GOLD PRICE PREDICTION SYSTEM

INNOVATIVE / MULTI-DISCIPLINARY

PROJECT REPORT

Submitted by

PRAKASH RAJ S	**722820104085**
PRAVIN D	**722820104092**
SABAREESWARAN G	**722820104108**
SANJITH R K	**722820104115**

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

SRI ESHWAR COLLEGE OF ENGINEERING

(AN AUTONOMOUS INSTITUTION) COIMBATORE – 641 202

MAY 2023

BONAFIDE CERTIFICATE

Certified that this project report **"GOLD PRICE PREDICTION SYSTEM**" is the bonafide work of

PRAKASH RAJ S	**722820104085**
PRAVIN D	**722820104092**
SABAREESWARAN G	**722820104108**
SANJITH R K	**722820104115**

who carried out the project work under my supervision

…… ……

SIGNATURE **SIGNATURE**

Dr. R. SUBHA M.E., Ph. D.,
Professor & Head,
Dept. of Computer Science & Engineering,
Sri Eshwar College of Engineering,
Coimbatore-641 202.

Mr. A. ANANDARAJ, M.E., (Ph.D.,)
PROJECT GUIDE,
Assistant Professor,
Dept. of Computer Science & Engineering,
Sri Eshwar College of Engineering,
Coimbatore-641 202.

Submitted for the End Semester practical examination –
Innovative/Multidisciplinary project work viva-voce held on ……

…….. ……

(Internal Examiner) **(External Examiner)**

DECLARATION

PRAKASH RAJ S	**722820104085**
PRAVIN D	**722820104092**
SABAREESWARN G	**722820104108**
SANJITH R K	**722820104115**

To declare that the project entitled "GOLD PRICE PREDICTION SYSTEM", submitted in partial fulfillment to University as the project work of Bachelor of Engineering (Computer Science and Engineering) Degree, is a record of original work done by us under the supervision and guidance of Mr. A. Anandaraj, Assistant Professor, Department of Computer Science and Engineering, Sri Eshwar College of Engineering, Coimbatore.

Place: Coimbatore

Date: [PRAKASH RAJ S]

[PRAVIN D]

[SABAREESWARAN G]

[SANJITH R K]

Project Guided by,

[Mr. A. ANANDARAJ, AP/CSE]

ACKNOWLEDGEMENT

We thank God for his grace and blessing that have been showered upon us to bring success for this project. The success we have accomplished in our project would not have been possible without the timely help and guidance rendered to us by many people. We take this opportunity to thank them all.

We wish to express our deep gratitude **Mr. R. Mohanram**, Chairman, Sri Eshwar College of Engineering, Kondampatti for being kindly permitted us to carry on the project. He has an abiding interest in our wellbeing.

We are grateful to **Dr. Sudha Mohanram M.E., Ph.D.,** Principal, Sri Eshwar College of Engineering, Kondampatti for having made all provision available to us.

We are thankful to **Mr. R. Rajaram**, Director, Sri Eshwar College of Engineering, Kondampatti for all his support to complete the project effectively.

We would like to express our heartfelt gratitude and humble thanks to Head of the Department **Dr. R. Subha M.E., Ph.D.,** for having facilitated us to complete the project successfully.

With the deep sense of gratitude, we extend our earnest and sincere thanks to our project guide **Mr. A. Anandaraj, M.E., (Ph.D.,)** Assistant Professor, Department of Computer Science and Engineering, for his kind suggestion and co-operation for doing the project. Finally, we would like to extend our thanks to our teaching and non-teaching staffs those who helped us in bringing out project successfully.

TABLE OF CONTENT

ABSTRACT

Gold is a widely traded commodity, and its price has significant impacts on global economies. Accurate prediction of gold prices is essential for investment and hedging decisions. Machine learning techniques have emerged as a powerful tool for predicting gold prices. These techniques employ mathematical algorithms to analyze vast amounts of historical data and identify patterns that can be used to make predictions about future prices. Various factors, such as economic indicators, geopolitical events, and supply and demand, can impact the price of gold, and machine learning models can incorporate such variables to make more accurate predictions. While no method can guarantee perfect prediction accuracy, machine learning can help investors make more informed decisions based on data-driven insights. However, these models need to be continually updated and refined to account for new data and changing market conditions.

Keywords: Machine Learning, Prediction, Future Prices, Prediction accuracy, Historical data.

CHAPTER 1

INTRODUCTION

1.1 OBJECTIVE

The objective of gold price prediction in machine learning is to develop accurate models that can predict the future prices of gold using historical data and other relevant factors. The goal is to provide investors with a tool to make informed investment and hedging decisions based on data-driven insights. Machine learning models aim to identify patterns and relationships between various factors that affect gold prices, such as economic indicators, geopolitical events, and supply and demand. Ultimately, the objective is to create models that can help investors and financial institutions manage risk and maximize returns in the commodity trading market.

