

Motorcycle Parts Sales Forecasting Using Auto-Regressive Integrated Moving Average Model

Martinus Maslim, Ernawati, and Komang Arinanda

Abstract—In the business world, the need for the availability of goods is critical, especially in motorcycle workshops. The goods availability is related to problems with customer trust, loss of capital, and storage warehouse capacity. Therefore, the ability of decision makers to predict the number of sales in the coming period is essential to be able to determine the procurement of goods more precisely. There is a method called Auto-Regressive Integrated Moving Average (ARIMA). This method is one model that can be used to forecast sales based on sales time series data in previous periods. The forecasting implementation with the ARIMA model can be done using the Pmdarima 1.1.0 library for Python. The test in this study uses sales data of 62 motorcycle parts from January 2017 to February 2019. Forecasting is done to help decision-makers in determining the amount of procurement of goods to meet the sales of the next three periods.

Index Terms—Auto-regressive integrated moving average, forecasting, motorcycle parts, time series.

I. INTRODUCTION

The motorcycle is a vehicle that is most in demand by the people of Indonesia. Apart from the relatively low price, the ease of maintenance is also a separate consideration. The Central Statistics Agency (BPS) shows the number of motorbikes in Indonesia in 2016 has reached more than 105 million units. Since 2012 - 2016, the number of bikes has increased by 8.32 percent per year [1]. These statistics show that every year, motorcycle users are growing. Along with the growth of motorcycle users, the need for repair services or so-called workshops and sales of motorcycle parts is also increasing.

In the field of selling motorcycle parts, stock availability is critical. In addition to increasing revenue, stock management is also essential to maintain consumer confidence. The sellers must ensure the availability of stock so as not to suffer losses. Auto parts must be available when the consumer needs the spare parts.

Too little amount of stock makes it possible to run out of stock, which results in not maximizing store revenue. Conversely, the amount of too much stock allows the capital turnover to be wrong. These happen because capital settles into stock for a long time. In addition to the related capital turnover, excessive amounts of stock can cause a buildup of storage warehouses.

Determination of the amount of optimal stock procurement

Manuscript received July 10, 2019; revised October 7, 2019.

Martinus Maslim, Ernawati, and Komang Arinanda are with the Department of Informatics Engineering, Universitas Atma Jaya Yogyakarta, Babarsari Street No. 43, Yogyakarta, Indonesia (e-mail: martinus.maslim@uajy.ac.id, ernawati@uajy.ac.id, 150708161@students.uajy.ac.id).

is complicated to do manually. According to Bounou et al., there are two parameters in inventory management, namely the date of ordering goods and order quantity. These two things greatly influence stock management in a motorbike workshop [2]. Besides those two things, other factors can influence the decision to carry out inventory such as product diversity, number of sales, and the amount of available stock. What is unique here is the diversity of a product, where one product can use on several types of motorbikes. Besides being complicated, determining the amount of stock procurement also requires quite a long time to analyze.

A decision support system is built to help the person to make a decision. In a decision support system, a method for forecasting is needed. One forecasting model is Auto-Regressive Integrated Moving Average (ARIMA). This model was chosen because the data used for prediction is data on the number of sales of each type of time series spare parts. This model will estimate the number of sales for every kind of product for the coming period. By knowing the estimated number of sales, the decision maker can be predicted to supply better parts.

II. RELATED WORKS

This ARIMA method has been widely used in research related to forecasting. One study that uses this method is research from Sabzi in 2017. This study shows the use of the ARIMA method in an expert system to predict the flow of seawater flows [3]. Sufahani et al. In 2017, researched a jewelry company using the Box-Jenkins method [4]. Dritsakis et al. In 2018, conducted a study using the Box-Jenkins method for the problem of predicting unemployment rates in the United States [5]. Ali et al. in 2016 used the Box-Jenkins method in the study to forecast gold prices [6]. Another study conducted by Adedia et al. in 2018 using the Box-Jenkins method to predict miscarriage rates in Ghana [7].

