

Analysis Prediction of Prambanan Temple Visitors with *Fuzzy Time Series Chen Model* and *Seasonal Auto Regressive Integrated Moving Average (SARIMA) Model*

Rohmah Artika¹, Adi Setiawan², Glagah Eskacakra Setyowisnu¹, Siti Uminasiah¹, Prihantini¹

¹Mathematics, ²Statistics, Faculty of Mathematics and Natural Science, Yogyakarta State University

Abstract. The number of visitors in tourist attractions are almost always changes each time, even for tourist attractions that are already well-known among local and foreign people, usually will tend to increase at certain times, as in the Prambanan Temple. Based on data from TWC (Taman Wisata Candi) unit office, the number of visitors of Prambanan Temple during holidays at the end of 2018 increased by 8% from the previous year. Because of its increase, the manager of tourist attractions must always try to provide the best service. Therefore, the manager of Prambanan Temple needs to know the prediction of the number of visitors in the future so that they can prepare services and innovations to increase its attractiveness. The data of Prambanan Temple visitors number is seasonal, so the visitors number prediction at Prambanan Temple will be determined using the method for seasonal data. This research tries to compare the two methods, namely Fuzzy Time Series Chen Model and Seasonal Auto Regressive Integrated Moving Average (SARIMA) Model. The results of these methods are the visitors number prediction with different errors, so it can be seen which method is better between the two.

Keywords: Prambanan Temple, Prediction, Fuzzy Time Series, SARIMA

Abbreviations: SARIMA(Seasonal Auto Regressive Integrated Moving Average), MAPE (Mean Absolute Percentage Error)

INTRODUCTION

Prambanan Temple tourism object is a cultural heritage that is managed by PT Borobudur Temple tourism park, Prambanan and Ratu Boko (Persero) which located on Jl Raya Yogya-Solo km 16 Prambanan Yogyakarta Indonesia. An inscription records the inauguration of the Prambanan temple in 856 AD. Prambanan temple has a special attraction for every visitor, both in terms of artistic value, history, religion, philosophy, and recreation. From the religious side, the Prambanan Temple is a Hindu holy building. The characteristics of the Hindu Temple is having a shape that is more tapered upwards than the building of a Buddhist Temple (for example the Borobudur Temple). In addition, Prambanan Temple has a legendary story where the temple was built by Bandung Bandawasa 1000 temples in one night as a condition to be able to marry the princess Rara Jonggrang. Because he can only build 999 temples, then Bandung Bandawasa curses Rara Jonggrang to become a temple so that there are even 1,000 temples.

The attraction of Prambanan Temple tourism has a huge impact on temple visitors that always increase every year, both local and foreign tourists. Tourists who come and even reach millions every year. At certain times, the number of visitors increases dramatically, for example during the end of semester holidays. An increase in the number of tourists that occur requires the manager of Prambanan Temple tourism objects to prepare facilities and infrastructure so that tourists can be served optimally, because if the service decreases, it can

have an impact on the number of clerks in the long range.

For that, it is necessary to analyze the number of tourists visiting the Prambanan Temple. This is done so that the results of existing research can be used as a reference for decision making by the management of Prambanan Temple attractions in order to prepare everything needed precisely so that the maximum. In this study two methods will be conducted to analyze the number of visitors to Prambanan Temple based on the data obtained.

Analysis of the number of visitors to the Prambanan Temple is done using the Fuzzy Time Series Chen Model and the seasonal auto regressive integrated moving average (SARIMA) model will produce two different outputs, both of which must have different errors. The two errors produced will be compared which will then result in a smaller error. The smaller error model means it is closer to the real than the other models.

METHODS

The data used in this study is secondary data obtained from the PT Borobudur Temple tourism park, Prambanan and Ratu Boko (Persero) which located on Jl Raya Yogya-Solo km 16 Prambanan Yogyakarta.

