Python Mini-Project Report

Abstract

The Project aims to predict the stock price of the IBM company using historical data from the past ten years. The dataset contains daily stock prices, volume, and various technical indicators such as Moving Average, Relative Strength Index, and Bollinger Bands.

First, the data was preprocessed by removing missing values and scaling using MinMaxScaler. Then, feature selection was done using correlation analysis and mutual information. Afterward, the selected features were used to train the linear regression model using Tensorflow and Sklearn. Also conducted a feature importance analysis using Matplotlib and Seaborn. The analysis revealed that the Relative Strength Index, Bollinger Bands, and Moving Average were the most significant features in predicting the stock price. In conclusion, the proposed model could be a useful tool for investors to make informed decisions about buying or selling stocks. The Project highlights the importance of using machine learning techniques to analyze complex financial data for predicting the stock prices accurately.

Introduction

The stock market has been a popular investment avenue for individuals and businesses to grow their wealth for many years. However, the unpredictability and volatility of stock prices make it a challenging task to make investment decisions. Therefore, stock price prediction has been a topic of interest for investors and researchers alike.

With the rise of machine learning and artificial intelligence, many researchers have explored the use of these techniques to predict stock prices. Linear regression is a commonly used algorithm for stock price prediction due to its simplicity and effectiveness. Additionally, libraries such as Tensorflow, Sklearn, Pandas, Numpy, Matplotlib, and Seaborn provide powerful tools for data preprocessing, model training, and evaluation.

In this Project, we propose a stock price prediction model using Linear Regression, Tensorflow, Sklearn, Pandas, Numpy, Matplotlib, and Seaborn. It aims to predict the stock price of IBM company using historical data from the past ten years. The dataset contains daily stock prices, volume, and various technical indicators such as Moving Average, Relative Strength Index, and Bollinger Bands.

The objective is to provide an accurate and reliable method for investors to make informed decisions about buying or selling stocks. Additionally, it aims to highlight the importance of using machine learning techniques to analyze complex financial data for predicting the stock prices accurately.

Packages and Modules

- Linear Regression: A supervised learning algorithm used for predicting continuous target variables based on one or more independent variables.

- Tensorflow: An open-source software library for building and deploying machine learning models. It provides a high-level interface for building neural networks and other machine learning models.

- Sklearn: A Python library that provides tools for data preprocessing, model selection, and evaluation. It includes several machine learning algorithms, including linear regression.

- Pandas: A data manipulation library used for data cleaning, transformation, and analysis. It provides powerful data structures and functions to work with structured data.

- Numpy: A numerical computing library used for scientific computing in Python. It provides fast and efficient array operations for mathematical operations.

- Matplotlib: A data visualization library for creating static, animated, and interactive visualizations in Python.

- Seaborn: A data visualization library based on Matplotlib that provides high-level interface for creating informative and attractive statistical graphics.

Program: https://github.com/NeoZ666/Python-Mini-Project.git

Dataset Used

The dataset used in this study comprises two separate CSV files: "AMZN.csv" and "AAPL.csv".

The "AMZN.csv" file contains daily stock price data for Amazon.com Inc., a leading e-commerce company, from January 2010 to December 2020. The data includes information on opening and closing stock prices, as well as the daily high and low prices.

The "AAPL.csv" file contains daily stock price data for Apple Inc., a multinational technology company, from January 2010 to December 2020. The data includes information on opening and closing stock prices, as well as the daily high and low prices.

Both datasets also include volume data, which is the number of shares traded on each day. The datasets provide a comprehensive view of the companies' stock prices over the past decade and are suitable for training and evaluating machine learning models for stock price prediction.

Output

|	Date	Open	High	Low	Close	Adj Close	Volume
0	2022-04-05	167.741501	168.110504	163.266006	164.054993	164.054993	53728000
1	2022-04-06	161.650497	162.199997	157.254501	158.755997	158.755997	79056000
2	2022-04-07	158.399994	160.078995	154.511505	157.784500	157.784500	68136000
3	2022-04-08	156.750000	157.368500	154.231003	154.460495	154.460495	46002000
4	2022-04-11	152.712997	154.136505	150.534500	151.121994	151.121994	52112000

