SHOES SALES FORECASTING USING AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) (CASE STUDY UD.WARDANA MOJOKERTO)

[1]ACHMAD KIKI QUSHAYRI WAHYU KUSUMA, [2]EKO PRASETYO, [3]RIFKI FAHRIAL ZAINAL
Informatic Engineering, University of Bhayangkara Surabaya
Email: kikiachmad50@gmail.com, eko@ubhara.ac.id, rifki@ubhara.ac.id

ABSTRACT

Shoes sales is increase of day by day along with growing trend in the society. This makes shoe manufacturers demand to fulfill the customer needs. UD. Ward as one of the shoe manufacturers in Mojokerto city trying to fulfill the customer needs efficiently in order that the make sales fit with production. To predict sales of shoes used Autoregressive Integrated Moving Average (ARIMA) method. ARIMA forecasting method is one of methods that According to historical data. Before go into the forecasting stage, differentiated the sales data per day during the year 2015-2016 ACF and PACF formula used Whose function is to Determine the value of p and q coefficient of the which will later be used in forecasting models in every formula that is AR , MA and ARMA. Result of this research shows that for the marching band category Obtained the best models that is MA with forecasting the result at the last period of 95.6432 and MSE of 472.4514. Obtained fashion category for the best models of forecasting that is AR with the result at the last period of 57.1872 and MSE of 304.8306. Obtained category for the best wedding that is AR models with forecasting the result at the last period of 21.4206 and MSE of 118.0681.

Keywords: ARIMA, forecasting, differentiation, MSE, sales, prediction

I.   INTRODUCTION

Every day people always do a routine activity such as work or study. Most of them definitely use shoes as a footwear. In addition to being more polite, have a diverse models of shoes that suits the tastes of its users. Currently the shoe manufacturers are competing for the make an attractive models with a affordable prices for satisfying consumer needs. UD. Ward is the one of shoe manufacturers in the Mojokerto City.

Shoe production at UD. Ward is erratic every month. There is 3 types of shoes that produced that is shoes for marching bands, wedding, and fashion. Consequences of the uncertainty of this production cause problems like a sales Several some shoes that were not sold or otherwise stock of shoes ran out while there were consumen needed it. One of the method for forecasting the shoes sale is an ARIMA method.

ARIMA method is an one of a method for forecasting. ARIMA method (Autoregressive Integrated Moving Average) is a forecasting method with a series of past using the data that is used to observe your an incident, event, or a variable in the data. This method is firstly made by George Box and Gwilym Jenkins for periodic series analysis. ARIMAis formed from three methods there is AR (autoregressive), MA (MovingAverage), and ARMA (Autoregressiveand Moving Average). The advantages onARIMA method is can be used for analyze the random situations, trends, seasons even cyclical nature in a time series Analyzed.

Based on the description above this research is conducted to apply ARIMA forecasting method for product sale shoes. Therefore an application was made entitled "Shoes Sale Forecasting Using Autoregressive Integrated Moving Average (ARIMA) Method" to make it Easier for shoes manufacturers forecast or estimate for the product to suit with a consumer needs.

II.   BASIC THEORY
2.1   forecasting

Based on the opinion from Jay Heizer and Barry Render (2001), forecasting is an art and science to predict a future events by taking a historical of data and projecting to the future with using Several forms of mathematical models.

Based on the Lerbin R. Arita R. (2002) opinion, forecasting is an activity of applying models that have been developed in the future.

Based on the then Sumayang (2003) opinion, forecasting is an objective calculation with using past data to Determine something in the future.

From the above understanding can be concluded that forecasting is an activity to Determine something in the future with models or using a particular method is based on the data or past historical data.

Based on the method or approach, forecasting is divided into 2 there is quantitative forecasting, using mathematical models or statistical models and historical or causal variables of data to forecast demand. And qualitative forecasting, using intuition, personal experience and based on opinion (judgment) from Reviews those who do the forecast.

2.2 *Autoregressive Integrated Moving Average* (ARIMA) Method

The time series of data can be used to analyze the patterns of the data that will be needed in forecasting in the future. One method used to forecast time series of data is ARIMA. ARIMA is a method that produces predictions based on the synthesis of historical patterns of data (Arsyad 1995). ARIMA has several advantages items, namely, easy in the formation of the models, faster in the model building, no need for training such as ANN (Artificial Neural Network), the results are Easily interpreted, Because The models coefficients are known, so it can be seen the influence of each predictor on the output of the model (Hagen 2006). ARIMA has a fairly high level of forecasting accuracy Because after experiencing the level of measurement of forecasting errors MAE (Mean Absolute Error) the value is close to zero (Francis and Hare 1995).

