

GOLD PRICE PREDICTION USING ML ALGORITHMS

**Radhamani V^{*1}, Manju D^{*2}, Bobby Prathikshana M^{*3}, Javagar M^{*4},
Nivetha V^{*5}, Rinubha P^{*6}**

^{*1*2} Assistant Professor, Department of Computing (Decision and Computing Sciences),
Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India

^{*3*4*5*6} M. Sc (Integrated) Decision and Computing Sciences, Department of Computing,
Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India

List of Authors

1. Radhamani V

Assistant Professor, Department of Computing (Decision and Computing Sciences),
Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India
Email ID: vradhamani@cit.edu.in Mobile: +919994314741

2. Dr. Manju D

Assistant Professor, Department of Computing (Decision and Computing Sciences),
Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India (Corresponding Author)
Email ID: manju@cit.edu.in Mobile: +919865229679

3. Bobby Prathikshana M

M. Sc (Integrated) Decision and Computing Sciences, Department of Computing,
Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India

4. Javagar M

M. Sc (Integrated) Decision and Computing Sciences, Department of Computing,
Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India

5. Nivetha V

M. Sc (Integrated) Decision and Computing Sciences, Department of Computing,
Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India

6. Rinubha P

M. Sc (Integrated) Decision and Computing Sciences, Department of Computing,
Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India

Abstract

Since ancient times, gold has been cherished for its value and worth. Back then, gold was primarily used for trading purposes and as a method of remuneration. But now, it is looked upon as an investment and is found to exhibit the wealth of a country. Expensive metals like gold, in critical times, are used to assure the reimbursement of money borrowed as well. Thereby, gold is not only found to behold the rich, but also the poor. During pandemic crises, like the Covid-19, investments on gold in early times, might have a beneficial impact. Hence, predicting gold rates with live data and investing on golds at the right time is quite useful. Various machine learning algorithms like the linear regression, decision tree and random forest have been used to predict the gold rates. By using different algorithms, we've come to a conclusion that the random forest method provides more accurate results.

Keywords: Gold rates, prediction, live gold rates, random forest, decision tree, linear regression, predictive analytics.

1. Introduction

Gold was the first well-known metal of our species. When we ponder the historical advancement of technology, we consider the development of iron and copper labor to be the greatest contributor to the economic and cultural advancement of our species, but gold came first. Gold has held its value and has been used as a means of assessing a country's financial strength. Big investors were attracted to this precious metal and invested large amounts in it.

In the early days, more money was invested in buying this basic product. Like most commodities, the price of gold is determined by supply and demand, including speculative demand. However, unlike most other raw materials, savings and disposal play a bigger role in influencing your price. Small investors also found this product to be a safe investment. Government investments in gold are largely determined by your financial conditions and interest rates as they are indicators of the strength of your economy. Activity is observed in the US, hence capital inflows into the gold market. Various phenomena are associated with gold rates and also affect the price. Gold spot prices are set twice a day based on the supply and demand of the gold market. A slight change in the price of gold can result in large gains or losses for both these investors and government banks.

The reason for the rise in the gold rate apparently lies in its incredible use and it is also a very rare metal to be found. There are several other reasons for the price intensification as well. Gold is used in various fields like finance, trade, mechanical. Industrial, dental and medical applications account for approximately 12% of the gold requirement. Gold has high thermal and electrical conductivity properties as well as high resistance to corrosion and bacterial colonization. It has fluctuated in recent years due to the constant expansion of the middle classes in emerging markets seeking western lifestyles. To extract a small amount of gold, a larger amount of gold ore could be used; along with a lot of staff associated with it. If

the ore is of lesser quality, only 5 grams of gold can be extracted from a large ton of ore. In fact, this metal can be ductile, fluffy, and malleable. easily in different ways. It becomes very flexible so that other metals can be added to make useful ornaments, and this valuable product is used for various purposes around the world. It has a decent cathode of power and heat. It also has the ability to keep through any atmosphere; not affected by moisture, air or the most dangerous or destructive elements. It is the best example of reliable materials. In addition, the recycling of used jewelry has grown into a multi-billion-dollar industry.

2. Literature Review

The literature has dealt extensively with the subject of gold due to the high level of interest not only from scientists and scholars, but also from investors and governments seeking to develop existing academic material related to the yellow metal. Several studies have examined the factors that affect gold. On the one hand, price fluctuations and, on the other hand, attempts were made to create predictive models. In fact, various machine learning techniques have been studied to predict the price of gold. Several studies have examined factors that can explain the variation in gold prices. For example, Qian et al. analyze the main factors that determine and influence the gold price.

