Gold Price Predicting with Machine Learning Algorithm

Athul Sreelesh
MCA Department
Amal Jyothi College of Engineering,
Kanjirappally,Kerala, India
athulsreelesh@mca.ajce.in

Ms.Grace Joseph
MCA Department
Amal Jyothi College of Engineering,
Kanjirappally,Kerala, India
gracejoseph@amaljyothi.ac.in

***Abstract*— The article is made on the basis of a study held to find out whether any relation of gold with other particular factors influence its price in the global market. Those factors are specifically SPX, USO, Silver price and Currency pair quotation of the Euro against the US. The price data of almost every working day from January 02 2008 to May 16 2018 were used to conduct the study. Three of the Machine Learning algorithms namely, Decision Tree Regressor, XGB Regressor and Random Forest Regressor, were used in analyzing the set of data. After analyzing the whole data, the best result was obtained by the decision tree regressor and the xgb regressor. So we took xgb for further data model building.**

Keywords: Price Prediction, Machine Learning

I. INTRODUCTION

Investments and savings are an important element of everyone's life. Investments are meant to earn favourable return in the future by employing our present funds in some trustworthy assets. In economic view, it is considered to be the purchasing of assets which are kept low in the present and to be used in the future for creating our wealth. According to finance, these investments can be sold for higher price which in turn increases the profit. The Indian economy ,one of the fastest growing in the world, has gained a huge shift in increasing investment avenues, from bottom to top. Many investment avenues are present nowadays for the investors like, Mutual funds, commodities etc. Risk is an inborn factor in every area of invesment. One such important asset is Gold. Gold is considered as an attractive investment avenue due to its area of usage and increase in value. Most of the people prefer gold as their assets, as they are not satisfied with the current scenario in the global market.

This will increase the preference for gold. It became a major asset. So we can say that, gold is a tool for the investors to hedge against the fluctuations in other markets. Gold price depends on other factors also. It behaves more as an asset than a commodity. The price of gold depends on various factors such as currency value, political issues, Transportation cost etc.

This increase in value of gold and down of price in other markets led more investors to be attracted towards gold. But at some point when gold price went down ,investing in gold became riskier. Investors were quite worried whether the price would remain same or fall again. Many research have been conducted on the relationship between gold and other commodities. The outcome reveals the elements that influence the price of gold. So, this paper is concentrating on the gold price with respect to other factors in the economy. Understanding this is very much helpful for investors. To analyse the data, we employ three machine learning algorithms: Decision Tree Regressor, XGB Regressor, and Random Forest Regressor. We can find accurate data by comparing these three algorithms under various circumstances.

II. DATA AND METHODOLOGY

The investigation is carried out using Machine Learning. In order to acquire the best possible result, both data training and testing were undertaken. The Decision Tree Regressor, XGB Regressor, and Random Forest Regressor were among the machine learning methods employed in this work. Regression analysis is a statistical tool for determining the relationship between two or more variables. When one of the independent variables changes while the other variables remain constant, regression analysis is performed to see how the value of the dependent variable changes.

Decision Tree Regression follows a predictive model. It uses a set of binary rules and then calculate target value. Every individual tree have its own branches, nodes and leaves. A decision tree is a type of tree that can be used to forecast and classify data. Trees that are grown to a great depth in order to learn extremely irregular patterns, on the other hand, tend to over-fit the training sets. The tree can develop in an unexpected way if there is a modest amount of noise in the data.

By fitting a number of classification decision trees on multiple sub-samples of the dataset, a random forest uses averaging to improve projected accuracy. It have the ability to prevent over-fitting The max samples parameter controls the sub-sample size. If bootstrap=True (the default), the entire dataset is used to create each tree; or, each tree will be build with whole dataset.

DOI: 10.5281/zenodo.6168294
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

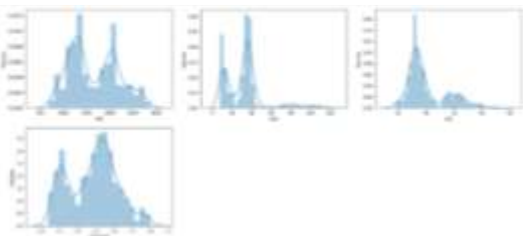
A Random Forest is an ensemble technique that uses several decision trees with a technique called Bootstrap Aggregation, sometimes known as bagging, to solve both regression and classification problems. Instead than depending on individual decision trees, the main idea is to aggregate numerous decision trees to determine the final outcome. The Random Forest reduces variance while keeping the low bias of a Decision Tree model by using bootstrapping on Decision Trees. The following are the advantages of Random Forest: When we utilise the random forest technique to solve any classification problem, the overfitting problem will never happen. The same random forest approach can be used to handle classification and regression problems. The random forest algorithm can also be used for feature engineering, which is the process of determining the most essential attributes from a set of options.

