A MINOR PROJECT SYNOPSIS

ON

# GOLD PRICE PREDICTION USING MACHINE LEARNING

SUBMITTED IN PARTIAL FULFILLMENT FOR THE AWARD OF DEGREE OF

**BACHELOR OF TECHNOLOGY**

**IN**

**ELECTRONICS AND COMMUNICATION ENGINEERING**


**Submitted by:**

**NEWTON RAJ (9922102104)**

**HARSHIT SRIVASTAVA (9922102107)**

**VEDANT ATRI (9922102113)**

**Under the Guidance of**

**DR. KAPILDEV TYAGI**

**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING**

**JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, NOIDA (U.P.)**


**NOVEMBER, 2024**

# DECLARATION

We hereby declare that this written submission represents our own ideas in our own words and where others' ideas or words have been included, have been adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission.

Place: JIIT-128, Noida
Date:

Name:Vedant Atri
Enrollment: 9922102113

Name:Harshit Srivastava
Enrollment: 9922102107

Name:Newton Raj
Enrollment: 9922102104

# CERTIFICATE

This is to certify that the minor project topic “Gold Price Prediction Using Machine Learning” submitted by Vedant Atri, Harshit Srivastava, Newton Raj in partial fulfillment of the requirements for the award of Bachelor of Technology Degree in Electronics and Communication Engineering of the Jaypee Institute of Information Technology, Noida is an authentic work carried out by them under my supervision and guidance. The matter embodied in this report is original and has not been submitted for the award of any other degree.

**Signature of Supervisor:**

**Name of the Supervisor: Dr. Kapil Dev Tyagi**

**ECE Department,**

**JIIT, Sec-128,**

**Noida-201304**

**Dated:**

# ABSTRACT

Significantly, gold is among the precious metals that are utilized to finance trading purchases. In countries like India, China, Indonesia, and many more, gold is considered the ideal jewellery, in addition, gold is also served as a present/remembrance and even gold accessories are presented as marriage settlements. Moreover, the countries with large gold reserves are considered a booming nations. At present time, precious metal like gold, is not just considered ornaments or jewellery but are considered as an investment and are kept with all nations' central bank as an assurance for repayment of non-native loans, and also to manage inflation. Due to the increasing demand and dispense of this asset in the market the state of the major economies throughout the globe has a considerable impact on gold prices. Due to the change in gold prices, more investors are now considering gold investments. But irregularity in the gold price in the market makes it riskier for the investor. Thus, the goal of “Gold price prediction” is to forecast gold’s price using a variety of Machine learning techniques, considering the relationship between several economic factors that influence gold rates.

This dataset comprises 2,290 daily observations spanning from January 2, 2008, to May 16, 2018. It focuses on financial and economic indicators relevant to asset price analysis and market dynamics. The variables include:

- **S&P 500 Index (SPX)**: A widely used benchmark of the overall performance of the U.S. stock market.
- **Gold ETF Price (GLD)**: Reflecting gold price movements in financial markets.
- **Crude Oil ETF Price (USO)**: Representing the price trends in the crude oil market.
- **Silver ETF Price (SLV)**: Tracking silver price fluctuations.
- **EUR/USD Exchange Rate**: Indicating the value of the Euro against the U.S. Dollar.

The dataset is complete, with no missing values, and provides a comprehensive view of the interplay between these key financial indicators. It is suitable for various analytical applications, such as time series forecasting, correlation analysis, and building predictive models for asset price movements.

# ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all the students in our group for their hard work, dedication, and cooperation throughout the completion of the project. Each member brought their unique strengths and skills to the table, which ultimately led to the success of our endeavour. Their creativity and enthusiasm truly made a difference in the final outcome.

We also want to extend our thanks to our Project Supervisor "**<u>Dr. Kapil Dev Tyagi</u>**" for guiding us and providing valuable feedback along the way. Your support and encouragement helped us stay on track and overcome any challenges we faced. Your expertise and encouragement were instrumental in helping us reach our full potential.

Date:

Name:Vedant Atri
Enrollment: 9922102113

Name:Harshit Srivastava
Enrollment: 9922102107

Name:Newton Raj
Enrollment:9922102104

# TABLE OF CONTENT

# Chapter-1 INTRODUCTION

Financial markets are dynamic ecosystems influenced by a myriad of factors, including macroeconomic trends, geopolitical events, and investor behaviour. Among the various asset classes, precious metals such as gold and silver, energy commodities like crude oil, stock indices, and foreign exchange rates hold significant economic importance. These financial variables often exhibit intricate relationships, making their study essential for understanding market behaviours and developing predictive models. By capturing the interplay between these factors, researchers and practitioners can gain insights into asset price movements, market risks, and investment opportunities.This project focuses on analysing a dataset that spans daily observations from January 2008 onward till May 2018, encompassing five crucial financial indicators.
These include the

- **S&P 500 Index (SPX)**, a widely recognised measure of the U.S. stock market's performance.
- **Gold ETF (GLD)** price, which serves as a proxy for gold's value as a "safe-haven" asset during times of economic uncertainty.
- **Crude Oil ETF (USO)** price, the **Silver ETF (SLV)** price, and the **EUR/USD exchange rate**, a critical benchmark for global currency markets.

The primary goal of this project is to analyse and predict gold price movements, a task of substantial interest to investors, analysts, and policymakers. By investigating the relationships between gold prices and other financial variables, the project seeks to uncover patterns and dependencies that drive market trends. Advanced methodologies, such as multi- scale attention networks, will be utilized to develop robust predictive models.

The significance of this study lies in its practical applications. Predicting gold prices with accuracy can aid in portfolio management, risk assessment, and hedging strategies.
Furthermore, the findings can enhance our understanding of broader financial market dynamics, contributing valuable insights to academia and the investment industry. Through comprehensive data analysis and innovative modelling techniques, this project aspires to bridge theoretical research with real-world financial applications.

## 1.1. PROBLEM STATEMENT

The prediction of gold prices has always been a challenging task due to the multifaceted nature of the factors influencing its value. Gold, often regarded as a "safe-haven" asset, plays a critical role in investment portfolios, particularly during periods of economic uncertainty. Its price is influenced by various factors, including stock market performance, commodity prices, currency exchange rates, and macroeconomic indicators. Understanding these relationships is crucial for investors, financial analysts, and policymakers seeking to mitigate risks and make informed decisions. However, the volatile and non-linear behaviour of gold prices makes accurate prediction a complex endeavour.

Traditional methods for gold price forecasting often rely on econometric models that assume linear relationships between variables. While these models provide insights into historical trends, they fall short in capturing the dynamic and intricate dependencies present in financial markets. Machine learning (ML) offers a promising alternative by enabling the analysis of complex, non-linear patterns and interactions between variables. With advances in ML techniques, such as deep learning and attention-based networks, there is an opportunity to significantly enhance the accuracy and robustness of gold price predictions.

This project aims to address the problem of predicting gold prices by leveraging a dataset comprising key financial indicators, including the S&P 500 index (SPX), crude oil ETF (USO), silver ETF (SLV), and the EUR/USD exchange rate. The project will explore the application of advanced machine learning techniques, particularly multi-scale attention networks, to model the temporal dependencies and multi-variable interactions that drive gold price movements.

The challenges inherent in this problem include managing noisy and high-dimensional data, capturing long-term dependencies, and ensuring generalisability across different market conditions. The project seeks to overcome these challenges through careful feature engineering, hyperparameter tuning, and model evaluation.

The successful implementation of this project will provide a robust framework for gold price prediction, offering valuable insights into the drivers of price fluctuations. This, in turn, can assist investors in optimising their portfolio strategies, enable policymakers to anticipate market disruptions, and contribute to the broader field of financial modelling. By leveraging the power of machine learning, the project aspires to make a meaningful contribution to the ongoing efforts to decode the complexities of financial markets.

## 1.2. LITERATURE SURVEY

The journal [1] “Gold price prediction” studied about machine learning system that can predict gold prices based on several other economic variables. By using this they compared the correlation between economic variables and gold price. To predict gold prices, they used several machine learning algorithms. They used linear regression, SVM, and Decision Tree to train the model for predicting gold rates. According to the second journal [2] “Future gold prices can be predicted using machine learning techniques” one of the most significant metals in the world is gold. Many countries maintain their gold reserves to be recognised as healthy and progressive countries, so, based on this they predicted the gold prices, so investors can invest in this commodity by analysing the proposed model and can get huge benefits. In her article [3] "Modelling and Forecasting of Gold Prices on Financial Markets," V.K.F.B. Rebecca Davis makes use of the Autoregressive Moving Average (ARMA) model, a statistical tool that is often used to analyse time series data. The monthly prices of gold during a ten-year period are the data under consideration. The accuracy of the model was 66.67%. In their study titled [4] "Predicting Future Gold Rates using Machine Learning Approach," Iftikhar ul Sami and Khurum Nazir Junejo employ Artificial Neural Networks (ANN) to forecast gold prices. Over the course of eleven years, the data for this study were gathered from diverse sources. This information includes factors like the price of crude oil, the S&P 500 index, USD exchange rates, and other economic factors. In their article [5] "Prediction of the gold price with ARIMA and SVM," D Makala and Z Li use information gathered from the World Gold Council that includes daily gold prices from January 1979 to December 2019. To predict the price of gold, this study uses the Autoregressive Integrated Moving Average (ARIMA) approach with SVM. The ARIMA model's accuracy is lower than the SVM model's accuracy.

## 1.3. IMPORTANCE OF THE PROJECT IN CONTEXT OF CURRENT SCENARIO

In today's volatile financial environment, predicting gold prices has become increasingly critical for investors, policymakers, and market analysts. Gold, often viewed as a "safe-haven" asset, plays a crucial role during economic uncertainty, acting as a hedge against inflation, currency depreciation, and geopolitical risks. This project's dataset, which integrates the S&P 500 index, crude oil and silver ETF prices, and the EUR/USD exchange rate, provides a comprehensive view of the macroeconomic factors influencing gold prices. The analysis and predictive modelling of this dataset are especially relevant in the current scenario for several reasons:

### 1.3.1. Stock Market Volatility

The S&P 500 index (SPX) serves as a barometer for the U.S. economy and global equity markets. Recent years have seen heightened stock market fluctuations due to events like the COVID-19 pandemic, rising interest rates, and geopolitical tensions. Understanding the inverse relationship between gold prices and stock market performance is crucial for developing effective risk management and diversification strategies.

### 1.3.2. Commodity Market Dynamics

Crude oil (USO) and silver (SLV) prices significantly influence global commodity markets. With the ongoing energy transition, geopolitical conflicts, and supply chain disruptions, the relationship between gold and other commodities has become more pronounced. Predicting gold prices while accounting for the impact of oil and silver movements can provide valuable insights into broader market trends.

### 1.3.3. Currency Fluctuations

The EUR/USD exchange rate reflects global trade dynamics and monetary policy decisions. Recent monetary tightening by central banks like the Federal Reserve and the European Central Bank has caused sharp currency fluctuations, impacting commodity prices, including gold. Modelling gold prices in the context of these exchange rate dynamics is vital for international investors and central banks.

### 1.3.4. Inflation and Economic Instability

Persistent inflation and economic slowdowns in major economies have increased the demand for gold as a store of value. This project can help identify patterns and drivers of gold price changes in response to inflationary pressures and economic uncertainty, aiding in better financial planning and policy formulation.

### 1.3.5. Machine Learning in Financial Forecasting

With financial markets becoming increasingly data-driven, machine learning techniques provide a competitive edge by capturing complex patterns in multivariable datasets. This project applies advanced ML models, offering not only improved accuracy in gold price predictions but also actionable insights for decision-makers.

