

Topic: Stock Price Prediction using Machine Learning

Name: Dhruvkumar Patel

Abstract

With big data at our fingertips, Deep Learning for predicting stock market prices and trends has become more popular than ever before. Artificial Intelligence has introduced numerous algorithms and with enhanced computational abilities and programmable methods of prediction have proven more effective at predicting stock prices. In this Study, Long Short Term Memory (LSTM) technique has been employed to predict the next few days price of Apple Stock over a variety of time periods. The models were evaluated by using MAE, MSE, RMSE which is a standard strategic indicator. The low values of these indicators show that the models are efficient in predicting stock closing price.

1. Introduction

Predicting stock prices is an uncertain task which is modelled using machine learning to predict the return on stocks. Stock market trading is considered to be extremely dynamic and complex. With Deep Learning, more and more accurate programs have been developed to help determine when it's best to buy or sell a stock, and many investment firms and listed companies rely heavily on algorithmic trading to make financial market decisions. Assuming that all news that can impact the performance of a stock is excluded, pricing shifts are in large part affected by the conclusions that the algorithms draw from price movements. Taking into account the increasing use of algorithms to guide financial market interactions, it seems easier to predict the closing price trend for a stock.

There are two main study in the Financial Markets, Technical Analysis and Fundamental Analysis. Based on the underlying factors that affect a company's business and its prospects, an analysis of a stock's fundamentals attempts to calculate its value is in Fundamental Analysis. On the other hand, Technical Analysis examines the price movement of a stock and predicts its future price movements. In this study, Technical Analysis is mainly focused in this study as we only have historical data of Apple Stock.

In this project, The Apple stock data is collected from yahoo finance website and examining data and their charts. The algorithm used is Long Short Term Memory(LSTM) method to predict the price of stock which built on various layer. The data is split into training and testing dataset. Data is visualize to see the movement of stock price and also to see the predicted prices on graph. The output of the algorithm and stock value is analyzed and it will be used to predict the stocks for futures. I have predicted the prices of next 5 days (next week) using last 1 year data and next 20 days using last 3 year data and for last 5 year data also. In addition I have also predicted next 2 days price using 1 months data. The minimum data that can be predicted is of 1 day only. We cannot predict data of next 1,5,10,15,60 mins, etc as there are no data available of it. To conclude results, I have used various indicators such as MAE, MSE, RMSE error.

2.1 Data

The Apple stock data is collected from yahoo finance website where I have downloaded specific one month (1/31/2021) to 2/11/2021), one year (3/15/21 to 3/11/22), last 3 year (1/1/19 to 3/11/22) and last 5 year (5/3/2017 to 4/28/22) data.

Here, We have used data of Apple stock of time period of Last 1 year(3/15/21 to 3/11/22) which is total 253 days.

	Date	Open	High	Low	Close	Adj Close	Volume
0	2021-03-15	121.410004	124.000000	120.419998	123.989998	123.257416	92403800
1	2021-03-16	125.699997	127.220001	124.720001	125.570000	124.828079	115227900
2	2021-03-17	124.050003	125.860001	122.339996	124.760002	124.022865	111932600
3	2021-03-18	122.879997	123.180000	120.320000	120.529999	119.817863	121229700
4	2021-03-19	119.900002	121.430000	119.680000	119.989998	119.281052	185549500

- Open: Opening stock price of the day
- Close: Closing stock price of the day
- High: Highest stock price of the data
- Low: Lowest stock price of the day
- Volume: Total transaction of stock

