Gold Price Prediction Using Machine Learning Techniques

Nandini Tripurana[1], Binodini Kar[2], Sujata Chakravarty[3], Bijay K. Paikaray[4] and Suneeta Satpathy[5]

[1,2] *Centurion University of Technology and Management, Paralakhemundi, Odisha*
[3] *Dept. of CSE, Centurion University of Technology and Management, Odisha, India.*
[4] *School of Information & Communication Technology, Medhavi Skills University, Sikkim, India*
[5] *Faculty. of Management Studies, Sri Sri University, Odisha, India*

Abstract

Significantly, gold is among the precious metals that are utilized to finance trading purchases. In countries like India, China, Indonesia, and many more, gold is considered the ideal jewelry, in addition, gold is also served as a present/remembrance and even gold accessories are presented as marriage settlements. Moreover, the countries with large gold reserves are considered a booming nations. At present time, precious metal like gold, is not just considered ornaments or jewelry but are considered as an investment and are kept with all nations' central bank as an assurance for re-payment of non-native loans, and also to manage inflation. Due to the increasing demand and dispense of this asset in the market the state of the major economies throughout the globe has a considerable impact on gold prices. Due to the change in gold prices, more investors are now considering gold investments. But irregularity in the gold price in the market makes it riskier for the investor. Thus, the goal of "Gold price prediction" is to forecast gold's price using a variety of Machine learning techniques, considering the relationship between several economic factors that influence gold rates. The machine learning algorithms that have been used are Random Forest regression, Decision tree (DT), Support vector regression (SVR), linear regression, and Artificial neural network (ANN).

Keywords

Machine Learning, Stock Price, Prediction, American Airlines, Support vector machine (SVR), Artificial neural network (ANN), Random Forest (RF), Decision tree (DT)

1. Introduction

For generations, gold has been a significant precious metal. It is one of the significant financial assets for nations and an important part of the world's monetary reserves for trade and currency hedging [1]. The inertness, undersupply, and difficulty in extraction of this metal gave itself a great value in the market. But shortage and stability are not the only reason for the increasing demand and supply of this valuable metal. Another quality that is making it more precious is the attention of investors towards it. It is one of the major assets being considered as an investment approach by many investors. Moreover, gold is reviewed to be "the final instance's assets" i.e., is the asset investors rely on [2].

ACI'22: Workshop on Advances in Computation Intelligence, its Concepts & Applications at ISIC 2022, May 17-19, Savannah, United States
EMAIL: nandinitripurana@gmail.com (A. 1); binodinikar28@gmail.com (A. 2); chakravartys69@gmail.com (A. 3); bijaypaikaray87@gmail.com (A. 4); suneeta.s@srisriuniversity.edu.in (A. 6)
ORCID: 0000-0001-5843-0335 (A. 4)


Investors rely on gold prices to check oscillation in other markets. But the unstable rise and fall in gold rates makes it riskier and develops a fear for the investors to invest in the gold market. Therefore, it is important to predict gold rates carefully. Thus, this paper aims on examining the interconnection between certain economic market variables and the price of gold, they are the S&P500 index, silver price, crude oil price, and US dollar exchange value (USD). Additionally, this research employs five Machine learning techniques, such as Support vector regression (SVR). Decision Tree (DT), Random Forest, linear regression, and Artificial neural network (ANN), and compares them to get the best fit model.

2. Literature survey

The journal [1] “Gold price prediction” studied about machine learning system that can predict gold prices based on several other economic variables. By using this they compared the correlation between economic variables and gold price. To predict gold prices, they used several machine learning algorithms. They used linear regression, SVM, and Decision Tree to train the model for predicting gold rates.

