Predicting Stock Market Prices Using LSTMs

Statement of Work, Version 1
AI Algorithms - AIDI 1002
Course Facilitator - Marcos Bittencourt

Submission Date:
November 6th, 2020

1. Executive Summary

Already almost 30% of traffic on stocks is generated by machines. This project I will predict the next day price of a stock. For this purpose I will use the New York Stock Exchange Dataset that contains stocks. The project has two main goals that are mentioned below:

- Finding important patterns from the historical data
- Predicting the next day price using the historical data

2. Problem Statement

We hear about the stock market everyday in the news, about it reaching an all time high or a new low. The investment and business in the stock market will change considerably if we are able to predict what the value can be of an individual stock the next day.

Stock Market is the most popular medium to earn and invest money but the nature of it is highly stochastic which makes making predictions very difficult. Stock prices are driven by new information and changes.

Previously, Artificial Neural Net and Convolutional Neural Nets have been used to make such predictions which give an error loss of 20%.

In this project I will use Recurrent Neural Networks which will predict the stock price with a less percentage of error. I will build a model that will look at the historical data fed to it and predict a likely outcome of the stock market in the future. There is no proper prediction model for stock prices. The price movement is highly affected by the demand and supply ratio.

3. Data Requirements

For this project following are the data requirements:

1. A large dataset that contains historical prices of the stock from the last decade.
2. The dataset should contain the abbreviated names of the stocks.
3. It should contain daily values of the features:
    - Open and Close, indicating the starting and final price at which the stock is traded on a particular day.
    - Low and High, indicating maximum and minimum price of the day

For this project, I will be using the New York Stock Exchange Dataset available on Kaggle[1]. It contains prices of the stocks which were taken from Yahoo Finance, fundamentals taken from NASDAQ Financials and some features were taken from EDGAR SEC databases.

4. Test Process

For testing the algorithm I will be implementing a number of testing algorithms. Since it is a time series problem which is why I can not implement the Cross Validation test. Instead I will try to test the model by Multiple Train-Test splits and Walk-Forward Validation[2].

References

[1] https://www.kaggle.com/dgawlik/nyse

[2] https://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting/