CHRIST (Deemed to be University)

Department of Data Science and Statistics

MDS 372

Report On Stock Market Analysis

Course Title: Machine Learning

Name: Mohammed Yasir Khattib

Jarin J V

Regno: 2348043

2348031

Class: 3MDS

Dated: 02-05-2024

Index

1. Introduction

2. Data description

3. Objective

4. Methodology

5. Exploratory Data Analysis

6. Models

7. Comparison

8. Conclusion

Introduction

The analysis in this report focuses on predicting the closing stock price of Apple Inc. using various machine learning models. The dataset spans a period of 11 years, from 2012 to 2023. We begin with exploratory data analysis (EDA) on four technology stocks: Apple, Google, Microsoft, and Amazon. Subsequently, we narrow down our focus to Apple for modeling purposes. The primary objective is to develop accurate models for predicting Apple's closing stock price based on historical data. The methodology involves data preprocessing, feature engineering, model training, and evaluation. The report concludes with a comparison of the performance of different models and a discussion of implications.

Data description

The dataset comprises daily stock prices for Apple, Google, Microsoft, and Amazon from January 2012 to April 2023. It includes features such as Open, High, Low, Close, Adjusted Close, and Volume. The data is obtained from Yahoo Finance using the pandas_datareader library. Additionally, the dataset contains company names for identification purposes. The primary focus is on the Adjusted Close price, which accounts for corporate actions such as dividends and stock splits.

Objective

The primary objective of this analysis is to develop accurate machine learning models for predicting the closing stock price of Apple Inc. The models will utilize historical stock price data along with relevant features to make predictions. The goal is to explore different modeling techniques and evaluate their performance based on metrics such as root mean squared error (RMSE) and accuracy.

Methodology

Exploratory Data Analysis (EDA): Conducted EDA on the four technology stocks to gain insights into their historical performance. Visualizations were used to analyze closing prices, sales volume, moving averages, daily returns, and correlations between stocks.

Data Preprocessing: The dataset was preprocessed by scaling the features using MinMaxScaler and splitting it into training and testing sets. For modeling purposes, a time window of 60 days was chosen for input features.

Model Selection: Three machine learning models were selected for predicting Apple's closing stock price: Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), Random Forest Regressor, and Support Vector Regression (SVR).

Model Training and Evaluation: Each model was trained on the training dataset and evaluated using the testing dataset. The performance of the models was assessed using RMSE and other relevant metrics.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) revealed several insights into the historical performance of the four technology stocks: Apple, Google, Microsoft, and Amazon. Visualizations of closing prices, sales volume, moving averages, daily returns, and correlations between stocks provided valuable information for understanding their behavior over time.

Closing Prices: The historical view of closing prices showed trends and patterns in the stock prices of each company. From the visualizations, it was apparent that all four stocks experienced fluctuations over the analyzed period, with some periods of growth and others of decline. For example, Apple's closing price showed a general upward trend over time, with occasional dips and peaks.

