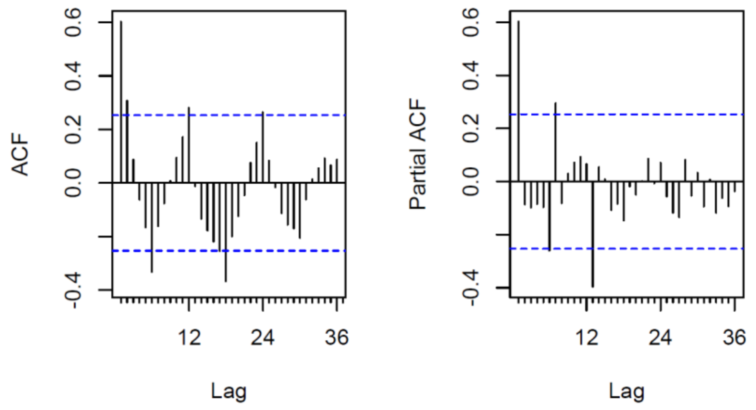

Fig. 3. ACF and PACF plot of the data.

The ACF pattern in Figure 3 shows that the data has a seasonal component. The blue dotted line indicates the correlation is significantly different from zero. The value of the autocorrelation coefficient at lags 12 and 24 is quite large which explains that the data on foreign tourist arrivals in Indonesia forms a seasonal pattern of s, 2s, 3s, and so on, with a value of s = 12, which is a seasonal period in the data.

Since the international tourist arrival data has a seasonal component, it takes seasonal differences. The seasonal difference data in Figure 4 appears to be non-stationary. In addition, based on the unit root test, it shows that the data is not stationary. Therefore, the researcher takes the additional first difference and then shows its ACF and PACF plots in Figure 5.

p-ISSN 2798-253X
e-ISSN 2798-3153

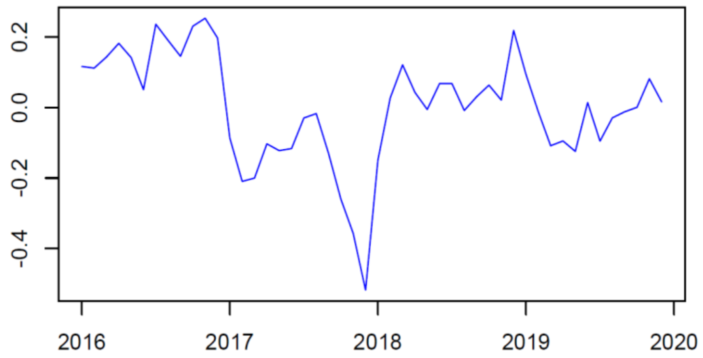

Fig. 4. Seasonality differnced of data.

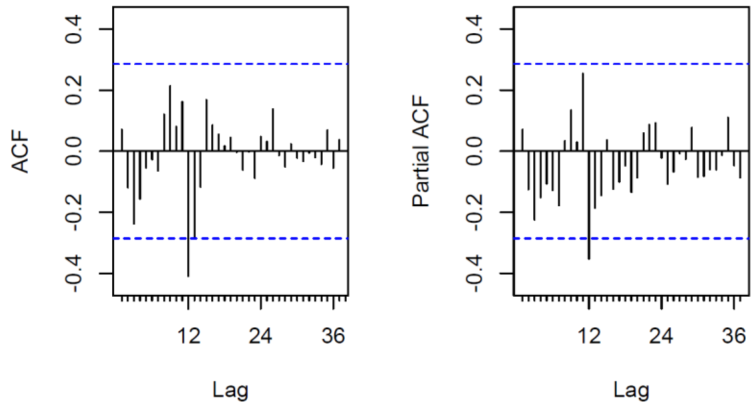

Fig. 5. ACF and PACF plot of double differenced data.

The appropriate ARIMA model is based on the ACF and PACF plots shown in Figure 5. Both ACF and PACF showed significant spikes at lag 12, indicating seasonal SMA(1) and SAR(1) components. Consequently, it starts with the ARIMA$(0,1,0)(1,1,1)_{12}$ model. Next, make several combinations of ARIMA$(0,1,0)(1,1,1)_{12}$ models. The estimation results from several models are summarized in Table 2.

