THE ELECTION OF THE BEST AUTOREGRESSIVE

by

Advances and Applications in Statistics
© 2018 Pushpa Publishing House, Allahabad, India
http://www.pphmj.com
http://dx.doi.org/10.17654/AS052040251
Volume 52, Number 4, 2018, Pages 251-265 ISSN: 0972-3617

THE ELECTION OF THE BEST AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODEL TO FORECASTING RICE PRODUCTION IN INDONESIA

Georgina Maria Tinungki

Department of Mathematics
Faculty of Mathematics and Natural Science
Hasanuddin University
Makassar 90245, Indonesia

Abstract

The objective of this research is to identify the best autoregressive integrated moving average (ARIMA) model for forecasting rice production in Indonesia. Rice crops are the most important agricultural commodity in Indonesia, because rice is the staple food of Indonesians. Therefore, the ability to forecast rice production is crucial. We used autoregressive integrated moving average (ARIMA) model to predict rice production in Indonesia. The data analyzed is the data of rice production in Indonesia from 1993 to 2012. The data of 2013 to 2015 got used to validate the forecasting results. The results show that the ARIMA model (2, 2, 0) is the best model for forecasting rice production in Indonesia.

1. Introduction

The main purpose of forecasting model research in statistics is a solution of a forecasting problem. The autoregressive integrated moving average

Received: September 18, 2017; Revised: January 15, 2018; Accepted: February 9, 2018
Keywords and phrases: agriculture, ARIMA model, rice production, time series model.

(ARIMA) model is a forecasting model that completely ignores independent variables in making forecasting. ARIMA is often also called the *Box-Jenkins time series method*. ARIMA is very good for short-term forecasting, while for long-term forecasting, the accuracy of forecasting is not good. Usually, it will tend to flat or constant for a long period. ARIMA uses past and present values of the dependent variable to produce accurate short-term forecasting. ARIMA is suitable if the observations of the time series are statistically related to each other (Bharathi et al. [2]).

Stationarity means that there is no growth and decline in data. A data can be said to be *stationary* if the data pattern is at equilibrium around the constant mean value and the variance around the mean is constant for a certain time (Prajneshu et al. [9]).

An eligible stationary time series is explained by $(Z_t) = E(Z_{t-k}); (Z_t) = Var(Z_{t-k})$; and $Cov(Z_t, Z_s) = Cov(Z_{t-k}, Z_{t-s})$. So if a data does not meet these requirements, then the time series is not stationary and to make the stationer needed differencing process (Paul and Das [8]).

The process of differentiation can be performed several times. Here is the ARIMA model that has experienced the differencing process on the one-time differentiation that is ARIMA (p, d, q) with value $d = 1$.

The time series model in the paddy production data can be determined by looking at the result of plot PACF (partial autocorrelation function) and ACF (autocorrelation function) from data which has stationer (Bolton [3]).

Gamal [6]: Indonesian National Logistics Agency or BULOG is an Indonesian government institution in charge of maintaining the stability of the price and the availability of basic food in Indonesia especially rice. Application of forecasting models is also used for the availability of food stock especially basic food such as rice. Stock of rice is very important in maintaining the stability of the food in those countries where the majority of the population consumes rice, such as Indonesia. Majority of Indonesian people consumes rice and makes it as the main basic foodstuffs.

2. Literature Review

There are several mathematical models that can be used to predict the availability of rice stocks in Indonesia, among others:

ARIMA model

ARIMA model is one of time series forecasting techniques which is only based on variable data behavior observed. The ARIMA model completely ignores the independent variables because it uses the present values and past values of the dependent variable to produce accurate short-term forecasting. Literally, ARIMA model is a combination of AR model (autoregressive) and MA model (moving average).

There are several stages in doing the forecasting using ARIMA model that is:

1. Model identification

The model identification is done to examine the autocorrelation and the stationary meaning of the data, so that the need for transformation or differencing process is done. Steps in identifying the model are:

(1) Map the data on the time and study of the data character to determine whether or not the transformation variance and / or distinguishing process is performed.

(2) Calculate and review the ACF and PACF original sample data (data before transformation and / or discrimination) to obtain information on the order of the differentiation process.

