

Trinity College Dublin School of Computer Science and Statistics

CS7CS4-20212022 Group Project

Machine Learning - Group Project
Future Stock price prediction

Report By

Kalaivani Kandasamy 21332195

Lokesh Selvakumar 21331969

Shubham Maurya 20310455

Motivation:

News, earnings reports, etc., are just some of the factors that cause stock markets to react quickly. Fundamental data may prove valuable in developing trading strategies, but rapid changes in stock markets are hard to predict and may not suit short-term traders. Using data science, this study attempts to forecast future prices in a way that maximizes an investor's chances of success.

Data Set:

Scraping is the process of extracting data from a website. Software tools like web scrapers are typically used to achieve this. Then the scraped data is exported in the form of an Excel spreadsheet or JSON.

We will scrape yahoo finance's web pages. The stock market data includes opening price, closing price, low price, high price, last price and Total Trade Quantity.

- Opening price and Closing price reflects the starting and closing price of the stock on a particular day.
- High price, Low price and Last price represent the maximum, minimum, and last price of the share for the day.
- Total Trade Quantity is the number of shares bought or sold in the day

Effect of Sentiment Analysis on stock prices:

In addition to the above features, we scrap the sentiment analysis data like sentiment, direction, mentions as additional features and it will be merged with existing data. These features are then used to predict the future stock price and the effect of sentiment analysis features on future stock price prediction will be analyzed.

Stock price trend among similar companies:

Another factor which we want to consider is the effect of one stock on another stock within the same industry. It is observed that when the stock price of one company increases it affects the stock prices of other companies within the industries.

For example, if we predict the stock price of power companies like Adani Power, Tata Power, Reliance Power, similar trend is observed across the stock prices of these companies as there is some correlation between them.

Yahoo finance:

Yahoo Finance is a media property that is part of the Yahoo! network. It provides stock quotes, press releases, financial reports, and original content, alongside financial news, data, and commentary. Personal finance management tools are also provided. In addition to posting partner content from other web sites, it posts original stories by its team of staff journalists. It is ranked 15th by Similarweb on the list of largest news and media websites.

Method:

The following models will be trained using the data scraped using yahoo finance and forecast stock prices.

Linear regression:

Linear regression is a basic approach for obtaining a linear trend. It is a supervised Machine Learning model that determines the best fit linear line between the independent and dependent variables, or the linear connection between the two variables. Linear regression with multiple variables otherwise known as multivariate Linear regression is used to predict the future stock prices.

Support vector machine:

SVMs, on the other hand, have benefits such as high accuracy and predictability compared with linear regression. We will use this as a second machine learning algorithm to predict the stock prices.

Intended Experiments:

We will be using jupyter notebook or google colab for data pre-processing, training and forecasting the data and model evaluation.

Performance evaluation:

The performance measures that were used to assess the predictive accuracy of the proposed system included the root mean square error (RMSE), the mean absolute error (MAE), the mean absolute percentage error (MAPE), the mean square error (MSE). These indexes are used to measure whether the predicted values are close to the actual values.

Baseline model: We will be using persistence algorithm as a baseline model for this time series problem. This algorithm uses the data at the previous time step to predict the outcome at the next time step.

Using all the above-mentioned machine learning techniques, future stock prices are predicted, and accuracy of each model is compared with each other to find the best model for future stock price prediction. Effect of sentiment analysis features on stock prices and correlation between different stocks in the same industry will be analyzed and shown.