Table 2. Model Estimation

Model	**Coefficient**	**Summary**		
		Estimates	***P-value***	***AIC***
ARIMA$(0,1,0)(1,1,1)_{12}$	SAR(1)	0.0535	0.7786	-82.51
	SMA(1)	-0.9997	0.1478	
ARIMA$(0,1,0)(0,1,1)_{12}$	SMA(1)	-0.924	0.4491	-84.43
ARIMA$(0,1,0)(1,1,0)_{12}$	SAR(1)	-0.4855	4e-04	-81.33

The results of Table 2 only ARIMA$(0,1,0)(1,1,0)_{12}$ which has a statistically significant model coefficient at the probability significance level = 5%. In addition, based on the results of the diagnostic test in Figure 6, the residual in ARIMA$(0,1,0)(1,1,0)_{12}$ is white noise. Researchers expect each autocorrelation close to zero. in Figure 5, there are no spikes beyond the blue dotted line above, and substantially more than 5% of spikes are outside this limit, so the series is white noise. Therefore, the ARIMA$(0,1,0)(1,1,0)_{12}$ model is feasible to use to predict the arrival of foreign tourists in Indonesia. The ARIMA ARIMA$(0,1,0)(1,1,0)_{12}$ model expression is

$$(1+0.4855B^{12})(1-B)(1-B^{12})y_t=(1-B)e_t.$$

p-ISSN 2798-253X
e-ISSN 2798-3153

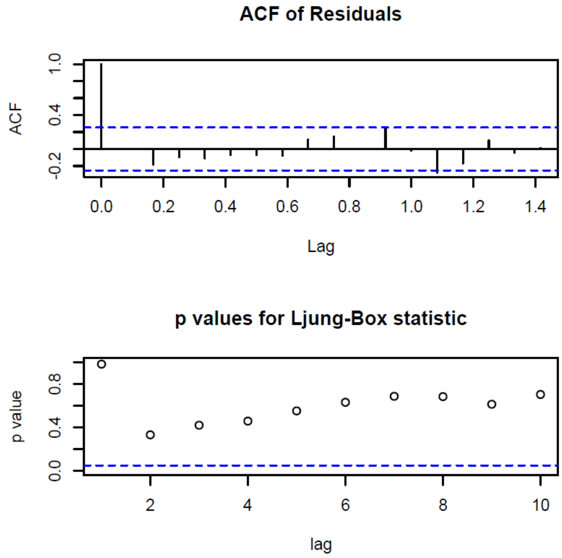

Fig. 6. Diagnostic test of ARIMA(0,1,0)(1,1,0)$_{12}$ model.

The ARIMA(0,1,0)(1,1,0)$_{12}$ is the best forecasting model, Researchers evaluate the model by comparing the actual data with the predicted data in Figure 7.

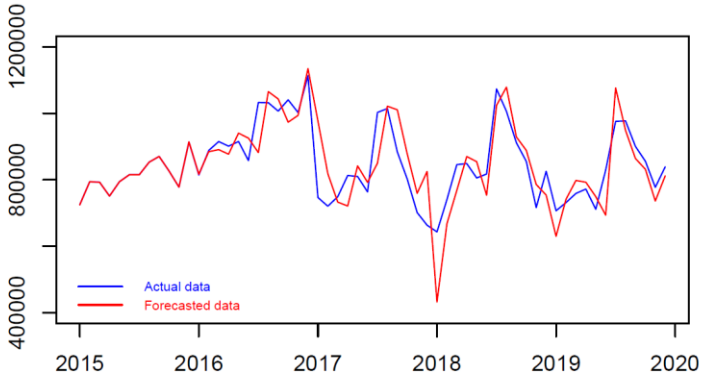

Fig. 7. Forecasted data of international tourist arrivals in Indonesia.

