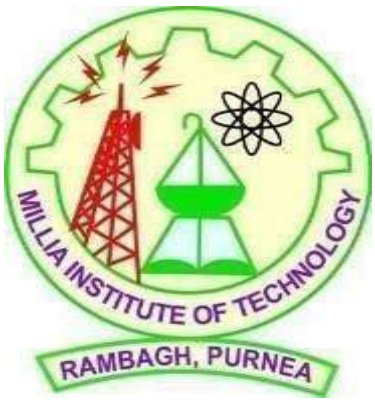
MILLIA INSTITUTE OF TECHNOLOGY

RAMBAGH, PURNEA (BIHAR) 854301, INDIA

Affiliated to Purnea University

Department of Computer Science & Engineering



A
Project Report on

GOLD PRICE PREDICTION USING MACHINE LEARNING

Submitted by

Abhinav Anand -9631

Pritam Kumar - 9630

Shivam Kumar- 9635

Sandhya kumari-9193

UNDER THE GUIDENCE OF

Professor : Mrs. Sayema Sadaf

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

MILLIA INSTITUTE OF TECHNOLOGY

RAMBAGH, PURNEA (BIHAR) 854301, INDIA

Department of Computer Science & Engineering

Certificate

This is certified that the major project entitled "**GOLD PRICE PREDICTION USING MACHINE LEARNING**" is a bonafied work carried out by Abhinav Anand-9631, Pritam kumar-9630, Shivam kumar-9635 & Sandhya kumari-9193 in partial fulfilment for the requirement of **4th** year **Computer Science & Engineering of Purina University, Purina, Bihar.**

Examiner- I **Examiner-II**

ACKNOWLEDGEMENT

First and foremost, we thank our God for his blessings.

We would like to express our special thanks of gratitude to the Principal **Dr. Saquib Shakil** and H.O.D of Computer Science & Engineering Department, **Mrs. Sayema Sadaf** who helped us in doing lots of study and research in doing this project on "**GOLD PRICE PREDICTION USING MACHINE LEARNING**". During this we came to know about so many new things.

We are also special thankful to our Internal Guide **Mrs. Sayema Sadaf** whose encouragement, guidance and support from the initial to the final level enabled us to develop an understanding of the topic.

We would also like to thank our parents and friends who helped us in finishing this project on time.

Submitted by

Abhinav Anand -9631

Pritam Kumar - 9630

Shivam Kumar- 9635

Sandhya kumari-9193

ABSTRACT

We predict future gold rates supported twenty-two market variables victimization machine learning technique. One machine learning algorithm, random forest regression was used in analyzing this knowledge. Historically, gold was used for supporting trade transactions around the world besides alternative modes of payment. Various states maintained and increased their gold reserves and were recognized as rich and progressive states. In present times, precious metals like gold area unit control with central banks of all countries to make sure re-payment of foreign debts, and conjointly to control inflation. Moreover, it conjointly reflects the I money strength of the country. Besides government agencies, varied transnational firms and people have conjointly invested with in gold reserves. In ancient events of Asian countries, gold is in addition presented as gifts/souvenirs and in marriages, gold ornaments are conferred as gift in Republic of India.

KEYWORDS: Price prediction, Machine Learning, Supervised Learning, Linear Regression, Python , Power Bi , Tableau.

CONTENT

INTRODUCTION

Investing refers to using existing fund with the aim of generating favorable returns in the future. From an economic point of view, investment can be thought of as the purchase of an asset that is not currently consumed but will be used to generate wealth in the future. In the world of finance, investing is buying a monetary asset with the idea that the asset will generate income in the future or be sold later for a higher price for a profit. India's economy is one of the fastest growing in the world, with higher levels of disposable income and abundant investment opportunities. There are many investment options available to investors such as stocks, deposits, commodities and real estate. Each has different risk and return characteristics. Gold is another asset that many investors see as an attractive investment route due to its rising value and range of uses. Investors are increasingly favoring gold as a safe haven due to negative expectations of developed foreign exchange and capital market conditions. Gold is also considered as "the asset of final instance" i.e., the asset that investors turn to when capital markets in developed countries fail to provide the desired level of profitability. Investors therefore see gold as a tool to hedge against other market volatility. Since gold is a precious metal, its price, like any other commodity, is subject to supply and demand. However, gold is storable and supplies have been accumulating over the centuries, so this year's production will have little impact on prices. Gold is used both as a commodity and as a financial asset. Gold works less as a commodity than long-term assets such as stocks and bonds. The price of gold is subject to a myriad of interrelated variables including inflation, currency fluctuations and political instability. This rising value of gold, combined with volatility and falling prices in other markets such as the capital and property markets, is attracting more and more investors to gold as an attractive investment. However, the price of gold has also witnessed high volatility recently, marking gold investments riskier. There are concerns about whether these high prices are sustainable and when prices

will reverse. Although there are numerous studies analyzing correlations between gold prices and several economic variables. It is still believed that studies clarifying the impact and impact of various macroeconomic factors on the price of gold in the current context will help determine the dynamic implications of current context will help determine the dynamic implications of these relationships. This work therefore aims to examine the relationship between gold prices and selected economic and market variables. Understanding such relationships not only helps monetary policy makers, but also investors, fund managers and portfolio managers to make better investment decisions in the market. Furthermore, in this study, he uses one machine learning algorithm when analyzing this data: random forest regression [1]. Gold was used for supporting trade transactions around the world besides other modes of payment. Various states maintained and enhanced their gold reserves and were recognized as wealthy and progressive states. Our

project will be beneficial for investors to decide when to invest in the market. Various big investor and individuals have also invested in gold and also day by day gold investment is increasing. Fractional change in gold price may result in huge profit and loss.

Investor preference for gold as protective asset.

LITERATURE REVIEW

There are many studies in the literature dealing with the price of gold. Although different variable was used in these studies, gold prices were generally observed to regress against US dollar and stock returns. The relationship between other macroeconomic variables and the price of gold has also been studied by many researchers. The relationship between the price of gold and the price of other commodities, especially crude oil, has also been extensively studied. However, the result of these studies is contradictory. The section below discusses some studies on the factors that influence the price of gold and various techniques used to study these relationships. Lawrence found no significant correlation between gold returns and changes in certain macroeconomic variables such as inflation and GDP. He also found that gold returns correlated less with stock and bond index returns than other commodities. However, Sjaastad and Scacciavillani reported that gold is a store of value against inflation, and Baker and Van-Tassei also found that the price of gold depends on future inflation rates. Regarding the relationship between the price of gold and inflation, Hanan Naser is an option based on a literature review that historical research on gold's effectiveness as a hedge against inflation is inconsistent. aIsmail et al. predict gold prices based on several economic factors, including: Standard and poor 500 index, Treasury bills and USD index. The rate of inflation, and the money supply have a significant impact on the price of gold. Khaemusunun examined the impact of selected country currencies, oil prices and interest rates on gold prices. Hammoudeh et al. concludes that tere is a correlation between gold price volatility and the exchange rates. Ai, et al. report empirical evidence that exchange rate is related to gold prices in both the long and short term. Ewing and Malik found evidence of volatility carryover between gold oil futures prices. Ghosh et al conclude that gold prices are related to US inflation, interest rate and dollar rates. As a result of cointegration analysis, they also reported a long-term relationship between the price of gold and US Consumer price index. A review of the relevant literature allows us to conclude that the relationship between the price of gold and the various factors considered to influence it are contradicting. Researchers have used a variety of techniques when studying the relationship between gold price volatility and the factors thought to influence it. Hossein and Abdolreza used artificial neural networks (ANN) and his ARIMA model to predict gold prices. Khaemusunun, (2009) uses multiple regression analysis and his ARIMA model to predict gold prices. Toraman has reported that various studies have been conducted using multivariate regression model to test the sensitivity of gold prices among various variables.

OBJECTIVE

Here, taking advantage of what has been proposed, we proposed adaptive, flexible, and scalable predictive model. Computationally intelligent neural network models to improve training learning processes and improve faster convergence. The proposed study is most likely to achieve high training rate predictions gold EFT price accuracy being monitored. In general, this work is done to make good suggestions. Predictive models to effectively display deemed money in different scenarios using the deemed data set. The respective database for the previous year. The purpose of this present is to correctly represent the future in its modified form. The closing price of the future Gold ETF for a specified time period. Machines maintained by this project. A learning algorithm and a solution model were used to decide whether to buy a gold ETF using a record of past values.

