Sugar Production Forecasting System in PTPN XI Semboro Jember Using Autoregressive Integrated Moving Average (ARIMA) Method

Januar Adi Putra [1], Saiful Bukhori [2], Faishal Basbeth [3]

Faculty of Computer Science,

University Of Jember

Jember, Indonesia

Januaradi.putra@unej.ac.id [1], saiful.ilkom@unej.ac.id [2], faishalbasbeth@unej.ac.id [3]

***Abstract*—. There is a lot of entrepreneurial competition in the production of goods or services in the world, especially in Indonesia, especially the production of staple goods, namely sugar. The problem that is often faced at Sugar Factory PTPN XI Semboro Jember is the lack of management that is neatly organized and efficient, which makes this company less working optimally. Often there is a lack and excess of sugar production which makes the sugar does not have the maximum value, the sugar has been damaged, and sales at a reduced price because the sugar is not as efficient as the initial product. From these various problems, it can reduce profits from the company. From these problems it can be concluded that the company needs a system that can organize the management of the company, and is able to forecast production in the future. In this research will make a forecasting system using the method of Autoregressive Integrated Moving Average (ARIMA), where this method is divided into three methods, namely the Autoregressive (AR) method, the Moving Average (MA) method, and the Autoregressive Integrated Moving Average (ARIMA) method, which preceded by checking stationary data, and modeling the Autoregressive Integrated Moving Average (ARIMA) method. Forecasting is done using production data for the previous 12 years from the company. The system is made to facilitate management that is less organized and displays predictions for the next production period. The results of this forecasting system are to determine the amount of production each year needed in this company. From the results of the ARIMA method modeling, the right ARIMA method is obtained by the ARIMA / AR (1,0,0), ARIMA / MA (0,0,1), and ARIMA (1,0,1) methods. The test results found that the average value of Mean Absolute Percentage Error (MAPE) in the Autoregressive (AR) method was 17%, the Moving Average (MA) method was 19%, and the Autoregressive Integrated Moving Average (ARIMA) method was 15%.**

Keywords—Forecasting, Sugar Production, ARIMA

I. INTRODUCTION

There is a lot of entrepreneurial competition in the production of goods or services in the world, especially in Indonesia, especially the production of nine basic commodities (basic food), namely sugar. It is one of the basic ingredients consumed by the people of Indonesia. Most of the sugar is consumed by the community as an energy source, flavor, and as a raw material for the food and beverage industry [5]. Sugar plays an important role in people's daily lives, one of the main sugar production areas in Jember is the Semboro Jember Sugar Factory PTPN XI Sugar Factory.

Sugar Factory PTPN XI Factory Sugar Semboro Jember is located in Semboro Village / District, Jember Regency. Operating since 1928 as a business unit owned by a private company in the era of colonialism. It is a factory built during the Dutch colonial period and was built by the Dutch. Located around 35 km west of the city of Jember.

According to the Head of the Production Section, the problem faced by this factory is the lack of neatly organized and efficient management, which makes this factory not work optimally. Often the instability of sugar production that does not produce sugar has the maximum possible value, the sugar is damaged, and sales at reduced prices because sugar is not as efficient as the initial product. So far the leader of this company has decided to produce sugar each year with a manual method, namely with the existing business estimates from previous production.

The problems that exist in Sugar Factory PTPN XI Semboro Jember can be overcome by forecasting the determination of sugar production. Forecasting can be an important thing in a decision making, namely factors that we cannot see when the decision is taken. Forecasting can also be done in the field of business, which can estimate the amount of production and use of products in the next period, so that products are made in the right quantity with the results of forecasting [1].

Forecasting the production of an item can use various methods, one of the right methods in this study is the ARIMA method. This method was chosen because of the previous sugar production data at Sugar Factory PTPN XI Semboro Jember y including stationary data [4]. In another study also explained that the ARIMA model is more flexible than the (Autoregressive) AR model, (Moving Average) MA, and (Autoregressive Moving Average) ARMA [3].

