

# Prediction of Daily Gold Prices Using an Autoregressive Neural Network

Mohamad As'ad<sup>1</sup>, Sujito<sup>2</sup>, Sigit Setyowibowo<sup>3</sup>

<sup>1,2,3</sup> Information Technology Department, STMIK Pradnya Paramita Malang, Indonesia

<sup>1</sup> asad@stimata.ac.id (\*)

<sup>2,3</sup> [sujito, sigit]@stimata.ac.id

**Abstract**— Gold is a precious metal that functions as a gem and also an investment. Gold investment is the reason for many people because it is practical, not easily damaged, easy cashed, not taxable, and other purposes. Based on this, many people choose gold as an investment. The problem for people who will invest in gold is related to uncertain gold price predictions so that the accuracy of forecasting methods are needed. The purpose of this paper is to forecast accurately daily gold prices using the Neural Network Autoregressive (NNAR) method. Training Data to find out the value of accuracy in the NNAR method uses secondary data obtained from Yahoo Finance in the form of daily gold prices. Test results on the NNAR method produce a better and more accurate level using the NNAR (25,13) model with a MAPE value of 0.370707, a MASE of 0.5851083, and an RMSE of 6.939331. The conclusion of the results of this paper is the daily price of gold is influenced by the daily price of gold a day ago to 24 periods ago with the NNAR (25,13) model.

**Keywords**— prediction; forecasting; daily gold prices; yahoo finance dataset; artificial neural network; neural network *autoregressive*

## I. INTRODUCTION

Gold is a precious metal that functions as a jewelry and also an investment and even is used as a measure of conversion in making zakat maal (if the calculation of zakat is complicated). Gold as an investment function is now booming in various online stores (Bukalapak, Tokopedia, etc.) or financial institutions (banks, pawnshops, and so on). This gold investment is mostly made by the general public because it is easy to cash, tax-free, simple transactions, not easily damaged, and other reasons. The price of gold always increases every year, although sometimes it goes down at a certain time, in the end in a long time interval will go up.

They are seeing the phenomenon of the price of gold above, very interested in predicting an accurate gold price so as not to lose the investment. From these problems, it is necessary to predict gold prices when buying at low prices and when selling at high prices to get a profit. Many researchers write forecasting models that are easy to use for forecasting, but sometimes forecasting accuracy is not good. This research will use the Neural Network Autoregressive (NNAR) forecasting model, which is relatively new to the Neural Network (NN) model and can be used in open-source software programs.

Many previous researchers have predicted gold prices, including the analysis of forecasting gold prices using the automatic clustering and fuzzy logic relationship method with a mean absolute percentage error of 0.053 [1]. Forecasting using this model has advantages and disadvantages. The advantages of this method include not including classical assumptions as in the probabilistic model, and this model can detect changes in fluctuating data to be able to predict data accurately. Besides this, the shortcomings of this method include that this method can work only on small data sets so that it can work well for short-term forecasting, this method is not able to provide a picture of the influence of past data such as autoregressive models or seasonal models [1]. Other studies

on forecasting the price of Indonesian gold use the classic fuzzy time series method with a mean absolute percentage error of 0.01 [2]. This research also has advantages and disadvantages. The advantage of this classic fuzzy time series method is that it does not have classical assumptions such as probability models, can make forecasting with good accuracy, this method is also not able to provide a picture of the influence of past data such as autoregressive models or seasonal models, this method is not able to work to detect the amount of data for a very long period so that it is only able to predict short-term [2]. Antam gold forecasting system research uses double exponential smoothing with forecast accuracy of 87.34% [3]. This double exponential smoothing method looks less accurate. Other than, the other drawback is that this method is only for data characters that have trends. This method is also only for short-term forecasting. The advantage of this method is that it does not use classical assumptions about errors that occur, another advantage of this method is that it is simple in the process that is influenced by smoothing exponential and trend alone. Otherwise, it can be done manually without special programs [3]. Research of gold price forecasting systems using the single exponential smoothing method results in an optimum exponential smoothing coefficient (alpha or  $\alpha$ ) of 0.9 and mean square error of 2021008.22 [4], this method has advantages and disadvantages almost the same as the double exponential model smoothing, but this model is only for models without trends [4]. Gold price prediction using Feed Forward Neural Network (FFNN) with the Extreme Learning Machine (ELM) method produces a mean absolute percentage error value of 0.05499 [5]. The ELM method is a development of the FFNN method, which has limitations. It takes a bit of time because the determination of network parameters is determined by iteratively, whereas in the ELM method, the determination of network parameters is determined by random means. Besides this, the ELM method does not use classical assumptions like probabilistic models. This ELM method also has high

accuracy. Besides that, this method can also describe the results of forecasting now influenced by past period data [5].

