

Forecasting the Items Consumption in the Hotel Storage with the Autoregressive Integrated Moving Average Method

Christopher Chandra^{1*}, Alfannisa Annurrullah Fajrin², Cosmas Eko Suharyanto³

^{1,2,3} Department of Informatics Engineering, Faculty of Engineering and Computers
Putera Batam University,
Kepulauan Riau, Indonesia 29444
pb170210005@upbatam.ac.id; alfannisa@puterabatam.ac.id; cosmas@puterabatam.ac.id

*Correspondence: pb170210005@upbatam.ac.id

Abstract - In this era, hotel has storage as a storing space for every kind of items. Items stored in the storage are items being used for the needs of the staffs, also for the needs of hotel's operational. The item consumption is running smoothly with resupply. However, there are often mistakes in resupplying the items. For preventing those several mistakes, a reference is needed to be used for controlling the amount of items arrival (monthly) with minding the amount of items in the storage should be. The reference to be used is the forecast of the item consumption every month. Forecasting was being done with Autoregressive Integrated Moving Average (ARIMA) method. There are five steps needed to build the ARIMA model, such as plot identification, model identification, model estimation, choosing the best model, and prediction (forecast). The input variable to be used in this research is the rime series from the data of storage's item consumption starts from January 2018 until October 2020, and the output variable is the result of the prediction of item consumption in the next period, such as in November to December 2020. The results is subtracted with the number of items left in storage to obtain the minimum amount of item to be entered for the month.

Keywords: Resupply; ARIMA; forecast; consumption; Python.

I. INTRODUCTION

Storage is not a facility that can be separated from operational activities in a company. At this time, the hotel has storage as a room that is used to store various kinds of item, both food and non-food items. Operational activities in storage greatly affect the company's operations, both on the performance and daily needs of employees and visitors.

A common problem that can occur in hotel storage is a sudden shortage or out of stock of item, and the worst thing is the lack of confidence in adding stock (resupply) when receiving news from suppliers that the stock of item that the storage wants to order is insufficient or empty, and even it could also be due to limited budget owned by the company. Even though they know the number of employee's needs constantly, Pacific Palace Hotel storage staffs cannot know directly the needs of the visitors, because the number of hotel's visitors changes from time to time, and even the number of events held at the hotel. Both of the number of visitors and events held are also affect the hotel's administration works. In addition, one of the factors that causes unsure of resupply at Pacific Palace Hotel storage is excessive or uncontrolled resupply of item that may have a low consumption level or a consumption level that has not changed too significantly, or for item that have not paid attention to their expiration date, so that resulting in a lack of budget in resupply. Therefore, it is necessary to take the preventive action by resupply early to prevent these problems.

This action can be supported by prediction of the level of consumption of certain item with the help of processing consumption data from the past recorded previously. Predictions that are usually carried out by every company are periodic predictions, either annually, monthly, weekly or daily. For problems in resupply, monthly predictions can be made, by processing consumption data which is made in the form of monthly periods. The most important thing in making a predictions is not to trigger the company to make the wrong decision if the prediction results obtained have a large error value. This data processing can be done by applying various methods, one of which is Machine Learning with the Autoregressive Integrated Moving Average (ARIMA) method. Machine Learning is a part of

Artificial Intelligence that plays a role in building a system with data. ARIMA is one of the excellent methods used to produce accurate short-term predictions, but not long-term predictions.

One of the studies involving the ARIMA method is a study entitled "Prediction of Short-Term National Staple Prices Using ARIMA". Research conducted by Rasyidi (2017) was based on his opinion that uncontrolled fluctuations in the price of basic commodities can cause losses to both consumers and producers. This research is one of the preventive steps in preventing these problems. The ARIMA model used in several prediction horizons produces a fairly accurate level of prediction, which is 97,78%, but the accuracy of the three main ingredients (shallots, red chilies, curly red chilies) which have the greatest error at the distance of the prediction horizon tends to change rapidly because of being influenced by the holidays on the Hijri calendar.

The researcher reiterates the problems that have been previously stated, that if no anticipation is made at the time of resupply, the hotel operational activities will certainly face obstacles.

II. METHODS

Machine Learning is an approach of artificial intelligence that mimics human behavior to solve problems or perform automation. Like Artificial Neural Network, Machine Learning also has a characteristic that is a training process, so it requires data to be studied (data training) (Putra, 2019).