CHAPTER 2
SYSTEM ANALYSIS AND DESIGN

2.1 EXISTING SCENARIO

In the current scenario, machine learning techniques are being increasingly used to predict gold prices. Various research studies and commercial applications have shown the effectiveness of these techniques in forecasting gold prices with a high degree of accuracy. Different models, such as regression models, decision trees, neural networks, and support vector machines, have been developed and tested for their efficacy in predicting gold prices. These models use a variety of data sources, including economic indicators, news articles, and social media sentiment, to identify patterns and relationships that may impact gold prices. Additionally, ensemble methods, such as stacking and blending, have been developed to combine multiple models for more accurate predictions.

Commercial applications of machine learning for gold price prediction include trading bots and recommendation systems that provide investors with timely and data-driven insights. However, challenges remain, such as the need for continuous refinement of models to account for changing market conditions and the potential for bias or limitations in the data used to train the models. Despite these challenges, the use of machine learning in gold price prediction is expected to continue to grow, as it offers significant potential for improving investment decisions and managing risk in the commodity trading market.

2.2 PROBLEM STATEMENT

The gold price prediction system is to develop models that can accurately forecast future gold prices based on historical data and other relevant factors. The challenge lies in identifying the most significant factors that impact gold prices and creating models that can capture the complexities of the gold market. Another challenge is to continually update and refine the models to account for changing market conditions and new data sources. Additionally, there may be limitations and biases in the data used to train the models, which can affect the accuracy of the predictions. The ultimate goal is to create models that can provide investors with data-driven insights to make informed investment and hedging decisions in the commodity trading market.

LITERATURE SURVEY

Gold is a highly valuable commodity that has been used as a store of value for centuries. Its price is influenced by various factors such as supply and demand, economic conditions, and geopolitical events. Predicting the price of gold accurately is a complex task, and traditional econometric models may not be suitable for this purpose. Machine learning techniques have emerged as a promising approach for predicting gold prices, due to their ability to handle large amounts of data and identify complex patterns in the data.

Machine Learning Techniques for Gold Price Prediction

Several machine learning techniques have been used for gold price prediction. These include artificial neural networks (ANNs), decision trees, support vector regression (SVR), and random forests. ANNs are widely used for time series prediction and can capture nonlinear relationships between variables. Decision trees can be used for feature selection and can handle both categorical and continuous data. SVR is a supervised learning algorithm that can model nonlinear relationships between variables. Random forests are an ensemble learning method that can improve prediction accuracy by combining multiple decision trees.

Data and Feature Selection

The accuracy of gold price prediction models depends on the quality of the data and the selection of relevant features. Historical gold price data, economic indicators, and geopolitical events are commonly used as input data for machine learning models. Feature selection techniques such as principal component analysis (PCA), correlation analysis, and mutual information can be used to select relevant features and reduce the dimensionality of the data.

Performance Evaluation Metrics

The performance of gold price prediction models can be evaluated using various metrics such as mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and R-squared (R2) value. MAE measures the average absolute difference between the predicted and actual prices, while MSE measures the average squared difference between the predicted and actual prices. RMSE is the square root of MSE and provides a measure of the standard deviation of the errors. R2 value measures the proportion of variance in the dependent variable that is explained by the independent variables.

In conclusion, machine learning techniques have shown promise in predicting gold prices due to their ability to handle large amounts of data and identify complex patterns. Several machine learning algorithms such as ANNs, decision trees, SVR, and random forests have been used for gold price prediction. The accuracy of gold price prediction models depends on the quality of the data and the selection of relevant features. Performance evaluation metrics such as MAE, MSE, RMSE, and R2 value can be used to evaluate the performance of these models. However, further research is needed to develop more accurate gold price prediction models that can account for the complex nature of the market.

CHAPTER 3
PROPOSED SOLUTION

3.1 OVERVIEW

The system for gold price prediction works by taking in historical data on various factors that impact gold prices, such as economic indicators and geopolitical events. The data is pre-processed to clean and prepare it for use by the machine learning model. The model is then trained on the pre-processed data to identify patterns and relationships between the input features and the target output, which is the future price of gold. Once the model is trained, it can be used to predict future gold prices based on new input data. This prediction system is a valuable tool for investors and financial institutions who rely on accurate price predictions to make informed investment and hedging decisions in the commodity trading market. The system provides data-driven insights that can help manage risk and maximize returns in a dynamic and rapidly changing market.

3.2 BLOCK DIAGRAM

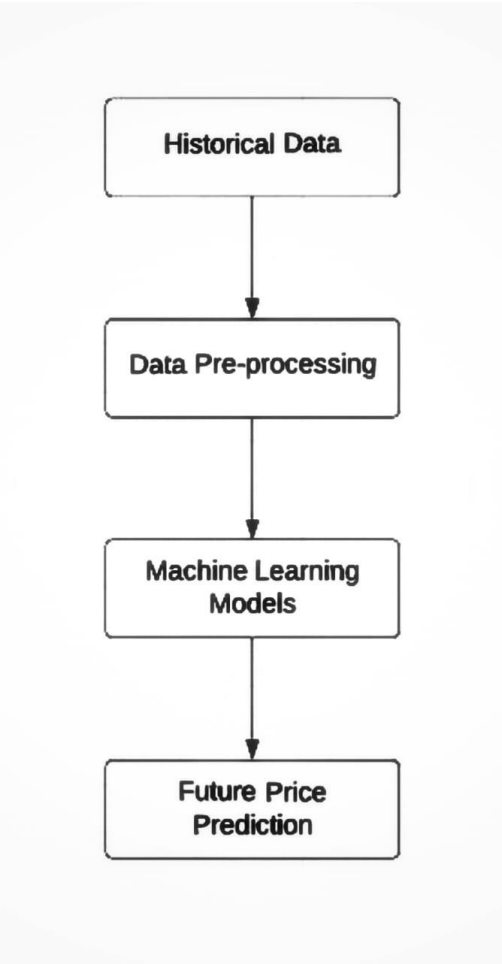


Figure 3.1 – Block Diagram

CHAPTER 4

SYSTEM SPECIFICATION

4.1 HARDWARE REQUIREMENTS

- Processor Type : Core i3
- Speed : 3.40GHZ
- RAM : 4GB DD2 RAM
- Hard disk : 500 GB
- Keyboard : 101/102 Standard Keys
- Mouse : Optical Mouse

4.2 SOFTWARE REQUIREMENTS

- Operating System : Windows 10
- IDE/ Coding Environment : Google Colab
- Coding Language : Python

CHAPTER 5
PROJECT DESCRIPTION

5.1 METHODOLOGY

The followings are the steps involved in the Social Media Follower Prediction System:

- **Data Collection:** Collect historical data on gold prices and relevant factors such as economic indicators, geopolitical events, supply and demand trends, etc.
- **Data Preprocessing:** Clean and preprocess the data to remove any noise, inconsistencies, and outliers, and prepare it for use by the machine learning model. This may involve steps such as data cleaning, data normalization, feature engineering, and data splitting.
- **Model Selection:** Choose the appropriate machine learning algorithm to use for predicting gold prices. This may involve selecting a regression model such as linear regression, decision tree regression, or support vector regression.
- **Model Training:** Train the machine learning model on the preprocessed data. This involves splitting the data into training and testing sets and using the training set to fit the model.
- **Model Evaluation:** Evaluate the performance of the model on the testing set using performance metrics such as mean absolute error (MAE), mean squared error (MSE), and coefficient of determination (R2).
- **Model Tuning:** Optimize the hyperparameters of the model to improve its performance. This may involve techniques such as grid search or random search.
- **Deployment:** Deploy the trained model to make predictions on new, unseen data. This may involve integrating the model with a web application or API.
- **Maintenance and Monitoring:** Continuously update and refine the model to account for new data and changing market conditions, and monitor its performance over time to ensure its accuracy and reliability.

