Implementation multiple linear regression in neural network predict gold price

Musli Yanto¹, Sigit Sanjaya², Yulasmi³, Dodi Guswandi⁴, Syafri Arlis⁵
¹,⁴,⁵Department of Informatics Engineering, Faculty of Computer Science, Universitas Putra Indonesia YPTK, Indonesia
²,³Department of Management, Faculty of Economic and Business, Universitas Putra Indonesia YPTK, Indonesia

ABSTRACT

The movement of gold prices in the previous period was crucial for investors. However, fluctuations in gold price movements always occur. The problem in this study is how to apply multiple linear regression (MRL) in predicting artificial neural networks (ANN) of gold prices. MRL is a mathematical calculation technique used to measure the correlation between variables. The results of the MRL analysis ensure that the network pattern that is formed can provide precise and accurate prediction results. In addition, this study aims to develop a predictive pattern model that already exists. The results of the correlation test obtained by MRL provide a correlation of 62% so that the test results are said to have a significant effect on gold price movements. Then the prediction results generated using an ANN has a mean squared error (MSE) value of 0.004264%. The benefits obtained in this study provide an overview of the gold price prediction pattern model by conducting learning and approaches in testing the accuracy of the use of predictor variables.

This is an open access article under the CC BY-SA license.

1. INTRODUCTION

Gold is a commodity that is most in demand by investors for investing. Besides, gold has also been considered as a means of investment and can be used as the most appropriate instrument for people to invest [1]. The problem that often arises when investing is how investors precise in making decisions to determine when is the right time to invest in gold based on the price that occurs in the next future. To solve this problem, we need a system that can carry out a predictive analysis process to observe the gold price movements that will occur [2]. The results of this process can provide an overview of the gold price fluctuations that will occur and are expected to be used as a reference for investment interests [3].

The forecasting process has developed a lot among researchers, especially discussing the topic of the gold price. One approach that has been widely used is the artificial neural network (ANN). This approach is an interpretation of the working picture of the human biological system which can carry out systematic calculations based on experience. Based on the research that has been done in predicting gold prices using ANN, it can provide output with good accuracy, so that it can be taken into consideration as well as important indicators of the economic sectors of various companies [4]. From the results of other prediction processes, ANN can produce maximum results in making predictions with a fairly low percentage of error values [5], [6]. The same research states that ANN is faster and has a high degree of accuracy in making
predictions [7], [8]. So that ANN is inside [9], states that the model describes a systematic model in the learning process from input and output [10]. In the prediction case, the ANN model can be developed to produce optimal predictive results based on performance which is influenced by the amount of data used [11]. This indicates that ANN can be used to predict based on the use of experimental data to obtain fairly good predictive results [12].

Based on the existing explanations, the study in the discussion still only focuses on the predictive output without ever paying attention to the predictor variables used. Therefore, this research discussion will also discuss the gold price prediction process by implementing multiple linear regression (MRL) in ANN. In this case, the novelty of previous research is that the implementation of MRL can be expected to measure the level of correlation between the predictor variables and the predicted output. To prove the relationship between variables in the prediction results, MRL is considered the most feasible to observe this relationship. In previous studies, it has also been explained that MRL can show a relationship between predictors. MRL is also called one of the measuring tools to test the relationship between the dependent variable and the independent variable. Initially, it can be seen that MRL can be used in a prediction model that is carried out in the case of predicting future gold prices. Such a model is known as “forecast-1” and is considered a benchmark model capable of evaluating model performance [13].

Several other studies have also stated that MRL can carry out the process of analyzing factors that affect changes in the price of gold in the future [14]. In parallel research also carried out in the process of predicting electricity prices, stating that the MRL method can analyze variables that can affect the predicted output results [15]. For this reason, MRL can take a statistical approach by looking at the relationship of the variables used in influencing the output [16].

The MRL approach will be able to provide the right predictor variables based on the results of the correlation test so that the prediction process that will be carried out with ANN is expected to give much better results than the previous prediction analysis process. Therefore, this study aims to produce a more accurate prediction result that can be used as information on the estimated gold price that will occur in the next period. Another goal to be achieved from this research based on the implementation of MRL in the ANN prediction process is to be able to propose a better form of the predictive analysis model. This analysis model can be used to predict other cases to obtain predictive results with a good level of accuracy. Therefore, this model will be able to improve the predictive analysis process that has been carried out by ANN previously.

