**Name:Rahul Kshirsagar**

**Class: CSE(AI) B div**

**Roll no.: 282018**

**PRN No.: 22310327**

**Assignment 6 – Regression Technique on Microsoft Stocks Dataset**

**Objective**

The objective of this assignment is to apply **Linear Regression** on the Microsoft Stocks dataset and evaluate the prediction performance using regression metrics such as **Mean Squared Error (MSE)**, **Mean Absolute Error (MAE)**, and **R-Square (R²)**.

**Dataset Used**

**Microsoft Stock Prices Dataset (MSFT.csv)**  
This dataset contains historical stock prices for Microsoft Corporation including:

* **Date**
* **Open**
* **High**
* **Low**
* **Close**
* **Adj Close**
* **Volume**

**a) Linear Regression Model**

**Data Preprocessing**

* The dataset was loaded and explored using .info() and .describe().
* The **‘Date’** column was converted to datetime format.
* Exploratory plots were created to visualize:
  + **Open vs Close stock prices**
  + **Volume trends**
* A **correlation heatmap** was plotted to understand relationships among features.

**Feature and Label Preparation**

* **Target variable (y)**: 'Close' price of the stock
* **Feature variables (X)**: All columns except 'Close'
* The **'Date'** column was dropped from features for regression.

**Train-Test Split**

* Dataset was split using train\_test\_split with **80% training** and **20% testing** data.

**Model Training**

* A **Linear Regression model** was built using LinearRegression() from sklearn.
* The model was trained on the training set and used to predict on the test set.

**b) Performance Evaluation Metrics**

| **Metric** | **Value** |
| --- | --- |
| **R² Score (train)** | Printed in output |
| **R² Score (test)** | r2\_score(y\_test, y\_pred) |
| **Mean Squared Error (MSE)** | mean\_squared\_error(y\_test, y\_pred) |
| **Mean Absolute Error (MAE)** | mean\_absolute\_error(y\_test, y\_pred) |

These metrics help evaluate how well the model is predicting stock prices.

**c) Visualization of Regression Model**

A simple linear regression plot was generated:

* **X-axis**: 'Open' price
* **Y-axis**: 'Close' price (actual vs predicted)
* A **scatter plot** shows the actual data points.
* A **regression line** indicates predicted values from the model.

**Conclusion**

Linear Regression was successfully applied to predict Microsoft stock closing prices based on various stock features. The model performance was assessed using standard regression metrics and visualized effectively. While linear regression gives a baseline, future enhancements may include advanced models like **Polynomial Regression**, **Random Forest**, or **LSTM** for better accuracy on time-series stock data.