International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 17 (2017) pp. 6806-6813

The Performance of ARIMAX Method in Forecasting Number of Tuberculosis Patients in Malang Regency, Indonesia

Wiwik Anggraeni[1], Niswati Pusparinda[2], Edwin Riksakomara[3], Febriliyan Samopa[4], Pujiadi[5]

Department of Information Systems, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia.
Department of Health, Malang, East Java, Indonesia.

[1]Orcid: 0000-0003-3010-3062

Abstract

The number of tuberculosis (TB) patients in Malang regency is ranked fourth in East Java with the number of cases as many as 1,932 in 2015. Given the high number, it is necessary to forecast the number of TB patients to know the number of TB patients in the coming period. This research is done to forecast the number of TB patients using ARIMAX and ARIMA method. In ARIMAX, the forecasting is done by involving the influence of other variables ie temperature and humidity. The data used in forecasting is monthly data period from January 2007 to September 2016 for each variable involved. ARIMAX method is more suitable to forecast the number of tuberculosis patients compared with ARIMA method. The ARIMAX method results in a lower MAPE value than using the ARIMA method. The result of model evaluation using MAPE proves ARIMAX model (3,1,0) gives very good forecasting result that is 4,70% and the ability average to follow a data pattern of ARIMAX method is 86.37%. The results of this forecasting can provide benefits especially for the health office of Malang Regency to determine the planning needs to tackle the case of TB.

Keywords: ARIMAX, Forecasting, Performance, Tuberculosis, Air humidity, Air temperature.

INTRODUCTION

Tuberculosis (TB) is an infectious disease caused by Mycobacterium Tuberculosis Bacillus [1]. Based on WHO data through Global Tuberculosis Report, the disease has resulted in the deaths of 1.5 million people by 2014. From these data, it can be seen that the number of TB is 300,000 is more than the number of deaths caused by HIV [2]. In Indonesia TB disease is ranked fourth as a deadly disease after Cerebrovascular, ischemic heart disease and Diabetes Mellitus with complications [3]. East Java Province was ranked second in Indonesia as the province with the highest TB disease cases after West Java province. According to Surabaya Tribun News TB cases found in East Java reached 40,185 cases and 2,475 of them were infectious pulmonary TB. Malang Regency which is one of districts in East Java with a population of 2,544,315 people [4] is the district with the fourth rank of TB cases in East Java with 1,932 cases, the rank is obtained after Surabaya with 4,754 cases, Jember 3,128 cases, and Sidoarjo with 2,292 cases [5]. Looking at the number of TB cases occur, making many researchers do forecasting to know the number of TB patients in the next periods. The results of it can be used to determine the planning needs to tackle TB cases. Research related to predicting the number of TB cases has been carried out by Mahmood Moosazadeh et al. in 2014 [6]. The forecasting done was to predict the number of cases of TB in Iran by using the method of Seasonal Autoregressive Integrated Moving Average (SARIMA). By 2015, Mahmood Moosazadeh et al. conducted research on Predicting the Incidence of Smear Positive Tuberculosis Case in Iran Using Time Series Analysis [7]. In the study Mahmood Moosazadeh and colleagues concluded that the SARIMA method is a suitable method for forecasting TB cases in Iran. Similar research was also done by Nothabo Dube [8]. They predict the number of TB incidents by 2015 in Zimbabwe. In the study, Nothabo did the forecasting by comparing three methods of Autoregressive Integrated Moving Average (ARIMA), Autoregressive Integrated Moving Average-Autoregressive Conditional Heteroscedasticity (ARIMA-ARCH), and Holt Winter. The results of the research show that ARIMA method is better than other methods [8].

The following research was held by Adeboye Azeez, et al. in 2016 [9]. In the study, Adeboyee Azeez and colleagues forecasted TB in the Eastern Cape by comparing the SARIMA method with the Seasonal Autoregressive Integrated Moving Average-Neural Network Auto-Regression (SARIMA-NNAR). From the research, it was found that SARIMA-NNAR combination method showed better result than SARIMA [9].