SARIMA (Seasonal Autoregressive Integrated Moving Average Method)

SARIMA is one of popular forecasting model for seasonal data, from (Nofinda Lestari, 2012) that SARIMA Model is an expanded of ARIMA Model, so the general notation is:

p, d, q : non seasonal part of model
 $(P, D, Q)^S$: seasonal part of model
 S : total periods per season

SARIMA has a general formula from ARIMA (p,d,q)(P,D,Q)^S below:

$$\Phi_p(B^S)\phi_p(B)(1-B)^d(1-B^S)^D Z_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t$$

Where:

$\phi_p(B)$: AR non seasonal
 $\Phi_P(B^S)$: AR seasonal
 $(1-B)^d$: differencing non seasonal
 $(1-B^S)^D$: differencing seasonal
 $\theta_q(B)$: MA non seasonal
 $\Theta_Q(B^S)$: MA seasonal
 a_t : residual value at t

Fitting a SARIMA model to data involves the following four-step iterative cycles: (Chen & Wang, 2007)

- Identify the SARIMA $(p, d, q)(P, D, Q)^S$ structure;
- Estimate unknown parameters;
- Perform goodness-of-fit tests on the estimated residuals;
- Forecast future outcomes based on the known data.

Fuzzy Time Series-Chen Model

The concept of fuzzy time series are describe as follows.

Let U be the universe of discourse, where $U = \{u_1, u_2, \dots, u_b\}$. A fuzzy set A_i of U is defined as $A_i = f_{A_i}(u_1)/u_1 + f_{A_i}(u_2)/u_2 + \dots + f_{A_i}(u_3)/u_3$, where f_{A_i} is the membership function of the fuzzy set A_i , $f_{A_i} : U \rightarrow [0,1]$. u_k is an element of fuzzy set A_i , and $f_{A_i}(u_i)$ is the degree of belongingness of u_k to A_i . $f_{A_i}(u_k) \in [0,1]$ and $1 \leq k \leq n$ (Chen, 1996).

The definitions of fuzzy time series are reviewed as follows (Huarng, 2001).

Definition 1. Let $Y(t) (t = \dots, 0, 1, 2, \dots)$, a subset of \mathbb{R} , be the universe of discourse on which fuzzy sets $f_i(t) (i = 1, 2, \dots)$ are defined and let $F(t)$ be a collection of $f_i(t) (i = 1, 2, \dots)$. Then, $F(t)$ is called a fuzzy time series on $Y(t) (t = \dots, 0, 1, 2, \dots)$.

Definition 2. If there exists a fuzzy relationship $R(t-1, t)$, such that $F(t) = F(t-1) \times R(t-1, t)$, where \times is an operator, then $F(t)$ is said to be caused by $F(t-1)$.

1). The relationship between $F(t)$ and $F(t-1)$ can be denoted by $F(t-1) \rightarrow F(t)$.

Definition 3. Suppose $F(t-1) = A_i$ and $F(t) = A_j$, a fuzzy logical relationship is defined as $A_i \rightarrow A_j$, where A_i is named as left-hand side of the fuzzy logical relationship and A_j the right-hand side. Note the repeated fuzzy logical relationships are removed.

Definition 4. Fuzzy logical relationships can be further grouped together into fuzzy logical relationship groups according to the same left-hand sides of the fuzzy logical relationships.

For example, there are fuzzy logical relationships with the same left-hand sides (A_i):

$$\begin{aligned} A_i &\rightarrow A_{j1}, \\ A_i &\rightarrow A_{j2}, \\ &\dots \end{aligned}$$

These fuzzy logical relationship group as follows:

$$A_i \rightarrow A_{j1}, A_{j2}, \dots$$

Definition 5. Suppose $F(t)$ is caused by $F(t-1)$ only, and $F(t) = F(t-1) \times R(t-1, t)$. For any t , if $R(t-1, t)$ is independent of t , then $F(t)$ is named a time-invariant fuzzy time series, otherwise a time-variant fuzzy time series.

The forecasting steps using Chen's fuzzy time series is as follows: (Fauziah, Wahyuningsih, & Nasution, 2016)

- Determine the universe of conversation U (*Universe of Discourse*).
- Determine the effective interval length using the average-based method and divide into several intervals that have the same interval.
- Determine the fuzzification linguistic value and define fuzzy set on U .
- Classification of Fuzzy Logical Relationship (FLR) that previously obtained into groups to form Fuzzy Logical Relationship Groups (FLRG).
- Perform a defuzzification process based on Chen's rules.
- Calculate the error value using MAPE, MAD, and MSE to evaluate the accuracy of the forecast.