Predictive research is quantitative research. Quantitative research methods are known as discovery methods because they can develop into a variety of new science and technology. The data used in this research method is in the form of numbers and statistics (Sugiyono, 2012). Predictive research can be carried out through trend studies, which is looking at developments over a certain period of time (present or past), which will reveal trends in the future (Sudaryono, 2015).

Inventory is a common term used to describe everything including the resources of an organization that are stored with the aim of anticipating demand fulfillment (Handoko, 1999). There is also an optimal supply, which means that the number of items stored is the best and profitable level (Lahu & Sumarauw, 2017). Furthermore, supply control is an activity aimed at not depleting existing supplies and maintaining an optimal level so that inventory costs remain stable (Indah et al., 2018).

The methods used in obtaining data were unstructured interviews and observations, which are the part of data collection method classified by Sugiyono (2012). The type of data used by researchers in this study is secondary data, where the data has been recorded in a report, and also it's time series data because of it's recorded with time history.

The method used in this research is the ARIMA method, known as Box-Jenkins. ARIMA is a prediction algorithm based on the idea that the information about past values contained in the time series itself can be used to predict future values (Prabhakaran, 2019).

The researcher use the programming language Python as a tools to do the data processing. Python has several advantages,

which is it is open source, ease of writing and reading, where the syntax is almost close to human language (English), and also quite a lot of packages can be used for scientific computing by scientists and engineers, artificial intelligence, machine learning, and data science (Adriyan, 2019).

The data processing steps in the construction of the ARIMA model are as follows (Nofiyanto et al., 2015) :

1. Identify data plots

The data obtained were plot identified beforehand, so that from the data plot it can be seen the stationarity of the data. Displaying the plot is done with the pandas package (to read the .csv file), the matplotlib by importing the pyplot (to process and display the plot). If the data plot is not stationary, it is necessary to do a differencing process. Differencing a series is simply finding the difference of observations that are calculated mathematically to form a new time series (Rasyidi, 2017). The stationary identification of a plot is done with the Adfuller (ADF) test imported from the statsmodels package in Python. The data used is the cumulative consumption data from January 2018 until October 2020.

2. Identify the model

After the mean data reaches stationarity, identification of the terms on the ACF (Autocorrelation Function) plot is carried out for the AR (Autoregressive) terms and PACF (Partial Autocorrelation Function) for the MA (Moving Average) terms. The two plots can be used to identify several possible models that are suitable for use in time series predictions. Identification of the terms of the plot is done by looking at the lag that is outside of the shading area (not past the significant limit). The two plots are also generated from the statsmodels package in Python by importing plot_acf and plot_pacf.

3. Estimating the model

If a suitable model has been determined and the estimation of the parameters has been made, the significance of the coefficient is tested. If the coefficient of the model is not significant, then the model cannot be used for prediction. Estimation on the model is done with the ARIMA summary generated from Python, by using ARMA and ARIMA which are imported from the statsmodels package.

4. Selection of the best model

Some things are needed to be considered in choosing a model:

- a. The parsimony principle, which means that the model must be as simple as possible, where the parameters contained in it are few so that the model is more stable. The model parameter chosen is the minimum AIC (Akaike's Information Criterion) value of all possible model parameters, which can be seen from the resulting ARIMA summary.
- b. Ensure that the model must approach the underlying assumptions. Each parameter chosen must be considered whether the p-value of each AR and MA term listed in the ARIMA summary

is significant or not. If there is a p-value in one of the orders that is not significant, then the model parameter cannot be selected.

The 3rd and the 4th step could also be switched by choosing a parameter with the minimum or the lowest AIC value of all possible parameters, then the significance of each terms (AR and MA) in the chosen parameter is tested.

5. Prediction (Forecasting)

The model that is considered the best can be used for prediction, especially the prediction of values in the next several periods. The predictions are processed also with the ARIMA imported from statsmodels package in Python, just like generating the ARIMA summary. The predicted value is used as a consideration to determine the amount of item that will enter in that period.

III. RESULT AND DISCUSSION

In this research, the highest level of consumption of 5 items was taken from 14 items used in hotel office operations, starting from January 2018 to October 2020. The highest consumption rate was taken from the total consumption of each item for 34 months, the highest consumption figure in between 34 months, and the average consumption per month is assisted by Microsoft Excel by using the Sum, Max, and Average functions. It turns out 6 items was taken because of the 1st rank until the 4th rank taken item are same (tissue, water gallon, pen and stapler filling) while the 5th rank taken item in rank based on Sum and Average (HVS paper) is different with the Max one (masking tape). All time series datasets of 6 items are processed to identify the plot.