The ARIMAX method is an extension method of ARIMA [10][11]. In forecasting, this method is not only pays attention to one variable alone, but also involves independent variables that will be used as eXogen variable [10]. The independent variables intended in this method are external factors that can affect the number of objects to be predicted [10]. According to Khairunnisa Siregar in his research in 2015, the temperature and humidity factors have a significant correlation in the influence of TB cases [12]. Subsequent research on tofu 2015

by Chadsuthi Sudarat et al. also showed that climate factors influenced the growth of influenza disease and then modeled the transmission using ARIMAX. [10]. In that year, there were other studies using ARIMAX method for Predicting Pest Population by Prawin Arya et al. [12].

Anggraeni et.al in 2017 also used ARIMAX method to predict the number of dengue fever patients in Indonesia by using google trend data and the results show that ARIMAX is suitable for trendy and irregular patterned data [13]. Then ARIMAX has also been used to forecast data that has a seasonal pattern and proves to be more suitable than ARIMA method [11].

According to the previous explanation, it is shown that the ARIMAX method works well for predicting data and forecasting the number of TB patients which is needed. The results of the forecasting are used as the supporting base of early warning system that is needed by the Health Office of Malang Regency to do anticipation planning soaring the number of TB patients. Given the fact that the development of TB is also strongly influenced by the temperature and humidity factors and previous research has shown that temperature and humidity have an effect on disease growth too [14], then in this study conducted the ARIMAX method to predict the number of TB patients in Malang regency with temperature and humidity as eXogen variables. Hope after doing this research, Malang Regency especially Health Office of Malang Regency can know the prediction number of TB case, so that government can do planning requirement to tackle TB case.

The contribution of this paper are:

1. The related studies mentioned above all forecast based on the number of cases occurring in each period either weekly or monthly [6] [7] [8] [9]and none have used ARIMAX and involves the influence of other variables.
2. The forecasting is done by involving the influence of other variables ie temperature and humidity.
3. The region of case study have specific characteristics, so the data which obtained will have different characteristics as well as data from other regions.

LITERATURE REVIEW

A. Tuberculosis

Tuberculosis (TB) is an infectious disease, which attacks the pulmonary parenchyma [1]. The cause of TB disease is a complex bacterium called Mycobacterium Tuberculosis Bacillus Invalid source specified Bacteria are 90% more often infect the lung organ than other organs in the human body and these bacteria including bacillas are very strong so that to treat patients due to these bacteria require a long time [1]. The higher RE value in a region indicates the region's, the better electricity availability. The percentage of the electrification ratio of 100 or more states that electricity has flowed throughout the household.

B. ARIMAX Method

ARIMAX is an ARIMA model involving eXogen variables. This model is a forecasting method that takes into account the independent variables. The general form of ARIMAX model (p, d, q) is [10] [11]:

$$(1-B)^d \emptyset_p(B) Y_t = \theta_q(B) e_t + a_1 X_{1,t} + a_2 X_{2,t} + \cdots + a_k X_{k,t}$$

C. Stationery Test

A time series data can be said to be stationary if the autocorrelation coefficient in all lags does not statistically show significant difference from zero. The stationarity test of a time series data can be performed in the following way [15][16]:

1. Detection in unstable data in averages using plots from the original data, autocorrelation function plot (ACF) and partial autocorrelation function plot (PACF). If the data already possessed contains a trend component, slowly non-stationary data in the mean and plot ACF and PACF will slowly decay.
2. Detection of unstated data in variants can be done using ACF and PACF plots of residual squares. In addition to these ways, to see it can be done by looking at the value of variance.

D. Differencing and Transforming

The process of discrimination is important if the data held is not stationary in the mean. The notation used is a reverse shift operator or that can be symbolized by B, by means of the following usage [15]:

$$BY_t = Y_{t-1}$$

The transformation process can be done if the data is not stationary in variance. One of the most commonly used transformations in this regard is the Box-Cox transformation.

