GOLD PRICE PREDICTION

[1]M.Sravani,[2]Ch.Abhilash,[3]T.Divya, [4]Ch.Vasthav,[5]D.Priyanka

[1,2,3,4]Student, Department of CSE, Gudlavalleru Engineering College, Gudlavalleru,India
[5]Assistant Professor, Department of CSE, Gudlavalleru Engineering College, Gudlavalleru, India

Abstract: Historically, gold is one of the valuable materials that is used for funding trading purchases. The countries with gold deposits are considered as prosperous nations. The financial strength of the country is reflected by its gold reserves. Several individuals and companies have participated in gold reserves in addition to the government departments. Gold rates are influenced by world's leading economies performance. This change in gold price has attracted more investors for investments in gold. Due to the uncertainity of gold market, it is dangerous for the transactions that are involving gold and there is uncertainty in the changes of gold price. To analyze the correlation between the economic variables and the gold prices, there are several studies available. The project "Gold price prediction" aims to predict the price of gold using machine learning. Generally, for predicting economic variables, machine learning is used. We use the Supervised learning algorithms like Multiple linear regression, Random forest regressor and Gradient boosting to train a model for predicting the gold price.

Index Terms - **Prediction, Supervised learning, Multiple linear Regression, Random Forest regressor, Gradient boosting.**

I. INTRODUCTION

Gold is one of the precious metals. It has been used as currency, for jewellery and other purposes. It is used as medium for money or exchange because of its limited supply and high value. This metal's scarcity and difficulty in extraction made it a valuable commodity. It also reflects the country's economical strength and hence many companies and individuals started to invest in gold reserves. Due to its increasing value, many people considered gold as an attractive investment.

Gold is preferred as protective asset by investors because of their negative expectations regarding the current situation in the foreign exchange and capital markets. Investors also consider gold as an asset to rely on, when the desirable profits are not achieved by the world capital markets. Since gold is stored and accumulated over years, the influence of an year's production on its price is less. The price of gold depends on currency fluctuations and other economic variables. The raise of gold prices and fall of prices in other markets has attracted more investors to invest in gold market. These changes in the price of gold made the investments risky and a fear has been developed that these prices would decrease.

There are several number of studies analyzing the relation between the gold price and other economic variables. Understanding the relation between these variables helps the investors to take better decisions. Hence, we use machine learning algorithms such as multiple linear regression, random forest and gradient boosting for analyzing the relation between the variables and predict the gold price.

II . RELATED WORK

V.K.F.B.Rebecca Davis in the paper "Modeling and Forecasting of Gold Prices on Financia Markets" uses Autoregressive Moving Average (ARMA) model, which is a statistical model that is frequently used for analyzing time series data. The data considered is the monthly prices of gold for over ten years. The model achieved an accuracy of 66.67%.

Iftikhar ul Sami and Khurum NazirJunejo in their paper titled "Predicting Future Gold Rates using Machine Learning Approach" uses Artificial Neural Networks(ANN) for predicting the gold prices. The data for this study is collected from various sources over a period of eleven years. This data included variables such as oil price, S&P 500 index , USD exchange rates and other economic variables.

D Makala and Z Li in their paper "Prediction of gold price with ARIMA and SVM" uses the data collected from World Gold Council containing the daily prices of gold from January 1979 to December 2019. This study uses Autoregressive Integrated Moving Average(ARIMA) method and SVM to forecast the gold price. The accuracy achieved by ARIMA model is less than the accuracy achieved by SVM.

"Big Data Analytics for Gold Price Forecasting Based on Decision Tree Algorithm and Support Vector Regression" by Navin And Dr.G.Vadivu uses the decision tree and support vector regression algorithms for predicting and forecasting gold prices using the data containing five attributes namely open, close, low, high values and volume of gold. In this study, the decision tree suited best for feature selection whereas the SVR suited best for large training data.

Sima P.Patil, V.M.Vasava and G.M.Poddar in their paper titled "Gold Market Analyzer using Selection based Algorithm" gave us the insight into the factors that are affecting the gold price. These factors include silver futures, copper futures, oil futures, EUR-USD,S&P/TSX composite. They used algorithms such as SVM, Logistic regression for prediction. The Logistic regression model achieved an accuracy of 61.92% outperforming the SVM model.

III . DATA AND METHODOLOGY

A. Dataset:

The data was sourced from the kaggle website consisting of ten years data from January 2008 to Decmeber 2018. It consists of the variables date, silver price, stock profit exchange, gold price ,US dollar rate and united states oil ETF. The dataset consists of 2290 records.

B. Machine Learning Algorithms:

For developing the model, we use the algorithms such as Multiple linear regression, Random forest and Gradient boosting.

Multiple Linear Regression is a statistical technique that uses multiple independent variables for predicting the outcome of the target variable. The mathematical expression is

$$Y = a + bX1 + cX2 + dX3 + \epsilon$$

where Y is the dependent variable, X1 and X2 and X3 are independent variables, a is the intercept, b ,c and d are slope values and ϵ is the error.

Random forest is a supervised learning algorithm which performs both classification and regression tasks. This algorithm operates by constructing multiple decision trees during training time and outputting the mean prediction of individual trees.

Gradient Boosting is a machine learning algorithm for classificiation and regression problems. This algorithm produces a prediction model in the form of an ensemble of weak prediction models, which are typically the decision trees. It builds the model in stage-wise fashion and generalizes them by enabling the optimization of an arbitary differentiable loss function. The gradient boosting trees usually outperforms random forest, but are prone to overfitting in some problems as the performance of this model improves over iterations.

