# GOLD PRICE PREDICTION USING MACHINE LEARNING

A report on
Mini Project (SRS and Design) submitted in partial fulfilment of the requirements for the award of the degree of

## Master of Computer Applications


**By**

**PRINS VERMA**

**MCA IV Semester**

**Exam Roll No.: NA**

**Enrolment No.: U1949059**

Under the Supervisions of

**Dr. VINOD KUMAR**

**Department of Electronics and Communication**
**(J K Institute of Applied Physics & Technology)**
**University of Allahabad**
**Prayagraj – 211002, India**


**July, 2021**

# Candidate's Declaration

I, **Prins Verma,** hereby certify that the work, which is being presented in the report/thesis, entitled "**Gold Price Prediction Using Machine Learning"**, in partial fulfillment of the requirement for the award of the Degree of **Masters of Computer Application** and submitted to the institution is an authentic record of my own work carried out during the period June, 2021 to July, 2021 under the supervision of **Dr. Vinod Kumar** at the Department of Electronics and Communication, University of Allahabad. The matter presented in this report/thesis has not been submitted elsewhere for the award of any other degree or diploma from any Institutions.

I declare that I have cited the reference about the text(s) /figure(s) /table(s) /equation(s) from where they have been taken. I further declare that I have not willfully lifted up some other's work, para, text, data, results, etc. reported in the journals, books, magazines, reports, dissertations, theses, etc., or available at web-sites and included them in this report/thesis and cited as my own work.

Date: July 27, 2021 Signature of the Candidate

**Prins Verma**

## Certificate From The Supervisor

This is to certify that the **Mr. Prins Verma** has carried out this project/dissertation entitled **Gold Price Prediction Using Machine Learning** under my supervision.

Date: ……..

Signature of the Supervisor
**Dr. Vinod Kumar**
(Supervisor)
Seal/Designation

## Acknowledgement

Project is sort of a bridge between theoretical and practical working. With this willing I joined this particular project. I am feeling oblige in taking the chance to sincerely thanks to **Dr. Vinod Kumar Sir** for his guidance, generous attitude, inspiration and constructive suggestions that help me in the preparation of this project. I am also thankful to all my friends who have always helping and encouraging me though out in successful completion of the project.

Prins Verma

# Abstract

This Project is based on a study conducted to know the relation between gold price and selected factors influencing it, namely stock exchange, petroleum price, rupee dollar rate of exchange, inflation and interest rate. In traditional events of Asian countries, gold **is** additionally presented as gifts/souvenirs and in marriages, gold ornaments are presented as Dowry in India, Pakistan and other countries. Monthly price data for the period of time- January 2000 to December 2020 was used for the study. The data is divided into two periods, period $I^{st}$ from January 2000 to October 2011 during which the gold price exhibits a raising trend and period $II^{nd}$ from November 2011 to December 2020 where the gold price is showing a horizontal trend. There are 3 machine learning algorithms- linear regression, random forest regression and gradient boosting regression is used for analyzing these data. It has been found that the correlation between the variables is robust during the period $I^{st}$ and weak during period $II^{nd}$. While these models show good fit with data during period $I^{st}$, the fitness isn't good during the period $II^{nd}$. While random forest regression is found to have better prediction accuracy for the whole period, gradient boosting regression is found to offer better accuracy for the 2 periods taken separately.



**Keywords:** Machine Learning, Regression, Prediction

(Sample)

# Table Of Contents

# CHAPTER 1

# Introduction

## 1.1 Background And Motivation

Now a days, Savings and Investments form an integral a part of everyone's life. Investments refer as the utilization of present funds with an objective of earning a positive return in future. In an economic sense, an investment is often considered as the purchase of assets that aren't consumed today but are utilized in the longer term to make wealth. In finance, an investment is purchase of a monetary asset with the thought that the asset will provide income within the future or will later be sold at a better price for a profit. The Indian economy being one among the fastest growing in the world has resulted in higher income level and a plethora of investment avenues. There are varieties of investments avenues available for investors, which include stock, deposit, commodity and land. Each of them differs in terms of risk and return characteristics. Gold is one of the attractive assets for Investment Avenue by many investors, due to its increasing value and therefore the area of usage. Investor's preference for gold as a protective asset, increases due to their negative expectations concerning things within the developed exchange markets and therefore the capital markets [1].