According to the second journal [2] “Future gold prices can be predicted using machine learning techniques” one of the most significant metals in the world is gold. Many countries maintain their gold reserves to be recognized as healthy and progressive countries, so, based on this they predicted the gold prices, so investors can invest in this commodity by analyzing the proposed model and can get huge benefits. In her article [3] "Modeling and Forecasting of Gold Prices on Financial Markets," V.K.F.B. Rebecca Davis makes use of the Autoregressive Moving Average (ARMA) model, a statistical tool that is often used to analyze time series data. The monthly prices of gold during a ten-year period are the data under consideration. The accuracy of the model was 66.67%. In their study titled [4] "Predicting Future Gold Rates using Machine Learning Approach," Iftikhar ul Sami and Khurum Nazir Junejo employ Artificial Neural Networks (ANN) to forecast gold prices. Over the course of eleven years, the data for this study were gathered from diverse sources. This information includes factors like the price of crude oil, the S&P 500 index, USD exchange rates, and other economic factors. In their article [5] "Prediction of the gold price with ARIMA and SVM," D Makala and Z Li use information gathered from the World Gold Council that includes daily gold prices from January 1979 to December 2019. To predict the price of gold, this study uses the Autoregressive Integrated Moving Average (ARIMA) approach with SVM. The ARIMA model's accuracy is lower than the SVM model's accuracy.

3. Data

The dataset, which includes ten years’ worth of data, is taken from the Kaggle website, dated from January 2008 to May 2018. It has six characteristics, such as gold rates, given as date, silver price, stock profit exchange, gold price, US dollar rate, and United States oil ETF. This dataset consists of 2290 records.

3.1 Data pre-processing

The dataset used should be cleaned at the very first step, in order to remove the noise, the null values, void values, etc.… to avoid the errors at the end. In order to do this firstly, the data should be normalized, next Feature extraction is done, where the attributes are identified as dependent and independent, and the unnecessary attributes should be removed, if any. In this paper ‘Date’ variable is dropped, as it is not important for the calculations.

3.2 Correlation analysis

Correlation is used to check the relationship among all the different variables that we have. For this paper, the correlation analysis has been shown in figure 1.

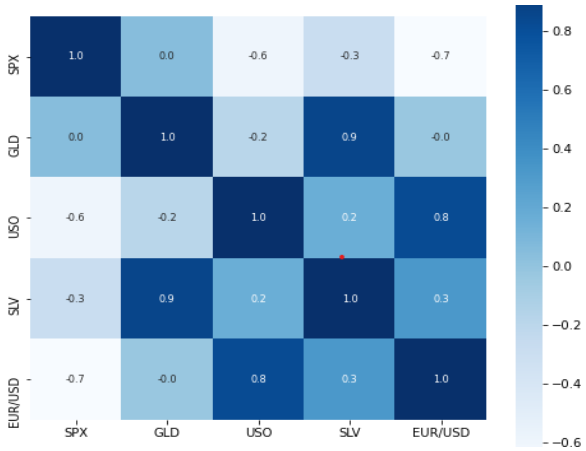

Figure 1: Correlation Analysis of different attributes.

4. Methodology

Machine learning algorithms used: For developing this model, we have used the algorithms listed as, Random Forest, DT, SVR, linear regression, and ANN.

4.1 Random Forest Regressor

Random forest regression is based on the concept of ensemble learning, let's go into more detail about ensemble learning. This technique uses the outputs of several machine learning algorithms or similar algorithms repeated many times. Ensemble learning techniques can be utilized for that reason because an exclusive lone model might not produce results that are as exact [6]. Similarly, it will use several decision trees trained together to get a final combined prediction. This supervised learning technique known as Random Forest uses ensemble learning to carry out both the regression and classification tasks. It creates

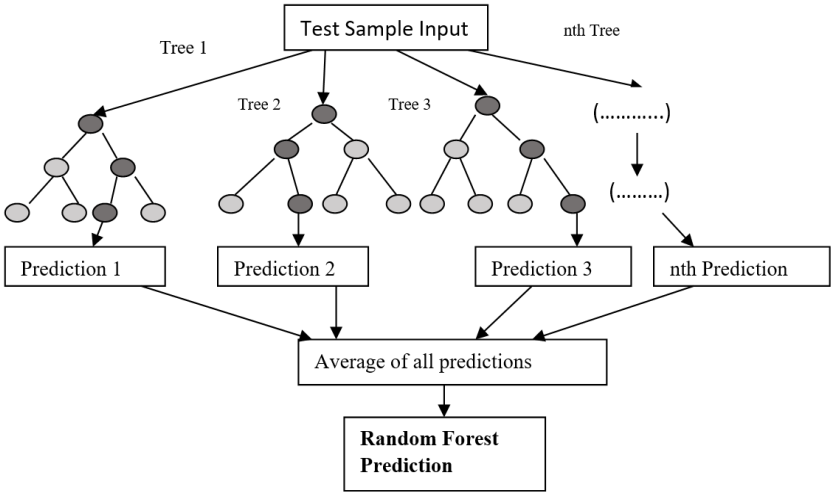

Figure 2: Random Forest regressor

several decision trees during the training phase and integrates the outputs from each to determine the final output, which might take the form of classification or regression [7].