2. RESEARCH METHOD

The research methodology is crucial in describing the research activities carried out by researchers. The research methodology starts from building a framework in the process of solving problems. The shape of the framework in this study is in Figure 1.

![Figure 1. Research method](image-url)
Figure 1 is presented, a prediction model developed in this study. The following steps are taken:

i) The first stage is the stage for analyzing the data that will be used as a variable for the prediction process; ii) The second stage is forming a predictive network pattern; iii) The third stage will test the correlation between predictive variables and also test the correlation with the prediction target using MRL; iv) After the correlation test phase with MRL, the next step is to analyze the test results. If the test results give optimal results, the best predictive pattern model is found. If not then repeat the data analysis process at the beginning; v) The fifth stage is the initial stage in carrying out the prediction process using ANN. At this stage the data will be normalized so that it can be carried out in the training process and network testing; and vi) The final stage in this research methodology is to measure the output level of the network so that in the end it will find a good predictive result.

2.1. Data analysis

In predicting the price of gold, the predictor variables used include: oil price (X1) is a variable that affects gold price movements [17], [18]. The exchange rate against the Dollar also has an influences on the price of gold [19], [20], and the inflation rate as an indicator that plays a role in fluctuating gold prices [21], [22]. The output is presented in Table 1.

| Table 1. Predictor variable
| Oil Price (X1) | Exchange Rate (X2) | Inflation (X3) | Gold Price (X4) |
|---------------|-------------------|---------------|----------------|
| 29.78         | 13,915.00         | 4.14 %        | 36.28          |
| 31.03         | 13,462.00         | 4.42 %        | 39.62          |
| 37.34         | 13,342.00         | 4.45 %        | 39.39          |
| 40.75         | 13,270.00         | 3.60 %        | 40.60          |
| 45.94         | 13,683.00         | 3.33 %        | 38.97          |
| 47.69         | 13,246.00         | 3.45 %        | 41.30          |
| 44.13         | 13,159.00         | 3.21 %        | 42.67          |
| 44.88         | 13,367.00         | 2.79 %        | 42.48          |
| 45.04         | 13,063.00         | 3.07 %        | 42.06          |
| 49.29         | 13,116.00         | 3.31 %        | 41.07          |
| 45.26         | 13,631.00         | 3.58 %        | 37.97          |
| 52.62         | 13,503.00         | 3.02 %        | 37.18          |

2.2. Multiple regression linear (MRL)

After the predictive network pattern is formed, the discussion will continue with the process of testing the correlation for the predictor variables. This process uses multiple linear regression methods. This method is very feasible in analyzing the relationship between variables as a parameter to measure the relationship between the predictor (X) and the predicted results (Y). After the predictive network pattern is formed, the discussion will continue with the process of testing the correlation for the predictor variables. This process uses multiple linear regression methods. This method is very feasible in analyzing the relationship between variables as a parameter to measure the relationship between the predictor (X) and the predicted results (Y) [23], [24]. After the predictive network pattern is formed, the discussion will multiple linear regression (MLR) is also a very simple method for looking at the relationship between predictor variables and response variables [25]. In multiple regression, there are more than two or more independent variables and one dependent variable. The equation is in the image [26]:

\[ Y_i = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 \ldots \beta_pxpx \]  \hspace{1cm} (1)

To develop multiple linear regression equations The parameters were obtained from the training data and the variables were extracted from the dataset using correlation. The quantity r, called the linear correlation coefficient, measures the strength and direction of the relationship between two variables. The math formula for r is (2):

\[ R = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2)(\sum y^2) - (\sum x)^2(\sum y)^2}} \]  \hspace{1cm} (2)

2.3. Artificial neural network (ANN)

In the process of predicting the price of gold. First, the data is converted into intervals [0, 1]. The data to be converted includes the input and output layers in the network. The input and output layers that have been changed, will be included in the activation function. This function is very necessary because the activation function is used as a feed-forward for both layers. The activation function used is sigmoid [27]. If a logistic activation function is to be applied where xi is the input value (predictor variable), the input value is converted to (0, 1). The math formula for r is (3) [28]:

Implementation multiple linear regression in neural network predict gold price (Musli Yanto)
\[ X_i = \frac{X_i - \text{min}(x_i)}{\text{max}(x_i) - \text{min}(x_i)} \]  

### 2.3. Prediction network patterns

In this stage, the research discussion will continue to build a gold price prediction network pattern. The pattern network consists of input layer neurons, 1 neuron in the hidden layer, and 1 neuron in the output layer [29]. The network pattern that is formed is shown in Figure 2.