E. Linearity and Multicollinearity Test

Linearity test is a test conducted to determine the relationship between dependent variable and independent variable. Linearity test can be done using stability diagnostics function with Ramsey RESET test method. A variable can be said to pass the linearity test if it has probaility value> 0,05 [16].

Multicollinearity test is a test conducted to determine the relationship between independent variables. The

multicollinearity test can be performed using the VIF (Variance Inflation Factors) method. How many variables cannot be said to pass multicollinearity test or said there is a connection between variables if it has VIF value > 10 [16].

F. Diagnostic Test

Diagnostic examination can be done in the following way [15][16]:

- White Noise Assumption Test

The White Noise assumption test is one of the testing methods used to find out the residual and homogeneous independent of residuals (homogeneous variance) [16]. To be able to see whether residual et is white noise or not, it can be done by performing serial correlation test, ie by testing the hypothesis with the following stages [15][16]:

- Hypothesis Test

  H_0: at least one $\rho_k \neq 0$ (model has not fulfilled white noise assumption)

  H_1: $\rho_1 = \rho_2 = ... = \rho_k = 0$ (the model meets the white noise assumption)

  The level of significance = $\alpha = 5\%$

  Statistics Test using Ljung-Box

$$Q = n(n+2)\sum_{k=1}^{m}\frac{\rho k^2}{(n-k)}$$

  m: number of maximum lag, n: number of original observations, ρ_k: autocorrelation coefficient

- Rule of decision making

  H_0 is rejected if $Q > X^2_{(1-\alpha,db)}$ or reject H_0 if p – value > α

  where db = m - M with M = p + q

- Normality test

  In addition to find the white noise assumption, residuals must also meet the normal distribution. Normal residual examination can be done using qq-plot. Normally, distributed data will display split points in a straight line if a plot with qq-plot is performed. If the points are slightly away from the straight line, then the data can still be said to be normal because what it need to be paid attention is the middle of the collection of data points. If the model does not meet these assumptions, then a new model must be redefined, the next estimated and the parameters retested [15][16]:

G. Selection of The Best Model

Once the model has met the assumptions on the diagnostic test, it is possible to obtain some appropriate models. So we can choose the best money model that can be used in forecasting. To select the best model can be done through the Akaike Information Criteria (AIC) Test, as follows [15][16]:

$$AIC = \left(\frac{2k}{n}\right) + \ln\left(\frac{SSE}{n}\right)$$

H. Accuracy Measures Forecasting Method

Mean Absolute Percentage Error (MAPE) is used to select the best method and know the accuracy in doing the forecasting. The MAPE formula is [15][16][17]:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}|PE_t|$$

n is the number of periods and PE_t is the percent error. Percentage Error can written as below [15][16]:

$$PE_t = \left(\frac{X_t - F_t}{X_t}\right) x100\%$$

X_t : observation at period t, F_t : forecast in period t. The smaller the MAPE value the approximate value closer to the actual value, or in other words the method chosen is the best method [16][17]:. A method has a very good performance when the MAPE value is below 10%, and performs well if the value is between 10% and 20% [15][17]:

DATA

Data on the number of TB patients used is obtained from the Health Office of Malang Regency. Data of district air temperature and humidity is gotten from online data of Meteorology, Climatology and Geophysics Agency (BMKG), namely Ploso Malang Climatology Station Indonesia. The data used is data from January 2007 to September 2016 in monthly period. Collected data will be pre-processed data, to ensure the data is ready to be processed. After that, the data is divided into three groups, namely training data by 70%, the first test by 20% and the second test by 10%.

RESEARCH METHOD

A. Data Preprocessing

The data preprocessing is to test outlier data. Outlier test was done using Grubbs test method with 95% significance level. The results of Grubbs test can be seen in Figure 1.

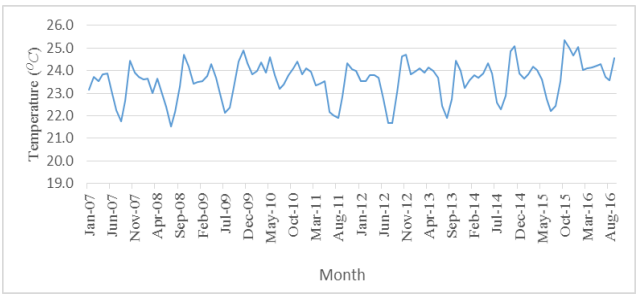
```
Grubbs' Test

Variable    N   Mean  StDev    Min     Max     G      P
BTA+      117  81,02  15,84  40,00  117,00  2,59  1,000

* NOTE * No outlier at the 95% level of significance
```

Figure 1: Outlier test result

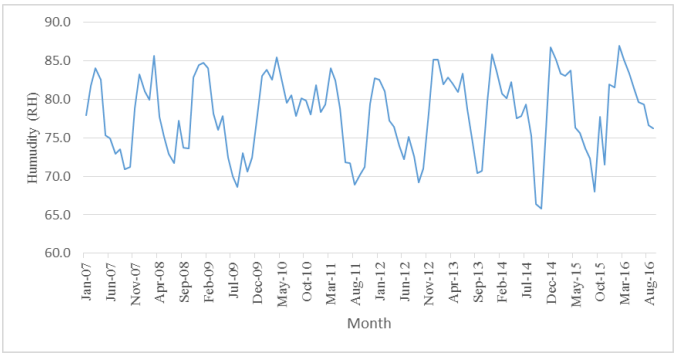
Figure 1. show that in the data used that there is no outlier in it. The increment of tubercolusis patient number can be seen ini Figure 2.

Gambar 2: Pattern of patient number

Figure 2. show that the number of tuberculosis sufferers in Malang district has been increasing trend anytime. Figure 3. demonstrates that in the average temperature data there is an upward trend and there is a seasonal factor, where the location of the seasonal occurs every 12 months. Figure. 4. deliberates that the average humidity data tends to be stable every year but there is a seasonal, where the seasons occur every 12 months.



Gambar 3: Pattern of temperature

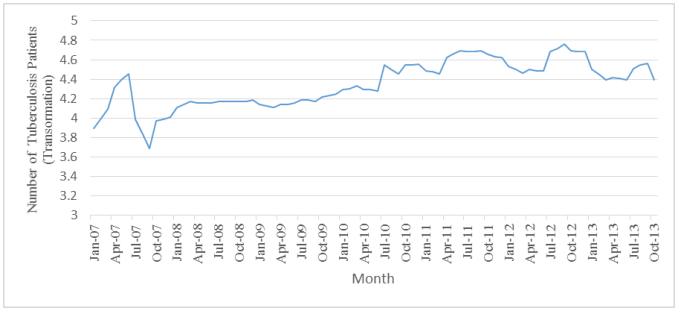


Gambar 4: Pattern of humidity

The data used for the training data is 70% of the total data owned by 82 data, from January 2007 - October 2013. Test data of 20% or 23 data from November 2013 to September 2015. Validation data of 10% or 12 data, from October 2015 to September 2016.

B. Variable Stationery Test

Variable stationary tests were performed using Barlett test with a significance level of 95%. If the result of rounded value shows ≠ 1, then the data is not stationary in variance, so data transformation must be done using Box-Cox. Data transformation result are shown is Figure 5.

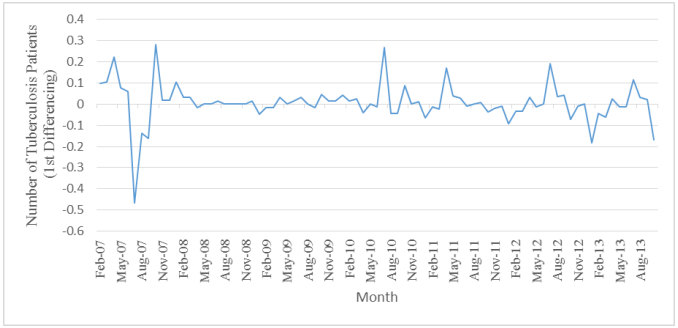


Gambar 5: Pattern of patient number transformation data

Figure 5. indicates that data on the number of tuberculosis sufferers after undergoing data transformation experienced an insignificant upward trend over time.