4.2 Decision Tree

The decision tree organized a series of rules in the tree structure. A set of rules were arranged in a tree structure via a decision tree. It is one of the most useful techniques for supervised non-parametric learning. The decision tree has a few key terminologies. The Root node, which is at the top of the tree, depicts the overview of the entire community as it splits into sub-nodes. Decision nodes are referred to when sub-nodes can be split into further nodes. The node of a tree with no additional split nodes is called a leaf node. Pruning is the act of removing a decision node's sub-node [8].

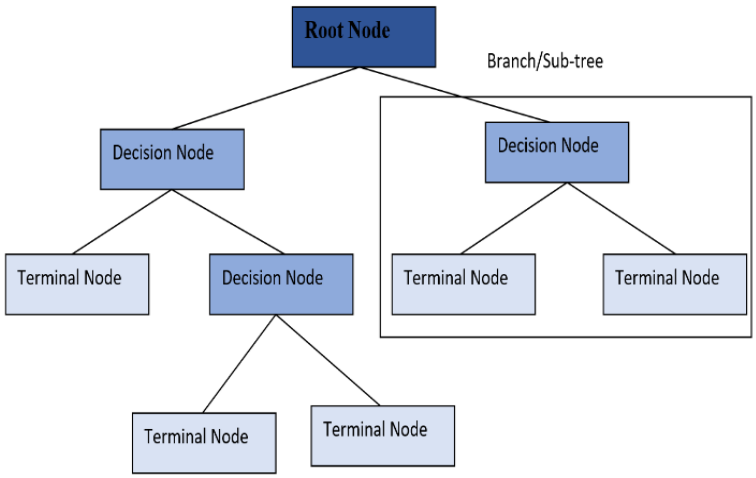

Figure 3: Decision Tree

4.3 Support Vector Regressor

Support vector regression is a part of a support vector machine, and this concept is applied for both regression as well as classification of data. Support vector machine is one of the fast and non-linear supervised machine learning models, given a set of labeled training data. SVM will help to find the optimum hyperplane which categorized new examples in one-dimensional space. The hyperplane is a point in two-dimensional space.

The hyperplane is a land in three-dimensional space, it is a surface that divides into two parts, where each class lies on either side. In two-dimensional linear separable cases, the data are separated by a line. In this case, the data is linearly separable [9].

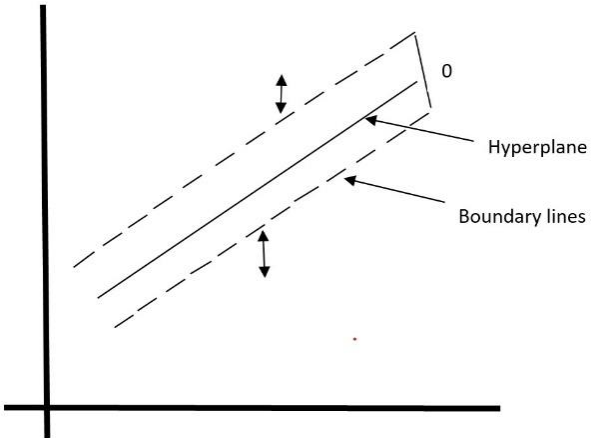

Figure 4: Support vector regression

4.4 Linear Regression

Linear regression is the linear estimation of a normal connection between two or more variables. Regression models are extremely important since they are among the most popular methods for reasoning and prediction and the activity goes like this [10].

$$\boldsymbol{Y = F\left(X_1, X_{2\cdots X_k}\right)} \qquad \textbf{..... (1)}$$

Independent variables labeled as X_1, and X_2, are the predictors. Y is the function of the X variable. The regression model is a linear estimation of this function. T.

