PHASE 5

INTRODUCTION:

Stock price prediction refers to the application of advanced technologies and novel approaches to forecast future stock prices or market trends. It involves leveraging innovative methods, data sources, and computational techniques to improve the accuracy and efficiency of predicting how the prices of financial assets, such as stocks and securities, will behave in the future

STEP 1:

DATA COLLECTION AND PREPROCESSING DATA COLLECTION: Data collection is the process of systematically gathering and acquiring relevant information or data from various sources to support analysis, modeling, and insights generation. PREPROCESSING: Preprocessing is the set of procedures and techniques used to clean, transform, and prepare raw data for analysis, modeling, or machine learning, ensuring it is in a suitable format for further processing.

STEP 2:

FEATURE ENGINEERING CREATING INFORMATIVE FEATURES: It is the process of generating new features from the existing dataset or domain knowledge that can enhance the predictive power of machine learning models. This can include mathematical transformations, aggregations, interactions between variables, and the creation of new variables that capture relevant information. DIMENSIONALITY REDUCTION: Feature engineering also includes tec

hniques to reduce the dimensionality of data while preserving essential information. Methods like Principal Component Analysis (PCA) or feature selection algorithms help eliminate irrelevant or redundant features, making models more efficient and interpretable.

STEP 3:

MODEL SELECTION ➢ Train and evaluate each candidate regression model using appropriate evaluation metrics. Using techniques like cross-validation to assess the models' performance on validation data. ➢ Consider how well each model handles the specific challenges of stock price prediction, such as capturing volatility and seasonality.

STEP 4:

MODEL TRAINING ➢ Train the selected regression model using the training data. The model learns to make predictions by finding the relationships between the input features and the target variable. ➢ The training process involves adjusting the model's internal parameters iteratively to minimize the chosen loss function (e.g., Mean Squared Error) on the training data.

STEP 5

: MODEL EVALUATION EVALUATION METRICS: Select appropriate evaluation metrics to measure the accuracy and reliability of your regression model's prediction. MEAN ABSOLUTE ERROR (MAE): Measures the average absolute difference between the predicted and actual values. Smaller MAE values indicate better accuracy MEAN SQUARED ERROR (MSE): Measures the average squared difference between the predicted and actual values. It penalizes larger errors more heavily and provides insight into the model's ability to capture variations in the data. ROOT MEAN ERROR (RMSE): The square root of MSE, which provides an interpretable metric in the same units as the target variable. R-SQUARED (R²): Also known as the coefficient of determination, R² quantifies the proportion of variance in the target variable explained by the model. A higher R² indicates a better fit to the data.

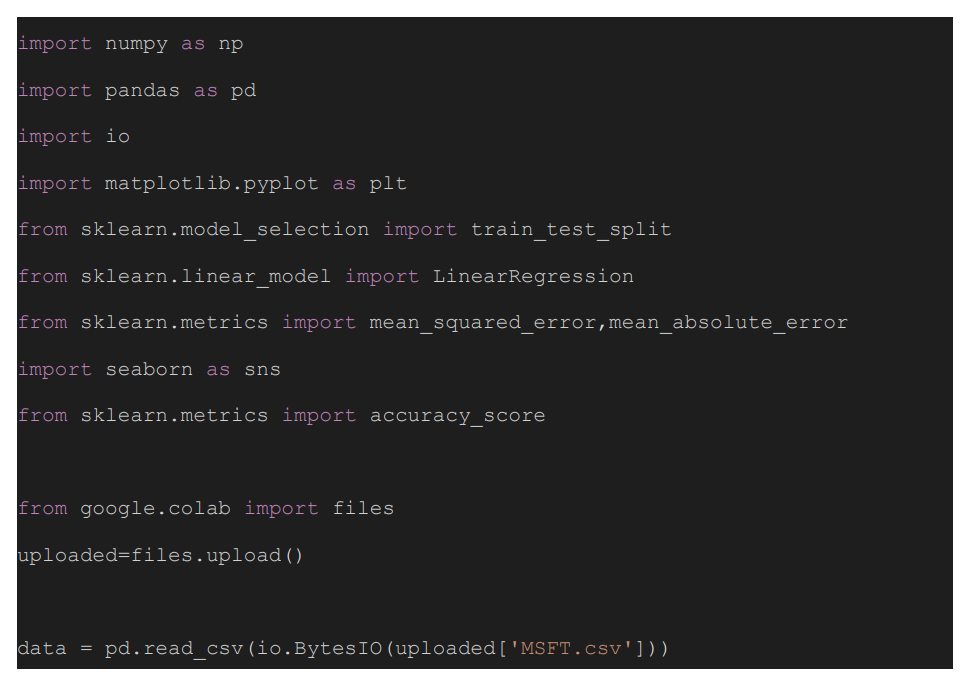
STEP 6:

