Gold Price Prediction using Machine Learning

Article *in* International Journal of Environmental Engineering · August 2022

DOI: 10.55041/IJSREM15027

5 authors, including:

Jyoti Kharade
Institute of Management and Information Technology
34 PUBLICATIONS **7** CITATIONS

SEE PROFILE

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Gold Price Prediction using Machine Learning

Rushikesh Ghule, Abhijeet Gadhave
Under the Guidance of Prof.Manish Dubey, Dr.Jyoti Kharade
Bharati Vidyapeeth's Institute of Management & Information Technology
Navi Mumbai

1. Abstract:-

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[1]. 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 Imoney 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.

KEYWORDS: Price prediction, Machine Learning, Supervised Learning, Linear Regression, Python , Power Bi , Tableau.

2. Introduction:-

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[5]. Gold could be a valuable, thus like all alternative merchandise, gold’s value ought to rely on offer and demand.

3. Literature Review:-

Gold is that the alone and that retains its worth even through the political and economic[14].

downturn. The gold values unit of measurement usually directly connected to different resources. Future gold worth forecast is

that the investors' alert mechanism thanks to unpredictable market risk. Therefore,precise prognostication of gold prices is needed to predict the market patterns[6]. several machine intelligent techniques for gold prognostication applications square measure noted over the past decade. the link between gold value and costs of different commodities particularly petroleum has conjointly been extensively studied. however the results from these studies square measure found to be contradicting. a number of the studies on the factors influencing gold value and numerous techniques used for learning these relationships[7].

Some of the research papers on Price prediction are listed below:

Stock Closing Price Pridiction :

This research based on accurate prediction of stock market returns is a very challenging task due to volatile and non-linear nature of the financial stock markets. With the introduction of artificial intelligence and increased computational capabilities, programmed methods of prediction have proved to be more efficient in predicting stock prices. In this work, Artificial Neural Network and Random Forest techniques have been utilized for predicting the next day closing price for five companies belonging to different sectors of operation

Bitcoin Price Prediction :

In this paper, we use the LSTM version of Recurrent Neural Networks, pricing for Bitcoin. To develop a better understanding of its price influence and a common view of this good invention, we first give a brief overview of Bitcoin again economics

House Price Prediction:

This research is carried out to analyze the relevant attributes and the most efficient models to forecast the house prices. The findings of this analysis verified the use of the Artificial Neural Network, Support Vector Regression and XGBoost as the most efficient models compared to others. Moreover, our findings also suggest that locational attributes and structural attributes are prominent factors in predicting house prices. This study will be of tremendous benefit, especially to housing developers and researchers, to ascertain the most significant attributes to determine house prices and to acknowledge the best machine learning model to be used to conduct a study in this field.

Used Car Price Prediction:
This research based on prediction price of used cars by using Machine Learning Algorithms such as Lasso Regression, Multiple Regression and Regression trees, we will try to develop a statistical model which will be able to predict the price of a used car, based on previous consumer data and a given set of features. We will also be comparing the prediction accuracy of these models to determine the optimal one.

4. Dataset:-

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 past.

Currency in USD

Date	Open	High	Low	Close*	Adj Close**	Volume
Mar 18, 2022	1,943.90	1,946.20	1,918.00	1,921.50	1,921.50	144,678
Mar 18, 2022	1,934.50	1,939.30	1,928.20	1,928.20	1,928.20	39
Mar 17, 2022	1,933.90	1,947.20	1,933.90	1,942.10	1,942.10	39
Mar 16, 2022	1,922.40	1,922.40	1,907.20	1,908.00	1,908.00	104
Mar 15, 2022	1,939.80	1,939.90	1,910.70	1,928.50	1,928.50	71
Mar 14, 2022	1,976.50	1,977.70	1,959.60	1,959.60	1,959.60	145
Mar 11, 2022	1,985.00	1,996.00	1,959.90	1,982.70	1,982.70	865
Mar 10, 2022	1,978.10	2,010.50	1,978.10	1,998.10	1,998.10	670
Mar 09, 2022	2,053.60	2,063.00	1,980.80	1,985.90	1,985.90	1,022

Figure-1.Data Sample

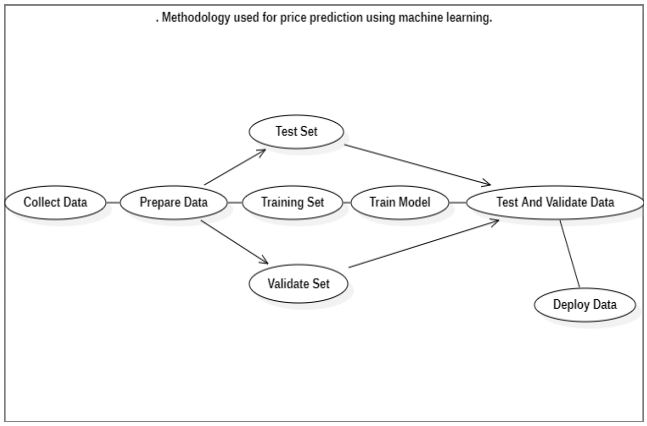

Figure-2.Methodology used for price prediction using machine learning

5. 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[13]. 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.

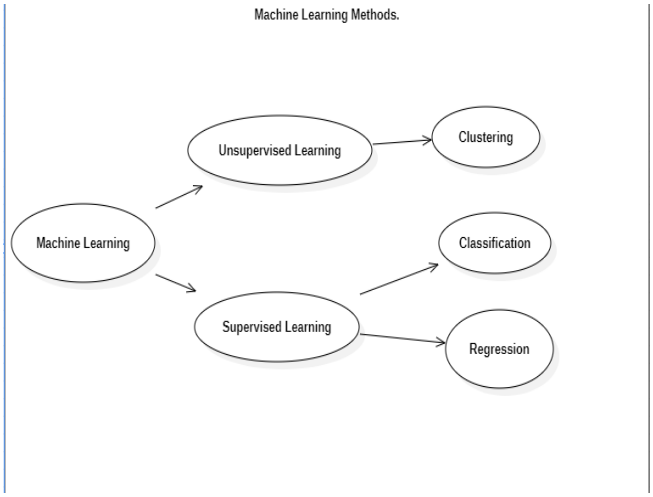

Figure-3. Machine Learning Methods

6. Problem Formulation:-

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 – foreign exchange rate prediction, stock market 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. 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

8) Project Deployment

7.1 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))
```

Figure-4. Main Imported Libraries

7.2 Data Understanding:-

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 506 entries, 2020-03-19 to 2022-03-18
Data columns (total 7 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Open          506 non-null    float64
 1   High          506 non-null    float64
 2   Low           506 non-null    float64
 3   Close         506 non-null    float64
 4   Volume        506 non-null    int64
 5   Dividends     506 non-null    int64
 6   Stock Splits  506 non-null    int64
dtypes: float64(4), int64(3)
```

Figure-5. Data Understanding

7.3 Data Pre-processing And EDA:-

Date	Open	High	Low	Close
2020-03-19	1481.500000	1491.599976	1471.000000	1478.599976
2020-03-20	1483.500000	1484.000000	1483.500000	1484.000000
2020-03-23	1505.300049	1567.000000	1490.199951	1567.000000
2020-03-24	1592.699951	1685.500000	1592.699951	1660.199951
2020-03-25	1676.500000	1676.500000	1631.099976	1632.300049
...	...	...	...	...
2022-03-14	1976.500000	1977.699951	1959.599976	1959.599976
2022-03-15	1939.800049	1939.900024	1910.699951	1928.500000
2022-03-16	1922.400024	1922.400024	1907.199951	1908.000000
2022-03-17	1933.900024	1947.199951	1933.900024	1942.099976
2022-03-18	1934.500000	1939.300049	1928.199951	1928.199951

Figure-6 . Column Deletion

Date	Open	High	Low	Close
2020-03-19	1481.500000	1491.599976	1471.000000	1478.599976
2020-03-20	1483.500000	1484.000000	1483.500000	1484.000000
2020-03-21	NaN	NaN	NaN	NaN
2020-03-22	NaN	NaN	NaN	NaN
2020-03-23	1505.300049	1567.000000	1490.199951	1567.000000
2020-03-24	1592.699951	1685.500000	1592.699951	1660.199951
2020-03-25	1676.500000	1676.500000	1631.099976	1632.300049

```
price= data.interpolate(method = 'linear')
price.head(7)
```

Date	Open	High	Low	Close
2020-03-19	1481.500000	1491.599976	1471.000000	1478.599976
2020-03-20	1483.500000	1484.000000	1483.500000	1484.000000
2020-03-21	1490.766683	1511.666667	1485.733317	1511.666667
2020-03-22	1498.033366	1539.333333	1487.966634	1539.333333
2020-03-23	1505.300049	1567.000000	1490.199951	1567.000000
2020-03-24	1592.699951	1685.500000	1592.699951	1660.199951
2020-03-25	1676.500000	1676.500000	1631.099976	1632.300049

Figure-7. Upsampling And Interpolation

7.4 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.

Figure-5. Gold Price By Months

Figure-6. Gold Prices Lag Plot

7.5 Statistical Measures (Mean, Standard-deviation, Kurtosis

Coding :

```
mean=df_d['GLD'].mean()
# computing standard deviation of Gold stock
std=df_d['GLD'].std()
kurt=df_d['GLD'].kurtosis()
print('Mean=',mean)
print('Standard Deviation=',std)
print('Kurtosis=',kurt)
#Plotting Histogram
df_d['GLD'].hist(bins=20)

plt.axvline(mean, color='w',linestyle='dashed',linewidth=2)
plt.axvline(std, color='r',linestyle='dashed',linewidth=2)
plt.axvline(-std, color='r',linestyle='dashed',linewidth=2)
plt.title(" Mean, Standard deviation and Kurtosis of Gold Prices")
plt.show()
```

Output:

```
Mean= -8.65698612128203e-05
Standard Deviation= 0.00961153616700639
Kurtosis= 8.606584924918355
```

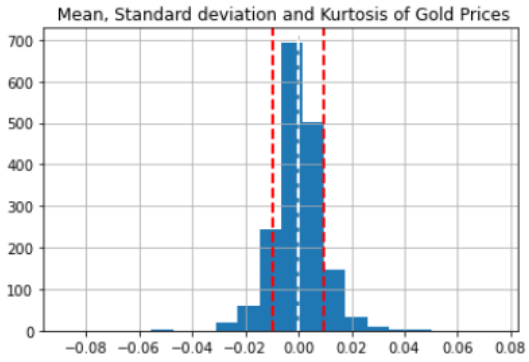

Figure-7. Mean , Standard deviation and Kurtosis of Gold Prices

7.6 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.

	Model	RMSE_Value
0	Simple Exponential Method	69.153642
1	Holt method	75.914794
2	HW exp smoothing add	78.903890
3	HW exp smoothing mult	74.328041
4	Linear Mode	67.420305
5	Exp model	67.198002
6	Quad model	66.932044
7	RandomForest	0.975509
8	LGBM and time series split	0.950000
9	LGBM Regressor with Repeated stratified K fold	0.982140

Figure-8. All Models And Their RMSE Value

We use Random Forest for model building because it's gives high accurancy and As we can see, the model's R-square is 97 % .

8. Final Results:-

Gold Trend Prediction

	Open	High	Low	Close	Volume	Dividends	Stock Splits
2022-06-03T00:00:00	1,867.6000	1,871.8000	1,845.4000	1,845.4000	53	0	0
2022-06-06T00:00:00	1,849.0000	1,854.1000	1,839.2000	1,839.2000	142	0	0
2022-06-07T00:00:00	1,836.9000	1,851.6000	1,835.0000	1,847.5000	1168	0	0
2022-06-08T00:00:00	1,844.8000	1,855.0000	1,844.4000	1,851.9000	218	0	0
2022-06-09T00:00:00	1,846.1000	1,850.1000	1,837.9000	1,848.8000	159	0	0
2022-06-10T00:00:00	1,842.7000	1,875.6000	1,823.9000	1,871.5000	1490	0	0
2022-06-13T00:00:00	1,873.8000	1,873.8000	1,818.7000	1,828.0000	127	0	0
2022-06-14T00:00:00	1,825.0000	1,825.5000	1,805.3000	1,809.5000	536	0	0
2022-06-15T00:00:00	1,814.1000	1,839.0000	1,810.8000	1,815.3000	536	0	0
2022-06-16T00:00:00	1,835.8000	1,838.7000	1,816.3000	1,823.1000	67047	0	0

Figure-9. Gold Trend Prediction

	count	mean	std	min	25%	50%	75%	max
Open	66.0000	1,831.3743	2.2842	1,725.5000	1,787.7250	1,812.2167	1,861.3000	2,053.6001
High	66.0000	1,840.0378	3.8857	1,735.2000	1,794.2750	1,819.9000	1,867.8167	2,072.0000
Low	66.0000	1,822.7148	1.0234	1,692.6000	1,781.0000	1,805.3750	1,852.1750	1,998.0000
Close	66.0000	1,831.1288	2.2054	1,721.5000	1,787.2500	1,811.5000	1,861.6000	2,040.1000
Volume	66.0000	4,138.8443	5.8676	0.0000	101.0000	237.0000	590.6667	209,783.0000
Dividends	66.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Stock Splits	66.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Figure-10 . Dataset Information

Data Visualization

Figure-11. Gold Price Prediction Data Visualization

Figure-12 . Monthly Prediction

Train Score : 99.75% and Test Score : 98.05% using Random Tree Regressor.

Accuracy: 99.68 %.

	Predictions
2022-06-16T00:00:00	1,788.1840
2022-06-17T00:00:00	1,783.4335
2022-06-18T00:00:00	1,850.6448
2022-06-19T00:00:00	1,749.3807
2022-06-20T00:00:00	1,983.8700
2022-06-21T00:00:00	1,838.4198
2022-06-22T00:00:00	1,802.4839
2022-06-23T00:00:00	1,788.7821
2022-06-24T00:00:00	1,895.0198
2022-06-25T00:00:00	1,785.3170

Figure-12 . Accurate Predictions With 99.69 % Accuracy

Figure-13 . Daily Gold Prices Index

9. Conclusion:-

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 April 2020 and Jan 2022. The results on proposed model is as per the following:

1)It is concluded that machine learning algorithms with Random Forest analysis are very useful in gold price prediction.

2)It is concluded that, the model's R-square is 97 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.

3)Results show that proposed Random forest method-machine learning beats customary and current predicting models

10. 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 and 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. International Journal of Trend in Scientific Research and Development (IJTSRD) @ www.ijtsrd.com eISSN: 2456-6470 @ IJTSRD | Unique Paper ID – IJTSRD33143 | Volume – 4 | Issue – 5 | July-August 2020 Page 1456

[13] NurulAsyikin Zainal and ZurianiMustaffa, "Developing A Gold Price Predictive Analysis Using Grey Wolf Optimizer", 2016 IEEE Student Conference on Research and Development, 2016.

[14] Hossein Mombeini and AbdolrezaYazdani-Chamzini, "Modeling Gold Price via Artificial Neural Network", Journal of Economics, Business and Management, 2015.

[15] ZurianiMustaffa and NurulAsyikin Zainal, "A Literature Review On Gold Price Predictive Techniques", 4th International Conference on Software Engineering and Computer Systems (ICSECS), 2015.