IOP Conf. Series: Materials Science and Engineering **407** (2018) 012148 doi:10.1088/1757-899X/407/1/012148

Forecasting Tourist Visits Using Seasonal Autoregressive Integrated Moving Average Method

R Fahrudin

Magister of Information System, Universitas Komputer Indonesia, Indonesia

rifqifhrdn@gmail.com

Abstract. Bandung, a city in Indonesia is one of the favorite tourist destinations for foreign tourists. The purpose of this writing is predicted to the abundance of foreign tourists who come to Bandung city using the time series methods. Data the data used are foreign tourists entering through the Husein Sastranegara airport from the year 2010 to 2017. This research using Seasonal Autoregressive Integrated Moving Average Method in forecasting foreign tourists who come to the city of Bandung. Model accuracy was measured by comparing the percentage of the value of forecasting with the true value. This value is called the Mean Absolute Deviation (MAD). Based on the results of the comparison, the best value of SARIMA model MAD SARIMA model is the smallest (0, 1, 1) $(1, 0, 0)^{12}$ with the value of the MAD 484,04. From the results it can be concluded that the model in the SARIMA model for forecasting made worth more.

1. Introduction

Bandung is a city in Indonesia which are most visited by tourists for foreign and domestic. The existence of Bandung is still a very strong relationship with the indigenous Sunda is still a reason of attraction for the tourists to visit. Other uniqueness of Bandung which is capable of adding to the attraction for tourists is the presence of the indigenous community of Sunda, but does not eliminate the authenticity of Sunda and the abundance of indigenous sorts of attractions in Bandung such as culinary, shopping and nature. In the school holidays and also new year's, the number of tourist visits in various places of tourism increased more than ordinary days. Data of foreign tourists who come to Bandung Husein Sastranegara airport through a five-year period 2010-2017 rise with a recurring pattern of each period features form the pattern of seasonal, source data obtained from BPS (Badan Statistik Indonesia) [1].

Time series data is one type of data is collected according to the order of time in a particular time span [2]. Research on forecasting has many uses, including Seasonal Holt Winters [3], Exponential Smoothing [4] and Moving Average [5]. Forecasting a time series data need to pay attention to the type or pattern data. Forecasting methods that can be used to predict the time series and data have a seasonal pattern is a SARIMA method [6]. SARIMA method has the data analysis patterns. Method of forecasting best viewed by way of comparing the value of the smallest error. Forecasting error rate obtained on the basis of the results of calculations using the Mean Absolute Deviation (MAD) [7].

The purpose of this research was trying to see the accuracy of forecasting results of SARIMA method in the case of the visit of foreign tourists in Bandung. This method was chosen because it is

considered as a method that can be used as a method of forecasting for traffic data. Moreover this method is a method of forecasting can be done easily and quickly [8-10].

2. Research and Methods

Time series data are data based on a specific interval, such as daily, weekly, monthly, and yearly. Forecasting is an attempt to predict future events to come. Forecasting have two properties that is based on the existing data (quantitative) and based on the opinion of experts (qualitative). Forecasting of time series data are predicted values will come based on past data at certain periods. Forecasting is done by observing the patterns of time series among others [2]; (trend) pattern time series have a tendency of long-term data for direction, (cyclical) time series pattern has waves/fluctuations for more than a year, (seasonal) time series pattern has a periodic fluctuation within a period of one year, (irregular) time series pattern has an unexpected occurrence and is random, but this data can affect the fluctuation of time series data.