![Figure 2. Gold price prediction network pattern](image)

3. **RESULTS AND ANALYSIS**

#### 3.1. MRL analysis

Based on the above equation, the multiple linear regression experiment to observe the correlation results of the predictor variables can be done using the SPSS program in data processing, processing using the variable world oil price (X1), the rupiah exchange rate (X2), and the inflation rate (X3) as an independent variable, the Price of Gold (Y) as the dependent variable. Following the results of the correlation generated are in Table 2.

|          | Correlations                  |
|----------|------------------------------|
|          | X1_OP | X2_ER | X3_INF | Y_GP  |
| Pearson Correlation |  | .723** | -.580** | .222  |
| Sig. (2-tailed) | 1     | .000  | .002   | .158  |
| N         | 42    | 42    | 42     | 42    |
| Pearson Correlation | .723** | 1     | -.473** | -.168 |
| Sig. (2-tailed) | .000  | .002  | .287   |
| N         | 42    | 42    | 42     | 42    |
| Pearson Correlation | -.580** | -.473** | 1      | -.350* |
| Sig. (2-tailed) | .000  | .023  |
| N         | 42    | 42    | 42     | 42    |
| Pearson Correlation | .222  | -.168 | -.350* |
| Sig. (2-tailed) | .158  | .287  | .023   |
| N         | 42    | 42    | 42     | 42    |

**. Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

From Table 2, it can be explained the relationship between variables. The relationship between variables X1 and X2 has a strong relationship (0.723), the relationship between variables X1 and X3 has a strong relationship (0.580). The relationship between variables X1 and Y has a weak relationship (0.222). The relationship between variable X2 and X3 has a weak relationship (0.473). The relationship between X2 and Y has a weak relationship (0.222). The relationship between variable X3 and Y has a weak category (0.350). To discover relationship of influence between X1, X2, X3 on Y, the R Square test was performed. The R Square test results are shown in Table 3.

From the results of the regression process that has been done, the authors found the results of R square 37.8%. This is the result of the percentage generated from the regression process in the gold price prediction process using 3 predictor variables. The remaining 62.2% is influenced by other variables not examined. Then the t test is used to see the effect of the variables X1, X2, X3 on Y partially done and the results can be seen in Table 4.
Based on the above output, the results obtained for X1, the results of the t test show that the oil price variable has a significant effect on the price of gold. where the sig value is 0.013<0.05. For X2, the t test results show that the exchange rate variable has a significant effect on the price of gold. As in the table where the value of sig 0.013<0.000. For variable X3, the results of the t test show that the inflation variable has a significant effect on the price of gold. This can be seen from the sig value of 0.017<0.05. For the relationship between the price of oil with the price of gold. The results of this study are in line with previous studies conducted by [30] stating that world oil prices are one of the factors that influence the price of gold. For the relationship between the exchange rate and the price of gold, the results of this study are in line with research conducted by [31] which proves that the exchange rate affects the movement of gold prices. For the relationship between inflation and the price of gold. The results of this study are also in line with research conducted by [32] which proved that inflation can also influence gold price movements. So those variables that have been tested can be used as variables in the gold price prediction process.