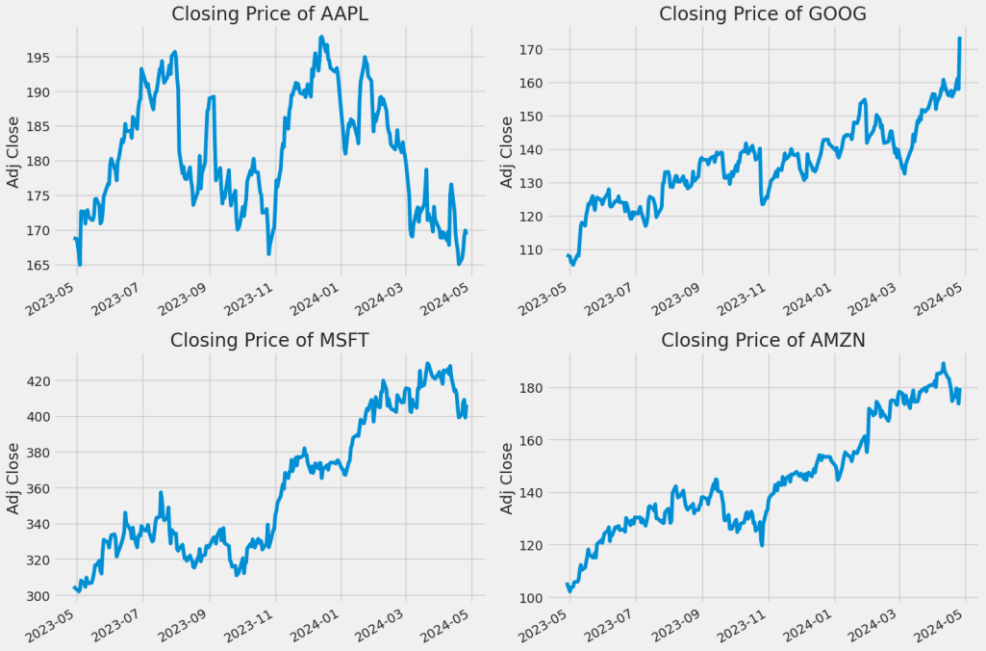

Fig1: Closing price of Apple, Google, Microsoft, Amazon

Sales Volume: Analysis of the total volume of stock traded each day provided insights into market activity. Higher trading volumes often coincide with significant price movements, indicating increased investor interest or trading activity. Observing sales volume alongside price movements can help identify trends and potential trading opportunities.

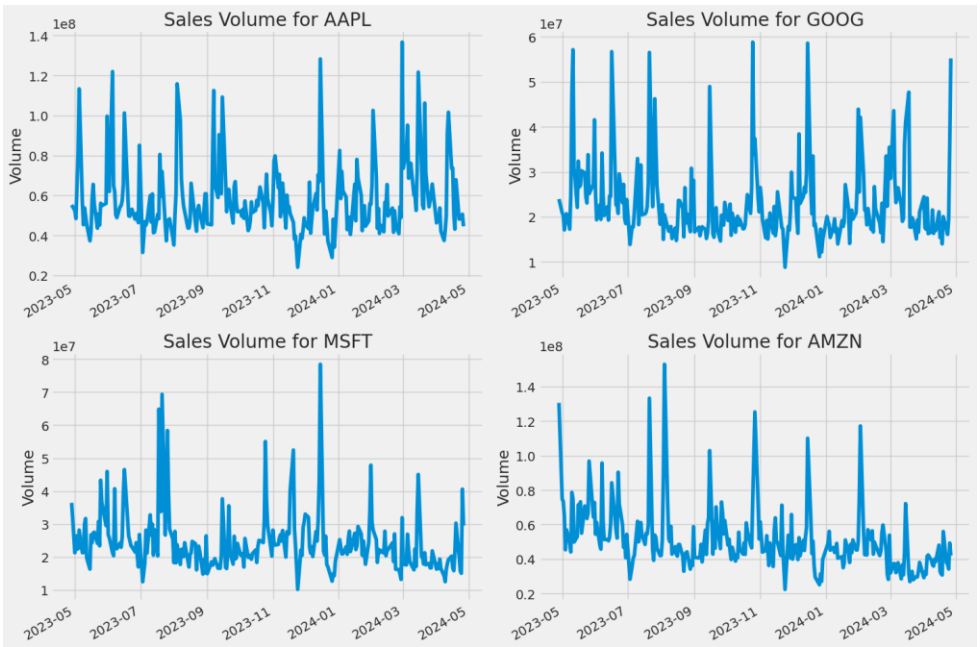

Fig2: Sales Volume of Apple, Google, Microsoft, Amazon

Moving Averages: The moving average plots for each company, calculated over different time periods (10, 20, and 50 days), helped smooth out price fluctuations and identify trends. Moving averages are commonly used technical indicators for determining the direction of a stock's trend. Crosses between short-term and long-term moving averages can signal potential buy or sell opportunities.

