

COMPARISON OF SUPPORT VECTOR REGRESSION AND RANDOM FOREST REGRESSION ALGORITHMS ON GOLD PRICE PREDICTIONS

*Yennimar Yennimar¹, Samuel Valentino Hutagalung², Erikson Roni Rumapea³, Michael Justin Gesitera Hia⁴,
Terkelin Sembiring⁵, Dhanny Rukmana Manday⁶*

^{1,2,3,4,6,5}Indonesian Prima University

Jl. Sampul No. 1, Sei Putih Tengah, Kec. Medan Petisah, Medan City, North Sumatra 20118

Email: samuelvalentino294@gmail.com

ABSTRACT- This research was conducted to test how the Support Vector Regression and Random Forest Regression algorithms predict gold futures prices. The data used in this research was taken from the Investing.com website which will later be processed into a prediction model by comparing the SVR and RVR algorithms. The Support Vector Regression and Random Forest Regression algorithms will be tested to see the performance of each prediction model. The test results show that the Support Vector Regression model is superior in terms of accuracy with a value of 83%. However, the Random Forest Regression algorithm is superior with a smaller error rate, namely with an MSE value of 270.85 and an MAE value of 12.53.

Keyword: Comparison, Prediction, Support Vector Regression, Random Forest Regression.

1. INTRODUCTION

Gold is a type of precious metal that is widely used in the production of jewelry and is also a financial asset because it can be used as a store of value. Apart from having high aesthetic value, gold can also be a profitable commodity for people to invest in due to its extraordinary performance which is able to survive in the midst of the world economic crisis and also its resistance to inflation rates which is quite good.[1]

Gold is an asset that is usually used as a long-term investment whose value is stable, liquid and safe in real terms. The nature of gold which is resistant to changes in the rate of inflation, easy to cash in and there is also no tax imposed on gold itself, these are what make investors interested in investing.[2]

Gold as an investment tool has several types, namely; investing in gold jewelry, investing in gold bars, investing in gold pieces, investing in gold certificates and investing in gold in online trading. In this research, the data used is historical gold trading investment data based on futures contracts with the code XAU/USD with the unit scale used being the Troy Ounce. [3]

In investing in gold, investors will not be separated from guessing the rise or fall of gold prices so they don't lose in investing. Investors must be able to predict prices which are certainly always changing so that investors are right in carrying out buying and selling activities. Prediction is to minimize errors that may occur, so that the difference between estimates and actual events can be minimized. Thus, the author will conduct research on gold futures price predictions by comparing the Support Vector Regression and Random Forest Regression algorithms. [4]

To predict the price of gold futures, a good calculation is needed. Calculation of gold data can be done using the Support Vector Regression (SVR) algorithm. SVR itself is a development method of Support Vector Machine (SVM) which includes regression attributes which can

provide calculations in the form of predictions like a statistical depiction. This algorithm works by matching data and lines, but while maintaining margins and epsilon and this algorithm will provide an overview of a line diagram by comparing the actual data from the previous period with the calculated data. Basically, this algorithm can work on nonlinear data by paying attention to kernel tricks which can make this algorithm very good at overcoming overfitting in training data and testing data which will give the smallest possible error results with a maximum hyperplane.[5]

The Random Forest algorithm used for regression modeling of historical gold futures data is also called Random Forest Regression or RFR. Random Forest itself builds trees using bootstrap samples from different data and changes the way regression builds trees. However, in Random Forest Regression, this algorithm works by building many regression trees and then calculating the average value of the predicted results from all the regression trees. [6]

2. RESEARCH CONTENT

2.1 Types of Research

The type of research used here is through a qualitative approach, which is a type of research using data that exists in the past over a consecutive and continuous period of time, with the aim of applying knowledge for problem solving. In this case, the research focuses on comparing the accuracy of the Support Vector Regression and Random Forest Regression algorithms in predicting gold futures prices with a time period starting from June 1 2021 to June 30 2023.

