Hybrid model of ARIMA-linear trend model for tourist arrivals prediction model in Surakarta City, Indonesia

Cite as: AIP Conference Proceedings **2114**, 060010 (2019); https://doi.org/10.1063/1.5112481
Published Online: 26 June 2019

Purwanto, Sunardi, Fenty Tristanti Julfia, and Aditya Paramananda

Hybrid Model of ARIMA-Linear Trend Model for Tourist Arrivals Prediction Model in Surakarta City, Indonesia

Purwanto[1,a)], Sunardi[1,b)], Fenty Tristanti Julfia[1, c)] and Aditya Paramananda[1,d)]

[1] *Universitas Dian Nuswantoro, Semarang, Central Java, Indonesia*

.

[a)] Corresponding author: purwanto@dsn.dinus.ac.id
[b)] sunardi@dsn.dinus.ac.id
[c)] fentytristanti@gmail.com
[d)] myaditya.nanda@gmail.com

Abstract. It is important to predict the tourist arrival to help the government in making appropriate decisions. Many models have been proposed to estimate the number of tourist arrivals in the future. An autoregressive integrated moving average (ARIMA) model, linear trend and Holt-Winter triple exponential smoothing are among successful models used in various fields. In the present study, we propose a hybrid model that combines ARIMA and linear trend model as a tourist arrivals prediction model. Experiment results show that the hybrid model produces better prediction performance compared to ARIMA, linear trend and Holt-Winter triple exponential smoothing models.

INTRODUCTION

The earnings gained from the tourism sector become the mainstay source of the current Indonesian government. In 2018, the tourism sector contributes the largest earnings to Indonesia, which is US$ 20 billion. Therefore, it is necessary to give easier access for foreign tourists in visiting Indonesia, including visas and the optimization of new tourist destinations in Indonesia [1].

Many cities in Indonesia with their natural and cultural objects can be potential tourism objects, and one of them is Surakarta or Solo city located in Central Java Province. It has a very strategic location and relatively accessible for foreign tourists since it already has an international airport. Solo with its 500,000 population, is well-known as the center for the development of Javanese traditions. Surakarta is a home of World Heritages set by UNESCO, they include *Keris* in 2005, *Batik* on October 2, 2009, Sangiran as an archaeological site in 1996, puppet shows as cultural heritage on 7 November 2003, and others [2].

Nevertheless, the frequency of tourists visit in Indonesia is relatively unstable, so is in Solo. It is not easy to keep the stability and estimate the number of tourists. Yet we expect there will be a large and constant number each year. For this reason, a good strategy is needed to achieve this aim in the future. In managing tourists, the prediction of tourist arrivals is very important, since it will help the government in arranging appropriate policies.

In the last few years, data mining models have been widely applied in solving problems, including in predicting the future events. There have been numerous forecasting models proposed by researchers to solve problems in various fields. In solving tourism problems, ARMA and ARIMA models have been applied to estimate the number of tourists. The ARIMA model applied for tourists prediction has been done in [3] -[5]. Furthermore, Holt-Winter triple exponential model has been successfully applied for forecasting some aspects in various fields. Lim et al. [6] applied Holt-Winter triple exponential to predict the number of nights spent by guests in New Zealand. Razali et al. [7] reported the application of the ARIMA and Holt-Winter methods for forecasting water consumptions

Exploring Resources, Process and Design for Sustainable Urban Development
AIP Conf. Proc. 2114, 060010-1–060010-8; https://doi.org/10.1063/1.5112481
Published by AIP Publishing. 978-0-7354-1850-9/$30.00

expenditure. Tirkeş et al. [8] successfully applied the Holt-Winter method for predicting monthly sales of a private company. Brügner [9] used the Holt-Winter method to do traffic prediction. Meanwhile, linear trend model has been successfully implemented by Janhvi et al. [10] to predict some export bound food items.

Today, researchers have improved many methods to increase predictive accuracy by using a hybrid model. A hybrid model is used to gain high prediction accuracy rates [11]- [12]. Zhang [11] has applied hybrid models by combining ARIMA and Neural Networks in which the result showed better accuracy. Purwanto [13] has also proposed a hybrid model from Neural Network and fuzzy inference system for investigating health data, and the accuracy of the produced hybrid model is better than a single model. In the present study, we propose a hybrid model by combining ARIMA and linear trend model for predicting tourist arrivals in Surakarta, Central Java, Indonesia.