The main objectives of the project are:

- This project is based on the applicability of the proposed machine learning algorithms that had demonstrated their efficiency to predict gold prices with a better predictive rate.
- We proposed the development of a prediction model for predicting future gold prices using Random Forest regression.

WORKING PROCEDURE

Steps:-

- First, we need the gold price data.
- Then we have to pre-process the data because we can't give the raw data to the machine learning algorithm so we have to process the data.
- After processing we have to analysis the data to find which features are important or which features are correlated with each other.
- After that, we will split our original data into train data & test data. We will use this training data to train our machine learning model and we need find how will our model is performing for that we need to use test data.
- Then we use random forest regressor model.
- Then we evaluate our model based on the test data.

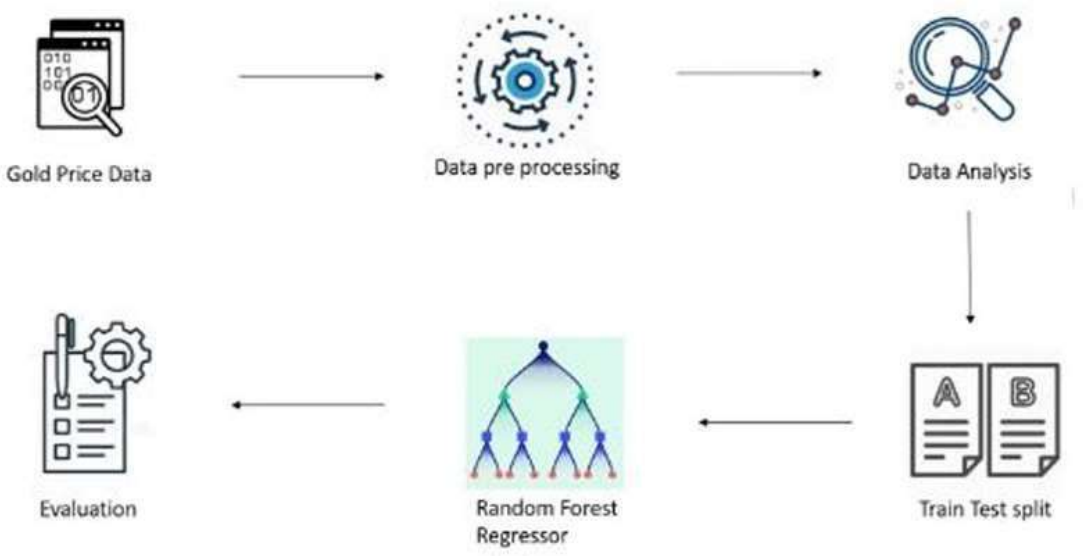


Fig-1

RANDOM FOREST REGRESSION: -

Random Forest Regression is a supervised learning algorithm that capable of both regression and classification with the use of multiple decision tree. The random forest uses bootstrap on the decision tree to reduce the variance also maintain the low bias which is the resulted. It is used to predict things like stock trading, e commerce etc. It can reduces overfitting problem. It can use for both classification and regression problems and I can handle categorical and continuous variable.

A Random Forest is an ensemble technique that can perform both regression and classification task using multiple decision trees. The basic idea behind this is a combine multiple decision trees to determine the final output, rather than relying on individual decision trees. The random forest algorithm has the following advantages over most other algorithm. The same random forest algorithm can be used for both classification and regression tasks. Also, the random forest algorithm can be used for feature engineering to identify the most important features from the available features of the training dataset.

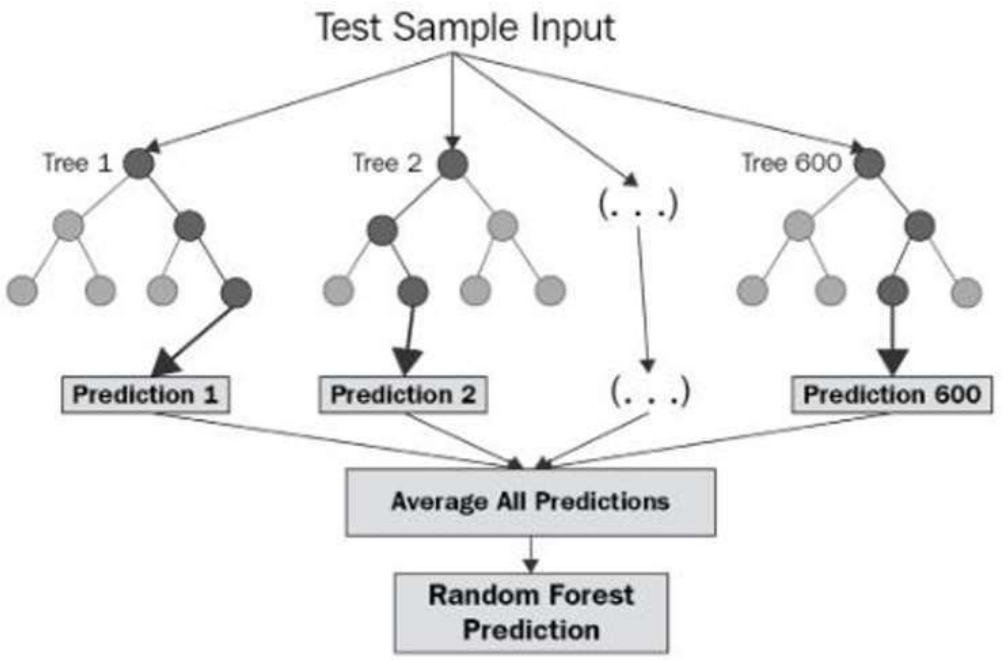


Fig-2

need data because data set need to train the model to help it predict then we need to test the data set how accurate is our model. In this project we have use the data set which is available in Kaggle. If you give more data to your model to feed then more accurate the result come out to be.

The statistical process for estimating the relationship between different variables is called regression analysis. Regression analysis is used to understand how the variable of the dependent variable changes when one of the independent variable changes, while other variables are fixed.

Decision trees can be used various machine learning applications. Decision trees constructs a tree that is used for classification and regression. But trees that are grown really deep to learn highly irregular patterns tend to over-fit the training sets. A slight noise in the data may cause the tree to grow in a completely different manner.

In our data set we have 6 column and 2290 rows, each column consists of ,

- Date: Date is the date of gold price.
- SPX: SPX stands for Standard & Poor's 500 Index.
- GLD: GLD is the price of the gold.
- USO: US economic data or Oil ETF.
- SLV: SLV is price of the silver.
- EUR/USD: The currency pair of Europe and United States.

RESULT AND ANALYSIS

The actual and the predicted prices are almost the same, as the two graphs overlap each other so, our model performed well.

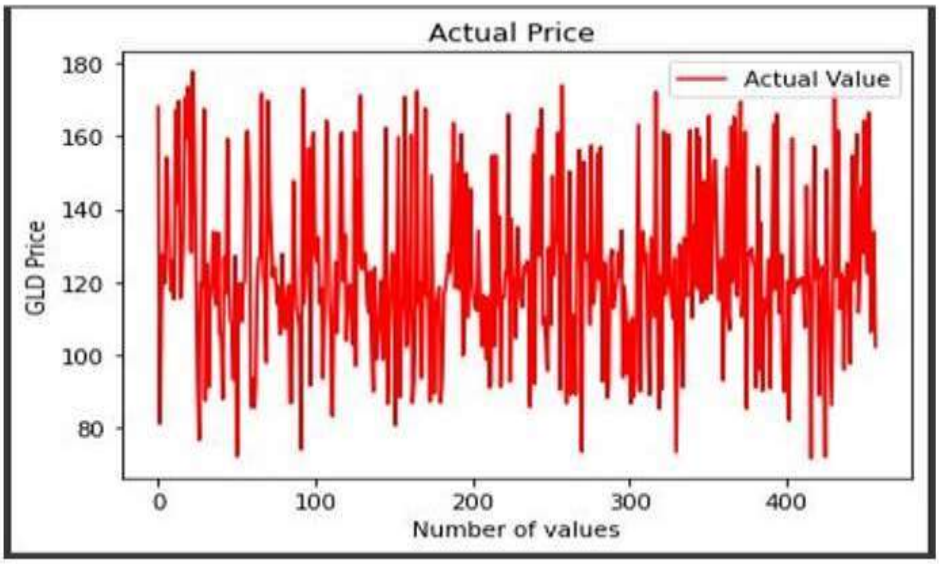


Fig-3

This graph shows the actual price of the gold.