2.2 Work Procedures

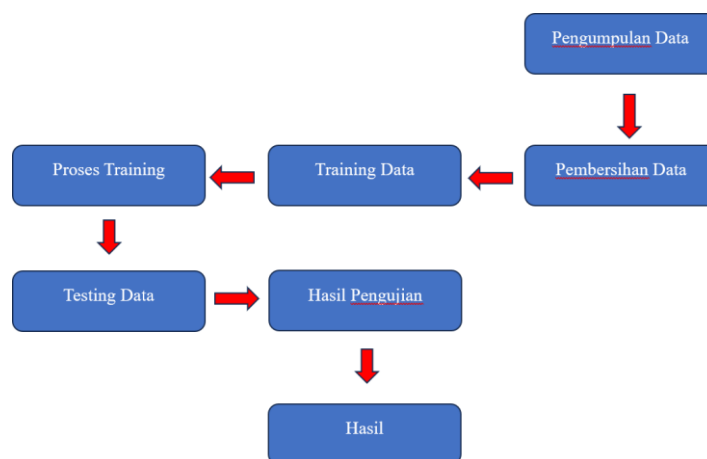


Figure 1. Research process

The stages of comparing the SVR algorithm and the RFR algorithm in predicting gold futures prices consist of several steps. The following are these stages:

1. Data Collection: The initial stage is to collect gold futures price data. In this gold price dataset there are 546 data records with 7 attributes, namely: Date, Last, Opening, Highest, Lowest, Vol., Change (%). This data must be in the appropriate format and ready for processing.

2. Pre-Processing: Data that has been collected requires additional processing, such as cleaning the data from empty values or invalid data.
3. Training Data: Data that has been normalized will be continued for training and the data that will be used for training data is from the "Last" column which is the last price of gold price trading in one day, and the data will be trained in the form of numpy data.
4. Training Process: Data that has been trained will be stored in a variable and will continue to divide the data by 70% for training data and 30% for testing data.
5. Testing Data: After the data training stage, the process will continue into the data testing process, then the data will be made into a prediction model.
6. Error Value Evaluation: After the data enters the data testing process, the data will then be evaluated for the predicted value with the actual value, by determining the MSE value and MAE value.
7. Prediction Results: At this stage the results of the data testing process will be visualized in the form of a graph of the results of the prediction model.

2.3 Data

The data that will be used for this research is historical gold futures data downloaded from the Investing.com website, with a troy ounce unit scale and in USD. Historical gold data was taken from the time period 1 June 2021 to 30 June 2023. In this gold price dataset there are 546 data records with 7 attributes, namely: Date, Last, Opening, Highest, Lowest, Vol., Change (%).

2.4 Method

Support Vector Regression

Support Vector Regression (SVR) is an algorithm that matches data and lines, but while maintaining margins and epsilon. Basically, this algorithm can work on non-linear data by paying attention to kernel tricks which can make this algorithm very good at overcoming overfitting in training data and testing data which will give the smallest possible error results with a maximum hyperplane.

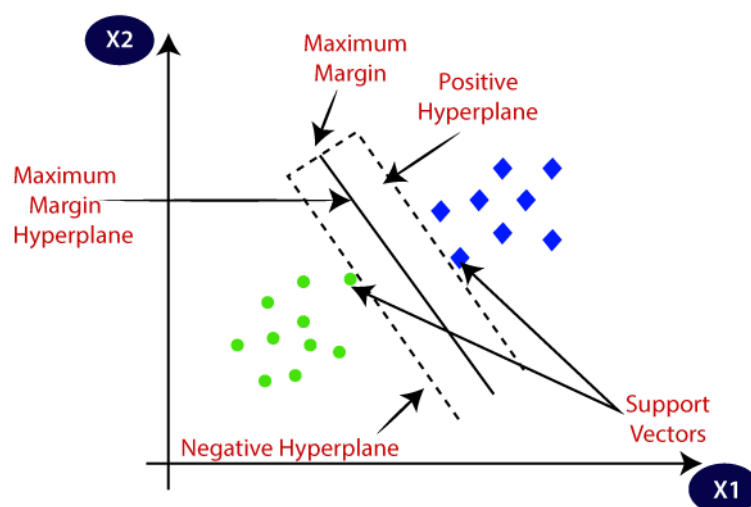


Figure 2. Support Vector Regression

Random Forest Regression

Random Forest Regression (RFR) is a machine learning algorithm whose testing process uses a supervised concept in building classifier classes. This algorithm combines predictions based on Multiple Decision Trees.

3. RESULTS AND DISCUSSION

3.1 Research Dataset

The data that will be used for this research is historical gold futures data downloaded from the Investing.com website. Investing.com is a website that provides complete information about Foreign Exchange, Indices & Shares at the url address: <https://id.investing.com/commodities/gold-historical-data>. This data was taken over a daily data period from 1 June 2021 to 30 June 2023. In this gold price dataset there are 546 data records with 7 attributes, namely: Date, Last, Opening, Highest, Lowest, Vol., Change (%).

The data was initially in CSV (Comma Spread Values) format, which is a data format commonly used to store datasets. However, this data format is not suitable for the data analysis process. So the historical data on gold futures prices which is still in CSV form is entered into data format in the form of Pandas Dataframe for processing internally Jupyter Notebook program based on the Python programming language. With the Pandas Dataframe format, historical data on gold futures prices can be visualized easily and can be further processed using an algorithm that we will test for levels. its accuracy.

3.2 Testing the Prediction Model

In the process of testing the prediction model, it includes split data, and Support Vector Regression and Random Forest Regression modeling. The limitations of this research are processing gold futures price data by testing two methods, Support Vector Regression and Random Forest Regression to obtain prediction model results with the highest accuracy and evaluating the resulting error values.

3.2.1 Testing the Support Vector Regression Prediction Model with the Kernel Radial Basis Function

At the testing stage with the Support Vector Regression algorithm, the gold price data that has been input will be processed data with a data division of 70: 30, with 70% for training data (378 data records from the "Last" attribute) and 30% for testing data (163 data records from the "Last" attribute). The "Last" attribute is the target column in testing the Support Vector Regression algorithm using the Radial Basis Function (RBF) kernel. As for testing this algorithm, the parameters set in the RBF kernel are, parameter value $C = 1000.0$ and parameter value $\gamma = 0.0001$, and get a prediction accuracy value of 0.83. In testing with the Support Vector Regression algorithm, it is evaluated based on the resulting error value, where the Mean Squared Error (MSE) value is 1214.

```
In [195]: from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.3)

In [196]: from sklearn.svm import SVR
svr_rbf = SVR(kernel='rbf', C=1e3, gamma=0.0001)
svr_rbf.fit(x_train, y_train)

Out[196]: SVR
SVR(C=1000.0, gamma=0.0001)

In [197]: svr_rbf_confidence = svr_rbf.score(x_test, y_test)
print('svr_rbf accuracy:', svr_rbf_confidence)

svr_rbf accuracy: 0.8324454274950489
```

Figure 1.Support Vector Regression algorithm testing

The prediction visualization results from testing the Support Vector Regression algorithm on the gold futures price dataset are:

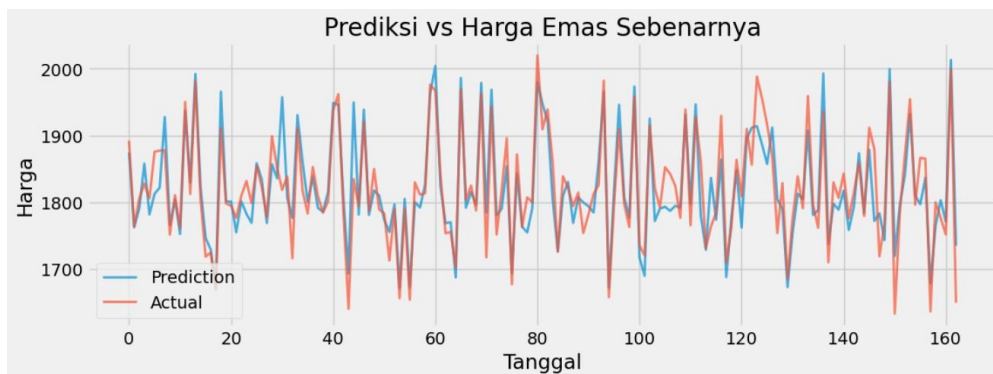


Figure 2.Prediction visualization with the Support Vector Regression algorithm

3.2.2 Testing the Random Forest Regression Prediction Model

In testing with the Random Forest Regression algorithm, gold price data will be divided 70: 30, 70% for training data (378 data records from the "Last" attribute) and 30% for testing data (163 data records from the "Last" attribute). ”).

