Journal of Mathematics Technology and Education Vol. 1, No. 2, 2022 | 173 - 182

The Effectiveness of ARIMA Method in Rice Production Forecasting of North Sumatera Province

Iqbal Al Farizy Siregar[1], Esther Sorta Mauli Nababan[1]

[1]*Departement of Mathematics, Faculty of Mathematics and Natural Science, Universitas Sumatera Utara, Medan, 20155, Indonesia*

Abstract. Rice production continues to be spurred to meet the increasing need for food. Meanwhile, based on data from the Badan Pusat Statistik, in 2019 there was a decline in national rice production, especially in North Sumatra Province. In this research, it was found that the estimated amount of rice production for the next 8 periods was obtained using the Box Jenkins ARIMA method with the ARIMA model (2,0,6) experiencing an up and down trend, The MAPE (Mean Absolute Percentage Error) deviation value is obtained by 12.84%, the forecast model is quite efective to be used.

Keyword: Rice Production, Box Jenkins Method, ARIMA, MAPE (Mean Absolute Percentage Error)

Abstrak. *Produksi padi terus dipacu untuk memenuhi kebutuhan pangan yang terus meningkat. Sementara itu berdasarkan data Badan Pusat Statistik pada tahun 2019 terjadi penurunan produksi padi nasional khususnya di Provinsi Sumatera Utara. Dalam penelitian yang dilakukan didapatkan prakiraan jumlah produksi padi selama 8 periode mendatang dengan Metode Box Jenkins ARIMA dengan model ARIMA (2,0,6) mengalami tren naik turun, diperoleh nilai deviasi MAPE (Mean Absolute Percentage Error) sebesar 12,84% maka model prakiraan cukup efektif untuk digunakan.*

Kata Kunci: *Produksi padi, Metode Box Jenkins, ARIMA, MAPE (Mean Absolute Percentage Error)*

Received 16 Sep 2021 | Revised 02 Dec 2021 | Accepted 17 Jan 2022

1. Introduction

Forecasting is needed in running a process because it can help to gain an overview and forecast values that will occur in the future. But most people still use intuition and subjective consideration

*Corresponding author at: Mathematics Department, Faculty of Mathematics and Natural Science, Universitas Sumatera Utara, Medan, Indonesia

E-mail address: iqbalalfarizy06@gmail.com, esthernababan@gmail.com

e-ISSN: xxxx-xxxx | DOI: 10.32734/jomte.v1i2.7069
Journal Homepage: https://talenta.usu.ac.id/jomte

when estimating the values in the future. The fundamental problems of forecasting is how high the level of accuracy / precision of forecasting results. This level of accuracy / precision of forecasting results can be seen as a measure of the effectiveness of a method of forecasting. One of the methods in forecasting is the ARIMA (Autoregressive Integrated Moving Average) method, where the effectiveness of the ARIMA method can be seen from how small the level of deviation that occurs from the forecasting results using this method.

The effectiveness of a forecasting method can be seen from the deviation that occurs from the forecasting results, so that whether or not a method is good depends on the resulting deviation value. If the resulting deviation value is getting smaller, then the forecasting results will be closer to the right. Azman [1].

Rice is the main source of carbohydrates in most Asian countries. National rice requirement is not met by domestic rice production because we are still always imported rice, rice production is still a priority to support food security and agribusiness program. However, all efforts to increase production always get interference, which include drought, floods, pests, and diseases. Meanwhile, based on data [2], there was a decline in national rice production, especially in North Sumatra Province.

2. Literature Review

2.1 Stationary

In conducting research using a time series analysis model, sometimes data that is not stationary is found. Stationarity means that there is no drastic change in the data. Data fluctuations are around a constant mean value, independent of the time and variance of these fluctuations [3]. So that the data can be said to be stationary if the mean and variance are constant from time to time.

2.2 Autoregressive (AR) Process

The autoregressive process means self-regression. More specifically, the autoregressive process Z_t of order p represents the equation:

$$Z_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots + \emptyset_p Z_{t-p} + e_t \quad (1)$$

it is assumed that Z_t stationary and $E(Z_t) = 0$

so, the value of the sequence Z_t is a linear combination of a number p of the last values Z_t in the past plus e_t and stating something that can not be explained by the values of the Z_t in past. In addition, e_t is an independent random variable with an average of zero [4].