ARIMA (Autoregressive Integrated Moving Average) is a models developed by George Box and Gwilyn Jenkinson the which is applied for time series analysis and forecasting of data, so this is a model Often Referred to as the Box-Jenkins models. ARIMA is a technique to find the most suitable pattern from a group of the data (curve fitting), by making full use of past and present the data for accurate short-term forecasting (Pankartz 1983). This method can be used only on stationary time series of data (Pankratz 1983). This method consists of three steps items, namely the model identification, parameter estimation, and forecasting (Pankratz 1983).

The ARIMA models is a combination of the autoregressive (AR) models and the Moving Average (MA) models with the Data having a differencing process (differentiation) as many times as d. Regression usually connects two different variables, but for Autoregressive (AR) regression is done to itself or connects the previous values to the variable itself, while the Moving Average (MA) models is one of the simple technical analysis methods by looking for moving averages from a variable over a period of time that is affected by errors or residuals at present and in the past.

2.3 *Autocorrelation Function* (ACF)

According to Makridakis, Wheelwright & Hyndman, (1998). ACF is a determinant of identification of archetypes that describe the data that can also identify the data that has stationary, random, and seasonal properties. ACF Explains how much the data correlation is sequential in the time series. Here are the steps of ACF:

1) ACF sample value at k lag:

$$\text{ACF } (rk) = \frac{\sum_{t=b}^{n-k} (X_t - \overline{X})(X_{t+k} - \overline{X})}{\sum_{t=b}^{n} (X_t - \overline{X})^2} \quad \dots (1)$$

$$\overline{X} = \frac{\sum_{t=b}^{n} X_t}{n-b+1} \quad \dots (2)$$

information:
rk : ACF sample value at lag-k
xt: Periodic series value at time t
b : Differencing order
$\overline{X}$: The data average
n : Amount of the data
k : Time lag

2) ACF error standard value (SACF) at lag k:

$$\text{SACF } (Srk) = \dots (3) \frac{\sqrt{1 + 2 \sum_{k=1}^{k-1} r_k^2}}{\sqrt{n-b+1}}$$

information:

srk : ACF standard error values

3) ACF value (TACF) at lag k:

TACF (trk) = ... (4)$\frac{r_k}{Sr_k}$

information:

trk: Statistical value of ACF

2.4   *Partial Autocorrelation Function* (PACF)

The partial autocorrelation coefficient measures the closeness level of the relationship between Xt and Xt-k while the effect of time lag is 1,2,3, ..., k-1 is Considered constant. In other words, the partial autocorrelation coefficient measures the degree of the relationship between current values and previous values (for A Certain time lag), while the influence of other variable time lag is Considered constant values. Following are the steps of PACF:

1) PACF sample value at k lag:

PACF (RKK) ... (5) $= \{ \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} \, r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} \, r_j} \}$

information:

RKK: PACF sample value at lag-k

rk : ACF sample value at lag-k

2) PACF error standard value (SPACF) at lag k:

SPACF (Srkk) ... (6)$= \frac{1}{\sqrt{(n-b+1)}}$

information:

Srkk : PACF standard error values

n: amount of the data

b: differencing order

3) PACF value (TPACF) at lag k:

TPACF (trkk) ... (7)$= \frac{r_{kk}}{Sr_{kk}}$

information:

trkk: Statistical value of PACF

2.5   AR (*autoregressive*) Method

The AR models is a models for predicting Xt past as a function of the data namely t-1, t-2, ..., tn. The AR model of equation can be written as follows (Wei, 1990):