## 1.2 Problem Statement And Objectives

Gold is also considered to be "the asset of final instance" i.e. is the asset investors rely on, when the developed world capital markets are not capable to provide desirable profitability [2]. Thus it can be said that investors see gold as a tool to hedge against the fluctuations in other markets. Gold is a precious metal, so like any other goods; gold's price should depend on supply and demand. We know that the gold price goes up and down daily in the share market. Due to this the gold rate in jewelry shop also varies, through this the common people can't decide when to buy or sell gold. Sometime people take wrong decision and makes loss.

For solving these crises, the “Gold Price Prediction” can be used. My objective is to build a model with the help of machine learning algorithm that will take the dataset and on the basis of previous data it will predict the price of gold. Based on the prediction prices given by the application people can make their decision to buy or sell gold for their own profit.

## 1.3 PROJECT CATEGORY

The project category of Gold Price Prediction is **Machine Learning, investment, stock exchange**, **probability** and **data analysis.**

## 1.4 Tools/Platform Requirement

A computer system having minimum hardware requirement:-

- Processor: Intel Core i3 or above
- Ram: 4 GB or more
- Operation System: Windows 7 or above
- Disk Size: Minimum 2 GB

Software Requirement:

- Python Version: 3.9.X
- Machine Learning

## 1.5 Contribution Of The Thesis/Report

In this project my aim is to design a improved, updated and effective model with the help of machine learning algorithm that will perform its calculations on the basis of previous data which is being provided and it will predict the price of gold. Based on the prediction prices given by the application people can make their decision to buy or sell gold for their own profit.

## 1.6 Structure Of The Thesis/Report

Chapter 1, introduced about the Gold Price Prediction, its background and motivation, how Gold price is to be concerned in the point of investment in its stock exchange, buy and sell by every person of India or any country. I also made a small contribution in this project by updating all current data by the help of reading various thesis/report.

Chapter 2, described the previous works done by others in this particular area. I considered the works analyze their merits and demerits, and shortcomings that motivate me to work further in this area. This chapter contains a specialized overview of Gold Price Prediction.

Chapter 3, introduced Gold Price Prediction, how its work in easy terms that a normal people can understand, who don't have any idea of investment and how they can invest. I have explained all the methodology I have used in this project with a basic work plan by the help of diagrams and algorithm.

Chapter 4, conclude about Gold Price Prediction and its scope for future work.

# CHAPTER 2

# LITERATURE REVIEW/SURVEY

A significant amount of research has been done to understand and predict gold prices. Various studies have developed different predictive models based on different techniques and factors. Some studies try to make predictions based on historical gold prices. Others focus on the economic aspects and try to explain the prices of gold with respect to a variety of economic factors. Therefore, we evaluate and summarize some widely used models into the following two categories:

1) Standard Time Series Models and
2) Structural Models Considering Economic Factors.

## 2.1) Standard Time Series Models (Technical Models) :

- Standard time series models attempt to predict the gold price using the current and historical gold prices. These models are useful in the following scenarios:
- The prices show autocorrelation and autoregressive behavior, i.e., there is a pattern or a significant correlation between current and the previous prices.
- There are a large number of explanatory variables and it is difficult to understand them well because they interact with each other in a very complicated manner.
- Forecasting the dependent variable may require predicting the explanatory variables. And prediction of the explanatory variable might in turn be a harder problem.
- Not all explanatory factors and variables are known.