Gradient boosting is a technique for selecting the best forecast from a large group. It helps to construct the model in a stage-by-stage manner, as other boosting techniques do. It also generalises them by allowing for the optimization of an object. Any differentiable loss function can be used. Trees that increase gradients in many cases, they are superior to random forests, although in some conditions, it's prone to overfitting. However, there are several exceptions. Techniques for overcoming the same and constructing a more widespread using a combination of characteristics such as learning rate, trees can be created.

Python is used to implement these machine learning methods (Binary Tree Regression, Random Forest Regression, and XGB Regressor) in this work. The regression algorithms' prediction accuracy was assessed using Mean Squared Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

III. RESULTS AND DISCUSSION

Distribution of Continuous Numerical Features is plotted:

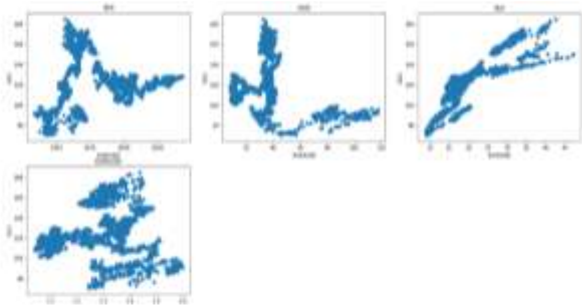


It seems that the SPX, a free-float weighted measurement stock market index of the 500 largest companies listed on

stock exchanges in the United States, Silver Price and Currency pair quotation of the Euro against the US are distributed normally. The United States Oil Fund is highly tilted to the right, with a few outliers.

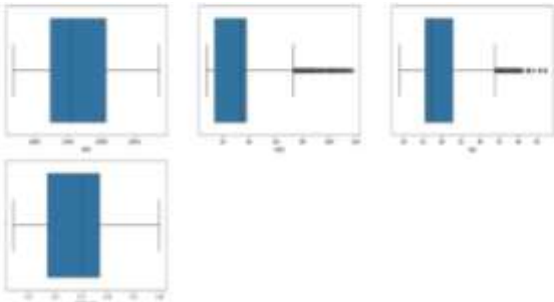
Relation between Continuous Numerical Features and Labels

Are plotted as given:

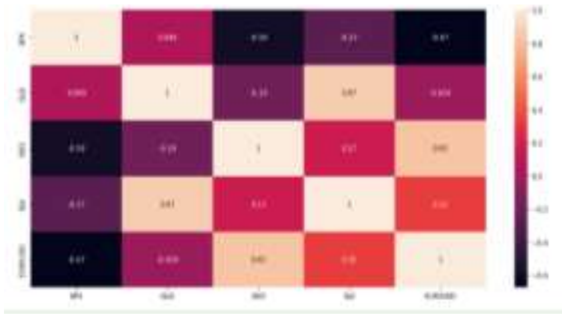


From the Continuous Numerical Features and Labels we can say that the Silver price is linearly progressing with Gold.

By checking Outliers in Numerical Features and we get:



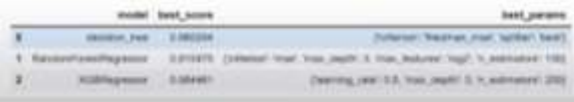
There are several outliers in the US Oil Fund and the Silver price. The following are the results of utilising heat maps to correlate numerical features:



```
GLD         1.000000
SLV         0.866632
SPX         0.049345
EUR/USD    -0.024375
USO        -0.186360
Name: GLD, dtype: float64
```

The silver feature is heavily correlated with Gold. We find that the XGB Regressor and Decision Tree Regressor produce the greatest results after implementing all three techniques.

DOI: 10.5281/zenodo.6168294
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam



So we randomly took XGB for building the model.