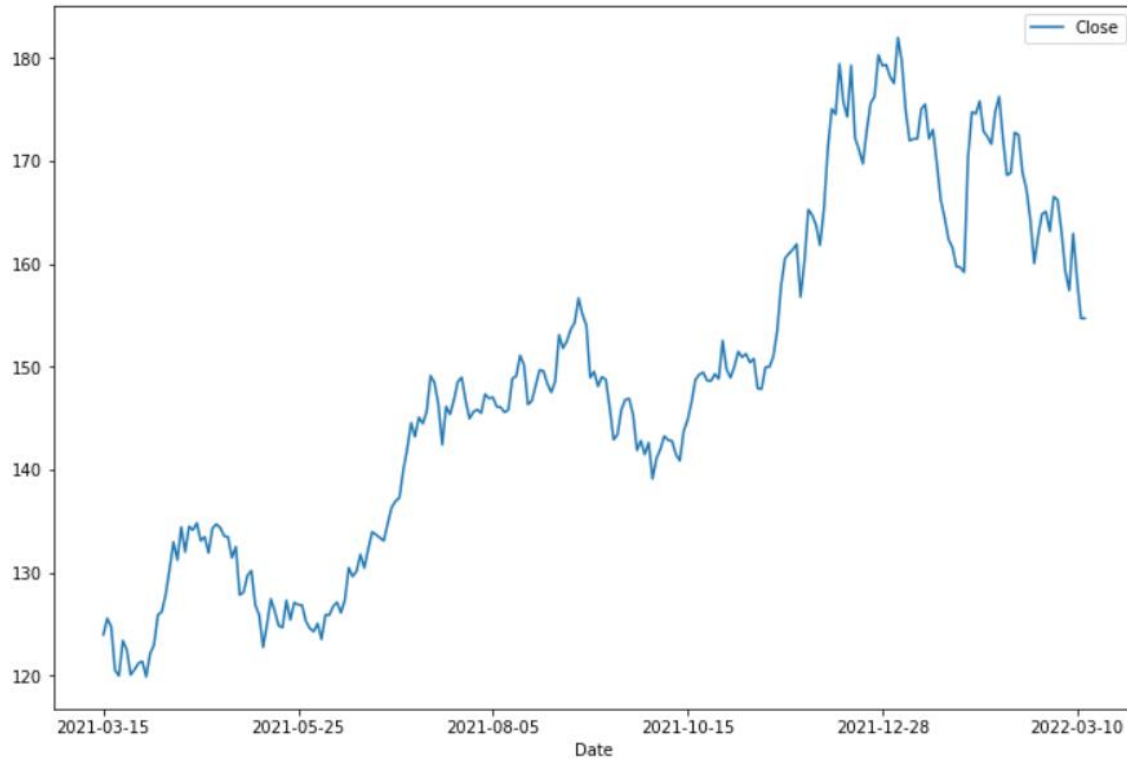
The data is split into 70% training data and 30% test data and for data of last 1 month data is split into 60% training data and 40% test data because of fewer number of data.

All Data with their Train and Test Data

	Data	Train Data	Test Data	Total
	Last 1 year	177	76	253
	Last 3 year	564	242	806
	Last 5 year	879	378	1257
	Last 1 month	17	12	29

2.2 Data Visualization

The data is visualized to display the movement of Close Price and Date



2.3 Normalization of Data

Normalization is a rescaling of the data from the original range so that all values are within the range of 0 and 1. It is required to know that we are able to accurately estimate the minimum and maximum observable values. I have use “Minmax Scalar” scales to normalize the data. Transform features by scaling each feature to a given range.