ARIMA method is a method used to forecast the next data using two methods combined, namely the AR method, MA method, and becomes the ARIMA method. The AR and MA methods perform forecasting calculations using previous data using the Cramer method matrix to find out the constant value, and the parameter value from the results of the calculation. After knowing the constant value and parameter values of the AR and MA methods, the ARIMA method of forecasting method will be carried out [4].

In addition to the detailed calculations, the ARIMA method can get three forecasting results, namely the results of forecasting the AR method, the MA method, and the ARIMA method. Based on the explanation above, this research aim to build an information system that to be able to forecast the amount of sugar production in the Sugar Factory using the ARIMA method by processing the previous production data. So that Sugar Factory PTPN XI Sugar Factory Semboro Jember can plan the required sugar production.

II. Method

1. Autoregressive Integrated Moving Average Method (ARIMA)

Autoregressive Integrated Moving Average (ARIMA) was first developed by George Box and Gwilym Jenkins for time series analysis modeling. It represents three models, namely from the Autoregressive (AR) model, Moving Average (MA), and the Autoregressive Integrated Moving Average model (ARIMA) [4]. The stages of implementation in the search method are:

1. Checking data, does the data include stationary data?
2. If the data includes data that is not stationary, differencing data will be carried out until the data becomes stationary data, if the data includes stationary data, differencing data will not be done.
3. Methods identified using Autocorrelation and Partial Autocorrelation.
4. The method is interpreted and estimated using past data using the least squares method or the Cramer method.
5. Testing is done to get a method that is feasible to use for the application of forecasting.
6. Application, which is forecasting the value of the periodic data series that will come using the method that has been tested.

Parameter Estimation

Determination of the estimation method of the Autoregressive Integrated Moving Average (ARIMA) (p, d, q) that can be determined by looking at the behavior of the Autocorrelation (ACF) plot and plotting Partial Autocorrelation (PACF) from periodic data series. In practice the values of p and q rarely have more than 2 values. After getting the value p, d, q, they can do forecasting. ARIMA Box-Jenkins Method (ARIMA) is divided into three groups, namely: the Autoregressive (AR) method, Moving Average (MA), and a mixed model Autoregressive Integrated Moving Average (ARIMA) that has the characteristics of the first two models [4].

1.1. Autoregressive (AR)

Determination of Partial Autocorrelation coefficient (PACF) is used to measure the level of closeness between Xt and Xt-k if the influence of time lag is 1,2, ..., k. The purpose in the analysis of periodic data is to establish an appropriate method of Autoregressive Integrated Moving Average (ARIMA) for forecasting, especially to determine the order p of the Autoregressive (AR) (p) model. Equation 1 is the formula of Autoregressive (AR) [4]:

$$Xt = \mu + \phi 1\ Xt\text{-}1 + \phi 2\ Xt\text{-}2 + ... + \phi p\ Xt\text{-}p + et \quad ..(1)$$

Information :
Xt : data t.
μ : the value of a constant.
ϕj : autoregressive parameter j.
et : error value t.

Estimating Autoregressive (AR) parameters can be used matrix multiplication method (Cramer method). Equations 2, 3, and 4 are formulas of the Cramer method:

$$\beta = (Z` Z )^{-1} Z` Y \quad . . . (2)$$

$$Z = \begin{bmatrix} 1 & Xp & ... & Xp-(p-1) \\ 1 & Xp+1 & ... & Xp-(p-1)+1 \\ ... & ... & ... & ... \\ 1 & Xn-1 & ... & Xn-p \end{bmatrix} ; . . (3)$$

$$Y = \begin{bmatrix} Xp+1 \\ Xp+2 \\ ... \\ Xn \end{bmatrix} ; \quad \beta = \begin{bmatrix} \mu \\ \phi_1 \\ .. \\ \phi_p \end{bmatrix} ; \quad . . . (4)$$

Information :
p : ordo of the AR model.
Xp : data p.
N : number of observation periods.
β : estimating parameter equations.

