JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Gold Price Prediction

A Phase 1 Report on Gold Price Prediction

[1]Sai Neha, [2]Dr. Shashidhar Kini K
[1]Student, [2]Professor & Head
[1]Department of Computer Application,
[1]Srinivas Institute of Technology, Valachil Mangaluru, Karnataka, India

Abstract : We predict gold prices using machine learning models by analyzing historical and financial data. Algorithms like Linear Regression, Random Forest, and LSTM are evaluated using MAE, RMSE, and R^2 scores. The results show improved prediction accuracy, providing valuable insights for investors and financial analysts. This approach helps in risk assessment, market analysis, and financial decision-making, making it a reliable tool for forecasting gold price movements in dynamic economic conditions.

IndexTerms - ***Gold Price Prediction, Regression Models, Random Forest, LSTM***

I. INTRODUCTION

Gold has long been regarded as a safe-haven asset and plays a crucial role in global financial markets. Its price is influenced by various factors such as inflation, interest rates, exchange rates, geopolitical events, and market demand**.** Predicting gold prices is a challenging task due to the complex and dynamic nature of financial markets**.** Traditional statistical methods often fail to capture non-linear patterns in price fluctuations, making machine learning (ML) a promising approach for more accurate predictions.Use the enter key to start a new paragraph. The appropriate spacing and indent are automatically applied.

This project aims to develop a machine learning model for predicting gold prices using historical data and key economic indicators. By using various algorithms such as linear regression, decision trees, and neural networks, the model will analyze past price trends, global financial variables, and market conditions to forecast future gold prices. Additionally, the project will explore the impact of geopolitical risks and inter-market relationships between gold and other assets like oil and stock indices. By integrating these diverse features, the goal is to create a comprehensive, data-driven tool for forecasting gold price movements, while also providing insights into market volatility and risk. The objective is to provide an accurate, data-driven tool that can assist investors, traders, and financial analysts in making informed decisions based on predictive insights.

Benefits of the Gold Price Prediction:

- Supports Informed Decision-Making
- Aids Economic and Financial Planning
- Enables Real-Time Forecasting
- Reveals Hidden Patterns and Trends

EASE OF USE

The proposed gold price prediction system is designed with simplicity and accessibility in mind, making it suitable for both technical and non-technical users. It can be implemented using platforms like Jupyter Notebook or integrated into a basic web interface, allowing users to interact with the model effortlessly. The system requires minimal input—mainly historical gold price data such as opening, closing, high, and low prices—making data entry straightforward. Once the data is provided, the entire

process of preprocessing, training, and prediction is automated, ensuring a smooth user experience. Even users without deep knowledge in finance or machine learning can interpret the results easily, as the output is presented in a clear and comprehensible format. Additionally, the model runs efficiently on standard computing resources, with quick prediction times, and can be easily extended or integrated with web applications or dashboards for broader use.

Abbreviations and Acronyms

Define each abbreviation or acronym the first time it appears in the text, even if it was already defined in the abstract.

- **ML** – Machine Learning
- **LR** – Linear Regression
- **RF** – Random Forest
- **RMSE** – Root Mean Squared Error
- **MAE** – Mean Absolute Error
- **R^2** – Coefficient of Determination
- **CSV** – Comma Separated Values (file format used for dataset)
- **EDA** – Exploratory Data Analysis
- **GUI** – Graphical User Interface (if you built one)
- **API** – Application Programming Interface (if data was fetched online)
- **CPU** – Central Processing Unit (mention if talking about system requirements)
- **EC2** – Elastic Compute Cloud (if AWS EC2 was used for deployment)

II.RESEARCH METHODOLOGY

The methodology section outline the plan and method that how the study is conducted. This includes Universe of the study, sample of the study, Data and Sources of Data, study's variables and analytical framework. The details are as follows;

2.1Population and Sample

In this study, the population refers to the complete set of historical gold price data available over a broad time frame, typically sourced from financial markets or commodity exchanges. The dataset may include gold prices over several years, containing features such as date, opening price, closing price, high, low, and volume. Due to the impracticality of processing the entire historical dataset in some cases, a representative sample is selected, usually covering a specific time period (e.g., from 2010 to 2023). The sample used in this project is a time-series dataset obtained from a reliable source such as Yahoo Finance or Kaggle, containing several thousand daily records. This sample is sufficient for training and testing the machine learning models to ensure accurate and meaningful predictions.