2.1. SARIMA Model

Seasonal Method Autoregressive Integrated Moving Average or better known as SARIMA method is Time Series forecasting method for stochastic model data with seasonal data pattern. The general formula of ARIMA (p, d, q) (P, D, Q) s is as follows [6,7,8]:

$$\Phi_{\mathrm{p}}\mathrm{B}^{\mathrm{S}}\phi_{\mathrm{p}}(\mathrm{B})(1-B)^{\mathrm{d}}(1-\mathrm{B}^{\mathrm{S}})^{\mathrm{D}}\mathrm{Z}_{\mathrm{t}} = \theta_{\mathrm{q}}(B)\Theta_{\mathrm{q}}(B^{\mathrm{S}})\mathrm{a}_{\mathrm{t}} \quad (1)$$

With:

${}_{\mathrm{p}}\mathrm{B}$	: AR *Non Seasonal*
$\Phi_{\mathrm{p}}\mathrm{B}^{\mathrm{S}}$	: AR *Seasonal*
$(1-B)^{d}$	: *differencing non seasonal*
$(1-B^{S})^{D}$	: *differencing seasonal*
$\theta_{q}(\mathrm{B})$	: MA *non seasonal*
$\Theta_{q}(\mathrm{B}^{\mathrm{S}})$	: MA *seasonal*

2.2. Measures of accuracy

It is necessary select a particular measure of accuracy in order to examine the accuracy of the forecasting method. Some of the measuring method which generally used are Mean Absolute Deviation (MAD) and Mean Absolute Percentage Error (MAPE) [7]. MAD Is the size of the overall forecasting error for a model. The MAD value is calculated by taking the sum of the absolute values of the forecasting error divided by the number of data periods.

$$MAD = \frac{\Sigma(forecast\ error)^2}{n} \quad (2)$$

Where:
n = number of data periods

3. Result and Discussion

Data on foreign tourists of Bandung city obtained from Badan Statistik Indonesia (BPS). The data in the form of monthly data from January 2010 to December 2017. To be able to analyse the time series of the data it needs original plots in advance so that the next step can be done appropriately. From the data obtained, the obtained graph time series of known data that the number of foreign tourists and the city of Bandung has a seasonal pattern (figure 1-2).

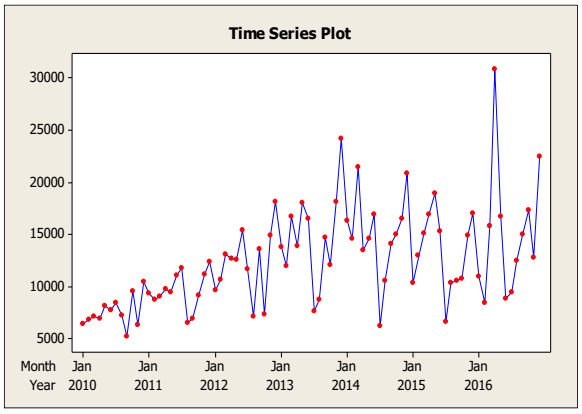

Figure 1. Time Series Plot.

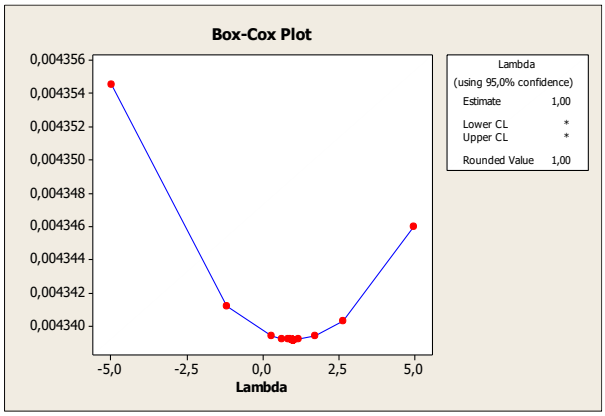

Figure 2. Box-Cox Plot.

From Figure 1 it looks that the data have a seasonal pattern that repeats annually with peaks looping occurs in December. So it can be noted that the pattern of seasonal is following seasonal patterns of 12 monthly. Figure 2 shows the result of data transformation process with Box-Cox Transformation [9]. The data is stationary on the variance because the value of rounded value on the Box-Cox plot is already worth 1.