Measure of Accuracy

To evaluate the accuracy of forecasting result, we used mean absolute percentage error (MAPE), the statistics is compute as follows:

$$MAPE = \frac{\sum |y_t - \hat{y}_t| / y_t}{n} (100\%) ; y_t \neq 0$$

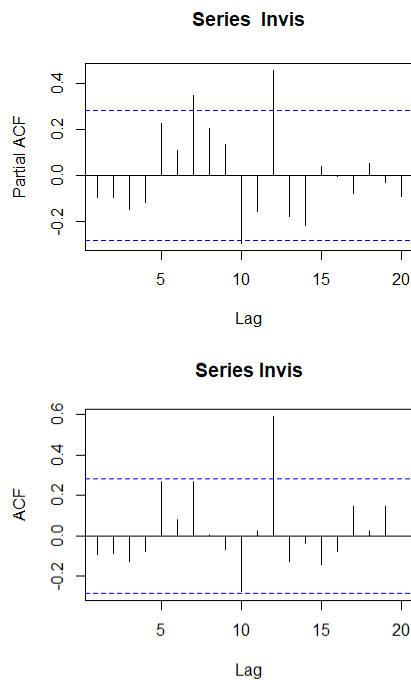
y_t is the actual value at time t , \hat{y}_t is the fitted value, and n is the number of observations. For all three measures,

the smaller the value, the better the fit of the model (Elena, et al., 2012).

RESULTS AND DISCUSSION

Result of SARIMA

We used R Studio to solve forecasting with SARIMA model. First, check the stationary data by plotting the ACF and PACF.

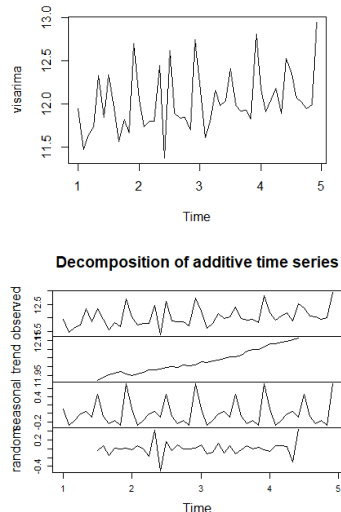


Augmented Dickey-Fuller Test

Data: Invis

Dickey-Fuller = -8.0576, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary

In Augmented Dickey-Fuller Test p-value is 0.01, its mean the plots is **stationary**. Then, determine the best model of sarima.



ARIMA(0,0,1)(0,1,0)[12] with drift

Coefficients:

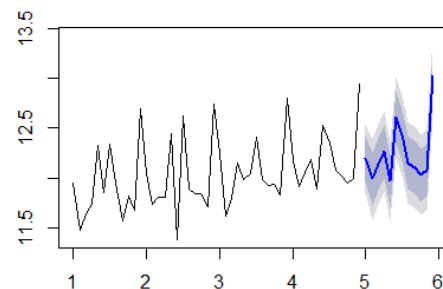
```
ma1 drift
-0.6184 0.0068
s.e. 0.1158 0.0010
```

sigma^2 estimated as 0.03122: log likelihood=12.14

AIC=-18.27 AICc=-17.52 BIC=-13.52

In this test, we know that the best SARIMA is SARIMA(0,0,1)(0,1,0)[12] and the results of SARIMA as follows

Forecasts from ARIMA(0,0,1)(0,1,0)[12] with dri



> finalforecastvalues

```
[1] 198750.7 161271.0 186812.8 212102.8 157985.6
299036.7 251256.9 189835.6 180740.1
[10] 167194.0 176036.6 453319.3
```

:

```
> sum(frequency$Freq)/48
```

```
[1] 0.3125
```

MAPE = 0.3125 = 31.25%

Result of Fuzzy Time Series

- Let D_{min} and D_{max} be the minimum and the maximum data of known historical data. Based on D_{min} and D_{max} , defined the universe of discourse U as $[D_{min} - D_1, D_{max} + D_2]$, where D_1 and D_2 are two proper positive numbers. According to the data of actual visitors of the Prambanan Temple, we get $D_{min} = 87520$ and $D_{max} = 417809$. Let $D_1 = 20$ and $D_{g1} = 91$, so that the universe of discourse $U = [87500, 417809]$.
- Partition the universe of discourse $U = [D_{min} - D_1, D_{max} + D_2]$ into even lengthy and equal length intervals u_1, u_2, \dots, u_m . In this paper, we partition $U = [87500, 417809]$ into seven intervals u_1, u_2, \dots, u_7 , where

Table 1. Partition the universe of discourse.