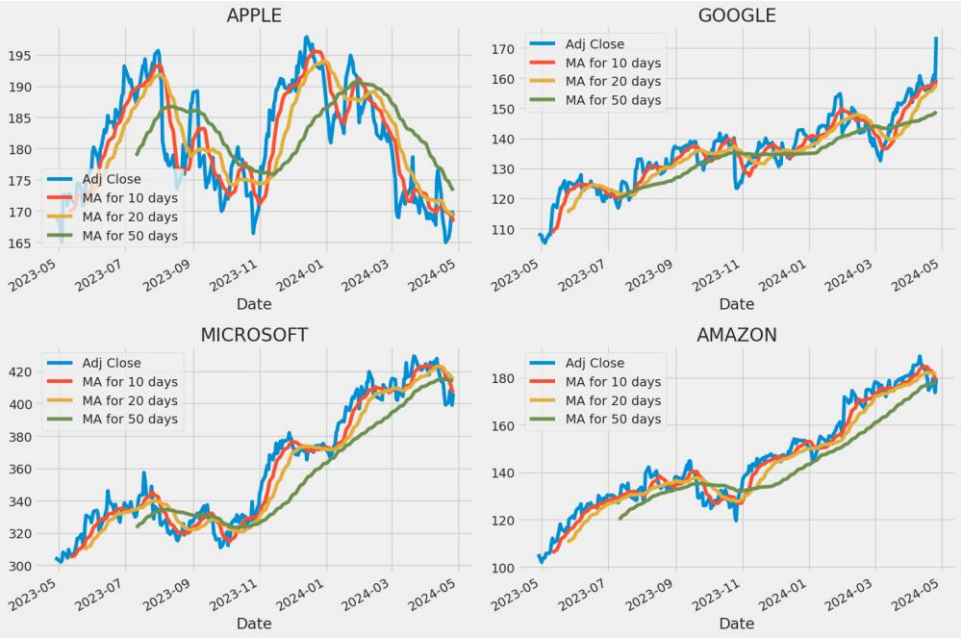

Fig3: Moving averages of Apple, Google, Microsoft, Amazon

Daily Returns: Analysis of the daily returns of the stocks on average revealed their volatility and performance relative to each other. Stocks with higher average daily returns may be considered riskier but could also offer higher potential returns. Histograms of daily returns provided insights into the distribution of returns and the frequency of different price movements.

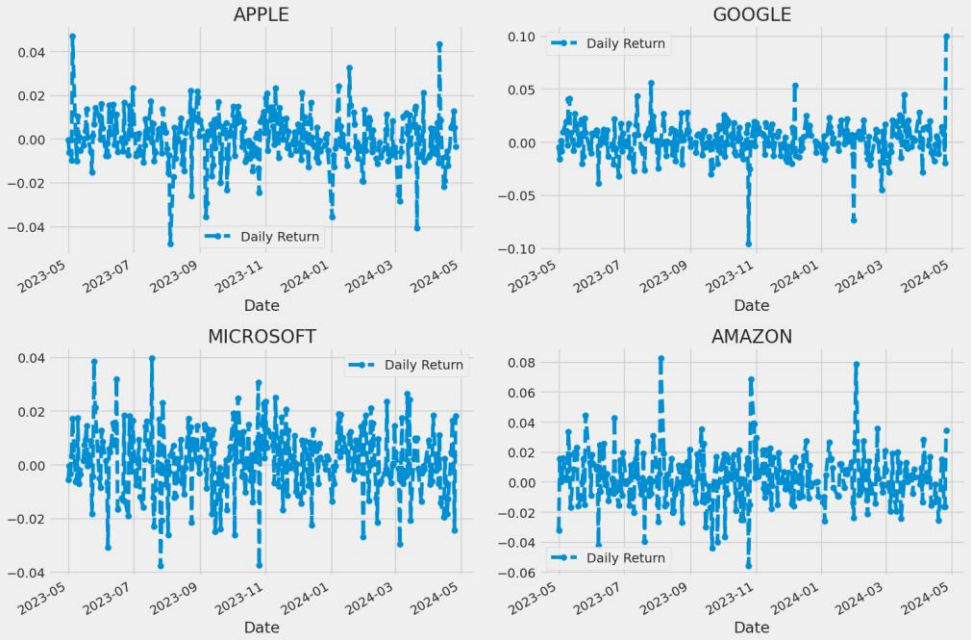

Fig4: Daily returns of Apple, Google, Microsoft, Amazon.