$xt = \mu + \emptyset_1 Xt\text{-}1 + \emptyset_2 Xt\text{-}2 + .. + \emptyset_p Xt\text{-}p + et$ ...(8)

information:

xt : Data at-t

μ   : constanta

∅p : Autoregressive parameters at-p

et: Error value

Estimation of autoregressive parameters can be used matrix multiplication method. Following is the formula of the matrix multiplication:

$$\hat{\beta} = (Z'Z)^{-1} Z'Y ...(9)$$

$$Z = \begin{bmatrix} 1 & X_p \\ 1 & X_{p+1} \\ \vdots & \vdots \\ 1 & X_{n-1} \end{bmatrix} ; Y = \begin{bmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_n \end{bmatrix} ; \hat{\beta} = \begin{bmatrix} \mu \\ \emptyset_1 \\ \vdots \\ \emptyset_p \end{bmatrix} ... (10)$$

information:

Xp : Data at-p

n : Amount of the data

β : Parameter estimation

μ : constanta

$\emptyset_p$: Autoregressive parameters at-p

2.6   MA (*Moving Average*) Method

In the order of q, it states that a model of the which is a distortion of past observations with observations of t-time. The function form of the equation for the MA models in order (q) is (Wei, 1990):

$Xt = \mu + et -\Theta1et-1 + \Theta2et-2 ... + \Theta qet-q... (11)$

information:

μ : constanta

Θq: moving average parameter at-q

et : Error value-at-t

Estimation of MA parameters can be determined by the matrix multiplication method. The following formula of the matrix multiplication method:

$\hat{\beta} = (Z'Z)^{-1}Z'Y... (12)$

$$Z=\begin{bmatrix} 1 & X_q\text{-}X_{q+1} \\ 1 & X_{q+1}\text{-}X_{q+2} \\ \vdots & \vdots \\ 1 & X_{n\text{-}1}\text{-}X_n \end{bmatrix} ; Y=\begin{bmatrix} X_{q+1} \\ X_{q+2} \\ \vdots \\ X_n \end{bmatrix} ; \hat{\beta}=\begin{bmatrix} \mu \\ \emptyset_1 \\ \vdots \\ \emptyset_p \end{bmatrix}... (13)$$

information:

Xq: Data at-q

n : Amount of the data

β : Parameter estimation

μ : constanta

Θq: moving average parameter at-q

2.7   ARMA (*Autoregressive and Moving Average*) Method

According to Hanke and Wichern (2009), the ARMA method of the order p and q (AR (p) and MA (q)) is a combination of the Autoregressive and Moving Average models. Here's the equation from the ARMA models:

$Xt = \mu + \emptyset1Xt-1 + \emptyset2Xt-2 + ... + \emptyset pXt-p + et -\Theta1et-1 + \Theta2et-2 ... + \Theta qet-q... (14)$

information:

xt : Data at-t

μ : constanta

op : Autoregressive parameters at-p

Θq: Moving average parameter at q

et : Error value-at-t

2.8   Error Calculation

Error calculation is used to test the forecasting errors in the ARIMA method. There are Several ways to test Reviews These errors, one of the which is using the MSE (Mean Square Error). The MSE models is calculated by summing the squared individual errors and dividing them by the number of observations. The following is the formula from MSE:

$MSE= \frac{\sum (X_t\text{-}\bar{X}_t)^2}{n}... (15)$

information:

n: amount of the data

xt : Periodic the data values

$\bar{X}_t$: Forecast value models

III. SYSTEM ANALYSIS AND SYSTEM DESIGN

3.1 Analysis

This system analysis stage has the task of defining the system problems, analyzing system requirements and related elements needed in the process of forecasting sales shoe using the ARIMA method According to consumer needs.

3.2 problem Analysis

Most people use every day shoes as footwear to do activities like work. For this reason, shoe manufacturers always try to create and produce shoes to meet consumer needs.

Erratic footwear production every day makes producers Often experience problems in terms of sales. For example, when sales are going up, but the stock of shoes available is small, and vice versa when sales are decreasing, some shoes stock is still very much. The problem is Because producers can not estimate the sales that will occur in the future, as a result they are less precise in producing shoes.

By utilizing the Autoregressive Integrated Moving Average or ARIMA method, a shoe sales forecasting application can be made that can help shoe manufacturers to estimate sales in the next period. With this application, it is expected that shoe production can be Carried out efficiently.