METHODOLOGY

For testing the model, data sets which is time series for tourist arrivals in Surakarta, Central Java, Indonesia in the period of 1991 to 2013. Data on tourist arrivals included data on monthly arrivals during January 1991 to December 2013 which is shown in Table 1.

TABLE 1. Time Series data of the time series tourist arrivals

No	Month	Year									
		1991	1992	1993	1994	1995	1996	1997	1998	...	2013
1	January	2,588	2,217	3,447	3,878	2,342	2,454	1,725	1,649	...	1,764
2	February	2,140	2,569	2,723	3,553	2,597	2,199	1,749	3,521	...	1,602
3	March	2,008	2,699	3,431	3,274	2,776	2,780	2,533	5,140	...	1,684
4	April	1,926	3,538	3,053	3,804	2,070	2,581	2,134	1,122	...	2,019
5	May	2,368	3,569	3,673	4,413	2,124	2,890	2,721	624	...	1,740
6	June	3,398	3,727	4,262	3,405	2,554	3,273	2,291	268	...	1,826
7	July	3,647	4,953	11,426	4,953	3,656	3,139	4,092	813	...	2,793
8	August	4,407	5,034	6,439	4,513	3,989	5,560	4,538	1,499	...	2,426
9	September	3,402	4,336	4,988	3,534	2,979	3,036	3,510	1,207	...	2,308
10	October	3,661	4,453	4,089	3,175	3,046	2,143	2,616	727	...	1,986
11	November	2,694	3,451	3,455	2,829	2,835	2,044	1,547	1,066	...	1,632
12	December	2,909	4,096	4,014	3,420	2,974	2,223	1,688	1,097	...	1,686

An autoregressive integrated moving average (ARIMA), linear trend, Holt-Winter triple exponential smoothing and a hybrid model that combines ARIMA and linear trend model are used as the tourist arrivals prediction models. The proposed hybrid model is illustrated in Fig. 1 as follow:

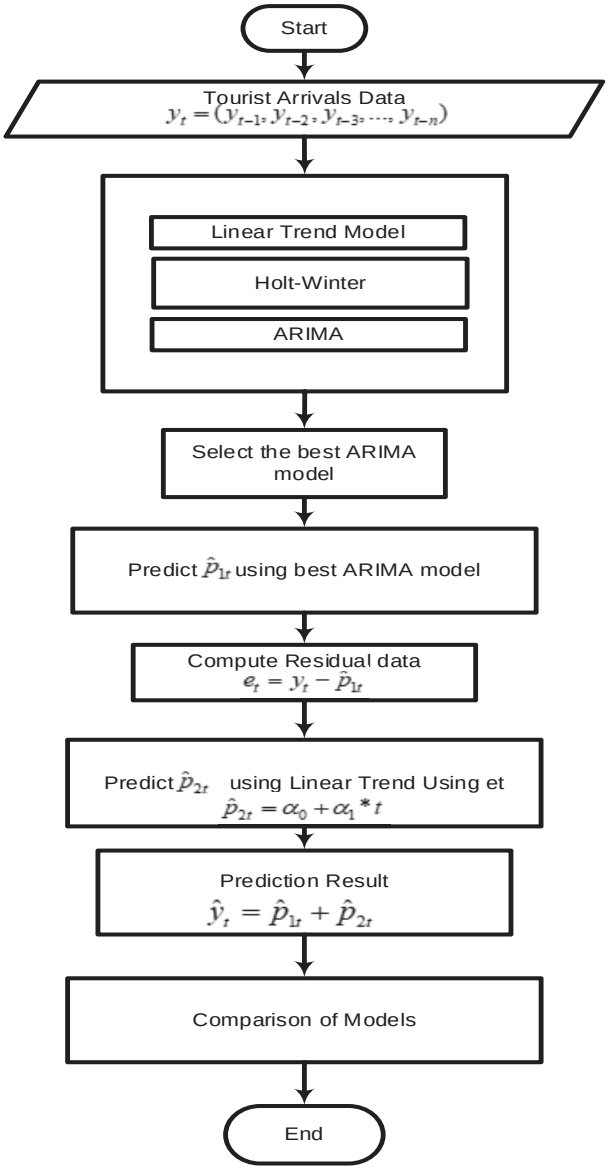

FIGURE 1. The flowchart of the proposed hybrid method for tourist arrivals prediction model

The proposed hybrid method is described as following steps:

STEP 1 : The time series data of tourist arrivals is collected as shown in Table 1.

STEP 2 : The experiments by using ARIMA model (*p, d, q*) with different values on *p, d,* and *q*, Linear trend model, and Holt-Winter triple exponential smoothing.

STEP 3 : The calculation by applying the hybrid model, which is the combination of the ARIMA and linear trend model. For the input, ARIMA model uses actual data while linear trend model uses residual data. The result of hybrid model is the total sum of prediction obtained by applying the ARIMA and linear trend models.

STEP 4 : The comparison of forecasting models, which are the ARIMA, Linear trend model, Holt-Winter triple exponential smoothing and a hybrid model of the ARIMA-linear trend model to assess the accuracy of the models. The best model is presumed to have the highest performance accuracy (the smallest Root Mean Square Error (RMSE)) that will be applied to predict the tourist arrivals.

RESULTS AND DISCUSSION

In the present study, we applied and investigated four prediction models, which are ARIMA, Linear Trend model, Holt-Winter triple exponential smoothing and a hybrid model of ARIMA-Linear Trend model. The results of the implementation are elaborated in the following section.

ARIMA model for forecasting tourist arrivals

In the present study, we predicted tourist arrivals by applying ARIMA (*p, d, q*) models with different values of the parameters. For ARIMA (*p, d, q*), we used the parameters of *p* (0, 1, 2), *d* (0,1) and *q* (0, 1, 2). The results are presented in Table 2.

TABLE 2. The tourist arrivals prediction model and RMSE by applying ARIMA

Month	Actual Data	ARIMA (1,0,1)	ARIMA (2,1,2)	ARIMA (1,0,2)	ARIMA (1,1,1)	ARIMA (1,1,2)	ARIMA (2,1,1)
Feb-91	2,140.00	2,452.44	2,584.54	2,452.04	2,584.59	2,584.63	2,584.65
Mar-91	2,008.00	2,189.64	2,253.48	2,188.67	2,247.18	2,243.33	2,242.01
Apr-91	1,926.00	2,051.21	2,123.54	2,161.79	2,130.00	2,131.25	2,131.51
May-91	2,368.00	1,969.56	2,052.83	2,075.87	2,051.22	2,052.76	2,053.32
Jun-91	3,398.00	2,229.90	2,307.74	2,304.13	2,316.59	2,321.62	2,323.45
Jul-91	3,647.00	2,934.58	2,972.91	2,805.03	2,974.87	2,979.89	2,981.92
Aug-91	4,407.00	3,244.50	3,196.42	2,918.01	3,183.64	3,176.70	3,174.62
Sep-91	3,402.00	3,790.08	3,703.38	3,437.12	3,704.68	3,700.93	3,699.52
Oct-91	3,661.00	3,276.40	3,205.28	2,949.34	3,174.93	3,156.93	3,150.65
Nov-91	2,694.00	3,327.65	3,336.36	3,313.83	3,353.85	3,353.44	3,352.62
Dec-91	2,909.00	2,730.37	2,834.48	2,754.31	2,807.84	2,798.66	2,795.43
...	...	...	...	...	...	...	...
13-Dec	1,686.00	1,785.88	1,902.85	1,971.09	1,908.11	1,909.16	1,909.49
RMSE		875.557	**849.626**	862.425	849.918	849.813	849.771

Table 2 shows that the smallest value of RMSE is obtained by the ARIMA (2, 1, 2). Therefore, it can be assumed that the ARIMA (2,1,2) model is the best model for further assessment.

Linear trend model for predicting tourist arrivals

The predicted values of the tourist arrivals by linear trend model are calculated as follow:

$$\hat{Y}_t = \alpha_0 + \alpha_1 \mathrm{t} \tag{1}$$

The parameters α_0 is the constant, α_1 is the average change from one period to the next, and *t* is value of the time unit. After the implementation of linear trend model by data tourist arrivals, the results can be expressed as follow:

$$\hat{Y}_t = 2873.4 - 6.0869 * \mathrm{t} \tag{2}$$

The results of prediction and the RMSE of the implementation of the linear trend model are presented in Table 3**.**

TABLE 3. The prediction results and the RMSE using linear trend model

Month	Actual data	Linear Trend Model
Jan-91	2588	2867.2683
Feb-91	2140	2861.1814
Mar-91	2008	2855.0945
Apr-91	1926	2849.0076
May-91	2368	2842.9208
Jun-91	3398	2836.8339
Jul-91	3647	2830.747
Aug-91	4407	2824.6601
Sep-91	3402	2818.5732
Oct-91	3661	2812.4864
Nov-91	2694	2806.3995
Dec-91	2909	2800.3126
....		
Dec-13	1686	1193.3767
RMSE		1,275.25

Holt-Winter triple exponential smoothing method for predicting tourist arrivals

In the present study, we implemented Holt-Winter triple exponential smoothing method for predicting the number of tourist arrivals. The accuracy and performance results using the model are shown in Table 4. We calculated Holt-Winter triple exponential smoothing method by applying different smoothing constants, which are α, γ and δ. The performance and accuracy of RMSE by applying the method is represented in Table 4.

TABLE 4. The performance assessment using Holt-Winter triple exponential smoothing method

MODEL	**Smoothing Constants**			**PERFORMANCE**	
	α	**γ**	**δ**	**MSE**	**RMSE**
Holt-Winter triple exponential smoothing	0.20	0.20	0.20	1,131,557.73	1063.75
	0.10	0.10	0.10	4,366,350.71	2089.58
	0.70	**0.10**	**0.10**	**923,822.41**	**961.16**
	0.70	0.10	0.20	931,259.26	965.02
	0.80	0.10	0.10	935,263.50	967.09
	0.70	0.10	0.30	938,191.32	968.60
	0.80	0.10	0.20	941,700.78	970.41
	0.60	0.10	0.20	941,418.46	970.27
	0.70	0.10	0.30	945,823.08	972.53
	0.80	0.10	0.20	946,756.93	973.01

Table 4 shows the smallest value of RMSE is obtained by Holt-Winter triple exponential smoothing method with smoothing constants of α=0.7, γ=0.1 and δ=0.1.

Hybrid model of ARIMA-Linear trend model for predicting tourist arrivals

Table 2 demonstrates that the smallest value of RMSE is obtained by the ARIMA (2, 1, 2) model. Subsequently, the best ARIMA (2, 1, 2) model is included to construct the hybrid model. Based on the proposed method as presented in Fig. 1, the hybrid model is applied to predict the tourist arrivals. The performance result is shown in Table 5 as follow:

TABLE 5. The prediction results and the RMSE of the tourist arrivals prediction using Hybrid model

Month	Actual	ARIMA (2,1,2)	Linear Trend	Hybrid ARIMA-Linear Trend
Feb-91	2140	2584.54	-39.62	2544.93
Mar-91	2008	2253.48	-39.26	2214.22
Apr-91	1926	2123.54	-38.90	2084.64
May-91	2368	2052.83	-38.55	2014.28
Jun-91	3398	2307.74	-38.19	2269.55
Jul-91	3647	2972.91	-37.84	2935.07
Aug-91	4407	3196.42	-37.48	3158.94
Sep-91	3402	3703.38	-37.13	3666.25
Oct-91	3661	3205.28	-36.77	3168.51
Nov-91	2694	3336.36	-36.41	3299.95
Dec-91	2909	2834.48	-36.06	2798.42
...				
Dec-13	1686	1902.85	57.83	1960.68
		RMSE		849.107

The comparison of the actual and predicted values is carried out and the results are specified in Fig. 2.

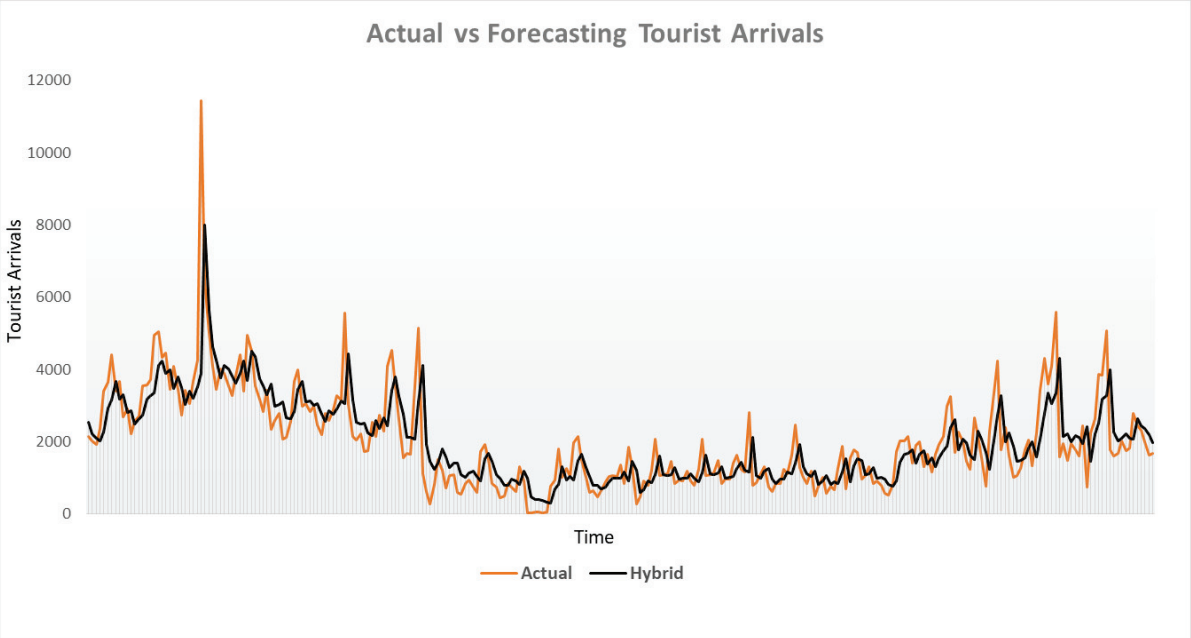

FIGURE 2. The comparison between the actual data and the tourist arrivals prediction model data

From figure of comparison actual and tourist arrivals prediction, it is shown that the tourist arrivals prediction values by applying the hybrid model combining ARIMA (2,1,2) and linear trend model are very close to the actual values.

Comparison of the Models

Table 6 shows the comparison of RMSE values obtained by applying the best ARIMA model, Linear trend, Holt-Winter triple exponential smoothing and Hybrid model of ARIMA(2,1,2)-Linear trend model.

TABLE 6. The comparison of the performance of tourist arrivals prediction models

NO	MODEL	RMSE
1	The Best ARIMA	849.626
2	The Holt-Winter triple exponential smoothing	961.160
3	Linear Trend	1,275.25
4	**Hybrid (ARIMA (2,1,2)-Linear Trend)**	**849.107**

From Table 5, it is shown that the hybrid model that combines ARIMA (2,1,2) and linear trend model gives the best results compared to other models. After using the Hybrid model, the RMSE value decreases by 0.519 from the ARIMA model, so the hybrid model has the smallest RMSE. The value of RMSE is used as the parameter to measure the prediction accuracy. Furthermore, since the hybrid model has the lowest RMSE, then this model can be used for tourist arrivals prediction. Fig. 3 compares the RMSE of different models.

