

# Predict Future Stock Prices (Short-Term)

## Task Objective:

The goal of this project is to predict the next day's stock closing price using historical stock market data. We use features such as Open, High, Low, and Volume to predict the Close price using a Linear Regression model.

## Dataset Used:

- Source: Yahoo Finance (via yfinance Python library)
- Stocks: Apple (AAPL)
- Time Range: January 1, 2015 – Present
- Features: Open, High, Low, Volume
- Target: Close

## Step-by-Step Code Explanation:

### 1. Install and Import Libraries:

- We use yfinance to fetch stock data, pandas and numpy for data manipulation, and sklearn for modeling.

### 2. Data Collection:

- Data from 2015 to today is fetched using yfinance's download function

### 3. Feature and Target Selection:

- We select 'Open', 'High', 'Low', and 'Volume' as features and 'Close' as the target variable.

### 4. Train-Test Split:

- 80% training and 20% testing, without shuffling to preserve time-series integrity.

### 5. Model Training:

- A Linear Regression model is trained on the training data.

### 6. Model Evaluation:

- RMSE and R2 Score are used to evaluate model performance.
- RMSE indicates prediction error in dollars.
- R2 shows how well the model explains price variability.

### 7. Visualization:

- A line plot is created to show actual vs predicted prices.

## 8. Prediction:

- The latest data point is used to predict the next day's closing price.

## Results and Insights:

Stock | RMSE | R2 Score | Next Day Price

Apple | 1.16 | 1.00 | \$225.34

## Insights:

- Linear Regression performs very well on Apple stock.
- Apple's smooth price changes make it slightly easier to predict with higher accuracy
- Very low RMSE and perfect R2 indicate strong model performance

## Conclusion:

This project demonstrates the effectiveness of Linear Regression in short-term stock price forecasting.

It uses real-time financial data, applies basic feature engineering, and produces highly accurate results.

Future improvements could include testing additional models like Random Forest or incorporating technical indicators.