3.3 Data Analysis

The Data used in this study Came from the sales of data of one of the shoe manufacturers in the city of Mojokerto items, namely UD. Ward that has three categories: marching band, fashion and marriage. Each of the categories Reviews These amounted to 731 Data Obtained over 2 years (daily) from 2015 and 2016. The following are examples of 12 the data of shoe of data taken from January 2015:

Table1 training Data

Data at-	date	month	Category		
			Drum band	fashion	Marriage
1	1	january	95	53	32
2	2		70	34	25
3	3		93	15	14
4	4		64	37	18
5	5		87	50	30
6	6		120	73	21
7	7		73	41	15
8	8		95	30	18
9	9		62	52	28
10	10		87	60	10
11	11		118	26	17
12	12		74	12	25

From the table above, it can be taken an example that in January 1, the Data Obtained were shoes for the marching band category of 95 pairs, the fashion category was 53 pairs and the wedding category was 32 pairs. The period of data is taken up to the 12th of data only as an example.

3.4 Flowchart of System Design

The overall flow chart of the system is a description of the processes related to the running of the system. This flow chart will show the stages of the ARIMA method used in the system. Here is a picture of the system flow diagram:
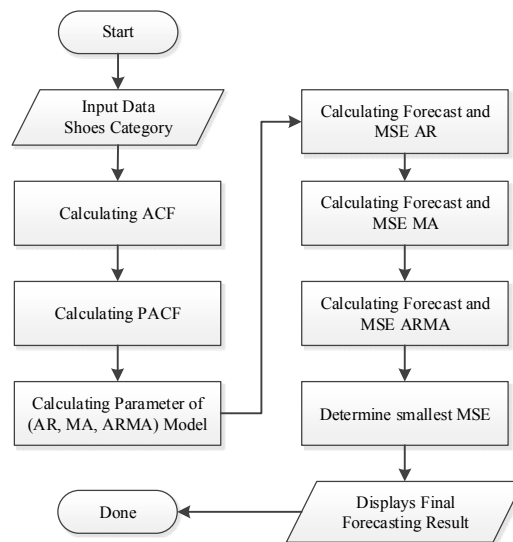
figure 1 *ARIMA* method Flowchart

From the flowchart image above it can be seen that the initial stage is to enter the shoe of data to be processed. After that the the data is processed by calculating the ACF, PACF, and Determining the models parameters. Then the system will calculate the forecasting of each models. After Obtaining the forecasting results for each models, the models with the smallest MSE value is selected. The smallest MSE models is the final result of the ARIMA forecasting.

3.5 context Diagram

Context Diagram is a general description of the system, context diagram system in this final project can be seen in Figure 2:
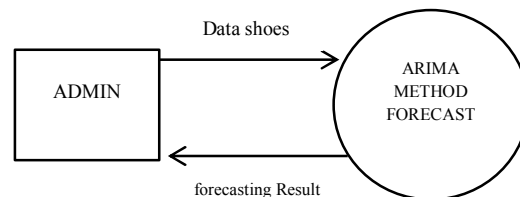
figure 2 context Diagram

The explanation of this context diagram in the system is that only one party has the right to use the application system items, namely admin. This admin manages the entire system from the data input to the generating forecasting.

3.6 Data Flow Diagram Level 1

This DFD level is the next process of the context diagram, in this level of the Data is explained in more detail, DFD Level 1 on this system can be seen in Figure 3:
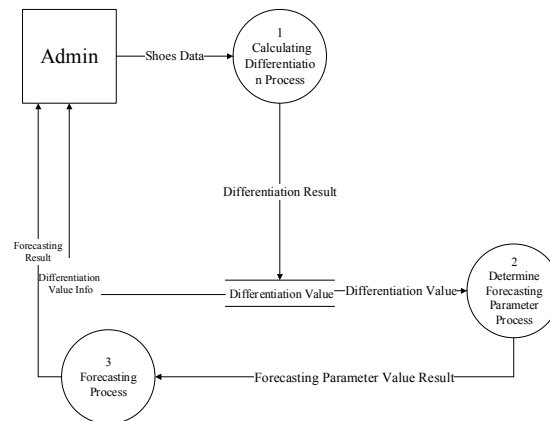
figure 3DFD Level 1