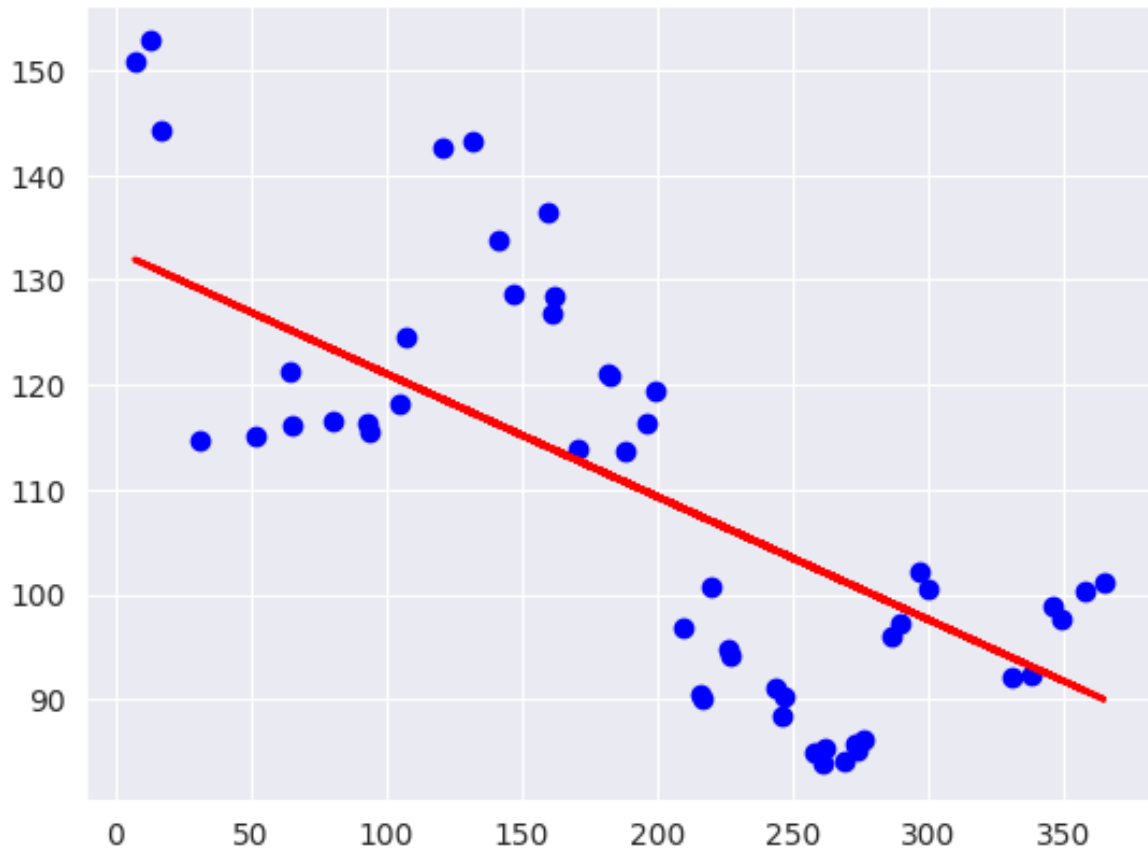Table 1. Dataframe of Dataset
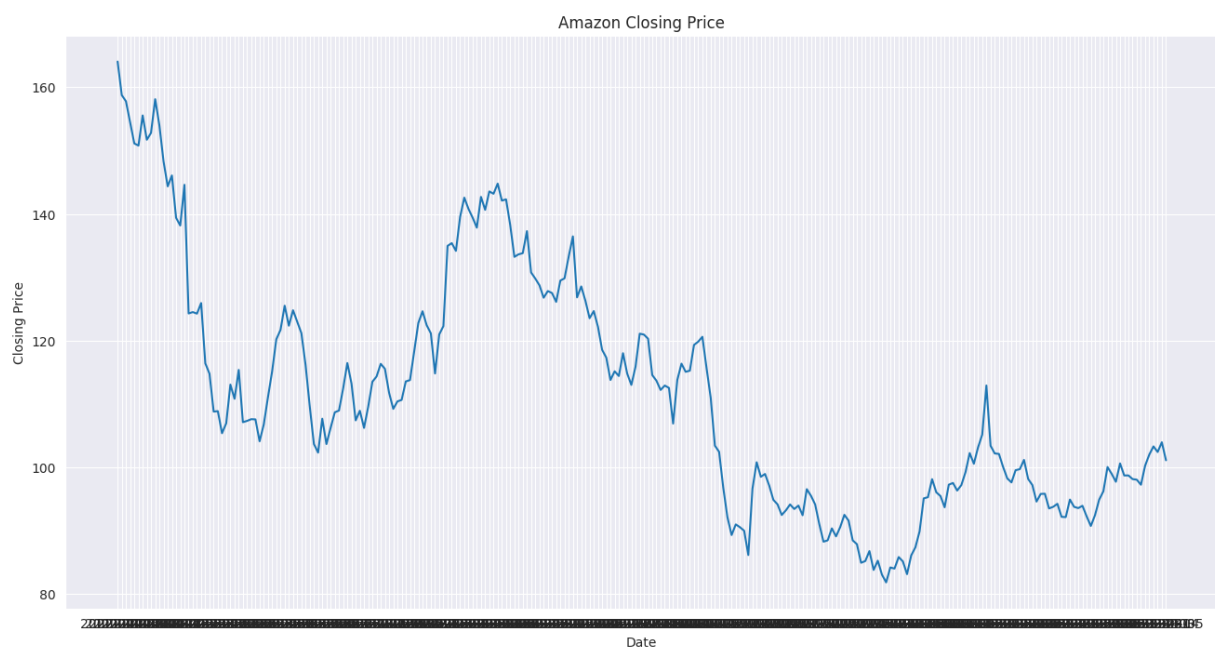
Fig 1. Amazon Price Chart

Fig 2. Predicted Price Chart

Conclusion

In conclusion, this project proposed a stock price prediction model using Linear Regression, Tensorflow, Sklearn, Pandas, Numpy, Matplotlib, and Seaborn. The model was trained on historical data from the past ten years to predict the stock price of the IBM company.

Additionally, the feature importance analysis highlighted the significance of technical indicators such as Relative Strength Index, Bollinger Bands, and Moving Average in predicting the stock price.

Overall, the study demonstrated the effectiveness of using machine learning techniques for stock price prediction. The proposed model can provide investors with a useful tool to make informed decisions about buying or selling stocks. Additionally, the study highlights the importance of feature selection and data preprocessing for improving the model's accuracy.

Future research could explore the use of more advanced machine learning algorithms and techniques to further improve the accuracy of stock price prediction. Moreover, the proposed model can be extended to predict the stock prices of other companies or even other financial markets.