Gold Price Prediction

Liyan Susan Kurian
[1]PG Scholar,Department Of Computer Application
Amal Jyothi College of
Engineering,kanjirapally,686518
liyansusankurian@mca.ajce.in

Merin Chacko
Assistant Professor
Amal Jyothi College of
Engineering,kanjirapally,686518
merinchacko@ajce.ac.in

***Abstract*—Gold is often used by investors as a barrier against inflation or adverse economic times. As a result, it is critical for investors to be able to accurately estimate gold prices. This article is based on a study of gold price prediction by relationship between gold price and selected factors influencing it , namely date, stock value, current gold price ,united state oil price, current silver price, currency medium(EUR/USD) using Colab by random forest regression algorithm. Comparing and Analyze R squared error graph and mean_absolute_error, and with linear regression algorithm. Monthly price data for the period January 2008 to May 2018 was used for the study. Two machine learning algorithms random forest regression and linear regression were used in analyzing these data. Random forest regression, on the other hand, has been found to have greater overall prediction accuracy.**

Keywords—Machine Learning, Random Forest Regression, Linear Regression , Prediction.

I. INTRODUCTION

Investments mention to the employment of current funds with an objective of earning a favorable return on it in future. In an profitable sense, an funding can be considered as the purchase of assets that are not consumed today but are used in the future to create good wealth. Number of investment avenues are available for investors, which include stocks, deposits, commodities, and real estate. Each of them has its own risk and reward characteristics. Gold is another asset which is being considered as an investment path by many investors due to its growth in value and the area of usage. Gold is a valuable metal, so like other than other goods, gold's price should depend on supply and demand. But, since gold is storable and the supply is accumulated, this year's production has influence on its prices. Gold behaves less like a product than long-lived assets like stocks or bonds. The mark price is the current market price at which commodity is purchased or sold for immediate payment and delivery. It is different from the future price, which is the price at which the two parties deal to transact on future date.

II. EASE OF USE

There are so many studies dealing with the price of gold in the world. Although various different kind of variables are used in these studies, it is predict the gold prices . gold price prediction by relationship between gold price and selected factors influencing it , namely date, stock value, current gold price ,united state oil price, current silver price, currency medium(EUR/USD) using Colab by random forest regression algorithm**.**

III. KEY FACTOR

- When the inflation is high, the demand for gold rises and so on.
- India is one of the world's top gold importers, and changes in import prices, as a result of global price movements, are mirrored in domestic gold prices.
- Central banks of most of the countries hold both currency as well as gold reserves

IV.PURPOSE OF THE STUDY

This document is to analyse variation in gold price and predict the gold price using machine learning , analyse and compare with two algorithms random force regression and linear regression. And study the different the value of R squared error graph and mean_absolute_error.

IV. METHODOLOGY

The purpose of this paper is find a Machine Learning model which can predict gold price with accuracy from the given dataset. The model should be able to classify correctly the dataset into actual value and predicted value.

A. Colab

Colaboratory, or "Colab" is a product from Google Research that runs entirely in the cloud. Colab allows us to execute python code through the browser platform, and is mainly well suite to machine learning, data analysis and algorithms. Colab is a hosted Jupyter notebook carrier that doesn't require any setup and gives you free get entry to to computing resources, along with GPUs.
You may input an image dataset into Colab, train an image classifier on it, and test the model, all in only some lines of code. Colab notebooks execute code on Googles cloud servers, which means you have an advantage of Google hardware, as well as GPUs and TPUs, nevertheless of the power of your machine. All you need is a browser.

DOI: 10.5281/zenodo.6183215
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

B. Machine language

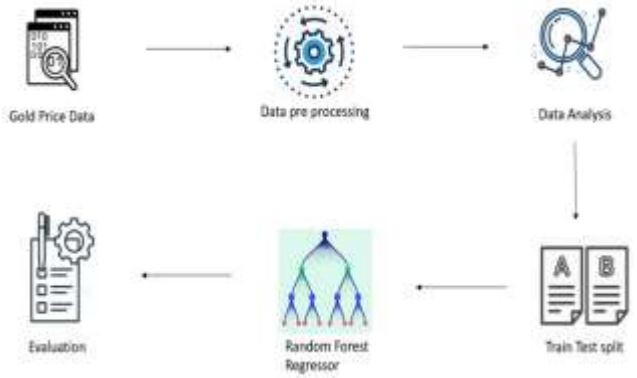
Machine learning is the technique of examine the data that automates analytical model building. It is related to artificial intelligence based on the aim that systems can learn from data, find patterns and make conclusion with very minimum human interaction. With the advent of new computing technologies, the present day machine learning is totally different from how it was in its inception. Machine learning was made from the pattern recognition methodology and from the theory that machine learning is possible without being programmed for specific tasks. The most important aspect of machine learning is the iterative aspect as the data models are uncover to new data they adapt independently.

C. Random Forest Reggression

1. A Random Forest is an ensemble technique that use several decision trees and a technique called Bootstrap Aggregation to solve both regression and classification problems, commonly known as bagging. Instead of depending on individual decision trees, the main idea is to integrate multiple decision trees to determine the final outcome. Random forests use the decision tree bootstrap to reduce the variance while maintaining the low bias that results from the decision tree model.A Random Forest algorithm has the following advantages when compared to most of the other algorithms - The overfitting problem will never come when we use the random forest algorithm in any classification problem. Both classification and regression tasks can be solved using the same random forest approach. You can also use the Random Forest algorithm for feature engineering to identify the most important features from the available features of your training dataset.

2. 1.***Workflow***

Figure 1



3. 2.***Dataset***

Date	SPX	GLD	USO	SLV	EUR/USD
0 . . .					

Date: Monthly price data for the period of January 2008 to May 2018.
SPX: Capitalization index of 500 companies.
GLD: Gold price
USO: United state oil price
SLV: Sliver price
EUR/USD: Currency medium.

4. There are two dataset csv files which should be uploaded to colab which is used for processing. We use data frames for analysis and processing. It obtain some statistical measures about the data.

```
# getting the statistical measures of the data
gold_data.describe()
```

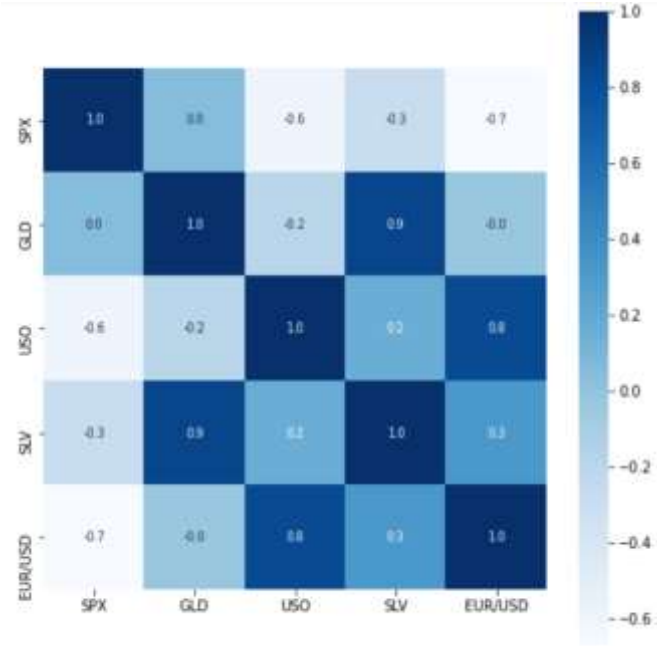
	SPX	GLD	USO	SLV	EUR/USD
count	2290.000000	2290.000000	2290.000000	2290.000000	2290.000000
mean	1654.315776	122.732875	31.842221	20.084997	1.283653
std	519.111540	23.283346	19.523517	7.092566	0.131547
min	676.530029	70.000000	7.960000	8.850000	1.039047
25%	1239.874969	109.725000	14.380000	15.570000	1.171313
50%	1551.434998	120.580002	33.869999	17.268500	1.303296
75%	2073.010070	132.840004	37.827501	22.882499	1.369971
max	2872.870117	184.589996	117.480003	47.259998	1.598798

From this we will get count, mean, std, min value of each column. mean value of gold price=122.7328.

Find the correlation between the various columns in a dataset, basically there are two types of correlation

a.Positive Correlation
b.Negative Correlation

DOI: 10.5281/zenodo.6183215
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam



From this map we analysis that silver is directly correlated with gold price and USO is slightly negatively correlated with gold price.

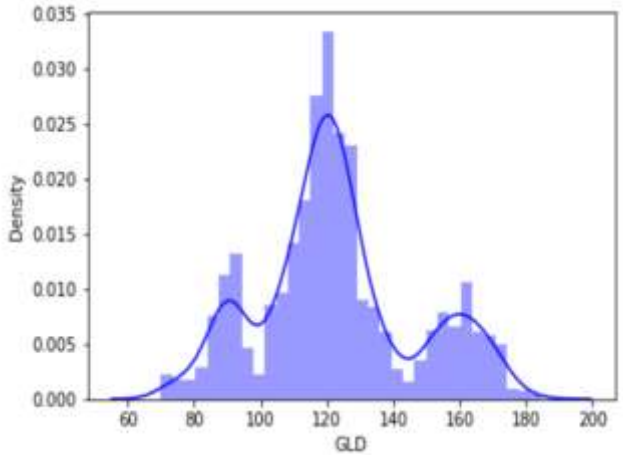


Figure 3:Distribution plot of gold price

So as you can see here,the most values lie in the range of 120 and less values in 180 range and bump in around 90 and in 160 but the majority of the values are in 120.

Then splitting the features and target by gold as one table and other attributes in another table.After this splitting into training data and test data and performing Random Forest regressor algorithm .Hereafter the random forest regressor is trained now we can use this trained model to make some predictions and we can evaluate our model based on the test data and we use predict() to predict the gold price.

```
[ ] error_score = metrics.r2_score(Y_test, test_data_prediction)
    print("R squared error : ", error_score)

    R squared error :  0.9891014138229717
```

The error score is 0.98,so from our data values we analyze that error value of 0.98 is kind of very less so that means our model is performind really well .

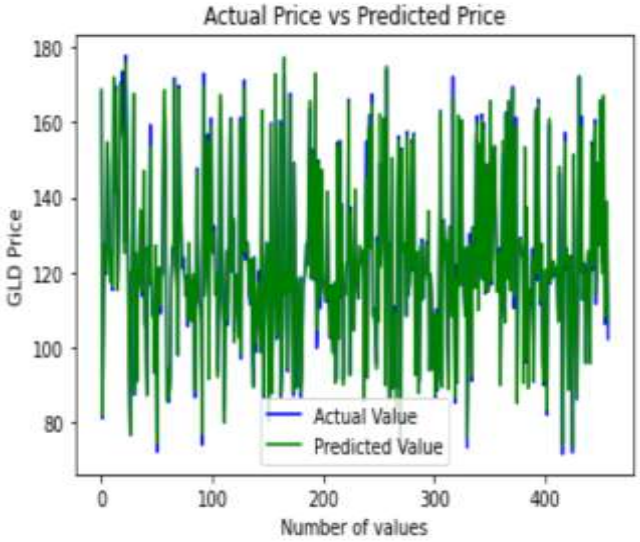


Figure 4:Plot of Actual price / Predicted price

From the above figure we understand that the predicted price and the values are you know very close to each other so the actual price is kind of slightly more than the predicted value .