The experiment was performed with six different variables, namely the Dollar Index U., the Federal Funds Rate U., the Consumer Price Index (CPI), and the U. The exchange rate of the dollar to the Chinese yuan, the price of oil and the S & P 500 index. When considering a significance level of 5%, the test result shows that without taking into account the CPI and the oil price, all of the above-mentioned determinants have a negative impact on the precious metal price. Singhal et al. In Mexico, examine the long-term relationship between the price of West Texas Intermediate (WTI) crude oil, the international spot price of gold, the US dollar to Mexican peso USD / MXN exchange rate, and the Mexican stock exchange "Index of Prices and Quotations (IPC). To assess the stochastic trend of the variables, the study was based on tests of autoregressive distributed delay limits (ARDL). The results show that the price of WTI crude oil has a negative impact on the price of gold, while the latter has a positive impact on Mexico. Assessment of share prices.² Bilgin et al. Analysis of the factors influencing the gold price, with an emphasis on uncertainty measures Four different steps were used in the study: the global economic policy uncertainty index (GEPUI) and the party conflict index (P.), the Skewness Index (SKEW) and the Volatility Index (VIX). The last two indicators mentioned relate to the risks and potential fluctuations in the financial markets.

The results of the study suggest that the price of gold reacts to economic and economic uncertainties politically and can be viewed as a safe haven as the yellow metal negatively correlates with political and economic conditions. Using the same logic, Fang et al. He also relied on the GEPUI to weigh the importance of the data covered by the index in predicting gold rates. The study was based on a generalized autoregressive model for mixed data samples with conditional heteroscedasticity (GARCH-MIDAS). A significant and positive influence on the

volatility of the valuation of gold futures. Not only economic and political factors influence the movement of the gold price, but also social celebrations at which gold gifts are offered. In fact, Schmidbauer and Rösch examine the effects of festivals on the volatility of spot gold prices on a daily basis.

The tests were based on 13 different holiday events around the world, such as Ramadan Eid, Eid Al Adha, Christmas, New Years Eve, and Chinese New Year. You build a model that combines regression with the GARCH model. The results of Schmidbauer and Rösch show that various celebrations correlate with fluctuations in the price of the yellow metal. In addition, several studies have implemented predictive models to predict the future price of gold. or the pattern of price volatility. Al-Dhuraibi and Ali. Like the one who tried to predict gold price fluctuations by implementing machine learning algorithms dedicated to ranking. To identify the future movement of the gold price, regression, decision tree, supervised machine learning (SVM) and K-Nearest Neighbor (KNN). For their analysis, Al-Dhuraibi and Ali relied on a single variable, which is a weekly difference in the price of gold. Because classification algorithms are used on categorical data, the weekly difference in gold price has been converted into two classes, "+1" if the difference is positive (increase in t-movements) and "-1" if the difference is negative (decreasing mobility). Finally, it was found that KNN is the only algorithm that achieves an acceptable accuracy of 60.26% at $K = 5$. Unlike classification models, Kemal used the Multilayer Neural Perceptron Network (MLPNN) to predict the price of gold.

The explanatory variables of the model were the spot rate of the USD, the main stock index of Turkey (BIST100), the price of Brent oil and silver. and copper price and the weekly interest rate of the Central Bank of the Republic of Turkey. Prices and various explanatory variables. In addition, the accuracy of the model reaches 98.17%. Sami and Junjo combined different prediction models. In fact, they use machine learning algorithms such as linear regression, neural networks, and autoregressive moving average (ARMA) to predict the price of gold. Data on banks and leading companies that have invested heavily in precious metals. Studies have shown that, compared with major companies that participate in the gold market through investment or mining, the US economy has little impact on gold interest rates. For investors and policy makers in Pakistan, the economy plays an important role in the decision-making process, Akbar et al. studied the relationship between gold prices and stocks and interest rates. Akbar et al. used an autoregressive Bayesian vector model (Bayesian VAR) to investigate the volatility of gold prices. The results show that gold can be considered a safe haven. Note that in a turbulent environment, the rupee and stock prices fall while the gold price rises. In our research, we will examine the relationship between explanatory variables and the price of gold, and predict the price of gold by implementing various machine learning algorithms. In the next part, we will further study the methods used and a description of the data used to build our predictive model.

3. Methodology

3.1 Gold Investments

Since ancient times, the value of the precious metal has been high in various periods despite the economic and financial crisis. For the past few years, the value of the currency has fluctuated depending on the currency market, the price of crude oil and inflation. The gold rate is also unstable. Investors or customers have a higher risk of investing or buying them. Since gold is viewed as essential to the core and a liquid asset around the world, trading is very easy. Hence, it is widely used as a precious metal. The forecast for gold is of great help in planning and implementing future investment results. Predicting the gold rate is not only intended to give hope to people, but also to protect money in this scenario as its value fluctuates dramatically. Units are traded like stocks, and predicting stocks is always a challenging problem because their non-linearity makes them non-trivial. As a result, shareholders will devote themselves to protecting themselves from political and monetary expansion and social fiasco.

Gold is benevolent because there is no "crop rotation fluctuation in the market". Therefore, along with individuals in multinational corporations, they also invested in gold reserves. In the meantime, this precious metal has become more of a currency, so the government is likely to even increase the gold reserve. Usually people assume that the benefit of gold is that it helps in tough times as there is a high level of liquidity control. According to government policy and gold regulators, the price of gold fluctuates daily. However, estimating the increases and decreases in monthly gold rates can help finance professionals decide when to buy or offer these products. Another payment method is gold, which is used around the world for additional business transactions. In this modern scenario, the central banks of all countries keep expensive metals to ensure the payment of foreign debt, regulate inflation and also reflect financial strength.

Alloy gold is required to have a more remarkable retention area in the placement of ultrasonic aluminum wire connections than ordinary alloy metallization. Understanding the factors affecting gold rates is imperative for any potential gold buyer so that they can more accurately predict interest rate patterns and, consequently, lead a company to a superior profit. Investors, researchers also have a great interest in understanding the gold business with the aim of a concrete study and an accurate forecast.

3.2 Dataset

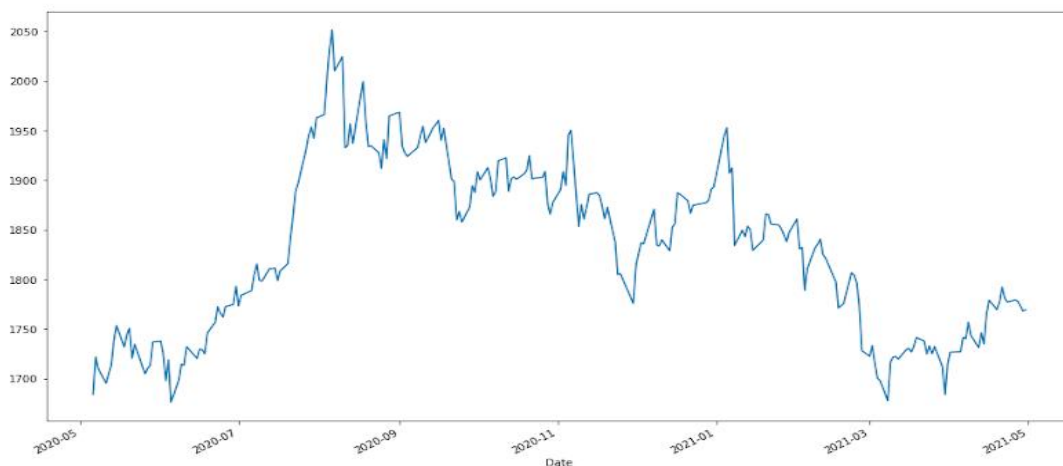
Data for this study is collected from yahoo finance which is a live dataset that collects the data for nearly 300 days from the present day of analysis that we make. Data for attributes, such as date, open and closing stock, high and low price of stock and adj.close and volume of stocks were gathered. Data of many other stocks in addition with gold are scrapped from yahoo

finance. Price of all the stocks during this period is also included in the analysis. Table below lists the attributes of data which are extracted for the past 300 days.

	Date	Open	High	Low	Close	Volume	Dividends	Stock Splits
0	2020-05-06	1704.000000	1704.000000	1684.199951	1684.199951	469	0	0
1	2020-05-07	1686.199951	1721.800049	1685.599976	1721.800049	853	0	0
2	2020-05-08	1726.199951	1728.900024	1703.800049	1709.900024	330	0	0
3	2020-05-11	1702.300049	1704.500000	1690.199951	1695.300049	749	0	0
4	2020-05-12	1699.300049	1713.099976	1699.300049	1704.400024	324	0	0
...
242	2021-04-26	1778.000000	1779.699951	1770.800049	1779.199951	549	0	0
243	2021-04-27	1780.000000	1782.000000	1777.699951	1778.000000	117	0	0
244	2021-04-28	1768.199951	1774.000000	1764.599976	1773.199951	756	0	0
245	2021-04-29	1781.000000	1789.000000	1758.599976	1768.099976	756	0	0
246	2021-04-30	1772.400024	1773.300049	1765.099976	1769.300049	50410	0	0

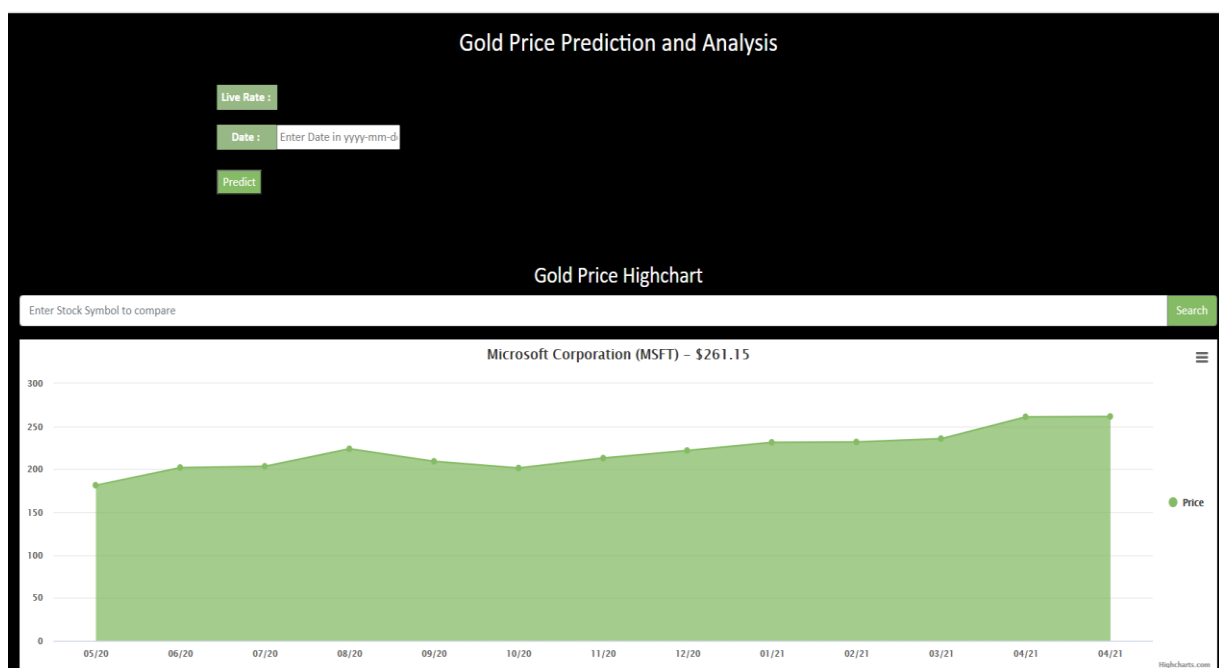
247 rows x 8 columns

The above listed dataset consists of 247 rows and 8 columns. Where 2 columns are used for the prediction of the future price. The two columns which are used are date and the high attributes. The price of gold that is to be predicted is taken in US Dollar. Here we do a lot of cleaning and preprocessing techniques to remove the unwanted attributes. Missing values are also treated to give the good prediction accuracy. Gold prices change day to day and the daily changes have been recorded in the analysis. The current gold rates higher in comparison with the previous years. As we see huge variations in price, we decided to split the dataset in a sequential split instead of random sampling. Therefore, the recent 30% of the data has been used as the test set, and the earliest 70% of data has been used for training. There is a major fluctuation in gold prices over the past years, so recent historical data would be more indicative of the future trend. This graph gives the trend of gold prices in the past 300 days.



3.3 User Interface creation

In order to make the analysis and prediction more interactive to the people, a user interface has been built. This user interface has been created with the help of html, css and javascript. Here the user will be able to see the live chart of the rate of each commodity. Javascript is used to display the high chart of the user interface. The search column allows the user to search for the ticker symbol. A ticker symbol or stock symbol is an abbreviated form for uniquely identifying publicly traded shares of a stock on a stock market. A stock symbol may be a combination of letters, numbers or both. The stock symbol for gold is GC = F. By searching this symbol it shows the trend and the past history of the following searched stock as a map. This search will get the history of prices of the stock for the past 300 days. Here the index will be reset and the date acts as an index later. As this chart consists of data and high value in the x and y axis respectively, dropping off the other attributes to give the trend in a clear manner. The price here is used to represent in USD. So it is easy for all the users around the world to use the same conviction. The chart is designed in such a way that, users can use to compare with any other stocks with the help of its symbol. This user interface has been deployed using the flask.



3.4 Machine Learning Implementation

We get the input in the form of tickle data and fit the models and return back the prediction result in the form of tickle, so there will be no change in the prediction and the accuracy.

Splitting of training and testing data

With the history of the dataset, the training set has been divided into 70% and the testing set has been divided into 30% in order to give more accuracy. The following are the machine learning models that are used to predict almost the exact value in the future.

Linear regression

Supervised machine learning algorithms define models that capture the relationships among data.

For example, we analyse the high prices of each stock and try to establish a dependence on the features or variables. The set of data related to a single stock is one observation. The features or variables can be of either of the two forms:

1. Independent variables, also called inputs or predictors, don't depend on other features of interest (or at least you assume so for the purpose of the analysis).
2. Dependent variables, also called outputs or responses, depend on the independent variables.

In the above where we're analyzing each stock, we might presume the rate of high, as being mutually independent, and consider them as the inputs. The rate could be the outputs that depend on the inputs.

Decision Tree

Decision tree classifiers try to form nodes in which they contain a high proportion of data points from a particular or single class by finding the values in features which divide the data into classes. It can be considered as a nonlinear model which is built by many linear boundaries. In this model we give both labels and features, and it classifies the points based on features. The accuracy has gone down due to overfitting.

Random forest model

Random forest is a bagging model that builds multiple decision trees and merges them together to get a more accuracy in prediction. Random forest has nearly the same hyperparameters as a decision tree classifier it is only a bagging classifier. Random forest adds additional randomness to the model that we fit the training data, while growing the trees. In Accordance with our dataset, the random forest model gives the best accuracy. For the application in the stock market, the Random Forest algorithm can be used to identify a stock's behaviour and the expected loss or profit can be forecasted from the prediction model.

4. Results and Discussion

With the help of the random forest classifier, we could get almost the exact predicted value. For example, here the input date is given as 2021/11/26. The predicted output by the random

forest model is given as \$1780.94. This prediction estimation will give the stock market users and even normal customers to easily use this interface.