2.3 Moving Average (MA) Process

The general form for an MA process of order q , written MA (q) is given by:

$$Z_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_p e_{t-p} \quad (2)$$

That is, the value of the sequence Z_t is a linear combination of the last number of e_t in the past. For identification purpose, if a time series has a graph of the autocorrelation function that breaks at the q-th lag and the partial autocorrelation function decreases exponentially, then the time series can be entered into the MA(q) process.

2.4 Mixed Process Autoregressive and Moving Average (ARMA)

If it is assumed that a time series has a model that is partly an Autoregressive process and partly a Moving Average process, the series will have a model that generally takes the form:

$$Z_t = Z_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots + \emptyset_p Z_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_p e_{t-p} \quad (3)$$

2.5 Autoregressive Integrated Moving Average (ARIMA) Model

A time series Z_t is said to follow the model of Autoregressive Integrated Moving Average (ARIMA) if the distinction d-order of Z_t is a stationary ARMA process, $W_t = (1 - B)^d Z_t$. Because W_t is the ARMA (p, q), then Z_t can be termed as a process ARIMA (p, d, q). In the form of the ARIMA model backshift operator can be written as follows

$$\emptyset(B)(1 - B)^d Z_t = \theta(B)\, e_t \quad (4)$$

Where,

$\emptyset(B)$ = $(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p)$ is the backshift operator AR

$\theta(B)$ = $(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^p)$ is the backshift operator MA

$(1 - B)^d$ = the d-order differencing operator.

2.6 Model Evaluation (Effectiveness)

In all situations the forecast contains a degree of uncertainty. The source of deviation (deviation) in forecasting is not only caused by the element of error, but the inability of a forecasting model to recognize other elements in the data series also affects the magnitude of the deviation in forecasting. So the magnitude of the deviation forecasting results can be caused by the amount of factors that are not expected (outliers) in which no forecasting method which is capable of producing accurate forecasting, or it can also be caused forecasting methods used can not predict exactly trend component, seasonal component or components cycle may be contained in the data series, which means that the method used is not appropriate. So the selection of the best forecasting method is based on the level of prediction error [5]. If the resulting error is getting smaller, then the forecasting results will be closer to the right. The measuring tool used to calculate prediction error is Mean Absolute Percentage Error (MAPE), MAPE has the advantage that absolute value provides the difference between actual data and forecasted data so that it can be made into a percent value that is easy to understand.

$$MAPE = \frac{100}{n}\sum_{t=1}^{n} \left| \frac{\hat{Z}_t - Z_t}{Z_t} \right| \qquad (5)$$

Where,

N = the number of observed data

$\hat{Z}_t$ = forecast result data at time *t*

Z_t = actual data at time *t*

The lower the MAPE value, the ability of the forecasting model used can be said to be good, and for MAPE there is a range of values that can be used as measurement material regarding the ability of a forecasting model, the range of values is as follows. [6].

Table 1 Parameter Estimation Results

Range MAPE	**Description**
<10%	Excellent forecasting model ability
10-20%	Good forecasting model ability
20-50%	Average forecasting model ability
>50%	Poor forecasting model ability

3. Research methods

The data will be processed with the following steps:

1. Stationary test of the data on the amount of rice production carried out by displaying the actual data plot, seeing the autocorrelation value and the form of the ACF and PACF plots of the data. To test kestasioneran more specific data used Augmented Dickey-Fuller test.
2. If the data does not meet the stationarity requirements, then the Box-Cox Transformation is performed to stationary the data against the variance and differencing to stationary the data to the mean.
3. Identification of seasonal lag data through ACF and PACF plots of stationary data. The identification of ACF and PACF plots is assisted by R Studio software.
4. Estimating the parameters of the model by means of an iterative algorithm using the Maximum Likelihood estimation method, namely by testing several different values. The estimation of these parameters is assisted by the R Studio software.

5. Performing a diagnostic check, this stage is used to check whether the estimated model has met the white noise and the residual normality test.
6. forecasts are made based on the equations of the selected model
7. Models that have met the requirements of the diagnostic examination are evaluated. The measuring instrument used to calculate prediction error is Mean Absolute Percentage Error (MAPE).