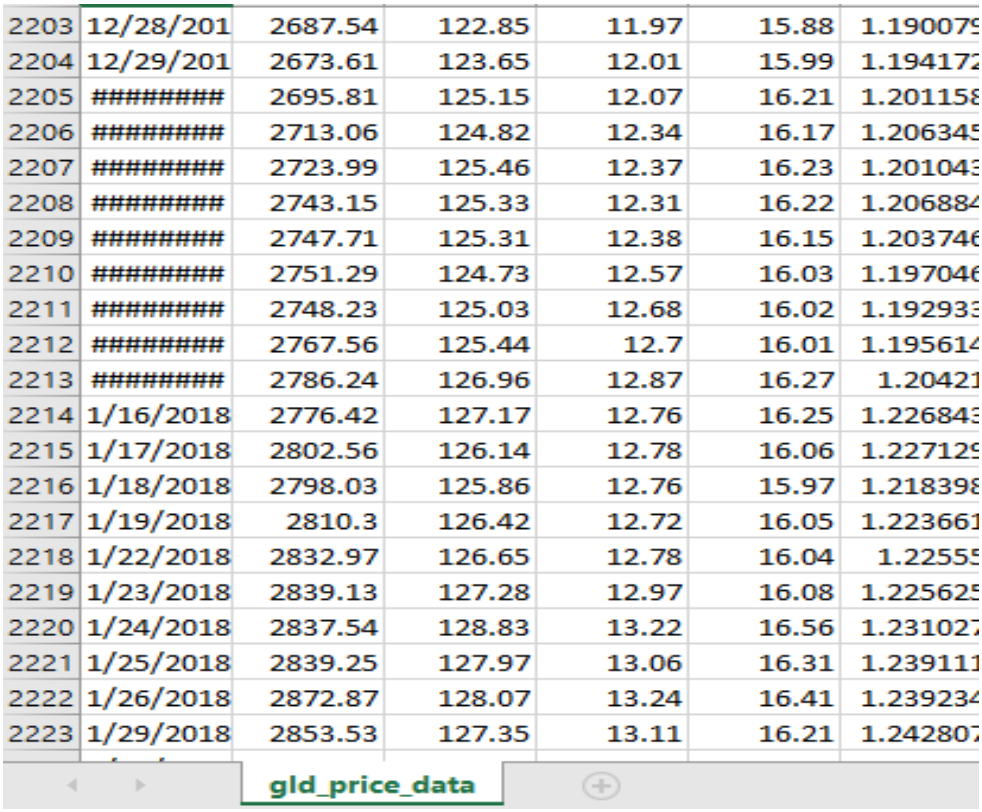
5.2 IMPLEMENTATION

To implement gold price prediction in machine learning, the first step is to collect and preprocess the data. Historical gold price data, economic indicators, and geopolitical events are commonly used as input data for gold price prediction models. Once the data is collected, it needs to be cleaned and preprocessed to remove any missing or inconsistent values. The next step is feature selection, where relevant features are selected to build the prediction model. Feature selection techniques such as principal component analysis, correlation analysis, and mutual information can be used to select relevant features and reduce the dimensionality of the data. After feature selection, a suitable machine learning algorithm needs to be selected to build the prediction model. Several machine learning algorithms such as artificial neural networks, decision trees, support vector regression, and random forests can be used. The performance of the prediction model needs to be evaluated using performance evaluation metrics such as mean absolute error, mean squared error, root mean squared error, and R-squared value. Finally, the prediction model can be deployed to make predictions on new data. However, it is important to note that the prediction accuracy of the model heavily depends on the quality of the data and the selection of relevant features. Therefore, further research is needed to improve the accuracy of gold price prediction models.

CHAPTER 6

RESULT AND IMPLEMENTATION

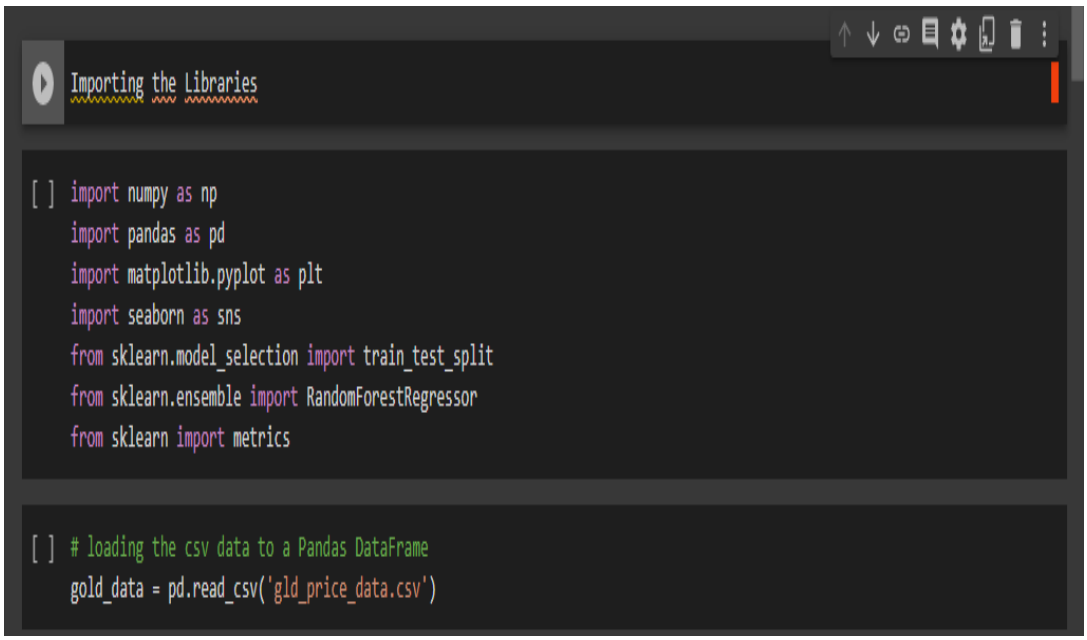
DATASET



2203	12/28/201	2687.54	122.85	11.97	15.88	1.190079
2204	12/29/201	2673.61	123.65	12.01	15.99	1.194172
2205	#########	2695.81	125.15	12.07	16.21	1.201158
2206	#########	2713.06	124.82	12.34	16.17	1.206345
2207	#########	2723.99	125.46	12.37	16.23	1.201043
2208	#########	2743.15	125.33	12.31	16.22	1.206884
2209	#########	2747.71	125.31	12.38	16.15	1.203746
2210	#########	2751.29	124.73	12.57	16.03	1.197046
2211	#########	2748.23	125.03	12.68	16.02	1.192933
2212	#########	2767.56	125.44	12.7	16.01	1.195614
2213	#########	2786.24	126.96	12.87	16.27	1.20421
2214	1/16/2018	2776.42	127.17	12.76	16.25	1.226843
2215	1/17/2018	2802.56	126.14	12.78	16.06	1.227129
2216	1/18/2018	2798.03	125.86	12.76	15.97	1.218398
2217	1/19/2018	2810.3	126.42	12.72	16.05	1.223661
2218	1/22/2018	2832.97	126.65	12.78	16.04	1.22555
2219	1/23/2018	2839.13	127.28	12.97	16.08	1.225625
2220	1/24/2018	2837.54	128.83	13.22	16.56	1.231027
2221	1/25/2018	2839.25	127.97	13.06	16.31	1.239111
2222	1/26/2018	2872.87	128.07	13.24	16.41	1.239234
2223	1/29/2018	2853.53	127.35	13.11	16.21	1.242807

gld_price_data

IMPORTING PACKAGES AND LOADING CSV FILE



Importing the Libraries

```
[ ] import numpy as np
    import pandas as pd
    import matplotlib.pyplot as plt
    import seaborn as sns
    from sklearn.model_selection import train_test_split
    from sklearn.ensemble import RandomForestRegressor
    from sklearn import metrics
```

```
[ ] # loading the csv data to a Pandas DataFrame
    gold_data = pd.read_csv('gld_price_data.csv')
```

GETTING SOME INFORMATION ABOUT DATA

```
[ ] # getting some basic informations about the data
    gold_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2290 entries, 0 to 2289
Data columns (total 6 columns):
 #   Column   Non-Null Count  Dtype  
---  ------   --------------  -----  
 0   Date     2290 non-null   object 
 1   SPX      2290 non-null   float64
 2   GLD      2290 non-null   float64
 3   USO      2290 non-null   float64
 4   SLV      2290 non-null   float64
 5   EUR/USD  2290 non-null   float64
dtypes: float64(5), object(1)
memory usage: 107.5+ KB
```

```
[ ] # checking the number of missing values
    gold_data.isnull().sum()
```

```
Date       0
SPX        0
GLD        0
USO        0
SLV        0
EUR/USD    0
dtype: int64
```

CORRELATION

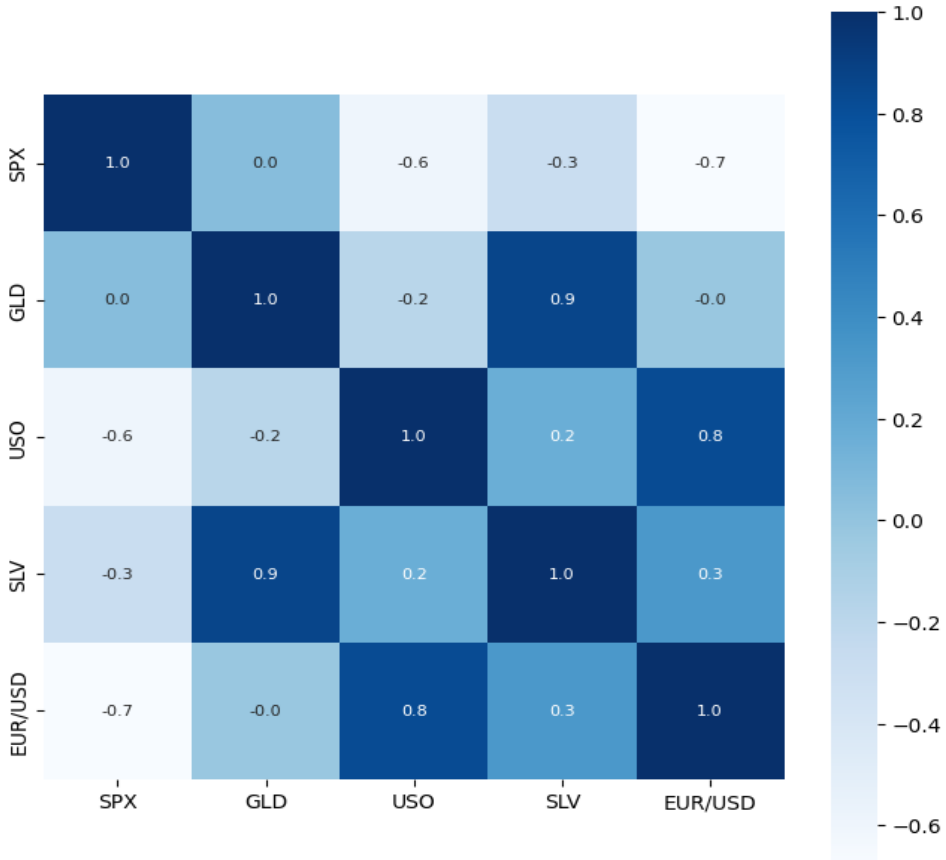
Correlation:

1.Positive Correlation 2.Negative Correlation

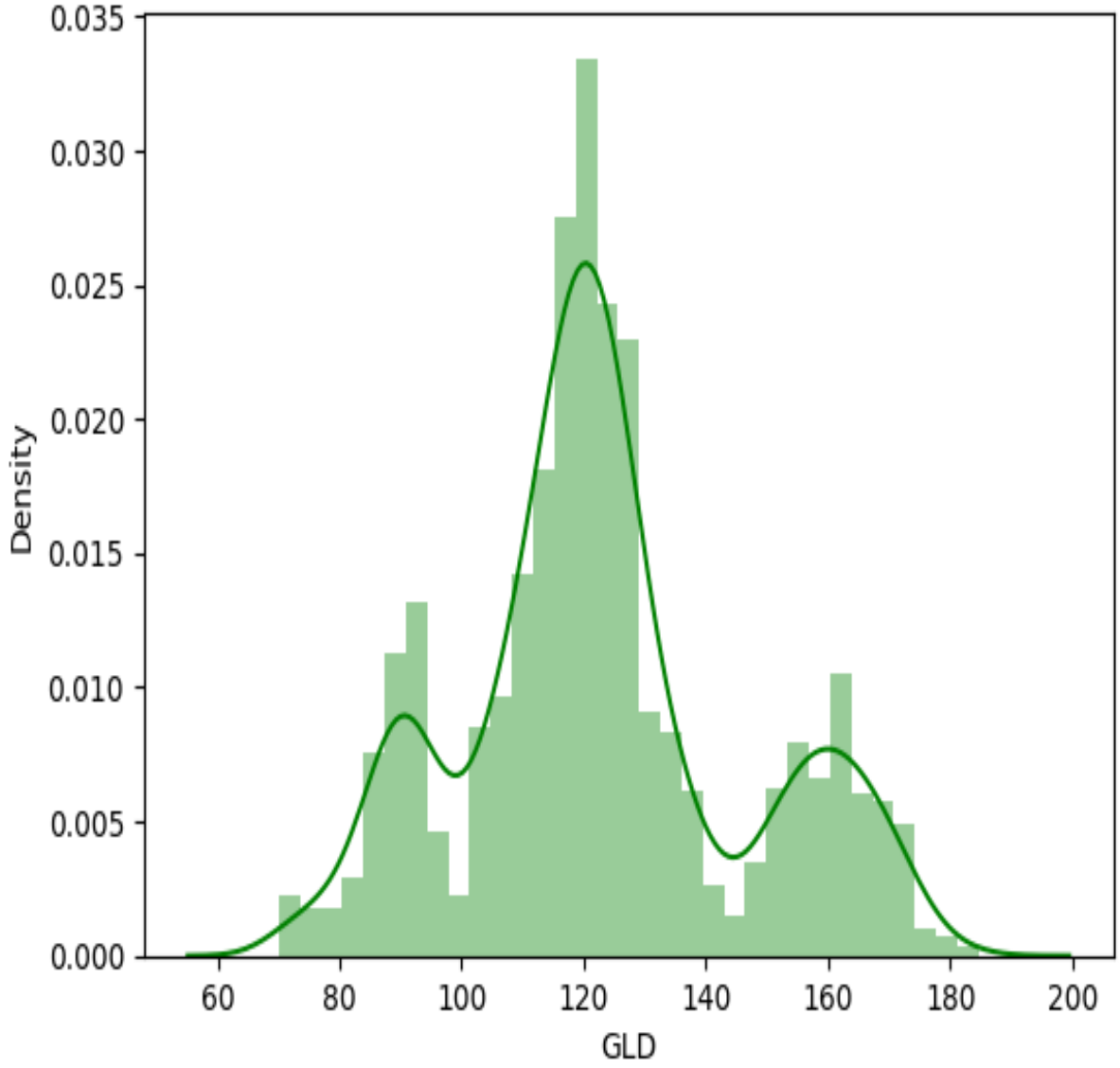
```
correlation = gold_data.corr()
```

```
# constructing a heatmap to understand the correlatiom
plt.figure(figsize = (8,8))
sns.heatmap(correlation, cbar=True, square=True, fmt='.1f',annot=True, annot_kws={'size':8}, cmap='Blues')
```