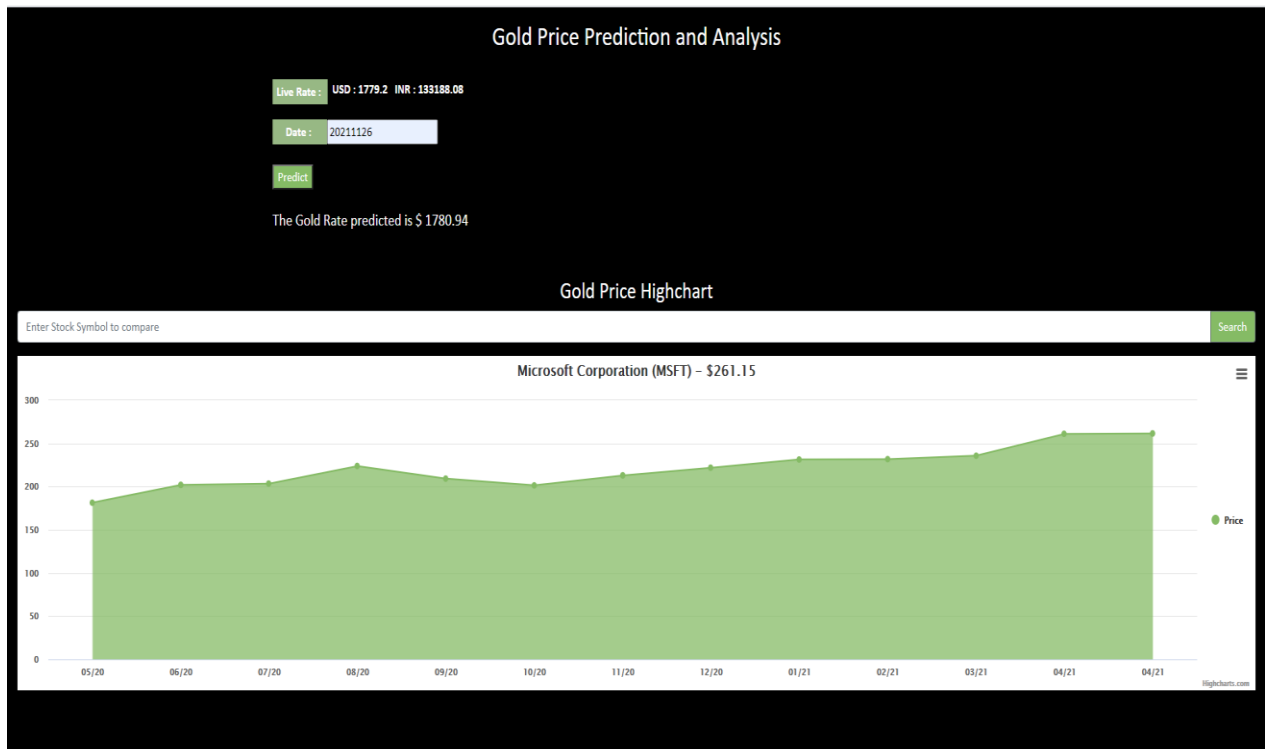


Fig. 1 User Interface – gold price chart

To check the accuracy of the predictor, we gave the input date as 2021/04/24, on the same day, so that the values can be compared. The predicted value that we got was \$1776.52 and the actual value was \$1777.0. This confirms that the accuracy and the precision of the model is quite high, as the predicted and actual values are coinciding to a great extent. The following image is the output of our result.

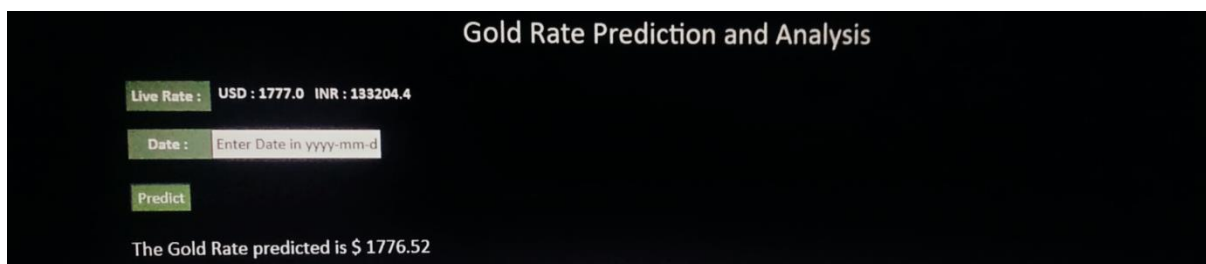


Fig 2. Gold Rate Prediction and Analysis

We've also predicted the gold rates for the following month of May. With the help of these values, the regular actual values are to be compared and tested for its acuteness.

5. Conclusion

As gold proves to be a viable source of investment, the investment opportunities has been expanded to huge numbers, and it arose a need for predicting the future highs and lows that the commodity(gold) might hit. By testing with different machine learning prediction models such as linear regression, decision tree regression, random forest regression, it has been established that random forest comes out with better accuracy in predicting future gold rates. By building a web application that could possibly merit people who are interested with investing in gold, it could benefit more number of people worldwide as it has been hosted in heroku as well. Investors must also include technical and fundamental analysis of the commodity to arrive at better decisions.

6. Future Enhancements

Our first step would be updating our web application by including other ornamental share prices and enhancing our app with better UI, so as to make it more user friendly for the users. We've also planned to include various share prices of Nifty fifty companies' data and their historical behaviour. Incorporating blogging features for the traders and the users to share their experiences and ideas would be the next step. And to reach out to a greater number of audience, we've planned to launch it as an app in playstore.

Reference

- [1] D. G. Navin, "Big Data Analytics for Gold Price Forecasting Based on Decision Tree Algorithm and Support Vector," *International Journal of Science and Research*, pp. 2026-2030, 2015.
- [2] V. K. F. B. Rebecca Davis, "Modeling and Forecasting of Gold Prices on Financia Markets," *American International Journal of Contemporary Research*, pp. Vol 4, No 3, 2014.
- [3] M. S. S. S. S. Kumar Chandar, "Forecasting Gold Prices Based on Extreme Learning Machine," *International Journal of Computers Communications & Control*, pp. 372-380, 2016.
- [4] Khan Ali, Forecasting of gold prices (Box Jenkins Approach). *International Journal of Emerging Technology and Advanced Engineering*, 662 -668 ,2013
- [5] Shafiee,Shahriar and Erkan Topal. An overview of global gold market and gold price forecasting. *Resources Policy* .Volume:35, 2010