3.6.1   Data Flow DiagramLevel 2 (Calculating Differentiation Process)

This DFD level is the next process of DFD Level 1, in this level of the Data is explained in more detail, DFD Level 2 in this system can be seen in Figure 4:
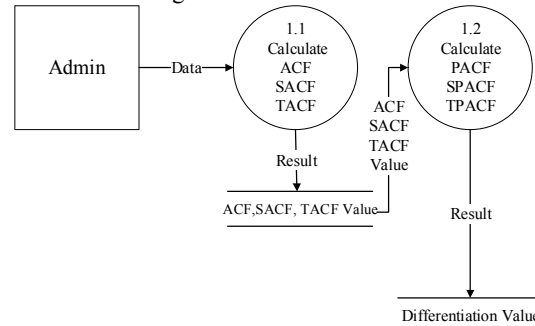
figure 4DFD Level 2 (Calculating Differentiation Process)

3.6.2   Data Flow DiagramLevel 2 (Determine Forecasting Process Parameters)
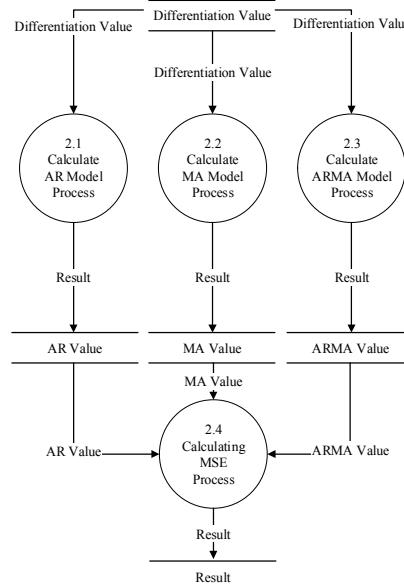
figure 5DFD Level 2 (Determine Forecasting Process Parameters)

3.6.2 Data Flow DiagramLevel 2 (Forecasting Process)
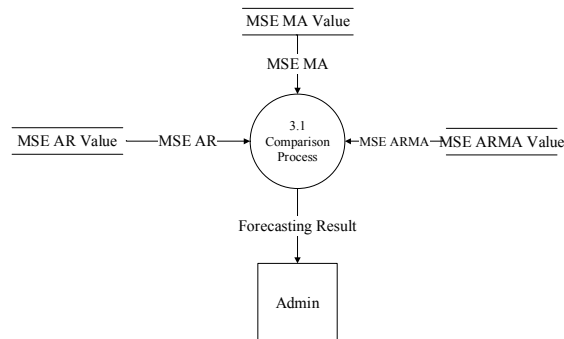
figure 6DFD Level 2 (Forecasting Process)

3.7 *Entity Relationship Diagram* (ERD)

ERD is a models to explain the relationship between the data in a database based on the basic data objects that have relationships between relations. ERD to the model the data structures and relationships between the data, to describe them are used some notations and symbols.

ERD design in this case Explains the relationships between attributes where this attribute will be used for calculation in the ARIMA (Autoregressive Integrated Moving Average) method, Including shoe types and others. For more details will be explained in Figure 7 below:
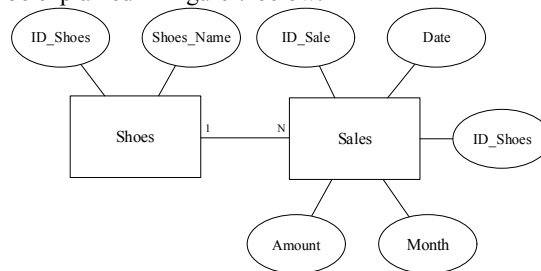
figure 7ERD diagram

IV. SYSTEM IMPLEMENTATION
4.1 interface Implementation

Interface implementation is a form or menu in a system. The form or menu consists of:
a) Main page
b) Master Data page
c) Data Form
d) First Process Form
e) ACF Process Form
f) SACF Process Form
g) PACF Process Form
h) SPACF Process Form
i) ARIMA AR Process Form
j) ARIMA MA Process Form
k) Final Result Form
l) graph Form

