JOSEIT JOURNAL OF SYSTEMS ENGINEERING AND INFORMATION TECHNOLOGY
http://jurnal.iaii.or.id/index.php/JOSEIT

Application of Data Mining for Visit Prediction at Amikom Creative Economy Park

Rumini[1*], Norhikmah[2]
[12]Universitas AMIKOM Yogyakarta, Indonesia

*rumini@amikom.ac.id

Abstract

A creative economy park is a place designed with strategic goals for technology skills collaboration, information and knowledge transfer, creation of innovative high-tech enterprises and entrepreneurs, introduction of new technology industries in creative economy enterprises to promote economic development. Yogyakarta Amikom University has been declared a Creative Economy Park and is known as Amikom Creative Economy Park (ACEP). ACEP includes multiple multimedia environments for targeting businesses such as software development, film, television, games, radio, animation, advertising, investment consulting, and project design. Every year, the number of institutions visiting Amikom Yogyakarta University carries the slogan Amikom Creative Economy Park with a fairly busy program of visits. The agenda for accepting this visit was carried out by Amikom's Public Relations Department (DKUI, Directorate of Public Relations and International Affairs). The evolution of visitor numbers from year to year, forecasts must be made to support the planning and preparation process when receiving visits. This research will discuss the trend of visitors having a comparative study in Amikom Creative Economy Park in the future. The data used in this study is visitor data from January 2019 to December 2019. This predictive data analysis uses the Autoregressive Integrated Moving Average (ARIMA) method and Exponential Smoothing as a comparison for the accuracy of the prediction. With the forecast of this visit, the planning and preparation for the Directorate of Public Relations and International Affairs and for the University AMIKOM Yogyakarta is to be done.

Keywords: Prediction, Visit, ARIMA, Exponential Smoothing, Creative Economy Park

1. Introduction

A creative economy park is a place designed using strategic goals for the cooperation of technological skills, the transfer of information and knowledge, the cultivation of innovative and entrepreneurial high-tech enterprises, and the emergence of new technology branches in creative economy enterprises to promote economic development. Amikom University Yogyakarta has been declared a Creative Economy Park and is known as Amikom Creative Economy Park (ACEP). ACEP includes several enterprise multimedia environments, e.g. Software development, film, television, games, radio, animation, advertising, investment advice and project design [2].

Every year the number of institutions visiting Amikom University Yogyakarta carrying the slogan Amikom Creative Economy Park is quite busy. The agenda for receiving these visits is managed by Amikom's Public Relations Department (DKUI, Directorate of Public Relations and International Affairs). The evolution of visitor numbers from year to year and each month must be forecast to support the planning and preparation process when receiving visits. Forecasting is the art and science of estimating future events. This can be done by taking past dates and putting them into the future with a mathematical form [5].

This study will discuss the trend of visitors studying in Amikom Creative Economy Park in the future. The data used in the research is data from visitors to ACEP from January 2019 to December 2019. This forecast data analysis uses Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing Holt-Winters as a comparison for the level of forecast accuracy by calculating the forecast error is with mean squared Error (MSE) and Mean Absolute Percentage Error (MAPE) so that the error becomes as small as possible. With the forecast of this visit, it is expected to provide planning and preparation to the Directorate of Public Relations and International Affairs of AMIKOM University Yogyakarta.

2. Method

This research uses a quantitative approach that emphasizes the analysis of numerical data (numbers) with statistical method processing. In this research, the research method is carried out by identifying problems, formulating problems, studying literature, analyzing it, carrying out research steps and drawing conclusions. Problem identification begins with a study of the literature. At this stage, the search and collection of secondary data is performed, and secondary data is selected, namely sampled data, which is used as the problem under study and analysis.

Accepted by editor: 25-08-2022 | Final revision: 15-09-2022 | Online publication: 25-09-2022

The problem focus phase in this research is research using Holt-Winters exponential smoothing and ARIMA methods to compute predictions. The research is supported with the help of evaluations and SPSS programs. This research data is data in terms of months and number of visits to ACEP for the period January 2019 - December 2019. The data is a type of time series data derived from the Directorate of Public Relations and International Affairs visit data. The techniques used to predict visits for the next 3 months are the Autoregressive Moving Average (ARIMA) and Exponential Smoothing methods.

2.1 Autoregressive Moving Average (ARIMA)

ARIMA is commonly referred to as the BoxJenkins time series method. ARIMA or Autoregressive Integrated Moving Average was discovered by George Edward Pelham Box and Gwilym Meirion Jenkins, a periodic series solving forecasting method for time series analysis. The ARIMA method is used very well for predicting data for short-term data, while the accuracy of the prediction is not good for long-term predictions. The ARIMA model is a univariate model, so this model is appropriate when the observations of the time series are not statistically related.

The phases of the ARIMA method include:

a. checking the stationarity of the data;
b. identification of the most suitable model by calculating and checking the ACF and PACF of the correlogram;
c. model estimation phase by estimating the parameters in the model;
d. diagnostic test phase: to test the adequacy of the model obtained in the previous phase, namely normality test, heteroscedasticity test and autocorrelation test;
e. using the model to make predictions, from which the MSE and MAPE values are then calculated;
f. performing predictions after obtaining a suitable model, then making predictions using the model according to the desired prediction period (3 periods in this study or months).

2.2 Exponential Smoothing

Besides using ARIMA, this research also uses the exponential smoothing method. Forecasts for seasonal data are developed using the Holt-Winters exponential smoothing method. The Holt-Winters method is the name of the triple exponential smoothing method that smooths three times and then forecasts. The Holt-Winters method is an extension of the two Holt parameters. The Holt-Winters method is a time series forecasting method that can handle seasonal behavior in data based on past data.

The stages of the Holt-Winters exponential smoothing method itself consist of several stages, namely:

a. taking seasonal data;
b. make a scatter diagram;
c. determining the seasonal length;
d. determine the initial estimated values (initialization) of smoothing (SL), trend (bL) and seasonality (IL);
e. determine the value of α, β, γ by trial and error which is in the range between 0 - 1;
f. finding the smallest RMSE value from the values of α, β, γ which is then tested for autocorrelation to get the best model;
g. calculate (St), (bt) and (It); (h) the forecast results are obtained and the MSE and MAPE values are calculated;
h. perform forecasting for the next 3 periods (3 months);

3. Result and Discussion

In this study, forecasts are made using visit data at ACEP for the period January 2019 to December 2019. Visit data is included in time series data that can be predicted using ARIMA and exponential smoothing methods. The visit data forms a trend pattern shown in Figure 1, which shows that visits decreased in May 2019 and increased in November 2019.

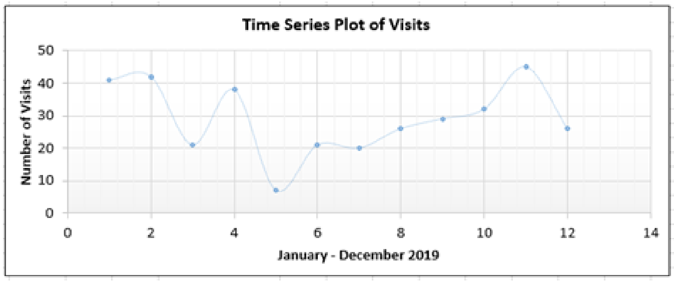
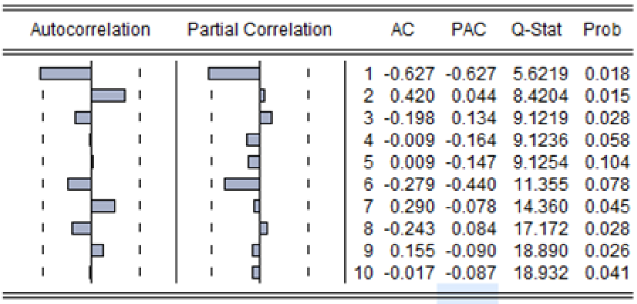


Figure 1. Trend Pattern of Visitation

3.1 Prediction Using ARIMA

In the ACEP visit data, it can be seen that the data is stationary and the data is trending upwards (there is a trend element). In Figure 1, it can be seen that the data has a high-low that is not the same for each month, indicating that the data does not have stationary variance. Because the data is not stationary on average and not stationary in variance, differentiation and transformation is required. The time series data are differentiated and transformed using E-Views (computer program), the results of the differentiation and transformation are shown in the time series diagram in Figure 2.



	AC	PAC	Q-Stat	Prob
1	-0.627	-0.627	5.6219	0.018
2	0.420	0.044	8.4204	0.015
3	-0.198	0.134	9.1219	0.028
4	-0.009	-0.164	9.1236	0.058
5	0.009	-0.147	9.1254	0.104
6	-0.279	-0.440	11.355	0.078
7	0.290	-0.078	14.360	0.045
8	-0.243	0.084	17.172	0.028
9	0.155	-0.090	18.890	0.026
10	-0.017	-0.087	18.932	0.041

Figure 2. Correlogram of Second Differencing and Logarithmic Transformation Results

Based on the graph of Figure 3 of the differentiation and transformation results, it is noted that the data tends to be much better. Stationary tests can be performed by looking at non-stationary data plots in mean and variance, correlogram of ACF and PACF plots, and unit root test values. The unit root test developed by Dickey-Fuller [3] is used to test the stationarity of data. This test is performed to test the assumption that time series data in Figure 3 is not stationary.

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.147471	0.0009
Test critical values:	1% level	-4.297073	
	5% level	-3.212696	
	10% level	-2.747676	

*MacKinnon (1996) one-sided p-values.

Figure 3. Unit Root Test

Based on Figure 3, the absolute value of the ADF test statistic is 6.147471, which is more than the value of the t-statistic with a critical value of = 5%, which is 3.212696, which means that H0 is rejected, which suggests indicates that the data were stationary. The above stationary data is differentiated twice because the first differentiation probability is 0.0618 while in the second differentiation the probability value is 0.0009. After performing the identification phase, it is known that the data on the number of visits to ACEP are non-stationary in mean and variance. The data are stationary after two differentiations for seasonal and non-seasonal patterns with logarithmic transformation. The next stage of identification is estimation. At this stage, the accuracy of several preliminary models is seen such that the best ARIMA model is obtained by the minimum error minimization criterion [4].

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.982619	3.463799	-0.283682	0.7862
AR(1)	-0.674436	0.410528	-1.642851	0.1515
AR(2)	0.295268	0.352088	0.838621	0.4338
AR(3)	0.362710	0.321525	1.128093	0.3024
SIGMASQ	105.0102	77.86817	1.348564	0.2262

R-squared	0.529015	Mean dependent var	-1.363636
Adjusted R-squared	0.215025	S.D. dependent var	15.66060
S.E. of regression	13.87511	Akaike info criterion	8.499786
Sum squared resid	1155.112	Schwarz criterion	8.680647
Log likelihood	-41.74882	Hannan-Quinn criter.	8.385778
F-statistic	1.684814	Durbin-Watson stat	1.917175
Prob(F-statistic)	0.270287		

Inverted AR Roots	.65	-.66-.35i	-.66+.35i

Figure 4. Equation Specification

3.2 Prediction using Exponential Smoothing

Time series model with exponential smoothing - Holt-Winters method with visit data plot. Referring to Figure 5, it can be seen that the red data pattern (Holt-Winters result) smooths out the seasonal data pattern that is present in the actual data, particularly in 2019. Descriptively, the number of visit dates will increase. Figure 6 shows the value of and the value = 0.52, = 0.200 and = 0.5384.

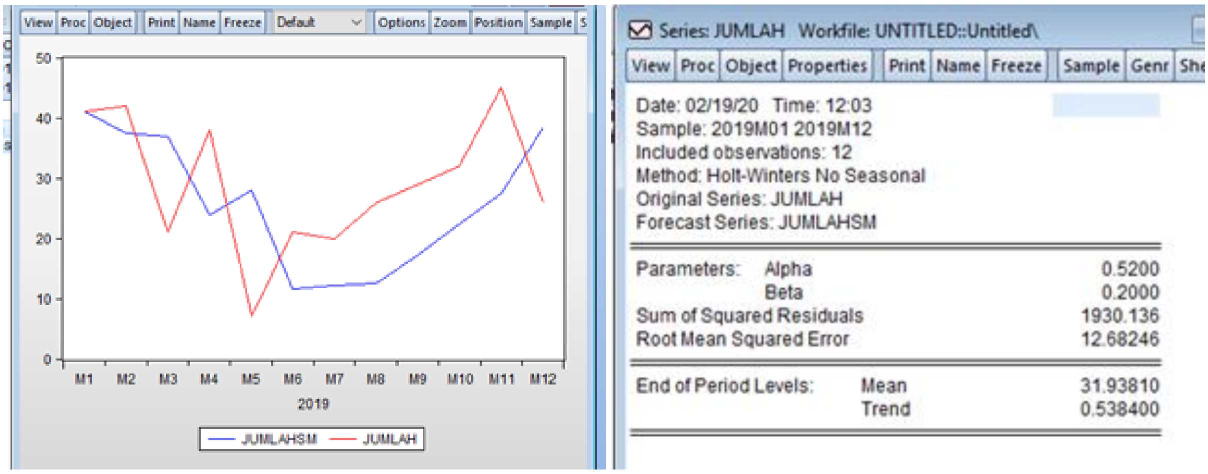


Figure 5. Output of Exponential Smoothing Results and Output of α, β and γ Values

The selection of the Holt-Winters exponential smoothing model is tested for autocorrelation. Autocorrelation tests for errors (residuals) must be performed to determine whether the errors (residuals) are random and there is no true correlation between errors. The 12 models were tested for correlation between errors (residuals). In the autocorrelation test, the larger the values of beta () and gamma (), the correlation occurs, so the model is not called the best model in autocorrelation. The model is considered the best model if it has the smallest RMSE value and no autocorrelation with the criterion that the chi-squared probability value of the R-squared observation is greater than or equal to the p-value of 0.05.

3.3 Effectiveness of Methods for Forecasting Visit Data

The following are the results of forecasting the next three months (13th, 14th, and 15th data) based on data from January 2019 to December 2019. Forecasting is done using the ARIMA and Exponential Smoothing methods.

Table 1. Forecasting Results

Forecasting	January (13)	February (14)	March (15)
ARIMA	40.01	39.03476	38.05214
EXPONENTIAL SMOOTHING	41	37.5	36.8

The MSE (Mean Square Error) value of the forecast results using the ARIMA and Exponential Smoothing methods is presented in Table 1. MSE is used to measure the accuracy of the estimated model value, which is expressed in mean square error.

Table 2. MSE Value

Method	MSE Value
ARIMA	11.160
Exponential Smoothing	11.475

Selecting the best model for the effectiveness of the two methods can use the MAPE score in addition to using the MSE score. MAPE is used to measure the accuracy of the model's estimated value, expressed in terms of an average absolute percentage error. The MAPE (mean absolute percentage error) value of the forecast results using the ARIMA method and the exponential smoothing method is presented in Table 3.

Table 3. MSE Value

Method	MAPE Value
ARIMA	48.949
Exponential Smoothing	47.197

Table 3 shows that the error of the ARIMA method is smaller than the exponential smoothing method, so the best model for predicting visit data in ACEP is using the ARIMA method. In the discussion above there are 12 months of data, while the following calculations are made using the SPSS software, based on the amount of data obtained and by observing the comparison of MSE and MAPE values in each method.

Table 4. Comparison of MSE and MAPE Values Based on Time Data

Comparison	Time 12 months (1 year)		Time 24 months (2 years)		Time 36 months (3 years)	
	Value MSE	Value MAPE	Value MSE	Value MAPE	Value MSE	Value MAPE
ARIMA	11.160	48.949	12.256	78.016	12.104	68.182
Exponential Smoothing	11.475	47.197	12.475	76.537	12.475	66.134

From Table 4 it can be seen that by using different data sets it can be seen that the exponential smoothing method has a smaller MAPE value than the MAPE value of the ARIMA method. The comparison in Table 4 shows that the exponential smoothing method is more effective in predicting the next visit.