(3) Calculate and study ACF and PACF transformed and / or differentiated data (if any transformation and / or differentiation are used), to estimate autoregressive (AR) and moving average (MA) to be retrieved. General guidance to examine whether the order of the stationary time series models is good enough based on the ACF and its PACF, as follows:

Model	ACF	PACF
$AR(p)$	Exponentially patterned or like a weakened sine wave	The difference in value between lag-1 and the value after the lag-p is quite large (cut off after lag-p)
$MA(q)$	The difference in value between lag-1 and the value after the lag-q is quite large (cut off after lag-q)	Exponentially patterned or like a weakened sine wave
$ARMA(p, q)$	Pattern decreased rapidly after lag-(q-p)	Pattern decreased rapidly after lag-(p-q)

Prajneshu et al. [9]

2. Model parameter estimation

Having obtained one or more temporary models, the next step is to look for estimates for the parameters in the model. Estimates can be done in several ways as follows:

(1) Method of moment.

(2) Maximum likelihood method.

(3) Ordinary least square method (OLS).

3. Diagnostic checking

Diagnostic checking is passed to check whether the estimated model matches the existing data. Diagnostic checking is based on residual analysis. The basic assumption of the ARIMA model is that residuals are normal independent distributed random variables with zero mean constant variance.

As for the ARIMA (p, 1, q) function is stated as:

$$w_t = \varnothing w_{t-1} + \varnothing_2 w_{t-2} + \cdots + \varnothing_p w_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}.$$

For $w_t = z_t - z_{t-1}$, then

$$z_t = (1 + \varnothing_1) z_{t-1} + (\varnothing_2 + \varnothing_1) z_{t-2} + (\varnothing_3 + \varnothing_2) z_{t-3} + \cdots + (\varnothing_p + \varnothing_{p-1}) z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

IMA(1, 1)

$$w_t = a_t - \theta_1 a_{t-1},$$

$$z_t = z_{t-1} + a_t - \theta_1 a_{t-1}$$

ARI(1, 1)

$$w_t = \varnothing w_{t-1} + a_t,$$

$$z_t = (1 + \varnothing_1) z_{t-1} + \varnothing z_{t-2} + a_t.$$

If we make a double distinction then the value $d = 2$, if three times the value $d = 3$, etc.

$$U_t = \nabla z_t = z_t - z_{t-1} \text{ for } d = 1,$$

$$W_t = \nabla^2 z_t = U_t - U_{t-1} \text{ for } d = 2.$$

Before forming an ARIMA model in a time series it must be done 4 processes are executed, i.e., model identification, model estimation, model checking and the last if the model has been formed then we will be able to use the model to predict the future.

3. Research Method

The data used are data taken from Central Statistics Agency Office of South Sulawesi Province on Rice Production Development in Indonesia from 1993 to 2015 (Million Ton). The data analyzed is the data of rice production in Indonesia from 1993 until 2012. The data of 2013 to 2015 is used as a comparison with forecasting results.

Stages of data analysis

Data analysis using ARIMA (autoregressive integrated moving average) method. Here are the stages of data analysis:

(1) Creating time series data plots to see if there is a trend element in the data.

(2) Perform stationary test of data consisting of two tests, namely stationary test of variance and stationary test against mean.

(3) Using Box-Cox transformation if the data has not been stationary to variance.

(4) Doing differencing to the data if not stationary to the mean.

(5) Conduct possible model selection based on ACF and PACF plots of stationer data.

(6) Testing significance and parameter estimation.

(7) Test residual white noise data.

(8) Testing the residual normality of the data.

(9) If there are two or more models that pass the test at points 4.5 and 6, then compare the mean square error (MSE) of the models. And if there is only one model then proceed to point (8).

(10) Forecasting data.

4. Results and Discussion

Stationary data test

Time series modeling is done by plotting data to see the stationarity of data. The plot results as follows:

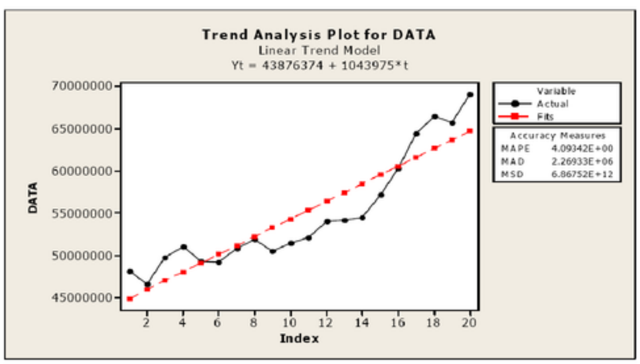

Figure 1. Plot time series data.

