

## GOLD PRICE PREDICTION

**Anisha Nayak, Sonali Sahoo**, Computer Science & Engineering, Gandhi Institute For Technology, Odisha, India anisha.nayak2020@gift.edu.in , sahoos2021@gift.edu.in

### *Abstract*

We want to predict gold price using python which involves a model and need to be forecast. Data collection is needed by using csv file which should show different stages gold price opening, closing price etc. Then by kdd process where data will be transferred processed, cleaned, model selection will happen. Training to the model needed by using python, pandas, NumPy, matplotlib so gold price can be predicted.

We predict future gold rates supported twenty two market variables victimization machine learning technique. One machine learning algorithm, random forest regression were used in analyzing these knowledge.

Historically, gold was used for supporting trade transactions around the world besides alternative modes of payment. Various states maintained and increased their gold reserves and were recognized as rich and progressive states.

In present times, precious metals like gold area unit control with central banks of all countries to make sure re-payment of foreign debts, and conjointly to control inflation. Moreover, it conjointly reflects the money strength of the country[2]. Besides government agencies, varied transnational firms and people have conjointly invested with in gold reserves. In ancient events of Asian countries, gold is in addition presented as gifts/souvenirs and in marriages, gold ornaments are conferred as gift in Republic of India.

### **I. INTRODUCTION**

A country is developed by many factors. Our reserve bank contains huge amount of many as well as gold if the gold is more with any country, then that country has good ranking in the world. Then gold is used when we are using foreign transaction as well as gold is used by domestic consumer of the country. In one side gold is needed to interact with foreign countries and at the same time gold is used by internal customers so whatever the gold stock is there with RBI it can't be float in the market as per wish. There is a balanced required between how much it can send to market for purchase. When inflation increases by any other factors follow price also increase so stock becomes less and market price goes high, so other comes the tracking for gold price that is needed.

Investing in gold has developed over a amount of your time in conventional forms by buying jewellery or through modern strategies, either by buying gold coins and bars (which area unit already accessible in scheduled banks). Historically, gold had been used as a sort

of currency in various components of the planet as well as USA[3]. In recent times also, gold has maintained its worth and has been used as a means for assessing the monetary strength of a rustic. Big investors have conjointly been interested in this valuable and invested vast amounts in it. Recently, rising world economies, like China, Russia, and

India are massive buyers of gold, whereas USA, South Africa, and Australia area unit among the large vender of this goods. Chinese and Indian traditional events conjointly have an effect on the worth of the gold. in this time more money is poured for purchase of this goods[4]. Small investors conjointly realize this goods for safe investment rather than alternate investment choices, that bear in-built investment risks. Internal monetary conditions of the aforementioned countries play an important role for setting spot rates for gold. Gold is another plus that is being thought-about as a lovely investment avenue by several investors thanks to its increasing worth and therefore the space of usage[15]. Investors preference for gold as a protecting plus will increase thanks to their negative expectations regarding things within the developed interchange markets and therefore the capital markets. Gold is additionally thought-about to be "the plus of ultimate instance"

i.e. is that the plus investors place confidence in, once the developed world capital markets aren't capable to supply fascinating profitability. so it is same that investors see gold as a tool to hedge against the fluctuations in alternative markets. Gold could be a valuable, thus like all alternative.

## 2. purpose

**Data Collection:** The first step is to gather historical data on gold prices. There are various sources available for this, including financial websites, APIs, or datasets from platforms like Yahoo Finance, Quandl, or the Federal Reserve Economic Data (FRED).

**Data Preprocessing:** Once you have the data, you need to preprocess it. This involves handling missing values, removing outliers, and transforming the data into a format suitable for analysis. You may also need to resample the data to a specific frequency (e.g., daily, weekly, monthly) depending on your analysis.

**Feature Engineering:** Feature engineering involves selecting or creating relevant features that can help predict gold prices. This might include technical indicators (e.g., moving averages, relative strength index), economic indicators (e.g., inflation rates, interest rates), or other factors that may influence gold prices (e.g., geopolitical events, stock market performance).

**Model Selection:** Choose a suitable machine learning model for prediction. Common models used for time series forecasting include linear regression, ARIMA (AutoRegressive Integrated Moving Average), SARIMA (Seasonal ARIMA), LSTM (Long Short-Term Memory) networks, and Prophet (developed by Facebook).