Then we compare the predicted value of the with the actual value.

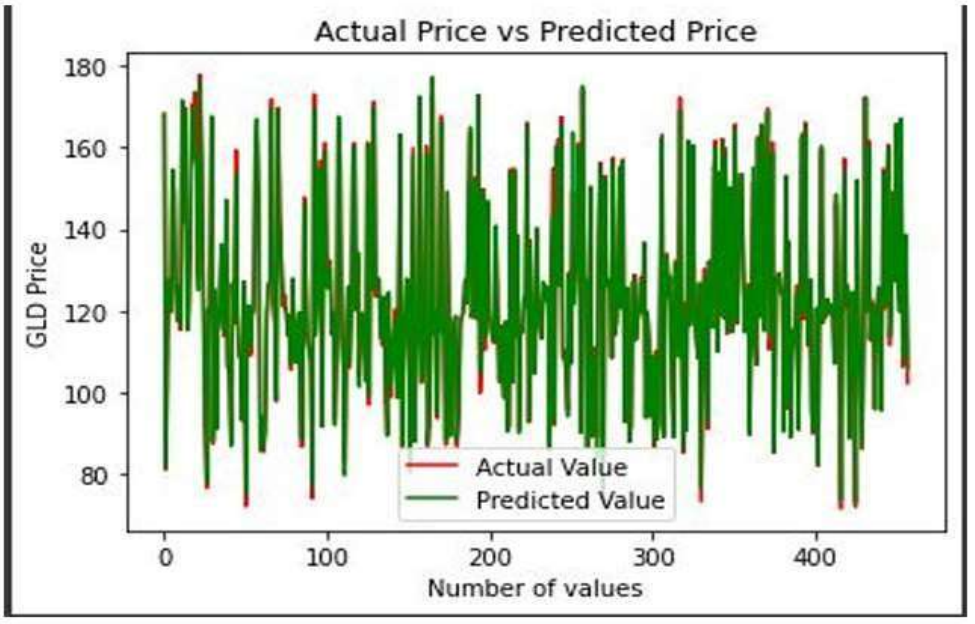


Fig-4

Here, you can see some time actual value is more than predicted value that is because of the 0.9891144286325335 error.

```
error_score = metrics.r2_score(Y_test, test_data_prediction)
print("R squared error : ", error_score)
```

```
R squared error :  0.9891144286325335
```

Fig-5

PROGRAM SOURCE CODE

```
# -*- coding: utf-8 -*-
"""Project 6. Gold Price Prediction.ipynb
Automatically generated by Colaboratory.

Original file is located at https://colab.research.google.com/drive/10h3A1F-i7q5C2YoHtu-Lmh81U_aMJNkW """

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

"""Data Collection and Processing"""

# loading the csv data to a Pandas DataFrame gold_data =
pd.read_csv('/content/gold price dataset.csv')

# print first 5 rows in the dataframe gold_data.head()

# print last 5 rows of the dataframe gold_data.tail()

# number of rows and columns gold_data.shape

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

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

# getting the statistical measures of the data gold_data.describe()
```

"""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')

# correlation values of GLD print(correlation['GLD'])

# checking the distribution of the GLD Price sns.distplot(gold_data['GLD'],color='green')

"""Splitting the Features and Target"""

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

print(X)

print(Y)

"""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)



"""Importing the Libraries

Model Training:
Random Forest Regressor
"""                    regressor                    =
RandomForestRegressor(n_estimators=100)
```

```
# training the model regressor.fit(X_train,Y_train)

"""Model Evaluation"""

# prediction on Test Data test_data_prediction
= regressor.predict(X_test)
print(test_data_prediction)

# R squared error error_score = metrics.r2_score(Y_test,
test_data_prediction) print("R squared error : ", error_score)

"""Compare the Actual Values and Predicted Values in a Plot"""