1.2. Moving Average (MA)

The Autocorrelation coefficient (ACF) with the correlation coefficient is the same. The difference is that it lies in the Autocorrelation coefficient (ACF) which describes the relationship (association) between the values of the same variable but differs in period. Autocorrelation (ACF) provides important information about the structure and pattern of the data. In a complete random data collection, Autocorrelation (ACF) between successive values will be close to 0 while the value of seasonal data and cycle patterns will have strong Autocorrelation (ACF) so that if this happens the data is not stationary. Autocorrelation (ACF) serves to find correlation between data and is useful for determining order (q) on Moving Average (MA) (q). Equation 5 is a formula from Moving Average (MA) [4]:

$$Xt = \mu + et - \theta 1\ et\text{-}1 - \theta 2\ et\text{-}2 - ... - \theta q\ et\text{-}q \quad ……(5)$$

Information :
μ : the value of a constant
ѳj : moving average parameter j.
et : error value t.

Estimation of the Moving Average (MA) parameter can be formed by the matrix multiplication method. Equations 6 and 7, are formulas for matrix multiplication methods [4]:

$$\beta = (Z` Z )^{-1} Z` Y \quad . . . (6)$$

$$Z = \begin{bmatrix} 1 & Xq+1-1-Xq+1 & Xq+1-2-Xq+1 & \dots & Xq+1-q-Xq+1 \\ 1 & Xq+2-1-Xq+2 & Xq+2-2-Xq+2 & \dots & Xq+2-q-Xq+2 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & Xn-1-Xn & Xn-2-Xn & \dots & Xn-q-Xn \end{bmatrix} \quad ; \dots (7)$$

Information :
q : ordo of the MA model.
Xq : data q.
n : number of observation periods.
β : estimating parameter equations.

1.3. Autoregressive Integrated Moving Average

The Autoregressive Integrated Moving Average (ARIMA) method p and q Autoregressive (AR) (p) and Moving Average (MA) orders are a combination of Autoregressive Model (AR) and Moving Average (MA). Equation 8 is a formula from ARIMA [4]:

$$Xt = \mu + \phi 1\ Xt\text{-}1 + \phi 2\ Xt\text{-}2 + \dots + \phi p\ Xt\text{-}p, + et - \text{ɵ}1\ et\text{-}1 - \text{ɵ}2\ et\text{-}2 - \dots - \text{ɵ}q\ et\text{-}q \quad . . . (8)$$

Information :
Xt : data t.
μ : constant value.
ϕj : autoregressive parameter j.
et : error value when t.
ɵj : moving average parameter j.

2. Error Calculation

The Mean Absolute Percentage Error (MAPE) is calculated using absolute errors in each period divided by the actual observation values for that period. Then, it states the absolute percentage error. The Mean Absolute Percentage Error (MAPE) is a measurement of error that calculates the percentage measure of deviation between actual data and forecasting data. Equation 9 is the formula of Mean Absolute Percentage Error (MAPE) [2]:

$$MAPE = \left(\frac{100\%}{n}\right) \sum_{t=1}^{n} \frac{[Xt-Ft]}{Xt} \quad \dots (9)$$

Information :
n : the amount of data.
Xt : periodic data value.
Ft : forecast value of the model.

3. Methodology

The sequence of research steps shown in Figure 1 and the forecasting flow Autoregressive Integrated Moving Average (ARIMA) are shown in Figure 2.

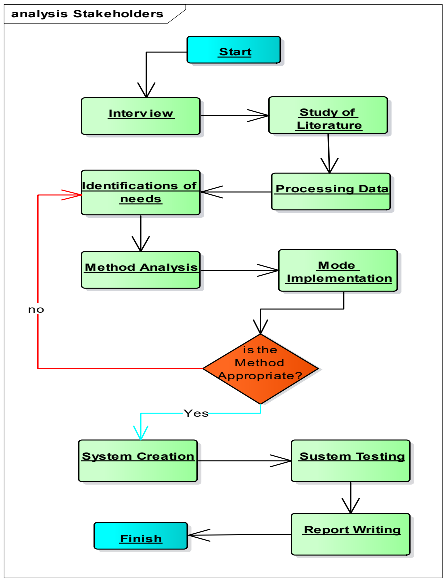

Figure 1. Research Step Chart

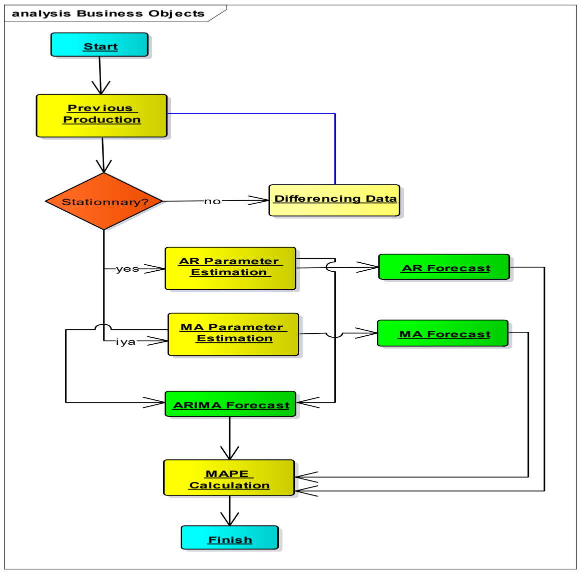

Figure 2. Flow Chart Forecasting Autoregressive Integrated Moving Average (ARIMA)

System Design

This chapter describes the design of the system to be built. The system design used in this study is as follows:

a. Input and Output Elemen

Information system for forecasting sugar production contains elements consisting of input elements, output elements, goals of information systems for forecasting sugar production, and uses or media used to make the system. Information system elements for forecasting sugar production can be seen in Figure 3.

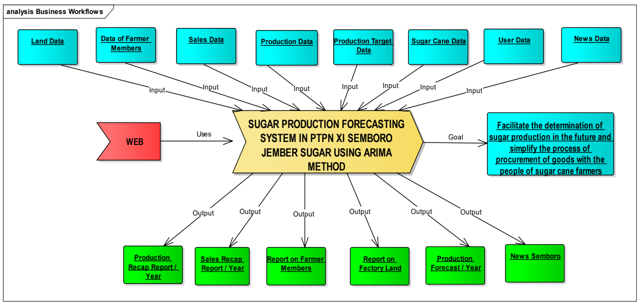

Figure 3. Bussiness Process

b. Use Case Diagram

Use Case Diagram of sugar production forecasting system can be seen in Figure 4.

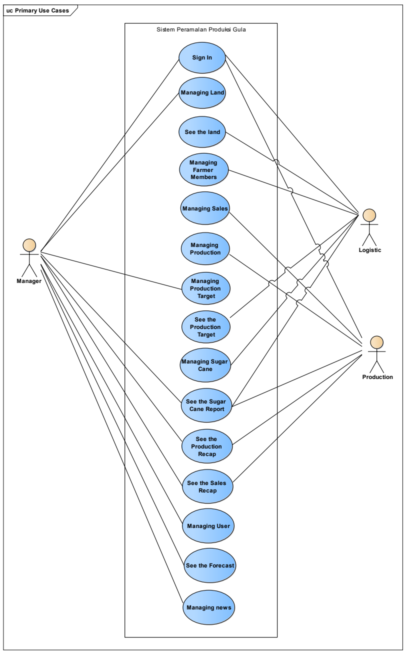

Figure 4. Use Case Diagram

c. Entity Relationship Diagram

Entity Relationship Diagram of sugar production forecasting system can be seen in Figure 5.