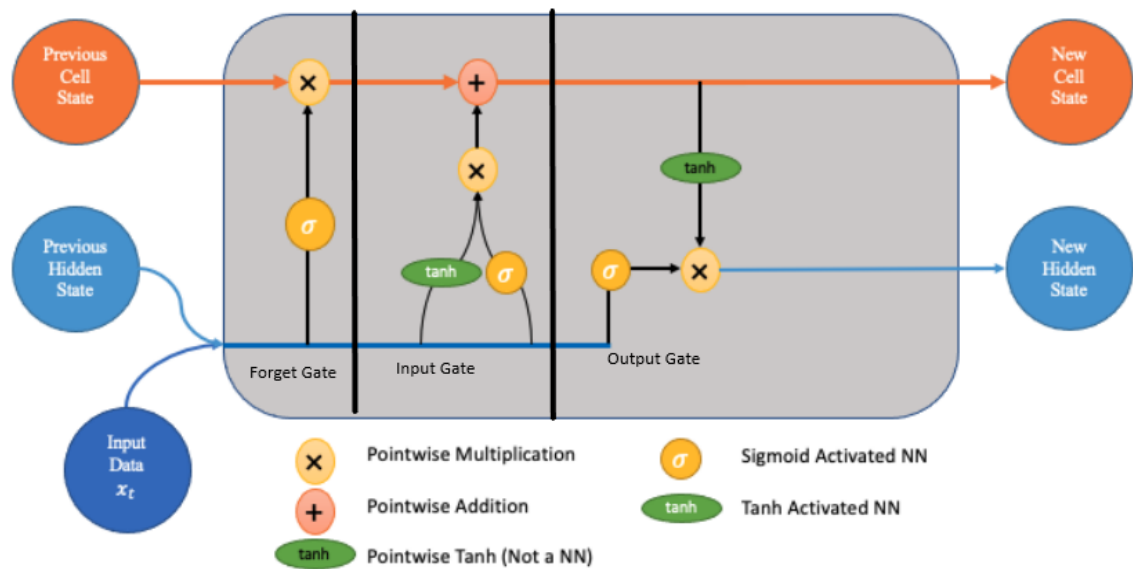
3. Methods

- **LSTM**

LSTM stands for Long Short-Term Memory networks. It is a special type of recurrent neural network (RNN) that is capable of learning long-term dependencies, which is particularly useful in prediction problems for sequences, and which exhibits outstanding performance on a wide range of problems. With its feedback connections, LSTM can process the entire sequence of data, with the exception of single data points such as images. It is capable of capturing the patterns of both long term such as a yearly pattern and short term such as weekly or daily patterns. They are extremely powerful time-series models. They can predict an arbitrary number of steps into the future.

With this type of model, it is possible to deal with short-term memory problems, disappearing and exploding gradients, as well as shrinking gradients as time passes. The cell remembers values over arbitrary time intervals, and the three gates regulate the flow of information into and out of the cell. The LSTM has gates that are able to

regulate what to keep or forget/throw away. By doing so, it learns to use relevant information to make predictions. The LSTM is able to store information from the past which helps especially predict stock price fluctuations based on past prices. The market changes constantly, so making timely accurate decisions and predictions is extremely crucial for accurate results.



LSTM Diagram

There are three gates:

1. Forget Gate:

It chooses whether the information coming from the previous timestamp is to be remembered or is irrelevant and can be forgotten.

2. Input Gate:

In this, the cell tries to learn new information from the input to this cell.

3. Output Gate:

In this, the cell passes the updated information from the current timestamp to the next timestamp.

Advantages:

One of the key advantages is that they address the vanishing gradient problem that makes network training difficult for a long sequence of words or integers. Using LSTM, time series forecasting models can predict future values based on previous, sequential data and it is proven to be extremely effective. The reason they work so well is that LSTM can store past important information and forget the information that is not. The different gates inside LSTM boost its capability for capturing non-linear relationships for

forecasting. Using LSTM, we can transform the data from a one-dimensional array to a two-dimensional by lagging the time series for a number of steps and then start train-test-validation. It is a method for having lower forecasting errors with LSTM is using STL decomposition. This way we can decompose the time series in trend, seasonal and residual and by performing forecast on each of those components and summing your results you can make your predictions much more robust.