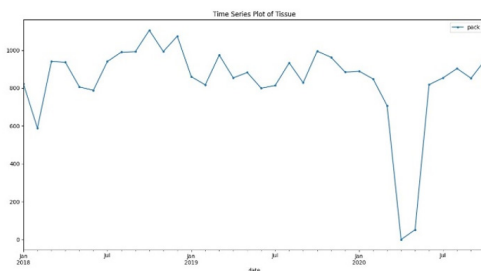


Figure 1 Time Series Plot of Tissue Consumption

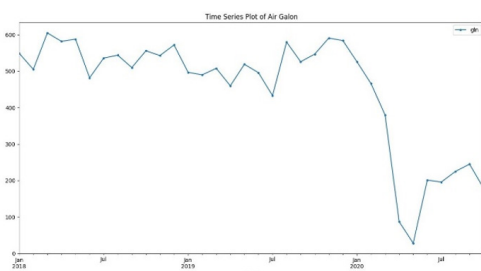


Figure 2 Time Series Plot of Water Gallon Consumption

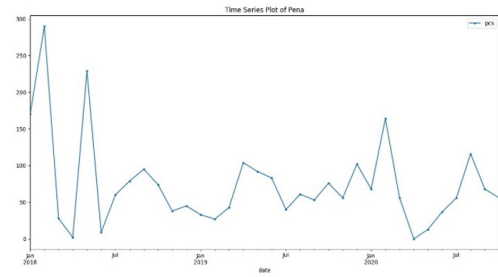


Figure 3 Time Series of Pen Consumption

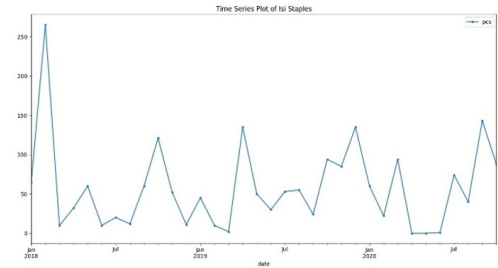


Figure 4 Time Series of Stapler Filling Consumption

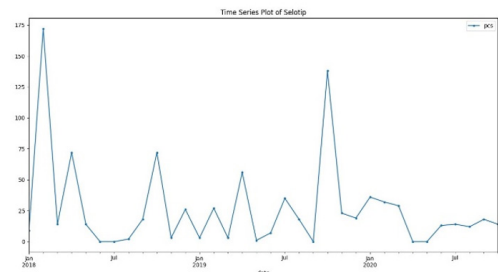


Figure 5 Time Series of Masking Tape Consumption

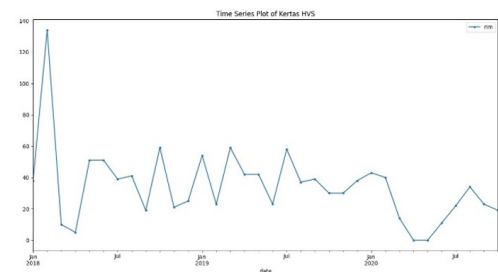


Figure 6 Time Series of HVS Paper Consumption

In the plot of the six items, there was a decrease in the level of consumption starting in March, until it reached 0 in April 2020 and May 2020. The decrease was caused by the hotel experiencing a bad business condition caused by the COVID-19 pandemic, so the reduced number of hotel visitors had an impact on administrative activities at the hotel office, as well as a reduction in the number of employees working there. Due to the declining state of the hotel business, Hotel Pacific Palace decided to temporarily suspend its operational activities starting from April to May 2020.

Identification of stationarity on plots can be done with the Augmented Dickey Fuller (ADF) test. The null hypothesis (h_0) of the ADF test is that the time series is non-stationary. If the p-value obtained from the test is lower than the significant level ($< 0,05$), then h_0 is rejected and it can be concluded that the time series is stationary (Prabhakaran,

2019). So, if the p-value > 0,05, it is followed by the search for the order of differencing. The following is a table of the results of the acquisition of p-value:

Table 1 ADF Test Result

Item	ADF Statistic	P-Value
Tissue	-3,371168	0,011982
Water Gallon	-1,020827	0,745593
Pen	-3,824425	0,002667
Stapler Filling	-5,510485	0,000002
Masking Tape	-6,590583	0
HVS Paper	-5,462222	0,000003

The only time series that requires differencing is the series of water gallon (p-value > 0). However, even though the five other series are stationary, differencing is still needed for pen, stapler filling, masking tape, and HVS paper series because the each AR and MA term obtained respectively from the original data plot are 0 and 0, so that the prediction model parameter becomes (0, 0). These parameters cannot be used to build a prediction model.

The ideal differencing term is the minimum amount of differencing needed to obtain a series that is close to stationary, where the autocorrelation plot reaches zero rapidly. If in autocorrelation there is a positive value in a large number of lags, then the series needs to be further differencing. But if the first lag of autocorrelation is negative which is very large then the series is excessively differenced. Determination on the differencing term can also be done with ADF test for the differenced series. One example of research that follows this method is a study conducted by As'ad et al. (2017), in which the ADF test was also carried out on the differenced series, to get the differencing term (3), with the p-value from each 1st, 2nd and 3rd differenced series are 0,4427, 0,06441 and 0,04239.

Table 2. 1st Lag Value of ACF

Table 2(a)

Item	ACF 1 st Lag Value
Tissue	0,5518278

Table 2(b)

Item	ACF 1 st Lag Value	
	1 st Difference	2 nd Difference
Water Gallon	-0,01022753	-0,43086099
Pen	-0,4484332	-0,52482728
Stapler Filling	-0,47857533	-0,51130662
Masking Tape	-0,60250095	-0,63886256
HVS Paper	-0,4688342	-0,46373431

Table 3. P-Value (ADF Test) in Differenced Series

Item	P-Value	
	1 st Difference	2 nd Difference
Water Gallon	0.000002	0.045159
Pen	0.000081	0.001745
Stapler Filling	0	0.144796
Masking Tape	0	0.000131
HVS Paper	0	0

The following is the table of the possible terms of difference of the five evaluated series :

Table 4. Differencing Term

Item	Differencing Term (d)
Water Gallon	1 and 2
Pen	1
Stapler Filling	1 and 2
Masking Tape	1 and 2
HVS Paper	1 and 2

Determining the AR (p) and MA (q) term respectively is done by identifying significant lag, which has a value that is outside the shading area (does not cross the significant limit) on the partial autocorrelation and autocorrelation plots. If the lag value is outside the shading area, it can be concluded that the lag is significant.

Table 5. Possible AR and MA Terms

Table 5(a)

Item	PACF (p)	ACF (d)
Tissue	0 and 1	0

Table 5(b)

Item	1 st Difference		2 nd Difference	
	PACF (p)	ACF (d)	PACF (p)	ACF (d)
Water Gallon	0	0	0, 1 and 2	0 and 1
Pen	0, 1 and 2	0 and 1	0, 1 and 2	0 and 1
Stapler Filling	0 and 1	0 and 1	-	-
Masking Tape	0 and 1	0 and 1	0, 1 and 2	0 and 1
HVS Paper	0, 1 and 2	0 and 1	0, 1 and 2	0 and 1

Based on the results of the evaluation of differencing, AR (PACF) and MA (ACF) terms, several possible parameters for the ARIMA model are :

Table 6. Possible ARIMA Parameter

Item	Parameter
Tissue	(0, 1)
	(1, 0)
	(1, 1)
	(0, 2, 1)
	(1, 2, 0)
Water Gallon	(1, 2, 1)
	(2, 2, 0)
	(2, 2, 1)
	(0, 1, 1)
	(0, 2, 1)
Pen	(1, 1, 0)
	(1, 1, 1)
	(1, 2, 0)
	(2, 1, 0)
	(2, 1, 1)
	(2, 2, 0)
	(2, 2, 1)

Stapler Filling	(0, 1, 1)
	(1, 1, 0)
	(1, 1, 1)
Masking Tape	(0, 1, 1)
	(0, 2, 1)
	(1, 1, 0)
	(1, 1, 1)
	(1, 2, 0)
	(1, 2, 1)
	(2, 2, 0)
	(2, 2, 1)
HVS Paper	(0, 1, 1)
	(0, 2, 1)
	(1, 1, 0)
	(1, 1, 1)
	(1, 2, 0)
	(1, 2, 1)
	(2, 1, 0)
	(2, 1, 1)
	(2, 2, 0)
(2, 2, 1)	