The purpose of this paper is to use the NNAR method to predict the daily gold price. This research is expected to contribute to the development of the NN method combined with the statistical method because, in the process of identifying this method, there are also statistical methods such as the Partial Autocorrelation Function (PACF) calculation. This NNAR model was developed by Hyndman in 2018 on R open-source software [6]. The results of this study are expected to be able to predict accurate daily gold prices using a large dataset.

## II. RESEARCH METHODOLOGY

The forecasting method is a method that is designed based on the analysis of the relationship pattern between variables to be estimated with time variables, which are periodic series. Based on the nature of the data forecasting methods, there are two, namely qualitative and quantitative. There are two types of quantitative methods, namely time series models and causal models [7]. The time series method has experienced rapid development, not only probabilistic (statistical) models, but also non-probabilistic models such as NN. NN models for forecasting include ELM, Multilayer Perceptron's (MLP), NNAR, and others.

### A. Artificial Neural Network

Artificial Neural Network (ANN) is a method that was created to resemble neural networks in humans. This model is used to help human work in many fields, including image processing, some sensor tools in households, offices, hospitals, quantitative forecasting methods, and others. Broadly speaking, ANN consists of inputs, processes, and outputs [6].

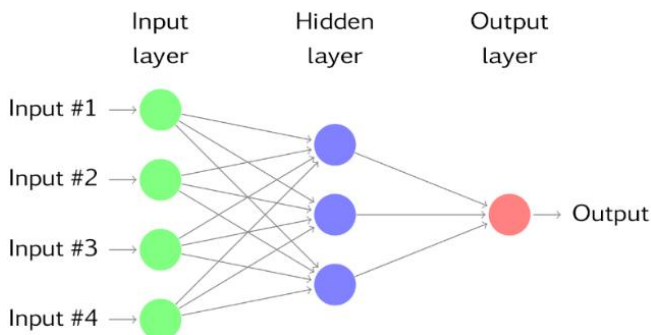


Figure 1. ANN with 4 input layers [6].

The multi hidden layer ANN model, between the input layer and output layer, is hidden by more than one hidden layer, so it is said to be a multi hidden layer ANN. In this study, we will use a hidden layer model, one hidden layer with three neurons and one output, as shown in Figure 1 [6].

### B. Model Backpropagation

The backpropagation model is an algorithmic model in artificial neural networks with supervised learning. This method is often used in the ANN algorithm for forecasting

models. Training or learning using backpropagation consists of three stages, namely feedforward (forward feed) from the input pattern, counting errors from the learning process, and adjusting the weights. This backpropagation model, the input of each node is a weighted linear combination. The weighted linear combination results are modified with nonlinear function to be the output of this ANN. This linear combination function can be written as [6]:

$$z_j = b_j + \sum_{i=1}^4 w_{ij}x_i \quad (1)$$

Where:

$z_j$  a variable is the sum function of the unit bias to  $j$  on the hidden layer,

$b_j$  a variable is a weight in the bias unit to  $j$ ,

$w_{i,j}$  a variable is the weight of the layer  $i$  bias to  $j$ ,

$x_i$  a variable is the network input to  $i$ .

Its activation function is binary sigmoid, which is a nonlinear function as in Equation (2) [6].

$$s(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

The nonlinear binary sigmoid function is part of the linear combination function in Equation (1). This binary sigmoid function is one of the functions for the backpropagation algorithm in a single layer network model [6].

### C. Neural Network Autoregression (NNAR)

The NNAR model is an ANN, where the input layer is just one variable input with the *lag 1*, *lag 2*, and so on models until *lag to p*, so it is called ANN Autoregressive (NNAR). NNAR was introduced by Hyndman and Athanasopoulos in 2018 with the application program R package program statistics in the "forecast" package with the net function. This model is only for feedforward networks in a single hidden-layer and is denoted by NNAR (p, k), where  $p$  denotes *lag-p* as input and  $k$  as nodes in hidden layer [6]. This the NNAR method uses a single hidden layer like in Figure 1 above and uses a nonlinear function as in Equation (1) to give weight and produce output from ANN. The activation function uses the binary sigmoid activation function as in Equation (2). This study uses the NNAR model (p, k) with the NNAR function in R package statistics [6].