DEPLOYMENT AND PREDICTION ➢ Deploy the chosen regression model to predict the stock price prediction. ➢ Assess the accuracy and reliability of the model's predictions on the new data. CONCLUSION: In phase 2 conclusion, we will summarize the key findings and insights from the advanced regression technique. We will reiterate the impact of these techniques on improving the accuracy

Step 1:

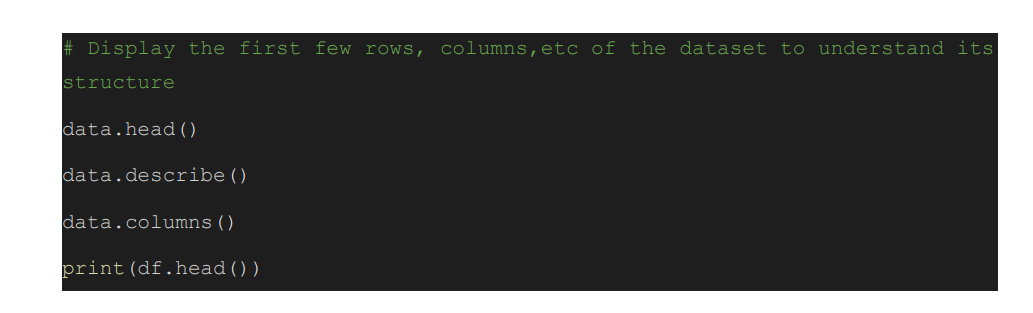
Loading Libraries and Dataset:

Data set link : https://www.kaggle.com/datasets/prasoonkottarathil/microsoft-lifetime-stocks-data set import numpy as np import pandas as pd import io import matplotlib.pyplot as plt from sklearn.model\_selection import train\_test\_spl



