**Gold Price Prediction Using Machine Learning Regression Techniques**

### **Introduction**

Gold rates are an important indicator of economic stability and financial markets worldwide. It is a popular investment asset due to its perceived stability and ability to act as a hedge against inflation and economic uncertainty. As such, accurately predicting future gold prices is of great interest to investors, financial analysts, and traders alike. The gold market is influenced by a range of factors such as global economic conditions, political instability, inflation, and currency exchange rates. In recent years, the use of machine learning and artificial intelligence has gained significant attention for its potential to predict gold rates with greater accuracy. By analyzing data and identifying patterns and trends, a predictive model can be developed to make informed forecasts about future gold prices, which could help investors make informed decisions.

### **Problem Definition**

To use data manipulation algorithms to predict gold rates based on historical data. The project will involve collecting and analyzing a large dataset of gold prices and relevant economic indicators. The project will then apply machine learning techniques to develop predictive models that can forecast future gold prices with a high degree of accuracy.

**Predictive Model Selection**

For this prediction model, we chose to use various techniques based on regression, such as Linear regression, Random Forest Regression, and Support Vector Regression. Also, we chose to perform data preprocessing, data visualization, and data analysis to make predictions about the gold prices.

Here, linear regression was used to determine the linear relationship between gold prices and various stock indexes. Support Vector Regression was used to fit the best hyperplane that maximizes the margin between the data points. Also, the Random Forest Regression was used for classification, regression, and other tasks, and operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes, or prediction, and mean prediction, or regression, of the individual trees.

Overall, we have found that Linear Regression seems to have a better (R2) score compared to the other regression models.

### **Analysis**

|  |
| --- |
| In this project, we forecasted gold rates using the most comprehensive set of features and applied various machine learning algorithms for forecasting and comparing results. |
|  |
| The excel data shown is provided from finance.yahoo.com to be used by our code to better predict gold prices. |
|  |
| “Effect of Index prices on gold rates” chart displays the results provided from the SPDR Gold Trust (GLD) Exchange Traded Fund data at finance.yahoo.com website. The data ranges between 2012 to 2019 showing the results between GLD, SPY, and DJ. |
|  |
| “Last 100 records of daily return of all features” chart displays the results of the daily returns of all the features within the last five months. The data shows all of the features together while RHO shows a steep decline of -1 daily returns. |
|  |
| “Last 100 records of daily return Stock Indexes” chart displays the results of the daily returns of GLD, SPY, and DJ within the last five months. The data shows all three of them fluctuating up and down as naturally expected due to the process of how the stock market works. |
|  |
| “Plotting of Mean, Standard deviation and Kurtosis of Gold Prices” bar graph displays the mean, standard deviation, and kurtosis of gold prices, as well as showing the highest results within the graph. |
| Evaluation Metrics: Used Root Mean Square (RSME) and R2 as our evaluation metrics. Root Mean Square Error (RSME) is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are. The formula for calculating RMSE is given below.  RSME is always non-negative, and a value of 0 (almost never achieved in practice) would indicate a perfect fit to the data.  R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regressions. |

### **Conclusion**

Forecasting rise and fall in the daily gold rates, can help investors to decide when to buy (or sell) the commodity.

We in this project forecasted gold rates using the most comprehensive set of features and applied various machine learning algorithms for forecasting and comparing their results. We also identify the attributes that highly influence the gold rates.

Machine learning algorithms can be used to forecast gold prices accurately. And this model can be very beneficial not only for the government agencies and investors but to common people also.

As you saw in this project, we first train a machine learning model, then use the trained model for prediction. Similarly, any model can be made much more precise, by feeding a very large dataset, to get a very accurate score.

### 

### **Recommendations and Citations**

[1] Ghule, Rushikesh & Gadhave, Abhijeet & Dubey, Manish & Kharade, Jyoti & Mumbai, Navi. (2022). Gold Price Prediction using Machine Learning. International Journal of Environmental Engineering. 6. 10.55041/IJSREM15027.

https://www.researchgate.net/publication/362491642\_Gold\_Price\_Prediction\_using\_Machine\_Learning

[2] Sami, Iftikhar & Nazir, Khurram. (2018). Predicting Future Gold Rates using Machine Learning Approach. International Journal of Advanced Computer Science and Applications. 8. 92-99. 10.14569/IJACSA.2017.081213.

<https://www.researchgate.net/publication/322222520_Predicting_Future_Gold_Rates_using_Machine_Learning_Approach>

[3] Ka, Manjula & .P, Karthikeyan. (2019). Gold Price Prediction using Ensemble based Machine Learning Techniques. 1360-1364. 10.1109/ICOEI.2019.8862557.

<https://www.researchgate.net/publication/326048272_Using_classification_techniques_to_predict_gold_price_movement>