

Stock Price Prediction — Apple (AAPL)

Introduction

The goal of this project is to predict the closing price of Apple Inc. (AAPL) stock using historical market data. Stock price prediction is important for investors, analysts, and researchers in finance as it helps in decision-making and understanding market dynamics.

I applied regression techniques (Linear Regression, Lasso Regression, and Random Forest) to build predictive models. Feature selection methods were used, and models were evaluated using metrics such as MAE, RMSE, and R^2 .

Dataset Description

The dataset contains 5 years of daily historical stock prices of Apple (AAPL) collected from Yahoo Finance. It consists of the following columns:

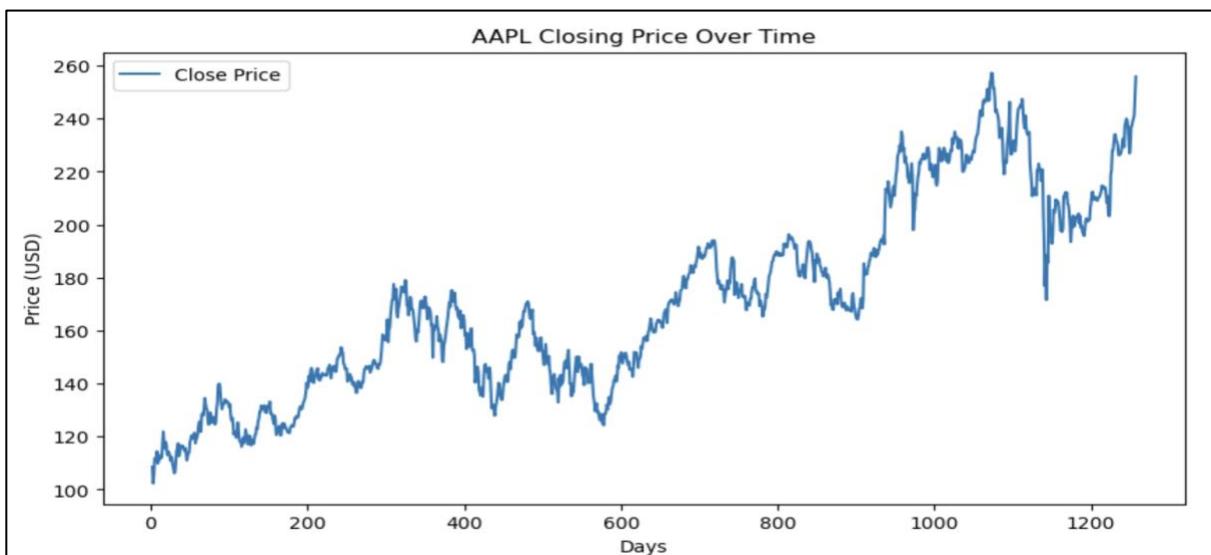
- Date: Trading day
- Open: Price at the start of trading
- High: Highest price of the day
- Low: Lowest price of the day
- Close: Price at market close (target variable)
- Volume: Number of shares traded

No missing values were found in the dataset after cleaning.

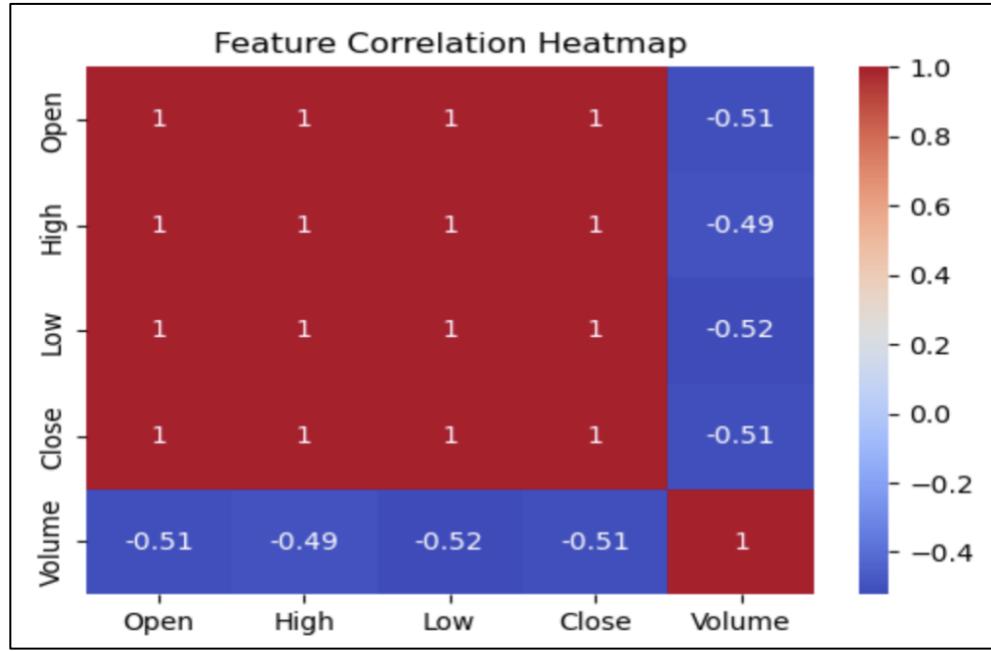
Exploratory Data Analysis (EDA)

I performed EDA to understand the data distribution and relationships:

- Plotted the trend of closing prices over time.
- Generated a correlation heatmap to examine feature relationships.
- Created scatter plots of Volume, Open, and High against Close to observe dependencies.



[Plot 1: Closing Price Trend]



[Plot 2: Correlation Heatmap]

Methodology

1. Data Preprocessing:

- Cleaned dataset and ensured numerical consistency.
- Split dataset into 80% training and 20% testing sets.
- Scaled features using StandardScaler for linear models.

