Strengthening economy through tourism sector by tourist arrival prediction

A Supriatna¹, B Subartini¹, E Hertini¹, Sukono¹, Rumaisha¹, N Istiqamah²

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, INDONESIA

*Corresponding author: agus.supriatna@unpad.ac.id

Abstract. Tourism sector has a tendency to be proposed as a support for national economy to many countries with various of natural resources, such as Indonesia. The number of tourist is very related with the success rate of a tourist attraction, since it is also related with planning and strategy. Hence, it is important to predict the climate of tourism in Indonesia, especially the number of domestic or international tourist in the future. This study uses Seasonal Autoregressive Integrated Moving Average (SARIMA) time series method to predict the number of tourist arrival to tourism strategic areas in Nusa Tenggara Barat. The prediction was done using the international and domestic tourist arrival to Nusa Tenggara Barat data from January 2008 to June 2016. The established SARIMA method was with MAPE error of 15.76. The prediction for the next six time periods showed that the highest number of tourist arrival is during September 2016 with 330,516 tourist arrivals. Prediction of tourist arrival hopefully might be used as reference for local and national government to make policies to strengthen national economy for a long period of time

1. Introduction

Indonesia, as the largest archipelago in the world is known as the second biggest mega biodiversity in the world for its natural resources, flora and fauna which are very highly diverse. Indonesia is located in the equator, which gives Indonesia the richness of tropical forests and diversity of local cultures. These potencies shows that Indonesia possesses a major resource and capital for the development of businesses and enhancing the tourism in Indonesia ([4]).

Lately, the tourism sectors become the pole of economic growth for countries with abundance of natural resources, such as Indonesia, Brazil, etc. This sector is affected by the number of domestic and international tourists, providing major income for the country. The number of tourist arrival to Indonesia is increasing annually, especially to potential tourism areas in Indonesia, i.e. Nusa Tenggara Barat (NTB). The increase of the number of tourist arriving to Nusa Tenggara Barat peaked significantly in 2012 (23.75%), hence noting that the year marked the start of a new era for tourism in Nusa Tenggara Barat. Thus, the prediction of tourist arrival to potential tourism areas in Nusa Tenggara Barat is needed.

Prediction of tourist arrival became an important element in decision making for tourism policies, especially in fulfilling product demands in tourism ([1]). The accuracy of prediction is important for the sake of planning efficiency ([6]), even though prediction shows results very close to facts in real life ([2]). Prediction could be divided into two types based on the data used in predicting; i.e.
Qualitative and quantitative prediction. Quantitative prediction uses quantified data from the past, whereas qualitative predictive uses expert opinion/judgment through their previous experience ([7]).

Quantitative method consists of time series and causal. Time series is a data collected based on time period (t). Time series method includes naïve, moving average, exponential smoothing and ARIMA (Autoregressive Integrated Moving Average). Among those methods, ARIMA is the best mode because ARIMA does not require assumption about the current structural association or economic model ([1]).

This study applied the seasonal ARIMA or the commonly known as SARIMA (Seasonal Autoregressive Integrated Moving Average) to international tourist arrival data to Nusa Tenggara Barat. The arrival data of year 2008 to June 2016 was obtained from Department of Culture and Tourism Nusa Tenggara Barat. The SARIMA model has been used as an alternative solution for forecasting problems, e.g. by [8] on Forecasting Tourism Arrivals in South Africa, in [9] the comparative analysis of Holt Winter and SARIMA for statistical prediction of foreign tourist arrival to Kraton Yogyakarta is discussed.

Based on the background explained above, the problems could be formulated as follows:

1. How many tourists’ arrivals to Nusa Tenggara Barat for the next six time periods?
2. What are the benefits for applying prediction methods in Nusa Tenggara Barat area?

This study aims to predict the number of tourist arrival to Nusa Tenggara Barat for the next six time periods. The result of study is hoped to be applied in policy making in tourism sector for the government in Indonesia, especially Nusa Tenggara Barat.

2. Research Methodology

This study used a four-step SARIMA method, consisting of model identification, parameter estimation, diagnostic tests and forecasting ([7]). These steps are presented in flowchart of SARIMA Modeling in Figure 1. The first step to be done was checking the data stationary. Transformation/differencing process would be done if the data were not stationary in terms of variance and mean. After the data was stationary, the next step would be model identification by looking at the Autocorrelation Function (ACF) plot and Partial Autocorrelation Function (PACF) do determine the possible temporary SARIMA model. The obtained model would be estimated for parameters and significance test would be done for each parameter. The insignificant model would be erased from the model. After all parameters had significant models, diagnostic test would be done to check whether the error was normally distributed and was not auto correlated. Kolmogorov-Smirnov test would be used to check the normality of the error; meanwhile, the test of autocorrelation for the error was L-Jung Box test. Should the model has not been yet obtained even after diagnostic tests, return to model identification step to look for another model. If all model assumptions had been fulfilled and if there were more than one model that could possibly be selected, the model with the least error was selected for forecasting. MAPE error was calculated then a conclusion was drawn.

