

Feature Engineering – Fall 2022

Project Increment-1

PREDICTION OF GOLD PRICES FOR FINANCIAL INSTITUTIONS

1.Team Members:

1. Yash Zauwar
2. Goutham Pallapothu
3. Veeranjanyulu Muppalla
4. Sai Manideep Reddy Pallerla

2. Goals and Objectives: To develop a machine learning model which will take the past prices data of Gold and predict the price for the future. To help investors to decide to buy or sell.

- **Motivation:** Predicting prices of commodities which are traded in the financial markets is very crucial. It plays a very important role in the success of any financial institution's business.
- **Significance:** More and more investors are finding gold to be an appealing investment due to the rise in gold's value as well as the volatility and declining prices in other sectors, such as the capital and real estate markets. Developing a model to predict the price will help the institutions to attract clients and to generate the best returns.
- **Objectives:** To apply the best machine learning algorithms. This research is based on the applicability of the suggested machine learning methods, which have proven to be effective at forecasting gold prices more accurately.
- **Features:** We are using features like numerical features, discrete numerical features, correlation between the numerical features, from the data to apply a Machine Learning Model using LSTM. With the help of data sources of price of gold, the data will be divided into training data and testing data then the Linear Regression Model will be applied to the data to generate few visualizations and forecast the price.

Related Work: Numerous studies have been done on the price of gold. Gold prices are regressing against the US dollar and stock returns despite the incorporation of several different variables in this study. Numerous academics have also investigated the connection between other macroeconomic factors and gold prices. The correlation between the price of gold and the pricing of other commodities, particularly crude oil, has also been the subject of numerous studies. The results of these investigations, however, appear to be contradictory. To investigate the volatility of the gold price and its relationship to the variables that are thought to have an impact on it, researchers have used a variety of techniques.

Dataset: We are using the Open, Close, High, and Low price of Gold from the python module named as yfinance: <https://pypi.org/project/yfinance/>. We are fetching the data from yahoo by providing the start date and end date to the datareader using pandas_datareader.

Detail Design of Features: We are working on the dataset by checking the missing values, checking for features, checking for categorical features in the data, checking for numerical features in the data, checking for Discrete Numerical features in the data, checking the Correlation between numerical features, Handling Missing Values, handling Categorical Values.

Analysis: It is observed that the correlation between the close price and the other features is

```
close          1.000000
Adj Close      1.000000
low            0.999865
high           0.999850
open           0.999710
volume         0.256708
Name: close, dtype: float64
```

For the training dataset of size 4234, the Root Mean Square is 2.7746350225417964.

Predictions using the Model Trained Above

```
In [39]: predictions = model.predict(x_test)
predictions = scaler.inverse_transform(predictions)
predictions[0:10]
```

```
8/8 [=====] - 1s 16ms/step
```

```
Out[39]: array([[166.87996],
                [166.67578],
                [166.59207],
                [166.6861 ],
                [166.24525],
                [166.01465],
                [166.60776],
                [167.14212],
                [167.23987],
                [167.1248 ]], dtype=float32)
```

```
In [40]: rmse = np.sqrt(np.mean(((predictions - y_test) ** 2)))
print("RMSE score of the model:", rmse)
```

```
RMSE score of the model: 1.8861561013304986
```

FE_Project_Increment_1 - Jupyter

localhost:8888/notebooks/Fall%202022/Feature%20Engineering/Project/FE_Project_Increment_1.ipynb#Feature-Engineering-Project---Fall-2022

jupyter FE_Project_Increment_1 Last Checkpoint: 13 minutes ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Run Code

```
min 41.360001 41.020000 41.029999 41.259998 3.193000e+05 41.259998
25% 93.677498 92.372498 93.017500 93.082497 5.320800e+06 93.082497
50% 121.349998 120.415001 120.844997 120.834999 8.044900e+06 120.834999
75% 150.380005 148.607506 149.652496 149.470005 1.194095e+07 149.470005
max 194.449997 192.520004 193.740005 193.889999 9.380420e+07 193.889999
```

In [5]: df #print the dataset

Out[5]:

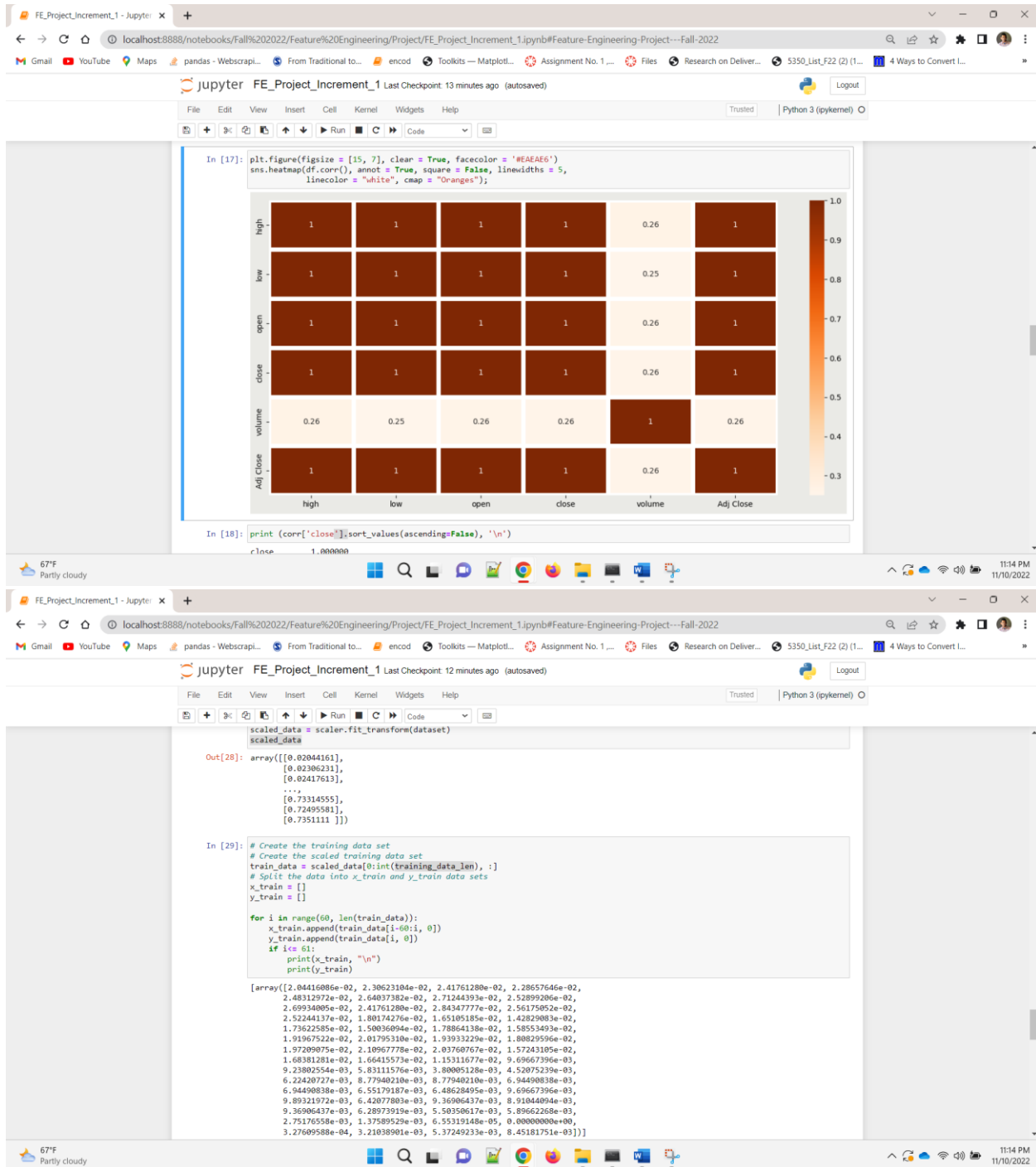
	High	Low	Open	Close	Volume	Adj Close
Date						
2004-11-18	44.490002	44.070000	44.430000	44.380001	5992000	44.380001
2004-11-19	44.919998	44.470001	44.490002	44.779999	11655300	44.779999
2004-11-22	44.970001	44.740002	44.750000	44.950001	11996000	44.950001
2004-11-23	44.919998	44.720001	44.880001	44.750000	3169200	44.750000
2004-11-24	45.049999	44.790001	44.930000	45.049999	6105100	45.049999
...
2022-10-26	155.729996	154.779999	154.779999	154.979996	4448200	154.979996
2022-10-27	155.089996	154.220001	154.710007	154.740005	3654900	154.740005
2022-10-28	153.539993	152.500000	152.960007	153.160004	4645400	153.160004
2022-10-31	152.869995	151.889999	152.160004	151.910004	6097800	151.910004
2022-11-01	153.979996	152.869995	153.820007	153.460007	4603600	153.460007

4520 rows x 6 columns

In [6]: df.rename(columns = {"Date": "date", "Open": "open", "High": "high", "Low": "low", "Close": "close", "Volume": "volume", "Adj Close": "adj_close"}, inplace = True)

#renaming the data frame columns to lower case

3. Checking for missing values in the data frame



Implementation: We are reading the data from yahoo finance with the help of yfinance module. The code for reading the data using **pandas_datareader** is written and the data is described. Then we are performing some Exploratory Data Analysis. Then we are performing feature scaling using the **MinMax** Scaler and transforming the dataset. Then we are dividing the dataset into training and testing dataset. After dividing the dataset, we are developing ML model using LSTM and apply the data to the model. Then we train the model with the training data. After this, we are applying the model to test data to predict the values. During this course, a few visualizations are generated for better understanding of the data.

Preliminary Results: We are able to generate the output for predicted price of gold using the data available from the yfinance module. For the dataset we used, the loss is found to be 0.0010.

Project Management:

Implementation Status Report:

S.No	Description of Task	Person	Contribution (%)
1	Understanding the dataset	Sai Manideep Reddy Pallerla	50
		Veeranjaneyulu Muppalla	50
2	Fetching the dataset	Sai Manideep Reddy Pallerla	50
		Veeranjaneyulu Muppalla	50
3	Exploratory Data Analysis	Yash Zauwar	25
		Goutham Pallapothu	25
		Sai Manideep Reddy Pallerla	25
		Veeranjaneyulu Muppalla	25
4	Feature Engineering	Yash Zauwar	25
		Goutham Pallapothu	25
		Sai Manideep Reddy Pallerla	25
		Veeranjaneyulu Muppalla	25
5	Machine Learning Model Training	Yash Zauwar	50
		Goutham Pallapothu	50
6	Prediction using Model Developed	Yash Zauwar	50
		Goutham	50
7	Documentation	Yash Zauwar	25
		Goutham Pallapothu	25
		Sai Manideep Reddy Pallerla	25
		Veeranjaneyulu Muppalla	25

Work to complete:

8	Visualizations of Predictions	Yash	25
		Goutham	25
		Sai	25
		Veeranjaneyulu Muppalla	25
9	Final Documentation	Yash	25
		Goutham	25
		Sai	25
		Veeranjaneyulu Muppalla	25

Github Link to the Project:

https://github.com/YashZauwar/Feature_Engineering_Fall_2022/tree/main/Team_Project

References:

- 1) chrome-extension://efaidnbmnnnibpcajpcgclefindmkaj/https://www.irjmets.com/uploadedfiles/paper//issue_5_may_2022/22138/final/fin_irjmets1651655991.pdf
- 2) <https://www.semanticscholar.org/paper/Predicting-Future-Gold-Rates-using-Machine-Learning-Sami-Junejo/d1ddf3a4df034e89919df73b8c367c33cf5fb3b9>
- 3) https://www.researchgate.net/publication/336414732_Gold_Price_Prediction_using_Ensemble_based_Machine_Learning_Techniques
- 4) <https://ieeexplore.ieee.org/document/7333120>
- 5) <https://www.kaggle.com/code/codestudioids/eda-and-model-building-gold-price-prediction>
- 6) <https://towardsdatascience.com/how-to-model-gold-price-940e809ac665>

----- End of Report -----