**Stock Market Prediction**

### **Approach and Methodologies:**

* 1. Data Loading and Initial Analysis:
  + The project starts by loading the stock market dataset using Pandas.
  + Exploratory Data Analysis (EDA) is conducted to understand the dataset's structure, size, and information.
* 2. Data Visualization:
  + Time series visualization of the 'close' price is presented to understand the overall trend.
  + Data cleaning is performed by removing the 'adjclose' column since it duplicates the 'close' column.
* 3. Feature Engineering:
  + The dataset is further processed through feature engineering to create additional relevant features for model training.
  + Two new features are created: 'Daily\_Return' (percentage change in closing price) and 'Rolling\_Avg\_Close' (20-day rolling average of closing price).
* 4. Data Distribution and Outlier Analysis:
  + Distributions and boxplots of features are visualized to understand their spread and identify potential outliers.
* 5. Model Training:
  + The dataset is split into training and testing sets.
  + A Gradient Boosting Regressor model is chosen for prediction, initialized with hyperparameters such as the number of estimators, learning rate, and maximum depth.
* 6. Model Evaluation:
  + The trained model is evaluated using Mean Squared Error (MSE) as the performance metric.
* 7. Prediction Visualization:
  + The actual and predicted stock prices are visualized over time.
  + A scatter plot is used to compare sampled actual and predicted values for better understanding.
* 8. Absolute Difference Analysis:
  + The absolute differences between actual and predicted prices are visualized using a bar graph and a heatmap.

### **Insights Gained:**

* 1.Feature Importance:
  + The choice of features, including 'Daily\_Return' and 'Rolling\_Avg\_Close,' suggests an emphasis on capturing daily price changes and trends.
* 2.Model Performance:
  + The MSE of 0.1841 is used to quantify the performance of the Gradient Boosting Regressor.
  + Visualization of actual vs. predicted prices provides insights into the model's ability to capture trends and patterns in the data.
* 3.Absolute Differences:
  + Absolute differences between actual and predicted prices are analyzed to understand the magnitude of prediction errors.
* 4.Heatmap Analysis:
  + The heatmap provides a visual representation of absolute differences across different time points, offering insights into when the model performs well or poorly.

### **Chosen Predictive Models:**

1. **Gradient Boosting Regressor:**
   * Gradient Boosting is an ensemble learning technique that combines the predictions of several weak learners (usually decision trees) to create a strong predictive model.
   * It is well-suited for regression tasks and is known for its high predictive accuracy and robustness.
   * The MSE of 0.18415409509283281 underscores the effectiveness of the Gradient Boosting Regressor in capturing the underlying patterns in the stock market data.
   * Its high predictive accuracy, as reflected by the low MSE, reinforces its suitability for the regression task.