As for testing this algorithm, with the RandomForestRegressor regression model, it gets a predicted accuracy value of 0.76. Testing with the Random Forest Regression algorithm will be evaluated based on the resulting error value, where the Mean Squared Error (MSE) value is 270.85 and the Mean Absolute Error (MAE) value is 12.53.

```
In [67]: from sklearn.model_selection import train_test_split
X_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
x_test.shape

In [68]: from sklearn.ensemble import RandomForestRegressor
model = RandomForestRegressor()

In [69]: LL_rf = model.fit(X_train, y_train)

In [70]: ytrain_pred = LL_rf.predict(X_train)

In [71]: print('The MAE is : %.2f%' mean_absolute_error(y_train, ytrain_pred))
The MAE is : 12.53

In [72]: print('The MAE is : %.2f%' mean_squared_error(y_train, ytrain_pred))
The MAE is : 270.85

In [73]: model_confidence = model.score(x_test, y_test)
print('rfr accuracy: ', model_confidence)
rfr accuracy: 0.768406052247757
```

Figure 3. Random Forest Regression algorithm testing

The prediction visualization results from testing the Random Forest Regression algorithm on the gold futures price dataset, namely:

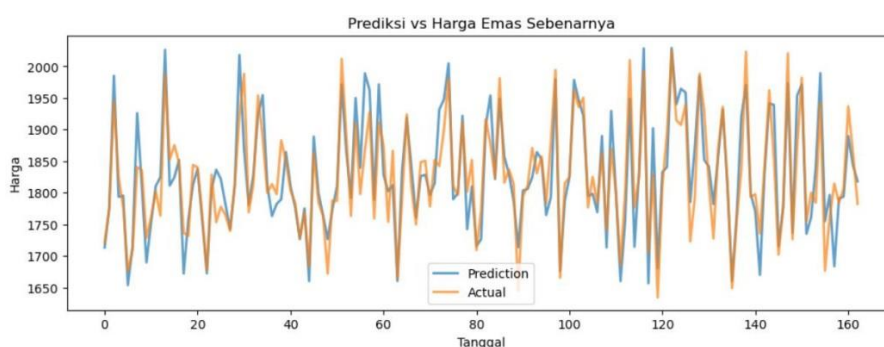


Figure 4. Prediction visualization with the Random Forest Regression algorithm

4. CONCLUSION

4.1 Conclusion

Based on the research that has been done to predict gold futures prices, the following conclusions can be drawn:

1. In the test results using the Support Vector Regression algorithm, the prediction accuracy value obtained was 0.83, with an MSE value obtained of 1214.32 and a Mean Absolute Error (MAE) value of 25.93. Then the test results use the Random Forest Regression algorithm for accuracy values predictionThe result obtained was 0.76, with an MSE value obtained of 270.85 and a Mean Absolute Error (MAE) value of 12.53.
2. This research shows that the comparison that has been carried out between the Support Vector Regression and Random Forest Regression algorithms on the gold futures price dataset, obtained the highest accuracy value obtained by the Support Vector Regression (SVR) algorithm of 0.83. However, the Random Forest Regression (RFR) algorithm has a lower error rate than the Support Vector Regression (SVR) with Mean Squared valuesThe error (MSE) is 270.85 and the Mean Absolute Error (MAE) value is 12.53.

4.1 Suggestions

Based on the results of this study, the authors suggest the following:

1. Do additional research by using the SVR algorithm, by trying other kernels such as polynomial and linear kernels to find out whether with these kernels the level of accuracy values will increase.
2. This research can be developed using other algorithms, such as Linear Regression or LSTM, which are not closed. It is possible that this algorithm can produce a smaller error rate compared to the algorithm that has been tested in this research.

