JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 3, June 2023, hlm. 339 - 344 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i3.1748
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

GOLD PRICE FORECASTING USING MULTIPLE LINEAR REGRESSION METHOD

Raras Tyasnurita[1*], Rifqi Luthfiansyah[1], M Rayhan Brameswara[1]
[1]Departemen Sistem Informasi, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut Teknologi Sepuluh Nopember
Email: raras@is.its.ac.id

Abstract – Price forecasting is a part of economic decision making. Forecasting the daily rise and fall of gold prices can help investors decide when to buy or sell the commodity. The price of gold depends on many factors such as the price of other precious metals, the price of crude oil, the performance of the stock exchange, and the exchange rate of currencies. This study discusses gold price forecasting using the multiple linear regression method. The results of this study indicate that the best model is in the data distribution of 70%: 30% for training and testing, with a MAPE of 4.7% Based on these results, it can be concluded that the use of multiple linear regression method produces a fairly good model for gold prices forecasting. Besides, the correlation analysis show that the price of other precious metals greatly influences the price of gold where in this case the silver price whose correlation value is 0.87.

Keywords: forecasting, gold investment, multiple linear regression

Abstrak: Peramalan harga merupakan bagian dari pengambilan keputusan ekonomi. Melakukan peramalan terhadap kenaikan dan penurunan harga emas harian dapat membantu investor memutuskan kapan harus membeli atau menjual komoditas. Harga Emas bergantung pada banyak faktor seperti harga logam mulia lainnya, harga minyak mentah, kinerja bursa saham, dan nilai tukar mata uang. Penelitian ini membahas peramalan harga emas dengan menggunakan metode regresi linear ganda. Hasil dari penelitian ini menunjukkan bahwa model terbaik terdapat pada pembagian data pelatihan 70% dan pengujian 30%, dengan MAPE sebesar 4.7%. Berdasarkan hasil tersebut dapat diambil kesimpulan bahwa penggunaan metode regresi linear ganda menghasilkan model yang cukup baik untuk peramalan harga emas. Selain itu, analisis korelasi menunjukkan bahwa harga logam mulia lainnya sangat mempengaruhi harga emas dimana dalam hal ini variabel harga perak yang nilai korelasinya 0.87.

Kata kunci: peramalan; investasi emas, regresi linear ganda

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 3, June 2023, hlm. 339 - 344 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i3.1748
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

Forecasting is a method for estimating predictive information to determine future directions with reference to historical data. Forecasting is also an important data science task for many activities in an organization. For example, organizations in all industry sectors should engage in capacity planning to efficiently allocate scarce resources and goal setting to measure performance relative to baselines [7].

Gold is one of the most recognized precious metals in the world, many people use gold as an investment asset. Arguably no asset reflects the transformation of financial markets over the last few decades more accurately than gold [1]. The price and production behavior of gold is different from most other mineral commodities. In the 2008 financial crisis, the price of gold increased by 6% while the prices of many major minerals fell, and other equities fell by around 40%. The unique and diverse drivers of demand and supply of gold are not highly correlated with changes in other financial assets [6]. However, the price of Gold depends on many factors such as prices of other precious metals, crude oil prices, stock market performance, bond prices, currency exchange rates, etc.

Time series forecasting is working with one variable of dataset that is recorded on several period of times [2]. The gold price is fluctuated in a timely manner. Besides, there are several variables which affect the gold price. Therefore, this study applies Multiple Linear Regression to generate forecast. There are previous studies which show the implementation of Multiple Linear Regression to generate forecast for corn [11] and house price [10].

METHOD

This research method was carried out in several stages starting from literature study, data collection, dataset analysis, and model building. Each stage will be explained in detail as follows.

To find out the accuracy of the forecasting results, an evaluation is carried out using the method of calculating errors in forecasting. Mean Square Error (MSE) is the average value of the number of errors produced by an estimation model. The lower the MSE value indicates the forecasting model has good abilities. The general MSE formula can be written in equation (1).

.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (1)$$

Where N = amount of data, yi = actual index value in period I and y^i = forecasting data value in period i.

In addition to using RMSE as a method for calculating errors, researchers also use Mean Absolute Percentage Error (MAPE). MAPE is a measure of relative accuracy used to determine the percentage of forecast deviation. The MAPE formula in general can be written in equation (2).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100\% \quad (2)$$

Where n = the amount of data, At = the actual value of the index in the t period and Ft = the value of forecasting data in the t period.

Regression analysis is performed to determine the correlation between two or more variables that have a causal relationship, and to make topic predictions using relationships [9].

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 3, June 2023, hlm. 339 - 344 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i3.1748
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Multiple linear regression extends simple linear regression to include more than one explanatory variable. In both cases the term 'linear' is still used because of the assumption that the response variable is directly related to a linear combination of the explanatory variables. The equation for multiple linear regression has the same form as the equation for simple linear regression but has more terms [8] as shown in equation (3). For the simple case, β0 is the constant that will be the predicted value of y when all explanatory variables are 0. In models with explanatory variables, each explanatory variable has its own β_coefficient.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + e_i \quad (3)$$

To carry out an evaluation related to the regression model that was made, it is necessary to measure the value of R2. The R2 value is the proportion of variance in the dependent variable that can be predicted from the independent variables. It can be seen in equation (4). A low value will indicate a low level of correlation, meaning that the regression model is not valid, but not for all cases.

$$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}} \quad (4)$$

Where SSR is the sum of the squares of the residuals and SST is the sum of the squares.

Investment is one of the determining factors in the rate of economic growth of a country. Investment is the mobilization of resources to create or increase production capacity or income in the future [3]. The main purpose of investment is to replace the part of the capital supply that is not very good and add to the existing capital supply. Investment is also known as investment activities in the form of money or objects with the aim of obtaining profits for one period [5]. Investments have an element of uncertainty or risk so that investors cannot predict with certainty the results of the profits or losses that will be obtained from the investments made.

One of the popular investments is gold investment. Gold has a metallic form that is dense, soft, shiny and is believed to be the most flexible metal among other metals. Gold has several advantages, namely it does not change color easily, does not rust easily, does not fade even though it has been stored for a long time and attracts people to own it [4].

Data Collection

At this stage, a data collecting was carried out regarding the price of gold, where a dataset called Gold Price Data is obtained from Kaggle in the year of 2018. The dataset consists of 2290 rows and 6 columns which has 6 variables in it, namely the date or the date that the gold price was taken, SPX or the capitalization index of 500 companies, SLV which is the price of silver, USO which is the price of oil, EUR/USD or the exchange rate euro/usd and GLD, namely the price of gold, with the variable to be predicted is GLD, namely the price of gold.

Data Analysis

In the next stage, pre-processing is carried out starting with displaying the dataset into a table containing the raw data to be used which can be seen in Image 1.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 3, June 2023, hlm. 339 - 344 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i3.1748
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

	Date	SPX	GLD	USO	SLV	EUR/USD
0	1/2/2008	1447.160034	84.860001	78.470001	15.1800	1.471692
1	1/3/2008	1447.160034	85.570000	78.370003	15.2850	1.474491
2	1/4/2008	1411.630005	85.129997	77.309998	15.1670	1.475492
3	1/7/2008	1416.180054	84.769997	75.500000	15.0530	1.468299
4	1/8/2008	1390.189941	86.779999	76.059998	15.5900	1.557099
...	...	...	...	...	...	...
2285	5/8/2018	2671.919922	124.589996	14.060000	15.5100	1.186789
2286	5/9/2018	2697.790039	124.330002	14.370000	15.5300	1.184722
2287	5/10/2018	2723.070068	125.180000	14.410000	15.7400	1.191753
2288	5/14/2018	2730.129883	124.489998	14.380000	15.5600	1.193118
2289	5/16/2018	2725.780029	122.543800	14.405800	15.4542	1.182033

2290 rows × 6 columns

Image 1. Raw Data

From Image 1 it can be seen that the dataset consists of 2290 rows and 6 columns. The dependent variable used is GLD (Gold Price), while there are 4 independent variables, namely SPX, USO, SLV, EUR/USD. After seeing the variables from the dataset that will be used, then the gold price variable will be mapped onto the graph to see the graphical form of the variable.