<Axes: >



DISTRIBUTION & PLOT



SPLITTING THE FEATURES

Splitting the Features and Target

```
[ ]  X = gold_data.drop(['Date','GLD'],axis=1)
     Y = gold_data['GLD']
```

```
     print(X)
```

```
                 SPX         USO       SLV   EUR/USD
0        1447.160034   78.470001  15.1800  1.471692
1        1447.160034   78.370003  15.2850  1.474491
2        1411.630005   77.309998  15.1670  1.475492
3        1416.180054   75.500000  15.0530  1.468299
4        1390.189941   76.059998  15.5900  1.557099
...              ...         ...       ...       ...
2285     2671.919922   14.060000  15.5100  1.186789
2286     2697.790039   14.370000  15.5300  1.184722
2287     2723.070068   14.410000  15.7400  1.191753
2288     2730.129883   14.380000  15.5600  1.193118
2289     2725.780029   14.405800  15.4542  1.182033

[2290 rows x 4 columns]
```

```
[ ]  print(Y)
```

```
0         84.860001
1         85.570000
2         85.129997
3         84.769997
4         86.779999
           ...
2285     124.589996
2286     124.330002
2287     125.180000
2288     124.489998
2289     122.543800
Name: GLD, Length: 2290, dtype: float64
```

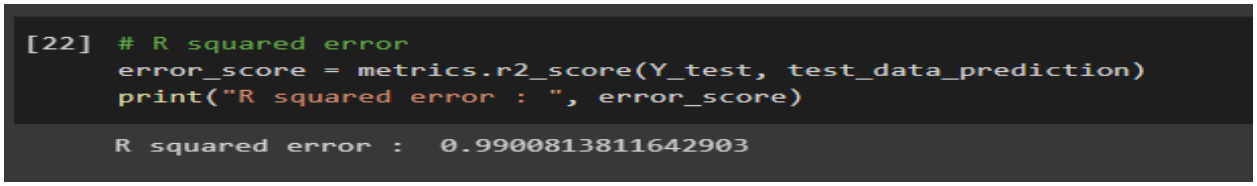
SPLITTING INTO TRAINING AND TEST DATA

Splitting into Training data and Test Data

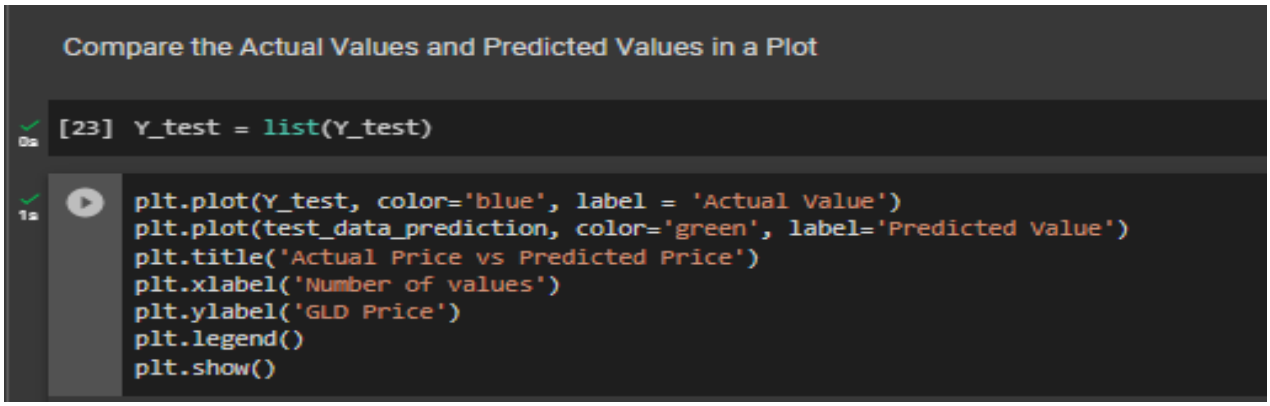
```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.2, random_state=2)
```