The study conducted by Fransisca aims to solve the company's problems in determining the optimal time and number of orders so that it can meet consumer demand. It is expected that the company can determine the exact amount of spare part orders and when rules must be made to minimize total inventory costs. The analysis was carried out using Microsoft Excel software. In this study analyzing spare parts with Honda brand category A. The data used is data from January 2010 - December 2010 [8].

The study conducted by Ervinasari explains the development of a currency exchange prediction application. The application developed can predict the exchange rate of a currency. The data analyzed is time series data. The methods used are Single Moving Average, Double Moving Average,

and Haar Wavelet [9].

Another study conducted by Hayuningtyas explains inventory forecasting uses the Weighted Moving Average method and the Double Exponential Smoothing method. The research was conducted at Arga Medical stores that are engaged in selling medical devices. The study was conducted with an analysis of sales data during 2016 using two methods, namely the Weighted Moving Average and Double Exponential Smoothing. In this study, the Weighted Moving Average method produces better forecasting than the Double Exponential Smoothing method, with a lower error value [10].

III. FORECASTING WITH ARIMA MODEL

A. Dataset Characteristics

Forecasting with ARIMA uses periodic data series. Regular series or time series is a series of observations or observations taken at successive times. The time of data retrieval can be in periods in seconds, days, years, and other time units. An example of periodic data is the number of goods sent from the factory every month the number of accidents on the road every week, the amount of daily rainfall and hourly observations in the chemical process [11]. Periodic series are written with Z_1, Z_2, \dots, Z_n with Z are variables obtained in a particular time sequence. The dataset used in this study is monthly sales data of 62 different parts for 25 months from January 2017 to February 2019 at XYZ auto parts stores. The sales data for these parts include parts, which have a low, medium, and high monthly sales level can be seen in Fig. 1.

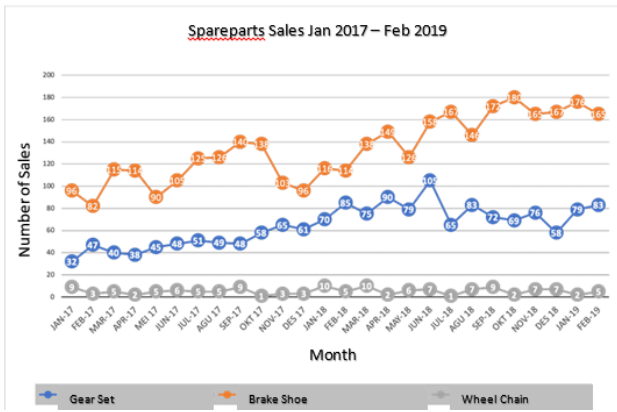


Fig. 1. Example of datasets.

Auto parts have sales data for 25 months (January 2017 - February 2019). Sales data for the first 22 months (January 2017 - November 2018) are used as training data to create ARIMA models. The next three months sales data (December 2018 - February 2019) are used to test the selected ARIMA model. So that it can be used to determine the value of the Mean Square Error (MSE) of the model.

B. Data Stationarity

The forecasting process is carried out if the periodic data is stationary. Information is fixed if the variant, average, and auto-correlation are constant over time. Data stationarity testing can be done using the Augmented Dickey-Fuller Test or ADF test [12].

H_0 : Not stationary data

H_1 : Fixed data

H_0 is rejected if the ADF test value is $<$ critical value, with a significance level.

Data that is not stationary can be handled by differentiating. The amount of differentiation made on the data is denoted by d . The first form of differentiation ($d = 1$) is:

$$\nabla Z_t = Z_t - Z_{t-1} \quad (1)$$

The second form of differentiation ($d = 2$) is:

$$\nabla^2 Z_t = \nabla Z_t - \nabla Z_{t-1} \quad (2)$$

with:

- Z_t : t -time observation
- Z_{t-1} : observation time ($t-1$)
- ∇Z_t : first differentiation at time t
- $\nabla^2 Z_t$: second differentiation at time t
- ∇Z_{t-1} : first differentiation at time ($t-1$)