4. Results and Discussion

4.1 Rice Production Periodic Data

The data to be analyzed in this study is data on the amount of rice production in North Sumatra in January 2018 to December 2019 which is insample data, where data in 2020 is still temporary data so that the data is used as outsample data to see deviations in forecasting. Data obtained through the official website of Badan Pusat Statistik Sumatera Utara (www.sumut.bps.go.id).

4.2 Stationary test of Variance

In stationary data against variance, the first step that must be done is to look at the value of λ.

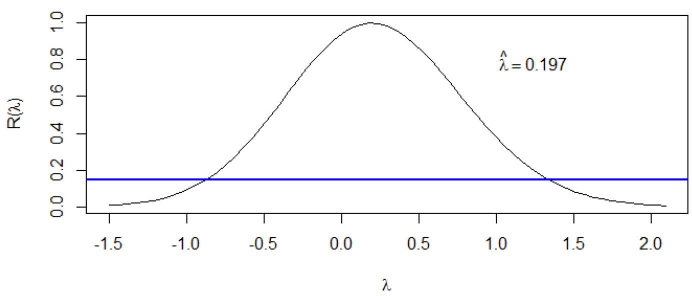

Figure 1 Box-Cox Transformation on Rice Production Data

In Figure 1 it can be seen that the value of $\lambda = 0{,}197 \approx 0$. Therefore, the data must be transformed into $ln\ Z_t$.

4.3 Stationary test of Mean

In stationing the data of mean, the first step that must be done is to look at the ACF plot of the transformed data.

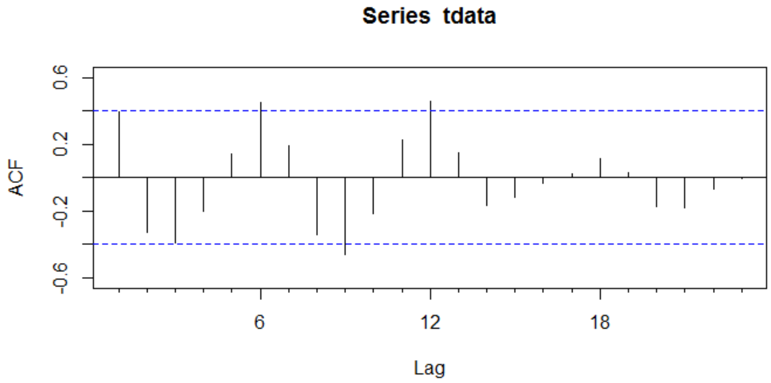

Figure 2 Box-Cox Transformation on Rice Production Data

Figure 2 shows that the data on the amount of rice production has an ACF pattern that does not decrease linearly or does not look sluggish so that the data is stationary on mean.

4.4 Model Identification

Based on Figure 2, the ACF pattern has cut offs at lag-1, 3, and 6 so that the MA (q) values are 1, 3, and 6. Next, to see the partial autocorrelation function (PACF) of the data that is stationary at lag-k. Here is the resulting plot.

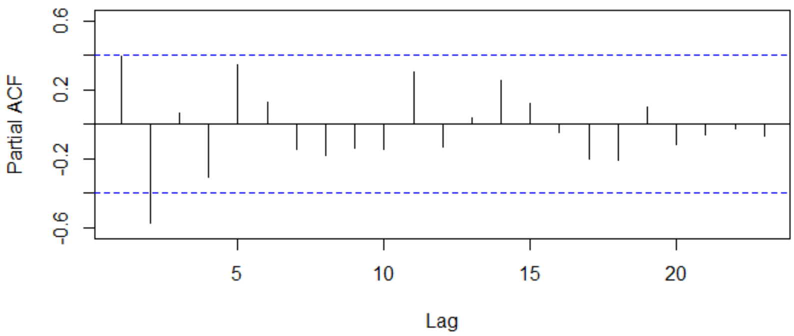

Figure 3 Box-Cox Transformation on Rice Production Data

Based on Figure 3, the PACF pattern has a cut off at lag-1 and 2 so that the AR (p) values are 1 and 2. Because the data is stationary without differencing, the temporary ARIMA models are ARIMA (1,0,1), ARIMA (1,0,3), ARIMA (1,0,6), ARIMA (2,0,1), ARIMA (2, 0.3), and ARIMA (2,0.6).