Seasonal Autoregressive Integrated Moving Average (SARIMA) has a common form of \((p, d, q) \times (P, D, Q)^s\) and was written as follows:

\[
\phi_p(B)\Phi_P(B^S) (1-B^d)(1-B^s)^D X_t = \theta_q(B)\Theta_Q(B^S)e_t
\]

\(e_t\): error

\(X_t\): Observed value at time of \(t (t = 1,2,...,n)\)

\((1-B)^d\): Mathematical operation of non-seasonal differencing

\((1-B^s)^D\): Mathematical operation of seasonal differencing

\(\phi_p(B)\): Autoregressive Operator = \((1-\phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)\)

\(\Phi_P(B^S)\): Seasonal Autoregressive Operator = \((1-\Phi_1 B^S - \Phi_2 B^{2S} - ... - \Phi_P B^{PS})\)

\(\theta_q(B)\): Seasonal Moving Average Operator = \((1+\theta_1 B + \theta_2 B^2 + ... + \theta_q B^q)\)
\[ \Theta_q(B^s) : \text{Seasonal Moving Average Operator} = (1 + \Theta_1 B^s + \Theta_2 B^{2s} + \ldots + \Theta_q B^{qs}) \] 

\[ ([5]) \]

3. Results And Discussions

SARIMA process consisted of four steps, i.e. model identification, parameter estimation, diagnostic test and forecasting.

During the first step, data plotting to determine the data characteristics and test of stationary assumption in time series analysis was done. Time series data plot was presented in Figure 2.

Figure 1. Flowchart of SARIMA Modeling

Figure 2. Tourist Arrival to NTB Year 2008-2016 Data Plot (Reference: http://www.disbudpar.ntbprov.go.id/)
Data plot in Figure 2 showed that there was an increasing trend and seasonality pattern, determining that it was a seasonal ARIMA (SARIMA) model. Stationary assumption in terms of variance and mean was needed for SARIMA model. Test of stationary in variance could be conducted using Box-Cox transformation ([1]), with results shown in Figure 3.

![Box-Cox Transformation](image)

**Figure 3.** Box-Cox Transformation

Figure 3 showed that the acquired lambda was -0.2, thus a logarithmic transformation was done. After the transformation, an Augmented Dickey Fuller (ADF) test was done to check data stationary in terms of mean. ADF test showed that p-value = 0.2676 > 0.05 (Appendix 1), thus, it could be concluded that the data was not stationary in terms of mean and a first differencing process on non-seasonal order should be done.

![Non-seasonal order first differencing result](image)

**Figure 4.** Non-seasonal order first differencing result

Figure 4 was a non-seasonal order first differencing result which showed that the data was stationary in terms of mean and to conform that it was stationary, the data could be re-tested with ADF test. The ADF test showed that the p-value = 0.01 < 0.05 (Appendix 2), meaning that the data was stationary.

**Model Identification**

After stationary was achieved, the ACF and PACF plot was checked to determine the temporary SARIMA model.

![ACF and PACF Plot](image)

**Figure 5.** ACF and PACF Plot

Figure 5 showed that the ACF plot was cut on lag 1, lag 12 and lag 24, meanwhile the PACF plot showed a sinus wave; hence, the possible SARIMA model was SARIMA $(0,1,1)(0,0,1)^{12}$ and SARIMA $(0,1,1)(0,0,2)^{12}$. 


**Parameter Estimation**

The parameter of the obtained models would be estimated using MLE method with software R 3.2.3 (Appendix 3), then would be tested for significance using t test. The result of parameter estimation was presented on Table 1.

| Temporary Model | Parameter | S.E  | Count t result | Significance test |
|-----------------|-----------|------|----------------|-------------------|
| SARIMA (0,1,1)(0,0,1)\textsuperscript{12} | $\theta_1 = -0.2083$ | 0.1086 | 1.918 | Significant |
| | $\Theta_1 = 0.3613$ | 0.1371 | 2.635 | Significant |
| SARIMA (0,1,1)(0,0,2)\textsuperscript{12} | $\theta_1 = -0.2720$ | 0.1168 | 2.328 | Significant |
| | $\Theta_1 = 0.3872$ | 0.1558 | 2.485 | Significant |
| | $\Theta_2 = 0.3656$ | 0.1845 | 1.981 | Significant |

Table 1 showed that both models were significant since \(t_{\text{count}} > t_{\text{table}} = 1.65\). The next step would be conducting diagnostic test on the models.