C. ARIMA Model

The Auto-Regressive Integrated Moving Average (ARIMA) model is the most common model class for forecasting periodic data. ARIMA can be made stationary by differentiating and possibly together with nonlinear transformations such as logging or deflating if needed. Lags of stable series in the forecasting equation are called "autoregressive." Delays of forecasting errors are called "moving average." A periodic series that requires a differentiation process to make the series stationary is referred to as the "integrated" version of the stationary set. The random walk model, casual trend, autoregressive, and exponential smoothing are individual cases of the ARIMA model [13]. The non-national ARIMA model has an ARIMA notation framework (p, d, q).

with:

- p : the number of autoregressive.
- d : the number of non-seasonal differences to make the series stationary.
- q : the number of lag forecasting errors in the forecasting equation.

The equation of the types of ARIMA models are:

a. ARIMA ($p,0,0$)

$$x_t = \mu' + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + e_t,$$

ARIMA process with μ' a constant value ϕ_p is an autoregressive parameter (p) and e_t is the error value at t .

b. ARIMA ($0,0,q$)

$$x_t = \mu' + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-k}$$

A moving average process θ_1 until θ_q parameters moving average e_{t-k} is the current error value $t - k$ and is a constant.

c. ARIMA (0,1,0)

$$x_t = \mu' + X_{t-1}$$

This ARIMA model is called a random walk. The random walk model can use to not fixed data. The model is a which is a particular condition in the AR model (1) with the autoregressive coefficient value is 1.

d. ARIMA (1,0,1)

$$x_t = \mu' + \phi_1 X_{t-1} + e_t - \theta_1 e_{t-1}$$

It is a general model for a mixture of AR (1) pure and MA (1) pure, which is not differentiated. This model is also called the ARIMA model.

e. ARIMA (1,1,1)

$$(1 - B)(1 - \phi_1 B)X_t = \mu' + (1 - \theta_1 B)e_t$$

First differentiation AR(1) MA (1)

D. Criteria for Selecting the Best Model

To assess the quality of a good model can use Akaike Information Criterion (AIC). AIC is a method that can be used to select the model found by Akaike based on the process of maximum likelihood estimation (MLE). By using the AIC method, the best model is the model that has the smallest AIC value. The AIC value can be obtained using the following formula:

$$AIC = -2(\log - likelihood) + 2K \quad (3)$$

with :

- K : the number of parameters estimated in the model
- $\log-likelihood$: model match size (obtained from statistical results)

E. Library Pmdarima 1.1.0 for Python

Pmdarima for Python is a library that is developed based on functionality. Previously existed in the language R. Pmdarima has a function to select the optimal ARIMA parameter then returns the ARIMA model that is already fit to the given data series. This function is called auto ARIMA. Auto ARIMA works by doing differencing tests to determine orders d. In this study, the differencing method used is Augmented Dickey-Fuller and then fitting the model to the given interval p and q values. In this study, the interval values of p and q have given 0 – 2. In this study, auto ARIMA uses AIC (Akaike Information Criterion) to determine the most optimal ARIMA model and returns ARIMA, which has the smallest AIC value. Next is the code snippet used to get the most optimal ARIMA model with auto ARIMA [12].

```
Model = auto_arima(train_set, trace=True,
                  error_action='ignore',
                  suppress_warnings=True,
                  seasonal=False,
                  information_criterion='aic',
                  return_valid_fits=False,
```

```
d=None, max_d=2,
start_p=0, start_q=0,
max_p=2, max_q=2,
test='adf')
```

Information:

- a. t train_set is a data set used to make a model selection. In this study, the number of train_set is:

$$\text{number of train set} = \text{number of series} - 4 \quad (4)$$

The last four series are not used for model selection but will later be used to see the performance of the model.