C. Average Stationery Test

An average stationary test was performed to determine the average data rate. The test is performed using the unit root test function. Data is said to be stationary if probability value ≤ 0.05 and value | t-Statistic | ≥ | test critical values | On the unit root test. The data already possessed required distinction of 1 because the data is not stationary in the mean. The result of the differences in Figure 6. shows that the data has been stationary in a mean.



Gambar 6: Pattern of 1st differencing result

D. Linearity Test and Multicollinearity

Linearity test is done to know the connection between

dependent variable and independent variable. Figure 7. is a result of linearity test between dependent variable and independent variable using RESET Test.

Ramsey RESET Test
Equation: UNTITLED
Specification: D(T_TRAINBTA) SUHU LEMBAB
Omitted Variables: Squares of fitted values

	Value	df	Probability
t-statistic	0.459185	78	0.6474
F-statistic	0.210851	(1, 78)	0.6474
Likelihood ratio	0.218665	1	0.6401

Gambar 7: Linearity test result of dependent variable and independent variable

Figure 7. shows that, between the dependent and independent variables are connected, since the resulting probability value is more than 0.05.

Multicollinearity test is done to know the relation between independent variable. Figure 8. demonstrates that the test results indicate a number of 289,2534 on the uncentered VIF, so that the independent variables did not pass in the multicollinearity test and inter independent variables are related because they have uncentered VIF value more than 10.

Variance Inflation Factors
Date: 12/17/16 Time: 15:19
Sample: 1 82
Included observations: 82

Variable	Coefficient Variance	Uncentered VIF
SUHU	2.275958	289.2534
LEMBAB	0.207186	289.2534

Gambar 8: Multicollinearity test result of independent variable

E. ARIMA Modelling

ARIMA modeling was performed after the data was passed in stationary and stationary test in the variety. Then, do some suitable models and perform significance tests, as well as diagnostic tests on the ARIMA models. Table 1 shows the significance test results on the models, the model that has passed in the significance test will be continued for the diagnostic test.

Table 1: The ARIMA significance test results

ARIMA Model	Significance Test		
	t-stat	Prob	Explanation
ARIMA(0,1,1)	0,793	0,430	Didn't Pass
ARIMA(0,1,2)	0,791	0,431	Didn't Pass
ARIMA(0,1,3)	-6,558	0,000	Passed
ARIMA1,1,0)	1,023	0,308	Didn't Pass
ARIMA(1,1,1)	0,211 -0,064	0,833 0,949	Didn't Pass
ARIMA(1,1,2)	0,884 0,655	0,379 0,524	Didn't Pass
ARIMA(1,1,3)	0,643 -6,433	0,522 0,000	Didn't Pass
ARIMA(2,1,0)	0,639	0,524	Didn't Pass
ARIMA(2,1,1)	0,828 0,930	0,409 0,355	Didn't Pass
ARIMA(2,1,2)	-0,107 0,152	0,914 0,880	Didn't Pass
ARIMA(2,1,3)	0,339 -6,473	0,736 0,000	Didn't Pass
ARIMA(3,1,0)	-5,928	0,000	Passed
ARIMA(3,1,1)	-6,015 0,650	0,000 0,518	Didn't Pass
ARIMA(3,1,2)	-5,864 0,596	0,000 0,553	Didn't Pass
ARIMA(3,1,3)	-0,288 -2,717	0,774 0,008	Didn't Pass

In Table 1, it can be seen that the ARIMA model (0,1,3) and ARIMA (3,1,0) is a model that passes the significance test because it has value | t-stat | Above 0.05 and probabilistic below 0.05. So that model can be continued the next stage that is diagnostic test. The diagnostic test results in Table 2. Show that the two models that have passed the significance test have also passed in the diagnostic test because they have a probability value on the randomness test and homogeneity test above 0.05. So both models are feasible for use in forecasting.