2. Feature Selection:

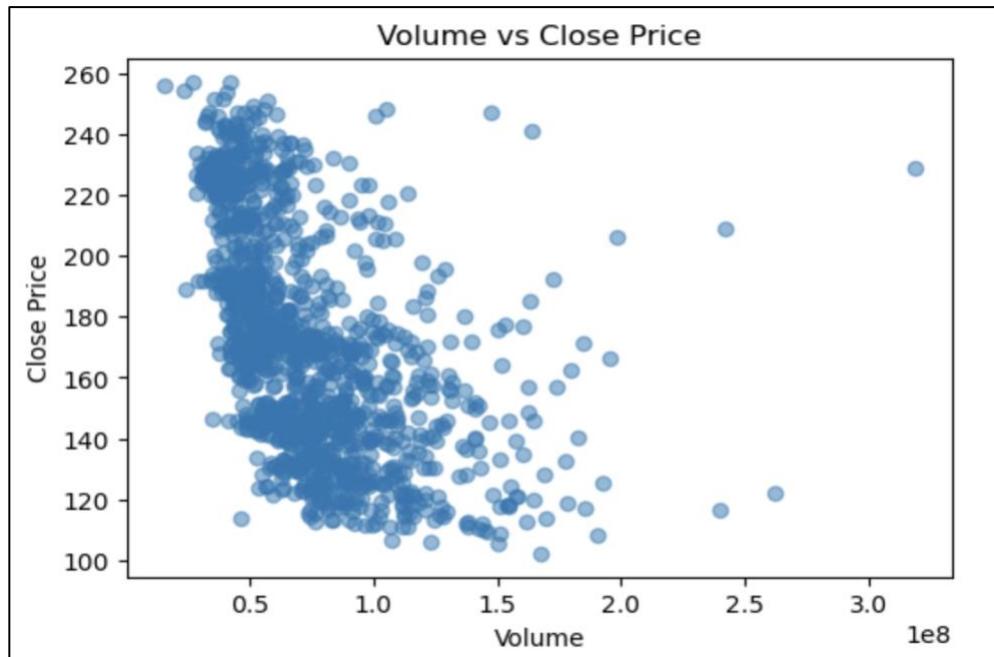
- Used SelectKBest (f_regression) to score all features.
- Applied Recursive Feature Elimination (RFE) to identify the best predictors.

3. Models Implemented:

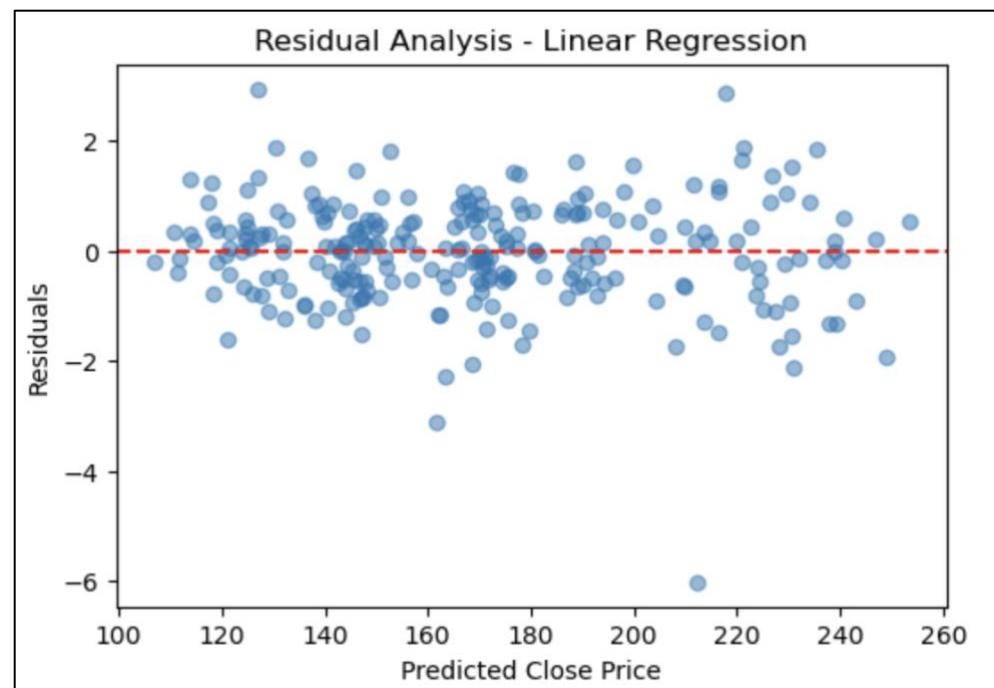
- Linear Regression (baseline model)
- Lasso Regression (L1 regularization)
- Random Forest Regressor (ensemble model)

4. Model Evaluation:

- Metrics: MAE, RMSE, R²
- Residual analysis performed for Linear Regression



[Plot 3: Model Comparison (R^2)]



[Plot 4: Residual Analysis]

Results

The performance of the models is summarized in the table below:

Model	MAE	RMSE	R ²
Linear Regression	0.80	1.07	0.999
Lasso Regression	0.93	1.21	0.999
Random Forest	1.09	1.52	0.998

Conclusion

In this project, I successfully predicted Apple's stock closing price using historical data. I compared Linear Regression, Lasso, and Random Forest models, and found that all achieved high accuracy. Linear Regression performed best due to strong linear relationships.

Future work could include adding external features such as market indices, news sentiment, and macroeconomic indicators, as well as testing advanced time-series models like ARIMA or LSTMs.