Stock price prediction Mini Project

**Title: Stock Price Prediction Using Linear Regression**

# Objective

The primary objective of this project is to develop a predictive model to forecast the stock prices of State Bank of India (SBIN.NS) using historical data and a linear regression approach. The aim is to provide investors with a reliable tool to anticipate future stock price movements, which is essential for making informed investment decisions in the dynamic stock market environment.

# Problem Statement

The problem at hand is to create a robust predictive model capable of accurately forecasting the future stock prices of SBIN.NS based on past historical data. This involves predicting the Adjusted Close price of the stock for the next trading day using various features such as opening, high, low, and closing prices from previous trading days. The challenge lies in developing a model that effectively captures the underlying patterns and trends in the data to make reliable predictions.

# Theory

**Linear Regression:** Linear regression is a statistical method used to model the relationship between a dependent variable (target) and one or more independent variables (features). In the context of stock price prediction, linear regression assumes a linear relationship between the features (such as opening, high, low, and closing prices) and the target variable (adjusted close price).

Mathematically, the linear regression model can be represented as:

𝑦=𝛽0+𝛽1𝑥1+𝛽2𝑥2+...+𝛽𝑛𝑥𝑛+𝜖*y*=*β*0+*β*1*x*1+*β*2*x*2+...+*βnxn*+*ϵ*

Where:

* + 𝑦*y* is the target variable (Adjusted Close price).
  + 𝛽0*β*0 is the intercept term.
  + 𝛽1,𝛽2,...,𝛽𝑛*β*1,*β*2,...,*βn* are the coefficients of the features 𝑥1,𝑥2,...,𝑥𝑛*x*1,*x*2

,...,*xn* respectively.

* + 𝜖*ϵ* is the error term representing the difference between the predicted and actual values.

The linear regression model aims to find the best-fitting line that minimizes the sum of squared differences between the actual and predicted values.

**Mean Squared Error (MSE):** Mean Squared Error (MSE) is a commonly used metric to evaluate the performance of regression models. It measures the average squared difference between the actual and predicted values. Mathematically, MSE is calculated as:

𝑀𝑆𝐸=1𝑛∑𝑖=1𝑛(𝑦𝑖−𝑦^𝑖)2*MSE*=*n*1∑*i*=1*n*(*yi*−*y*^*i*)2 Where:

* + 𝑛*n* is the number of observations.
  + 𝑦𝑖*yi* is the actual value of the target variable for observation 𝑖*i*.
  + 𝑦^𝑖*y*^*i* is the predicted value of the target variable for observation 𝑖*i*. A lower MSE indicates better accuracy in predicting the target variable.

# Implementation

**Data Collection:** Historical data for SBIN.NS spanning from January 1, 2016, to January 1, 2018, was obtained using the Yahoo Finance API. The dataset includes various attributes such as Date, Open, High, Low, Close, and Adjusted Close prices.

**Data Preprocessing:**

* + The Date column was transformed into a regular column for ease of handling.
  + Features (Open, High, Low, Close) were selected as predictors for the target variable, Adjusted Close price.
  + The dataset was split into training and testing sets, with a 75:25 ratio, to train and evaluate the model.

**Model Training:** A linear regression model was trained on the training set using the scikit-learn library's LinearRegression class. The model was fitted to the training data to learn the underlying relationships between the features and the target variable.

**Model Evaluation:** The performance of the trained model was evaluated using Mean Squared Error (MSE) on the test set. This metric provides an indication of how well the model is performing in predicting the next day's stock prices.

**Model Prediction:** After training and evaluating the model, it was utilized to predict the next day's Adjusted Close price based on the latest available data. The model takes the opening, high, low, and closing prices of the most recent day and predicts the Adjusted Close price for the next trading day.

# Conclusion

In conclusion, the linear regression model exhibited reasonable performance in predicting the stock prices of SBIN.NS based on historical data. Although the model's accuracy is acceptable, there is still room for improvement by exploring more sophisticated machine learning algorithms and incorporating additional features. However, this project demonstrates the potential application of machine learning techniques in stock price prediction, offering valuable insights for investors to make informed decisions in the financial markets. With further refinement, such models can play a significant role in optimizing investment strategies and managing risks effectively.

# Future Work

In future iterations of this project, several enhancements and explorations can be pursued:

* + **Advanced Algorithms:** Investigate the performance of more advanced machine learning algorithms such as Random Forest, Gradient Boosting, or even deep learning models like Long Short-Term Memory (LSTM) networks for improved prediction accuracy.
  + **Feature Engineering:** Explore the inclusion of additional features such as trading volume, technical indicators (e.g., Moving Averages, Relative Strength Index), and external factors (e.g., news sentiment analysis, economic indicators) to capture more nuanced patterns in the data.
  + **Real-time Prediction:** Develop a real-time prediction system that continuously updates based on the latest market data, providing investors with up-to-date forecasts to adapt their investment strategies accordingly.

# References

* + Academic papers, online tutorials, or documentation of libraries used.