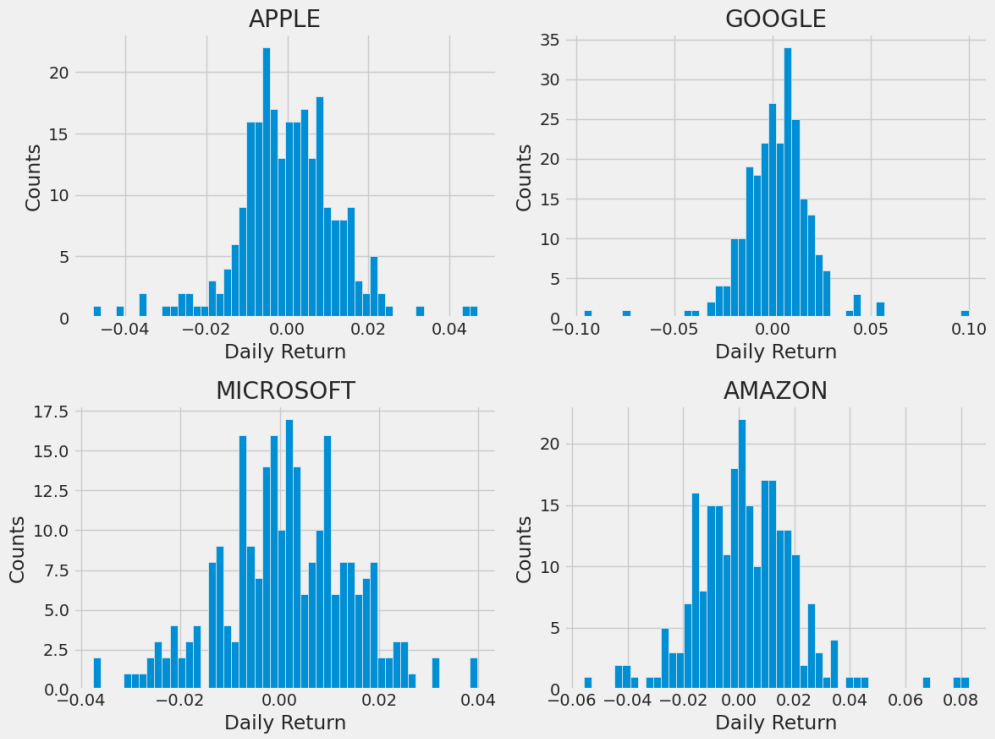

Fig5: Daily returns Histogram of Apple, Google, Microsoft, Amazon.

Correlation Analysis: Correlation analysis between the closing prices of the four stocks helped identify relationships and dependencies. High positive correlations suggest that the stocks tend to move in the same direction, while negative correlations indicate inverse movements. Understanding correlations between stocks is crucial for portfolio diversification and risk management.