2.2 Data and Sources of Data

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. 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.

2.3 Theoretical framework

The **variables** in this study consist of **dependent** and **independent variables**. The study uses a pre-specified method for the selection of variables related to **gold price prediction** using machine learning. The dependent variable is the **gold price**, which is the target output the model aims to predict. The price of gold is typically influenced by factors such as historical prices, economic indicators, and market sentiment. These factors are used as independent variables in the predictive models.

Independent Variables:

- **Inflation Rate**: The inflation rate, measured using the **Consumer Price Index (CPI)**, is considered an independent variable. Inflation can significantly impact gold prices, as investors often turn to gold as a hedge against inflation. It is assumed that as inflation rises, the demand for gold increases, thus driving up its price (Iqbal et al., 2010). Higher inflation often results in a decrease in the purchasing power of money, leading to a higher demand for gold as a store of value.
- **Interest Rate**: Interest rates play a crucial role in determining gold prices. When interest rates rise, the opportunity cost of holding gold increases, as investors may prefer interest-bearing assets. This negatively affects the demand for gold, causing its price to decrease (Nguyen, 2010). Conversely, when interest rates are low, gold becomes a more attractive investment.
- **Oil Prices**: Oil prices have a significant effect on gold prices, as the two are often positively correlated. When oil prices rise, inflationary pressures increase, leading to a higher demand for gold as a safe haven (Iqbal et al., 2012). However, there is some debate on this relationship, with studies like Dash & Rishika (2011) suggesting that oil prices have no significant effect on gold prices.

- **Exchange Rate**: The exchange rate, particularly the **U.S. Dollar to other currencies**, influences gold prices. A stronger U.S. Dollar typically leads to lower gold prices, as gold becomes more expensive for holders of other currencies. Conversely, a weaker dollar generally leads to higher gold prices, as it becomes cheaper for foreign investors to buy gold (Jecheche, 2010).
- **Market Sentiment**: Investor sentiment and market volatility, captured through features such as **KSE-100 Index** or other stock market indices, influence the price of gold. In times of uncertainty, investors tend to shift their investments into gold, which drives up its price. Systematic risk, measured through the **Beta coefficient**, is used to quantify the market risk associated with the gold price movements (Fama and MacBeth, 1973).

Dependent Variable:

The **price of gold** is the primary dependent variable. The model aims to predict the gold price based on historical data, including the independent variables mentioned above. Gold prices are influenced by a mix of economic factors, market dynamics, and investor behavior. Thus, accurate prediction of gold prices requires an understanding of the complex relationships between these factors.

The **Gold Price Prediction** model considers both economic indicators and market factors as independent variables that influence gold prices. Theoretical underpinnings from **financial economics** and **machine learning** suggest that these variables have a direct or indirect relationship with gold price movements, making them critical to the prediction process.

2.4Statistical tools and econometric models

This section elaborates the proper statistical/econometric/financial models which are being used to forward the study from data towards inferences. The detail of methodology is given as follows.

2.4.1 Statistical Tools

- **Python:** The primary programming language used in this study is **Python**, due to its extensive libraries and frameworks for data analysis, machine learning, and statistical modeling. Key libraries include:
  - **Pandas:** Used for data manipulation and cleaning, especially for handling time-series data and creating dataframes.
  - **NumPy:** Provides support for large multi-dimensional arrays and matrices, which is essential for numerical computations.
  - **Matplotlib and Seaborn:** These libraries are used for data visualization, helping to identify trends, patterns, and relationships in the dataset.
  - **SciPy:** Employed for scientific computing, including statistical tests and optimizations required in model building.
  - **Scikit-learn:** The primary library for implementing machine learning algorithms, including Linear Regression, Random Forest, and evaluation metrics (e.g., RMSE, MAE, R^2).
  - **Jupyter Notebook:** This tool is used to document the analysis and provide a user-friendly environment for running Python code and generating outputs in a presentable format.