## 2.2) Structural Models Considering Economic Factors:

Structural models considering a variety of economic factors apply to the gold price prediction. According to Ismail et al.(2009), They design models with the gold price as the only dependent variable, alongside different numbers of independent variables. Initially, they propose that the gold price is dependent on the following 8 factors: Commodity Research Bureau future index (CRB), USD/Euro Foreign Exchange Rate (EUROUSD); Inflation rate (INF); Money Supply (M1); New York Stock Exchange (NYSE); Standard and Poor 500 (SPX); Treasury Bill (T-BILL) and US Dollar index (USDX). Therefore, their first-order regression model, which they call it naive model, is like :

$$\hat{Y} = -560.618 + 0.712X_1 + 161.740X_2 - 7.836X_3 + 0.424X_4 - 0.010X_5 + 0.010X_6 + 3.198X_7 + 0.580X_8$$

where

$\hat{Y}$ is the predicted gold price;

$X_1$ is CRB;

$X_2$ is EUROUSD;

$X_3$ is INF;

$X_4$ is M1;

$X_5$ is NYSE;

$X_6$ is SPX;

$X_7$ is T-Bill;

$X_8$ is USDX.

Then they show that using stepwise regression, the number of independent variables can be reduced from 8 to 4. Their enhanced model is

$$\hat{Y} = -301.509 + 0.676X_1 + 114.651X_2 - 5.563X_3 + 0.309X_4$$

Where

$\hat{Y}$ is the predicted gold price,

$X_1$ is CRB;

$X_2$ is EUROUSD;

$X_3$ is INF;

$X_4$ is M1.

# Chapter 3

# Design Details, Proposed Approaches, System And Algorithms

## 3.1 Introduction

Based on the review of literature five major factors that is considered to have influence on the gold price were identified. The factors that are considered for this study are stock market, crude oil price, rupee dollar exchange rate, inflation and interest rate. Values of Nifty500 index is taken as a representation of the stock prices. Nifty500 index represents the top 500 companies based on full market capitalization from the companies listed in National Stock Exchange. Inflation is represented by Consumer Price Index with base year as 2001. The term deposit rate for deposits above 1 year is taken as a representation for interest rate. Spot gold price in rupees per ounce is used to represent gold price. Monthly data from January 2000 to December 2020 were collected for all these variables. The databases of Centre for Monitoring Indian Economy were used for getting these data. There were 228 observations for each of these variables.