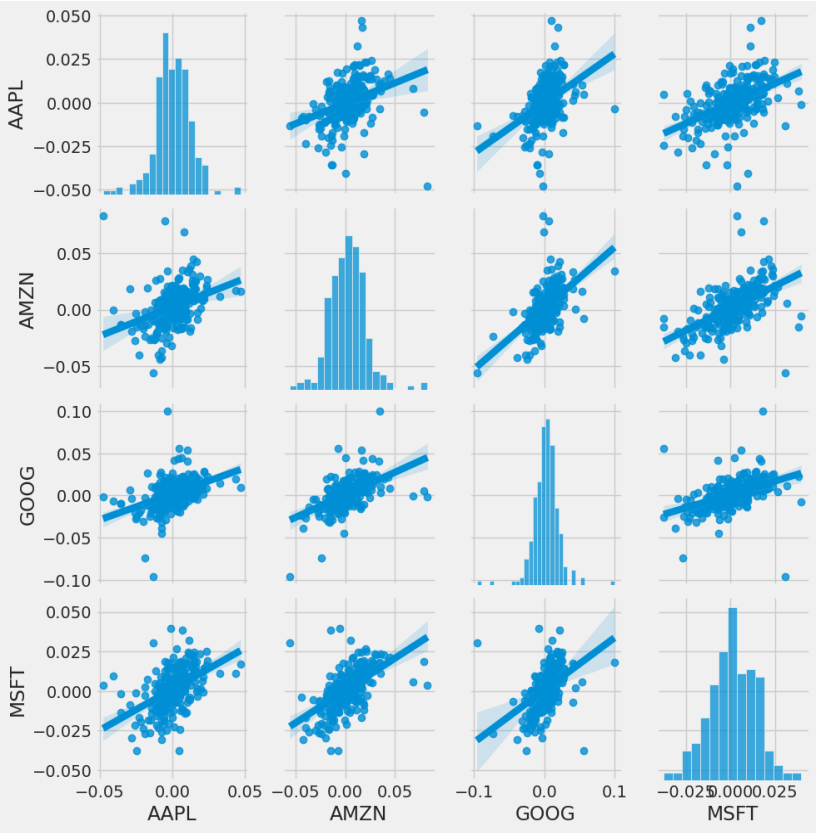

Fig6: Correlation (pairplot)

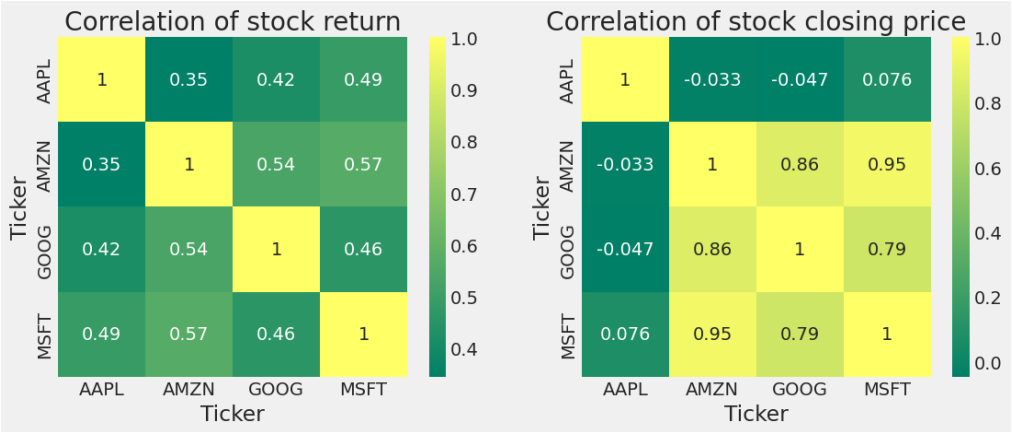

Fig7: Correlation (heatmap)

Moving to the modeling phase focused specifically on Apple, four machine learning models were trained and evaluated for predicting its closing stock price: Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), Random Forest Regressor, and Support Vector Regression (SVR).

Models

LSTM Model: The LSTM model achieved an RMSE of approximately 3.44, indicating relatively accurate predictions. LSTMs are well-suited for sequence prediction tasks like time series forecasting, making them effective for modeling stock prices that exhibit sequential dependencies.

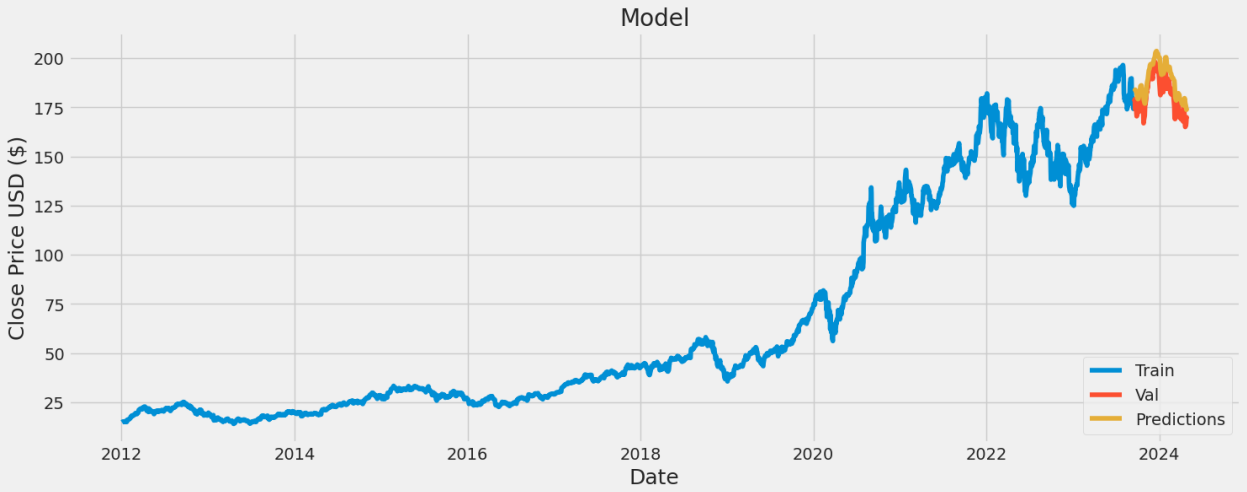

Fig8: LSTM on Apple Inc

CNN Model: The CNN model had an RMSE of around 7.86, indicating slightly lower prediction accuracy compared to the LSTM model. CNNs are primarily used for image processing tasks but can also be applied to sequential data like time series with appropriate data transformations.

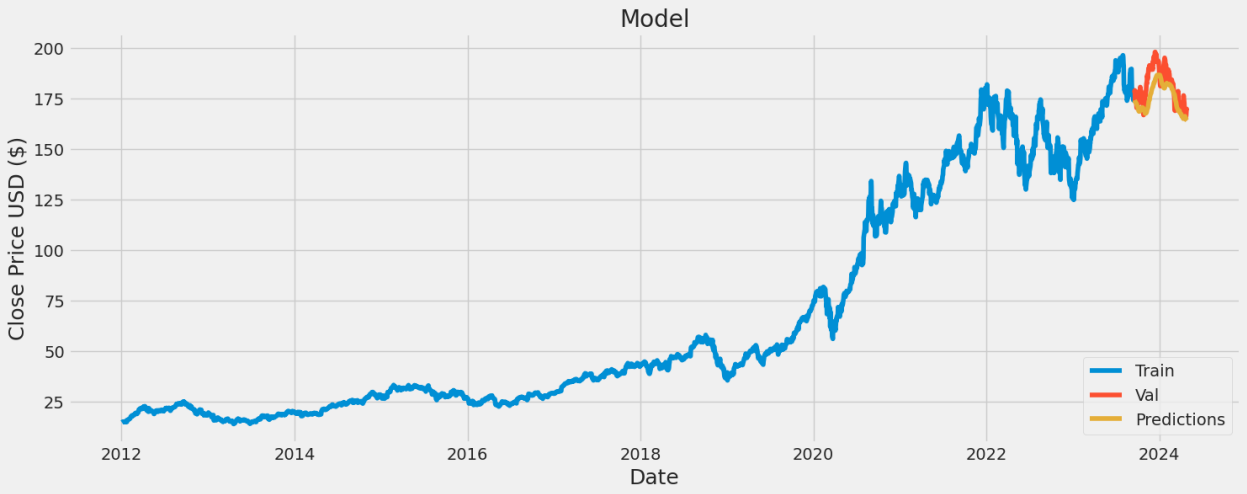

Fig9: CNN on Apple Inc

Random Forest Regressor: The Random Forest model performed well, with an RMSE of approximately 2.95. Random Forests are versatile and robust ensemble learning models capable of handling non-linear relationships and high-dimensional data.

