

Predicting Stock Market Prices Using LSTMs

Statement of Work - Version 2

AI Algorithms - AIDI 1002

Course Facilitator - Marcos Bittencourt

Submission Date:

December 1st, 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. **Methodology**

a. **Normalization**

The dataset was normalized for faster convergence on learning and more uniform influence for all weights. The range was set from -1 to 1. I chose this range because there are some negative values in the dataset. For this I used `MinMaxScaler()`, which is a sklearn package.

The features that were normalized are open, high, close and low, as these are the main features and are fed to the model.

b. **Feature engineering**

For training and testing the model, I removed the features that would be considered noise, in this case, date, symbol and volume. The features we are left with are open, close, high and low. These features are the important feature form which we will be getting the prediction from.

For predicting the next day price, we have no use for volume value as it only gives details about stock trading. The stock symbol is not needed because we will be focusing on one particular stock and its prices. The date column becomes irrelevant since all the previous dates are historical, knowing their date stamp has no use in predicting the next day's price.

By doing so I was able to reduce the dimensions of the data with only relevant information intact.

c. Model Used

For this project I have used LSTMs (Long Short Term Memory) RNNs. The model consists of a basic LSTM layer. Even with one LSTM layer, the model is considered a deep neural network since it is able to process sequential data and memorize the hidden states through time.

With dimensionality reduction and feature reduction, the data is ready to be fed to the LSTM layer.

The model will be fed price data of a particular stock to get the price trend prediction result. The LSTM layer is composed of 2 LSTM layers with 50 batch size. There are a total of 200 neurons in the model. The model is made by using Tensorflow. I fed four inputs to the model and the output was four features.

References

- [1] <https://www.kaggle.com/dgawlik/nyse>
- [2] <https://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting/>