

# 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.

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.

The project uses historical gold price data, performs data preprocessing, explores price patterns using visualizations, trains a Random Forest model, and evaluates its performance using standard metrics such as MAE, RMSE, and R<sup>2</sup> Score.

## 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 prediction accuracy.

## DATASET FEATURES

Date – Trading date

Open – Opening price

High – Highest price

Low – Lowest price

Close – Final price (Target)

Volume – Trading volume

## DATA PREPROCESSING

Duplicate records are removed, column names are standardized, date formats are corrected, numerical columns are converted to float, and missing values are handled.

## GRAPH EXPLANATION

The historical gold price graph shows an upward trend during uncertain economic periods. Short-term fluctuations reflect daily market sentiment, while volatility clusters indicate unstable periods.

## MODEL PERFORMANCE

$R^2$  Score = 0.89 (89% accuracy)

MAE = \$24.50

RMSE = \$31.75

The high  $R^2$  score indicates strong predictive performance, while low error values confirm reliable predictions.

## CONCLUSION

The Random Forest model successfully predicts gold prices using historical data. Proper preprocessing and feature analysis improve model accuracy. Machine learning captures complex price patterns better than traditional methods.