Pen	(0, 1, 1)	372,682
	(0, 2, 1)	384,442
	(1, 1, 0)	384,001
	(1, 1, 1)	374,672
	(1, 2, 0)	401,487
	(2, 1, 0)	12223,441
	(2, 1, 1)	377,161
	(2, 2, 0)	2249,994
Stapler Filling	(2, 2, 1)	370,619
	(0, 1, 1)	367,957
	(1, 1, 0)	376,454
Masking Tape	(1, 1, 1)	380,2
	(0, 1, 1)	341,688
	(0, 2, 1)	360,708
	(1, 1, 0)	347,783
	(1, 1, 1)	347,101
	(1, 2, 0)	367,803
	(1, 2, 1)	349,962
	(2, 2, 0)	357,805
	(2, 2, 1)	347,335
	HVS Paper	(0, 1, 1)
(0, 2, 1)		326,34
(1, 1, 0)		321,425
(1, 1, 1)		327,256
(1, 2, 0)		342,555
(1, 2, 1)		318,331
(2, 1, 0)		315,822
(2, 1, 1)		309,307
(2, 2, 0)		327,691
(2, 2, 1)		314,381

Parameter (1, 2, 1) was included in one of the parameters from pen's ARIMA model, but this parameter cannot be used for the ARIMA model because there is an error when processing ARIMA summary with the parameter itself, which the error message itself is : "The computed initial MA coefficients are not invertible You should induce invertibility, choose a different model order, or you can pass your own start_params."

Parameter estimation is done by observing the AIC (Akaike's Information Criterion) value of each model, and also paying attention to the p-value in each of the AR and MA terms. The following is the AIC table of the parameter for each models in each item :

Table 7. AIC Value of All Parameters

Item	Parameter	AIC
Tissue	(0, 1)	453,649
	(1, 0)	458,786
	(1, 1)	455,621
Water Gallon	(0, 2, 1)	381,007
	(1, 2, 0)	393,307
	(1, 2, 1)	383,005
	(2, 2, 0)	389,269
	(2, 2, 1)	384,707

Table 8. Chosen Parameter

Item	Parameter
Tissue	(0, 1)
Water Gallon	(0, 2, 1)
Pen	(2, 2, 1)
Stapler Filling	(0, 1, 1)
Masking Tape	(0, 1, 1)
HVS Paper	(0, 1, 1)

From the table above, the best model parameter for each item is selected based on the smallest AIC value. The p-value of each AR and MA terms in each chosen models are as follows :

Table 9 P-Value of AR and MA Terms

Item	P-Value			
	AR 1	AR 2	MA 1	MA 2
Tissue	-	-	0	-
Water Gallon	-	-	0	-
Pen	0	0	0	-
Stapler Filling	-	-	0	-
Masking Tape	-	-	0	-

HVS Paper	-	-	0	-
-----------	---	---	---	---

The p-value of each AR and MA terms from the chosen models are far below the significant level ($\ll 0,05$). The coefficient value of each terms (AR and MA) for each parameter of the selected model are as follows :

Table 10. Coefficient Value

Item	Coefficient				
	Const.	AR 1	AR 2	MA 1	MA 2
Tissue	847,288	-	-	1,110	-
Water Gallon	-0,989	-	-	-1	-
Pen	0,642	-0,694	-0,611	-1	-
Stapler Filling	-0,109	-	-	-1	-
Masking Tape	-0,738	-	-	-1	-
HVS Paper	-0,949	-	-	-1	-

From the coefficients above, the ARIMA equation for each model can be written as follows :

Table 11. ARIMA Equation

Item	Equation
Tissue	$Z_t = 847,288 + \epsilon_t - 1,110\epsilon_{t-1}$
Water Gallon	$Z_t = -0,9893 + 2Z_{t-1} - Z_{t-2} + \epsilon_t + \epsilon_{t-1}$ $Z_t = 0,642 + 2(1 - 0,694)Z_{t-1} + (-0,611 + 0,694)Z_{t-2} + 0,611Z_{t-3} + \epsilon_t + \epsilon_{t-1}$
Pen	$Z_t = 0,642 + 0,612Z_{t-1} + 0,083Z_{t-2} + 0,611Z_{t-3} + \epsilon_t + \epsilon_{t-1}$
Stapler Filling	$Z_t = -0,109 + Z_{t-1} + \epsilon_t + \epsilon_{t-1}$
Masking Tape	$Z_t = -0,738 + Z_{t-1} + \epsilon_t + \epsilon_{t-1}$
HVS Paper	$Z_t = -0,949 + Z_{t-1} + \epsilon_t + \epsilon_{t-1}$

The chosen parameters are used to create the prediction of item consumption number of those 6 items (tissue, water gallon, pen, stapler filling, masking tape and HVS paper) for 2 months forward (November - December 2020). The prediction result are as follows, in which all decimal numbers are rounded up :

Table 12. Prediction Result (Nov-Dec 2020)

Item	November 2020		December 2020	
	Result	Round	Result	Round
Tissue	1091,77	1092	847,29	848
Water Gallon	152,19	153	124,39	125
Pen	111,58	112	99,7	100
Stapler Filling	54,7	55	54,59	55
Masking Tape	13,57	14	12,83	13
HVS Paper	17,92	18	16,97	17

The prediction results above can be implemented in the resupply of item at the Pacific Palace Hotel's storage.