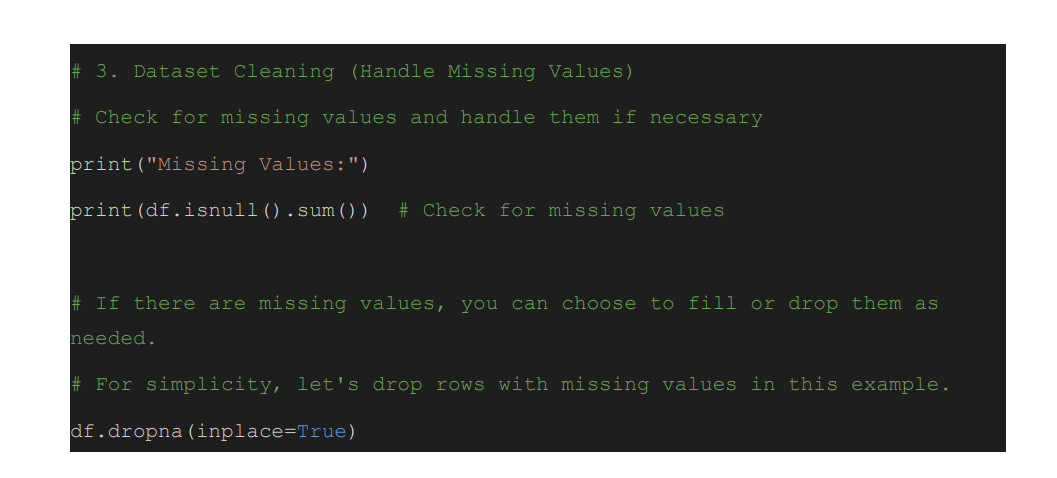


Step 2: Exploring the Dataset



Explanation: The dataset was loaded and the first few rows, descriptive statistics, and column names were displayed for initial exploration and to understand its structure.

Step 3: Handling Missing Values:



Explanation: Identified specific columns with missing values and filled them or dropped them as needed

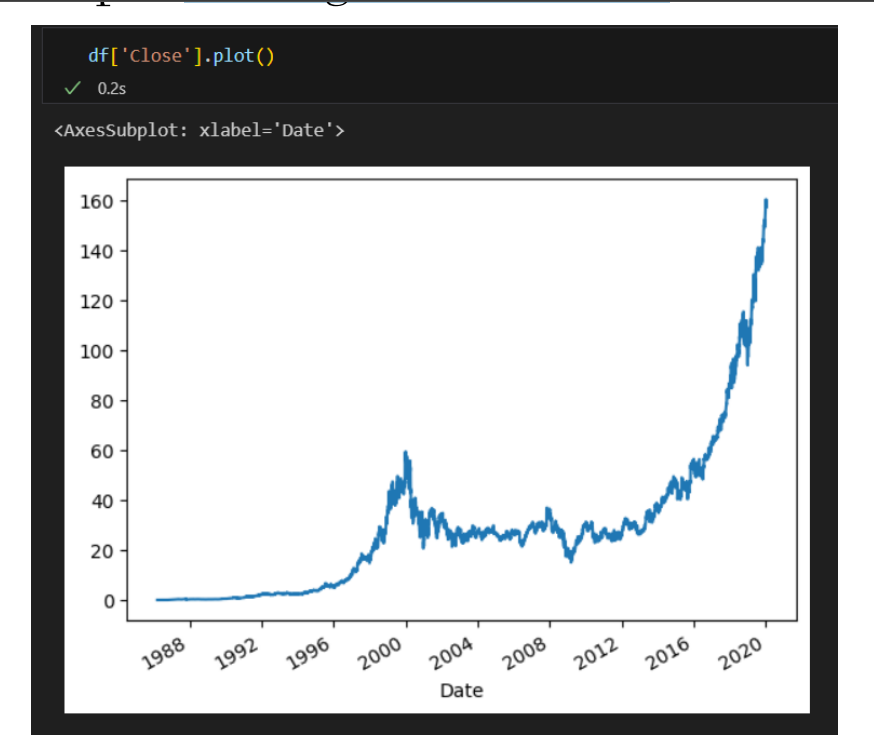
Step 4: We save and load the Preprocessed Data

Preprocessed dataset link: MSFT.xlsx

Conclusion: In this phase, we successfully loaded the dataset, explored its initial structure, handled missing values, and saved the preprocessed data for future analysis. The dataset is now ready for in-depth analysis, modeling, and prediction in the subsequent phases of the project

FEATURE ENGINEERING:

Plotting the Close value



Explanation: To plot the value of the close price of total time period of stock for acknowledgement