4.5 Artificial Neural Network

A neural network means, the connection of neurons. It constructs the base of deep learning. The working of Human neural networks and Artificial neural networks are almost the same, although there is a difference in their mechanism.

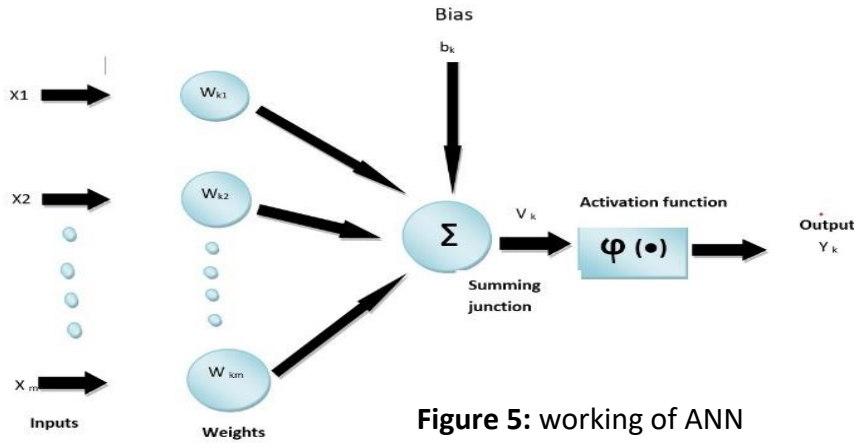

Figure 5: working of ANN

An artificial neural network works in three layers, the Input layer, and the output layer, and in between there exists the Hidden layer. The input layer receives the data, the output layer provides the final result, and the hidden layer is in charge of handling all of the weight distribution, calculating bias, and calculations [11][12]. The formula is mentioned below:

$$\Sigma_{i=1}^{n} W * X_i + B \qquad \text{..... (2)}$$

$$y_{1=Activation(W1*X_1+W_2*x_2+\cdots W_n*Xn)} \qquad \text{..... (3)}$$

4.6 Work-flow diagram

Presentation metrics of model: The four standard assessment metrics, the Root mean squared error (RMSE), Mean squared error (MSE), coefficient of determination (R-squared error), and mean absolute percentage error (MAPE) are used to determine the anticipated outputs of all five machine learning models.

-RMSE can be calculated by the given formula:

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \cdot \frac{(\hat{y}_i - y_i)^2}{n}} \qquad \text{..... (4)}$$

-MSE can be calculated by the given formula:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad \text{..... (5)}$$