BIBLIOGRAPHY

- [1] Azzahra. M., BD Setiawan, and PP Adikara, "Optimization of Support Vector Regression Parameters Using Genetic Algorithms for Gold Price Prediction," *Journal of Information Technology and Computer Science Development*, vol. 2, no. 1, pp. 273-281, 2018.
- [2] Wati. LRS, I. Cholissodin and PP Adikara, "Implementation of the Extreme Learning Machine (ELM) Algorithm for Predicting Gold Prices for Investors," *Journal of Information Technology and Computer Science Development*, vol. 3, no. 3, pp. 2408 - 2415, 2019.
- [3] Chrysanthemum Y., and NA Rizal "Effect of Bond Yield Prices, Dollar Index, Level" United States Inflation and Interest Rates on Gold Futures Contract Prices in London Gold Loco Investments during the Covid-19 Pandemic," *Journal of Management & Business*, vol. 6, no. 2, pp. 329-338, 2023.
- [4] Halimi. I., Y. Azhar and GI Marthasari, "Gold Price Prediction Using Univariate Convolutional Neural Network," *National Faculty Student Seminar Information Technology (SENAFTI)*, vol. 1, no. 2, pp. 105 - 116, 2019.
- [5] Pradana. FNSPM, and FS Papilaya, "Gold Price Prediction Analysis With "Probability of Recession Using the SVR Method," *Journal of Science and Information Technology (SINTECH)*, vol. 6, no. 1, pp. 37-46, 2023.
- [6] Fitri. E., and D. Riana, "Comparative Analysis of Prediction Models in Stock Price Prediction Using Linear Regression, Random Forest Regression Forest Methods and Multilayers Preception," *Journal of Informatics & Computerization Management Accounting*, vol. 6, no. 1, pp. 69 - 78, 2022.
- [7] Adriansyah. I., MD Mahendra, E. Rasywir, and Y. Pratama, "Comparison of Methods Random Forest Classifier and SVM On Classification of Students' Distance Learning Adaptation Level Ability." *Journal Bulletin of Informatics and Data Science*, vol. 1, no. 2, pp. 98 – 103, 2022.
- [8] Ristiano. EF, Nurmalasari, and A. Yoraeni, "Implementation of the Naive Bayes Method For Gold Price Prediction," *Journal of Computer Science (CO-SCIENCE)*, vol. 1, no. 1, pp. 62 - 71, 2021.
- [9] Agusmawati. NK, F. Khoiriyah and A. Tholib, "Gold Price Prediction Using LSTM and GRU Methods," *Journal of Informatics and Applied Electrical Engineering (JITET)*, vol. 11, no. 3, pp. 620-627, 2023.
- [10] Rakhmawati. D., and M. urhalim, "Prediction of Gold Futures Prices during the Covid-19 Pandemic Using Deterministic Trend Models," *ACCOUNTABLE*, vol. 18, no. 1, pp. 146–152, 2019.
- [11] thunder. M., J. Santony, and Yuhandri, "Gold Price Prediction Using Naïve Bayes Method in Investment to Minimize Risk," *Journal of Information Systems Engineering and Technology (RESTI)i*, vol. 2, no. 1, pp. 354–360, 2018.
- [12] Aulia. A., B. Aprianti, Y. Supriyanto., and C. Rozikin "Gold Price Prediction Using the Support

- Vector Regression (Svr) and Linear Regression (LR) Algorithms," Wahana Pendidikan Scientific Journal, vol. 8, no. 5, pp. 84–88, 2022.
- [13] Fitri. E, "Comparative Analysis of Linear Regression Methods, Random Forest Regression and Gradient Boosted Trees Regression Methodfor House Price Prediction," Kaputama Informatics Engineering Journal (JTIK), vol. 4, no. 2, pp. 208–213, Nov. 2020.
- [14] Givari. MR, MR Sulaeman, and Y. Umaidah. "Comparison of SVM, Random Forest and XGBoost Algorithms for Determining Approval of Credit Applications," Journal of Nuansa Informatics, vol. 16, no. 1, pp. 141–149, 2022.