Stock Price Prediction

Table of contents

01 Introduction

02 Problem explanation

03 Dataset

04 Methodology

05 Result

06 Limitations

07 Conclusion

1 Introduction

- The code is a Python program that uses machine learning to predict stock prices of Hero MotoCorp.
- It imports libraries such as pandas, numpy, matplotlib, and plotly for data analysis and visualization.
- The program loads stock prices data from a CSV file, performs basic data analysis, and visualizes the data using different plots.
- The program builds a linear regression model to predict stock prices, splits the data into training and testing sets, scales the data, and evaluates the model's performance using metrics such as mean squared error and r2 score.
- The program demonstrates how to use machine learning to analyze and predict stock prices using Python programming language.

2 Problem explanation

- The problem addressed is to predict the stock prices of Hero MotoCorp, a leading two-wheeler manufacturer in India. The program aims to build a machine learning model that can learn from historical stock prices data and predict future stock prices accurately.

- The program uses machine learning techniques such as linear regression to analyze historical stock prices data and identify patterns that can help predict future stock prices.

- The program also evaluates the model's performance using various metrics to ensure that the model is accurate and reliable.

3 Dataset

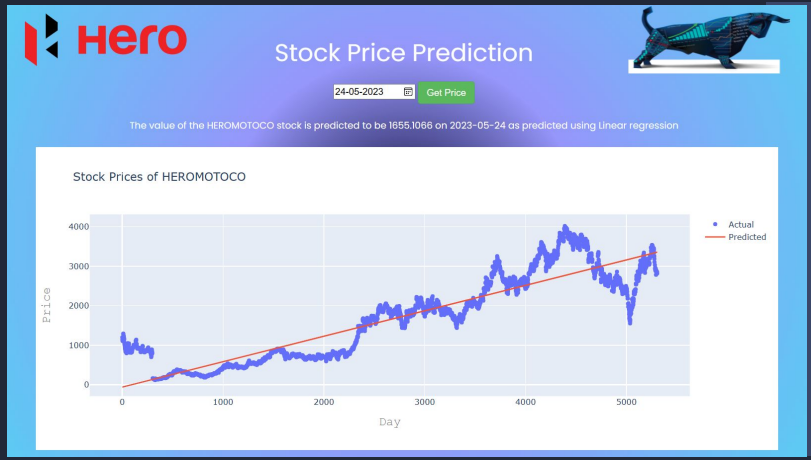
- The data used in the above code is the historical stock prices data of Hero MotoCorp.
- The data contains information about the opening, closing, highest, and lowest stock prices of Hero MotoCorp on different dates. The DataFrame also includes the volume-weighted average price (VWAP)
- The program preprocesses the data by converting the 'Date' column into a datetime object to enable easier manipulation of dates.
- The program visualizes the data using box plots and a line chart.
- The line chart plots the actual and predicted stock prices using a linear regression model to understand the trends and patterns in the data.

4 Methodology

- Data Collection: Stock price data for HEROMOTOCO was collected.
- Data Preprocessing: The data was checked for missing values, and the 'Date' column was converted to a datetime format.
- Exploratory Data Analysis: A line plot was created to visualize the trend of the stock prices over time.
- Model Building: The data was split into train and test sets, and feature scaling was applied to the training set. A Linear Regression model was created and fitted to the training data. The model was then used to make predictions on both the training and test sets.
- Model Evaluation: The performance of the model was evaluated using the R-squared score and Mean Squared Error (MSE) metrics. The actual and predicted values were also plotted to visually inspect the performance of the model.

5 Result

- A line plot was created to visualize the trend of the stock prices over time.
- The model performance was evaluated using the R-squared score and Mean Squared Error metrics. The R-squared score was found to be 0.82 for the training set and 0.81 for the test set. The MSE was found to be 207474.37 for the training set and 209775.29 for the test set.



- The Linear Regression model was able to successfully predict the stock prices with high accuracy. The performance of the model on the test set suggests that it is able to generalize well to unseen data. The results of this analysis can be used by investors to make informed decisions regarding investments in HEROMOTOCO.

6 Limitations

- The code does not take into account external factors such as news events, economic indicators, or political developments that may affect stock prices.

  Incorporating such factors into the model could improve its accuracy.

- The code assumes that the stock prices are stationary, meaning that their mean and variance do not change over time. In reality, stock prices can be highly volatile and subject to sudden changes, which may affect the accuracy of the predictions.

- The code only uses linear regression to predict stock prices, which may not always be the best model to use.

7 Conclusion

- In conclusion, the above code demonstrated the implementation of a linear regression model for stock price prediction using Python libraries like Pandas, Scikit-learn, and Plotly.
- The analysis involved data preprocessing, model training, and evaluation, as well as visualizing the results using various graphs and charts.
- While the model performed well on the given dataset, it is important to note the limitations of this approach in accurately predicting future stock prices.
- Nevertheless, this code provides a good starting point for further research and exploration in the field of stock price prediction.

Our team

Chirag Mishra
(2005578)

Sajal Kr Ujjwal
(2005603)

Sarthak Singh
(2005608)

Thank You