DOI: 10.5281/zenodo.6183215
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

```
abc=metrics.mean_absolute_error(Y_test,test_data_prediction)

[ ] print(abc)

    1.3234271305021816
```

Mean absolute error of this model is 1.323 so the Random Forest Regression model will give more accurate result.

V. IMPLEMENTATION

In order to compare the mean absolute error and R squared error that is obtained from Random Forest Regression Model, we use Linear Regression Model.

A. Linear Regression Model

The simplest and most well-known Machine Learning algorithm is linear regression. It's a predictive analysis method that uses analytic techniques. Linear regression takes predictions for continuous, actual or numeric variables along with sales, salary, age, product price, etc.
The linear regression algorithm displays a linear relationship between a dependent(y) and one or more independent(y) variables. Since linear regression shows the linear relationship, It determine how the values of the dependent variable changes when the value of the individual variable change.

Then preforming the prediction with linear regression model We found R squared error=0.88

```
[ ] error_score = metrics.r2_score(Y_train, training_data_prediction)
    print("R squared Error : ", error_score)

    R squared Error :  0.8863976496408241
```

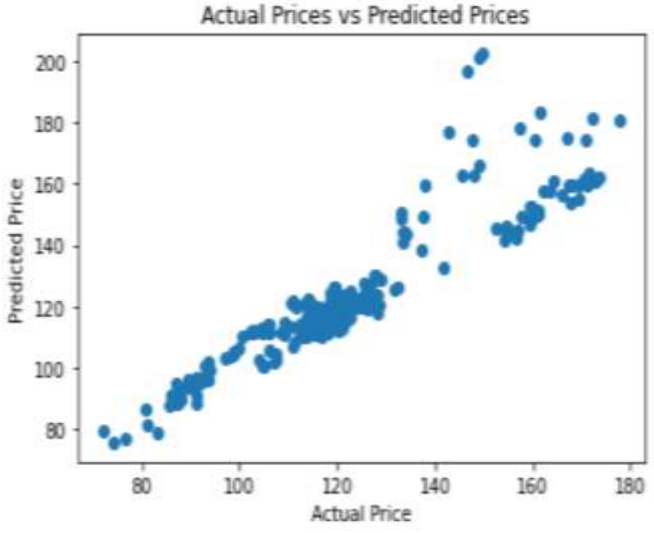


Figure 5:scattered graph of Actual price/Predicted price

From the above figure by using linear regression we understand that the predicted price and the values are you know very close to each other so the actual price is kind of slightly more than the predicted value .

```
abc=metrics.mean_absolute_error(Y_test,test_data_prediction)

[ ] print(abc)

    6.280344071340783
```

Mean absolute error of this model is 6.28 so Linear Regression model will give less accurate result than Random Forest Regression model.

VI. CONCLUSION

Machine language	Mean absolute error	R_Square_Error
Random Forest Regression	1.3234	0.9891
Linear Regression	6.28034	0.8863

DOI: 10.5281/zenodo.6183215
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

From this we can conclude that Mean absolute error of random forest regression is lower than linear regression so random forest regression is better than linear regression model.

REFERENCES

[1] J. Jagerson and S. W. Hansen, "All about investing in gold", McGraw-Hill Publishing, 2011.

[2] Z. Ismail, A. Yahya, and A. Shabri, "Forecasting gold prices using multiple linear regression method," Am. J. Appl. Sci., vol. 6, no. 8, p. 1509, 2009.

[3] H. Mombeini and A. Yazdani-Chamzini, "Modeling gold price via artificial neural network," J. Econ. Bus. Manag., vol. 3, no. 7, pp. 699–703, 2015.

[4] D. Ghosh, E. J. Levin, P. Macmillan, and R. E. Wright, Gold as an inflation hedge?," Stud. Econ. Finance, vol. 22, no. 1, pp. 1–25, 2004

DOI: 10.5281/zenodo.6183215
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam