PRICE PREDICTION OF GOLD USING RANDOM FOREST

| By	By	By
Kothuri Vamshi	**Kalyanam Chakradhar**	**Lonkala Sandeep**
Student	*Student*	*Assistant Professor*
B V Raju Institute of Technology	*B V Raju Institute of Technology*	*B V Raju Institute of Technology*
Narsapur, Hyderabad 502313	*Narsapur, Hyderabad 502313*	*Narsapur, Hyderabad 502313*

ABSTRACT

Gold has been a valuable resource for funding trade and economic prosperity. Nations with gold reserves are often viewed as financially stable, and both government departments and individuals have invested in gold. The price of gold is influenced by the performance of major world economies, and this has attracted many investors to the market. However, the gold market is known for its uncertainty, which makes it risky for transactions and predicting price changes. Many studies have been conducted to analyze the correlation between economic variables and gold prices, and one such project is "Gold Price Prediction", which aims to use machine learning to predict the price of gold. Supervised learning algorithm such as Random Forest Regressor are commonly used for predicting economic variables and will be used to train the model for this project.

1.INTRODUCTION

Over time, investing in gold has evolved from traditional methods like buying jewelry to modern strategies such as purchasing gold coins and bars available in banks. Historically, gold was used as a form of currency in various parts of the world, including the USA, and continues to be used to assess the monetary strength of a country. Large investors have invested vast amounts of money in gold, while rising global economies like China, Russia, and India are now major buyers of gold, with the USA, South Africa, and Australia being major sellers. The value of gold is also influenced by Chinese and Indian traditional events, leading to increased demand and investment. Small investors consider gold a safe investment compared to other options with inherent risks. The economic conditions of these countries also play a role in determining gold's spot rates. Gold is increasingly being seen as an attractive investment due to its rising value and versatility. Investors use gold as a protective asset when they have negative expectations about the performance of developed exchange and capital markets. It is considered the "ultimate asset" when other markets fail to provide desirable returns. Therefore, investors view gold as a tool to hedge against fluctuations in other markets. As a valuable commodity, gold's value is based on supply and demand.

Numerous studies have been conducted to examine the correlation between gold prices and other economic factors. Investors can make better decisions by understanding this relationship. To analyze this relationship and predict gold prices, machine learning algorithms such as multiple linear regression and random forest are utilized.

Work Flow
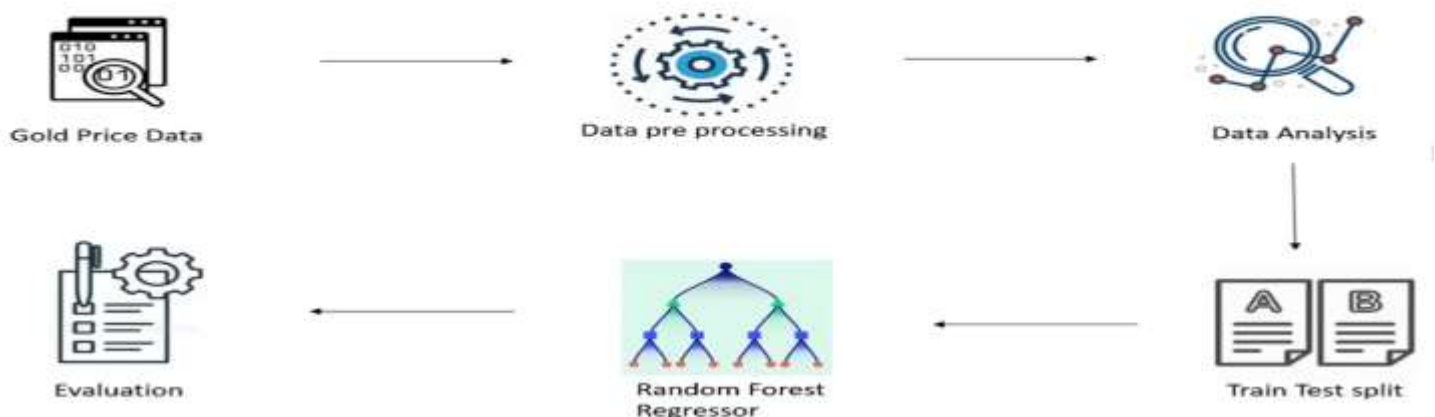
2.RESEARCH METHODOLOGY

Our project aims to improve the accuracy of gold price prediction by utilizing a dataset with various attributes in addition to the existing system. The Random Forest algorithm, which is a classifier that employs multiple decision trees on different subsets of the dataset, is utilized to enhance predictive accuracy. Instead of relying on a single decision tree, Random Forest calculates the average of predictions from each tree and makes the final output prediction based on the majority vote. A higher number of trees in the forest

enhances accuracy and mitigates overfitting issues. This approach results in higher predictive accuracy.

2.1 Assumptions for Random Forest: -

The random forest algorithm utilizes multiple decision trees to predict the class of the dataset. Due to the use of multiple trees, it is possible that some trees may predict the correct output while others may not. However, by combining the predictions of all trees, the random forest algorithm produces the correct output. To improve the accuracy of the random forest classifier, there are two assumptions that need to be considered. Firstly, the feature variable of the dataset should have actual values to ensure the classifier predicts accurate results, rather than random or estimated results. Secondly, the predictions from each tree should have minimal correlations to avoid any biases and improve the overall accuracy of the classifier.

2.2Why Random Forest? :-

- The algorithm requires less time for training in comparison to other algorithms.
- It has a high accuracy rate in predicting output and performs efficiently even when dealing with large datasets.
- Additionally, it can maintain accuracy even when a considerable amount of data is missing.

2.3 Working process of Random Forest: -

Random Forest operates in two stages: the first stage involves creating the forest by combining multiple decision trees, and the second stage involves making predictions for each tree that was created in the first stage. The following steps and diagram illustrate the process:

Step 1: Choose K data points at random from the training set.

Step 2: Build decision trees using the selected data points (subsets).

Step 3: Select the number of decision trees, N, that you want to build.

Step 4: Repeat Steps 1 and 2.

Step 5: When presented with new data points, find the predictions of each decision tree, and assign the new data points to the category with the highest number of votes.
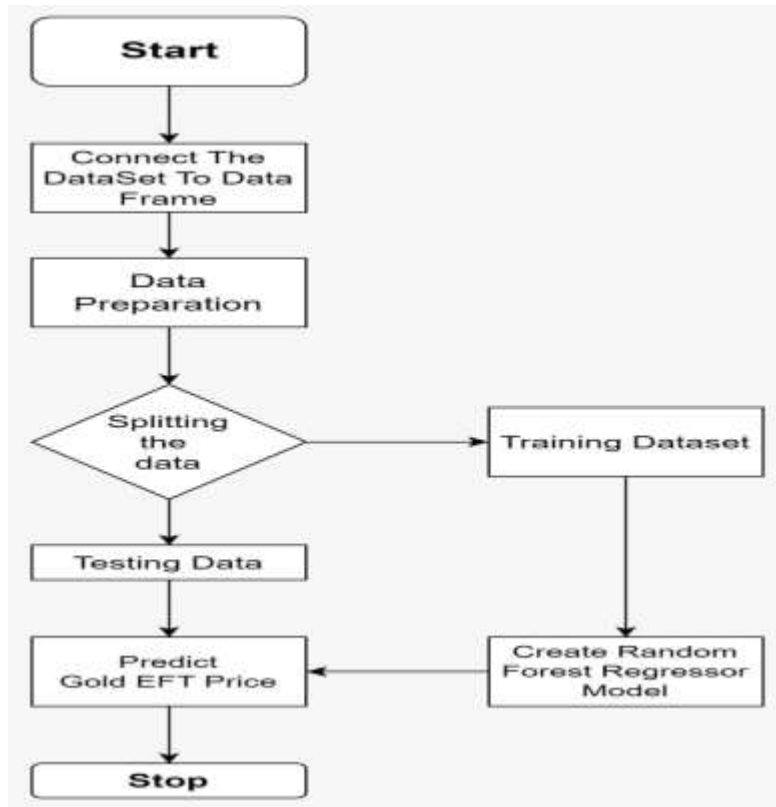
2.4 Description of algorithm: -

