Stock Closing Price Prediction Using Artificial Neural Networks

1. INTRODUCTION

Stock market basically refers to the collection of markets and exchanges, where shares of publicly owned companies are regularly bought, sold and issued under a formal setup. Prediction of returns associated with stock market is still a challenging task as the market dealing in such financial activities are often dynamic, volatile and non-linear in nature. Stock market predictions are affected by multiple factors including but not limited to political scenarios, international economy, the reports and performances of the company, anticipated takeover or merger, employee layoffs, future estimated earnings, etc. Hence, in order to maximise the company's profit and lowering down the losses, stock market predictions are necessary. This could be analysed through the trend in the previous periods.

The prediction of stock market direction is crucial and assists the short-term investors as a recommendation system and as a financial distress early warning system for the long-term shareholders. Forecasting accuracy is fundamental in the selection of any forecasting methods. Selection of profitable stocks as a suitable investment is the most challenging task. The key factor for any investor is to earn the maximum return with minimal investment. Before an investor invests in any stock, he needs to be aware of the behavioral traits of any stock. Investing in good stock at inappropriate time can be fatal whereas investing in mediocre stock at the right time can reap benefits. Financial investors of today are facing this problem of trading as they do not understand which stocks, they should trade to reap maximum profits and/or return. Predicting long term price of the stocks is comparatively easier than predicting the day-to-day stock prices due to their highly volatile nature which is based on world events. The solution to this problem demands the use of tools and technologies related to the field of data mining, pattern recognition, machine learning and data prediction. The algorithm will predict the stock prices at the closing and for the next trading day.

In our project paper, the focus was to use artificial intelligence and the computational capabilities to efficiently predict stock market prices. The Long Short-Term Memory (LSTM), XGBoost Regressor and Random Forest Regressor were the machine learning algorithms which was implemented to analyse the hidden patterns and complex relations between the variables affecting the stock market in 10 companies, namely, Starbucks, Microsoft, Cisco, Qualcomm, Facebook, Amazon, Tesla, Apple, Zynga, AMD. It is already proved that the use of machine learning algorithms in this field can raise the efficiency by

60-86 percent, so the purpose of this analysis is to contribute more towards accurate prediction pertaining to the stock market. We have also factored in the time series in our models to predict the closing price of the stock as it gives us better insights compared to normal regression models which does not consider the date feature which is an important aspect of the data, into the analysis.

2. BENCHMARK

In a study on "Stock Closing Price Prediction using Machine Learning Techniques", stock market predictions are made using two important approaches - technical analysis and qualitative analysis. For the technical analysis, historical price associated with stocks were used. Those include price both closing and opening, traded volume, adjacent close values etc. The qualitative analysis can be executed on the basis of external factors. These external factors are the profile of the company, market situation, factors lined to politics and economics, news articles, etc. This paper has used the data from 5 companies over a period of 10 years. Using this, Artificial Neural Network (ANN) and Random Forest (RF) techniques were used in the paper to analyse trend and make predictions. To evaluate the performance of the models, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Bias Error (MBE) were computed, which further showed that ANN has given better results.

This paper can be served as a potential guide in predicting stock closing price prediction, where we will be implementing more machine learning algorithms along with the Random Forest model already used and will then evaluate them by using Root Mean Square Error (RMSE) and Mean Square Error (MSE) methods. These applied machine learning algorithms have already been covered in the course. The focus will be to generate least error. When searched online, we found that RMSE rate was between 0 to 2 for the best models.

3. METHODS

We have used the dataset available from Yahoo Finance for the 10 companies mentioned in our 'Introduction'. The dataset consists of features like opening stock prices, High, Low and closing stock prices of that day, volume of stocks of companies in the market, etc., which will be used to predict the closing price of the stock for the day. This dataset consists of details from the beginning date of the company, but we will be using the data from 1st January 2017 to 31st of October 2020.

To gain more insight about the data, we have used exploratory data analysis (EDA) and have observed the descriptive statistics for all the features of the 10 companies that we have selected. We have analysed different trends using the line charts for the average stock price, average price fluctuation per day and per day volume sale for stocks. In addition, we have also compared the closing price trends for all the mentioned companies using a single graph. These results are elaborated in the further section.

Finally, before using the data for our models we have split the data into 80:20 train-test ratio. Also, we have applied MinMaxScaler(feature_range = (0,1) to range the data between 0 and 1 to avoid biasedness.

After making appropriate changes to the data we applied the following machine learning algorithms:

1. Long Short-Term Memory (LSTM) which is a type of Recurrent Neural Network (RNN) which is used to predict data using Time Series by using feedback connections, making it perform better than other Neural Network techniques. So as our data is dependent much on time variability, it was a good option to consider a model using series of data rather than single data points. Our neural network has 9 layers with first 4 Layers of LSTM whereas following four layers are Dense layers with various combination of hyperparameters. The last layer is basically the activation layer which gives us the final output.

2. Random Forest Regression- Random Forest Regression is a machine learning technique based on ensemble method. It is made up of many small decision trees known as estimators, where each of these produce their own predictions. These predictions are combined in random forest to produce more accurate results. So, this model will basically use our data to make several small decision trees taking into consideration different parameters of our data and finally combining the outputs to give the regression prediction.

3. Extreme Gradient Boost (XGBoost)- Extreme Gradient Boosting method, is an ensembling method which combines multiple weak performing decision trees to make a better prediction. It does that by optimizing the cost function.

We chose Random Forest and XGBoost because these are the best performing basic regression models which gave us better results than other regression models.

After applying all the models, we used Root Mean Squared Error (RMSE) for all the models to check their efficiency. This technique is a widely used tool to measure the performance of the regression models. Our aim is to keep the RMSE value close to 0.

$$RMSE = \sqrt{\frac{\sum_1^n (Original\ value - predicted\ value)^2}{n}}$$, where n is number of observations

4. RESULTS

The findings from the exploratory data analysis is presented, which have provided us the insights about the main characteristics and other relevant information for the stock prices.

Fig. 3.1: Average Stock Price of Starbucks for the period 01-01-2017 to 10-31-2020

In Fig. 3.1 the average stock price of Starbucks for each day from 01-01-2017 to 10-31-2020 have been computed to see the price trend of the stocks. The horizontal axis represents the period and price at the vertical axis. This chart shows the Moving Average (MA) of the average closing price of the stock over the 3 years period. From this the trend of closing price of stocks can be determined. This helps to forecast the long-term trends of the stock. For each of the companies, we have followed this EDA method to gain insight on the average closing price of the stocks.

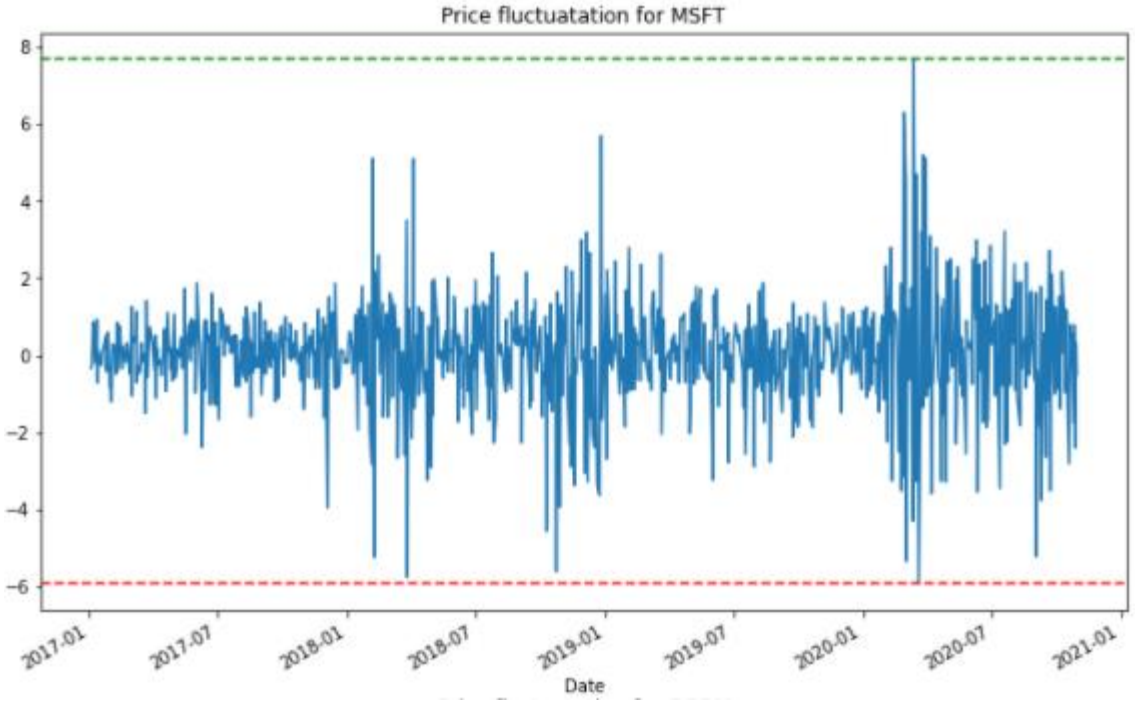


Fig. 3.2: Daily stocks closing price fluctuation of Microsoft for the period 01-01-2017 to 10-31-2020

In fig. 3.2 we have computed the daily closing price fluctuation for the stocks of Microsoft. From this we get a glimpse that how often the stock closing price had price increase and decreased. This indicates the stability and sensitivity of the stock. When the price increase it indicates that the demand for the stocks of Microsoft on a particular day or period was high since a greater number of people were purchasing the stocks in that particular period and vice versa. There can be several other reasons which can trigger the fluctuation of the stock prices, but we have limited our scope of analysis for the benefit of training and testing our models.

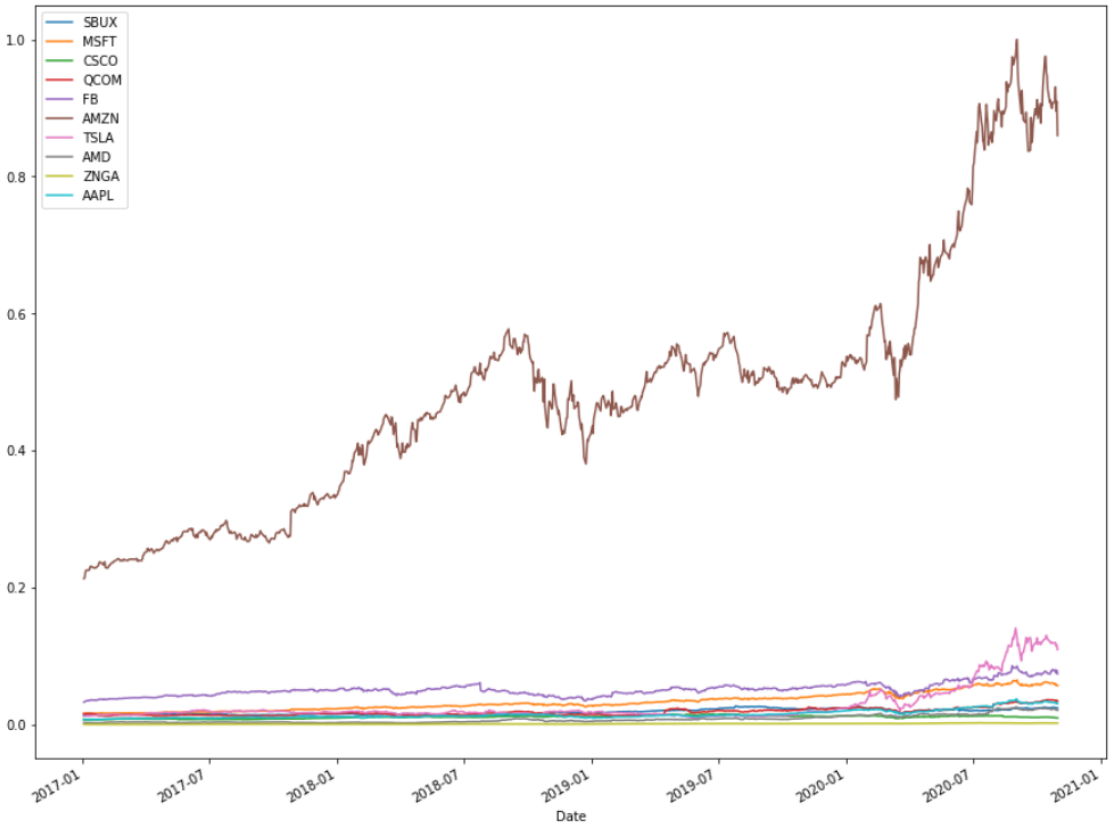


Fig. 3.3: Comparison of stock closing price trend of all 10 companies for the period 01-01-2017 to 10-31-2020

In this chart we have compiled together the stock closing price trend of all the companies so as to have a comprehensive and comparative view. From this we can understand the trend of stock prices of all companies in one picture. So, we observe that trend of all the companies were similar. This indicates that we can use the same model for the prediction of closing prices for all the companies.

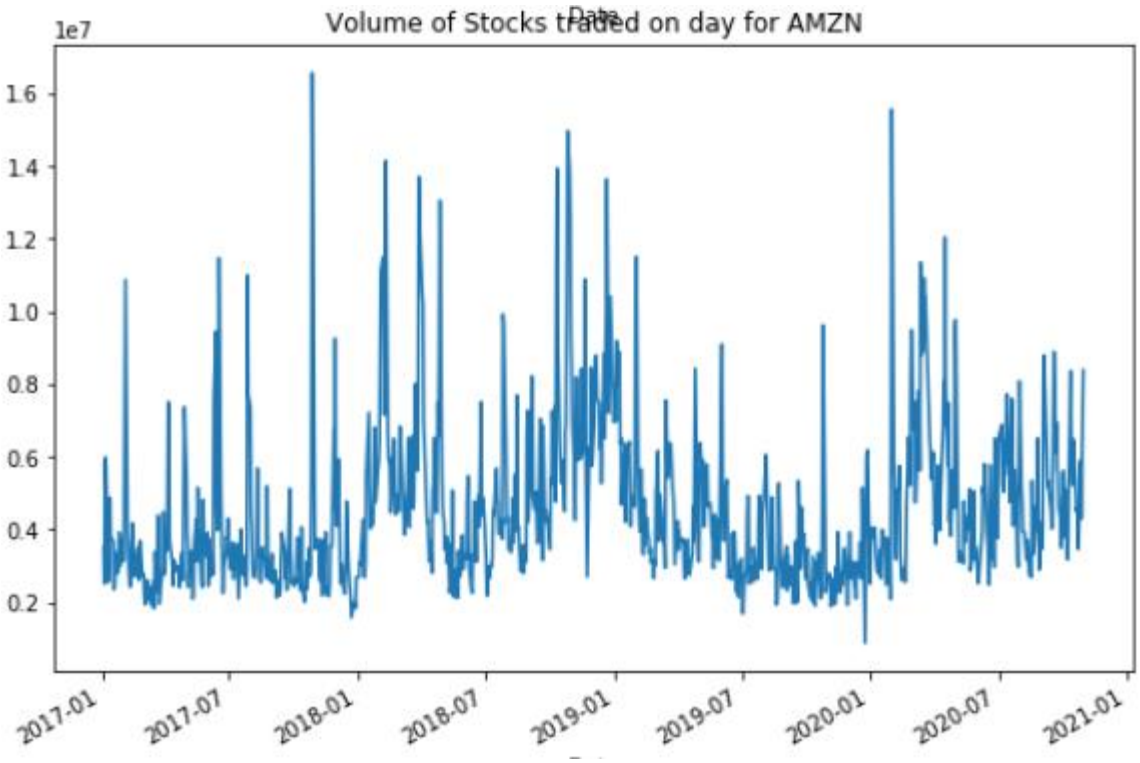


Fig. 3.4: Volume of stocks of Amazon traded per day between period 01-01-2017 to 10-31-2020

Fig. 3.4 shows the quantity of stocks that has been traded in the given period of Amazon. This indicates the liquidity position of the stocks and how conveniently the price of the stock can change from its current price. A company with liquidity in its stock prices is more in demand over others.

The EDA has shown us the dynamicity of stock market and thereby, has highlighted the need for analysis in this field.

In our analysis, we have created a function in all the applied machine learning algorithm which will provide us the RMSE value and the graph for the predicted values of the target variable, which is closing stock price. The reference for the created functions for each technique is shown below:

For LSTM:

```
#Prediction of any company from the list
companies = ['SBUX','MSFT', 'CSCO','QCOM', 'FB', 'AMZN', 'TSLA','AMD','ZNGA', 'AAPL']

company_name = 'CSCO'
rmse_lstm=predict_stock(company_name)
```

For XGBoost:

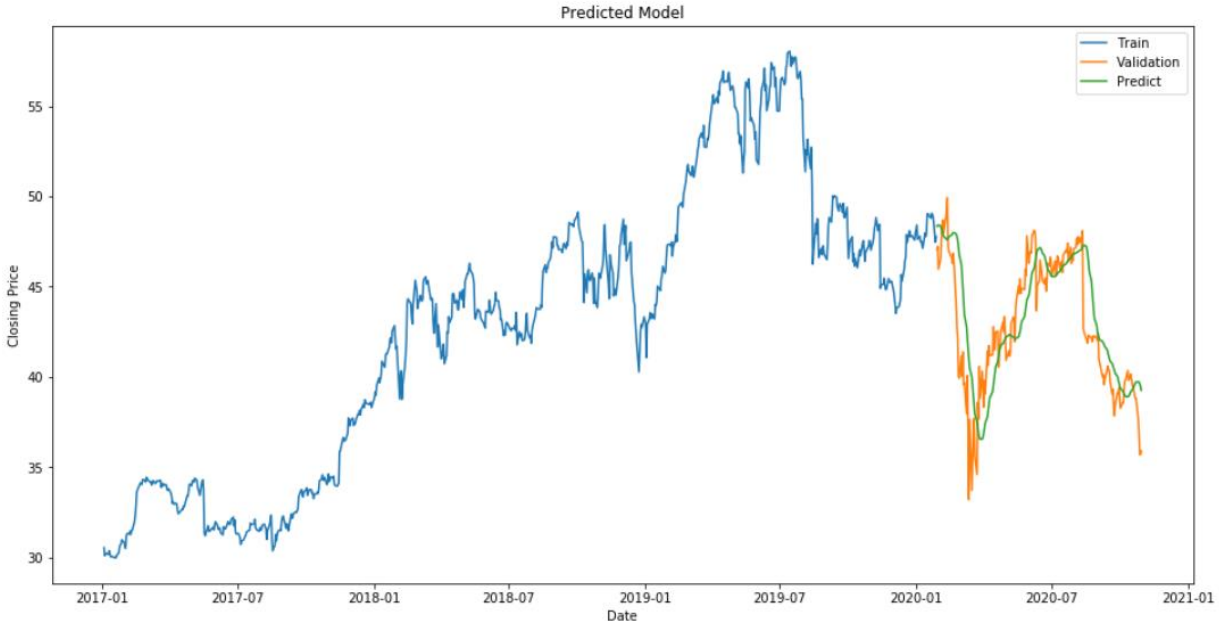
```
#Calling the function to train and predict CSCO stock price

x_train, y_train, x_test, y_test, company = prep_data('CSCO', scale)
y_pred, rmse_xgb = xgb_model(x_train, y_train, x_test, y_test, company)
plot_result(company, y_pred)
```

For Random Forest:

```
# Calling all the function to predict the CSCO stock using Random Forest
x_train, y_train, x_test, y_test, company = prep_data('CSCO', scale)
y_pred, rmse_rf = rfr_model(x_train, y_train, x_test, y_test, company)
plot_result(company, y_pred)
```

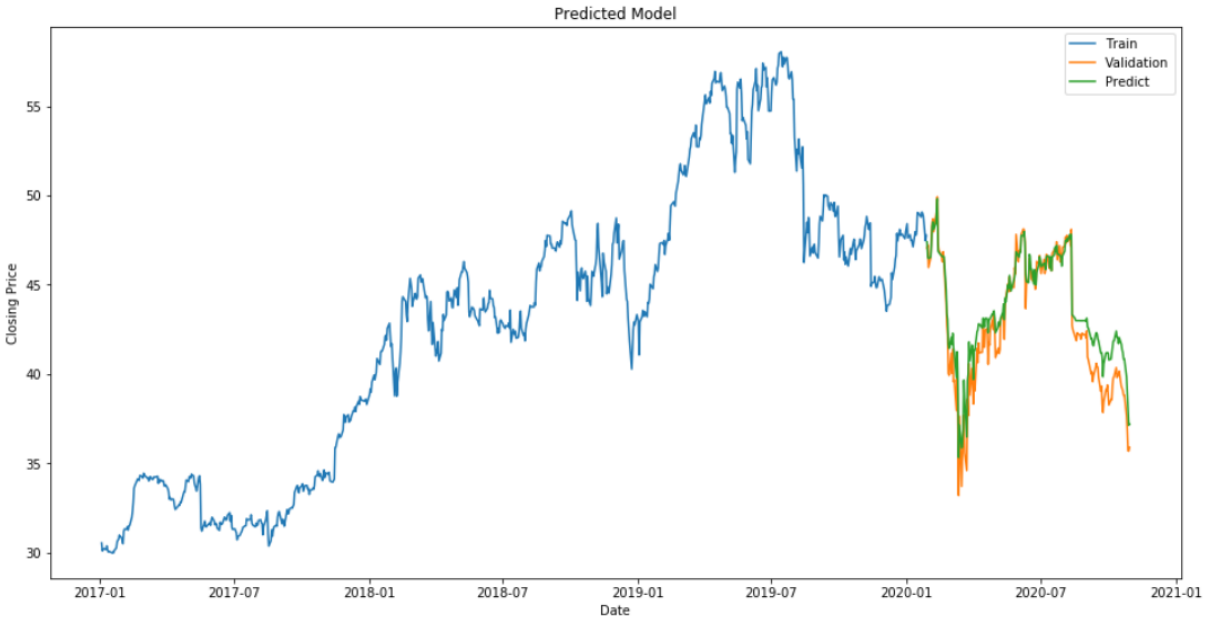
When the function for LSTM was used, specifically for the company CISCO, we obtained the following results:

```
The Root Mean Squared Error: Close    2.311429
dtype: float64
```



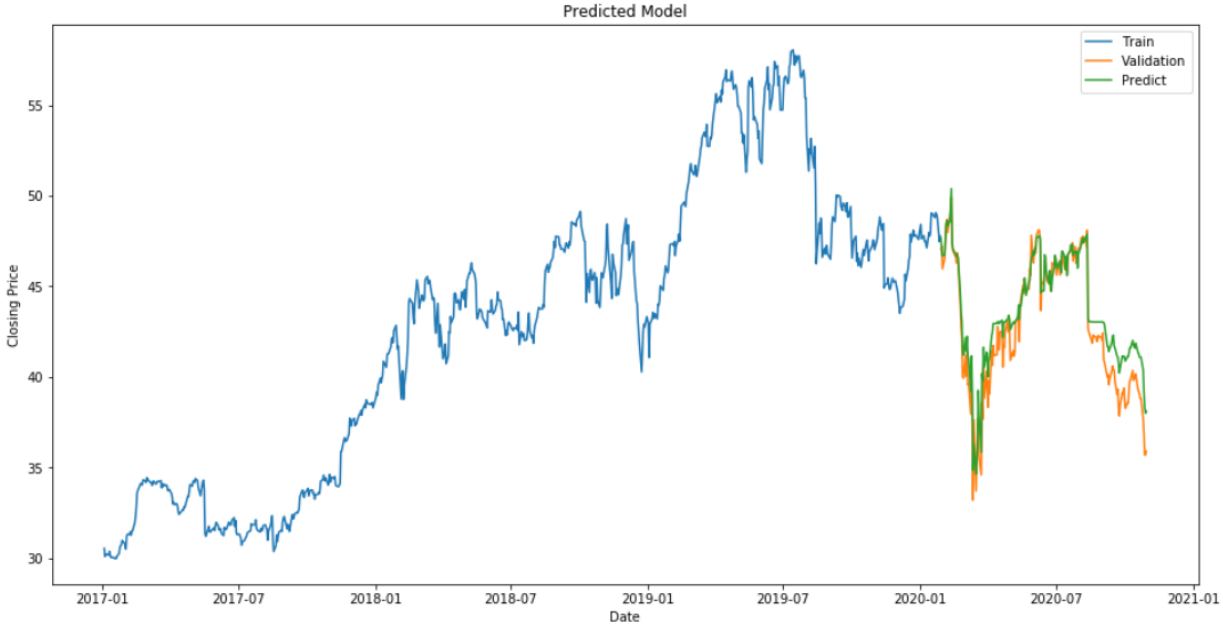
When the function for XGBoost was used, specifically for the company CISCO, we obtained the following results:

```
The Root Mean Squared error : 0.045007782092448365
```



When the function for Random Forest was used, specifically for the company CISCO, we obtained the following results:

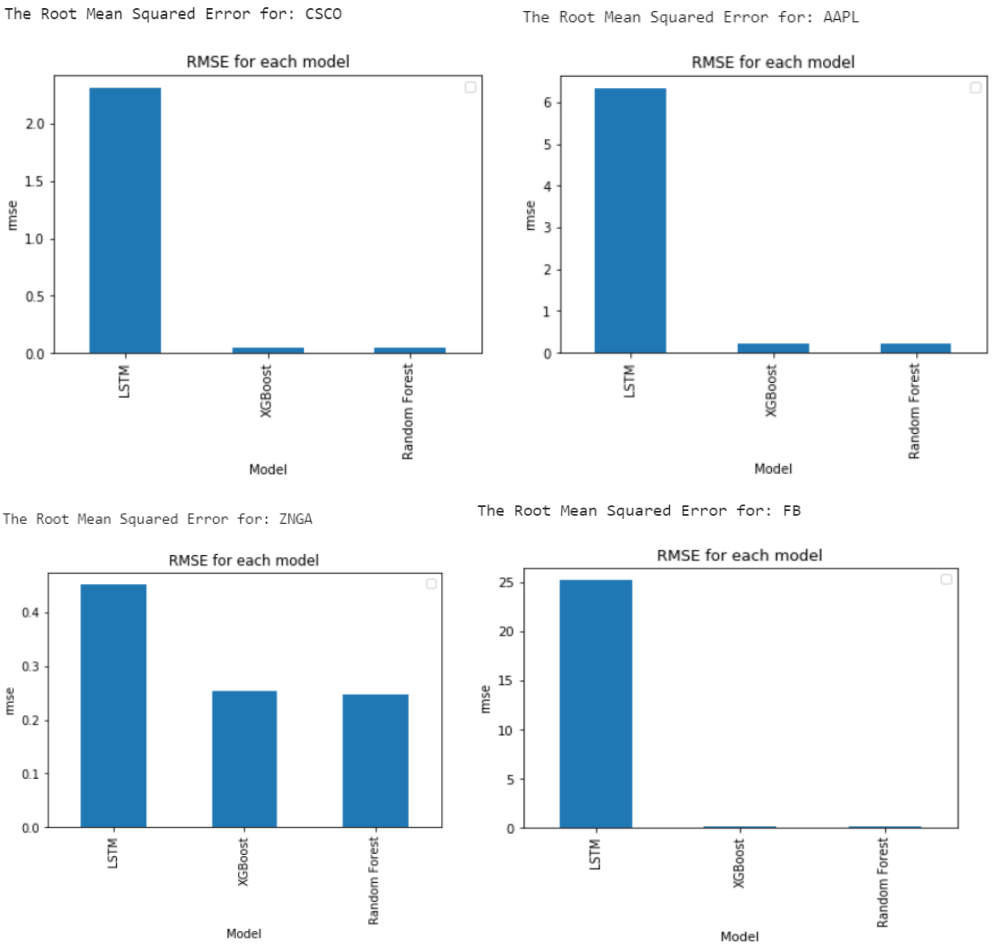
```
The Root Mean Squared error : 0.044002565799163784
```



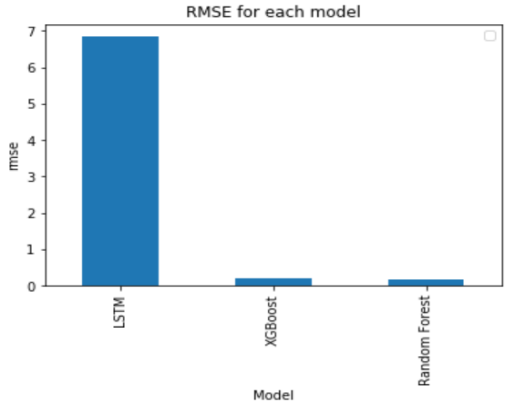
All the above results for different machine learning algorithms show us that the green line almost overlaps the orange line. This means that the predicted values are almost equal to the values in the original dataset. This result can also be confirmed by the RMSE values, which are displayed above the respected graphs. In the graph, it can be observed that the 80% data belongs to training dataset and 20% belongs to testing dataset, for which we have displayed both the original as well as the predicted values to clearly analyse the difference between them.

5. DISCUSSION

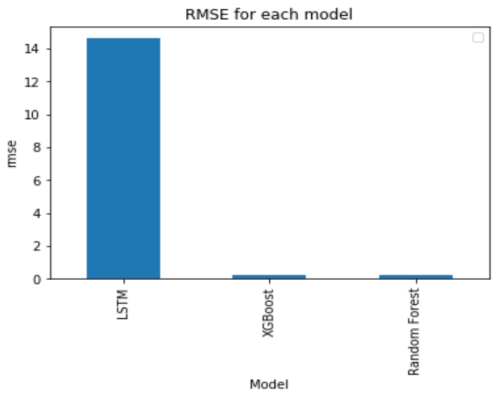
We have shown the results i.e. the graph and RMSE value for all the companies so as to make accurate inferences about the performance of models on them.



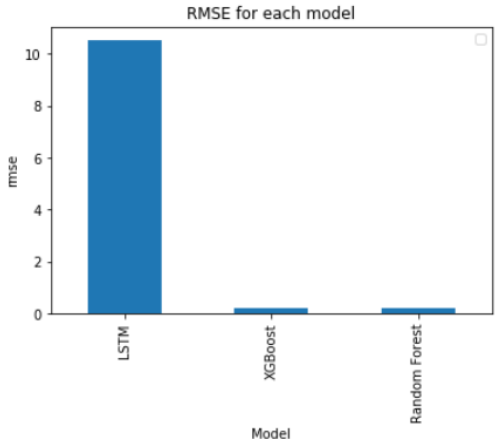
The Root Mean Squared Error for: QCOM



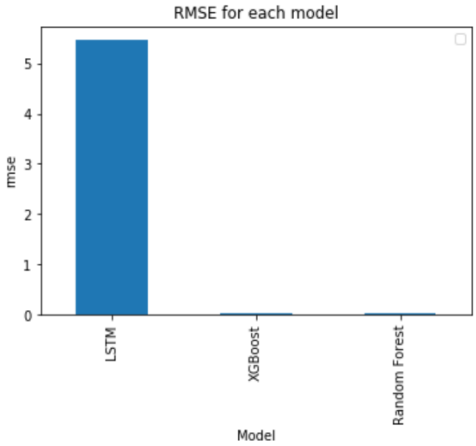
The Root Mean Squared Error for: MSFT



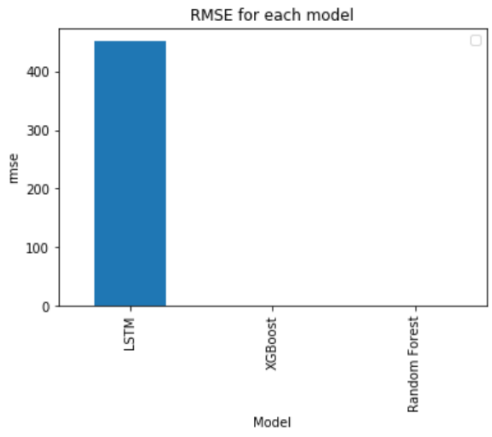
The Root Mean Squared Error for: AMD



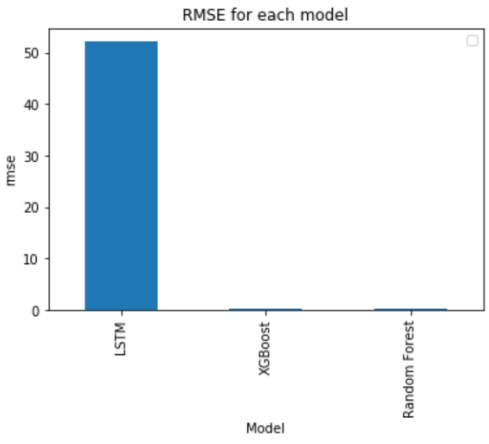
The Root Mean Squared Error for: SBUX



The Root Mean Squared Error for: AMZN



The Root Mean Squared Error for: TSLA



The results obtained above shows us that the predictions made for the stock closing price were very accurate for the companies in XGBoost and Random Forest as these models were not based on time series. In LSTM algorithm, model training was done taking into consideration the time series and it performed well for the companies having similar price

trends. The RMSE values for XGBoost and Random Forest for all the companies were almost 0, which indicates a good fit.

For all the companies except Amazon, Facebook and Tesla, the RMSE values for LSTM model is less than 15. It shows that the model predicts the target variable effectively.

In real life scenarios, these models can be directly used to predict the closing stock prices of any company provided the similar features.

The LSTM gave lower result in comparison to basic regression models because we have not considered a large time series dataset due to the system limitations.

6. CONCLUSION

The stock market returns are highly dynamic in nature, which creates difficulty in making accurate decisions by the investors but at the same time is a crucial aspect as it serves as a recommendation system and as a financial distress early warning system for the long-term shareholders. To overcome these concerns, we have created the models using LSTM, XGBoost and Random Forest algorithms which can predict the closing stock price accurately. Also, the timeframe which was used in the analysis in the testing part contained the pandemic period, which had high and unusual fluctuations in the market. Even after these challenges, the model performed well on the considered data, which shows the efficiency of the built models.

For future work, neural network layers for LSTM can be manipulated and implemented along with more time series data, which would help in achieving better performance of the model. Also, models can be built using more machine learning algorithms including Moving Average Convergence / Divergence (MACD) and Autoregressive Integrated Moving Average (ARIMA) and the model performance can be analysed based on these.

7. REFERENCES

- Research Paper: https://www.sciencedirect.com/science/article/pii/S1877050920307924
- Dataset : https://drive.google.com/file/d/0B2lCmt16L_r3UFJ1a0dmUmFLUGs/view
- https://www.investopedia.com/terms/s/stockmarket.asp#:~:text=The%20stock%20market%20refers%20to,publicly%2Dheld%20companies%20take%20place.&text=There%20can%20be%20multiple%20stock,and%20other%20forms%20of%20securities.

- https://www.getsmarteraboutmoney.ca/invest/investment-products/stocks/factors-that-can-affect-stock-prices/
- https://www.indiainfoline.com/investment-guide/why-do-stock-prices-fluctuate
- https://en.wikipedia.org/wiki/Moving_average
- https://www.thebalance.com/day-trading-based-on-buying-volume-and-selling-volume-1031209