3.2. Predict ANN

In conducting the prediction process, the ANN model used is backpropagation. The process begins with a transformation process. These results can be seen in Table 5. After the data is transformed, the prediction process is continued to conduct training and testing on the network architecture that has been built based on the predictor variables from the data analysis process. This process will produce the optimal weight in the form of performance and epoch values based on the previous network pattern [33], [34]. The training and testing results table can be seen in Table 6.

| Table 3. Multiplicative regression results          |
|-----------------------------------------------|
| Model Summary                                  |
| Model  R  R Square  Adj R Square  Std. Error of Estimate |
| 1    .615  .378  .329  1.39588                  |

| Predictors: (Constant), X3_INF, X2_ER, X1_OP |

| Table 4. T test result                  |
|----------------------------------------|
| Coefficients*                         |
| Model    Unstandardized Coefficients  Standardized  t  Sig.  |
|          B        Std. Error  Beta  |
| 1   (Constant)  72.922    7.742   -  9.419  .000           |
| X1_OP  .076   .029   .525  2.608  .013           |
| X2_ER  -.002  .001  -.734  -3.944  .000           |
| X3_INF  -1.388  .556  -.393  .017           |

Dependent Variable: Y_GP

| X1  | X2  | X3  | X4  | T  |
|-----|-----|-----|-----|----|
| 0.1000 | 0.1000 | 0.9000 | 0.183916 | 0.268504 |
| 0.17078 | 0.120626 | 0.831429 | 0.746154 | 0.150394 |
| 0.264222 | 0.224404 | 0.717143 | 0.323776 | 0.10000 |
| 0.325378 | 0.284162 | 0.557143 | 0.287413 | 0.227559 |
| 0.480908 | 0.241436 | 0.237143 | 0.301399 | 0.274803 |
| 0.520023 | 0.266483 | 0.591429 | 0.530769 | 0.566142 |
| 0.676484 | 0.166298 | 0.18000 | 0.90000 | 0.90000 |
| 0.590493 | 0.384346 | 0.10000 | 0.693007 | 0.753543 |
| 0.612534 | 0.420442 | 0.351429 | 0.726573 | 0.855906 |
| 0.755957 | 0.509576 | 0.362857 | 0.693007 | 0.745669 |
| 0.90000 | 0.564825 | 0.157143 | 0.51958 | 0.629134 |
| 0.854986 | 0.90000 | 0.031429 | 0.10000 | 0.41811 |

| Table 5. Result of convert data predict |
|----------------------------------------|

| Network Pattern | Error | MSE |
|-----------------|-------|-----|
| Pattern 4-4-1   | 0.0062| 0.006364 |
| Pattern 4-6-1   | 0.0041| 0.004264 |
| Pattern 4-8-1   | 0.005 | 0.004864 |
| Pattern 4-10-1  | 0.0193| 0.019264 |
| Pattern 4-12-1  | 0.0206| 0.020664 |

Table 6 shows that the best network pattern in the prediction process is in the 4-6-1 pattern with a MSE value of 0.004864. The pattern resulting from this process, in this case, needs to be reconsidered to ensure the...
The MSE value can be used as an approach in seeing the error value of the network output. Based on the prediction analysis process carried out using Matlab software, the prediction results show the accuracy in describing the gold price movements that will occur in the next period can be seen in the Figure 3 and a graph of the acquisition of the prediction results is shown in Table 7.

From the prediction process carried out, this research can provide good prediction results in predicting the gold price that will occur in the next period. Based on the prediction table, shows that the error rate is minimal based on the specified target. Another finding obtained in this study is that the proposed analysis model with the implementation of MRL in the prediction process using ANN can describe a better predictive analysis model and can improve the previous prediction process.

![Graph of the performance of the prediction output](image)

| Target | Prediction Result | Target | Prediction Result |
|--------|------------------|--------|------------------|
| 36.26  | 35.90            | 41.3   | 39.70            |
| 39.62  | 37.56            | 42.67  | 42.98            |
| 40.6   | 38.03            | 42.48  | 42.98            |
| 38.97  | 38.16            | 41.07  | 41.11            |

4. CONCLUSION

This study concludes that the prediction model proposed with the implementation of multiple linear regression methods in the gold price prediction process using an artificial neural network can present a precise and structured prediction process that provides a very high accuracy value and a fairly low error rate. The results obtained by the multiple linear regression method can prove the correlation between the predictor variables. The calculation results obtained are 62% of the correlation variable correlation (x1, x2, and x3) affects the results based on the results of the t-test which shows that each variable affects output (y). Meanwhile, the prediction process using artificial neural networks produces prediction results with an average MSE error value of 0.004264%.