The number of an item's quantity left at that time must be known. So, to count the quantity number needed for resupply, it was done by subtracting the quantity number got from the prediction result with the item's quantity left in the storage at that time. This predictions is conducted as a practice of applying the economic lot sizing (minimalize the costs per unit) and anticipation supply purpose (prepare the seasonal supply) (Handoko, 1999).

IV. CONCLUSION

From the ARIMA parameters obtained, it can be assumed that there are 3 kind of prediction models, which consist of an MA (Moving Average) model (for tissue (0, 1)), 4 IMA (Integrated Moving Average) models (for water gallon (0, 2, 1); stapler filling (0, 1, 1); masking tape (0, 1, 1), and HVS paper (0, 1, 1)) and an ARIMA model (for pen (2, 2, 1)).

ARIMA processing in Python uses several packages that are usually used in general, which consist of pandas, matplotlib and statsmodels, then adding several other packages as a complement so that processing performance can run normally.

Management of the item resupplying in storage can be done by calculating the number of item that must be entered at that time (the month).

ARIMA method is a predictions method, but it is not suitable to be used to make a long term period. For the example, when the researcher processed the prediction for the following months, there are some predictions result in negative value. The researcher suggest that the ARIMA process redo is needed when there is any update in item consumption data.

For the research, it was the last 10 months (January - October 2020) of consumption data used before the researcher updated it to the last 34 months (January 2018 - October 2020) data. The sample data was added (updated) because there are not any of AR and MA terms identified from the plots, even though the plots was differenced. In the end, those terms was identified after the sample data updated. ARIMA is the one of Machine Learning method, so it is better to use the large scale sample data to make the result of data processing more accurate. The researcher reaffirm that this research can be used as a reference to make the design of resupply decision making system.

REFERENCES

- Adriyan. (2019). Instalasi Python 3 dan IDE atau Anaconda Distribution pada Windows 10. DQLab.
- As'ad, M., Wibowo, S. S., & Sophia, E. (2017). PERAMALAN JUMLAH MAHASISWA BARU DENGAN MODEL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA). JIMP - Jurnal Informatika Merdeka Pasuruan, 2, 20-33.

- Handoko, T. H. (1999). Dasar-dasar manajemen produksi dan operasi (Edisi 1). BPF UGM.
- Indah, D. R., Purwasih, L., & Maulida, Z. (2018). Pengendalian Persediaan Bahan Baku pada PT. Aceh Rubber Industries Kabupaten Aceh Tamiang. *JURNAL MANAJEMEN DAN KEUANGAN*, 7.
- Lahu, E. P., & Sumarauw, J. S. . (2017). ANALISIS PENGENDALIAN PERSEDIAAN BAHAN BAKU GUNA MEMINIMALKAN BIAYA PERSEDIAAN PADA DUNKIN DONUTS MANADO. *Jurnal EMBA*, 5.
- Nofiyanto, A., Nugroho, R. A., & Kartini, D. (2015). Peramalan Permintaan Paving Blok dengan Metode ARIMA. *Konferensi Nasional Sistem & Informatika*.
- Prabhakaran, S. (2019). ARIMA Model – Complete Guide to Time Series Forecasting in Python. <https://www.machinelearningplus.com/time-series/arma-model-time-series-forecasting-python/>
- Putra, J. W. G. (2019). Pengenalan Konsep Pembelajaran Mesin dan Deep Learning. Tokyo Institute of Technology.
- Rasyidi, M. A. (2017). Prediksi Harga Bahan Pokok Nasional Jangka Pendek Menggunakan ARIMA. *Journal of Information Systems Engineering and Business Intelligence*, 3, 107–112.
- Sudaryono. (2015). Metodologi Riset di Bidang TI. Penerbit ANDI.
- Sugiyono. (2012). METODE PENELITIAN KUANTITATIF KUALITATIF DAN R&D (Cetakan Ke). Penerbit Alfabeta.