As for one menu display on the application:
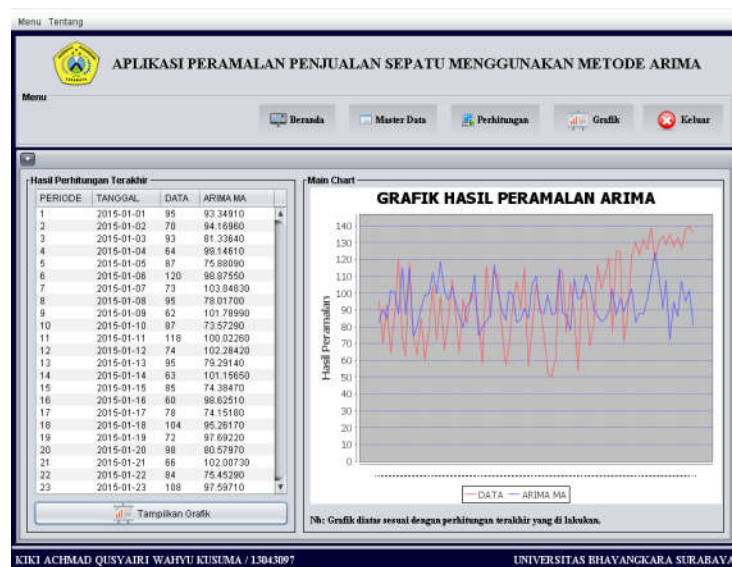
figure 8 graph Form

V. TESTING AND DISCUSSION
5.1 Testing

The previous chapter has explained the implementation of the online shoe sales forecasting with ARIMA method of data daily from 2015 to 2016. Previously UD. Ward had never shoe sales forecast for the next period, using past data. They only focus on production According to consumer demand without taking into account the possibility of subsequent sales. After applying the system using the ARIMA forecasting method, the forecast results for the following year are Obtained.

The following is the test results from the marching band category of data using the ARIMA method:

figure 9 Drumband Category Forecasting Result

From the results of the forecasting calculation above, then the smallest MSE is determined for the marching band category the which will be the final result of forecasting. The following is the smallest MSE result display for the marching band category:
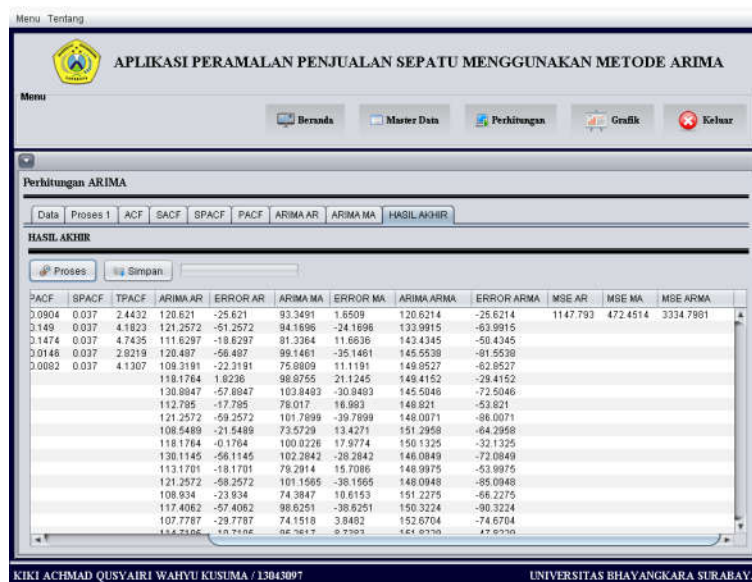
figure 10 The Smallest MSE for Drumband Category

Based on the results above, the smallest MSE falls in the MA models of 472.4514. So the MA models Becomes the final result of the forecasting of the marching band category. From the final results can be seen through the graph as shown below:
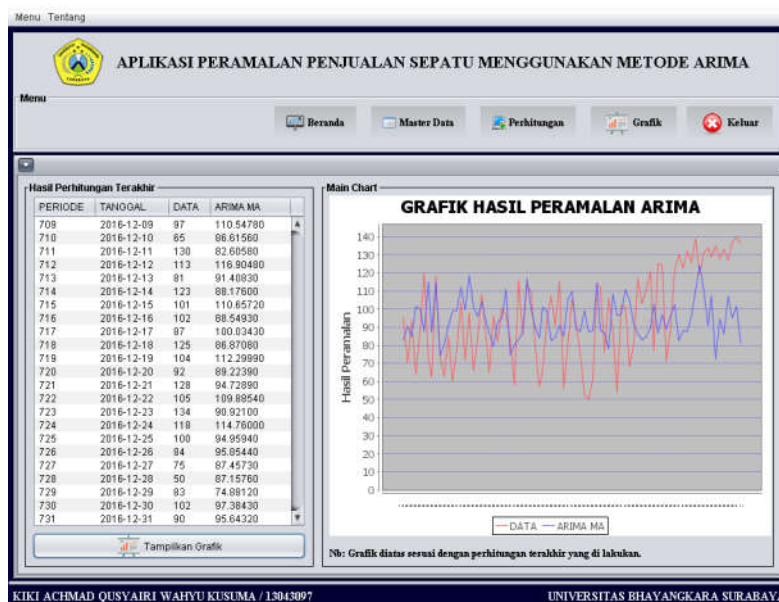