MODEL TRAINING

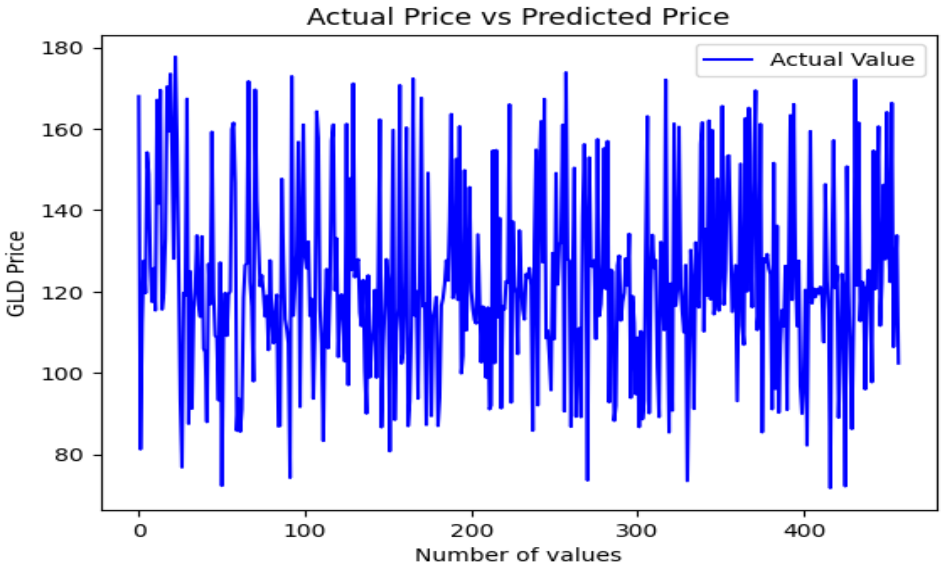
R SQUARED ERROR

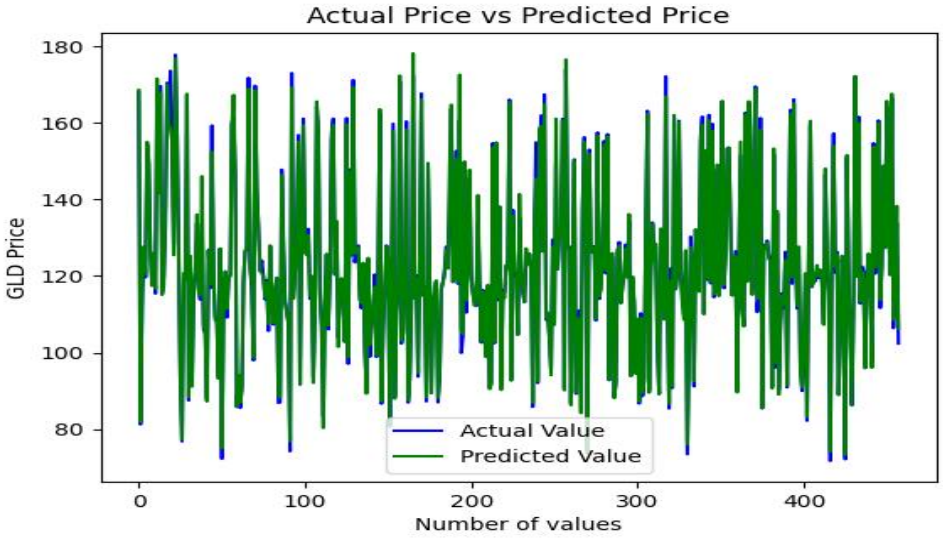


PREDICTION



RESULT


Actual Price vs Predicted Price
Actual Value
GLD Price
Number of values
180
160
140
120
100
80
0
100
200
300
400


Actual Price vs Predicted Price
Actual Value
Predicted Value
GLD Price
Number of values
180
160
140
120
100
80
0
100
200
300
400

CHAPTER 7

CONCLUSION AND FUTURE SCOPE

7.1 CONCLUSION

In conclusion, the Gold Price Prediction System is a useful tool for investors and businesses to estimate future gold prices. Machine learning algorithms and historical data are leveraged to make accurate predictions based on various factors such as economic indicators, geopolitical events, and supply and demand. The use of this system can aid in investment and hedging decisions, and help businesses plan their financial strategies more effectively. However, it is important to note that the accuracy of the predictions may vary depending on the quality and quantity of data available, as well as changes in market conditions. Therefore, it is recommended to periodically review and update the system to ensure optimal performance. Overall, the Gold Price Prediction System has the potential to provide valuable insights and help investors and businesses succeed in the competitive world of finance.

7.2 FUTURE SCOPE

The future scope of Gold Price Prediction in machine learning is very promising as it can provide valuable insights to businesses and investors who rely on gold as an investment asset. With the constantly changing economic and political landscape, predicting the future price of gold can be challenging but important for decision-making. Some potential advancements in this system include incorporating more advanced machine learning algorithms such as neural networks and deep learning, integrating more data sources such as global economic indicators and government policies, and expanding the system to work with other precious metals and commodities. Additionally, this system could also be used for other purposes such as predicting the demand for gold in different industries, which could be useful for businesses in supply chain management and production planning. Overall, the future of Gold Price Prediction in machine learning is bright and has the potential to be an important tool for businesses and investors in the ever-changing market.

CHAPTER 8

REFERENCES

1. https://www.youtube.com/watch?v=9ffkBvh8PTQ
2. https://www.geeksforgeeks.org/gold-price-prediction-using-machine-learning/
3. https://www.analyticsvidhya.com/blog/2021/07/building-a-gold-price-prediction-model-using-machine-learning/