2.4.2 Econometric Models

2.4.2.1 Linear Regression (LR)

The **Linear Regression** model is used to explore the relationship between the dependent variable (gold price) and one or more independent variables (such as inflation, interest rates, exchange rates, etc.). This model assumes a **linear relationship** between the predictors and the target variable. In the context of gold price prediction, the model will help identify how much each independent variable (e.g., inflation rate, exchange rate) contributes to the fluctuation of gold prices over time.

- **Assumptions**: Linearity, independence, homoscedasticity, and normality of errors.
- **Equation**: $\text{Gold Price}=\beta 0+\beta 1(\text{Inflation})+\beta 2(\text{Interest Rate})+\cdots+\epsilon$

2.4.2.2 Random Forest (RF):

Random Forest is an **ensemble learning** method that builds multiple decision trees to make predictions. Unlike Linear Regression, Random Forest does not assume any linear relationship between the features and the dependent variable. It can handle complex, non-linear relationships and is robust against overfitting, making it suitable for predicting gold prices when there are intricate patterns and interactions between variables.

- **Assumptions**: No assumptions about the distribution of the data.
- **Working:** Randomly selects subsets of data and features to create decision trees, and then averages their predictions for a more stable result.

2.4.2.3 Econometric Models for Financial Forecasting:

- **CAPM (Capital Asset Pricing Model):** Although not directly applicable in a machine learning context, CAPM could be used as a theoretical framework for understanding how systematic risk (beta) impacts the price of gold, which could be incorporated into a broader econometric approach.
- **APT (Arbitrage Pricing Theory):** APT, similar to CAPM, incorporates multiple macroeconomic factors (e.g., inflation rate, interest rates) to determine how they affect gold prices. A multivariate regression or machine learning approach may be used in parallel with APT to predict future price movements based on these factors.

2.4.3 Model Evaluation Metrics

The performance of the predictive models is assessed using various **statistical metrics**, including:

- **Root Mean Squared Error (RMSE):** Measures the average magnitude of prediction errors, with a penalty for larger errors. A lower RMSE indicates a better fit of the model.
- **Mean Absolute Error (MAE):** Calculates the average absolute error between predicted and actual values. It is easier to interpret than RMSE and does not give more weight to larger errors.
- **R^2 (Coefficient of Determination):** Indicates how well the model explains the variance in the target variable (gold price). An R^2 value closer to 1 suggests the model explains most of the variance.

III. RESULTS AND DISCUSSION

Table 4.1: Descriptive Statistics

Variable	Minimum	Maximum	Mean	Std. Deviation
Gold Price (USD)	1200	2025	1650.45	190.34
Inflation Rate (%)	1.2	7.4	4.03	1.58
Interest Rate (%)	3.5	9.2	5.84	1.23
Exchange Rate (USD)	65.2	82.4	73.08	4.25
Oil Price (USD)	32.1	115.3	74.62	21.34

Table 4.1: The descriptive statistics show that the dataset covers a wide range of values across all economic variables. Gold price fluctuates between 1200 and 2025 USD, reflecting volatility in the global market. The inflation and interest rates show moderate variation, which could significantly impact investor decisions. The exchange rate and oil price also exhibit variability, potentially affecting gold price trends. These variations provide a solid base for predictive modeling.

IV. ACKNOWLEDGMENT

The author wishes to express sincere gratitude to the Project Guide and Head of the Department of MCA, Dr. Shashidhar Kini K, for his invaluable guidance, constant encouragement, and kind support throughout this research work. Appreciation is also extended to the Principal, Dr. Shrinivasa Mayya D, for fostering an environment conducive to completing this project within the institution. The author thanks the management of Srinivas Institute of Technology for their direct and indirect support. Gratitude is also due to all the faculty members and non-teaching staff of the MCA department for their constant help and support. Finally, the author is indebted to parents and friends for their unwavering support and belief throughout this endeavor.

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.