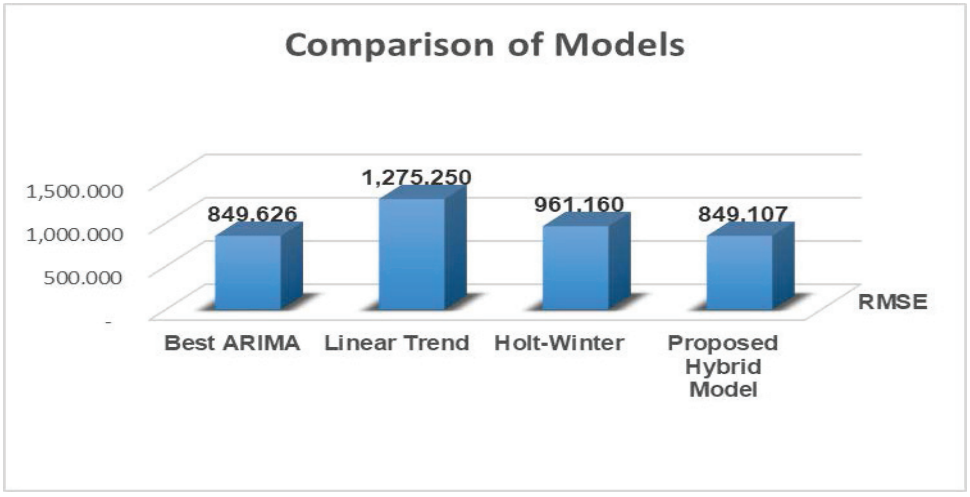

FIGURE 3. The comparison of the RMSE value of the examined models

CONCLUSION

This study has proposed a hybrid model that combines ARIMA and linear trend model as a tourist arrival prediction model. The performances of each model as well as the hybrid model are determined and then compared. The performance of the models as expressed in the value of Root Mean Square Error (RMSE) has been assessed for comparison. It found that hybrid model of ARIMA-Linear trend model obtained the smallest RMSE value among those of ARIMA model, linear trend model, and Holt-Winter triple exponential smoothing model. Briefly, the developed hybrid model can be a reliable model for predicting the tourist arrivals.

ACKNOWLEDGMENT

This paper is the result of a study under research grant supported by the Ministry of Research, Technology and Higher Education, Republic of Indonesia in 2018.

REFERENCES

[1] E. Simorangkir, *Tourism Becomes the Mainstay of Foreign Exchange Contributors of US $ 20 Billion*, (https://finance.detik.com/berita-ekonomi-bisnis/d-3844660/pariwisata-jadi-andalan-penyumbang-devisa-us-20-miliar, 2018)
[2] Tourism Department of Surakarta City, *World Heritage in Solo*, (http://pariwisatasolo.surakarta.go.id/warisan-dunia-di-kota-solo, 2018)
[3] Choden and S. Unhapipat, "ARIMA model to forecast international tourist visit in Bumthang, Bhutan", *IOP Conference Series: Journal of Physics: Conf. Series 1039* (2018)
[4] A. Papatheodorou and H. Song, *Tourism Economics* 11, 11–23 (2005).
[5] L. Nanthakumar and Y. Ibrahim, *South Asian Journal of Tourism and Heritage* 3 (2), (2010)
[6] C. Lim, C. Chang, M. McAleer, *International Journal of Hospitality Management 28*, 228-235 (2009)
[7] S. N. A. M. Razali, M. S. Rusiman, N. I. Zawawi, N. Arbin, "Forecasting of Water Consumptions Expenditure Using Holt-Winter's and ARIMA", *IOP Conf. Series: Journal of Physics: Conf. Series 995* (2018)
[8] G. Tirkeş, C. Güray, N. Celebi. 2017, *Tehnički vjesnik* 24 (2),503-509 (2017)
[9] H. Brügner, "Holt-Winters Traffic Prediction on Aggregated Flow Data", *Seminars FI / IITM SS 17, Network Architectures and Services*, September 2017, pp. 25-32.
[10] U.Janhvi, and P. J. Jha, *International Journal of Academic Research and Development* 2(3), 153-157 (2017)
[11] G. P. Zhang, *Neurocomputing* 50, 159 – 175 (2003).
[12] D. O. Faruk, *Engineering Applications of Artificial Intelligence* 23, 586–594 (2010).
[13] Purwanto, C. Eswaran and R. Logeswaran, "Improved Adaptive Neuro-Fuzzy Inference System for HIV/AIDS Time Series Prediction", *in Informatics Engineering and Information Science 253*, (Malaysia, 2011), pp. 1–13.