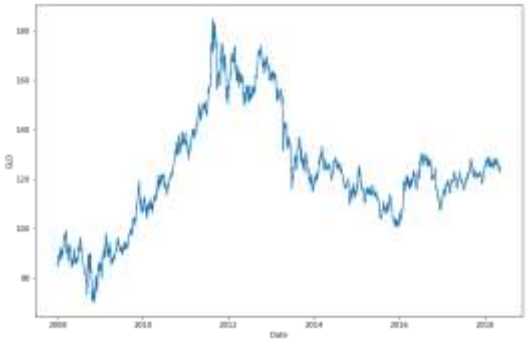

Image 2. GLD value

From Image 2, it can be seen that there is a trend in the dataset used. After knowing the form of the dataset, we will then find out the relationship between each variable in the dataset and the correlation map.

Image 3. Correlation Map

Image 3 is a correlation map made from the gold price dataset. This study only focused on the price of gold. This study obtained a relationship between the price of gold and the SPX variable worth 0.049, the relationship with the USO variable was worth -0.19, the relationship with the SLV variable was worth 0.87 and the relationship with other variables EUR/USD is worth -0.024. It is known that the price of other precious metals greatly influences the price of gold where in this case the SLV variable or silver price has a value close to 1.

Modelling

At this stage, the model will be made from the method used, namely multiple linear regression. Starting with creating the x and y variables, where the x variable consists of the independent variables, namely SPX, USO, SLV, EUR/USD and y consists of the dependent variable, namely GLD. Furthermore, dividing the train and test data, in this study 3 scenarios will be made, namely the first 80% train data: 20% test data, the second scenario is 70% train data: 30% test data, and the third scenario is 60% train data: 40% test data.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 3, June 2023, hlm. 339 - 344 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i3.1748
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

RESULTS AND DISCUSSION

After the multiple linear regression model has been made, predictions are made regarding the price of gold using the model and the prediction results for each scenario are obtained as shown in Image 4 for the 80:20 scenario, Image 5 for the 70:30 scenario, and Image 6 for the 60:40 scenario .

Image 4. Results of 80:20 train test

Image 5. Results of 70:30 train test

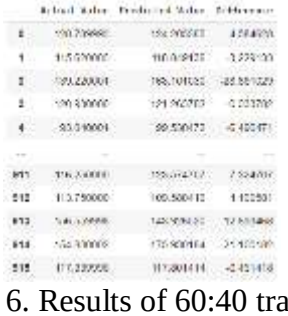

Image 6. Results of 60:40 train test

After getting the prediction results, then the prediction results will be mapped onto the chart to see a comparison of the original gold price chart with the predicted price in Image 7, Image 8, and Image 9, where the orange line is the original gold price data and the line that is the blue color is predictive data from the price of gold.

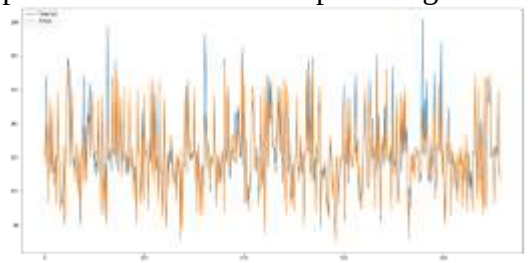

Image 7. Actual vs Prediction 80:20

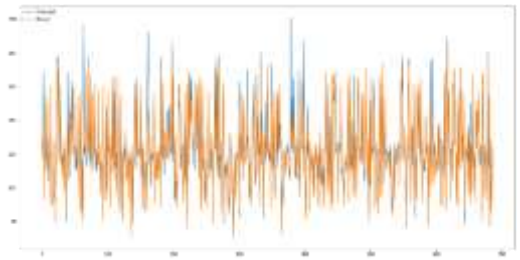

Image 8. Actual vs Prediction 70:30

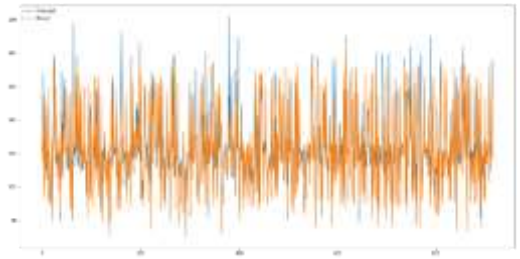

Image 9. Actual vs Prediction 60:40

After obtaining the predicted value of the gold price and seeing the graphical form of the gold price comparison with the predictions from each scenario, an evaluation of the model used using MSE, R^2, and MAPE is carried out which are shown in Table 1.

Tabel 1. Error Prediction Results

	Scenario		
	80:20	70:30	60:40
MSE	75.616	67.122	72.224
R2	0.85	0.87	0.86
MAPE	0.048	0.047	0.048

From Table 1, it can be seen that for the 80% data train scenario, the MSE value is 75.6176, the R2 value is 0.8530, and the MAPE value is 0.04898. For the data train 70%, the MSE value is 67.1227, the R2 value is 0.8743, and the MAPE value is 0.0475. For data train 60%, the MSE value is 72.2244, the R2 value is 0.8698, and the MAPE value is 0.0485.

CONCLUSION

From the results of the evaluation, between the 3 scenario trials, the best value was obtained in the 70% training and 30% testing distribution scenario with the smallest MAPE value of 4.8%. It is indicated that the percentage of errors is small enough that the gap is low beteen actual and prediction. Besides, the high R2 value (87%) indicated that the independent variable (silver price) highly affected the dependent variable value (gold price).

REFERENCES

[1] Beckmann, J., Berger, T., & Czudaj, R. (2017). Gold Price Dynamics and the Role of Uncertainty. econstor.

[2] Chatfield, C. (2000). TIME-SERIES FORECASTING. Florida: Chapman & Hall/CRC.

[3] R. V. Kawengian (2002). Analisis Pengaruh Investasi dan Tenaga Kerja dalam Sektor Pertanian dan Sektor Industri Guna Menentukan Strategi Pembangunan Irian Jaya. Institut Pertanian Bogor.

[4] S. Dipraja (2011). Siapa Bilang Investasi Emas Butuh Modal Gede? Proyek. Jakarta : Tangga Pustaka

[5] S. S. Husnan (2000). Studi Kelayakan Proyek. Yogyakarta : UPP AMP YKPN

[6] Shafiee, S., & Topal, E. (2010). An overview of global gold market and gold price forecasting. ELSEVIER, 178-189.

[7] Taylor, S. J., & Letham, B. (2017). Forecasting at Scale. PeerJ Preprints.

[8] Tranmer, M., Murphy, J., Elliot, M., & Pampaka, M. (2020). Multiple Linear Regression (2nd Edition). The Cathie Marsh Centre for Census and Survey Research (CCSR).

[9] Uyanik, G. K., & Guler, N. (2013). A study on multiple linear regression analysis. ELSEVIER, 234-240.

[10] Madhuri, C. R., Anuradha, G., & Pujitha, M. V. (2019). House price prediction using regression techniques: A comparative study. In *2019 International conference on smart structures and systems (ICSSS)* (pp. 1-5). IEEE.

[11] Ge, Y., & Wu, H. (2020). Prediction of corn price fluctuation based on multiple linear regression analysis model under big data. *Neural Computing and Applications, 32*, 16843-16855.

.