Class	Lower bounded	Upper bounded	Midpoint
A1	87,520	134,719	111,120
A2	134,720	181,919	158,320
A3	181,920	229,119	205,520
A4	229,120	276,319	252,720
A5	276,320	323,519	299,920
A6	323,520	370,719	347,120
A7	370,720	417,919	394,320

- c. Let A_1, A_2, \dots, A_k be a fuzzy sets, and defined it on the universe of discourse U as a follows:

$$A_1 = \frac{a_{11}}{u_1} + \frac{a_{12}}{u_2} + \dots + \frac{a_{1m}}{u_m}$$

$$A_2 = \frac{a_{21}}{u_1} + \frac{a_{22}}{u_2} + \dots + \frac{a_{2m}}{u_m}$$

$$\vdots$$

$$A_k = \frac{a_{k1}}{u_1} + \frac{a_{k2}}{u_2} + \dots + \frac{a_{km}}{u_m}$$

where $a_{i,j} \in [0,1]$, $1 \leq i \leq k$, and $1 \leq j \leq m$. The fuzzified historical visitors are shown below.

Table 2. Fuzzified historical visitors.

Year	Month	Visitors	Fuzzification
2015	JAN	154150	A2
2015	FEB	96178	A1
2015	MART	113133	A1
2015	APRIL	125102	A1
2015	MEI	226374	A3
2015	JUNI	139539	A2
2015	JULI	228048	A3
2015	AGST	154715	A2
2015	SEPT	105439	A1
2015	OKT	135648	A2
2015	NOP	116837	A1
2015	DES	326083	A6
2016	JAN	170221	A2
2016	FEB	124999	A1
2016	MART	133762	A1
...
2017	JULI	244304	A4
2017	AGST	159750	A2
2017	SEPT	150105	A2
2017	OKT	151971	A2
2017	NOP	137231	A2
2017	DES	364490	A6
2018	JAN	194331	A3
2018	FEB	148638	A2
2018	MART	172179	A2
2018	APRIL	195488	A3
2018	MEI	145610	A2
2018	JUNI	275612	A4
2018	JULI	231575	A4
2018	AGST	174965	A2
2018	SEPT	166582	A2
2018	OKT	154097	A2
2018	NOP	162247	A2
2018	DES	417809	A7

Table 3. Fuzzy Logical Relationship.

$A_1 \rightarrow A_1$	$A_1 \rightarrow A_2$	$A_1 \rightarrow A_3$	$A_1 \rightarrow A_4$	$A_1 \rightarrow A_5$
$A_1 \rightarrow A_6$	$A_2 \rightarrow A_1$	$A_2 \rightarrow A_2$	$A_2 \rightarrow A_3$	$A_2 \rightarrow A_4$
$A_2 \rightarrow A_6$	$A_2 \rightarrow A_7$	$A_3 \rightarrow A_1$	$A_3 \rightarrow A_2$	$A_4 \rightarrow A_1$
$A_4 \rightarrow A_2$	$A_4 \rightarrow A_5$	$A_5 \rightarrow A_2$	$A_6 \rightarrow A_2$	$A_6 \rightarrow A_3$

Table 4. Fuzzy Logical Relationship Groups.

FLRG	LH	RH	total
G1	A1	A1,A2,A3,A4,A5,A6	12
G2	A2	A1,A2,A3,A4,A6,A7	21
G3	A3	A1,A2	6
G4	A4	A1,A2,A5	4
G5	A5	A2	1
G6	A6	A2,A3	3

- d. Calculate the forecasted outputs. The calculations are carried out by the following principles: (Chen, 1996)

- 1) If the fuzzified visitors of year i is A_j , and there is only one fuzzy logical relationship in the fuzzy logical relationship groups, which is shown as follows:

$$A_j \rightarrow A_k,$$

where A_i and A_k are fuzzy sets and maximum membership value of A_k occurs at interval u_k , and midpoint of u_k is m_k , then the forecasted visitors of year $i + 1$ is m_k .

- 2) If the fuzzified visitors of year i is A_j , and there is only one fuzzy logical relationship in the fuzzy logical relationship groups, which is shown as follows:

$$A_j \rightarrow A_{k1},$$

$$A_j \rightarrow A_{k2},$$

$$\vdots$$

$$A_j \rightarrow A_{kp}.$$

Where $A_j, A_{k1}, \dots, A_{kp}$ are fuzzy sets, and the maximum membership values of $A_{k1}, A_{k2}, \dots, A_{kp}$ occurs at interval u_1, u_2, \dots, u_p and the midpoint of u_1, u_2, \dots, u_p are m_1, m_2, \dots, m_p , then the forecasted visitors of year $i + 1$ is $(m_1 + m_2 + \dots + m_p)/p$.