Table 2: The ARIMA diagnostic test results

Model	Residual Randomness Test		Homogeneity Test		Explanation
	Average Prob.	Explanation	Average Prob.	Explanation	
ARIMA (0,1,3)	0,577	Passed	0,369	Passed	Feasible
ARIMA (3,1,0)	0,139	Passed	0,335	Passed	Feasible

F. ARIMAX Modelling

ARIMAX modeling was performed after the data was passed in significance test, and diagnostic test on ARIMA modeling. Then in the ARIMA model are included independent variables and tested significance, as well as diagnostic tests on the ARIMAX model. Table 3 shows the significance test results in the ARIMAX model. The model that has passed in the significance test will be continued for diagnostic test.

Table 3: The ARIMAX model Significance test result

Model ARIMAX	Significance Test		
	t-stat	Prob	Explanation
ARIMAX(0,1,3)	-4,739	0,000	Passed
ARIMAX(3,1,0)	-5,387	0,000	Passed

In Table 3, it can be seen that the model of ARIMA (0,1,3) and ARIMA (3,1,0) is a model that pass the significance test because it has value | t-stat | Above 0.05 and probabilistic below 0.05, so that model can be continued the next stage that is diagnostic test.

The results of the diagnostic tests in Table 4 indicates that the two models that have passed the significance test have also passed in the diagnostic test because they have a probability value on randomness test and homogeneity test above 0.05. So both models are feasible for use in forecasting.

Table 4: The ARIMA Model Diagnostic Test Result

Model	Residual Randomness Test		Homogeneity Test		Explanation
	Average Prob.	Explanation	Average Prob.	Explanation	
ARIMAX (0,1,3)	0,638	Passed	0,491	Passed	Feasible
ARIMAX (3,1,0)	0,151	Passed	0,392	Passed	Feasible

RESULT AND DISCUSSION

ARIMAX models that have met the model feasibility test are forecasted in all three data groups. The results of model and validation tests in Table 5 indicates the MAPE value for each model <10%, so the ability of forecasting is very good and the model can be used for further forecasting.

Table 5: The MAPE of ARIMAX Model

Model ARIMAX	Training	1st Validation	2nd Validation
ARIMAX(0,3,1)	5,40%	3,82%	2,70%
ARIMAX(3,1,0)	5,17%	3,02%	2,41%

The results of the error analysis in forecasting (MAPE) in Table 5 shows that the ARIMAX model (3.1,0) has better accuracy than other models, so the ARIMAX model (3.1,0) can be used for data forecasting and generate forecasts Tuberculosis patients in October 2016 - December 2017.

To determine the performance of data forecasting results, analysis of the ability to follow the data pattern must be done.

The result of capability analysis follows the data pattern of Table 6 demonstrating that the two ARIMAX models produced have the ability to follow a good pattern that is more than 70%.

Table 6: Ability to Follow Data Pattern

Model ARIMAX	Training	1st Validation	2nd Validation
ARIMAX(0,3,1)	82,72%	72,73%	72,73%
ARIMAX(3,1,0)	83,33%	73,68%	100%

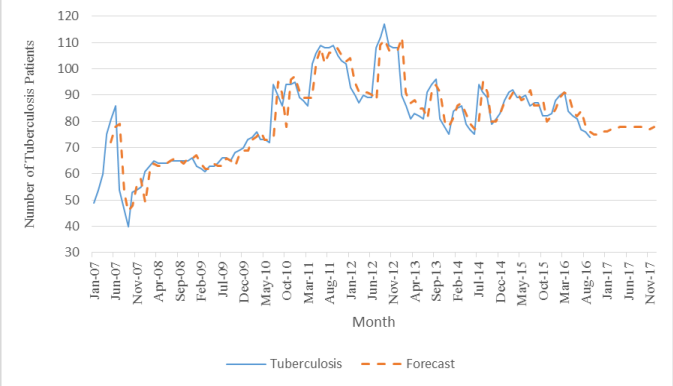
To know the quality of method used in forecasting, hence method compared with comparison method that is ARIMA method. The chosen arima method is the ARIMA method that has passed significance and diagnostics tests, so the ARIMA model is feasible.

Table 7: The MAPE comparison between ARIMAX and ARIMA model

Model ARIMAX	Training	1st Validation	2nd Validation
ARIMAX(3,1,0)	5,17%	3,02%	2,41%
ARIMA (0,1,3)	5,23%	3,44%	2,87%
ARIMA (3,1,0)	5,35%	3,19%	3,07%

Based on the comparison of the three data groupings in Table 7, it was found out that ARIMAX method with ARIMAX model (3,1,0) is the best method that can be used to forecast the number of tuberculosis patients in Malang Regency. It can be seen based on the MAPE value generated that MAPE value of ARIMAX (3.1,0) always yields the smallest when it is compared with other comparison methods in all three groups of data.

Figure 9. illustrates that the result of forecasting using ARIMAX model (3,1,0) obtain a good result because the data forecasting results coincide with the actual data.



Gambar 9: The number of patient forecasting result

Multicollinearity test results demonstrate that inter independent variables have dependence on each other, and it can affect forecasting result. To know the effect of the multicollinearity, then analyze the influence of multicollinearity which is summarized in Table 8.

Table 8: The MAPE comparison between ARIMAX model with 1 and 2 independent variables

Independent Variabel	Model	Trai-ning	1st Validation	2nd Validation
Temperature, Humidity	ARIMAX(3,1,0)	5,17%	3,02%	2,41%
Temperature	ARIMAX (0,1,3)	4,99%	3,41%	2,81%
	ARIMAX (3,1,0)	5,12%	2,98%	2,41%

Based on the MAPE comparison performed on the training data, the first test and the second test, summarized in Table 8, it's show that ARIMAX forecasting uses one variable (the average of temperature) has smaller MAPE values when compared to the ARIMAX model using two variables. It also can be seen in comparison of MAPE value in first training data and testing data, MAPE value of ARIMAX using one variable smaller when compared using two variables. While in the second testing data, ARIMAX model (3.1,0) uses one variable and two variables have the same MAPE value. The result of MAPE comparison have been done and proof that If the independent variables affect each other then the quality of forecasting results obtained will be less good.

CONCLUSION

Based on the results of this research, it is concluded that ARIMAX method is better used in forecasting the number of tuberculosis patients compared with ARIMA method. ARIMAX (3,1,0) has a very good rate of 4.70%. The ability average to follow a data pattern of ARIMAX method is 86.37%. Forecasting results for the October 2016 - December 2017 period shows that the number of people with tuberculosis in Malang district gradually increased. Forecasting uses the average temperature variable only, yielding a higher level of accuracy when compared using the average temperature and air humidity variables. It occurs because of the influence of multicollinearity among independent variables. For further research, forecasting should be done using independent variables that are not mutually dependent on each other. This can worsen the accuracy of forecasting results. If it finds a dependency in independent variables, it should be removed.

ACKNOWLEDGMENT

The authors thanks to Ministry of Research, Technology, and Higher Education, Republic of Indonesia for funding this research under Excellent Research Universities grant schema and the Health Office of Malang Regency which has supported the implementation of this research.

REFERENCES

[1] Nicole Fogel, Tubercolusis: A Disease Without Boundaries, Tuberculosis, Volume 95:5, 2015, pp. 527-531

[2] AS and LT, "WHO: TBC Akibatkan Kematian Setara dengan HIV," 05 November 2005. [Online]. Available: http://www.voaindonesia.com/a/who-tbc-akibatkan-jumlah-kematian-setara-dengan-hiv/3036848.html. [Accessed January 17, 2017].