Y_test = list(Y_test) plt.plot(Y_test, color='blue', label = 'Actual
Value') plt.plot(test_data_prediction, color='green', label='Predicted
Value') plt.title('Actual Price vs Predicted Price')
plt.xlabel('Number of values') plt.ylabel('GLD Price') plt.legend()
plt.show()
```

SCREEN SHORT

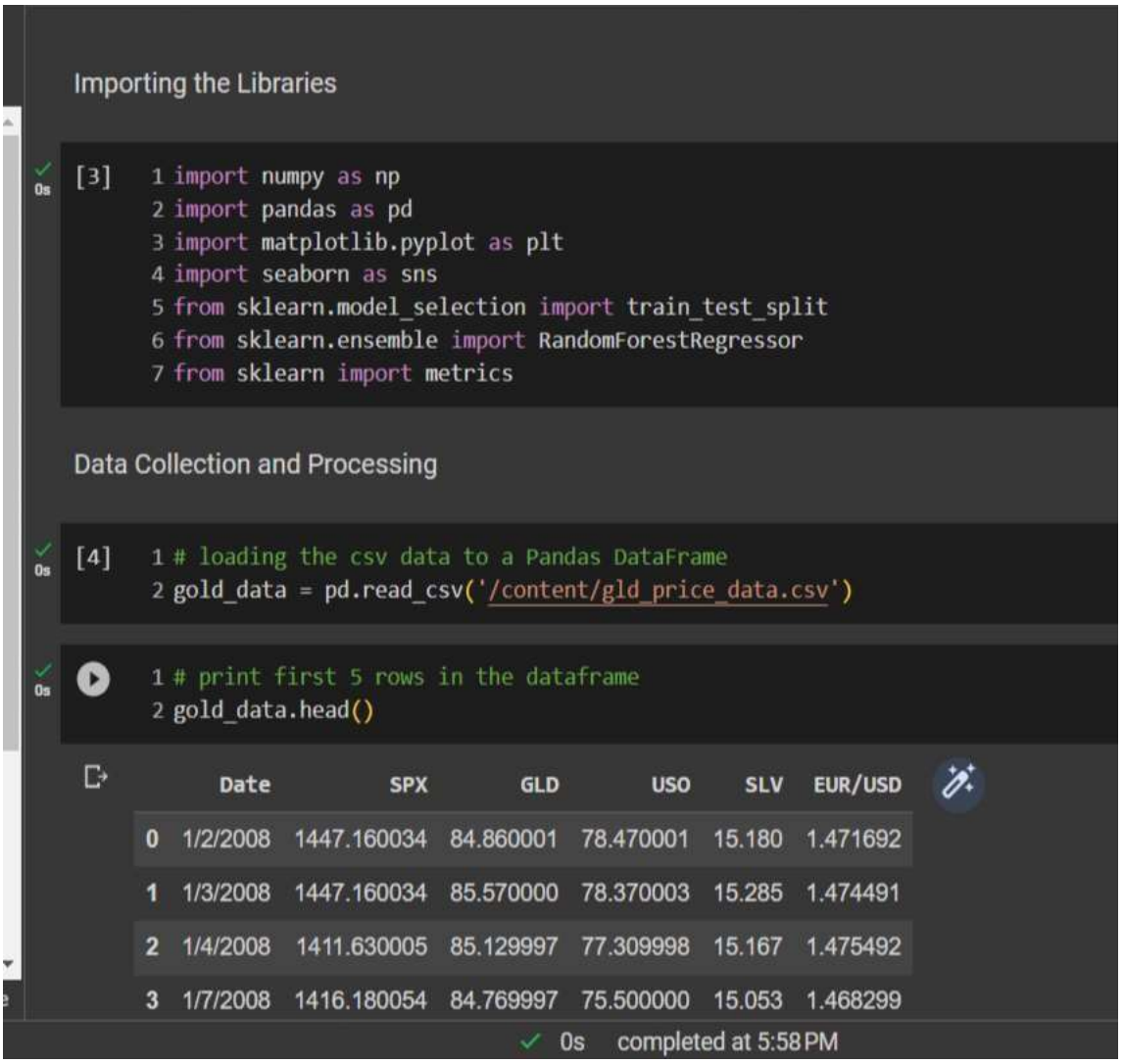


Fig-6

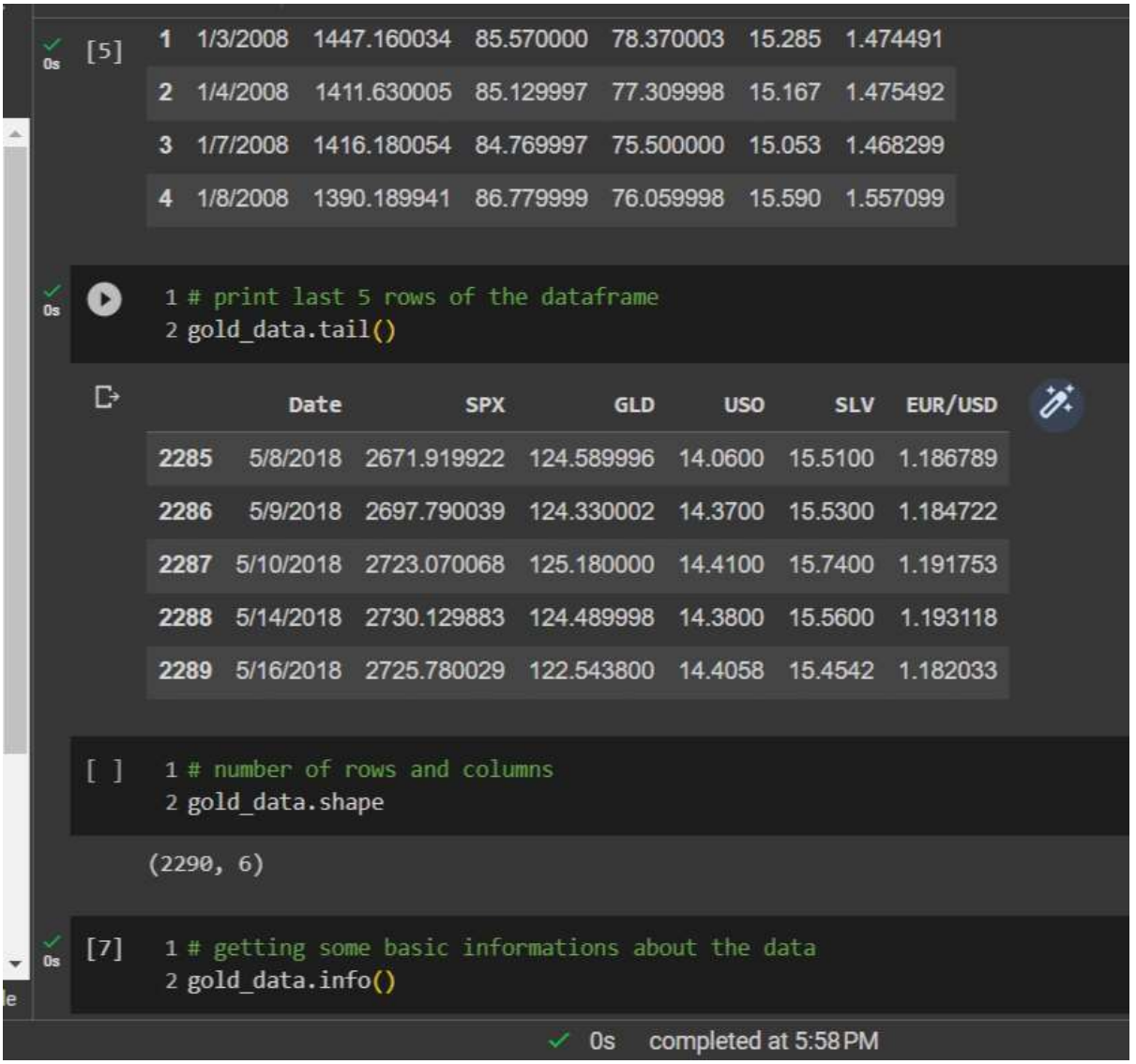

[5]
1 1/3/2008 1447.160034 85.570000 78.370003 15.285 1.474491
2 1/4/2008 1411.630005 85.129997 77.309998 15.167 1.475492
3 1/7/2008 1416.180054 84.769997 75.500000 15.053 1.468299
4 1/8/2008 1390.189941 86.779999 76.059998 15.590 1.557099
1 # print last 5 rows of the dataframe
2 gold_data.tail()
Date SPX GLD USO SLV EUR/USD
2285 5/8/2018 2671.919922 124.589996 14.0600 15.5100 1.186789
2286 5/9/2018 2697.790039 124.330002 14.3700 15.5300 1.184722
2287 5/10/2018 2723.070068 125.180000 14.4100 15.7400 1.191753
2288 5/14/2018 2730.129883 124.489998 14.3800 15.5600 1.193118
2289 5/16/2018 2725.780029 122.543800 14.4058 15.4542 1.182033
1 # number of rows and columns
2 gold_data.shape
(2290, 6)
[7]
1 # getting some basic informations about the data
2 gold_data.info()
0s completed at 5:58 PM

Fig-7

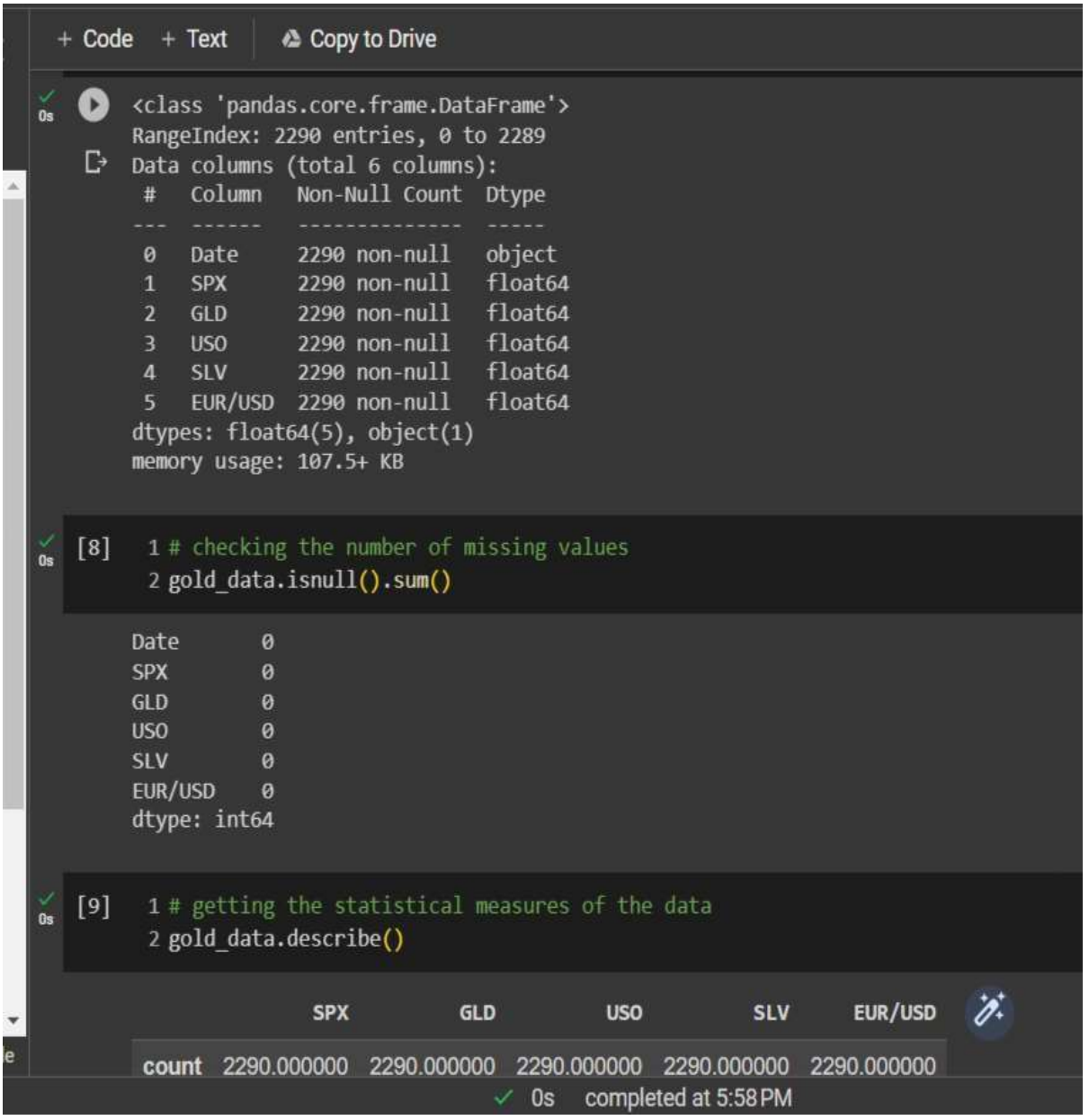

+ Code
+ Text
Copy to Drive
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2290 entries, 0 to 2289
Data columns (total 6 columns):
Column Non-Null Count Dtype
0 Date 2290 non-null object
1 SPX 2290 non-null float64
2 GLD 2290 non-null float64
3 USO 2290 non-null float64
4 SLV 2290 non-null float64
5 EUR/USD 2290 non-null float64
dtypes: float64(5), object(1)
memory usage: 107.5+ KB
[8] 1 # checking the number of missing values
2 gold_data.isnull().sum()
Date 0
SPX 0
GLD 0
USO 0
SLV 0
EUR/USD 0
dtype: int64
[9] 1 # getting the statistical measures of the data
2 gold_data.describe()
SPX GLD USO SLV EUR/USD
count 2290.000000 2290.000000 2290.000000 2290.000000 2290.000000
0s completed at 5:58 PM

Fig-8

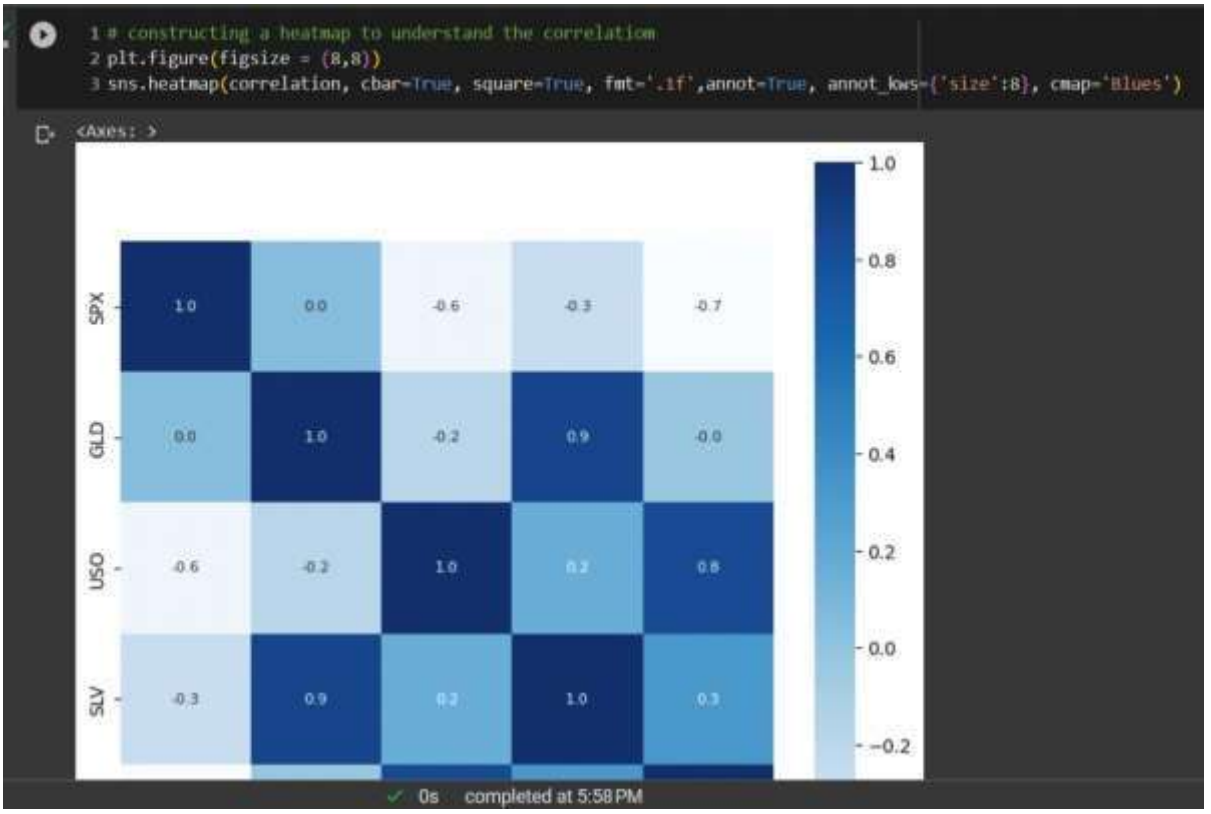

1 # constructing a heatmap to understand the correlation
2 plt.figure(figsize = (8,8))
3 sns.heatmap(correlation, cbar=True, square=True, fmt='.1f',annot=True, annot_kws={'size':8}, cmap='Blues')
<Axes: >
SPX
GLD
USO
SLV
1.0
0.0
-0.6
-0.3
-0.7
0.0
1.0
-0.2
0.9
-0.0
-0.6
-0.2
1.0
0.2
0.8
-0.3
0.9
0.2
1.0
0.3
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0s completed at 5:58 PM

Fig-9

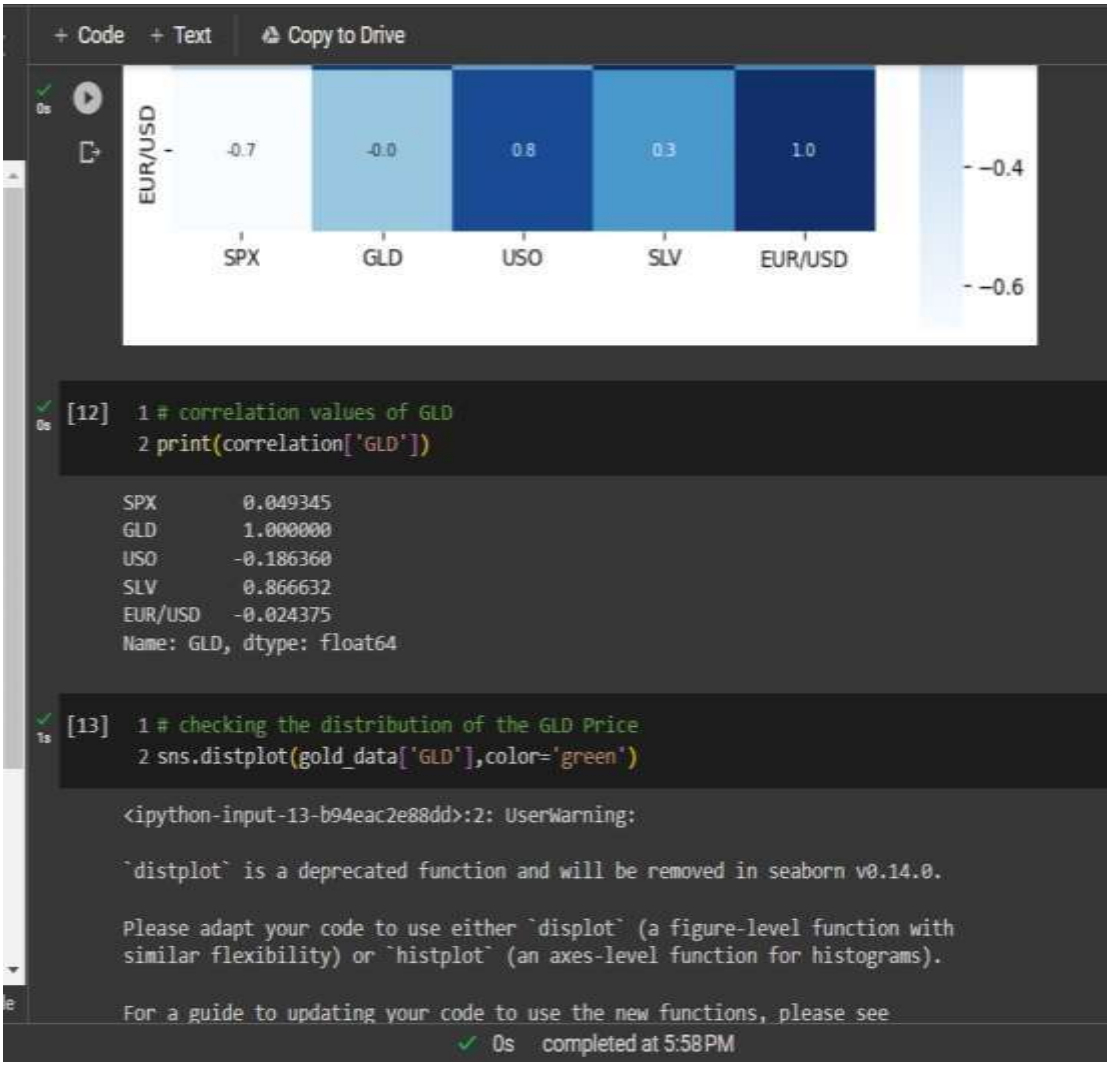

+ Code
+ Text
Copy to Drive
EUR/USD
-0.7
-0.0
0.8
0.3
1.0
SPX
GLD
USO
SLV
EUR/USD
−0.4
−0.6
[12] 1 # correlation values of GLD
2 print(correlation['GLD'])
SPX 0.049345
GLD 1.000000
USO -0.186360
SLV 0.866632
EUR/USD -0.024375
Name: GLD, dtype: float64
[13] 1 # checking the distribution of the GLD Price
2 sns.distplot(gold_data['GLD'],color='green')
<ipython-input-13-b94eac2e88dd>:2: UserWarning:
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).
For a guide to updating your code to use the new functions, please see
0s completed at 5:58 PM

Fig-10

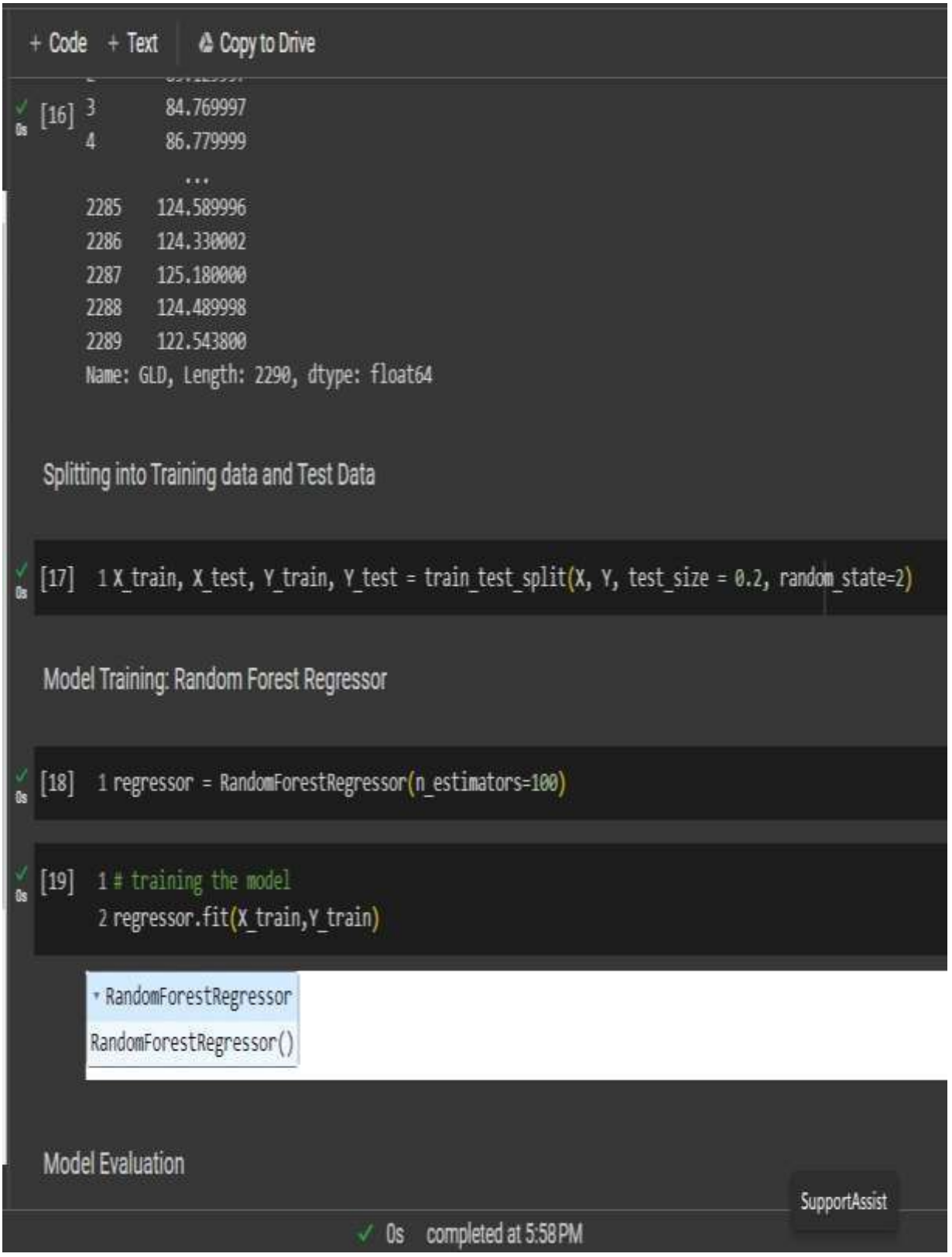

+ Code
+ Text
Copy to Drive
[16]
3 84.769997
4 86.779999
...
2285 124.589996
2286 124.330002
2287 125.180000
2288 124.489998
2289 122.543800
Name: GLD, Length: 2290, dtype: float64
Splitting into Training data and Test Data
[17] 1 X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.2, random_state=2)
Model Training: Random Forest Regressor
[18] 1 regressor = RandomForestRegressor(n_estimators=100)
[19] 1 # training the model
2 regressor.fit(X_train,Y_train)
RandomForestRegressor
RandomForestRegressor()
Model Evaluation
SupportAssist
0s completed at 5:58 PM

Fig-11

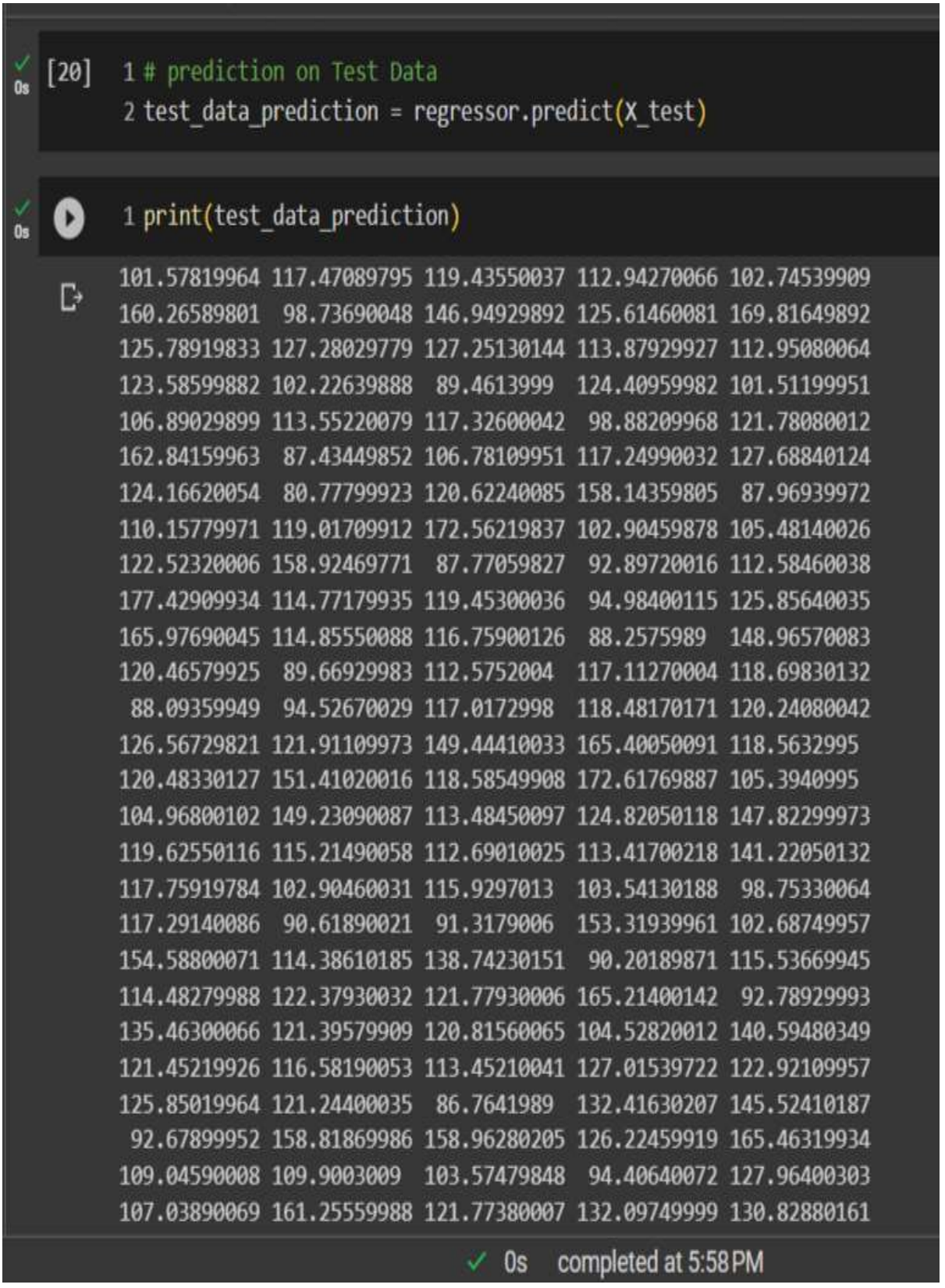

[20] 1 # prediction on Test Data
2 test_data_prediction = regressor.predict(X_test)
1 print(test_data_prediction)
101.57819964 117.47089795 119.43550037 112.94270066 102.74539909
160.26589801 98.73690048 146.94929892 125.61460081 169.81649892
125.78919833 127.28029779 127.25130144 113.87929927 112.95080064
123.58599882 102.22639888 89.4613999 124.40959982 101.51199951
106.89029899 113.55220079 117.32600042 98.88209968 121.78080012
162.84159963 87.43449852 106.78109951 117.24990032 127.68840124
124.16620054 80.77799923 120.62240085 158.14359805 87.96939972
110.15779971 119.01709912 172.56219837 102.90459878 105.48140026
122.52320006 158.92469771 87.77059827 92.89720016 112.58460038
177.42909934 114.77179935 119.45300036 94.98400115 125.85640035
165.97690045 114.85550088 116.75900126 88.2575989 148.96570083
120.46579925 89.66929983 112.5752004 117.11270004 118.69830132
88.09359949 94.52670029 117.0172998 118.48170171 120.24080042
126.56729821 121.91109973 149.44410033 165.40050091 118.5632995
120.48330127 151.41020016 118.58549908 172.61769887 105.3940995
104.96800102 149.23090087 113.48450097 124.82050118 147.82299973
119.62550116 115.21490058 112.69010025 113.41700218 141.22050132
117.75919784 102.90460031 115.9297013 103.54130188 98.75330064
117.29140086 90.61890021 91.3179006 153.31939961 102.68749957
154.58800071 114.38610185 138.74230151 90.20189871 115.53669945
114.48279988 122.37930032 121.77930006 165.21400142 92.78929993
135.46300066 121.39579909 120.81560065 104.52820012 140.59480349
121.45219926 116.58190053 113.45210041 127.01539722 122.92109957
125.85019964 121.24400035 86.7641989 132.41630207 145.52410187
92.67899952 158.81869986 158.96280205 126.22459919 165.46319934
109.04590008 109.9003009 103.57479848 94.40640072 127.96400303
107.03890069 161.25559988 121.77380007 132.09749999 130.82880161
0s completed at 5:58 PM

Fig-12

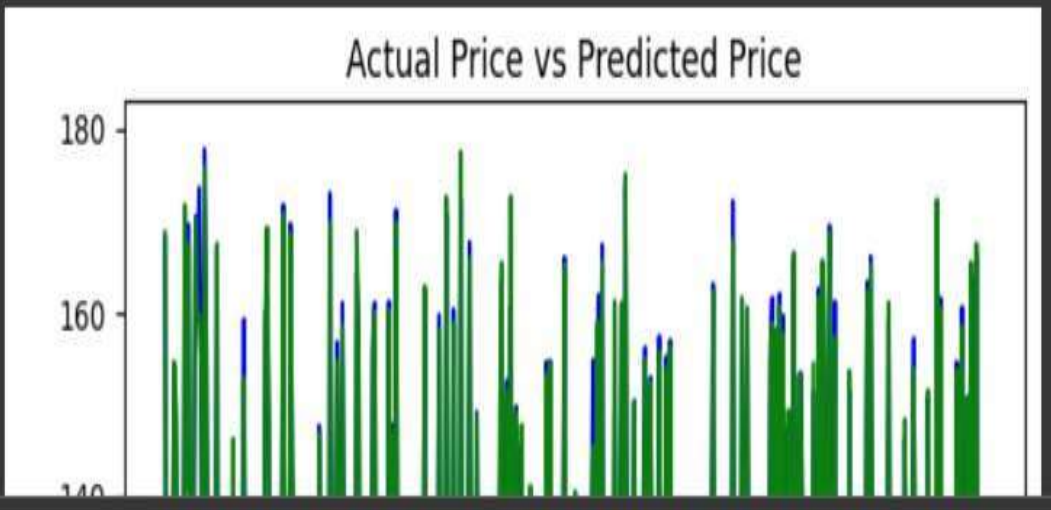
```
[22]  1 # R squared error
      2 error_score = metrics.r2_score(Y_test, test_data_prediction)
      3 print("R squared error : ", error_score)
```