figure 11 Final Result Drumband Category Graph

5.2 Result Discussion

Based on the picture above it can be seen that the forecasting graph tends to move down Compared to the original of data roomates Initially tended to fall but moved up in the final periods. For forecasting results using the MA models as in the picture above, the forecast charts per day has a significant range of Increase or Decrease Compared to the AR models. In the graph the results can be seen from the results of the forecasting that has been done. The forecasting value for the last period or in the period 731 95.6432 from the original is the data of 90 with an MSE of 472.4514.

VI.   CONCLUSION

From the analysis and discussion that has been done in the case study research in the previous chapter, some Conclusions can be drawn based on the research objectives items, namely:

1) From the application of the method used, the ARIMA method can be used to build a shoe sales forecasting application at UD. Ward.
2) From the three forecasting results of each models, the AR, MA and ARMA models are selected one based on the smallest MSE value to be used as the final result of forecasting.

The final result for the marching band category uses the MA models with the forecast value in the last period of 95.6432 and MSE of 472.4514, the fashion category uses the AR models with the forecast value in the last period of 57 172 and the MSE of 304.8306 and marriage categories uses the AR models with forecast values in the last period of 21.4206 and MSE of 118.0681.

REFERENCES

[1] Ali Machmudin. 2012. "Forecasting Air Temperature in Surabaya Using the ARIMA and Artificial Neural Network". Journal of Science and Arts of ITS. Vol.1, No.1: 118-123.

[2] Amira Herwindyani Hutasuhut. 2014. "Making the Decision Support Applications For Raw Material Inventory Forecasting Production and Inject Plastic Blowing Method Using ARIMA (Autoregressive Integrated Moving Average) CV. Asia". Journal of Science and Arts POMITS. Vol.3, No. 2: 169-174.

[3] DT Wiyanti. 2012. "Time Series Forecasting Model Using Radial Basis Function (RBF) and Autoregressive Integrated Moving Average (ARIMA)". Journal of Mathematics. Vol.35, No.2: 175-182.

[4] Helmi Wibowo. 2012. "Short-Term Electric Load Forecasting Based Unclassified Autoregressive Integrated Moving Average Method". ElecTrans. Vol.11, No.2: 44-50.

[5] Syaiful Isnan Robby. 2014. "Temporal Predictions For Emerging Hotspots in Riau Province Using Autoregressive Integrated Moving Average (ARIMA)". Essay. Natural Sciences, Bogor Agricultural University.

[6] Mey Lista Tauryawati. 2014. "Comparison of Methods of Fuzzy Time Series Cheng and Box-Jenkins Method To Predict JCI". Journal of Science and Arts POMITS. Vol.3, No. 2: 34-39.

[7] Rori Muflih Son Harahap. 2014. "Motorcycle Sales Forecasting Analysis in Ngawi with ARIMA and ARIMAX". Journal of Science and Arts POMITS. Vol.3, No. 2: 122-127.

[8] Ni Luh Ayu Kartika Yuniastari. 2014. "Demand Forecasting Method Using Silver Products Simple Moving Average and Exponential Smoothing". Journal of Systems and Information Technology. Vol.9, 1: 97-106.

[9] Nindita Sekar Dini. 2012. "Forecasting Needs Premium Method To Optimize Inventory ARIMAX in Madiun TBBM Territory". Journal of Science and Arts of ITS. Vol.1, No.1: 230-235.

[10] Oktavianis Kusumaningrum. 2012. "Forecasting Requirements Premium Fuel Depot Ampenan-ARIMA Method with Hybrid Neural Network For Inventory Optimization". Journal of Science and Arts of ITS. Vol.1, No.1: 194-201.

[11] Samhis Setiawan. 2016. "8 Understanding Sales According to Experts and Their Kind", In www.gurupendidikan.com, accessed May 15, 2016.