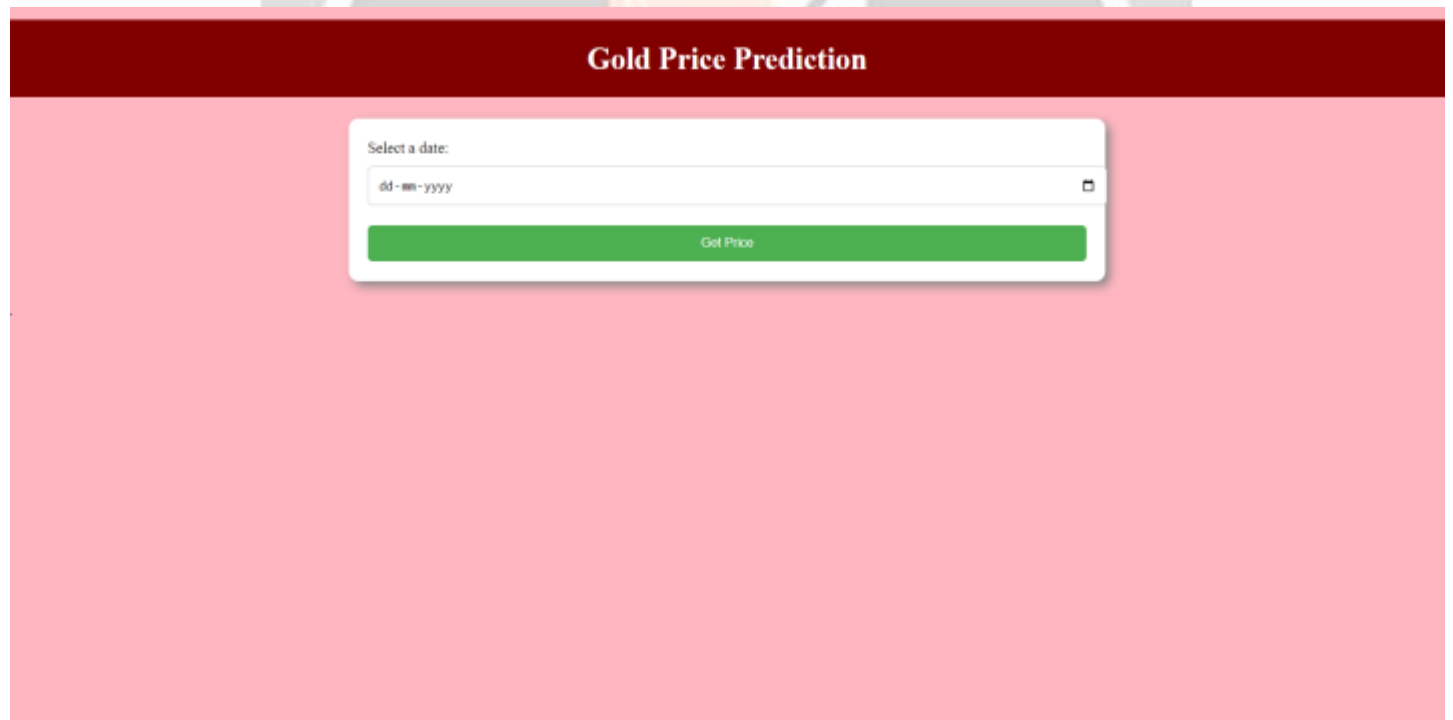Step 1: Start with a dataset containing various data points.

Step 2: Assign equal weights to each data point as input to the model.

Step 3: Identify the data points that are incorrectly classified and increase their weight.

Step 4: If the desired output is obtained, stop the process. Otherwise, go back to Step 2   and
Repeat the process.

3. RESULTS:

4.CONCLUSION:

Random Forest is a powerful algorithm that can handle a large number of features and nonlinear relationships between them. The model was able to capture the patterns in the historical data and make accurate predictions on the test set. The feature importance analysis indicated that the most important factors affecting the price of gold were the USD/INR exchange rate, the S&P 500 index, and the crude oil price. The model's accuracy could be improved further by incorporating additional relevant features and data sources.

Overall, the Random Forest algorithm showed promising results in predicting the price of gold and could be a useful tool for investors and analysts looking to make informed decisions in the gold market. However, as with any predictive model, it is important to note that the predictions are not guaranteed to be accurate and should be used in conjunction with other forms of analysis and market research.

5.REFERENCES

1.   K. R SekarManav Srinivasan, K. S. Ravichandran and J. Sethuraman, "Gold Price Estimation Using A Multi Variable Model", International Conference on Networks & Advances in Computational Technologies, 2017.

2.   Sima P. Patil, Prof. V. M. Vasava, Prof. G. M. Poddar, " Gold Market Analyzer using Selection based Algorithm", International Journal of Advanced Engineering Research and Science, 2016.

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.