- b. Trace, error_action, and suppress_warnings are parameters to display the status of the fitting process and how the system will react if there are a status error and warning.
- c. Seasonal is a boolean that states whether auto ARIMA will run the SARIMA process (Seasonal Autoregressive Integrated Moving Average). In this study, seasonal value is False because the SARIMA model is outside the scope of the study.
- d. information_criterion = 'AIC' indicates that AIC is used to find the most optimal ARIMA model.
- e. return_valid_fits = False will return the most optimal ARIMA model.
- f. d = None indicates that the value of d will be automatically selected based on the results of the test differencing, while max_d = 2 indicates the maximum value given is 2.
- g. start_p and max_p indicate the interval of the given p-value.
- h. start_q and max_q indicate the interval of the given q value.
- i. test = 'adf' indicates that the choice of value d or test differencing used is Augmented Dickey-Fuller.

After obtaining a valid ARIMA model that is already fit, the forecasting process can be carried out using the predict method from the ARIMA model as follows.

```
forecast = model.predict(n_periods=period)
```

As stated earlier, the most optimal ARIMA model will be returned and used to forecast or predict. The n_periods value indicates the number of future periods to be predicted. In this study, the number of future periods predicted is always six, namely the previous three periods, the current period, and the next two periods.

IV. MODEL EVALUATION

A forecast always contains uncertainty. Deviation from forecasting results (Table I) other than due to errors can be caused because no forecasting method can produce forecasting accurately. The method can also be caused because the method or model used cannot predict the trend, seasonal, or cycle components found in the data series. Mean Squared Error (MSE) is one method that can be used to evaluate the model using the formula:

$$MSE = \frac{\sum_{t=1}^n e_t^2}{n}$$

To assess how well the performance of the model chosen in forecasting is done, an error value is calculated. The calculation of the error value used is MSE (Mean Square Error) using the following code.

```
MSE = mean_squared_error(test_set[:test_length]
forecast[:test_length])
```

The test_set parameter is a data frame that contains a set test. The value of this set test is the value of the previous three-period series. The forecast code [: test_length] takes the predicted value for the same period with the test set so that both can be used to calculate the MSE value.

TABLE I: FORECASTING RESULTS

No	Spareparts	Brand	Model	Comparasion						Sales Prediction		
				Dec-18		Jan-19		Feb-19		Mar-19	Apr-19	May-19
				Actual	Prediction	Actual	Prediction	Actual	Prediction			
1	Headlight Back Cover	Federal	ARIMA (1,2,1)	5	4	3	3	2	4	4	4	4
2	Front V'heel Cover	Federal	ARIMA (0,0,1)	2	2	2	2	1	2	2	2	2
3	Headlight Back Cover	Honda	ARIMA (0,2,1)	3	2	1	2	2	2	2	1	1
4	Carburator Repair Kit	Federal	ARIMA (0,2,1)	2	3	2	3	4	3	3	4	4
5	CDI Unit	Federal	ARIMA (0,0,0)	1	1	0	1	1	1	1	1	1
6	CDI Unit	Yamaha	ARIMA (0,0,0)	1	1	2	1	1	1	1	1	1
7	CDI Unit	Suzuki	ARIMA (1,2,1)	2	0	1	0	0	0	0	0	0
8	Cool Starter	Federal	ARIMA (0,1,1)	2	2	3	2	2	2	2	2	2
9	Cool Starter	Yamaha	ARIMA (0,0,1)	1	2	0	2	2	2	2	2	2
10	Disc Pad	Federal	ARIMA (0,2,1)	142	152	106	154	104	156	159	161	163
11	Disc Pad	Yamaha	ARIMA (0,0,1)	109	102	112	93	96	93	93	93	93
12	Disc Brake	Honda	ARIMA (1,2,2)	3	2	2	2	1	2	2	2	2
13	Diso Brake	Yamaha	ARIMA (0,1,1)	2	2	3	2	1	2	2	2	2
14	Kick Starter	Honda	ARIMA (0,1,1)	2	2	3	2	3	2	2	2	2
15	Gasket Full Set	Federal	ARIMA (0,1,1)	4	3	3	3	5	3	3	3	3
16	Gasket Full Set	Yamaha	ARIMA (0,0,0)	2	3	3	3	2	3	3	3	3
17	Gasket Full Set	Suzuki	ARIMA (0,1,1)	3	2	2	2	3	2	2	2	2

V. CONCLUSION

Forecasting sales of motorcycle parts can be made using the ARIMA (Auto-Regressive Moving Average) model. Forecasting is done using sales time series data from January 2017 to February 2019, which is stationed using the Augmented Dickey-Fuller Test (ADF). The selection criteria for the ARIMA model is done using the Pmdarima 1.1.0 library for Python. Based on the results of testing the sales dataset for 62 different parts in the January 2017 to February 2019 period, obtained by the average MSE value (Mean Squared Error) of 61,737.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Martinus Maslim conducted the research and searching the data; Ernawati analyzed the data and make the model; Komang Arinanda make the program and synchronize the model into program; Martinus Maslim and Komang Arinanda wrote the paper; all authors had approved the final version.

REFERENCES

- [1] *Number of Types of Motorized Vehicles in Indonesia*. Jakarta Pusat: Central Bureau of Statistics 2016.
- [2] O. Bounou *et al.*, "Inventory models for spare parts management: A review," *International Journal of Engineering Research in Africa*, vol. 28, pp. 182-198, 2018.
- [3] H. Z. Sabzi *et al.*, "Developing an intelligent expert system for streamflow prediction, integrated in a dynamic decision support system for managing multiple reservoirs: A case study," *Expert Systems with Applications*, vol. 4, pp. 145-163, 2017.
- [4] S. F. Sufahani *et al.*, "Descriptive statistics with Box-Jenkins and marketing research for jewellery company in Malaysia," *Far East Journal of Mathematical Sciences*, vol. 101, no. 10, pp. 2151-2161, 2017.
- [5] N. Dritsakis and P. Klazoglou, "Forecasting unemployment rates in usa using Box-Jenkins methodology," *International Journal of Economics and Financial Issues*, vol. 8, no. 1, pp. 9-20, 2018.
- [6] A. Ali *et al.*, "Forecasting of daily gold price by using Box-Jenkins methodology," *International Journal of Asian Social Science*, vol. 6, no. 11, pp. 614-624, 2016.
- [7] D. Adedia *et al.*, "Box-Jenkins' methodology in predicting maternal mortality records from a public health facility in Ghana," *Open Journal of Applied Sciences*, vol. 8, pp. 189-202, 2018.
- [8] Fransisca, *Analisis Pengendalian Persediaan Spare Part Sepeda Motor Honda di PT Menara Agung*, 2011.
- [9] R. Y. Ervinasari, *Pengembangan Aplikasi Prediksi Nilai Tukar Mata Uang*, 2008.
- [10] R. Y. Hayuningtyas, "Peramalan persediaan barang menggunakan metode weighted moving average dan metode double exponential smoothing," *Pilar Nusa Mandiri*, vol. 13, no. 2, pp. 217-222, 2017.
- [11] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken: John Wiley & Sons, Inc., 2016.
- [12] T. G. Smith. (2018). ARIMA estimators for Python. [Online]. Available: <https://www.alkaline-ml.com/pmdarima>
- [13] R. Nau, "Introduction to ARIMA models," Fuqua School of Business, Duke University, 2018.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).



Martinus Maslim was born in 1990 and received his master of engineering in Information Engineering Universitas Atma Jaya Yogyakarta, Indonesia in 2012. His area of interest includes artificial intelligence and data mining. He is a lecturer of Information Engineering Department in Universitas Atma Jaya Yogyakarta since 2013.



Ernawati was born in 1968 and received her master of engineering in Information Engineering Institut Teknologi Bandung, Indonesia in 1998. Her area of interest includes artificial intelligence and data mining. She is a lecturer of Information Engineering Department in Universitas Atma Jaya Yogyakarta since 1992.



Komang Arinanda was born in 1996 and received his bachelor of computer in Information Engineering Universitas Atma Jaya Yogyakarta, Indonesia in 2019. His area of interest includes software development and data mining. Now he is an employee in startup company in Indonesia since 2019.