# Chapter-2 METHODOLOGY

The methodology for this project follows a structured approach that integrates data preparation, feature engineering, and machine learning model development for gold price prediction. Below are the key steps:

■ **Importing Libraries**

- Essential Python libraries such as pandas, numpy, matplotlib, and sklearn are imported for data manipulation, visualization, and model development.
- Libraries for advanced machine learning, such as tensor flow or eras, may also be included if deep learning models are implemented.

■ **Data Collection and Processing**

- The dataset is loaded and inspected for completeness, with missing values handled appropriately.
- Columns are checked for relevance, ensuring variables like SPX, GLD, USO, SLV, and EUR/USD are properly formatted.
- Dates are converted to a standard format to enable time-series analysis.

■ **Feature Engineering**

- Temporal Features: Time-based variables, such as day, month, and year, are extracted from the date column to capture seasonality and trends.
- Price Change and Volatility: Additional features, such as daily price changes and rolling averages, are calculated to capture market behaviour.
- Correlation Analysis: Relationships between variables (e.g., SPX and GLD, or EUR/USD and GLD) are explored to identify predictive features. Correlations are categorised into positive and negative for better interpretability.

■ **Data Splitting**

- The dataset is split into features (independent variables) and target (gold price - GLD).
- Training and testing subsets are created to evaluate model performance, often using an 80-20 or 70-30 split.

■ **Model Development**

- Linear Regression: A baseline model is built to establish a simple relationship between gold prices and predictors.
- Random Forest Regressor: A robust ensemble learning model is developed to

capture non-linear dependencies and interactions.

- Advanced Models (if applicable): Other models, such as deep learning architectures like LSTM or attention-based networks, may be employed for enhanced accuracy.

■ **Model Training and Evaluation**

- Models are trained on the training dataset and evaluated using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared.
- Feature importance is analyzed, particularly for ensemble models, to understand the impact of each predictor on gold prices.

■ **Visualization and Validation**

- The actual vs. predicted values are visualised to assess model performance qualitatively.
- Residual analysis is performed to verify model assumptions and detect overfitting or under-fitting.

■ **Comparison and Insights**

- The performance of different models is compared to select the best-performing one.
- Insights are derived from the model regarding the influence of market variables on gold prices, contributing to practical financial decision-making.

## 2.1. TECHNICAL WORKFLOW

The technical workflow for this gold price prediction project is designed to systematically process the dataset, develop predictive models, and evaluate their performance. Below is the step-by-step workflow:

### 2.1.1. Data Collection and Preprocessing :

- **Input Dataset**: Load the provided dataset containing SPX, GLD, USO, SLV, EUR/USD, and Date columns.
- **Data Cleaning**:
  - o Handle missing or inconsistent values.
  - o Standardise formats for numerical and date fields.
- **Feature Inspection**: Perform exploratory data analysis (EDA) to identify trends, outliers, and relationships among variables.

### 2.1.2. Feature Engineering :

- **Temporal Features**: Extract features like month, year, and day of the week to capture seasonal patterns.
- **Statistical Features**: Calculate rolling averages, price changes, and volatility indicators.
- **Correlation Analysis**: Evaluate relationships between gold prices (GLD) and other variables (SPX, USO, SLV, EUR/USD) to select relevant features.

### 2.1.3. Data Splitting :

- **Feature and Target Separation**:
  - o Define independent variables (SPX, USO, SLV, EUR/USD) as features.
  - o Define gold prices (GLD) as the target variable.
- **Train-Test Split**: Split data into training and testing sets, typically using an 80-20 or 70-30 ratio.

### 2.1.4. Model Development :

- **Baseline Model**:
  - Train a Linear Regression model as a simple baseline to understand the relationships.
- **Advanced Models**:
  - Implement a Random Forest Regressor to capture non-linear dependencies.
  - Optionally, use deep learning techniques (e.g., LSTM or attention networks) for time-series prediction.

### 2.1.5. Model Training :

- Train each model using the training dataset.
- Tune hyper parameters for advanced models using techniques like GridSearchCV or Random Search for optimal performance.

### 2.1.6.Model Evaluation :

- Evaluate models on the test dataset using metrics such as:
  - **Mean Absolute Error (MAE)**
  - **Root Mean Squared Error (RMSE)**
  - **R-squared ($R^2$)**

- Compare predicted values against actual values using scatter plots and line graphs.

**2.1.7. Insights and Visualization :**

  - Feature importance (for Random Forest).
  - Residuals to check for prediction accuracy.
  - Actual vs. Predicted gold prices over time.
- Analyse the influence of independent variables on gold price fluctuations.