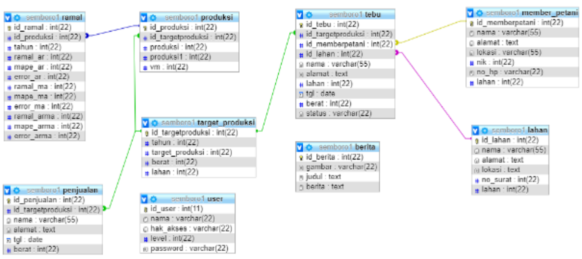

Figure 5. Entity Relationship Diagram

III. Result and Discussion

This chapter explains the results of the research that has been done and the discussion of the system that has been made.

1. Data

The previous 12 years of sugar production data, 2006 - 2017, are used as reference data in calculating production forecasting in the next period, that is 2018. The previous 12 years of production data are in Table 1 and the production graph in Figure 6.

Table 1. Production Data

Year	Production (Ton)
2006	21132
2007	24668
2008	28083
2009	29578
2010	30828
2011	29889
2012	23407
2013	22388
2014	23023
2015	31438
2016	32242
2017	36860

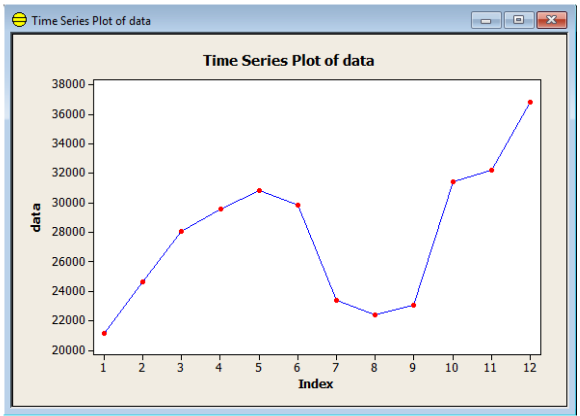

Figure 6. Production Data Chart

From the production data 12 years earlier, the data pattern is known as stationary data. Based on the data pattern, the right method of calculating the prediction of sugar production is the Autoregressive Integrated Moving Average (ARIMA) method because it is suitable for stationary data plots [4].

2. Modeling the Autoregressive Integrated Moving Average (ARIMA) Method

In the ARIMA forecasting method modeling will be carried out several stages. The first stage will be carried out calculation of Partial Autocorrelation (PACF), and the second is Autocorrelation (ACF) calculations use production data 12 years earlier.

The calculation of Partial Autocorrelation (PACF) and Autocorrelation (ACF) using the Minitab-14 application. The results of the graph and calculation of Partial Autocorrelation (PACF) can be seen in Figure 7 and Figure 8. The results of the graph and calculation of Autocorrelation (ACF) can be seen in Figure 9 and Figure 10.

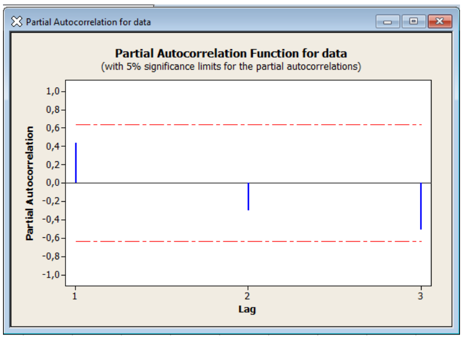

Figure 7. Partial Chart of Autocorrelation functions (PACF)

Partial Autocorrelation Function: data

Lag	PACF	T
1	0,433695	1,50
2	-0,300885	-1,04
3	-0,507527	-1,76

Figure 8. Partial Autocorrelation (PACF) Data

The calculation process for Partial Autocorrelation (PACF) can determine the value of p, where in Figure 8 the result of lag 1 is 0.433695 during period 1. Because the first period > period 2, then p is given a value of 1. The value p will be used for calculation forecasting Autoregressive Integrated Moving Average (ARIMA) [4].

