DSC680-T301 Applied Data Science
3/19/2023
Soma Shekar Vayuvegula
Project 1: Stock Market Analysis and Prediction

Topic

Stock market analysis and prediction will bear huge financial payoffs as this will decide the profit-making for many individuals/firms. With the amount of data available on the internet, Stock market analysis and prediction have gained a lot of momentum of late.

Business Problem

Analyzing data related to the stock market can be tricky because of the amount of data available, but it is equally interesting. Creating a proper prediction system will help in profit maximization.

Stock market data is complicated, and predictions are not straightforward as they depend on various factors. The stock market prediction is not always accurate, but we need some prediction to be done to invest in stocks.

The main aim of this project is to build a prediction system using different models to see past trends and forecast if the value of a stock will rise or fall.

Datasets

The datasets used in this project will be sourced from Kaggle, a public-domain dataset. I am planning to use data from two different sources and utilize the following data:

Date – Trade date.

Open – Open price of the day.

High – The highest price in a day.

Low – Lowest price in a day.

Close – Close price of the day.

Volume - Number of shares traded.

Name - Stock's ticker name

Methods

As part of this project, I will be building the below models:

1. Linear Regression: Linear Regression is used to predict the value of a variable based on the value of one or more explanatory variables. If we use one explanatory variable, it's called Simple Linear Regression; for more than one explanatory variable, it's called Multiple Linear Regression.
   Linear Regression will be run using close price as the dependent variable and open, high, low, and volume as the independent variables.
2. Decision Tree Approach: Decision Tree is a Supervised Learning used to make predictions based on how the previous set of questions is answered.
   Constructing a proper decision tree can be helpful in stock price prediction. We can see if the price of the stock will rise or fall.

I want to use Data Visualization to show the trends using Line graphs for any two stocks to show how the visualizations look like.

Ethical Considerations

We don't foresee any privacy issues as we are sourcing this data from public sites. However, stock related data is always posted on websites that are accessible to the public. Hence, there are no ethical issues.

The only issue is that data is not the latest so this might impact the model predictions.

Challenges/Issues

I will be using two different datasets, and if the data size creates a problem with running on my local machine, I might have to delete a few records, which can impact the analysis/prediction.

To build a system, we need business expertise along with technical knowledge. As I am not a Subject matter expert, the model implementation might need to be revised.

While investing in stocks, people will look at data for the last five years from the current date; we will only have visualizations on the current date.

The models built here in a short period might not be accurate; we might have to tweak them to make them accurate.

References

- Damarla, R. (n.d.). *Stock Market Prediction using Decision Tree*. Kaggle. Retrieved March 18, 2023, from https://www.kaggle.com/datasets/camnugent/sandp500?resource=download
- Mooney, P. (n.d.). *Stock Market Data (NASDAQ, NYSE, S&P500)*. Kaggle. Retrieved March 18, 2023, from https://www.kaggle.com/datasets/camnugent/sandp500?resource=download