-MAPE can be calculated by the given formula

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \qquad \text{..... (6)}$$

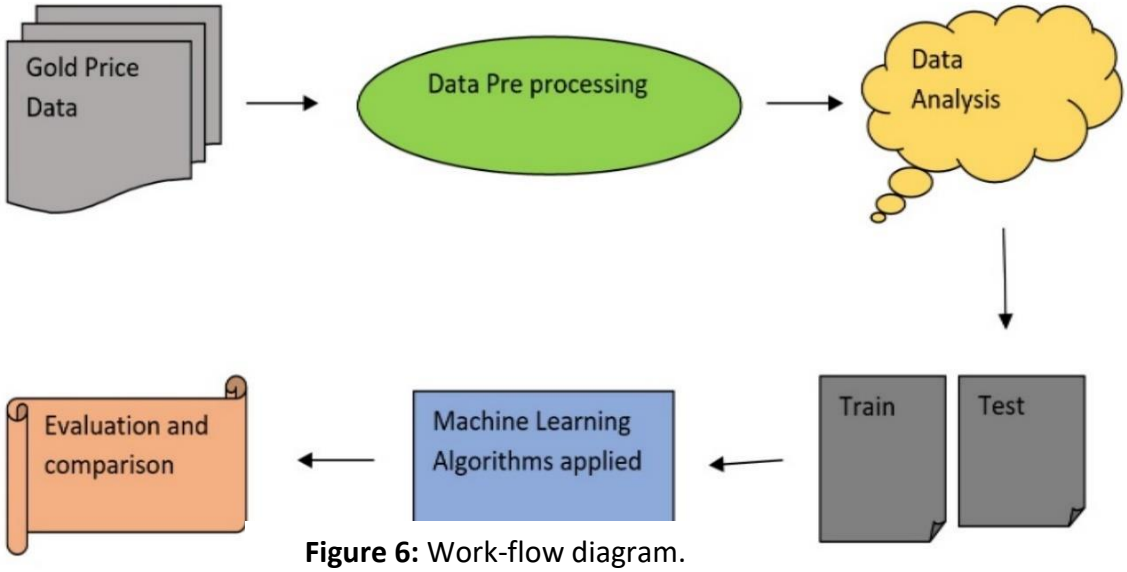

Figure 6: Work-flow diagram.

5. Results and discussions

A comparison of the model's performances based on the evaluation metrics has been listed below: After applying all the five machine learning algorithms, the least RMSE is calculated as 0.377 for ANN, and the least MAPE is calculated as 0.151, it is also for ANN.

Table 1
Machine Learning Algorithms

Serial No.	Machine Learning Algorithms applied	RMSE difference	R-squared error	MAPE difference
01.	ANN	0.377	0.65	0.151
02.	Linear regression	0.707	0.77	14.62
03.	SVR	0.585	0.83	78.72
04.	Random forest	0.248	0.98	83.26
05.	Decision tree	0.357	0.96	83.13

The best fit model on the basis of errors calculated, is an artificial neural network, with the least RMSE, MAPE, and R^2 error.

6. Conclusion

From this paper, it can be concluded that Machine learning algorithms can be used to forecast gold prices accurately. And this model can be very beneficial not only the government agencies and investors but to common people also. Moreover, for future work related to this, we can analyze and include some more variables such as diamond price, some other precious metal prices, or some other stock exchange indices that strongly influence the gold rates, to make our predicted values more accurate.

7. References

[1] M. Sravani, Ch. Abhilash, T. Divya, Ch. Vasthav, D. Priyanka, "Gold price prediction", International journal of creative research thoughts (IJCRT), Issue- 6 (June 2021).

[2] Iftikhar ul Sami, Khurum Nazir Junejo, "Predicting Future Gold Rates using Machine Learning Approach", International Journal of Advanced Computer Science and Applications, Issue-12(2017).

[3] V. K. F. B. Rebecca Davis, "Modeling and Forecasting of Gold Prices on Financia Markets," American International Journal of Contemporary Research, 2014.

[4] Iftikharul Sami and Khurum Nazir Junejo, "Predicting Future Gold Rates using Machine Learning Approach", International Journal of Advanced Computer Science and Applications, 2017.

[5] D Makala and Z Li, "Prediction of gold price with ARIMA and SVM", Journal of Physics: Conference Series, 2021.

[6] S. Chakravarty, B. K. Paikaray, R. Mishra and S. Dash, "Hyperspectral Image Classification using Spectral Angle Mapper," 2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), 2021, pp. 87-90, doi: 10.1109/WIECON-ECE54711.2021.9829585.

[7] Manjula K A, Karthikeyan P," Gold Price Prediction using Ensemble based Machine Learning Techniques", IEEE 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Issue-23-25 (April 2014).

[8] Z. Ismail, A. Yahya, A. Shabri," Forecasting Gold Prices Using Multiple Linear regression Method", American Journal of Applied Sciences, Issue-2009.

[9] L.A Sjaastad, F. Scacciavillani," The price of gold and exchange rate", Journal of International Money and Finance, Issue-December, 1996

[10] loannis E. Livieris, EmmanuelPintelas, Panagiotis Pintelas," A CNN-LSTM model for gold price time-series forecasting", Neural computing and Applications, Issue- 21(November 2019).

[11] Dr. Abhay Kumar Agarwal, Swati Kumari, "Gold Price Prediction using Machine Learning", International Journal of Trend in Scientific Research and Development (ijtsrd), 2020.

[12] Sami Ben Jabeur, Salma Mefteh-Wali, Jean-Laurent Viviani," Forecasting gold price with the XGBoost algorithm and SHAP interaction values, Annals of Operations Research, Issue-24(June 2021).