Based on Figure 1, it can be concluded that the data has not been stationary in the average because there is an average change over time. In addition to using the time series plot, stationary of data can also be viewed using the ACF plot. Based on Figure 2, it appears that the autocorrelation value tends to fall slowly or down linearly, in other words the value of autocorrelation in a lag is not much different from the previous lag. So it can be concluded using image 2 that the data has not been stationary in the average (Aswi and Sukarna [1]).

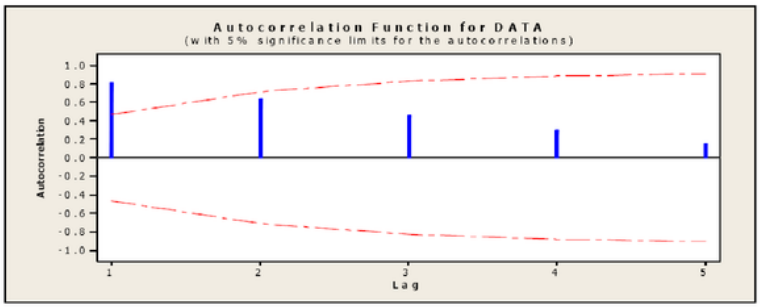

Figure 2. Plot ACF data.

If so, the data cannot be directly used to get the best ARIMA model, but first the data is stationary. But before stationing the data against the mean, the data needs to be stationary to the variance with Box-Cox transformation, after the stationer to the variance, then performs the differencing method to stationary against the average. Differencing is done a maximum of three times (Aswi and Sukarna [1]).

Here is the data plot to see whether the data is stationary to variance or not.

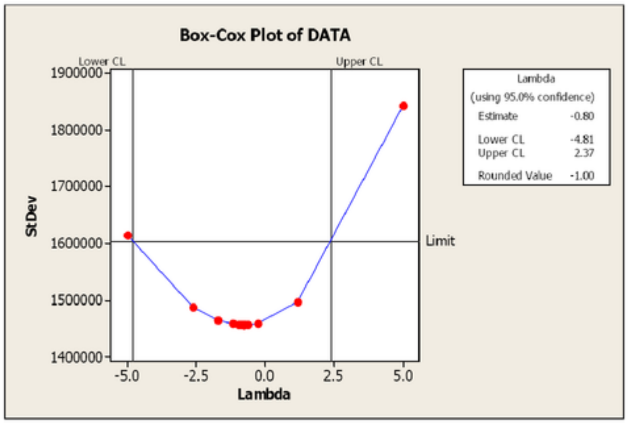

Figure 3. Plot Box-Cox data.

Data is said to be *stationary* in variance if rounded value (Lambda) = 1 in Box-Cox graph. Based on Figure 3, the data has not been stationary to the variance because the Lambda = -1. Therefore, the data will be transformed by:

Table 1. Lambda value and transformation form

Lambda value	Transformation
-1.0	$1/Z_t$
-0.5	$1/\sqrt{Z_t}$
0.0	LnZ_t
0.5	$\sqrt{Z_t}$

Based on Table 1, then the data will be transformed by 1 divided by data. Furthermore, the transformed data will be tested the stationarity by reviewing the lambda values in the Box-Cox graph.

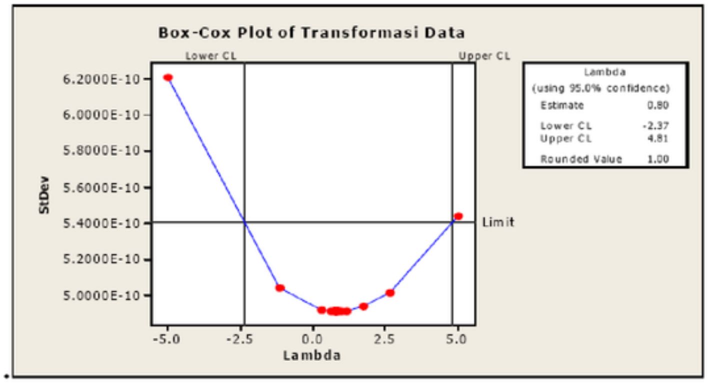

Figure 4. Plot Box-Cox results of data transformation.

Based on Figure 4, the rounded value (Lambda) is equal to 1. This shows that the transformed data has been stationary to its variance.

Next is done differencing to the data so that the data stationer against the average. Box-Jenkins method can be used if stationary data. One technique to stationary the average data is to do differencing, i.e., by subtracting Y_t by Y_{t-1} (differencing first),

$$D_t = \Delta Y_t = Y_t - Y_{t-1}.$$

Data from the first differencing results next plotted ACF.

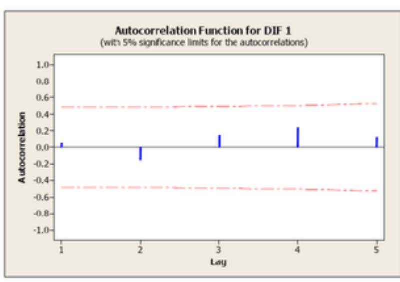

Figure 5. Plot ACF results first differencing results.

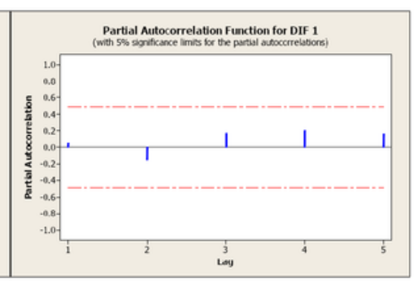

Figure 6. PACF plot first differencing result.

Based on Figures 5 and 6, it is seen that the first differencing data has been stationary to the average. However, we still cannot continue the next stage of data analysis that sets the model temporarily because there is no lag out so it cannot be predicted order. Similarly, on its PACF plot. Therefore, it will be done differencing again for data differencing first. Data differencing results hereinafter we call as data result of differencing second.

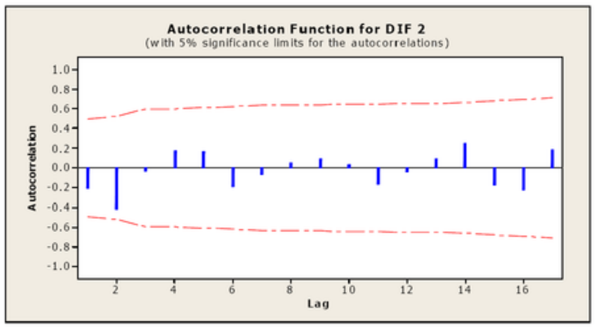

Figure 7. ACF plot of second differencing results.

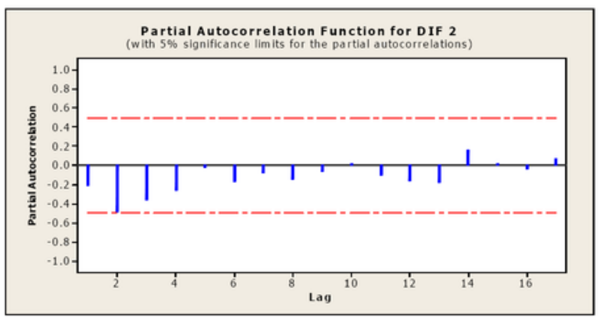

Figure 8. PACF plot of second differencing results.

Based on Figures 6 and 7, it appears that there is already a lag coming out. So we can continue the next stage of analysis.

Identify ARIMA

The time series model in the paddy production data can be determined by looking at the plot results of PACF (partial autocorrelation function) and

ACF (autocorrelation function) from the stationer data after the second differencing.

Based on plot and ACF value

Figure 7 shows the ACF plot is sinusoidal and there is no cut off. In addition can be seen in the value $|T|$ on the ACF of the data after the first differencing.

Hypotheses

$$H_0 : p = 0 \text{ (no cut-off occurs)},$$

$$H_1 : p \neq 0 \text{ (cut-off occurs)}.$$

Testing criteria

$$\text{Reject } H_0 \text{ if value } |T| > 2,$$

$$\text{Accept } H_0 \text{ if value } |T| < 2.$$

Based on ACF data and value $|T|$ in Appendix 6, it is known that autocorrelation for all lags has $|T| < 2$ so H_0 is accepted which means no cut-off occurs.

Based on plot and PACF value

In Figure 8, the PACF plot is sinusoidal and significant in the 1st and 2nd time lag. In addition can be seen in the value $|T|$ on the PACF of the data after the third differencing.

Hypotheses

$$H_0 : p = 0 \text{ (no cut-off occurs)},$$

$$H_1 : p \neq 0 \text{ (cut-off occurs)}.$$

Testing criteria

$$\text{Reject } H_0 \text{ if value } |T| > 2,$$

$$\text{Accept } H_0 \text{ if value } |T| < 2.$$

Based on PACF data and its $|T|$ value in Appendix 7, it is known that autocorrelation lag 2 (-0.558229) differs significantly with zero where the value $|T| = 2.05 > 2$. This indicates a cut-off on lag 2. The possible model is AR (2).

Based on the above, the identification model of ARIMA model tentative is ARIMA model (2, 2, 0) or AR (2, 2).

There are two assumptions that must be met in determining the adequacy of the model, i.e., residual white noise and normal distribution. Residual white noise testing can be done by using Ljung-Box test with $\alpha = 5\%$.

Parameter estimation

After setting the model temporarily, the next step is to estimate the parameters of each model temporarily. The following parameter estimates from ARIMA Model (2, 2, 0) or AR (2, 2).

Table 2. Estimation and test of significance of rice production in Indonesia

Model	Parameter	*P*-value	Decision
ARIMA(2, 2, 0)	$\Phi 1 = -0.5763$	0.005	Significant
	$\Phi 2 = -0.8510$	0.000	Significant
	$\mu = 247.5$	0.530	Not significant

Source: Data processed 2017

From the estimation Table 2 and significance test, there is only 1 significant model with negligible cost because it is not significant.

Table 3. Residual white noise test

Model	Lag	P-value	Decision
ARIMA(2, 2, 0)	12	0.907	White noise

Source: Data processed 2017

Only one model tested white noise is ARIMA (2, 2, 0) or AR (2, 2), then we test the residual (the normal test using Kolmogorov-Smirnov test).

Table 4. Residual normality test

Model	P-value	Decision
AR(2, 2)/ARIMA(2, 2, 0)	>0.150	Residual normal distribution

Source: Data processed 2017

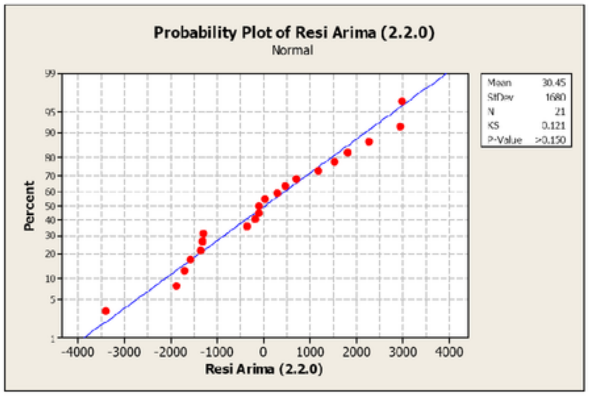

Figure 9. Residual test plot.

From Table 2, estimation and significance test, residual white noise test table, and residual normal table and plot of residual normality test, it is clear that there is one model is the best model that is AR model (2, 2) or ARIMA (2, 2, 0).

Table 5. Hasil forecasting

Year	Forecasting
2013	71280907
2014	71032657
2015	72937721
2016	75861325
2017	76752116

Source: Data processed 2017

From the above forecasting results can be compared with the original data that is data Year 2013, 3014 and 2015.

Table 6. Forecasting results 2013-2015 (million tonnes)

Year	Actual data	Forecasting
2013	71279709	71280907
2014	70846465	71032657

The data on the results of the app does not differ significantly from the original data. As seen in Table 6 that the data in 2013 that is 71279709 and the results forecast in 2013 that is 71280907 not much different, as well as for the year 2014 is 70846465 and the forecast 71032657, and the year 2015 is 75397841 and the results of the forecast 72937721. So it can be concluded that the model ARIMA (2, 2, 0) or AR (2, 2) is the best model for predicting rice production data in Indonesia. The results of this data analysis yielded rice production data in Indonesia in 2016 amounted to 75861325 tons and in 2017 amounted to 76752116 tons.

5. Conclusion and Suggestions

Based on the analysis and the results of the discussion, it can be concluded that

(a) The ARIMA model (2, 2, 0) is the best model for forecasting rice production in Indonesia. Forecasting is done to forecast 2013, 2014, 2015, 2016 and 2017.

(b) From the forecasting results, it can be seen that rice production has increased and not much different results from actual data in 2013 to 2015. For that, it is expected that the rice production can be maintained and even improved in the future.

Acknowledgement

The author thanks the anonymous referees for their valuable suggestions and constructive criticism which improved the presentation of the paper.

References

[1] Aswi and Sukarna, Analisis Deret Waktu, PenerbitAndira, Makasar, 2006.

[2] R. Bharathi, Y. N. Havaldhar, S. N. Megeri and G. M. Patil, Forecasting of arrivals and prices in Ramnagar and Siddlagatta market, J. Indian Soc. Agricultural Statist. 63(3) (2009), 247-257.

[3] G. Bolton, Game theory's role in role-playing, International J. Forecast. 18 (2002), 353-358.

[4] M. A. Hamjah, Rice production forecasting in Bangladesh: an application of Box-Jenkins ARIMA model, Mathematical Theory and Modeling 4(4) (2014), 1-11.

[5] Jonathan D. Cryer and Kung-Sik Chan, Time Series Analysis with Applications in *R*, 2nd ed., Springer, 2008.

[6] M. D. H. Gamal, Forecasting of rice stock using Winter's exponential smoothing and autoregressive moving average models, Internat. J. Eng. Res. Tech. 4(9) (2015).

[7] Nasiru Suleman and Solomon Sarpong, Forecasting milled rice production in Ghana using Box-Jenkins approach, Internat. J. Agricultural Management and Development (2011).

[8] R. K. Paul and M. K. Das, Statistical modelling of inland fish production in India, J. Inland Fisheries Society of India 42(2) (2010), 1-7.

[9] Prajneshu, S. Ravichandran and Savita, Structural time series models for describing cyclical fluctuations, J. Indian Soc. Agricultural Statist. 55(1) (2002), 70-78.

[10] Subagyo, Pangestu, Forecasting Konsepdan Aplikasi, Yogyakarta: BPFE, 1986.

[11] W. W. S. Wei, Time Series Univariate and Multivariate Methods, Addison Wesley Publishing Company, Inc., 1994.

Georgina Maria Tinungki: ina_matematika@yahoo.co.id

THE ELECTION OF THE BEST AUTOREGRESSIVE

ORIGINALITY REPORT

%23	%16	%9	%7
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.pphmj.com Internet Source	%9
2	www.ijert.org Internet Source	%4
3	Yenni P. Pasaribu, Hariani Fitrianti, Dessy Rizki Suryani. "Rainfall Forecast of Merauke Using Autoregressive Integrated Moving Average Model", E3S Web of Conferences, 2018 Publication	%2
4	Ansari Saleh Ahmar, Achmad Daengs GS, Tri Listyorini, Castaka Agus Sugianto, Y Yuniningsih, Robbi Rahim, Nuning Kurniasih. " Implementation of the ARIMA(p,d,q) method to forecasting CPI Data using in R Software ", Journal of Physics: Conference Series, 2018 Publication	%1
5	eprints.utm.my Internet Source	%1
6	Henning Titi Ciptaningtyas, Bagus Jati Santoso, Muhammad Fahrur Razi. "Resource elasticity	%1

controller for Docker-based web applications", 2017 11th International Conference on Information & Communication Technology and System (ICTS), 2017
Publication

7	Rizki Mahrivandi, Lienda Noviyanti, Gatot Riwi Setyanto. "Black-Litterman model on non-normal stock return (Case study four banks at LQ-45 stock index)", AIP Publishing, 2017 Publication	%1
8	Sonny Yuhensky, Rendy Munadi, Hafiddudin. "Forecasting formulation model for amount of fault of the CPE segment on broadband network PT. Telkom using ARIMA method", 2016 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC), 2016 Publication	%1
9	Sirajuddin, M Cut Gebrina Hisbach, Ratna Ekawati, SM Ade Irman. "Analysis of Container Yard Capacity In North TPK Using ARIMA Method", IOP Conference Series: Materials Science and Engineering, 2018 Publication	<%1
10	www.slideshare.net Internet Source	<%1
11	Submitted to Perry High School	

	Student Paper	<%1
12	Submitted to University of Zakho Student Paper	<%1
13	krishikosh.egranth.ac.in Internet Source	<%1
14	Submitted to Manchester Metropolitan University Student Paper	<%1
15	www.keele.ac.uk Internet Source	<%1
16	Ratna Sulistyowati, Suhartono, Heri Kuswanto, Setiawan, Erni Tri Astuti. "Hybrid forecasting model to predict air passenger and cargo in Indonesia", 2018 International Conference on Information and Communications Technology (ICOIACT), 2018 Publication	<%1

EXCLUDE QUOTES	ON	EXCLUDE MATCHES	< 5 WORDS
EXCLUDE BIBLIOGRAPHY	ON		