3.1. Application of SARIMA model

The first step is the identification of process models. The process of identification of the model was first tested whether stationary or not data that is by looking at the plot of ACF and PACF of original data [10]. The steps used to view the pattern data. Figure 1 shows that the data have a repeating pattern every year. So the plot has the data seasonal trend pattern. Further data were analyzed using the methods of time series data, SARIMA first identified the stationary in variant and mean. From the results of the analysis, the plot of the original data of the number of foreign tourists and the city of Bandung has not been stationary in the variants so that the necessary transformation (figure 3-4).

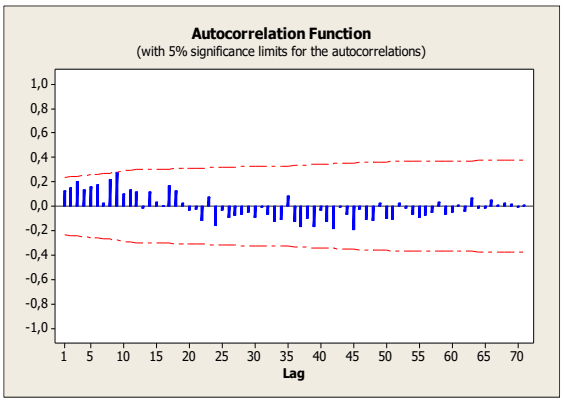

Figure 3. ACF.

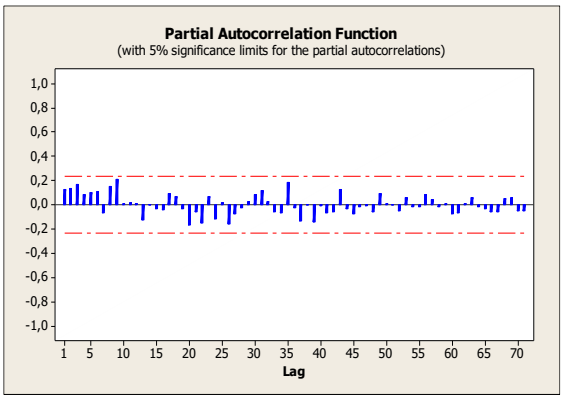

Figure 4. PACF.

Figure 3 and 4 shows the plot of the ACF and PACF, where the data is stationary to the range because there is no lag coming out of the dashed line (confidence interval). Through the ACF and PACF plots can be identified SARIMA model order appropriate for the data number of tourists.

3.2. Parameter Estimation

After time series data is processed has been stationary, then the application of SARIMA model $(p,d,q)(P,D,Q)^{12}$ in the first step and determines the most efficient one. The candidate is shown in table 2.

Table 1. SARIMA model

MODEL	MAD	EXPLANATION
SARIMA $(0,1,1)(1,0,0)^{12}$	484,04	(SAR) and (MA) Significant
SARIMA $(0,1,1)(1,1,1)^{12}$	2849,36	(SAR) and (MA) Significant
SARIMA $(0,1,1)(1,0,1)^{12}$	4514,24	(MA) and (SMA) Significant

From table 1 it can be summed up best SARIMA model for forecasting foreign tourists was SARIMA model with (0, 1, 1) $(1, 0, 0)^{12}$ has a value of 12 for the small MAD 484,04 (table 2).

Table 2. Forecasting 12 months.

Period	Actual Value	Forecast	Absolute Value of Error
January 2017	10098	12988	2890,13
February 2017	13221	10994	2227,26
March 2017	17568	16877	691,08
April 2017	15856	28716	12860,39
Mei 2017	15162	17616	2454,42
June 2017	11978	11477	501,39
July 2017	11479	11954	475,02
Augustus 2017	15196	14383	812,53
September 2017	14025	16378	2352,55
October 2017	13773	18253	4480,33
November 2017	15170	14663	506,60
December 2017	17766	22305	4538,60
Sub Total			5808,47
MAD			484,04

Table 2 shows the forecasting result with SARIMA model $(0,1,1)$ $(1,0,0)^{12}$ for January to December 2017.

The results show that this model can predict foreign tourists with good accuracy, while in other studies [11] that have limitations. For example, a predictable literature but an unacceptably poor accuracy for long-term prediction (figure 5).

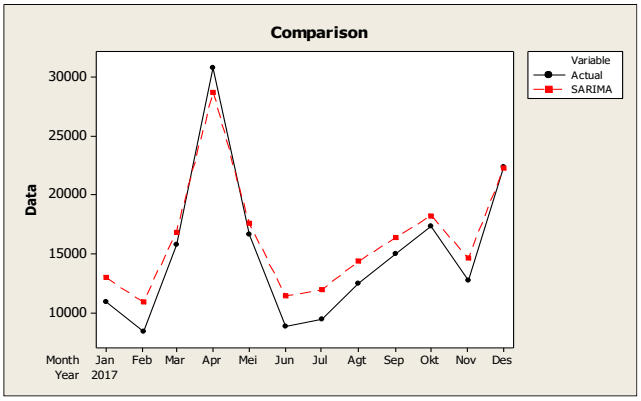

Figure 5. Comparison Data Actual and Forecast.

Figure 5 shows comparison graphs of forecasting data with actual data. The graph uses forecasting with the SARIMA model $(0,1,1)(1,0,0)^{12}$. Comparison shows a pretty good performance seen from its proximity to the actual data.

4. Conclusions

SARIMA model can be used for forecasting the number of foreign tourists with MAD of 484,04. Use a SARIMA model $(0, 1, 1)\ (1, 0, 0)^{12}$ with the smallest error value in the bulletin if the incoming foreign tourism. Based on the results of data analysis, the authors use a SARIMA Method for predicting the number of foreign tourists and Bandung. Furthermore, the reader is expected to try to use other methods such as Holt Winters, Double Seasonal Holt Winters, Exponential Smoothing and Naive model for forecasting. So that the reader can know which method is the best and right to use.

References

[1] Badan Pusat Statistik 2010-2017 *Statistik Pariwisata Kota Bandung* (Kota Bandung: BPS)
[2] Ma E, Liu Y, Li J and Chen S 2016 Anticipating Chinese tourists arrivals in Australia: A time series analysis *Tourism management perspectives* **17** pp.50-58
[3] Sudheer G and Suseelatha A 2015 Short term load forecasting using wavelet transform combined with Holt–Winters and weighted nearest neighbor models *International Journal of Electrical Power & Energy Systems* **64** pp.340-346
[4] Crevits R and Croux C 2016 Forecasting with Robust Exponential Smoothing with Damped Trend and Seasonal Components *Working Paper*
[5] Baldigara T and Mamula M 2015 Modelling international tourism demand using seasonal ARIMA models *Tourism and hospitality management* **21**(1) pp.19-31
[6] Granger C W J and Newbold P 2014 *Forecasting economic time series* Academic Press
[7] Kurian G and Chi H 2017 Predict Florida Tourism Trend via Using Data Mining Techniques *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact* 7 pp. 68-70
[8] Box G E, Jenkins G M, Reinsel G C and Ljung G M 2015 *Time series analysis: forecasting and control* John Wiley & Sons.
[9] Bergmeir C, Hyndman R J and Benítez J M 2016 Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation *International journal of forecasting* **32**(2) pp.303-312
[10] Moffat I U and David A E 2016 Modeling inflation rates in Nigeria: Box-Jenkins' approach *International Journal of Mathematics and Statistics Studies* **4**(2) pp.20-27
[11] Lestari N and Wahyuningsih N 2012 Peramalan Kunjungan Wisata dengan Pendekatan Model Sarima (Studi Kasus: Kusuma Agrowisata) *Jurnal Sains dan Seni ITS* **1**(1) pp.A29-A33