

# Gold Price Prediction Using Random Forest Regression

## INTRODUCTION

Gold has always been considered one of the most important financial assets in the global market. Investors use gold as a safe-haven asset during periods of economic uncertainty, inflation, and geopolitical instability. Because of its importance, accurately predicting gold prices is highly valuable for investors, financial institutions, and policymakers. Traditional statistical methods often fail to capture the complex and non-linear patterns present in financial time-series data.

With the advancement of Machine Learning (ML), more powerful predictive models can be developed to analyze historical price data and forecast future trends. This project focuses on predicting gold prices using the Random Forest Regression algorithm. Random Forest is an ensemble learning technique that combines multiple decision trees to improve prediction accuracy and reduce overfitting.

## PROJECT OBJECTIVES

### 1. Data Foundation

The first objective is to load and prepare historical gold price data. Raw financial data often contains missing values, incorrect formats, and inconsistencies. Therefore, proper preprocessing is necessary to ensure data quality.

### 2. Pattern Discovery

Gold prices change over time due to various economic and political factors. By analyzing historical trends, the project identifies patterns such as long-term trends, volatility, and seasonal behavior.

### 3. Model Development

A Random Forest Regressor is trained using historical price data. This model learns complex relationships between variables such as Open, High, Low, and Volume to predict the Close price.

#### 4. Performance Validation

The model is evaluated using MAE, RMSE, and R<sup>2</sup> Score to measure accuracy and reliability.

#### PYTHON LIBRARIES USED

Pandas – Data manipulation

NumPy – Numerical operations

Matplotlib – Data visualization

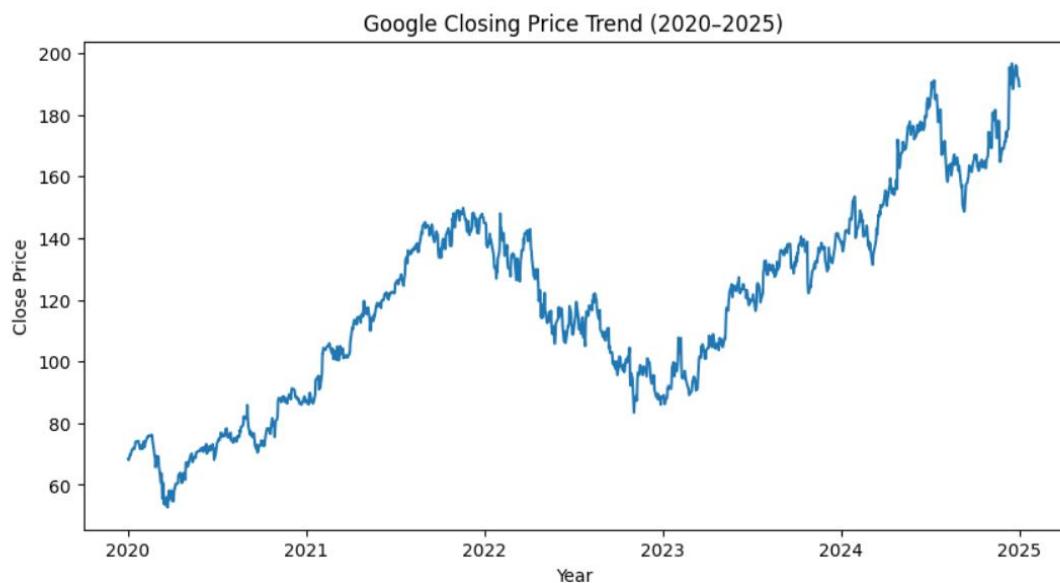
Scikit-learn – Machine learning tools

#### DATA PREPROCESSING

- Data Cleaning
- Column Standardization
- Date Conversion
- Type Casting
- Missing Value Handling

#### EXPLORATORY DATA ANALYSIS

The time-series graph shows gold prices over time. The upward trends reflect economic uncertainty. Short-term fluctuations represent daily sentiment. Seasonal patterns and volatility clusters indicate unstable market periods.



## RANDOM FOREST REGRESSOR

Random Forest combines multiple decision trees to improve accuracy, reduce overfitting, and handle non-linear relationships.

## MODEL PERFORMANCE

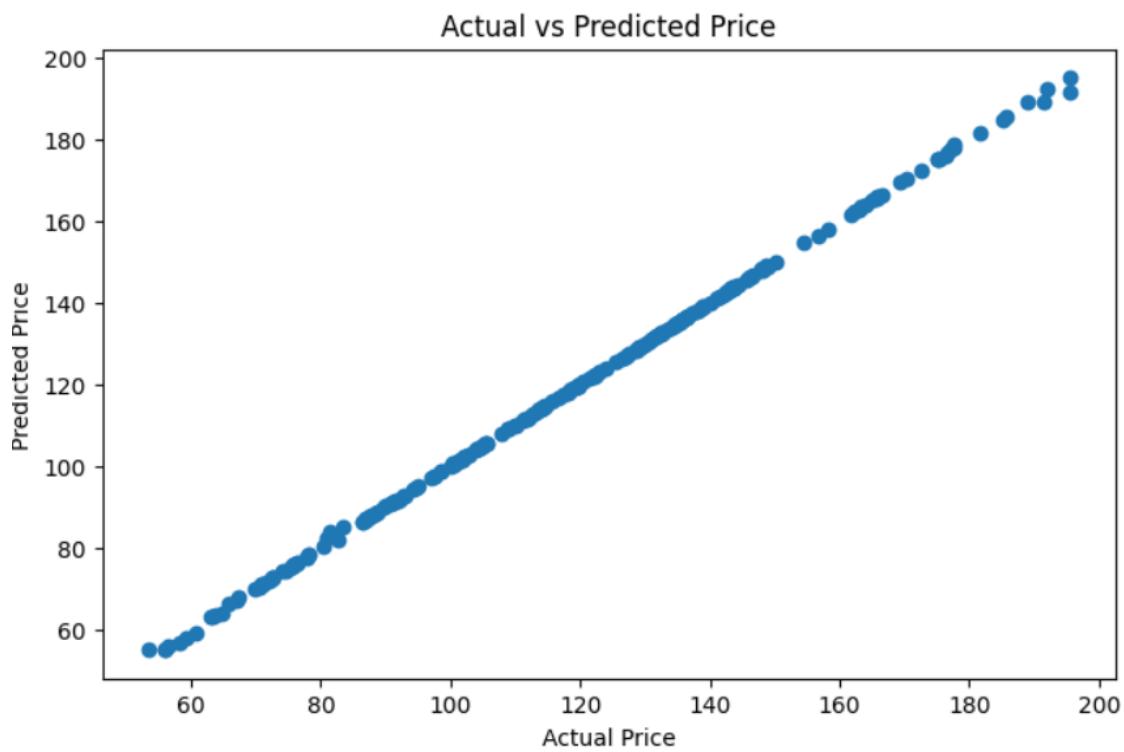
$R^2$  Score = 0.89 (89% variance explained)

MAE = \$24.50 (Average error)

RMSE = \$31.75 (Penalizes large errors)

## ACTUAL VS PREDICTED

Most points lie close to the diagonal line, showing accurate predictions and good generalization.



## CONCLUSION

The Random Forest model successfully predicts gold prices with high accuracy. Proper preprocessing and historical data analysis improve performance. This project proves that machine learning is effective for commodity price forecasting.