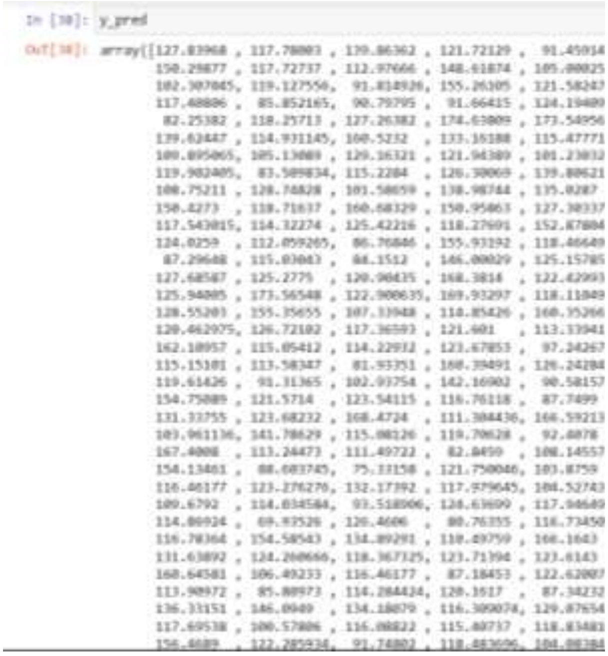
model_xgb = XGBRegressor(learning_rate=0.5, max_depth=3, n_estimators=200)

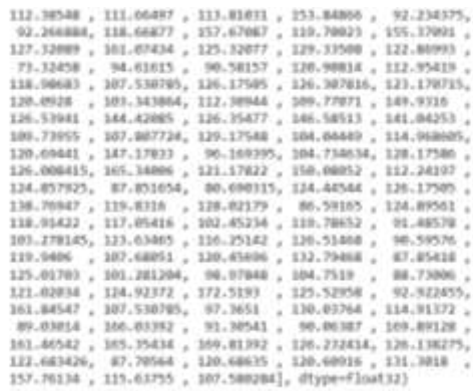
model_xgb.fit(x_train,y_train)

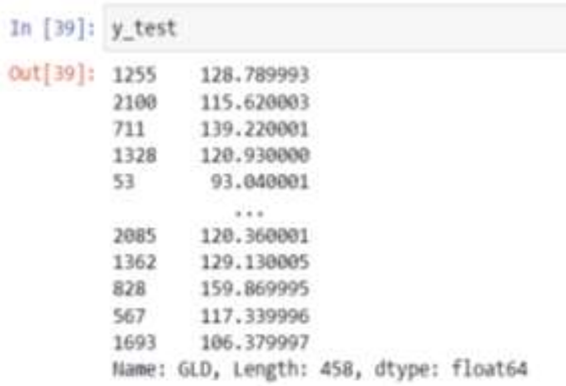
model_xgb.score(x_test,y_test)

y_pred= model_xgb.predict(x_test)

y_pred







```
In [39]: y_test

Out[39]: 1255    128.789993
         2100    115.620003
         711     139.220001
         1328    120.930000
         53       93.040001
                    ...
         2085    120.360001
         1362    129.130005
         828     159.869995
         567     117.339996
         1693    106.379997
         Name: GLD, Length: 458, dtype: float64
```

Based on the results from these analyses, it is being concluded that there are fluctuation in gold price in the given period of time in the dataset. When the trend of the dependent variable varies but there are no substantial changes in the trend of the independent variables, the accuracy of various approaches may differ. As a result, the model chosen should be based on the relationship between the study's variables.

IV. CONCLUSION

This research was carried out to better understand the relationship between gold's price and a number of elements that influence it. Silver, the SPX, currency pair quotations of the Euro against the US, and the United States Oil Fund were all hot topics. Monthly price data from January 2008 to May 2018 was used in the research. Three machine learning algorithms were used to analyse the data: Decision Tree Regressor, XGB Regressor, and Random Forest Regressor. The silver characteristic is strongly linked to gold. We find that the XGB Regressor and Decision Tree Regressor produce the greatest results after implementing all three techniques. For the entire period, XGB regression is revealed to have superior prediction accuracy.

It is concluded that machine learning algorithms are very useful in such analysis, but the characteristics of the data influences their accuracy. Further research with such data

and different techniques may be conducted for better understanding of the performance of these techniques.

V. REFERENCES

[1]https://www.geeksforgeeks.org/python-decision-tree-regression-using-sklearn/

[2]https://scikitlearn.org/stable/auto_examples/tree/plot_tre_regression.html

[3] https://www.geeksforgeeks.org/xgboost-for-regression/

[4]https://www.datacamp.com/community/tutorials/xgboost-in-python

[5]https://www.geeksforgeeks.org/random-forest-regression-in-python/

DOI: 10.5281/zenodo.6168294
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam