Capstone Proposal – Stock Price Prediction

Domain Background

Stocks are traded voluminously across the world almost every day. With increasing IPOs and firms looking to gain more capital, the investment and finance world will likely see this volume traded increasing throughout the years with even more traders trying to predict which are stocks worthy to buy. This project will be under the "Investment and Trading" domain.

Problem Statement

With the volatility of the price of stocks increasing nowadays, the risk of investing via stocks has risen. With the use of LSTM which uses an artificial recurrent neural network, it will help to predict the stock price based on historical stock price attributes.

Datasets and inputs

Historical Stock Data Prices - Open, High, Low – Close, Volume and Adjusted Close for the various stocks with their tickers and the date ranges can be obtained from Yahoo! Finance or the Bloomberg API. These will be the inputs and the scope will be stocks under the S&P 500.

Solution Statement

This model aims to predict the adjusted close price for a future date of a certain stock with the LSTM model. LSTM models are useful when dealing with time-series data.

Benchmark model

The benchmarking model in this case would be the linear regression model, a model that has also been used for time series data which will also have the same inputs as LSTM.

Evaluation metrics

To evaluate the model, r-square and RMSE (root-mean-squared-error) will be used as this is a regression problem. R-square measures the proportion of variance for a dependent variable that is contributed by an independent variable while RMSE measures the error of a model in terms of how far the predicted values are away from the actual values.

Project Outline

First step would be to set up the notebooks and incorporate the required libraries.

Next would be to prepare and process the dataset which in this case will be the S&P 500 stocks. Data will be normalised and split into training/testing data sets.

The benchmark model – Linear Regression will then be developed which will be set up with Scikit-learn. The results will be plotted using matplotlib and the model will be evaluated with the r2 and RMSE.

Next, the LSTM model will be developed using Keras. Similarly, the results will be plotted using matplotlib and the model will be evaluated with the r2 and RMSE.

From there, we will compare and assess the effectiveness of the LSTM model in predicting stock prices.