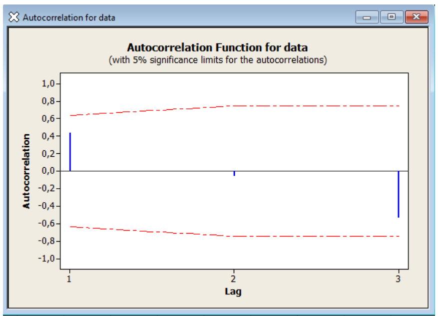

Figure 9. Chart of Autocorrelation functions (ACF)

Autocorrelation Function: data

Lag	ACF	T	LBQ
1	0,433695	1,50	2,87
2	-0,056200	-0,17	2,93
3	-0,536960	-1,58	8,31

Figure 10. Autocorrelation (ACF) Data

The process of Autocorrelation calculation (ACF) can determine the value of q, where in Figure 9 the results of lag 1 are 0.433695 during period 1. Because the first period> period 2, then q is given a value of 1. This q value is later used for forecasting calculations Autoregressive Integrated Moving Average (ARIMA) [4].

The next process is to determine the method of Autoregressive Integrated Moving Average (ARIMA) (p, d, q). The value of p obtained from the calculation of Partial Autocorrelation (PACF) where the calculation results in Figure 8 get a value of 1, the value of d in this study has a value of 0 because the value of d in the Autoregressive Integrated Moving Average method (ARIMA) is used to station the existing data [12]. While in this study the data is stationary, so the value of d is given a value of 0, and the value of q obtained from the calculation of Autocorrelation (ACF) where the results of calculation in Figure 10 get a value of 1. Then obtained modeling Autoregressive Integrated Moving Average (ARIMA) (p, d, q) for this sugar production data are ARIMA (1,0,0) / AR, ARIMA (0,0,1) / MA, and ARIMA (1,0,1).

3. Forecasting Results

Testing of forecasting data to evaluate the results of forecasting data available on the system using the Mean Absolute Percentage Error (MAPE) method. Forecasting ability is very good if it has a MAPE value of less than 10% and has good forecasting ability if the MAPE value is less than 20%. [2]. The results of the calculation of ARIMA forecasting results (1.0,0) / AR using the Mean Absolute Percentage Error (MAPE) method are in Table 2, the results of the calculation of ARIMA (0.0,1) / MA forecasting test results are in Table 3, and the results of the calculation of testing of ARIMA forecasting results (1.0,1) are in Table 4.

Table 2. ARIMA (1,0,0) / AR Test Results

Year	Actual Data	Forecaseting Data	MAPE %
2012	23407	30363	23%
2013	22388	26448	15%
2014	23023	24805	7%
2015	31438	24674	27%
2016	32243	29166	11%
2017	36860	30432	21%

From the calculation of the Mean Absolute Percentage Error (MAPE) in table 2 the biggest mean Absolute Percentage Error (MAPE) in forecasting in 2015 is yielding a value of 27%, and the smallest Mean Absolute Percentage Error (MAPE) obtained in the forecasting in 2014 is 7 %. The average of all Mean Absolute Percentage Error (MAPE) produces a value of 17%. this shows that the ARIMA (1,0,0) / AR method has a good accuracy value in this study.

Table 3. ARIMA (0,0,1) / MA Test Results

Year	Actual Data	Forecaseting Data	MAPE %
2012	23407	33053	29%
2013	22388	24451	8%
2014	23023	26058	12%
2015	31438	25217	25%
2016	32243	29501	9%
2017	36860	28327	30%

From the calculation of the Mean Absolute Percentage Error (MAPE) in Table 3 the biggest mean Absolute Percentage Error (MAPE) in forecasting in 2017 is yielding a value of 30%, and the smallest Mean Absolute Percentage Error (MAPE) obtained in the forecasting in 2013 is 8 %. The average of all Mean Absolute Percentage Error (MAPE) produces a value of 19%. this shows that the ARIMA (0,0,1) / MA method has a good accuracy value in this study.