```
R squared error :  0.9889161525754209
```

Compare the Actual Values and Predicted Values in a Plot

```
[23]  1 Y_test = list(Y_test)
```

```
1 plt.plot(Y_test, color='blue', label = 'Actual Value')
2 plt.plot(test_data_prediction, color='green', label='Predicted Value')
3 plt.title('Actual Price vs Predicted Price')
4 plt.xlabel('Number of values')
5 plt.ylabel('GLD Price')
6 plt.legend()
7 plt.show()
```



✓ 0s completed at 5:58 PM

Fig-13

```
plt.xlabel('Number of values')
plt.ylabel('GLD Price')
plt.legend()
plt.show()
```

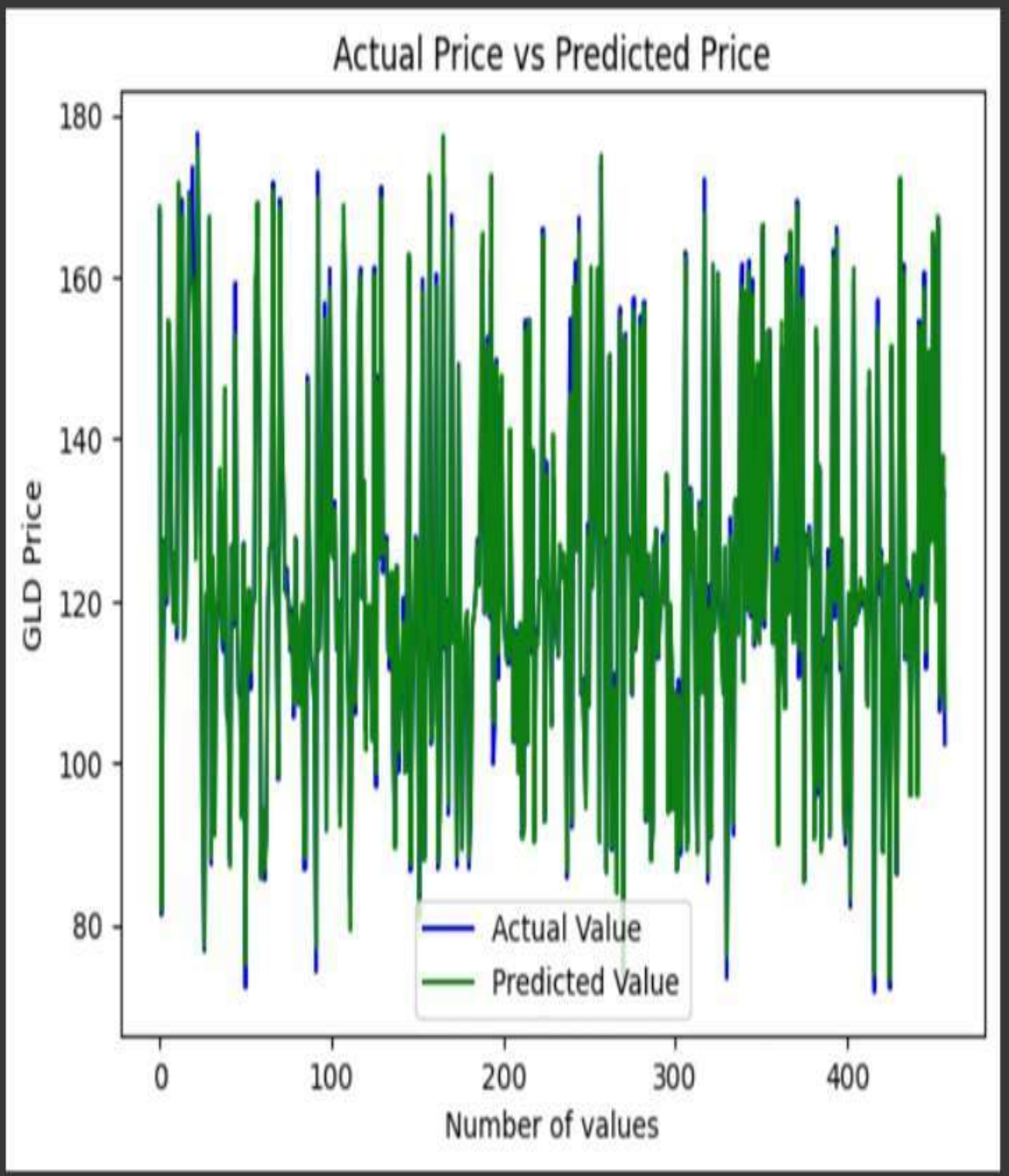


Fig-14

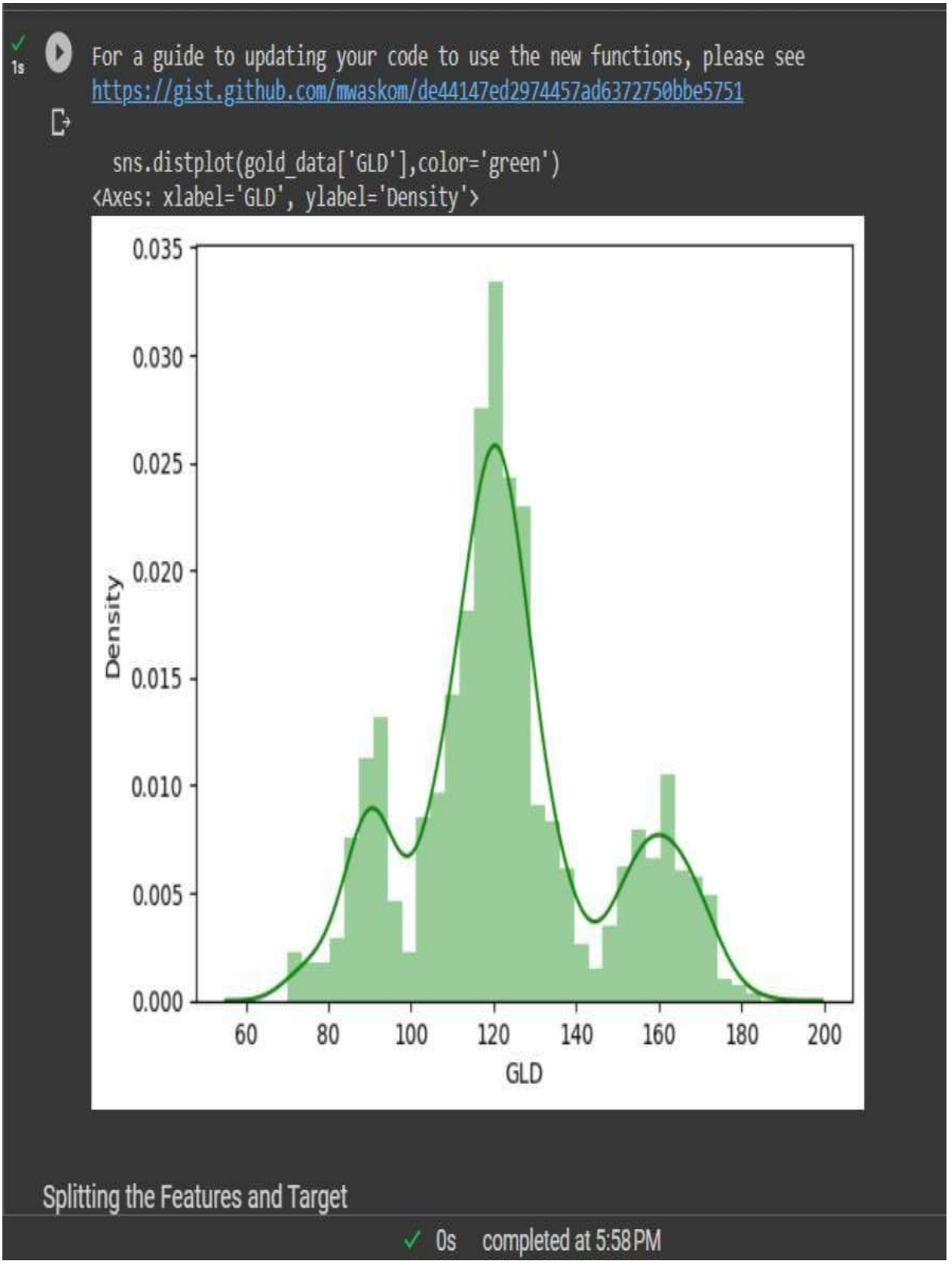

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
sns.distplot(gold_data['GLD'],color='green')
<Axes: xlabel='GLD', ylabel='Density'>
Density
GLD
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
60
80
100
120
140
160
180
200
Splitting the Features and Target
0s completed at 5:58 PM

Fig-15

CONCLUSION AND FUTURE SCOPE

As we saw in this project, we created a machine learning random forest regression model. We first train our machine learning model by giving past data. Any model can be more accurate depending upon the data. By this gold rate forecasting it will be very easy for the investor to invest right amount of money at the right time and thus maximize their income.

This study is conducted to understand the relationship between gold price and selected factors influencing its price, namely stock market, crude oil price, rupee dollar exchange rate, silver price. The machine learning algorithm random forest regression model used in analysis the data.

Random forest regression is found to have better prediction accuracy for the entire period. In future we can use it in digital gold treading platforms.

It is concluded that machine learning algorithms are very useful in such analysis, but the characteristics of the data influences their accuracy. Further research with such data and different techniques may be conducted for better understanding of performance of these techniques.

References

- https://www.irjmets.com/uploadedfiles/paper//issue_5_may_2022/22138/final/fin_irjmets1651655991.pdf
- https://www.researchgate.net/publication/336414732_Gold_Price_Prediction_using_Ensemble_based_Machine_Learning_Techniques
- https://www.analyticsvidhya.com/blog/2021/07/building-a-gold-price-prediction-model-using-machine-learning/
- https://www.youtube.com/watch?v=9ffkBvh8PTQ