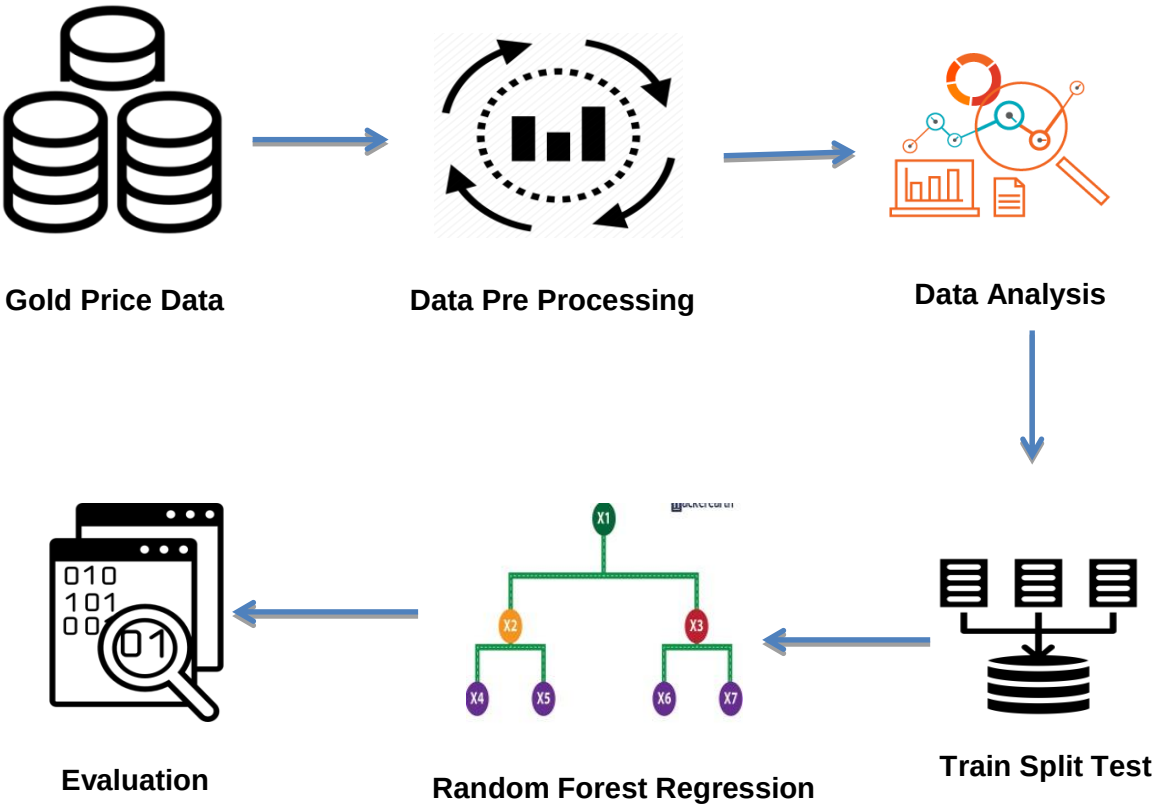
Values of Nifty500 index is taken as a representation of the stock prices. Nifty500 index represents the top 500 companies based on full market capitalization from the companies listed in National Stock Exchange. Inflation is represented by Consumer Price Index with base year as 2001. The term deposit rate for deposits above 1 year is taken as a representation for interest rate. Spot gold price in rupees per ounce is used to represent gold price. Monthly data from January 2000 to December 2020 were collected for all these variables. The databases of Centre for Monitoring Indian Economy were used for getting these data. There were more than 250 observations for each of these variables.

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.

Linear regression models with more than one independent variable are called multiple linear models.

In this study these machine learning algorithms (Linear Regression, Random Forest Regression and Gradient Boosting Regression) are implemented using python. The prediction accuracy of the regression methods used were measured using Mean Squared Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

## 3.2 Methodology Used




❖ **Work Plan:**

- **Gold Price Data:** We have collected data consisting of gold prices for several days in a period of 10 years. We will use this data to train our machine learning algorithm. When we train this algorithm, so we give new value it can predict the gold prices for that specific value.

- **Data Pre –Processing:** We cannot feed this raw data to our machine learning algorithm so we need to preprocess this data properly.

- **Data Analysis:** To know which feature is important and which features are correlated with each other i.e. visualization techniques.

- **Train Test Split:** We will split our original data set into training data and test data so we will use this training data to train our machine learning model and we need to evaluate our model so we need to find how well our model is performing so for that we will be using a test data to test our machine learning model.

- **Random Forest Regression:** Once we feed this training data to our random forest regression model, it can make new predictions.
- **Evaluation:** We will evaluate our model based on the test data we are splitted.

❖ **Data Description:** Our study aims to explore the significant variables arising from the financial and geopolitical spheres that will enable the future gold price prediction. Hence, the dependent variable is the gold price (GOLD), on the other hand our explanatory variables are oil spot price (OIL), US dollar spot price (USD), S&P500 index (S&P500), Euro against USD pair (EUR/USD), copper price (COPPER), global policy uncertainty (GPU) [18], trade policy uncertainty in the India (EPU) [19], world uncertainty index (WUI) [20], and finally world trade uncertainty (WTU) [21].

We will concentrate on the new millennium time span. Thus, our analysis will be based on the period running from 01/01/2000 to 31/12/2020 with a monthly frequency. The dataset didn't include any missing values.

I have provided the data set in Appendices section.