4. Results:

1) For Last 1 Year(3/15/21 to 3/11/22)

	Adj Close	Predicted
248	157.440002	157.120544
249	162.949997	157.776886
250	158.520004	157.623230
251	154.729996	157.275970
252	154.729996	156.615372

Mean Absolute error: 4.112533629550783

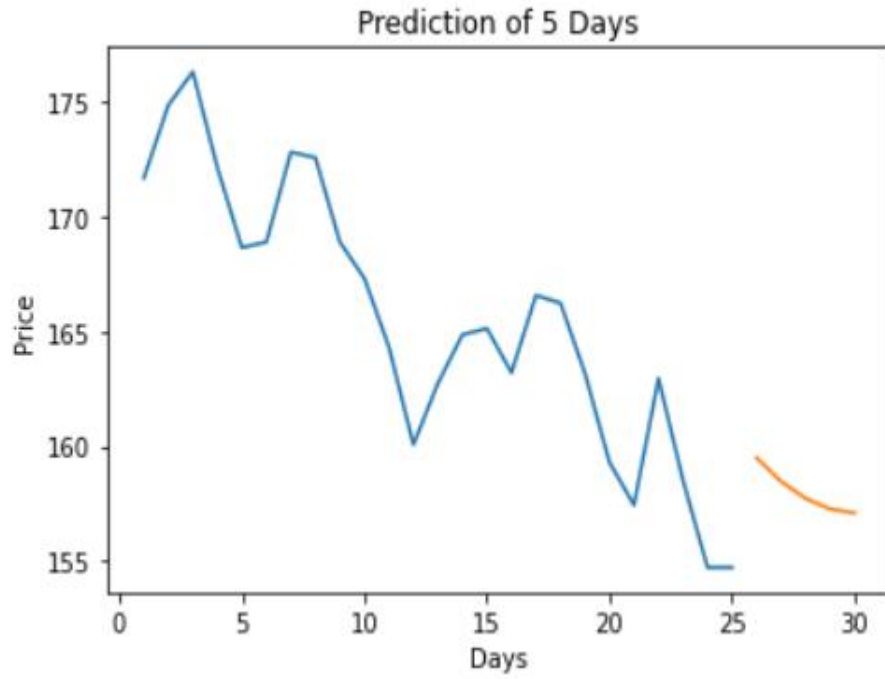
Mean Squared Error: 27.173769304870703

Root Mean Squared Error: 5.212846564485733

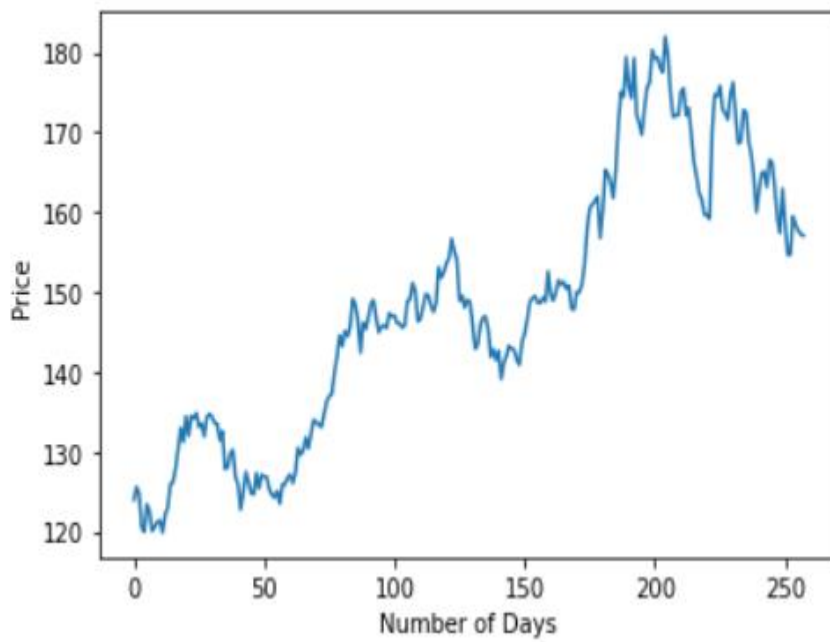
Prediction of Close Price Next 5 Days

	Date	Predicted Price
0	2022-03-14	159.498654
1	2022-03-15	158.470614
2	2022-03-16	157.729499
3	2022-03-17	157.271741
4	2022-03-18	157.093310

Visualization



The orange line indicates the Predicted Price on graph.



2) For Last 3 years From 1/1/19 to 3/11/22

	Adj Close	Predicted
801	157.440002	155.994049
802	162.949997	156.050674
803	158.520004	155.419418
804	154.729996	154.913391
805	154.729996	154.224670

Mean Absolute error: 7.346088813137187

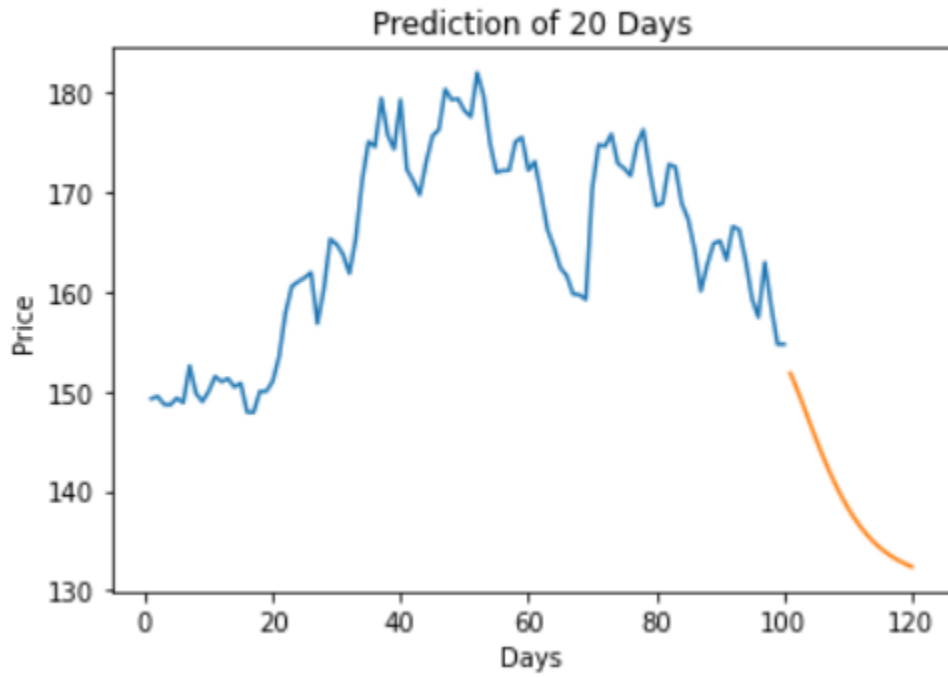
Mean Squared Error: 76.74109184285778

Root Mean Squared Error: 8.76019930383195

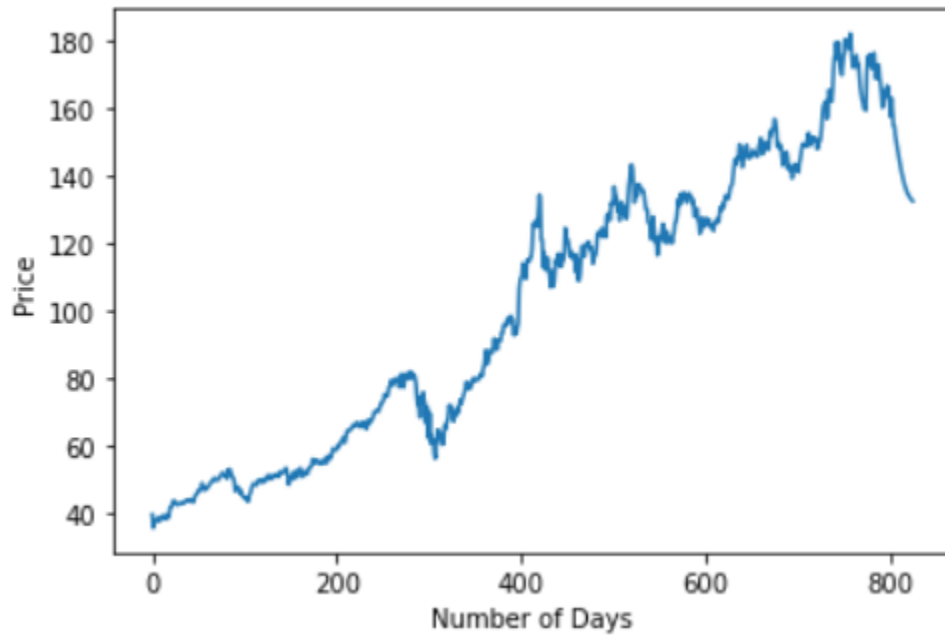
Prediction of Close Price Next 20 Days

Predicted Price		Predicted Price	
0	159.375742	10	137.274306
1	158.481055	11	136.372147
2	157.932943	12	135.582331
3	157.664230	13	134.896288
4	157.589624	14	134.304186
5	157.635273	15	133.795463
6	157.746884	16	133.359058
7	157.888054	17	132.983884
8	158.035431	18	132.659117
9	158.174367	19	132.374393

Visualization



The orange line indicates the Predicted Price on graph.