REFERENCES

[1] K. Aden, “Prediction of Gold Prices Using Artificial Neural Networks,” Uluslararası Mühendis. Arastirma ve Geliştirme Derg., no. April, pp. 83-89, 2017, doi: 10.29137/umagd.350596.
[2] M. Hamdi and C. Aloui, “Forecasting crude oil price using artificial neural networks: A literature survey,” Econ. Bull., vol. 35, no. 2, pp. 1339-1359, 2015.
[3] W. A. Al-Dhuraihi and J. Ali, “Using classification techniques to predict gold price movement,” in 2018 4th International Conference on Computer and Technology Applications, ICCTA 2018, 2018, pp. 127-130, doi: 10.1109/CATA.2018.8398669.
[4] I. ul Sami and K. Nazir, “Predicting Future Gold Rates using Machine Learning Approach,” Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 12, 2017, doi: 10.14569/ijacsa.2017.081213.
[5] P. Zhang and B. Ci, “Deep belief network for gold price forecasting,” Resour. Policy, vol. 69, 2020, doi: 10.1016/j.resourpol.2020.101806.
[6] Y. Sari, E. S. Wijaya, A. R. Baskara, and R. S. D. Kasanda, “PSO optimization on backpropagation for fish catch production prediction,” TELKOMNIKA Telecommunication, Computing, Electronics and Control, vol. 18, no. 2, pp. 776-782, 2020, doi: 10.12928/TELKOMNIKA.V18I2.14826.
Implementation multiple linear regression in neural network predict gold price (Musli Yanto)
[35] X. Pang, Y. Zhou, P. Wang, W. Lin, and V. Chang, “Stock market prediction based on deep long short term memory neural network,” in COMPLEXIS 2018 - Proceedings of the 3rd International Conference on Complexity, Future Information Systems and Risk, vol. 2018-March, pp. 102-108, 2018, doi: 10.5220/0006749901020108.

[36] M. Hironaga and N. Shimano, “Evaluating the noise variance of an image acquisition system with various reconstruction matrices,” in Conference on Colour in Graphics, Imaging, and Vision, 2012, pp. 247-252.

BIOGRAPHIES OF AUTHORS

**Musli Yanto** born in Jakarta on July 7, 1989. His undergraduate study was completed in 2012 at Universitas Putra Indonesia YPTK. He completed his Masters degree at Putra Indonesia University, YPTK Padang. He currently serves as a lecture in the Informatics Engineering study program at the Universitas Putra Indonesia YPTK Padang. Teaching history that has been carried out starting from 2014 until now, such as object oriented programming, mobile programming, and computer graphics. Published research history places more emphasis on the artificial neural network in the process of identification and prediction.

**Sigit Sanjaya** born in Padang, July 7, 1989. Undergraduate program in Accounting major Padang state University completed in 2013. Master program in Management major in Padang state University completed in 2015. Currently, He is a lecturer in Faculty of Economic and Business in Universitas Putra Indonesia YPTK Padang. His research focus on Accounting, Business Management and Financial management.

**Yulasmi** born in Padang, August 19, 1974. Undergraduate program in Marketing Management major at Putra Indonesia University, Padang. Master Program in the Department of Marketing Management at Putra Indonesia University, Padang and Doctor of Marketing Management Science Doctorate at UPI YAI Jakarta. Currently working as a Lecturer and Chairperson of Magister Management Universitas Putra Indonesia YPTK Padang.

**Dodi Guswandi** born in Barung-BarungBelantai, Agustus 3,1986. Undergraduate program in Information System Universitas Putra Indonesia YPTK completed 2012. Master program in Computer Science also at Universitas Putra Indonesia YPTK. Currently, he is a lecturer in Faculty of Computer Science in Universitas Putra Indonesia YPTK. His research focus on decision support system and data mining.

**Syafri Arlis** born in Padang on October 23, 1986. His undergraduate study was completed in 2009 at Universitas Putra Indonesia YPTK. He completed his Masters degree at Putra Indonesia University, YPTK Padang. He currently serves as a lecture in the Informatics Engineering study program at the Universitas Putra Indonesia YPTK Padang. Teaching history that has been carried out starting from 2011 until now, such as database and digital image processing. Published research history places more emphasis on the artificial neural network and digital image processing.