**Training the Model:** Split the data into training and testing sets. Use the training data to train the model on historical patterns and relationships between features and gold prices.

**Model Evaluation:** Evaluate the performance of the trained model using the testing data. Common evaluation metrics for regression tasks include Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

**Prediction:** Once you have a trained and evaluated model, you can use it to make predictions on future gold prices. Depending on the model used, this may involve feeding in new data and obtaining forecasts for future time points.

**Visualization:** Visualize the predicted gold prices along with actual prices to understand how well the model is performing and to communicate the results effectively.

## 3.SCOPE

The scope of gold price prediction using Python involves collecting historical data, preprocessing it, and conducting exploratory data analysis. Feature engineering extracts relevant factors like economic indicators or sentiment analysis. Various machine learning algorithms, including linear regression and neural networks, are employed for prediction. Model performance is evaluated using metrics like Mean Absolute Error. Deployment as a web service or integration into applications is optional. Continuous improvement entails monitoring and periodic retraining with new data. The scope encompasses a comprehensive process from data collection to model deployment, aiming to accurately forecast gold prices in response to market dynamics.

## 4. MOTIVATION

People want to guess what gold will cost in the future. It's like looking at clues in a treasure hunt to guess where the treasure is buried. Using Python, we can look at lots of data about things like the economy and world events to make smarter guesses about gold prices. If we can predict gold prices accurately, it helps investors make better choices about buying and selling gold. It's kind of like having a crystal ball for the gold market, helping everyone from regular folks to big businesses make decisions that could save or make them a lot of money.

“Gold price prediction” is to forecast gold’s price using a variety of techniques , considering the relationship between several economic factors that influence gold rates.

## 5. OVERVIEW

Gold price prediction using Python is like solving a puzzle. First, we gather lots of information about gold prices and other important stuff like the economy. Then, we organize this data to find patterns and clues. Next, we use special tools in Python to make guesses about what gold prices might be in the future. We check how good our guesses are by comparing them to real gold prices. If our guesses

are good, it can help people make smarter decisions about buying and selling gold. It's all about using technology to understand and predict what might happen in the gold market.

## **6.PROBLEM STATEMENT**

From the above comprehensive literature review on different prediction methods in business applications utilizing analytical intelligent techniques throughout the past decades, it has been noted that the following problems are found when carrying out prediction processes for the business applications considered price prediction, gold price prediction:

- 1)Unattained Scalability
- 2)Premature Network Convergence model
- 3)Trapping ourselves in local and global optima
- 4)Stilling
- 5)Large prognostic bias
- 6)Overlapping computing energy
- 7)The computational load of the algorithms increased
- 8)No assurance on system's interpretability

## **7. OBJECTIVES**

The objectives of predicting gold prices with Python are to accurately forecast future prices, help investors make smart choices, and manage risks effectively. By analyzing historical data and relevant factors like economic indicators and geopolitical events, we aim to create models that offer practical insights for buying, selling, or holding gold assets. Continuous monitoring and updating of these models ensure they adapt to changing market conditions, giving stakeholders reliable tools to navigate the gold market confidently. Ultimately, it's about using technology to improve decision-making and maximize returns in gold investments, making it accessible and understandable for all involved.

## **LITERATURE SURVEY**

There are many studies dealing with the price of gold in the literature. Although various different variables are used in these studies, it is observed that gold prices are regressed against USA dollar and stock return in general. The relationship between other macroeconomic variable and gold prices has also been studied by many researchers. The relationship between gold price and prices of other commodities especially crude oil has also been extensively studied. But the results from these studies are found to be contradicting. Some of the studies on the factors influencing gold price and various techniques used for studying these relationships are discussed in the following sections. Lawrence has found that there is no significant correlations between returns on gold and changes in certain macroeconomic variables such as inflation and GDP.

He has also found that that gold returns are less correlated with returns on equity and bond indices than returns of other commodities. But, Sjaastad and Scacciavillani reported that gold is a store of value against inflation and Baker and Van Tassel also have found that the price of gold depends on the future inflation rate. With respect to the relationship between gold price and inflation, based on the review of literature Hanan Naser is of the opinion that historical studies with regards to the effectiveness of gold as a hedge against inflation are contradicting. Ismail have forecasted gold prices based on multiple economic factors such as commodity research bureau future index, USD/Euro foreign exchange rate, inflation rate, money supply, New York Stock Exchange Index; Standard and Poor 500 index, Treasury bill and USD index. The study finds that Commodity Research Bureau future index, USD/Euro foreign exchange rate, Inflation rate and money supply have a significant impact on gold price. Khaemusunun has examined the impact of currencies of selected countries, Oil Prices and Interest Rate on the gold price. Hammoudeh conclude that there is an interdependent exist between the volatility of gold price and the exchange rate. Ai, et al. report empirical evidence that the exchange rate relates to the gold price both in the long-run and short-run. Ewing and Malik find evidence of volatility transmission between gold and oil future prices. Ghosh et al. have concluded that gold prices are related with US Inflation level, interest rates and dollar exchange rate. They have also reported a long run relationship

between gold prices and US Consumer Price Index as a result of the cointegration analysis. From the review of related literature, it can be concluded that the relationship between gold price and various factors considered to influence it are contradicting. In studying volatility in gold price and the relationship with the factors considered to influence it, researchers have used a variety of techniques. Hossein and Abdolreza have predicted the gold price by using artificial neural networks (ANN) and ARIMA been conducted using multivariate regression models to test the sensitivity of gold prices among various variables. In this regard Ismail et al have used multiple linear regression (MLR) model for forecasting the gold prices and are of the opinion that MLR model appeared to be useful for predicting the gold price. From the review of literature, it can be seen that multiple linear regression is widely used technique for understanding relationship among such variables.

Steps:-

- 1) Importing libraries
- 2) Data Understanding
- 3) Data Pre-processing & EDA
- 4) Checking Missing Values
- 5) Data Visualization
- 6) Statistical Measures (Mean, Standard deviation, Kurtosis)
- 7) Model Building

Imported Libraries:-

```
import pandas as pd
import numpy as np
%matplotlib inline
import matplotlib.pyplot as plt
import matplotlib from sklearn.preprocessing
import MinMaxScaler from sklearn.ensemble
import RandomForestRegressor from matplotlib.pyplot
import figure
import seaborn as sns
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import TimeSeriesSplit
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.dates as mdates
from sklearn import linear_model
from sklearn.model_selection import TimeSeriesSplit
from sklearn.svm import SVR
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

Data Understanding:-

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2904 entries, 0 to 2903
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype
---  ---
 0   Date        2904 non-null  object
 1   Open        2904 non-null  float64
 2   High        2904 non-null  float64
 3   Low         2904 non-null  float64
 4   Close       2904 non-null  float64
 5   Adj Close   2904 non-null  float64
 6   Volume      2904 non-null  int64
dtypes: float64(5), int64(1), object(1)
memory usage: 158.9+ KB
```

Data Pre-processing And EDA:

	Date	Open	High	Low	Close	Volume	Adj Close
0	6/29/2010	19.000000	25.00	17.540001	23.889999	18766300	23.889999
1	6/30/2010	25.790001	30.42	23.299999	23.830000	17187100	23.830000
2	7/1/2010	25.000000	25.92	20.270000	21.959999	8218800	21.959999
3	7/2/2010	23.000000	23.10	18.709999	19.200001	5139800	19.200001
4	7/6/2010	20.000000	20.00	15.830000	16.110001	6866900	16.110001

Data Visualization:

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

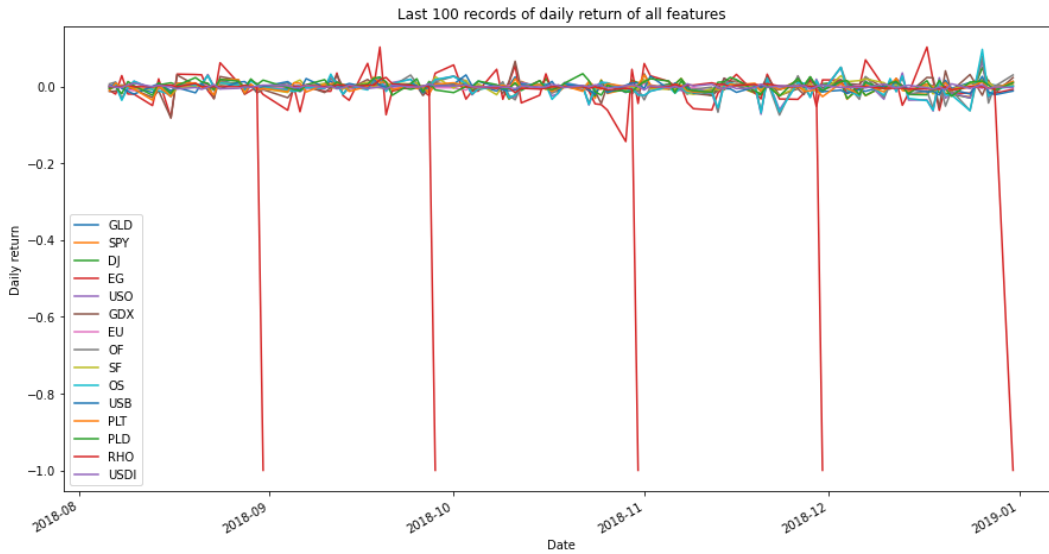


Fig1:Gold Price By Months

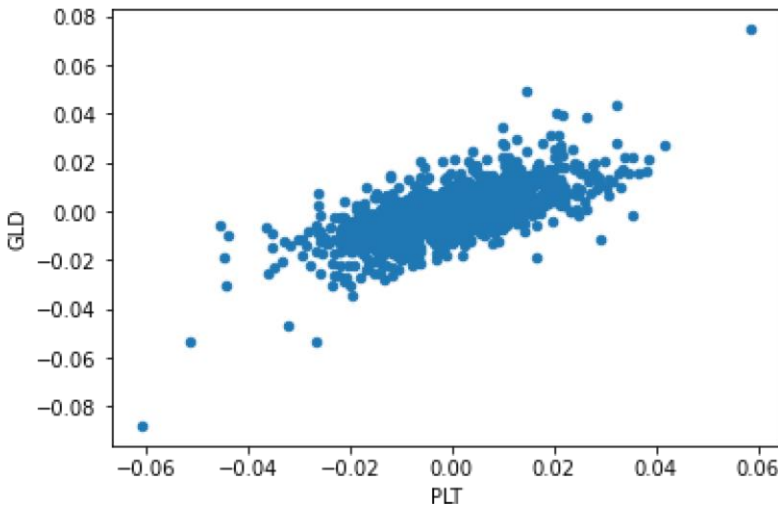


Fig2:Gold Prices Lag Plot

Plotting of Mean, Standard deviation and Kurtosis of Gold Prices

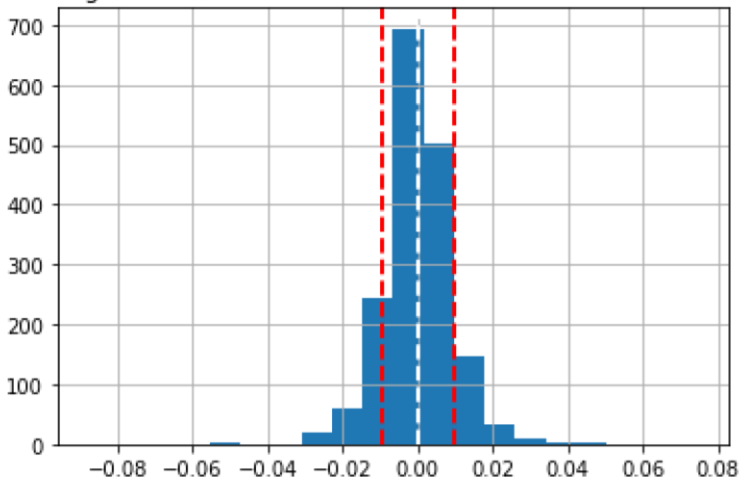


Fig3: Mean , Standard deviation and Kurtosis of Gold Prices

### Model Building:-

The model building process involves setting up ways of collecting data, understanding and paying attention to what is important in the data to answer the questions you are asking, finding a statistical, mathematical or a simulation model to gain understanding and make predictions.

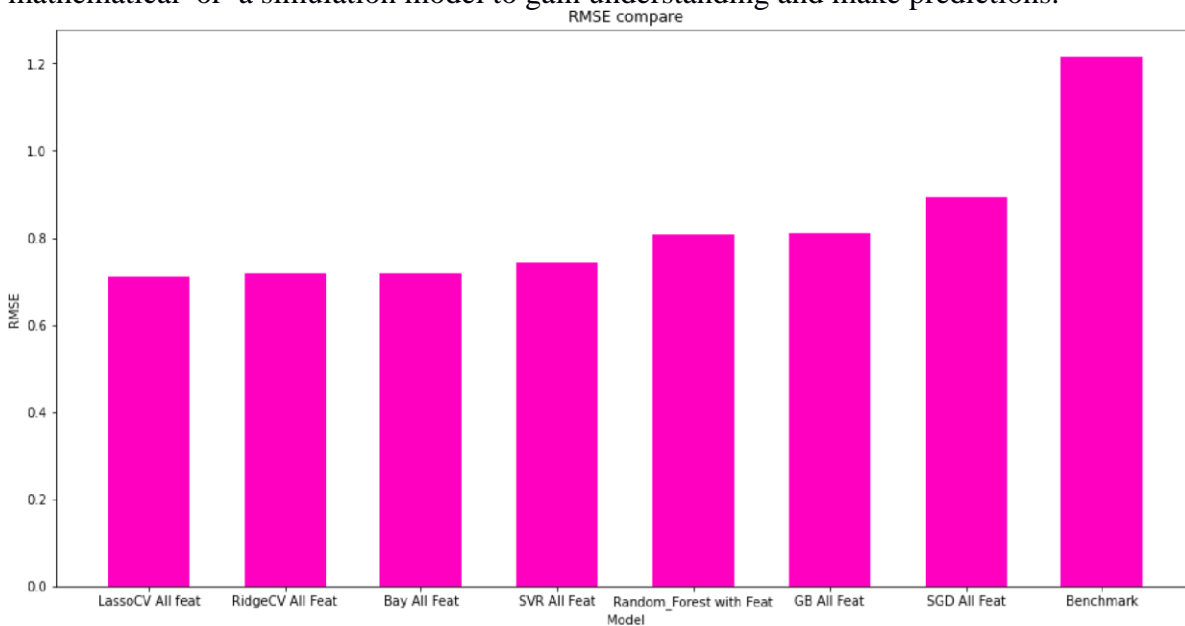


Fig4: RMSE Model

### METHODOLOGY

model. Khaemusunun, predicts the Thai gold price by using Multiple Regression and ARIMA model. Toramanhas reported that various studies have

### METHODOLOGY

Predicting gold prices typically involves collecting historical data on various factors that influence gold prices, such as economic indicators, geopolitical events, currency movements and supply and demands dynamics. Machine learning algorithms are then trained on this data to identify patterns and relationships that can be used to forecast future gold prices. Techniques such as regression analysis, time series analysis, and sentiment analysis may be employed to develop predictive models. It's also crucial to continuously update the models and incorporate new data to improve their accuracy over time

Based on the review of literature five major factors that is considered to have influence on the gold price were identified. The factor that is considered for this study are Historical Gold Prices Values of India.

Machine learning algorithms were used to train and model the collected data. From the data collected, eighty percentage of the data was used for training and remaining twenty percentages for testing the model. The machine learning algorithms used in this study are linear regression, random forest regression and LSTM Model.

The statistical process for estimating the relationship between different variables is called regression analysis. Regression analysis is used to understand how the value of the dependent variable changes when one of the independent variables changes, while other variables are fixed.





A Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap Aggregation, commonly known as bagging. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees. The Random Forest uses bootstrapping on Decision Trees to reduce the variance while maintaining the low bias that is resulted from a 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. The same random forest algorithm can be used for both classification and regression task. And, the random forest algorithm can be used for feature engineering for identifying the most important features out of the available features from the training dataset.

The Long Short-Term Memory network is a RNN that is trained using Backpropagation. It takes care of the disappearing gradient problem encountered earlier. LSTM networks have their own memory and so they prove to be efficient in creating large RNNs and handle time specific scheduling problems. The memory blocks in LSTM network are connected through recurrent layers rather than having neurons.

A block has many basic and a few complex components that make it smarter as compared to the standard neuron. It consists of many gates that coordinate relative input functions with output functions. Whenever a block receives an input, a gate is triggered which takes decision about whether or not to pass the block forward for further processing. The standard LSTM block, in its simplest form, consists of an input gate, an output gate, a cell and a forget gate.

Cell: It is used to remember the values over arbitrary time intervals.

Input Gate: It decides which information to keep in the cell.

Output Gate: It is used to decide which part of cell state should be given as an output. Forget Gate: It is used to decide which information to throw away from the cell.

In the Line Redistribution model, the calculation line calculation is used to combine a set of input data values ( $x$ ) into a predicted output data set of input values ( $y$ ). Both the input and output variables and values are considered integers. The unique number given by the Line Rotation equation is represented using the Greek capital letter Beta ( $B$ ) and is commonly known as a coefficient.

In addition to this, another coefficient is added to give the line additional degrees of freedom. This additional term is often referred to as the bias coefficient. Typically, the bias coefficient is calculated or otherwise measured by finding the distance of our mathematical points from the most relevant line. This can be displayed as a straight line at right angles to the vertex and calculated using the line bias. Statistically, a line tangent is used to measure its proximity to the relative linear Regression.

A problem model model in Linear Regression will be provided as follows:  $y = B_0 + B_1 * x + E$   
 This same line is also called a plane or plane when we are dealing with more than one input. This is often the case with high-volume data. The Linear Regression model is therefore represented by the mathematical and introverted values measured by the specific coefficients. However, before using this line number, we are faced with a number of issues. These issues often increase the complexity of the model which makes accurate estimates difficult. This complexity is often discussed in terms of the number of dependent and independent factors.

**Data Set:**

Data for this study is collected from (kaggle.com) Dec 2011 to September 2016 from numerous sources. information for attributes, such as Oil worth, NYSE, normal and Poor's (S&P) five hundred index, US Bond rates (10 years), EuroUSD exchange rates were gathered[8]. Data of the many government central banks and 5 giant companies that have invested with Brobdingnagian amounts in gold have conjointly been collected. worth of precious metals throughout this era is also enclosed within the analysis. Table I lists the net sources from that this information was extracted[9]. Table II lists of these attributes. The price of gold that we tend to try to predict is taken in US Dollar. plenty of cleansing and preprocessing was performed on the dataset. the matter of missing values was handled in appropriate manner to finish the dataset. Gold costs amendment on everyday and are full of major world events. Current gold rates ar a lot of beyond a few years pas

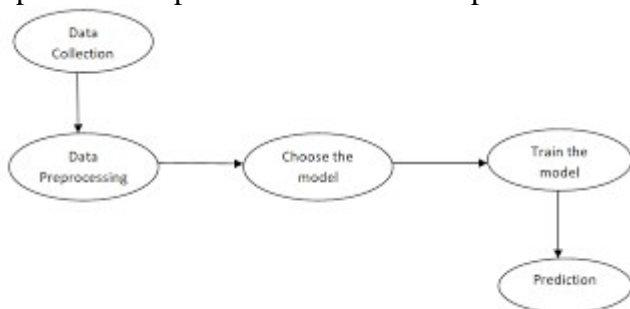
	Date	Gold	Return_22	Gold-T+22	Date-T+22
84	2020-04-13	1769.4	0.0406	1841.2	2020-05-05
85	2020-04-14	1751.3	0.0188	1784.2	2020-05-06
86	2020-04-15	1743.3	-0.0068	1731.4	2020-05-07
87	2020-04-16	1730.1	-0.0096	1713.5	2020-05-08
88	2020-04-17	1694.5	-0.0233	1655.0	2020-05-09

**Figure1: Data Sample**

**Machine Learning Models:-**

We use 2 milliliter models, particularly Random Forest, and linear regression. Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems.

In addition to the input and output layers, they carries with it one or more hidden layers of neurons that try and learn non-linear decision boundaries that separate totally different categories of information[10]. It can also be accustomed predict continuous valued attributes like gold costs in our case. Fig. four depicts a sample ANN. Linear regression (LR) is AN approach utilized in statistics to model relationship between dependent (class variable) and one or additional freelance variables (attributes). regression can be used for predicting continuous valued attribute. We use the implementation of LR and ANN that's provided by RapidMiner tool[11]. each models area unit optimized exploitation the RMSE performance live.



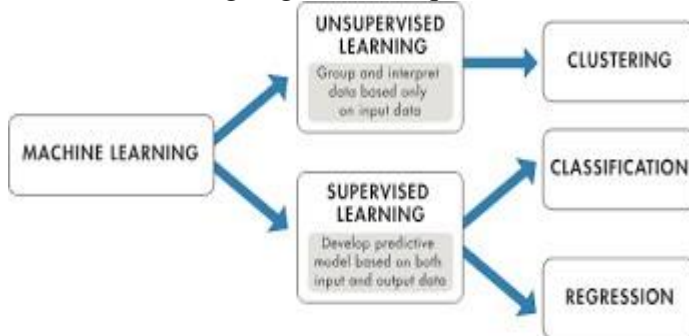
The specialists then collect, pick, plan, pre-process and convert this information. Upon completion of this stage, the specialists begin building prognosticative models[12]. A model that



predicts costs at the utmost preciseness rate would be chosen to drive a tool or program. and therefore the worth prediction perform system would possibly appear as if this:

- 1)Statement on drawback.
- 2)Knowing the peculiarities of economies. respondent the question: What variables square measure poignant artefact / product / service prices?
- 3)Gathering, preparing, and preprocessing information.
- 4)Testing and modelling.
- 5)Deploying a model into a code or application

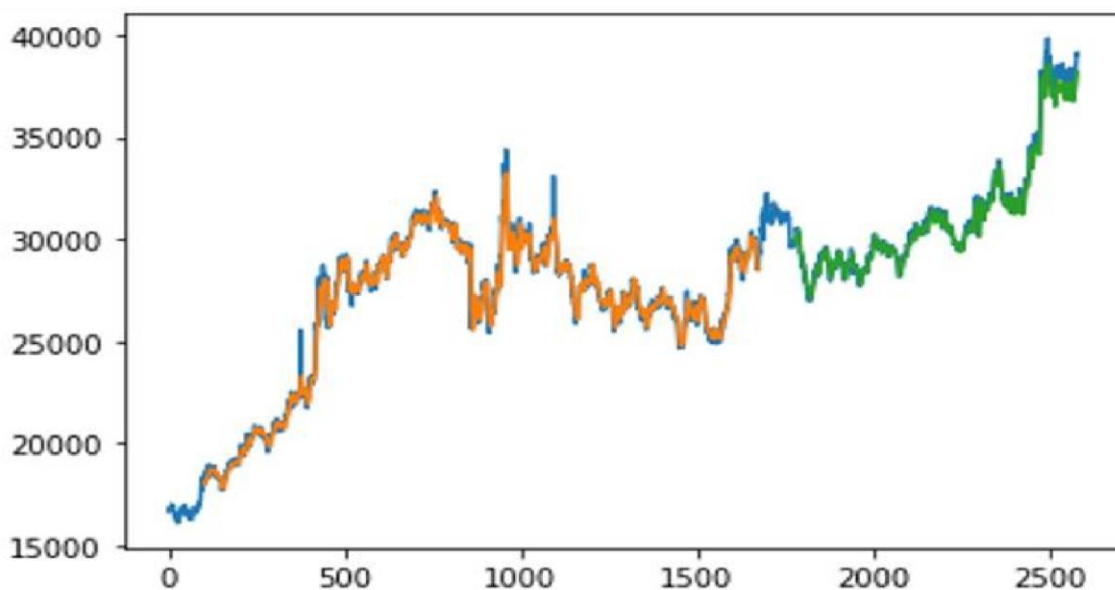
Machine learning algorithms square measure oftentimes classified as supervised or unsupervised

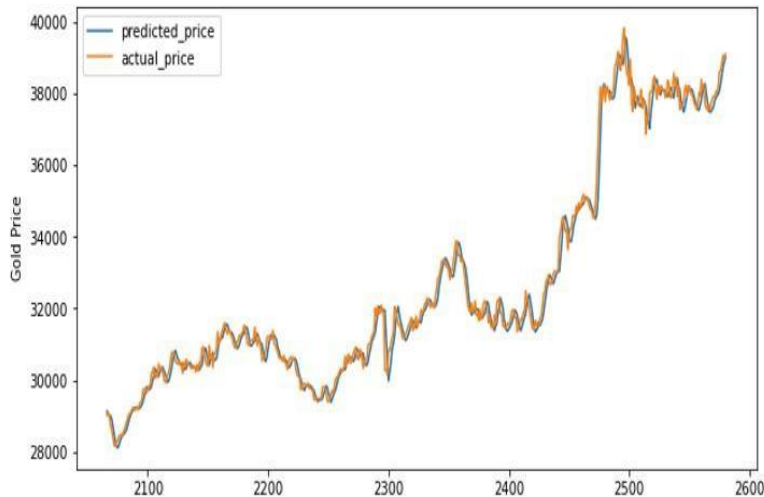


## RESULT & DISCUSSION

The results of a gold price prediction project can vary depending on factors such as the quality of data, choice of models, feature engineering techniques, and market conditions. The prediction model may achieve certain accuracy metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE) when compared to actual gold prices. Visualization, real-world testing, compare with baselines, robustness testing, ultimately, the success of the project may be evaluated based on its business impact, such as its ability to assist traders, investors, or financial institution making informed decisions.

Interpretation of the results and key findings. Analysis of factors influencing gold prices and their impact on the prediction model. Assessment of the model's strengths and limitations. Discussion on the practical implications of the prediction model for traders, investors, or financial institutions. Consideration of ethical or societal implications, if applicable.





## CONCLUSION

Gold has been one of history's most significant commodities. Maintaining central banks' gold reserves is essential to maintaining the world's existing economic system. Some big firms and investors are now spending large amounts of money in gold. While forecasting the rate of gold is not very easy, it will allow investors and central banks to determine better when to sell and buy them and thus maximize their income. Furthermore, an attempt has been made in this study by using machine learning algorithms to accurately predict the gold prices and when to sell them and purchase them. This research was done in order to clarify the gold ETF price predictions using machine learning using Python. The research was carried out for data between January 2005 and August 2020. The results on proposed model is as per the following: □ It is concluded that machine learning algorithms with linear regression analysis are very useful in gold price prediction. □ It is concluded that, the model's R-square is 98.76 percent. R-squared is usually 0 to 100 per cent. A score close to 100 per cent indicates that the Gold ETF prices are well explained by the given model. □ It is concluded that, the model's Sharpe proportion is 0.84495 The Sharpe proportion which is a measure for figuring hazard balanced return. The Sharpe proportion is the normal return per unit of vulnerability increased over hazard free expense. □ Results show that proposed linear regression method machine learning beats customary and current predicting models. □ The new estimating model updated on the gold cost dataset, and the results showed that the forecast model is beating other benchmark models, such as ARIMA, ANN and ANFIS.

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.

## REFERENCES

1. Xiaohui Yang, "The Prediction of Gold Price Using ARIMA Model", 2nd International Conference on Social Science, Public Health and Education 2019.
2. Manjula K. A., Karthikeyan P, "Gold Price Prediction using Ensemble based Machine Learning Techniques", Third International Conference on Trends in Electronics an Informatics, 2019.
3. Mrs. B. Kishori 1, V. Preethi, "Gold Price forecasting using ARIMA Model", International Journal of Research, 2018.

4. R. Hafezi\* , A. N. Akhavan, "Forecasting Gold Price Changes: Application of an Equipped Artificial Neural Network", AUT Journal of Modeling and Simulation, 2018.
5. Xiaohui Yang, "The Prediction of Gold Price Using ARIMA Model", 2nd International Conference on Social Science, Public Health and Education (SSPHE 2018).
6. Shian-Chang Huang and Cheng-Feng Wu, Energy Commodity Price Forecasting with Deep Multiple Kernel Learning, MDPI Journal, 2018.
7. Wedad Ahmed Al-Dhuraibi and Jauhar Ali, "Using Classification Techniques to Predict Gold Price Movement", 4th International Conference on Computer and Technology Applications, 2018.
8. Iftikharul Sami and KhurumNazirJunejo, "Predicting Future Gold Rates using Machine Learning Approach", International Journal of Advanced Computer Science and Applications, 2017.
9. NalinipravaTripathy, "Forecasting Gold Price with Auto Regressive Integrated Moving Average Model", International Journal of Economics and Financial Issues, 2017.
10. 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.
11. 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.
12. S. Kumar Chandar, M. Sumathi and S. N. Sivanadam, "Forecasting Gold Prices Based on Extreme Learning Machine", International Journal of Computers Communications & Control, 2016.
12. International Journal of Trend in Scientific Research and Development (IJTSRD) @ [www.ijtsrd.com](http://www.ijtsrd.com) eISSN: 2456-6470 @ IJTSRD | Unique Paper ID – IJTSRD33143 | Volume – 4 | Issue – 5 | July-August 2020 Page 1456