Ryan Senoune

260989415

Deliverable 2

Project

As my final project, I will attempt to predict the price of stocks. My model will use past data of the S&P500 Index fund to attempt to predict future prices of its stocks.

Data Preprocessing

Kaggle dataset: https://www.kaggle.com/camnugent/sandp500

The dataset tracks the price of 505 stocks inside the S&P500 for 5 years. For now, I have decided to only use the closing price and volume. I will add more features when I have a better understanding of my model. For certain stocks, some information seems to be missing on specific days. Since there are only 11 of those days, I simply dropped all rows containing Nan values. In addition, since the closing prices and volumes are on a very different scale, I had to normalize the data.

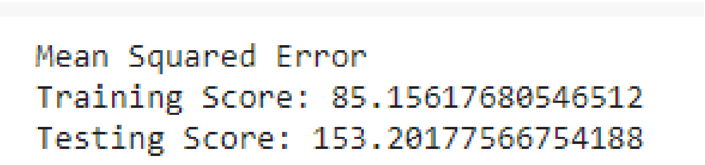
I created a feature vector (X) and label vector (y). The feature vector consisted of three sequential closing prices and volumes. The target vector associated with it was the closing price of the next day. So my model is predicting the closing price of a day using the information of the three previous days.

Finally, I decided to split my data 80% for training and 20% for testing.

Machine Learning Model & Preliminary results

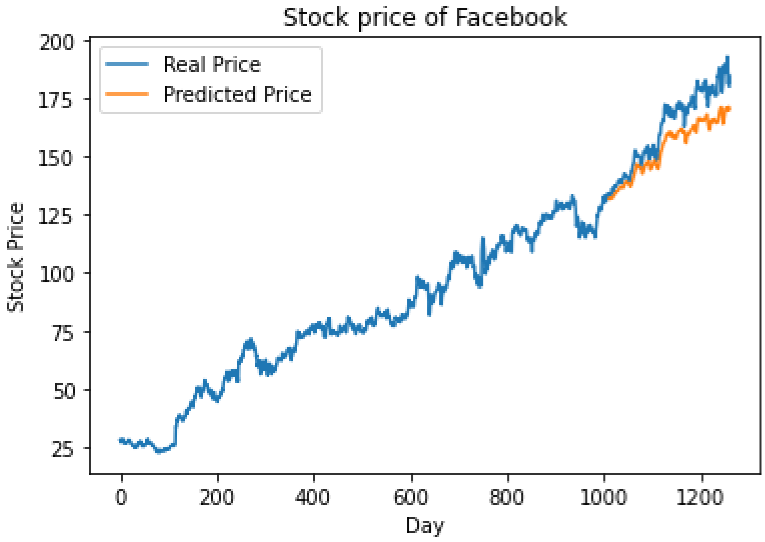
Upon further research, I have decided to opt for the LSTM model using the Keras library. Since I am not completely familiar with this model, I have decided to test it using only one stock. I have used only the default parameters for now. I added one LSTM layer and one Dense layer in my model.

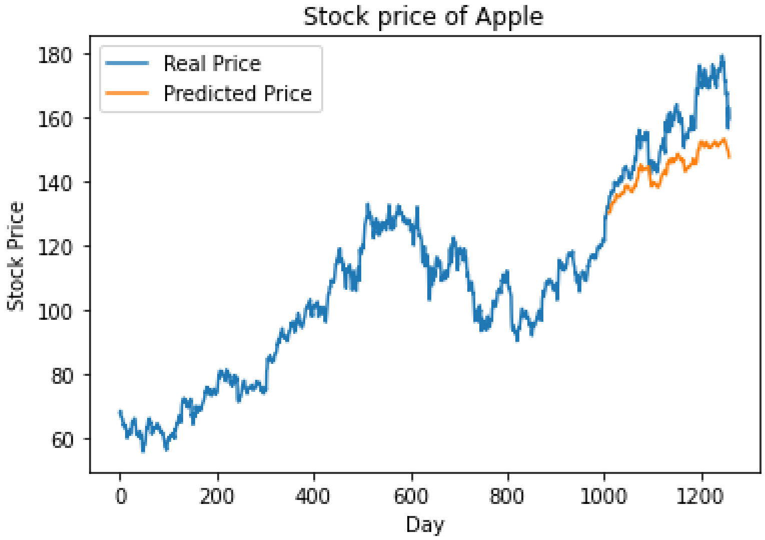
Trying to predict the stock prices for Facebook, I got the following Mean Squared Error



```
Mean Squared Error
Training Score: 85.15617680546512
Testing Score: 153.20177566754188
```

Here are two graphs of the predicted price of Facebook and Apple





*Note that the price of the apple stock is lower today because of stock splits

As we can see, my predictions are not so far from the real price. However, they do seem to always be lower. I would like to better understand how LSTMs work. That will allow me to fine tune the model and experiment with the layers. I have also only used two features (closing price and volume), so adding more features could help the accuracy.