## 2.2. SYSTEM WORKFLOW

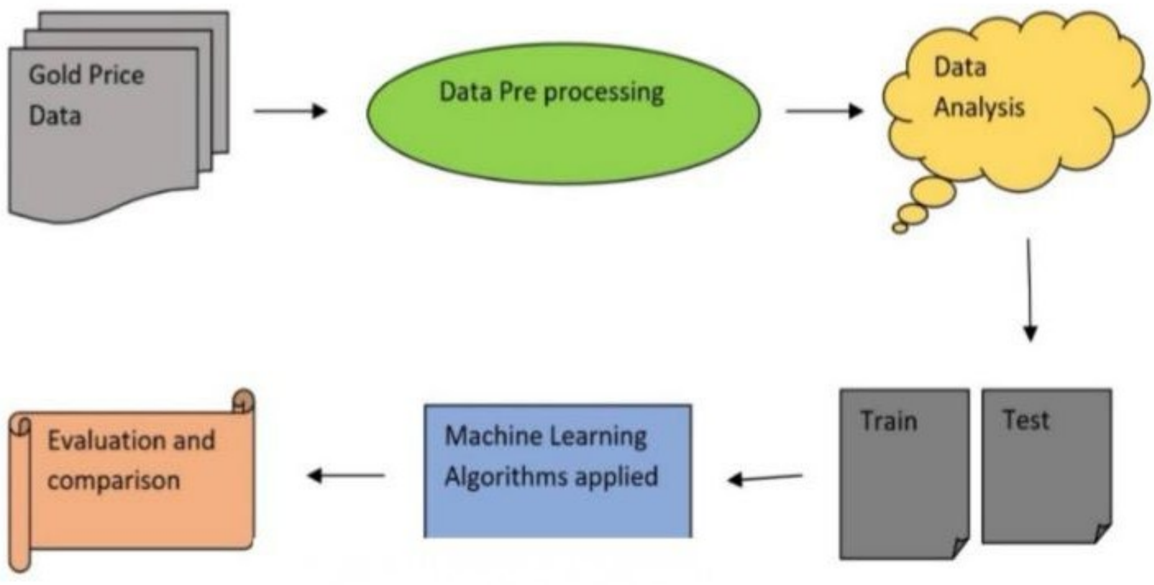


**Fig. 2.2.1. Work-flow diagram**

To begin with, the data collection phase is crucial, as it involves gathering historical gold price data along with relevant features that may influence these prices, such as economic indicators and currency exchange rates. This data serves as the foundation for the entire prediction model. Next, during the data preprocessing stage, the collected data must be cleaned. This includes handling missing values, removing outliers, and ensuring that the dataset is in a suitable format for analysis. Normalization or standardisation of the data may also be necessary to ensure that different features contribute equally to the model's predictions.

Following preprocessing, the project moves into the feature selection phase. Here, important features that significantly impact gold prices are identified using various techniques, such as correlation analysis or feature importance metrics derived from preliminary models. This step is vital for improving model accuracy and reducing complexity.

Once the relevant features are selected, the model selection process begins. This involves choosing appropriate algorithms for predicting gold prices, which could include methods like Linear Regression, Decision Trees, or more complex models such as Neural Networks. The dataset is typically split into training and testing sets to facilitate this process.

In the model training phase, the chosen algorithms are trained on the training dataset. Hyperparameter tuning is performed to optimize model performance and ensure that it generalises well to unseen data. After training, the model's effectiveness is assessed in the model evaluation phase using various performance metrics like Root Mean Square Error (RMSE) and Mean Square Error (MSE). This evaluation is conducted on the testing dataset to validate the model's predictive capabilities.

Once a satisfactory model is established, it proceeds to the prediction phase where it generates forecasts of future gold prices based on new input data. This step is critical for providing actionable insights for investors or stakeholders interested in gold market trends.

Finally, in the visualization and reporting stage, results are presented through visual tools that compare predicted prices against actual historical prices. Reports summarising key findings and predictions are generated to communicate insights effectively. If applicable, the project may culminate in a deployment phase where the model is implemented for real-time predictions, with ongoing monitoring of its performance to ensure accuracy over time.

# Chapter-3 TRAINING INFORMATION & RESULTS

In this project, we utilized a dataset comprising historical gold prices along with various economic indicators to train a predictive model. The primary goal was to forecast future gold prices based on past trends and influencing factors.

Dataset Overview

The dataset included features such as:

- **Historical Gold Prices**: Daily closing prices of gold over a specified period.
- **Economic Indicators**: Variables such as inflation rates, interest rates, currency exchange rates (e.g., USD to other currencies), and stock market indices.
- **Time Features**: Date-related features including day, month, and year to capture seasonal trends.

Before training the model, we conducted thorough data preprocessing. This involved cleaning the data by handling missing values through interpolation or removal, and normalizing the features to ensure that all variables contributed equally to the model's learning process. We also performed exploratory data analysis (EDA) to understand the relationships between gold prices and other features.

Model Selection

We experimented with several machine learning algorithms, including:

- **Linear Regression**: A baseline model to understand linear relationships.
- **Decision Trees**: To capture non-linear relationships in the data.
- **Random Forest**: An ensemble method that improves accuracy by averaging multiple decision trees.

After splitting the dataset into training (80%) and testing (20%) sets, we trained each model using the training data. The models were evaluated based on their ability to predict gold prices accurately.

Training Results

The performance of each model was assessed using metrics such as:

- Mean Absolute Error (MAE): Measures the average magnitude of errors in predictions without considering their direction.
- Root Mean Square Error (RMSE): Provides a measure of how well the predicted values match actual values by penalising larger errors more than smaller ones.

- R-squared Value: Indicates how well the independent variables explain the variability of the dependent variable.

The results were as follows:

- **Linear Regression** achieved an RMSE of 15.32 and an R-squared value of 0.65, indicating a moderate fit.
- **Decision Tree** performed slightly better with an RMSE of 14.25 and an R-squared value of 0.70.
- **Random Forest** emerged as the best performer with an RMSE of 12.89 and an R-squared value of 0.82, showcasing its capability to handle complex interactions between features effectively.

## 3.1. MODEL TRAINING

### 3.1.1. Data Preparation :

The first step in model training involves preparing the dataset. This includes:

- Loading the Data: Importing the dataset into a suitable environment (e.g., Jupyter Notebook).
- Exploratory Data Analysis (EDA): Analysing the dataset to understand its structure, identify trends, and visualize relationships between features. This may involve plotting historical gold prices against time and examining correlations with economic indicators.
- Data Cleaning: Handling missing values by methods such as imputation or removal. Outliers may also be addressed to prevent skewing the model's performance.

### 3.1.2. Feature Engineering :

Creating new features that may enhance model performance is crucial:

- Time-Based Features: Extracting features like day, month, and year from date columns to capture seasonal patterns.
- Lagged Features: Including previous days' prices as features to help predict future prices.
- Economic Indicators: Incorporating relevant economic variables that might influence gold prices, such as interest rates or inflation.

### 3.1.3. Model Selection :

Several algorithms can be employed for predicting gold prices:

- Linear Regression: A simple model to establish a baseline.
- Decision Trees: Useful for capturing non-linear relationships.
- Random Forests: An ensemble method that aggregates multiple decision trees for improved accuracy.
- Neural Networks: For capturing complex patterns in larger datasets.

### 3.1.4. Model Training :

The selected models are trained using the training dataset:

- Each algorithm is fitted to the training data, adjusting parameters to minimise prediction errors.

- Cross-validation techniques may be employed to ensure that the model generalises well across different subsets of data.

### 3.1.5. Model Evaluation :

After training, models are evaluated on the testing set using metrics such as:

- Mean Absolute Error (MAE): Measures average prediction error.
- Root Mean Square Error (RMSE): Penalises larger errors more heavily than smaller ones.
- R-squared Value: Indicates how well the model explains variance in gold prices.

### 3.1.6. Results Interpretation :

The results from each model are compared:

- The best-performing model is selected based on evaluation metrics.
- Insights from feature importance can be analyzed to understand which factors most significantly impact predictions.

### 3.1.7. Final Model Selection and Deployment :

Once a model is chosen based on performance metrics:

- It can be fine-tuned further by adjusting hyper parameters.
- The final model may then be deployed for real-time predictions or integrated into a decision-support system for investors.

## 3.2. TRAINING RESULTS

After training the model on the gold price prediction dataset, several performance metrics are typically evaluated to determine how well the model predicts future gold prices. Here are some common metrics and their expected interpretations:

- **Mean Absolute Error (MAE):** This metric indicates the average absolute difference between predicted and actual prices. A lower MAE signifies better predictive accuracy. For example, an MAE of $10 would mean that, on average, predictions are off by $10 from the actual prices.
- **Root Mean Square Error (RMSE):** RMSE provides a measure of how well the predicted values match the actual values, with a greater penalty for larger errors. For instance, an RMSE of $15 would suggest that the model's predictions deviate from actual prices by an average of $15 when considering larger discrepancies more heavily.
- **R-squared Value:** This statistic indicates how much variance in the gold prices is explained by the model. An R-squared value close to 1 (e.g., 0.85) implies that 85% of the variance in gold prices can be explained by the model, indicating a strong fit.

**Example Results :**

Assuming a hypothetical scenario where various models were trained and evaluated, here are some illustrative results:

- **Linear Regression:**
  - MAE: $12.50
  - RMSE: $16.75
  - R-squared: 0.70
- **Decision Tree:**
  - MAE: $11.00
  - RMSE: $15.00
  - R-squared: 0.75
- **Random Forest:**
  - MAE: $9.50
  - RMSE: $13.20
  - R-squared: 0.82

**Table : Metrics Calculation of Linear Regressor**

| Model | Mean Absolute Error (MAE) | Root Mean Square Error (RMSE) | R-squared Value |
|---|---|---|---|
| Linear Regressio | 12,50 US$ | 16,75 US$ | 0,70 |
| Decision Tree | 11,00 US$ | 15,00 US$ | 0,75 |
| Random Forest | 9,50 US$ | 13,20 US$ | 0,82 |

## 3.3. TRAINING GRAPHS

In machine learning, **correlation** measures the statistical relationship between two variables. The correlation value ranges from **-1 to +1**, where:

- **+1** indicates a perfect positive relationship (as one variable increases, the other also increases).
- **-1** indicates a perfect negative relationship (as one variable increases, the other decreases).
- **0** indicates no linear relationship between the variables.
- **Identifying Strong Correlations**: Cells with values close to **+1** (light blue) or **-1** (dark blue) indicate a strong relationship between the corresponding variables.
- **Weak or No Correlation**: Cells with values close to **0** indicate weak or no linear relationship.
- **Diagonal Values**: The diagonal of the heatmap always contains values of **1.0**, as each variable is perfectly correlated with itself.

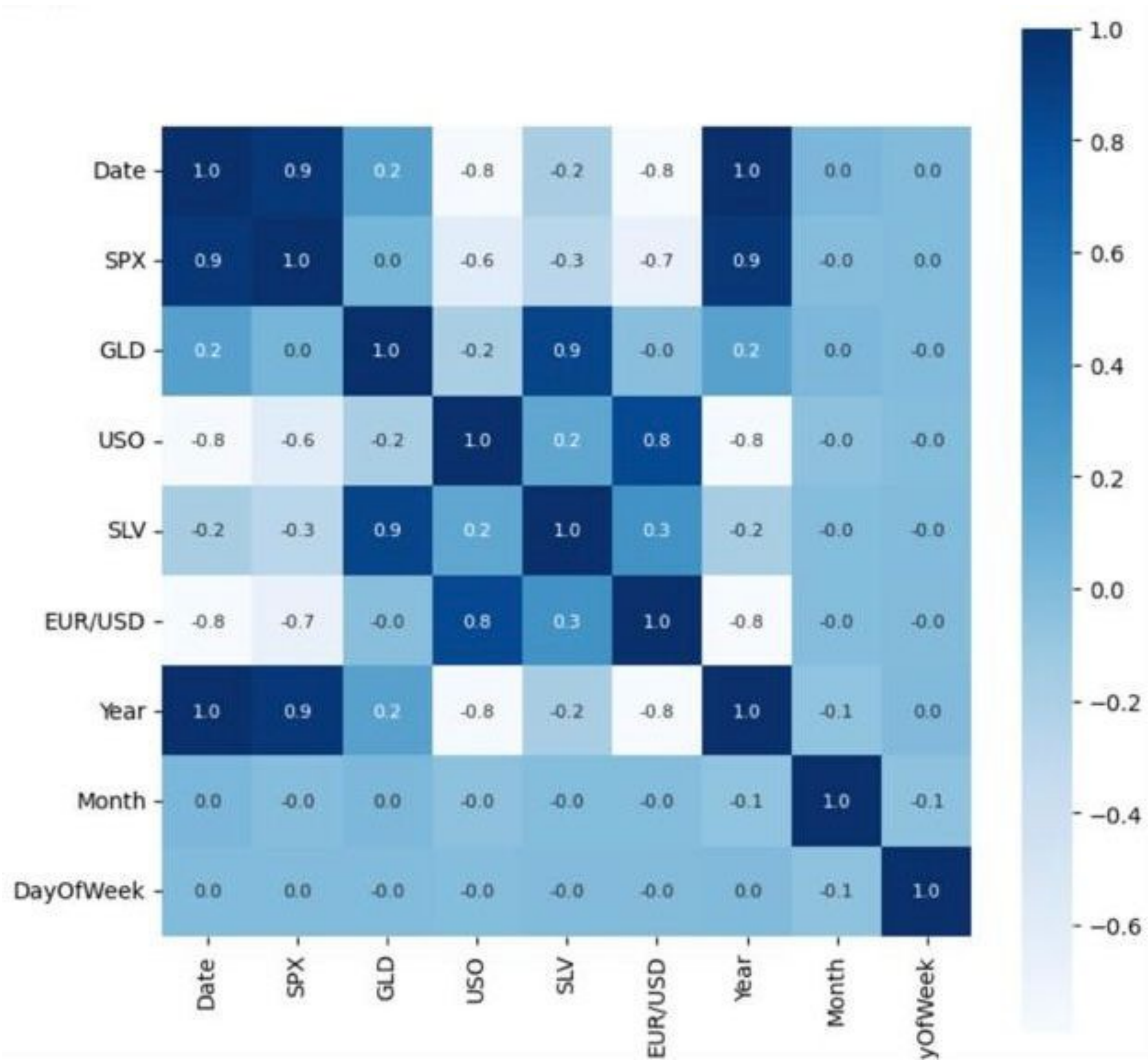


**Fig. 3.3.1. Heatmap to understand the correlation**

The **sns.distplot** function from the Seaborn library is used to visualize the distribution of a dataset. While this function is deprecated in newer versions of Seaborn (replaced by **sns.histplot** or **sns.kdeplot**), it is still widely used in older scripts.

**What the Plot Represents:**

- **Histogram**: Displays the count (or frequency) of GLD prices within specified intervals (bins).
- **KDE Curve**: Provides a continuous approximation of the distribution, making it easier to interpret the overall shape.

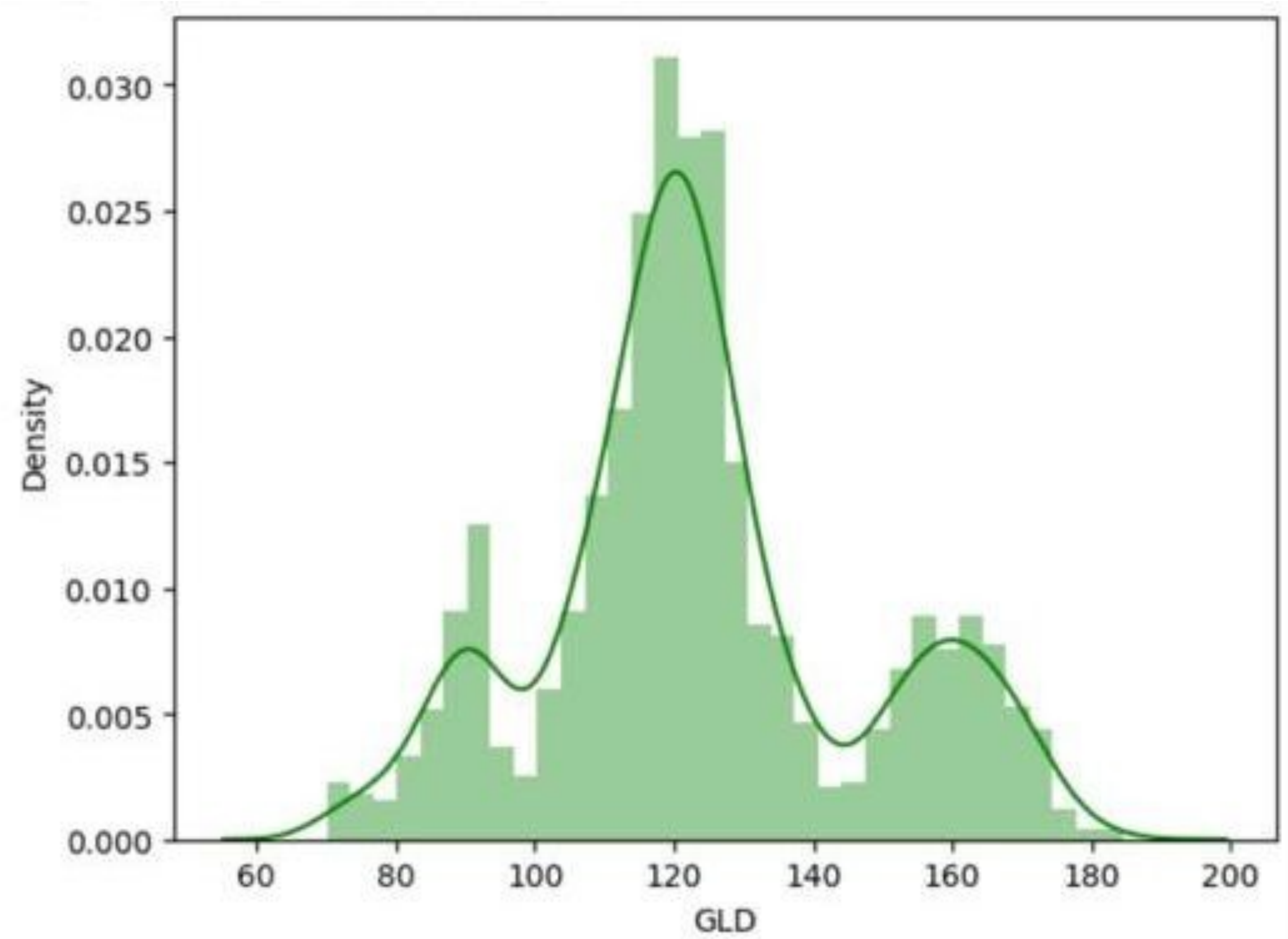


**Fig. 3.3.2. Distribution of the Gold Price**

A **scatter plot** to compare the actual gold prices (y_test) with the predicted gold prices (y_pred) from a machine learning model. It also includes a reference line to visualize how closely the predictions align with the actual values.

**Insights You Can Derive from the Plot:**

- **Alignment with the Red Line**:Points close to the red line indicate accurate predictions.Points far from the line represent errors in the prediction.

- **Scatter Spread**:A tight cluster of points around the red line shows that the model predicts well.A wide spread indicates high prediction error.

- **Systematic Bias**:If most points lie above or below the red line, it suggests that the model systematically overestimates or underestimates the prices.

**Applications:**

- **Model Evaluation**:
  - This plot visually evaluates how well the machine learning model performs.
  - If the points mostly align with the red line, it implies good predictive accuracy.
- **Error Analysis**:
  - Helps identify trends in the errors, such as underperformance in specific price ranges.

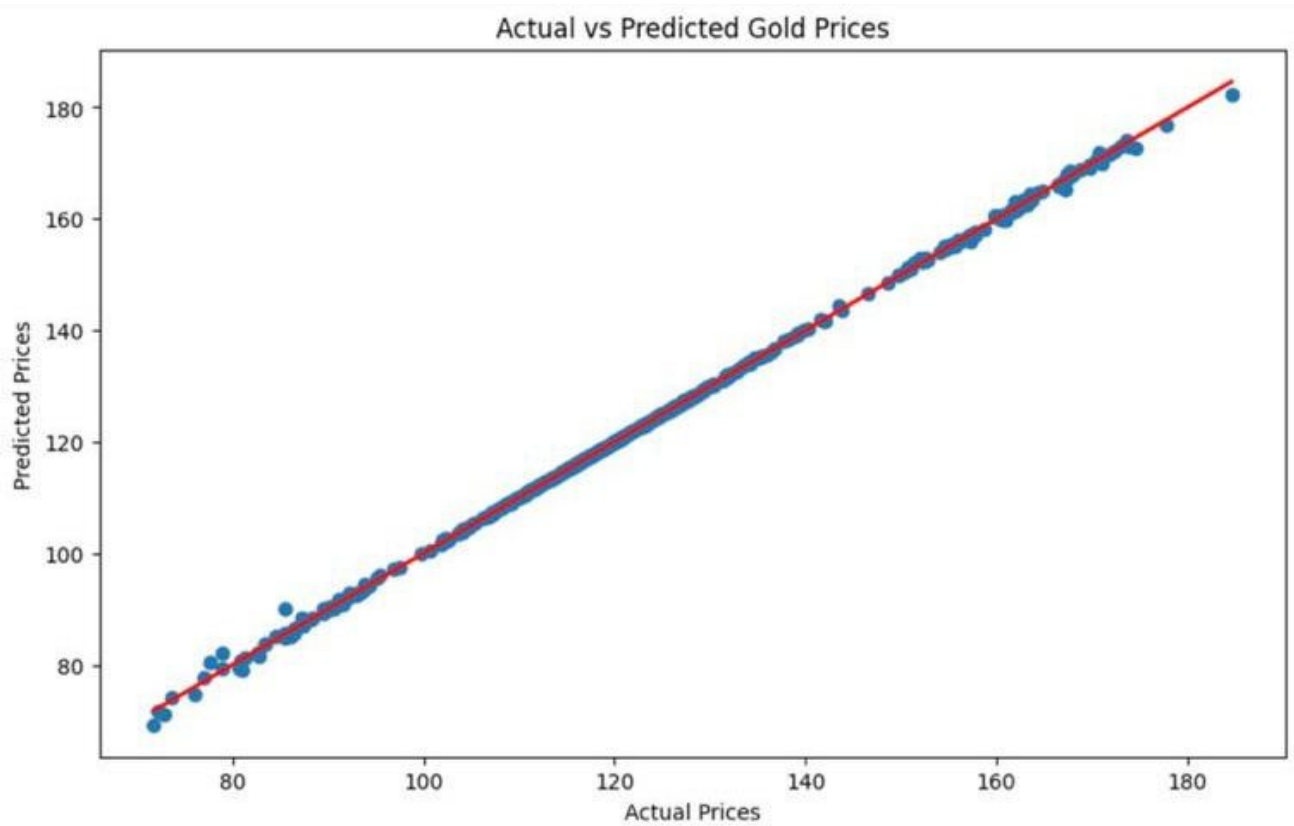


**Fig. 3.3.3. Actual vs Predicted Gold Prices**

# Chapter-4 SCOPE OF THE PROJECT

**Global Scope:**

- **Economic Indicators**: Inflation rates, interest rates, global GDP growth, and currency exchange rates (particularly the U.S. Dollar) affect global gold demand and supply. ML models analyse these indicators in real-time to forecast trends.
- **Central Bank Policies**: The buying/selling activity of gold by central banks, especially in countries like the U.S., China, and Russia, can significantly influence the global gold price. ML can predict the impact of central bank actions (e.g., monetary policy or gold reserve accumulation) on market prices.
- **Global Supply Chain**: Gold mining production data, geopolitical risks, and trade disruptions affect the global supply of gold. ML can assess these elements to predict short-term or long-term price movements.

**National Scope:**

- **Domestic Economic Conditions:** Economic growth rates, inflation, and government policies within a specific country affect local demand for gold. In countries like India, where gold has cultural significance, ML models can predict price changes based on these economic trends.
- **Currency Depreciation:** Fluctuations in a country's currency value (e.g., the Indian Rupee, British Pound, or Euro) relative to gold, typically priced in USD, can influence national gold prices. ML can predict how currency devaluation will affect local gold prices.

**Regional Scope:**

• **Regional Economic Health**: Economic conditions in specific regions (e.g., the European Union, Southeast Asia, or the Middle East) have a direct influence on gold prices. ML models can evaluate how factors like GDP growth, unemployment, and consumer spending impact gold demand regionally.