4.5 Literature Review

Table 2 Parameter Estimation Results

Model	Parameter	Koefisien	SE. Koefisien	AIC
ARIMA(1,0,1)	$\emptyset_1$	0,0041	0,2377	11,08
	θ_1	0,8219	0,1446	
ARIMA(1,0,3)	$\emptyset_1$	0,2785	0,2914	9,57
	θ_1	0,2645	0,2644	
	θ_2	-0,7883	0,2074	
	θ_3	-0,476	0,244	
ARIMA(1,0,6)	$\emptyset_1$	0,4724	0,4864	14,16
	θ_1	-0,1484	0,4736	
	θ_2	-1,1529	0,3383	
	θ_3	-0,4445	0,5129	
	θ_4	0,2328	0,6028	
	θ_5	0,3580	0,2748	
	θ_6	0,1939	0,3866	
ARIMA(2,0,1)	$\emptyset_1$	0,3668	0,5787	10,02
	$\emptyset_2$	-0,5339	0,4219	
	θ_1	0,3944	0,8397	
ARIMA(2,0,3)	$\emptyset_1$	0,9204	0,3339	9,83
	$\emptyset_2$	-0,6259	0,2589	
	θ_1	-0,4853	0,4459	
	θ_2	-0,6208	0,2271	
	θ_3	0,1061	0,3600	
ARIMA(2,0,6)	$\emptyset_1$	0,9673	0,0726	9,67
	$\emptyset_2$	-0,9700	0,0761	
	θ_1	-0,6767	0,2901	
	θ_2	0,0588	0,2862	
	θ_3	0,4656	0,2640	
	θ_4	-0,5624	0,3449	
	θ_5	-0,3518	0,2599	
	θ_6	0,0670	0,2599	

Based on the results of the significant level test in Table 1, the ARIMA(1,0,3) and ARIMA(2,0,6) models have an AIC error value that is smaller than the other models, which will then be tested for diagnostic tests.

4.6 Diagnostic Test

A good model would have the properties of white noise, that meet the assumptions of residual random and normally distributed. The residual randomness of a model can be tested using the Ljung Box statistical test with the following hypothesis:

$H_0 : \rho_1 = \rho_2 = \rho_3 = \cdots = \rho_m = 0$ (white noise)

$H_1 : \rho_k \neq 0\ , k = 1, 2, 3 \ldots m$ (not white noise)

With $\alpha = 0.05$ and statistical test:

$$Q = n\,(n+2)\sum_{k=1}^{i}\frac{\hat{\rho}_k^{\;2}}{(n-k)}$$

The criteria are decided if the *p-value* > α which means it meets the White Noise process and if the *p-value* < α which means it does not meet the White Noise process. In simple terms, $\boldsymbol{H_0}$ is rejected if $Q > \boldsymbol{X^2}_{\boldsymbol{\alpha},\ \mathbf{df=K-p-d}}$. To see if the residuals are normally distributed or not, it can be seen using the Shapiro Wilk test with the following hypothesis:

H_0 : meet the normal distribution

H_1 : *did not* meet the normal distribution

With $\alpha = $ *0.05 and* statistical test:

$$W = \frac{\left(\sum_{i=1}^{t} a_i\, x_{(i)}\right)^2}{\sum_{i=1}^{t}(X_i-\bar{x})^2}$$

Where $x_{(i)}$ is sample data that has been sorted from smallest to largest and a_i is the Shapiro Wilk index, As well as test criteria: reject H_0 if $W < W_a\ (p-value)$

Table 3 Model Diagnostic Test Result

Diagnostic Check	ARIMA (1,0,3)	ARIMA (2,0,6)
White Noise	fulfil	fulfil
Normal Distribution Test	fulfil	fulfil

4.7 Prediction with the Best Model

The mathematical form of the ARIMA model (1,0,3) can be written as follows.

$$Z_t = 0{,}2785Z_{t-1} + e_t - 0{,}2645e_{t-1} + 0{,}7883e_{t-2} + 0{,}476e_{t-3}$$

The mathematical form of the ARIMA model (2,0,6) can be written as follows.

$$Z_t = 0{,}9673Z_{t-1} - 0{,}970Z_{t-2} + e_t + 0{,}6767e_{t-1} - 0{,}0588e_{t-2} - 0{,}4656e_{t-3} + 0{,}5624e_{t-4} + 0{,}3518e_{t-5} - 0{,}067e_{t-6}$$

4.8. Prediction with the Best Model

In the evaluation of the model, based on formula (5), outsampled data was used from January to August 2020 so that MAPE obtained the evaluation value of the ARIMA model as follows.