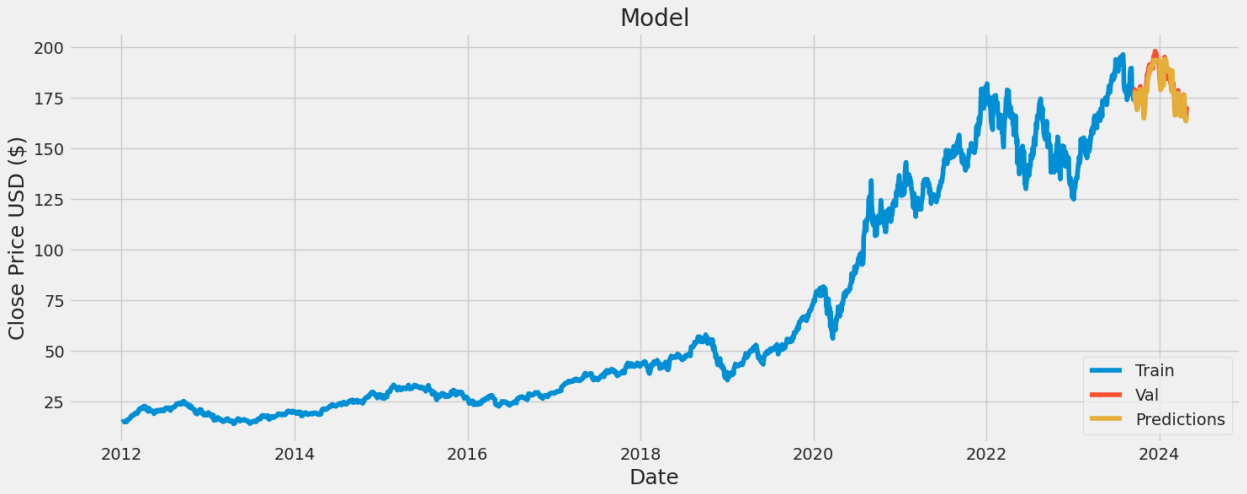

Fig10: Random Forest Regressor on Apple Inc

SVR Model: The SVR model performed the worst among the models evaluated, with an RMSE of approximately 9.74. SVR is sensitive to parameter tuning and may not perform well with noisy or non-linear data.