## 3.3 ALGORITHM

There are 3 algorithms of the Machine Learning will be used in this project:

- **Linear Regression:**
  1. Linear Regression is a machine learning algorithm based on supervised learning.
  2. It performs a regression task.
  3. Linear regression is a linear model, i.e. a model that assumes a linear relationship between the input variables (x) and the single output variable (y).
  4. More specifically, that y can be calculated from a linear combination of the input variables (x).
  5. When there is a single input variable (x), the method is referred to as simple linear regression.
  6. When there are multiple input variables, the method is referred as

multiple linear regressions.

7. A representation of multiple linear regressions is where, Y is dependent
8. Variable and $X_1$, $X_2$ … are independent variables are as seen below.
9. $Y = a + b1*X_1 + b_2*X_2 + ... + bp*Xp$.
10. As such, linear regression was developed in the field of statistics and is studied
11. as a model for understanding the relationship between input and output
12. Numerical variables, but has been borrowed by machine learning.
13. It is both a Statistical algorithm and a machine learning algorithm now.
14. This algorithm is used to show the relationship between the factors (Independent variables) which is affecting the price of gold (such as Petroleum, stock exchange and Rupee/dollar rate) and Dependent variables (Gold).

- **Random Forest Regression:**

1. 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.
2. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.
3. 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.
4. A Random Forest algorithm has the following advantages when compared to most of the other algorithms - The over fitting problem will never come when we use the random forest algorithm in any classification problem.
5. For classification and regression task, the same random forest algorithm can be used. 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.

- **Gradient Boosting Regression:**

1. Gradient boosting is a machine learning technique which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees.
2. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.
3. Gradient boosting trees are better than random forests in many situations, but they are prone to over fitting in some situations. However, there are strategies to overcome the same and build more generalized trees using a combination of parameters like learning rate (shrinkage) and depth of tree. Generally, these two parameters are kept on the lower side to allow for slow learning and better generalization.

# Chapter 4

# Conclusion And Future Work/Extension

## 4.1 Introduction

In this chapter will summarize the whole project and conclude the main aim of the project that i.e. training our model and also explain the future scope of our model and its real life implementation.

## 4.2 Conclusions

The main aim of this study is to predict the gold price that is influenced by the economic variables such as stock profit exchange, silver price, EUR/USD, United States oil ETF. In this study, we used the machine learning algorithms such as linear regression, random forest and gradient boosting to predict the price of gold accurately. Considering the results obtained, we conclude that the random forest model performed better than the other models.

## 4.3 Scope For Future Work

For future work, we can improve the results and predict the price more accurately by incorporating the other factors such as gold production, crude oil price, platinum price, inflation to the data and by using deep learning.

# References

[1] W. Du and J. Schreger, “Local Currency Sovereign Risk,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2976788, Dec. 2013.

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

[3] 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.

[4] C. Toraman, Ç. Basarir, and M. F. Bayramoglu, “Determination of factors affecting the price of gold: A study of MGARCH model,” Bus. Econ. Res. J., vol. 2, no.4, p. 37, 2011.

[5] C. Lawrence, “Why is gold different from other assets? An empirical investigation,” Lond. UK World Gold Council. 2003.

[6] L. A. Sjaastad and F. Scacciavillani, “The price of gold and the exchange rate,” J. Int. Money Finance, vol. 15, no.6, pp. 879–897, 1996.

[7] S. A. Baker and R. C. Van Tassel, “Forecasting the price of gold: A fundamentalist approach,” Atl. Econ. J., vol. 13, no. 4, pp. 43–51, 1985.

[8] H. Naser, “Can Gold Investments Provide a Good Hedge Against Inflation? An Empirical Analysis,” Int. J. Econ. Financ. Issues, vol. 7, no. 1, pp. 470–475, 2017.

[9] P. Khaemasunun, “Forecasting Thai gold prices,” Available Http://www Wbiconpro Com3-Pravit. Pdf Acess, vol. 2, 2014.

[10] S. M. Hammoudeh, Y. Yuan, M. McAleer, and M. A. Thompson, “Precious metals–exchange rate volatility transmissions and hedging strategies,” Int. Rev. Econ. Finance, vol. 19, no. 4, pp. 633–647, 2010.

[11] A. Han, K. K. Lai, S. Wang, and S. Xu, “An interval method for studying the relationship between the Australian dollar exchange rate and the gold price,” J. Syst. Sci. Complex., vol. 25, no. 1, pp. 121–132, 2012.

[12] B. T. Ewing and F. Malik, “Volatility transmission between gold and oil futures under structural breaks,” Int. Rev. Econ. Finance, vol. 25, pp. 113–121, 2013.

[13] 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. [14] 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.

[14] Ahir, H., Bloom, N., & Furceri, D. (2019). Caution: Trade uncertainty is rising and can harm the global economy. VoxEU.org, July, 4.

[15] SHARP, T. O. M. (2020, March 3). An Introduction to Support Vector Regression (SVR). Retrieved May 3, 2020, from https://towardsdatascience.com/an-introduction-to-support-vectorregression-svr-a3ebc1672c2

[16] Iftikhar ul Sami, Khurum Nazir Junejo Graduate School of Science and Engineering ,Karachi Institute of Economics & Technology Karachi, Pakistan , Nov 12,2017

[17] Ayush Sengupta, Benjamin Lin, Komal Sanjeev, and Sreevathsan Ravichandran, Department of Computer Science, Stony Brook University, Stony Brook, NY 11794-4400

# Appendices

The Data set used in this project is very large so, there is a link to access it: https://drive.google.com/drive/folders/1DaWsglLyoMLkpAyMJOLu4liDkwicWYbI?usp=sharing

# Plagiarism Report

PLAGIARISM SCAN REPORT

| | |
|---|---|
| **Date** | July 27, 2021 |
| **Exclude URL:** | NO |

| | | | |
|---|---|---|---|
| Unique Content | 96% | Word Count | 1,375 |
| Plagiarized Content | 4% | Records Found | 5 |
| Paraphrased Plagiarism | 0 | | |

CONTENT CHECKED FOR PLAGIARISM:

GOLD PRICE PREDICTION USING MACHINE LEARNING A report on Mini Project I (SRS and Design) submitted in partial fulfilment of the requirements for the award of the degree of Master of Computer Applications By PRINS VERMA MCA IV Semester Exam Roll No.: NA Enrolment No. : U1949059 Under the Supervisions of Dr. VINOD KUMAR Department of Electronics and Communication (J K Institute of Applied Physics & Technology) University of Allahabad Prayagraj –

PLAGIARISM SCAN REPORT

| | |
|---|---|
| **Date** | July 27, 2021 |
| **Exclude URL:** | NO |

| | |
|---|---|
| Unique Content | 94% |
| Plagiarized Content | 6% |
| Paraphrased Plagiarism | 0 |

| | |
|---|---|
| Word Count | 412 |
| Records Found | 1 |

CONTENT CHECKED FOR PLAGIARISM:

CHAPTER 2 LITERATURE REVIEW/SURVEY A significant amount of research has been done to understand and predict gold prices. Various studies have developed different predictive models based on different techniques and factors. Some studies try to make predictions based on historical gold prices. Others focus on the economic aspects and try to explain the prices of gold with respect to a variety of economic factors. Therefore, we evaluate and summarize some widely

PLAGIARISM SCAN REPORT

| | |
|---|---|
| **Date** | July 27, 2021 |
| **Exclude URL:** | NO |

| | |
|---|---|
| Unique Content | 85% |
| Plagiarized Content | 15% |
| Paraphrased Plagiarism | 0 |

| | |
|---|---|
| Word Count | 517 |
| Records Found | 5 |

CONTENT CHECKED FOR PLAGIARISM:

3.3 ALGORITHM

There are 3 algorithms of the Machine Learning will be used in this project:

 Linear Regression:

1. Linear Regression is a machine learning algorithm based on supervised learning.

2. It performs a regression task.