Table 4 Model Evaluation Result

Model	MAPE Score
ARIMA (1,0,3)	25,65948
ARIMA (2,0,6)	12,84379

Because the MAPE deviation value from the ARIMA model (2.0.6) is 12.83% smaller than the MAPE deviation value from the ARIMA model (1.0.3) which is 25.65%, the model used is ARIMA (2.0.6). Therefore, the ARIMA model (2,0.6) is an effective model because it is able to predict the amount of rice production in North Sumatra province for the next few months.

5. Conclusion

Based on the results of the tests and discussions that have been described in the previous section, it can be concluded that as an answer to the formulation of the problem in this study, the model used is the ARIMA model (2,0,6) with forecasting results that have an up and down trend, where the deviation results (deviation) using MAPE (Mean Absolute Percentage Error) on the results of forecasting rice production in 2020 is 12.84% where the model has good forecasting ability. Therefore, the ARIMA model (2,0.6) is quite effective in predicting the amount of rice production in North Sumatra province for the next few months.

REFERENCES

[1] Maricar, M Azman. 2019. "Analisa Perbandingan Nilai Akurasi Moving Average dan Exponential Smoothing untuk Sistem Peramalan Pendapatan pada Perusahaan XYZ". Jurnal Sistem dan Informatika. Bali: 2019.

[2] Badan Pusat Statistik. 2020. Luas Panen dan Produksi Padi di Indonesia: Maret 2020.

[3] Makridakis, S., Wheelwright, S.C., dan Victor, E.M. 1999. Metode dan Aplikasi Peramalan, second edition. Jakarta: Erlangga.

[4] Akolo, Ingka Rizkyani. 2019. "Perbandingan Exponential Smoothing Holt-Winters dan Arima pada Peramalan Produksi Padi di Provinsi Gorontalo". Jurnal Teknologi Pertanian Politeknik Gorontalo. Gorontalo: JTech 7(1),20 - 26, 2019.

[5] Badan Pusat Statistik Sumatera Utara. 2020. Luas Panen dan Produksi Padi di Provinsi Sumatera Utara : Maret 2020.

[6] Chatfield, Christopher. 2000. Time-Series Forecasting. United States of America: CRC Press LLC.

[7] Cryer, Jonathan D dan Chan, Kung-sik. 2008. Time Series Analysis With Applications in R Second Edition. United States of America: Springer Science+Business Media, LLC.

[8] Fattah, Jamal, Latifah Ezzine, et al. 2018. "Forecasting of demand using ARIMA model". Jurnal Internasional Engineering Business Management. Maroko:Volume 10: 1–9, 2018.

[9] Meli Pranata. 2020. Prediksi Pencurian Sepeda Motor Menggunakan Model Arima (Studi Kasus: Polres Kotabumi Lampung Utara). Skripsi: Universitas Islam Negeri Raden Intan, Lampung.

[10] Riandiantika Prameswari. 2020. Peramalan Ekspor Kendaraan Bermotor di Indonesia Menggunakan Metode Arima (Autoregressive Integrated Moving Average) Box-Jenkins. Skripsi: Universitas Islam Negeri Sunan Ampel, Surabaya.

[11] Rosadi, Dedi. 2016. Analisis Runtun Waktu dan Aplikasi dengan R. Yogyakarta: Gajah Mada University Press.

[12] Suhartono. 2008. Analisis Data Statistik dengan R. Surabaya: Lab. Statistik Komputasi, ITS.

[13] Supranto J. 2000. Statistik (Teori dan Aplikasi), Edisi Keenam. Jakarta: Erlangga.

[14] Wei, W.S.W. 2006. Time Series Analysis Univariate and Multivariate Methods Second Edition. United States of America: Person Education, inc.

[15] Wijaya, Adi dan Suhartono. 2012. "Peramalan Produksi Padi dengan Arima, Fungsi Transfer, dan Adaptive Neuro Fuzzy Inference System". Jurnal Matematika Institut Teknologi Sepuuh Nopember. Surabaya: 2012