- 3) If the fuzzified visitors of year i is A_j , and there do not exist any fuzzy logical relationship groups whose current state of visitors is A_j , where the maximum membership value of A_j occurs at interval u_j , and the midpoint of u_j is m_j , then the forecasted visitors of year $i + 1$ is m_j .

Thus, based on the principles we can get forecasted visitors.

Table 5. The results of forecasting visitors of Prambanan Temple.

Year	Month	Visitors	Forecasted Visitors
2015	JAN	154150	
2015	FEB	96178	244852.8
2015	MART	113133	229119.5
2015	APRL	125102	229119.5
2015	MEI	226374	229119.5
2015	JUNI	139539	134719.5
2015	JULI	228048	244852.8
2015	AGST	154715	134719.5
2015	SEPT	105439	244852.8
2015	OKT	135648	229119.5
2015	NOP	116837	244852.8
2015	DES	326083	229119.5
2016	JAN	170221	158319.5
2016	FEB	124999	244852.8
2016	MART	133762	229119.5
...
2017	JULI	244304	244852.8
2017	AGST	159750	189786.2
2017	SEPT	150105	244852.8
2017	OKT	151971	244852.8
2017	NOP	137231	244852.8
2017	DES	364490	244852.8
2018	JAN	194331	158319.5
2018	FEB	148638	134719.5
2018	MART	172179	244852.8
2018	APRL	195488	244852.8
2018	MEI	145610	134719.5
2018	JUNI	275612	244852.8
2018	JULI	231575	189786.2
2018	AGST	174965	189786.2
2018	SEPT	166582	244852.8
2018	OKT	154097	244852.8
2018	NOP	162247	244852.8
2018	DES	417809	244852.8
2019	JAN	-	181919.5

MAPE = 47.08%

Discussion

In previous research by (Elena, et al., 2012) about the comparison of forecasting results using fuzzy time series and SARIMA, the results is fuzzy time series-chen model is the best result because it produces a smaller MAPE value. Meanwhile, in this paper the MAPE of the SARIMA model is better than the fuzzy time series chen.

The data plot in this study shows that there is data outreach which will certainly have an effect on the results of forecasting, so it is expected that in future there are research about the classification of suitability of the time series data to a method, like SARIMA model, fuzzy time series model, and others.

CONCLUSION

The data plot in this study shows that there is data outreach which will certainly have an effect on the results of forecasting, so it is expected that in future there are research about the classification of suitability of the time series data to a method, like SARIMA model, fuzzy time series model, and others. The results of these methods are the visitors number prediction with different errors, so it can be seen which method is better between the two.

ACKNOWLEDGEMENTS

This research was partially supported by PT. Taman Wisata Candi Borobudur Prambanan Ratu Boko (persero) who gave us the opportunity to proceed the data, although they may not agree with all of the interpretations/conclusions of this paper.

REFERENCES

- Chen, K.-Y., & Wang, C.-H. 2007. A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. *Expert Systems with Applications*, 245-264.
- Chen, S.-M. 1996. Forecasting Enrollments Based on Fuzzy Time Series. *Fuzzy Sets and System*, 311-319.
- Elena, M., Lee, M. H., Suhartono, Hossein, Haizum, N., & Bazilah, N. A. 2012. FUZZY TIME SERIES AND SARIMA MODEL FOR FORECASTING TOURIST ARRIVALS TO BALI. *Jurnal Teknologi*, 69-81.
- Fauziah, N., Wahyuningsih, S., & Nasution, Y. N. 2016. Peramalan Menggunakan Fuzzy Time Series Chen (Study Kasus: Curah Hujan Kota Samarinda). *Statistika*, Vol. 4, No. 2.
- Huang, K. 2001. Heuristic Models of Fuzzy Time Series for Forecasting. *Fuzzy Set and System*, 369-386.
- Nofinda Lestari, N. W. 2012. Peramalan Kunjungan Wisata dengan Pendekatan Model SARIMA (Studi kasus: Kusuma Argowisata). *JURNAL SAINS DAN SENI ITS*, 29-30.