4. Conclusion

The effectiveness of the model is evident from the MSE and MAPE scores for each method. Predicting the data on the number of visits to the Amikom Creative Economy Park using the Holt-Winters exponential smoothing method yields a Mean Square Error (MSE) value of 11.475 and a Mean Absolute Percentage Error (MAPE) value of 47.197. While the ARIMA method produces a Mean Square Error (MSE) value of 11,160 and a Mean Absolute Percentage Error (MAPE) value of 48,949. Therefore, predicting the number of visits data for the period January to December 2019 using the Holt-Winters exponential smoothing method is more effective than the ARIMA method because the MAPE value is smaller than the MAPE value produced by the ARIMA method. Value.

References

[1] Direktorat Kehumasan dan Urusan Internasional Universitas AMIKOM Yogyakarta, 2019. http://www.humas.amikom.ac.id. [02 Januari 2020].
[2] Suyanto, Mohammad., Wibowo, F.W. “Animation Opportunities of Intellegent Multimedia System in Developing a Creative Economy Park”. International Conference on Information and Communications Technology (ICOIACT) pp. 69-74., 2018.
[3] Anggriningrum, D.P., Hendikawati, P., dan Abidin, Z. 2013. Perbandingan Prediksi Harga Saham dengan Menggunakan Jaringa Syaraf Tiruan Backpropagation dan ARIMA. Unnes Journal of Mathematics. 2(2): 105-109.
[4] Machmudin, A., Ulama, B.S.S. 2012. Peramalan Temperatur Udara di Kota Surabaya dengan Menggunakan ARIMA dan Artificial Neural Network. Jurnal Sains dan Seni ITS. 1(1): 118-123.
[5] Istiqomah, Widia., Darsyah, M. Yamin., “Efektifitas Metode Arima dan Exponential Smoothing Untuk Meramalkan Nilai Tukar Petani di Jawa Tengah”, Seminar Nasional Mahasiswa Unimus Vol. 1, 2018.
[6] Hendranata, Anton. ARIMA (Autoregressive Moving Average), Manajemen Keuangan Sektor Publik FEUI, 2003.
[7] Makridakis, Spyros. , Steven C. Wheelwright, dan Victor E. McGee. Metode dan Aplikasi Peramalan, Jakarta: Erlangga, 1999.
[8] M. Suyanto, The Success Secrets of Taiwan’s Worldwide SME in Creative Industries. Yogyakarta: Penerbit Andi, 2014.
[9] Hanke, J.E., danWinchern, D.W. 2005. Business Forecasting. New Jersey: Pearson Education International.
[10] Aziz, Syarfi., dkk. 2017. Penerapan Metode ARIMA untuk Peramalan Pengunjung Perpustakaan UIN Suska Riau. Seminar Nasional Teknologi Informasi, Komunikasi dan Industri(SNTIKI) 9, FST, UIN Sultan Syarif Kasim Riau.
[11] Haryati, Siska. 2016. Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). Jurnal Media Infotama Vol. 11 No. 2, September 2015, hlm 130-138.
[12] Lingitz, Lucas. Dkk. 2018. Lead time prediction using machine learning algorithms : A case study by semiconductor manufactur. Procedia CIRP 72 (2018) 1051 – 1056. Elsevier.
[13] Putro, Bossarito, dkk. 2018. Prediksi Jumlah Kebutuhan Pemakaian Air Menggunakan Metode Exponential Smoothing. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol. 2, No. 11, November 2018, hlm. 4679-4686.
[14] Safitri, Tias. 2017. Perbandingan Peramalan Menggunakan Metode Exponential Smoothing Holt-Winters Dan Arima. UNNES Journal of Mathematics UJM 6 (1), hal. 48-58.
[15] Suyanto, Dr. 2017. Data Mining Untuk Klasifikasi dan Klasterisasi Data. Bandung: Informatika.

[16] Safitri, Tias., Dwidayati, N., Sugiman. 2017. Perbandingan Peramalan Menggunakan Metode Exponential Smoothing Holt Winters dan ARIMA. UNNES Journal of Mathematics, UJM 6(1) 2017.