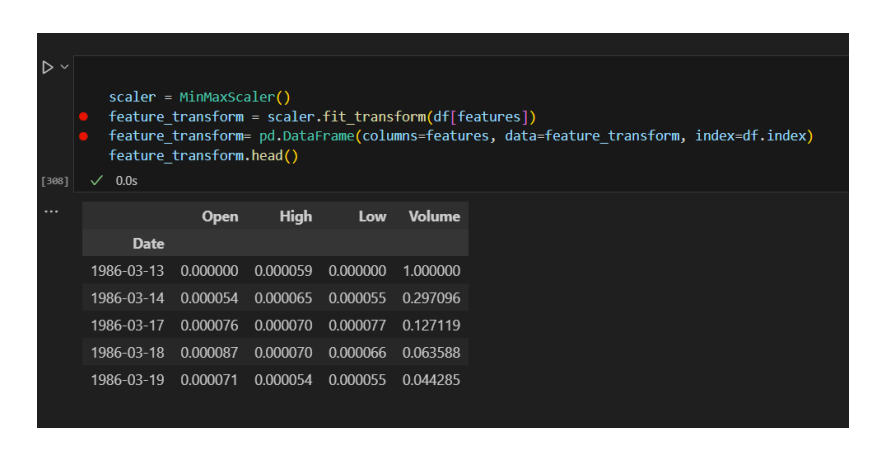
**select the feature and variable**

****

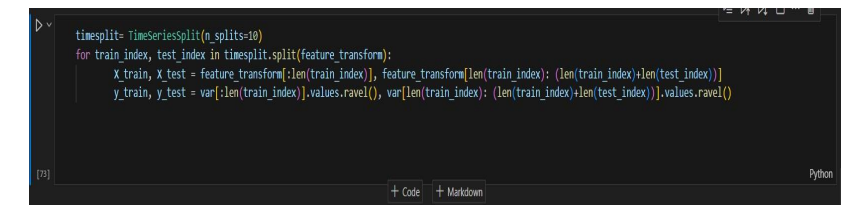
**Explanation: The column for output is assigned to target variables .The feature is being serve as the independent variable to the dependent variable. On those Open, High, Low, Volume as features**

**Data normalization Explanation:**

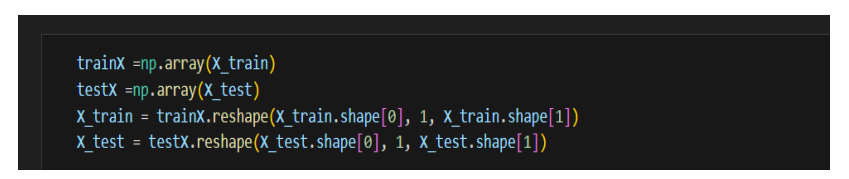
**normalizing the data for better efficient to model. By scaling the data. The values of each feature are scaled to a specific range, typically between 0 and 1**

****

**Creating a test set, training set and processing the data for model**

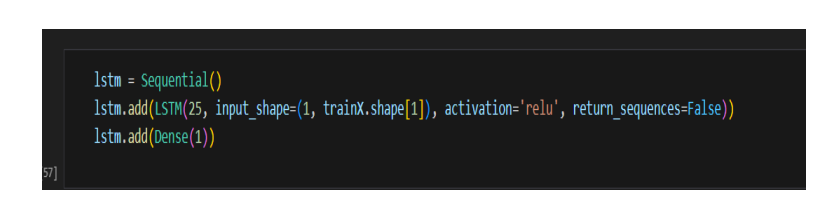
****

**Explanation: splitting the data into training set and test set for the model and split the 10% of the data for the test sets. And 90% of data for training the model**

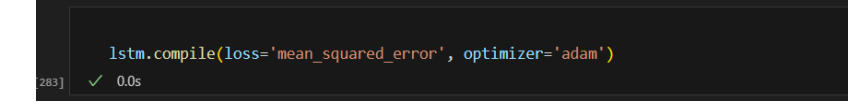
****

**Explanation: model can interact with data in its format. The data is being reshaped as arrays by numpy modules.**

**Building the model and Training the model**

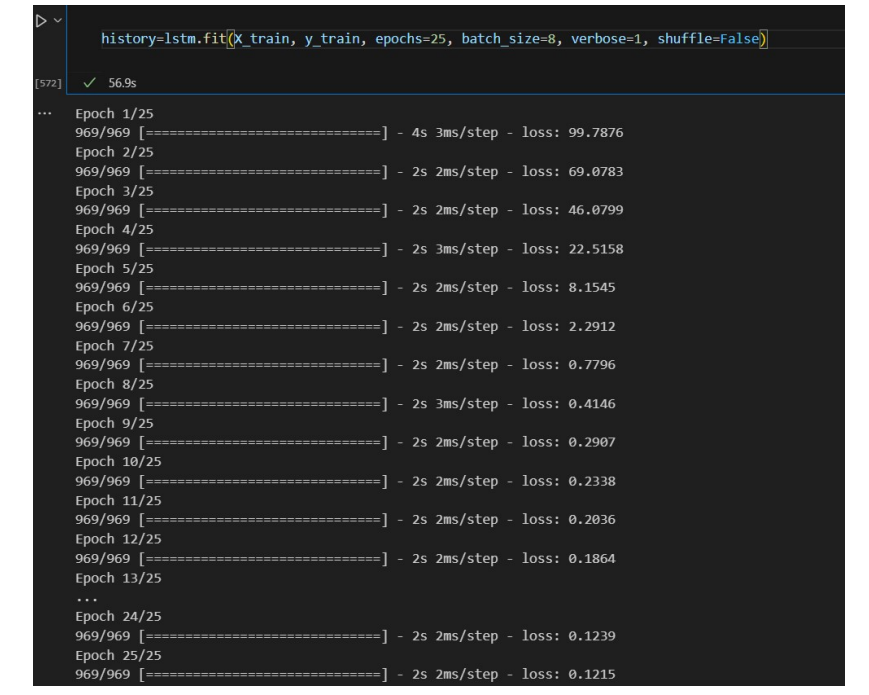
****

**Explanation: for a model we use an sequential keras with LSTM. The LSTM has 25 units and Dense layer of one neuron**

****

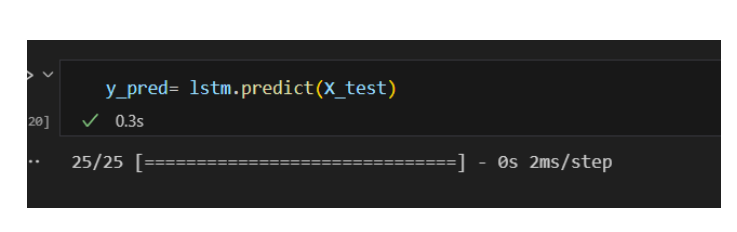
**Explanation: then compiling the model by adam optimizer and mean squared error**

**Training the LSTM model**

****

**Explanation: using the fit() function to train the LSTM model on the data for epochs with batch size of 8**

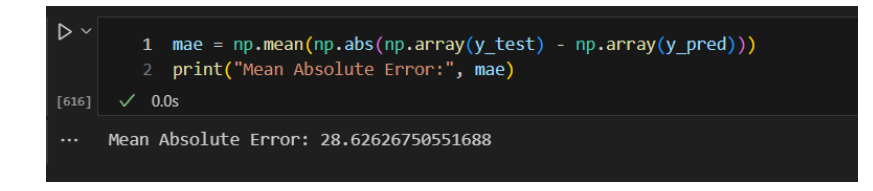
**LSTM prediction for data**

****

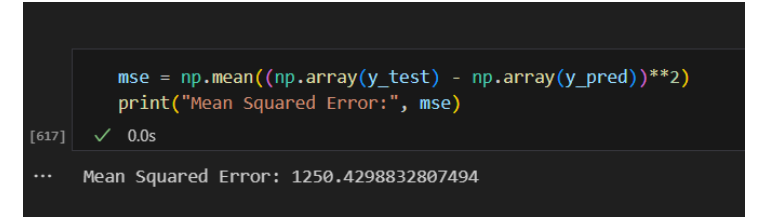
**Explanation: Using the LSTM model on the test data set for the data prediction.**

**Evaluations of data evaluation metrics to measure the accuracy and reliability of regression model's prediction. MEAN ABSOLUTE ERROR (MAE):**

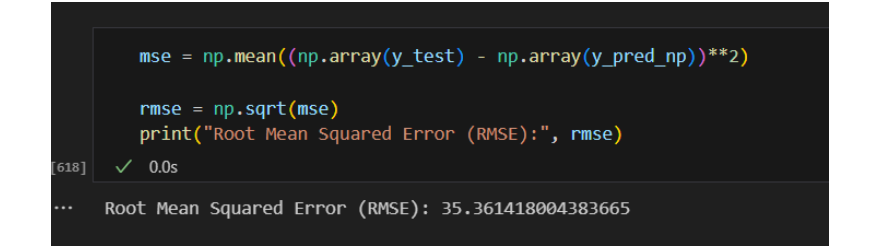
**Measures the average absolute difference between the predicted and actual values. Smaller MAE values indicate better accuracy**

****

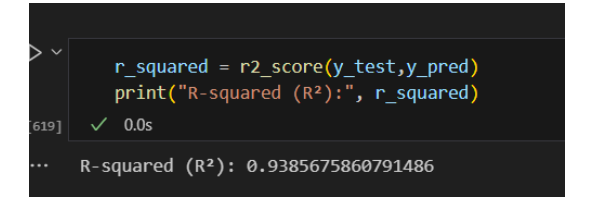
**MEAN SQUARED ERROR (MSE): Measures the average squared difference between the predicted and actual values. It penalizes larger errors more heavily and provides insight into the model's ability to capture variations in the data**

****

**ROOT MEAN ERROR (RMSE): The square root of MSE, which provides an interpretable metric in the same units as the target variable**

****

**R-SQUARED (R²): Also known as the coefficient of determination, R² quantifies the proportion of variance in the target variable explained by the model. A higher R² indicates a better fit to the data**

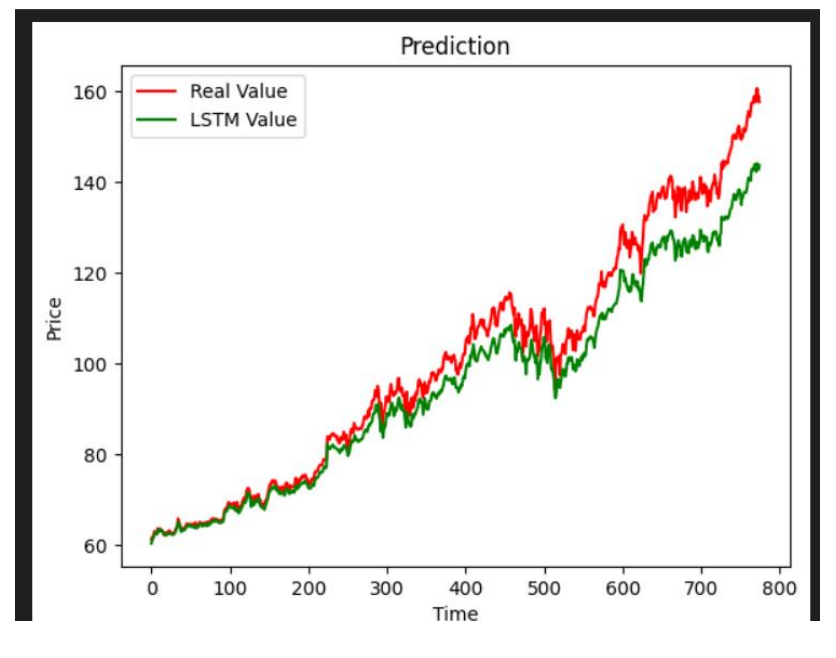
****

**: Comparing the data by visually**

****

**Explanation: plotting the data from the prediction**

**FinaAL PRDICTION**

****

**THANK YOU**