IV. IMPLEMENTATION

To predict the gold price, we need to build a machine learning model which includes the following steps.
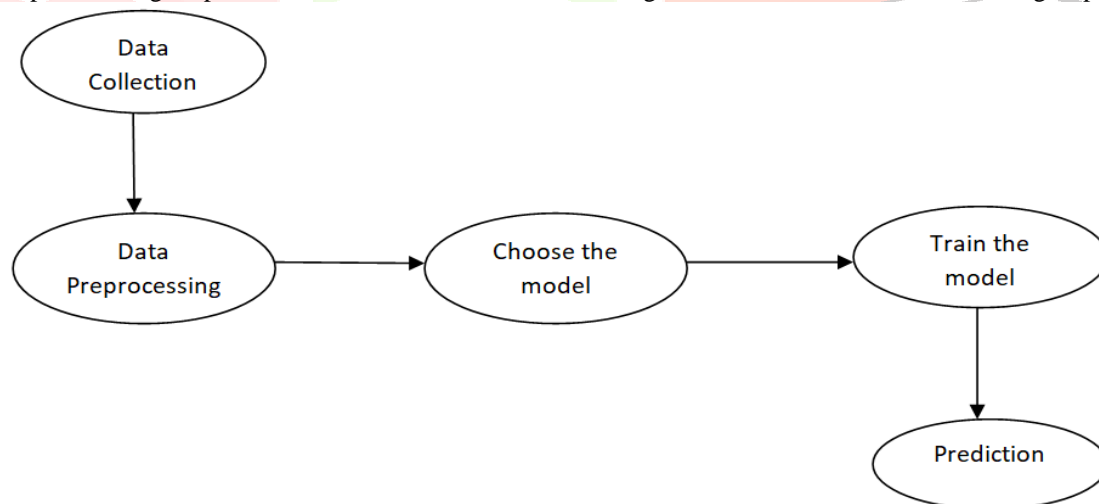

Figure 1 Proposed Method

1. Data Collection:

The first thing required while building a machine learning model is the data. The data is collected from kaggle website consisting of 2290 records and 6 attributes.

2. Data Preprocessing:

Data preprocessing is required when the data is incomplete, inconsistent or noisy. The data collected was noisy, so we performed outlier analysis and removed the noisy data. The data transformation is also done by performing normalization in which the data in each attribute is scaled between the range 0 to 1.

3.Choose the model:

Prediction of gold price is a regression task, so we consider the regression algorithms such as Multiple Linear Regression, Random Forest Regressor and Gradient Boosting for building the model.

4. Training the model:

    The model is trained by importing the required model and by passing the training data to it. The dataset is splitted into train and test data with test_size=0.20. The linear model is imported from sklearn and the Random forest regressor and Gradient boosting regressor modules are imported from sklearn.ensemble. These models are trained by passing the train data.

While conducting training, it is also important to record the metrics of each training process. The metrics that are tested are mean absolute error, root mean square error and r2 score.

5. Prediction:

    The trained model is checked by predicting the test data of the dependent variable using the test data of the independent variables.

IV. RESULTS

    After applying different regression techniques on the data, the results are as follows: When multiple regression is applied on the data, the accuracy (r2 score) obtained is 91% and RMSE is 5.56 which is high.

Table 1 Comparision of results

Regressor Model	MAE difference	RMSE difference	Accuracy on train data	Accuracy on test data	Accuracy difference
Random Forest	-0.04	-0.15	99.83	99.77	0.06
Gradient Boosting	-0.01	-0.13	98.71	98.59	0.12

    In random forest model, the accuracy obtained for training data is 99.83% and the accuracy obtained for test data is 99.77%. The accuracy difference is very less. The RMSE (-0.15) and MAE (-0.04) have only slight differences between the train and test data. For gradient boosting model, the accuracy obtained for training data is 98.71% and the accuracy obtained for test data is 98.59%. The accuracy difference is very less. The RMSE (-0.13) and MAE (0.01) have only slight differences between the train and test data.

    Hence Random Forest and Gradient Boosting best suited to this data. However, the accuracy is higher for random forest and the accuracy difference is also very less compared to gradient boosting. Hence, random forest regressor model is considered.

V. CONCLUSION AND FUTURE SCOPE

    The main aim of this study is to predict the gold price that is influenced by the economic variables such as stock profit exchange, silver price, EUR/USD, United States oil ETF. In this study, we used the machine learning algorithms such as multiple linear regression, random forest and gradient boosting to predict the price of gold accurately. Considering the results obtained, we conclude that the random forest model performed better than the other models.

    For future work, we can improve the results and predict the price more accurately by incorporating the other factors such as gold production, crude oil price, platinum price,inflation to the data and by using deep learning.

REFERENCES

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

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

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

[4] Navin, Dr. G. Vadivu, "Big Data Analytics for Gold Price Forecasting Based on Decision Tree Algorithm and Support Vector Regression (SVR)", International Journal of Science and Research (IJSR), 2013.

[5] P. V. M. Vasava, P. G. M. Poddar, Sima P Patel, "Gold Market Analyzer using Selection based Algorithm", International Journal of Advanced Engineering Research and Science (IJAERS), 2016.

[6] Megan Potoski,"Predicting Gold prices", CS229, Autumn 2013.

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