3) For last 1 month from 1/31/2021 to 2/11/2021

In this only 2 day stock price was predicted as there were only 28 days data in total.

	Adj Close	Predicted
22	166.229996	167.000732
23	163.169998	167.399155
24	159.300003	167.438843
25	157.440002	167.400085
26	162.949997	167.675507
27	158.520004	166.861740
28	154.729996	165.768250

Mean Absolute error: 5.053473360491073

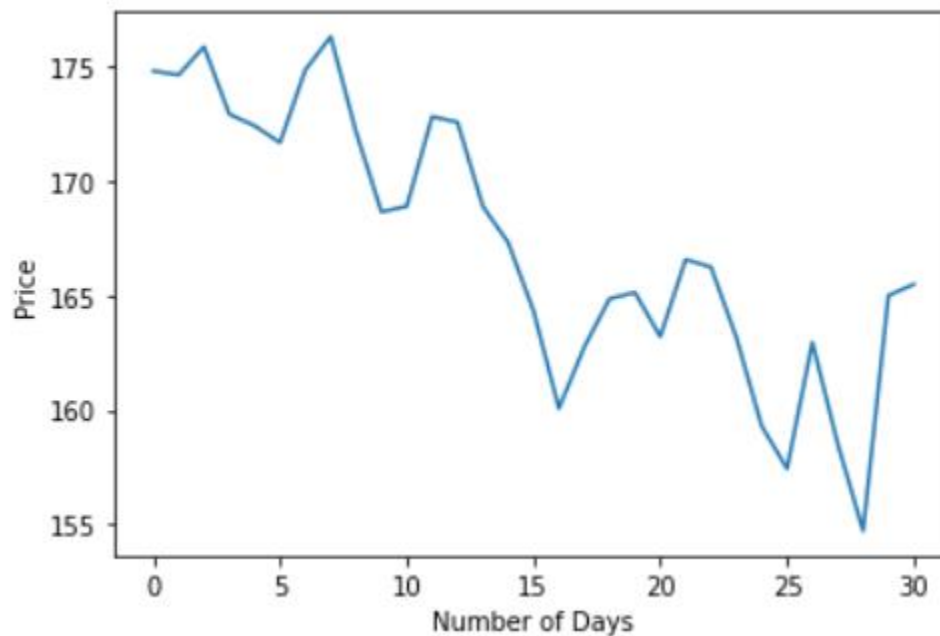
Mean Squared Error: 36.85712880249285

Root Mean Squared Error: 6.071007231299667

Prediction of Close Price Next 2 Days

	Date	Predicted Price
0	2022-03-14	164.990738
1	2022-03-15	165.477428

Visualization



4) For Last 5 year from 5/3/2017 to 4/28/22

	Adj Close	Predicted
801	157.440002	155.994049
802	162.949997	156.050674
803	158.520004	155.419418
804	154.729996	154.913391
805	154.729996	154.224670

Mean Absolute error: 4.264961843770777

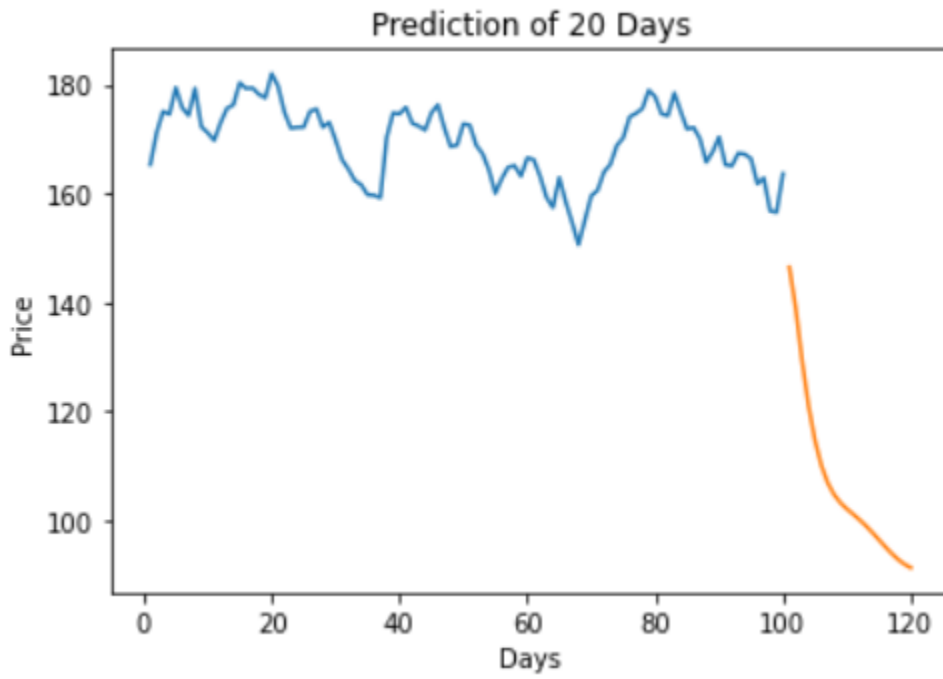
Mean Squared Error: 31.757880286101514

Root Mean Squared Error: 5.635413053725655

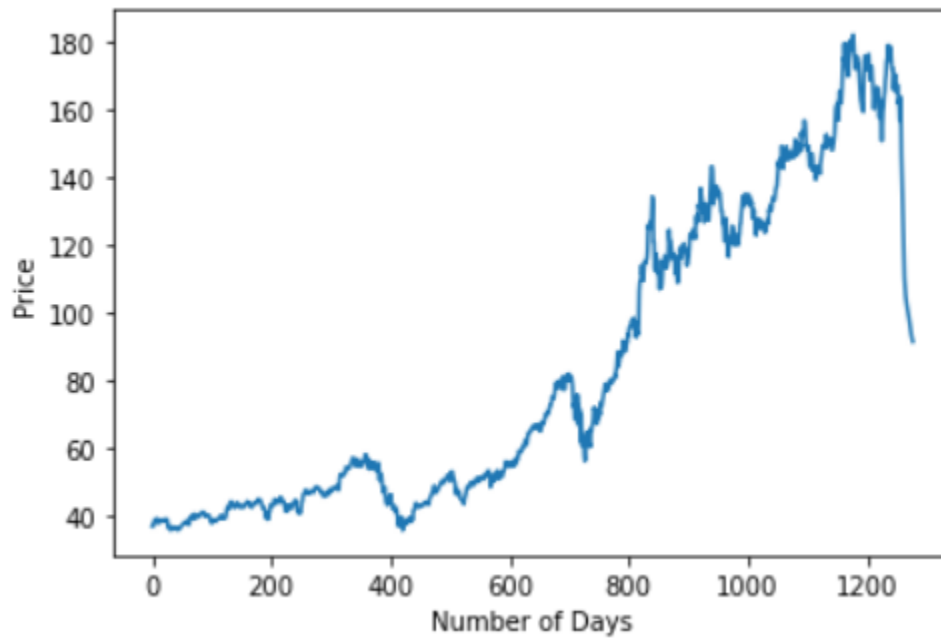
Prediction of Close Price Next 20 Days

Predicted Price		Predicted Price	
0	146.513276	10	101.143901
1	138.919009	11	100.135202
2	129.282194	12	99.024910
3	121.081902	13	97.800632
4	114.759242	14	96.501125
5	110.147492	15	95.195851
6	106.933259	16	93.964524
7	104.753988	17	92.879579
8	103.255593	18	91.994312
9	102.131298	19	91.337177

Visualization



The orange line indicates the Predicted Price on graph.



MAE, MSE, RMSE Error Comparison of all 4 Data Set

		MAE	MSE	RMSE
	Last 1 year	4.11	27.17	5.21
	Last 3 year	7.34	76.74	8.76
	Last 5 year	4.26	31.75	5.63
	Last 1 month	5.05	36.85	6.07

The Last 1 year data has lowest MAE, MSE, RMSE which we can say is more efficient in compare to other 3 dataset. Last 3 year data has highest MAE, MSE, RMSE. Here, We can also say that Last 3 year data might have more fluctuations in compared to other 3 methods which can be the reason it MAE, MSE, RMSE is slightly higher than the other dataset.

5. Conclusion

Stock market prices are highly unpredictable and volatile. This means that there are no consistent patterns in the data that allow you to model stock prices over time near-perfectly. We don't need the exact stock values of the future, but the stock price movements, i.e. if it is going to rise or fall in the near future. The LSTM Model was built with various layer which was used to predict future stock price. The model predict the data in smooth curve. The model predict data with MAE and RMSE error between 4 to 8 which is compared to other models is very small. I have predicted the prices of next 5 days (next week) using last 1 year data and next 20 days using last 3 year data and for last 5 year data also. In addition I have also predicted next 2 days price using 1 months data. We cannot predict stock price for next 1,5,10 minutes and we can predict of at least 1 day. By comparing the data of all 4 different time periods, the data used of 1 last year has provided the best MAE, MSE, RMSE error which also help to predict the future stock price accurately.

6. Reference

- Prashant Sharma, Machine Learning for Stock Market Prediction (<https://www.analyticsvidhya.com/blog/2021/10/machine-learning-for-stock-market-prediction-with-step-by-step-implementation/?custom=TwBL912>)
- Stock Price Prediction- Machine Learning (<https://data-flair.training/blogs/stock-price-prediction-machine-learning-project-in-python>)
- Stock Market Prediction using Machine Learning (<https://www.mdpi.com/2079-9292/10/21/2717/pdf>)
- Avijeet Biswal, An Easy Guide to Stock Price Prediction Using Machine Learning
- <https://www.sciencedirect.com/science/article/pii/S1877050920307924>
- <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00333-6>
- <https://www.analyticsvidhya.com/blog/2021/12/stock-price-prediction-using-lstm/>
- <https://thinkingneuron.com/predicting-stock-prices-using-deep-learning-lstm-model-in-python/>
- <https://www.hindawi.com/journals/jmath/2021/1275637/>
- <https://www.econ.berkeley.edu/sites/default/files/Selene%20Yue%20Xu.pdf>
- <https://www.emerald.com/insight/content/doi/10.1108/IJCS-05-2020-0012/full/html#sec001>