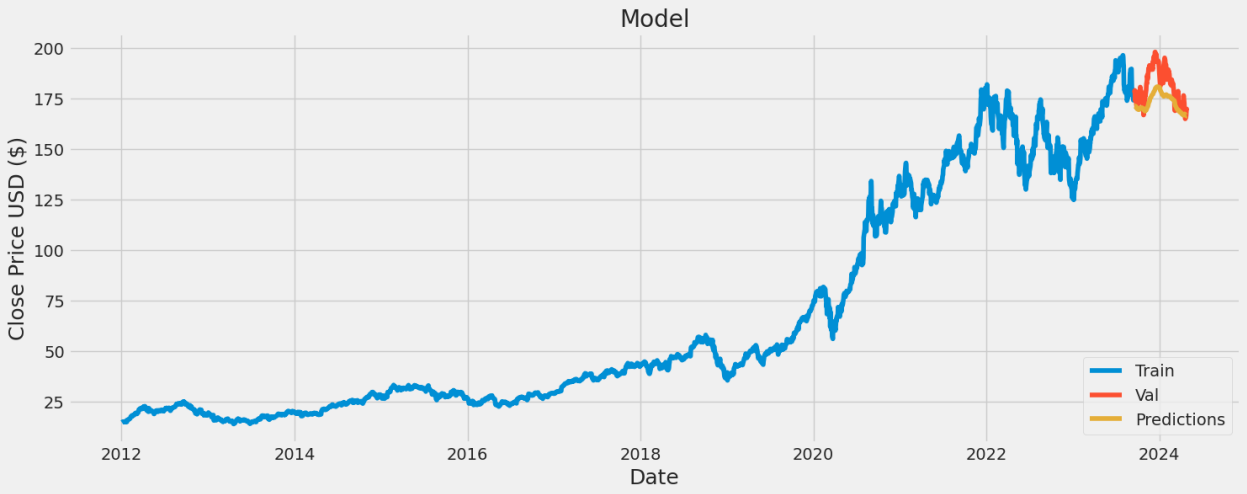

Fig11: SVR on Apple Inc

Comparison

The performance of the different models varied significantly. The LSTM and Random Forest models demonstrated relatively accurate predictions, with RMSE values below 5. On the other hand, the CNN model showed slightly higher RMSE, indicating less accuracy in prediction. The SVR model performed the worst, with the highest RMSE among all models. Overall, the LSTM and Random Forest models are preferred for predicting Apple's closing stock price due to their superior performance.

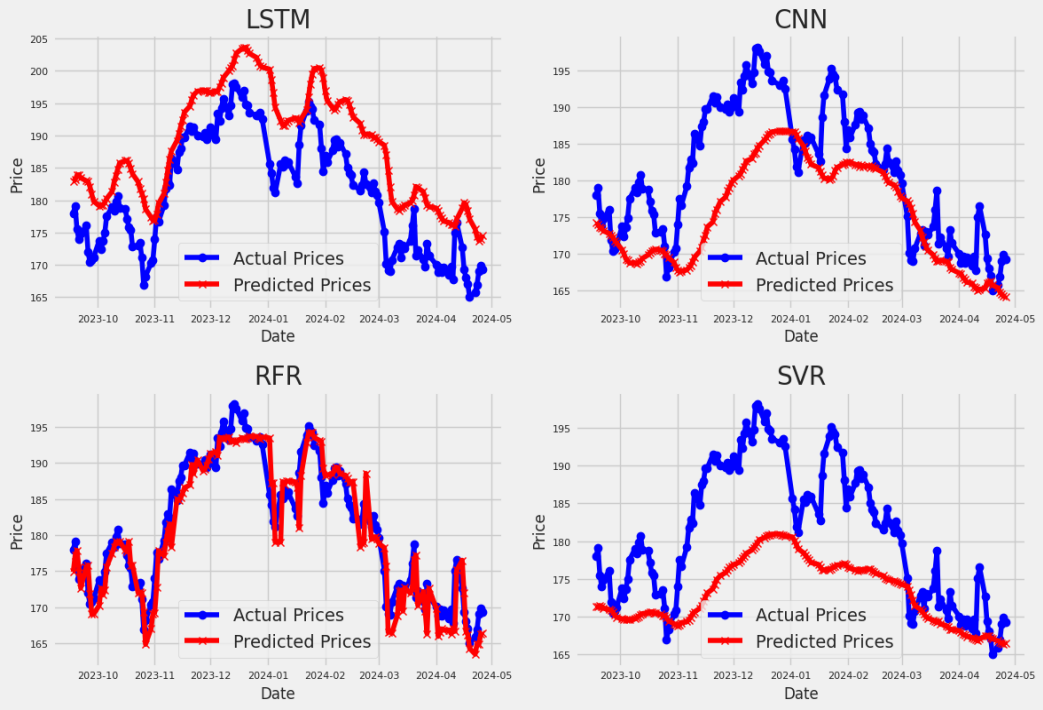

Fig12:

Comparison of LSTM, CNN, RFR, SVR

Conclusion

In conclusion, this analysis demonstrates the feasibility of using machine learning models to predict the closing stock price of Apple Inc. based on historical data. The LSTM and Random Forest models emerged as the top performers, achieving relatively accurate predictions with RMSE values below 5. These models can be valuable tools for investors and financial analysts in making informed decisions. However, further research and refinement of the models are recommended to enhance prediction accuracy and reliability. Overall, this study highlights the potential of machine learning in stock price prediction and its applications in the financial industry.