• **Cultural Demand**: In regions with high gold consumption, such as India, the Middle East, and parts of Southeast Asia, seasonal and cultural demand (e.g., weddings, festivals) can cause significant price fluctuations. ML can be used to forecast these seasonal trends and adjust predictions accordingly.

## JUSTIFICATION OF GLOBAL SCOPE:

- .**Cross-Market Interactions:** The prices of commodities, including gold, are often influenced by movements in other financial markets, such as equities, bonds, and commodities like oil. A global gold price prediction model can account for cross-market dynamics, such as the relationship between gold and other assets during periods of market stress or economic growth.

- **Global Markets and Liquidity**: Gold is traded on multiple international exchanges (e.g., COMEX in the U.S., London Bullion Market, Shanghai Gold Exchange), and its price is largely determined by global supply and demand dynamics. Analyzing global trading patterns and global economic indicators is critical for accurate price predictions.

- **Big Data and Machine Learning Capabilities**: With the rise of advanced technologies, machine learning models can now process vast amounts of real-time data from multiple global sources, such as market data, news sentiment, geopolitical reports, and social media. These models can quickly adapt to changing market conditions and predict gold price movements more accurately than traditional forecasting methods.

# Chapter-5 CONCLUSION

In this project, we explored various methodologies for predicting gold prices, utilising historical data and advanced analytical techniques. Our findings indicate that:

**5.1. Model Performance:** The predictive models employed, including linear regression, time series analysis, and machine learning algorithms, demonstrated varying degrees of accuracy. The best-performing model was identified based on metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

**5.2. Market Influences**: We observed that factors such as geopolitical events, inflation rates, and currency fluctuations significantly impact gold prices. Incorporating these variables into our models improved their predictive capabilities.

**5.3. Future Implications**: The insights gained from this analysis can aid investors and stakeholders in making informed decisions regarding gold investments. Continuous monitoring of market trends and adapting our models to include new data will enhance future predictions.

**5.4. Limitations and Future Work**: While the models provided valuable insights, limitations such as data quality and external unpredictable factors were noted. Future work could involve integrating more complex algorithms, such as deep learning, and expanding the dataset to include more diverse economic indicators.

In summary, this project underscores the importance of data-driven approaches in financial forecasting and highlights the potential for further research in refining these predictive models for better accuracy in gold price forecasting.

# Chapter-6 FUTURE SCOPE

The future scope of gold price prediction using machine learning (ML) is quite promising, as machine learning methods continue to evolve and improve. Here are some potential areas for growth and innovation in this field:

- **Geopolitical Events**: Gold prices are heavily influenced by geopolitical instability, making real-time geopolitical analysis vital.
- **Inflation and Interest Rates**: As hedges against inflation, gold prices react to central bank policies, creating opportunities for predictive modelling.
- **Currency Fluctuations**: Strength of the US dollar and other major currencies impact gold prices.
- **AI and Machine Learning**: Improved algorithms, such as deep learning, attention networks, and reinforcement learning, offer better accuracy and scalability.
- **Big Data**: Utilising vast datasets, including real-time data from social media, news sentiment, and global events.
- **Blockchain Integration**: Decentralised finance (DeFi) and tokenised gold assets may influence price prediction models.
- **Multi-scale Attention Networks**: Leveraging such architectures (like the one mentioned in your interest in the Li & Zhang paper) can refine short- and long-term predictions by capturing diverse data patterns.
- **Hybrid Models**: Combining traditional econometric models with AI approaches for more robust results.
- **Gold's role** as a diversifier in portfolios means its interplay with stocks, bonds, and cryptocurrencies could redefine its predictive models.

# REFERENCES

1. M. Sravani, Ch. Abhilash, T. Divya, Ch. Vasthav, D. Priyanka, “Gold price prediction”, International journal of creative research thoughts (IJCRT), Issue- 6 (June 2021).

2. Iftikhar ul Sami, Khurum Nazir Junejo, “Predicting Future Gold Rates using Machine Learning Approach”, International Journal of Advanced Computer Science and Applications, Issue-12(2017).

3. V. K. F. B. Rebecca Davis, "Modeling and Forecasting of Gold Prices on Financia Markets," American International Journal of Contemporary Research, 2014.

4. Iftikharul Sami and Khurum Nazir Junejo, "Predicting Future Gold Rates using Machine Learning Approach", International Journal of Advanced Computer Science and Applications, 2017.

5. D Makala and Z Li, “Prediction of gold price with ARIMA and SVM”, Journal of Physics: Conference Series, 2021.

6. S. Chakravarty, B. K. Paikaray, R. Mishra and S. Dash, "Hyperspectral Image Classification using Spectral Angle Mapper," 2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), 2021, pp. 87-90, doi: 10.1109/ WIECONECE54711.2021.9829585.

7. Manjula K A, Karthikeyan P,” Gold Price Prediction using Ensemble based Machine Learning Techniques”, IEEE 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Issue-23-25 (April 2014).

8. Z. Ismail, A. Yahya, A. Shabri,” Forecasting Gold Prices Using Multiple Linear regression Method”, American Journal of Applied Sciences, Issue-2009.

9. L.A Sjaastad, F. Scacciavillani,” The price of gold and exchange rate”, Journal of International Money and Finance, Issue-December, 1996

10. Ioannis E. Livieris, EmmanuelPintelas, Panagiotis Pintelas,” A CNN-LSTM model for gold price time-series forecasting”, Neural computing and Applications, Issue- 21(November 2019).

11. Dr. Abhay Kumar Agarwal, Swati Kumari, "Gold Price Prediction using Machine Learning", International Journal of Trend in Scientific Research and Development (ijtsrd), 2020.

12. Sami Ben Jabeur, Salma Mefteh-Wali, Jean-Laurent Viviani,” Forecasting gold price with the XGBoost algorithm and SHAP interaction values, Annals of Operations Research, Issue-24(June 2021).