Table 4. ARIMA (1,0,1) Test Results

Year	Actual Data	Forecaseting Data	MAPE %
2012	23407	32114	27%
2013	22388	23589	5%
2014	23023	24313	5%
2015	31438	24175	30%
2016	32243	32697	1%
2017	36860	30217	22%

From the calculation of the Mean Absolute Percentage Error (MAPE) in Table 4, the biggest mean Absolute Percentage Error (MAPE) in forecasting in 2015 is yielding a value of 30%, and the smallest Mean Absolute Percentage Error (MAPE) obtained in forecasting in 2016 is 1 %. The average of all Mean Absolute Percentage Error (MAPE) produces a value of 15%. this shows that the ARIMA method (1,0,1) has good accuracy in this study [2].

IV. Conclusion

The results of the research that has been done, conclusions can be taken as follows:

1. Forecasting sugar production at Sugar Factory PTPN XI Sugar Factory Semboro Jember using the ARIMA method. The calculation of forecasting begins with calculating production data 6 years before. The sugar production data used is data 12 years ago. In this study forecasting is carried out in annual intervals using the ARIMA method to calculate the value of production forecasting for the following year. The ARIMA method is able to properly calculate production forecasting for the following year.
2. The sugar production forecasting system at Sugar Factory PTPN XI Sugar Factory Semboro Jember using the ARIMA method was made based on the sequence of management of Sugar Factory PTPN XI Sugar Factory Semboro Jember. The forecasting system performs a forecasting process after the production actor adds production data every year.
3. The model of ARIMA (p, d, q) obtained in this study is ARIMA (1,0,0) / AR, ARIMA (0,0,1) / MA, and ARIMA (1,0,1). The result of lag 1 PACF is 0.433695 during period 1, then p is given a value of 1, the value of d has a value of 0, and the result of lag 1 ACF is 0.433695 during period 1, q is given a value of 1.
4. From the results of the ARIMA / AR calculation (1.0,0) it is found that the average of all MAPE scores 17%. It shows that the ARIMA (1,0,0) / AR method has a good accuracy value in this study. From the results of the ARIMA / MA calculation (0,0,1), the average of all MAPE results in a value of 19%. It shows that the ARIMA (0,0,1) / MA method has a good accuracy value in this study. From the results of ARIMA calculation (1.0,1), the average of all (MAPE) results in a value of 15%. It shows that the ARIMA method (1,0,1) has good accuracy in this study.

Acknowledgment

We would like to thank the Sugar Factory Sugar Factory PTPN XI for supporting this research.

References

[1] Habibi, M. Y., & Riksakomara, E. (2017). Forecasting Price of Consumption Salt Using Artificial Neural Network Feedforward-Backpropagation (Case Study: PT. Garam Mas, Rembang, Central Java).*Teknik ITS.*

[2] Kristiean Margi, S. P. (2015). Analysis and Application of Single Exponential Smoothing Method for Prediction of Sales in Specific Periods (Case Study: PT. Media Cemara Kreasi). *Prosiding SNATIF.*

[3] Permatasari, C. I., Sutopo, W., & Hisjam, M. (2018). Sugar Producion. Department of Industrial Engineering, Faculty of Engineering, Sebelas Maret University, Surakarta 57126, Indonesia . *030017-1.*

[4] Setiawan, A., Wibowo, A., & Wijaya, S. (2013). Application of Forecasting Cosmetics Sales Using the ARIMA Method. *KNSI-274.*

[5] Suwarno, D, R. R., & Indah, H. (2015). The Process of Making Inverted Sugar From Sucrose With Citric Acid, Tartric Acid and Chloride Acid Catalysts.. Jakarta