Figure 7 shows that the predicted data is closer to the actual data. Visually, the researcher concludes that the model produces good predictions. Accurately, this is indicated by the MAPE value. MAPE prediction result of ARIMA(0,1,0)(1,1,0)$_{12}$ model is 6.00%. according to [12] this value is very accurate. Forecasting on the ARIMA(0,1,0)(1,1,0)$_{12}$ model is very good for predicting the arrival of foreign tourists in Indonesia. Researchers predict for the next 7 periods to see the number of tourists in the future. Table 3 shows the predicted results of foreign tourists arrivals increasing in each period.

Table 3. Forecasted Values of International Tourist Arrivals in Indonesia

Period	Forecasted Value
January 2020	770125.6
February 2020	831680.3
March 2020	898485.2
April 2020	906571.0
May 2020	854650.5
June 2020	918791.0
July 2020	1120961.0

p-ISSN 2798-253X
e-ISSN 2798-3153

4. Conclusion

Based on the results of the analysis, the researcher concludes that ARIMA$(0,1,0)(1,1,0)_{12}$ is the best model to predict the arrival of foreign tourists in Indonesia. ARIMA$(0,1,0)(1,1,0)_{12}$ has a statistically significant model coefficient and the forecasting results show that the arrival of foreign tourists has increased in each period.

References

[1] Indonesia, "Undang-Undang Republik Indonesia Nomor 10 Tahun 2009 Tentang Kepariwisataan," Indonesia, Jakarta, 2009.

[2] Indonesia, "Peraturan Pemerintah Republik Indonesia Nomor 67 Tahun 1996 Tentang Penyelenggaraan Kepariwisataan," Indonesia, Jakarta, 1996.

[3] Indonesia, "Undang-Undang Nomor 9 Tahun 1990 Tentang Kepariwisataan," Indonesia, Jakarta, 1990.

[4] Indonesia Investments, "Industri Pariwisata Indonesia," [Online]. Available: https://www.indonesia-investments.com/id/bisnis/industri-sektor/pariwisata/item6051. [Accessed 20 12 2021].

[5] BPS, "Istilah," Badan Pusat Statistik, [Online]. Available: https://www.bps.go.id/istilah/index.html?Istilah_page=68&Istilah_sort=keyword_ind. [Accessed 20 12 2021].

[6] R. J. Hyndman and G. Athanasopoulos, Forecasting: principles and practice, 2nd edition, Melbourne, Australia: OTexts, 2018.

[7] D. K. Silalahi, "Forecasting of Poverty Data Using Seasonal ARIMA Modeling in West Java Province," *Jurnal Teori dan Aplikasi Matematika,* vol. 4, no. 1, pp. 76-86, 2020.

[8] P. Y. A. Andoh, C. K. K. Sekyere, L. D. Mensah and D. E. K. Dzebre, "Forecasting Electricity Demand In Ghana With The Sarima Model," *Journal of Applied Engineering and Technological Science,* vol. 3, no. 1, 2021.

[9] T. Berhane, N. Shibabaw, G. Awgichew and T. Kebede, "Modeling and Forecasting Rainfall in Ethiopia," *International Journal of Computing Science and Applied Mathematics*, vol. 4, no. 2, pp. 42-46, 2018.

[10] J. D. Urrutia, J. L. B. Diaz and F. L. T. Mingo, "Forecasting the Quarterly Production of Rice and Corn in the Philippines: A Time Series Analysis," in *Journal of Physics: Conference Series, Volume 820, 5th International Conference on Science & Engineering in Mathematics, Chemistry and Physics 2017 (ScieTech 2017),* Bali, Indonesia, 2017.

[11] BPS, "Jumlah Kunjungan Wisatawan Mancanegara per bulan ke Indonesia Menurut Pintu Masuk 2017-2020," Badan Pusat Statistik, [Online]. Available: https://www.bps.go.id/indicator/16/1150/3/jumlah-kunjungan-wisatawan-mancanegara-per-bulan-ke-indonesia-menurut-pintu-masuk-2017---sekarang.html. [Accessed 27 12 2021].

[12] J. J. M. Moreno, A. P. Pol, A. S. Abad and B. C. Blasco, "Using the R-MAPE Index as a Resistant Measure of Forecast Accuracy," *Psicothema,* vol. 25, no. 4, pp. 500-506, 2013.

p-ISSN 2798-253X
e-ISSN 2798-3153