[3] U. Widowati, "10 Penyakit Paling Mematikan di Indonesia," 14 Mei 2015. [Online]. Available: http://www.cnnindonesia.com/gaya-hidup/20150513163407-255-53129/10-penyakit-paling-mematikan-di-indonesia/. [Accessed january 17, 2017].

[4] Kabupaten Malang, "Selayang Pandang," 10 Mei 2016. [Online]. Available: http://www.malangkab.go.id/site/read/detail/79/selayang-pandang.html. [Accessed September 19, 2016].

[5] S. Sofiana, "Penderita TB di Surabaya Tertinggi di Jatim, Kedua Jember, dan Ketiga Sidoarjo, Ini Data Lengkapnya," 23 Maret 2016. [Online]. Available: http://surabaya.tribunnews.com/2016/03/23/penderita-tb-di-surabaya-tertinggi-di-jatim-kedua-jember-dan-ketiga-sidoarjo-ini-data-lengkapnya. [Accessed September 19, 2016].

[6] Moosazadeh, M., Nasehi, N. Khanjani, A. B. N., Sharafi, S., and Ahmadi, S, 2014, "Forecasting

Tuberculosis Incidence in Iran Using Box-Jenkins Models".

[7] Moosazadeh, M., Nasehi, N. Khanjani, A. B. N., and Ahmadi, S,, 2015, "Predicting Incidence of Smear Positive Tuberculosis Case in Iran Using Time Series Analysis".

[8] Dube, N., 2015, "Aplication and Comparation of Time Series Methods on Tuberculosis Incidence Data: A case study of Zimbabwe 1990-2013".

[9] Azeez, A., Obaroni,D., Odeyemi, A. Ndege, J., and Muntabayi, R., 2016, "Seasonality and Trend Forecasting of Tuberculosis Prevalence Data in Eastern Cape, South Afrika, Using Hybrid Model".

[10] S. Chadsuthi, S. Iamsirithaworn, W. Triampo and C. Modchang, 2015, "Modeling Seasonal Influenza Transmission and Its Association with Climate Factors in Thailand Using Time-Series and ARIMAX Analysis".

[11] Anggraeni, W., Vinarti, R.A., and Kurniawati,Y.D.,2015,"Performance Comparisons Between ARIMA and ARIMAX Method in Moslem Kids Clothes Demand Forecasting: Case Study", Procedia Computer Science, 72, pp. 630-637.

[12] Arya,P. Kumar Paul, R. Kumar, A., Signh, K.N., Sivaramne, N. and Chaudhary,P., 2015, "Predicting Pest Population Using Weather Variables: An ARIMAX Time Series Framework".

[13] Anggraeni, W., Aristiani, L.,2016,"Using Google Trend Data in Forecasting Number of Dengue Fever Cases with ARIMAX Method Case Study: Surabaya, Indonesia", 2016 International Conference on Information, Communication Technology and System, Surabaya, pp 114-118.

[14] Anggraeni,W., Riksakomara, E., Samopa, F. Wibowo, R..P., Vinarti, R.A.,, and Surahman Arif, 2015, "Prediction The Appearance of Dengue Fever Dsease Based on Weather Using Rule Base System", International Journal of Applied engineering Research, 10, pp. 75-79

[15] Spyros Makridakis, Steven C. Wheelwright, and Rob J. Hyndman, Forecasting: Methods and Applications, 3rd ed.: John Wiley & Sons, 1998

[16] O'connell Bowerman and Koehler , Forecasting, Time Series, and Regression, 4th ed. United States of America: Curt Hinrichs, 2005.

[17] Anggraeni,W., Aditya I.P.A., Samopa, F., Riksakomara, E., Wibowo, R.P., Trikoratno, L.C., Pujiadi. Forecasting The Number of Dengue Fever Cases in Malang Regency Indonesia Using Fuzzy Inference System Models. Journal of Theoretical and Applied Information Technology 2017; **95**:1-8.