**Diagnostic Test**

Another assumption that must be fulfilled from SARIMA method was that the errors on the models were normally distributed and were not auto correlated. Test of normality on the error was conducted using Kolmogorov-Smirnov test and test for autocorrelation using L-jung Box test. The result of the calculation using software R 3.2.3 (Appendix 4) was shown on Table 2.

| Model | P-value of Kolmogorov-Smirnov Test | P-value of L-Jung Box Test |
|-------|-----------------------------------|---------------------------|
| SARIMA (0,1,1)(0,0,1)\textsuperscript{12} | 0.9459 | 0.06828 |
| SARIMA (0,1,1)(0,0,2)\textsuperscript{12} | 0.9332 | 0.09853 |

Kolmogorov-Smirnov test showed the p-value for SARIMA (0,1,1)(0,0,1)\textsuperscript{12} was 0.9459 and the p-value for SARIMA (0,1,1)(0,0,2)\textsuperscript{12} was 0.9332; thus, it could be concluded that the \(H_0\) (normally distributed error) was failed to be rejected since the p-value > 0.05 and that both models have normally distributed errors. Meanwhile, using L-Jung Box test, it was known that the p-value for SARIMA (0,1,1)(0,0,1)\textsuperscript{12} was 0.06828 and the p-value for SARIMA (0,1,1)(0,0,2)\textsuperscript{12} was 0.09853, hence, it could be concluded that the \(H_0\) (error was not auto correlated) was failed to be rejected since the p-value > 0.05 and that both models have non-auto correlated errors.

**Forecasting**

Appropriate models were then evaluated by reviewing the error.

| Model | AIC  | RMSE  | MAPE  |
|-------|------|-------|-------|
| SARIMA (0,1,1)(0,0,1)\textsuperscript{12} | 2264.78 | 57496.97 | 21.46042 |
| SARIMA (0,1,1)(0,0,2)\textsuperscript{12} | 2262.67 | 49293.35 | 18.95509 |
Based on error calculation on both models, the SARIMA model \((0,1,1)(0,0,2)^{12}\) was selected as the model with the least error and was mathematically written as:

\[
(1 - 0.2720B)(1 - B)X_t = (1 - 0.3872B^{12} - 0.3656B^{24})e_t
\]

\[
X_t = X_{t-1} + 0.2720X_{t-1} - 0.2720X_{t-2} + e_t - 0.3872e_{t-12} - 0.3656e_{t-24}
\]

Thus, the prediction result of tourist arrival was as shown below.

![Arrival Prediction Result](image)

**Figure 6. Arrival Prediction Result**

| Time  | Prediction of Arrival |
|-------|-----------------------|
| Jul-16| 299.917               |
| Aug-16| 293.695               |
| Sep-16| 330.516               |
| Oct-16| 221.119               |
| Nov-16| 210.195               |
| Dec-16| 223.870               |

Table 4 showed that the predicted tourist arrival to NTB (Appendix 5) peaked on September 2016 and hit the lowest on November 2016.

**4. Conclusion**

The selected SARIMA model in this study was SARIMA \((0,1,1)(0,0,2)^{12}\). This model showed that the predicted tourist arrival to NTB during July 2016 was 299,917 tourist arrivals, August 2016 was 293,695 tourist arrivals, September 2016 was 330,516 tourist arrivals, October 2016 was 221,119 tourist arrivals, November 2016 was 210,195 tourist arrivals, December 2016 was 223,870 tourist arrivals with MAPE error of 15.76. This prediction of tourist arrival was hoped to be applied as a reference for local and national government in making more comprehensive and efficient policies in tourism sector, hence, strengthening the national economy as a long term effect.

**References**

[1] Bigovic M 2012 Demand forecasting within montenegrin tourism using box-jenkins methodology for seasonal ARIMA models *J. Tourism and Hospitality Management* **18** 1-18.

[2] Clements M P and Hendry D F 2004 *A companion to economic forecasting* Blackwell Publishing Oxford.

[3] Dinas Budaya dan Pariwisata Provinsi NTB 2016 *Data kunjungan wisatawan ke NTB tahun*
2008 - Juni 2016 (http://www.disbudpar.ntbprov.go.id/).

[4] Fandeli C and Muhammad 2009 Pengembangan ekowisata berbasis konservasi di taman nasional Yogyakarta: UGM Press.

[5] Ibrahim N et al. 2016 The fitting of SARIMA model on peads patients coming at outpatients medical laboratory (OPML) Mayo hospital, Lahore, (https://www.researchgate.net/publication/299458058, accesed 5 September 2016)

[6] Li G et al. 2006 Forecasting tourism demand using econometric models”, in Buhalis, D. and Costa, C.(Ed.), Tourism management dynamics: trends, management and tools, Butterworth-Heinemann, Oxford 219-220.

[7] Makridakis et al 1999 Metode dan aplikasi peramalan, jilid 1 Jakarta: Binapura Aksara

[8] Saayman A and Saayman M 2010 Forcasting Tourist Arrivals In South Africa Acta commerce 2010 280-293.

[9] Widiarsih I N and Subekti R 2015 Analisis Komparasi Holt Winter Dan Sarima Pada Peramalan Statistik Wisatawan Asing Kraton Yogyakarta Paper Presented at Seminar Nasional Matematika dan Pendidikan Matematika, Universitas Negeri Yogyakarta, Yogyakarta, 2015.