### D. Partial Autocorrelation Function (PACF)

PACF or the so-called partial autocorrelation function is a function used to identify the order of autocorrelation on the *lag to p*. This PACF is written with Equation (3) [8]:

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j}} \quad (3)$$

### E. Performance Evaluation

To find out how accurate the forecasting is done, it can be evaluated by calculating the value of forecasting accuracy.

Forecasting accuracy used in this study is the mean absolute square error (MASE), mean absolute percentage error (MAPE), and root mean square error (RMSE). The calculation of the three accuracy uses Equation (4-6) [8].

$$MASE = \frac{\sum_{i=1}^n |e_{it}|^2}{n} \quad (4)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{e_{it}}{y_i} \right| \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n e_{it}^2}{n}} \quad (6)$$

### III. RESULT AND DISCUSSION

The time series plot for secondary data from Yahoo Finance is in the form of daily gold prices from 1 February 2018 to 30 April 2020, as shown in Figure 2 [9].

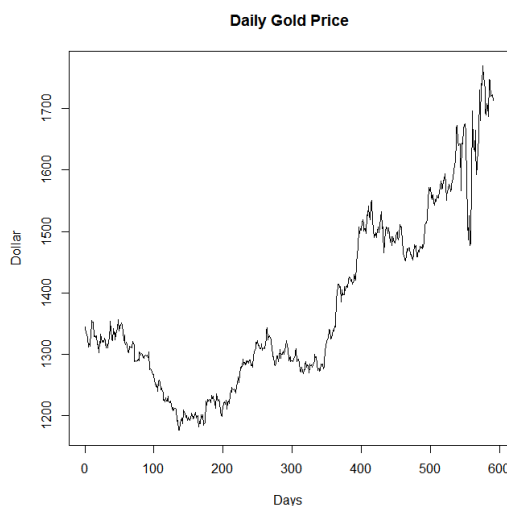


Figure 2. Time series plot of daily gold price data

Figure 2 shows that the daily gold price data from 1 February 2018 to 30 April 2020, there were 591 daily gold price data. Data contains trends and fluctuations that cannot be predicted in plain view, so NNAR forecasting methods are needed to make these predictions as accurately as possible.

#### A. Determination of Network Input

To determine the network input by doing the Partial Autocorrelation Plot (PACF) of the daily gold price data, does the data contain lag (autocorrelated) with the previous data [8], as the results in Figure 3.

Figure 3 shows the autoregressive model in lag 1 to lag 24, and this shows that there are spikes (long nails/lines extending) that are the Upper Conviction Band or the Lower Conviction Band to lag 24. Furthermore, as input is the daily gold price data at lag 1 to with lag 24. In the NNAR (p, k) model, the p variable shows the input of the autoregressive model that is lag 1 to lag 24. Whereas the number of neurons in a single hidden layer is the sum of the variable p plus one,

then divided by two, so we get the value k equal to  $(24 + 1) / 2$  is equal to 12.5 [6]. The proposed tentative (temporary) NNAR model is the NNAR (24,12) model. Other models that will be tested are the models around the proposed tentative model are NNAR (23,11), NNAR (23,12), NNAR (23,13), NNAR (24,11), NNAR (24,13), NNAR (25,11), NNAR (25,12), and NNAR (25,13).

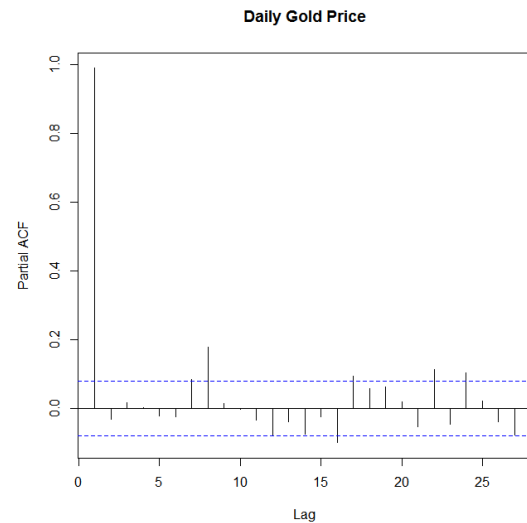


Figure 3. PACF plot of daily gold price data

#### A. Data Training

To make good of the ANN model must be done the training of part for data. There are several criteria observed to get a good training model based on the supervised backpropagation model, including a small error value [10]. Training data samples were taken randomly using 295 data from 591 daily gold price data. The best training model to be chosen is the training model that has the smallest RMSE value. Next will also be selected models that have the second and third smallest RMSE training values as a comparison for the selection of the best NNAR models based on RMSE, MAPE, and MASE in forecasting the daily gold price. The results of the calculations are presented in the Table I.

#### B. Data Testing

Testing data used (randomly taken) as many as 296 of 591 daily gold price data, this is done because the training data and testing data should be balanced in number because it is a sample taken to build the model and at the same time validate the model [11]. The criteria used to determine the model tested are the same as the criteria in the training model are the smallest error that is RMSE with the results of the model selection in the training model, and testing model is presented in table I. The smallest RMSE value in training is 0.1149614 namely the NNAR (25,13) model. This means that the best testing model is the NNAR (25,13) model. The smallest

RMSE value in testing is 0.1237697, so the best testing model is NNAR (25,13).

TABEL 1

RMSE VALUE IN TRAINING AND TESTING MODELS			
Model	Neuron	RMSE	
		Training	Testing
NNAR(23,11)	11	3.3172410	3.8155880
NNAR(23,12)	12	0.9722487	1.4322430
NNAR(23,13)	13	0.2412572	0.2976733
NNAR(24,11)	11	2.1641130	2.2804160
NNAR(24,12)	12	0.6062693	0.7154208
NNAR(24,13)	13	0.1836898	0.2416972
NNAR(25,11)	11	1.4170530	1.2386680
NNAR(25,12)	12	0.3830478	0.3792693
NNAR(25,13)	13	0.1149614	0.1237697

The two smallest RMSE training and testing values, the best forecasting model can be chosen, the NNAR model (25,13). The model that has the second smallest RMSE value is the NNAR (24,13) model with an RMSE training value of 0.1836898 and an RMSE testing value of 0.2416972. The model that has the third smallest RMSE value is the NNAR (23,13) model with an RMSE training value of 0.2412572, and an RMSE testing value of 0.2976733. Of the three selected NNAR models, namely NNAR (25,13), NNAR (24,13), and NNAR (23,13), which means that the number of neurons used is 13 neurons and network input has autocorrelation models  $x_1 \dots x_{25}$ ,  $x_1 \dots x_{24}$  and  $x_1 \dots x_{23}$ . Among the three models that will be used for forecasting models, namely the ANNAR model, which has the smallest RMSE, MAPE, and MASE, when used for forecasting.

C. Forecasting Models

Daily gold price predictions starting from 1 February 2018 - 30 April 2020 will be conducted on the three models with the smallest RMSE values, namely NNAR (23,13), NNAR (24,13), and NNAR (25,13). Forecasting results in Figure 6, Figure 7, and Figure 8 and the accuracy value will be displayed in Table II.

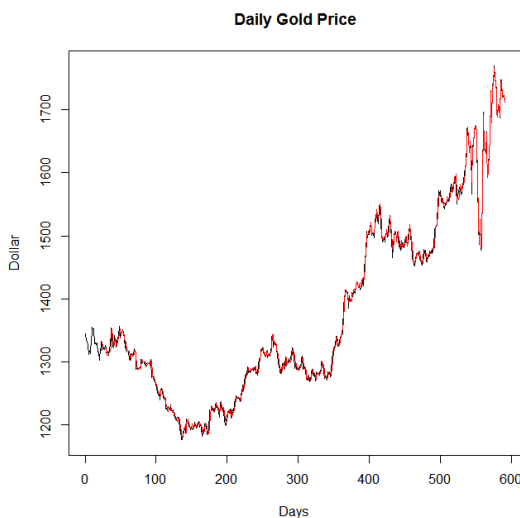


Figure 6. Model of NNAR(23,13)

