Long-Term Stock Prediction Based on Financial Statements

Machine Learning Project (Group 4)

Done By:
Abhay Ajith - AM.EN.U4CSE19301

Abstract

Informal Definition:

To develop a long-term stock prediction model that uses various financial statements from the past few years. We use the **Supervised learning approach.**

Formal Definition:

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

Task(T) :

A long-term stock prediction based on the financial statements.

Experience(E):

The opening price, closing price, highest value, lowest values of the stock market over the past few years.

Performance(P):

To accurately predict the price of the stock market to the actual stock market.

.

Assumptions:

We assume that the stock prices vary depending on the timeline. The prices may vary as per the region as well as the time of that particular stock being sold. For example, ,start of the week, end of the week, holidays etc can have significance on the price of the stocks.

Introduction

Motivation:

The stock market is an area where a huge volume of data is generated and updated every second. People who invest in the stock market know the risks associated with stock markets. There are people who lost all of their money and became broke while there are people who double or triple their money. This is the main risk associated with investing in the stock market. Our project makes an attempt to accurately predict the stocks. This can be done in mainly two ways, one is daily prediction and the other is monthly prediction. In this project, we focus more on long term stock predictions.

For Example,

Consider that you are interested in buying stocks and have recorded the prices for some stock the past month. Now you want to gain profit by predicting when to buy or sell a particular stock.

Benefits of the solution:

This solution helps in reducing the risk associated with investing in the stock market. Even though the predicted output may be inaccurate, it provides a good form of assurance when it comes to the individual's decision of buying or selling a stock.

Dataset

We have an estimate of 25K of instances, divided into 12 dataset again divided into 4 students So each student has a total of 6500 instances of data.

The dataset is composed of 25,000 items: 21,000 in the training set, 2000 in dev set,1000 in test set. The dataset is crawled and parsed from https://www.nasdaq.com/.

The dataset has 6 columns: [Date, closing price, volume, open, high, low].

Date -> It denotes the closing date of a particular stock of a market.

Closing price -> This is considered as our target variable, it denotes the final price at which the following stocks are sold.

Volume -> This denotes the quantity at which a stock is sold.

Open -> It is the starting price of a stock

Low -> Lowest price recorded for a stock

High -> Highest price recorded for a stock