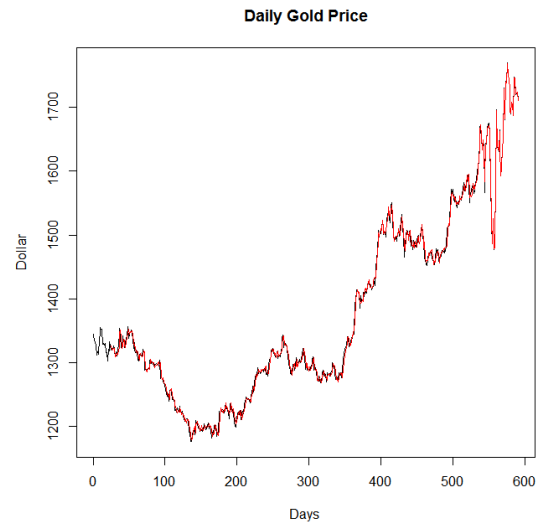


Figure 7. Model of NNAR(24,13)

Figure 6 for the NNAR model (23,13), where daily gold data is drawn as a black line, while the fitted or predicted results are drawn as a red line. The black line (daily gold data) starts at number one because this is the real data, while the fitted red line starts at number 24 because the NNAR model (23,13) has input  $x_1 \dots x_{23}$ . So the fitted value is at starting at number 24.

The results are fitted from Figure 7 for the NNAR model (24,13). The black line is the daily gold data, while the results fitted or predicted are drawn as red lines. The daily gold data (black line) starts at number one because this is the real data, while the fitted red line starts at number 25 because the NNAR model (24,13) has input  $x_1 \dots x_{24}$ . So the fitted value is at starting at number 25.

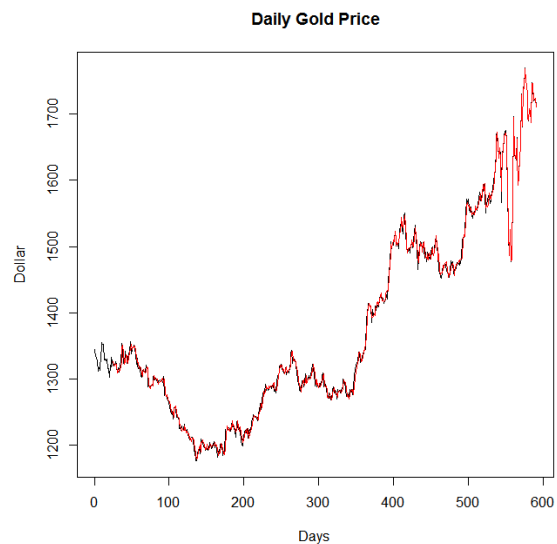


Figure 8. Model of NNAR(25,13)

Figure 8 can be interpreted for the NNAR model (25,13). The black line is the daily gold data, while the fitted or

predicted results are drawn as a red line. Daily gold price or black line starts from number one because this is the real data, while the fitted red line starts at number 26 because the NNAR model (25,13) has input  $x_1 \dots x_{25}$ . So the fitted value starts at number 26.

In figure 6-8, it appears that the daily gold data and its fitted data coincide once with the daily gold price data, which means it is difficult to distinguish, which is the most accurate. Next will be the selection of the best model based on performance evaluation. Performance evaluation is based on forecasting accuracy by selecting the smallest value for the three accuracy values, namely RMSE, MAPE, and MASE Table II. The value of the NNAR model (23,13), NNAR (24,13), and NNAR (25,13) with the smallest value in the NNAR model (25,13) is 0.370707; 0.5851083 and 6.939331, respectively. The NNAR model (25,13) will be recommended to predict the daily gold price because it has the highest value.

TABEL II  
 ACCURACY VALUE OF FORECASTING

Model	MAPE	MASE	RMSE
NNAR(23,13)	0.3874226	0.6142237	7.273566
NNAR(24,13)	0.373663	0.5889504	6.958934
NNAR(25,13)	0.370707	0.5851083	6.939331

#### D. Results of Forecasting

The results of forecasting the next five periods are visually shown in Figure 9. At the beginning of the period, the black lines appear as daily gold price data, then coincide with red as fitted or prediction from number 26 to number 591. Furthermore, from number 592 to number 596, appear red color alone as the forecast value of the next five periods. For more details, the forecast value of the next five periods, shown in table III.

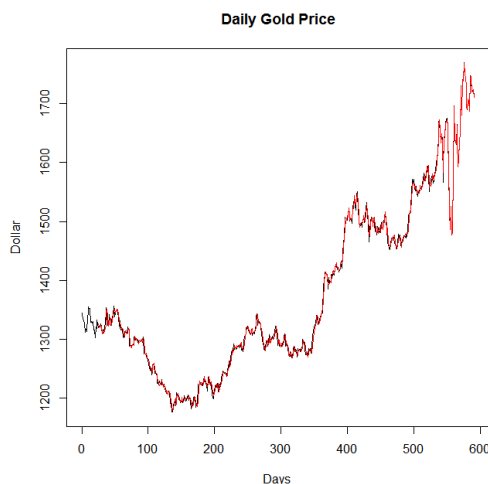


Figure 9. Plots of daily gold prices and forecasting five from the NNAR (25,13) model

Table III is the result of forecasting the next five periods from position number 591, namely 592 to 596. Numerically in the first period (1726.05) and the second (1751.24) predicted an increase in the daily gold price,

in the third period (1722.31) predicted a decline in daily gold prices and the fourth period (1771.67) and fifth (1824.99) an increase in daily gold prices.

TABEL III  
 FORECASTING OF DAILY GOLD PRICE FOR FIVE AHEAD

future period	1	2	3	4	5
forecasting	1726.05	1751.24	1722.31	1771.67	1824.99

#### IV. CONCLUSION

The NNAR method can predict the daily price of gold, as well. The NNAR model (25,13) has a network architecture, input data in the form of daily gold prices with lag-1 to lag-25. The NNAR model (25,13) has 13 neurons with a single hidden layer and uses a binary sigmoid activation function. Forecasting results have the MASE forecasting accuracy value is 0.5851083, MAPE is 0.370707, and RMSE is 6.939331. The future work on daily gold price predictions is to compare this NNAR method with other ANN time series methods such as Extreme Learning Machine (ELM) or Short-Term Memory Machine Learning models so that a better model will be obtained.

#### REFERENCE

- [1] S. Sitohang and A. M. Siringo, "Analisis Peramalan Harga Emas Dengan Metode Automatic Clustering And Fuzzy Logic Relationship", *Jurnal Information System Development*, vol. 3 no. 2, p. 104-115, 2018.
- [2] F. Aditya, D Devianto and Maiyastri, "Peramalan Harga Emas Indonesia Menggunakan Metode Fuzzy Time Series Klasik", *Jurnal Matematika UNAND*, Vol. 8 No. 2 p.45-52, 2019.
- [3] T. Andriyanto, "Sistem Peramalan Harga Emas Antam Menggunakan Double Exponential Smoothing", *Jurnal INTENSIF* vol. 1 no. 1, p.1-9, 2017.
- [4] F. M. Yuma, "Sistem Peramalan Harga Emas Menggunakan Metode Single Exponential Smoothing", *Seminar Nasional Royal (SENAR)*, hal. 299-302. 2018.
- [5] N. A. Izati, B. Warsito and T. Widiarih, "Prediksi Harga Emas Menggunakan Feed Forward Neural Network Dengan Metode Extreme Learning Machine", *Jurnal Gaussian*, vol. 8 no. 2, p.171-183, 2019.
- [6] R.J. Hyndman, and G. Athanasopoulos, *Forecasting: principles and practice, 2nd edition*, OTexts: Melbourne, Australia. 2018. [Online]. Available: OTexts.com/fpp2. Accessed March, 1st 2019.
- [7] S. Makridakis, S. C. Wheelwright, and R.J. Hyndman, *Forecasting: Methods and Applications*, John Wiley & Sons, Inc., 1998.
- [8] M. As'ad, "Finding the Best ARIMA Model to Forecast Daily Peak Electricity Demand". 5<sup>th</sup> Annual Applied Statistics Education and Research Collaboration (ASEARC) - Conference Papers 2012 page 39-42. 2012.
- [9] <https://finance.yahoo.com/quote/GC%3DF/history?p=GC%3DF>. Yahoo finance Accessed 1 May 2020.
- [10] D. Sena, & N.K. Nagwani, "A Neural Network Autoregression Model To Forecast Per Capita Disposable Income", *Rpn Journal Of Engineering And Applied Sciences*, Vol. 11, No. 22. 2016
- [11] L. Safitri, S. Mardiyati, H. Rahim, "Forecasting the mortality rates of Indonesian population by using neural network", *Journal of Physics: Conference Series*, IOP Publishing, 974-012030. 2018