Stock Price Predictor

Project Report

For Seminar Course (CS709)

by

Sumedh Kamble (212IS012)

Under the Guidance of

Kushagra Gupta

DEPARTMENT OF COMPUTER SCIENCE

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALORE-575025

Jan, 2022

DECLARATION

I hereby declare that the Report of the P.G. Project Work entitled **Stock Price Predictor** which is being submitted to the National Institute of Technology Karnataka Surathkal, in Completion of course work of seminar(CS709) of the requirements for the award of the Degree of Master of Technology (M.tech) in Computer Science Engineering(Information Security) in the Department of Computer Science, is a bonafide report of the work carried out by me. The material contained in this report has not been submitted to any University or Institution for the award of any degree.

Sumedh Kamble (212IS012)

Department of Computer Science

Place: NITK, SURATHKAL

Date: 10-1-2022

CERTIFICATE

This is to certify that the Seminar Project Work Report entitled **Stock Price Predictor** submitted by ***Sumedh Kamble (212IS012)*** as the record of the work carried out by her, is accepted as the Seminar Project Work Report submission in fulfilment of the requirements for the Seminar Course(CS709) Completion in Computer Science Engineering(Information Security) in the Department of ***Computer Science***.

Internal Guide

Mrs. Radhika B.S and Kushagra Gupta

Chairman - DPGC
(Signature with Date and Seal)

ACKNOWLEDGEMENT

I would like to express my kindest appreciation to all those who provided me the possibility to complete this report. Special gratitude to my guide for providing me the opportunity to work on a project in virtual reality. Working in the field was a great experience and want to explore this field more. I take this opportunity to express my deep sense of gratitude towards those, who have helped me in various ways.

Last but not the least, I would like to express my gratitude towards my parents and friends for their kind cooperation and encouragement which helped me in the completion of work.

ABSTRACT

A stock market, equity market, or share market is the aggregation of buyers and sellers of stocks (also called shares), which represent ownership claims on businesses. Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit. The efficient-market hypothesis suggests that stock prices reflect all currently available information and any price changes that are not based on newly revealed information thus are inherently unpredictable. Others disagree and those with this viewpoint possess myriad methods and technologies which purportedly allow them to gain future price information. Accurate prediction of stock market returns is a very challenging task due to volatile and non-linear nature of the financial stock markets. With the introduction of artificial intelligence and increased computational capabilities, programmed methods of prediction have proved to be more efficient in predicting stock prices. In this work, Artificial Neural Network and Random Forest techniques have been utilized for predicting the next day closing price for companies belonging to different sectors of operation. As part prediction model, historical prices are combined with sentiments and then trends are predicted. The financial data: Open, High, Low and Close prices of stock are used for creating new variables which are used as inputs to the model. The models are evaluated using standard strategic indicators: RMSE and MAPE. The low values of these two indicators show that the models are efficient in predicting stock closing price.

Contents

1 INTRODUCTION

A stock market, equity market, or share market is the aggregation of buyers and sellers of stocks (also called shares), which represent ownership claims on businesses. Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit. The efficient-market hypothesis suggests that stock prices reflect all currently available information and any price changes that are not based on newly revealed information thus are inherently unpredictable. Others disagree and those with this viewpoint possess myriad methods and technologies which purportedly allow them to gain future price information.Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. Prediction methodologies fall into three broad categories which can (and often do) overlap.

1.1 History

Driven by the desire to predict market movements and reap profits, there are three different trading schools of thought: fundamental, technical, and quantitative technical analysis.

Fundamental analysis: Fundamental analysis involves the examination of economic factors that influence the price of a stock. Such factors include a balance sheet and income statement. The balance sheet is a financial statement that provides information about a company's assets, liabilities as well as the equity of their shareholders at a specific point in time. Basically, the balance sheet gives intel into what a company owns and owes and the amount investors have invested in it. The income statement is another type of financial statement that gives a synopsis of a company's performance by providing information about their revenues, expenses, and net profit/loss over time. These reports are released quarterly throughout the year. Because fundamental analysis relies on reports that are issued on the basis of a slower time frame, this type of analysis is often used to project long-term price movements.Evaluate a company's past performance as well as the credibility of its accounts. Many performance ratios are created that aid the fundamental analyst with assessing the validity of a stock, such as the P/E ratio.

Technical Analysis: The goal of technical analysis is to anticipate what other stock holders are thinking based on available information about the price and volumes of stock. Technical analyists use a number of different types of indicators calculated from the past history of stock price and volume to predict future prices. Overall, the key to technical analysis is trend. Practitioners of technical analysis argue that trends in stock prices are caused by an imbalance between the supply and demand of stocks, which is reflected in the bid and ask prices. From the noisy data of stock prices, technical analyists attempt to extract patterns. Technical analysis is largely qualitative because it relies on the visual analysis of stock charts [9]. Two examples of such stock charts are shown below. Fig. 1 represents historical price data for stocks for the IT sector for one year. (It should be noted that I have chosen the IT to train an algorithm because stocks in the same sector usually exhibit similar price movements). Fig. 1,2 and 3 represent candlestick charts. Each bar or "candle" represents one day's high, low, and closing prices. The additional lines on the charts are examples of different technical indicators. (Please note brief explanations of how technical traders would utilize these symbols are

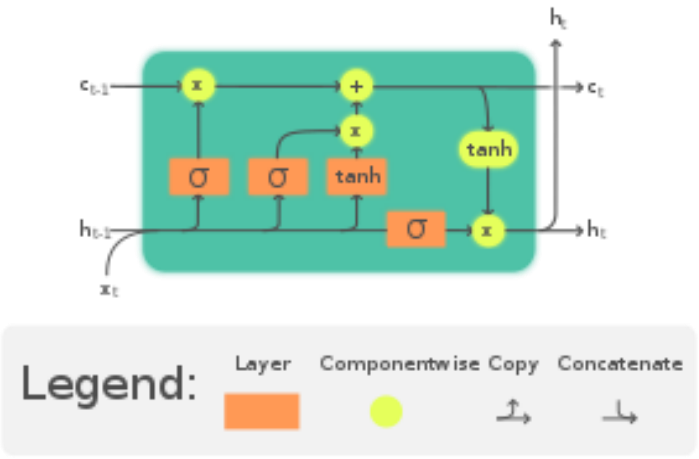
shown on the charts)They seek to determine the future price of a stock based solely on the trends of the past price (a form of time series analysis).

Quantitative Technical Analysis: This qualitative aspect to our second school of thought is what differentiates it from our next methodology. (Please note that we will revisit Quantitative Technical Analysis My capstone project explores this form of stock prediction. As suggested by its name, this form of stock prediction relies on quantitative methods of prediction rather than visualizations on graphs. Specifically, I will explore the usage of machine learnings algorithms to predict future stock prices.

Technological Methods: Due to advances in computer technology,stock market prediction has since moved into the technological realm. The most prominent technique involves the use of artificial neural networks (ANNs) and Genetic Algorithms(GA).For Example, It uses historical data to learn the patterns and predict future prices .A Recurrent neural network (RNN) is most commonly used Neural network for training.LSTM, a type of neural network is used in this project. It uses a certain time window to learn the features.

2 Long Short Term Model

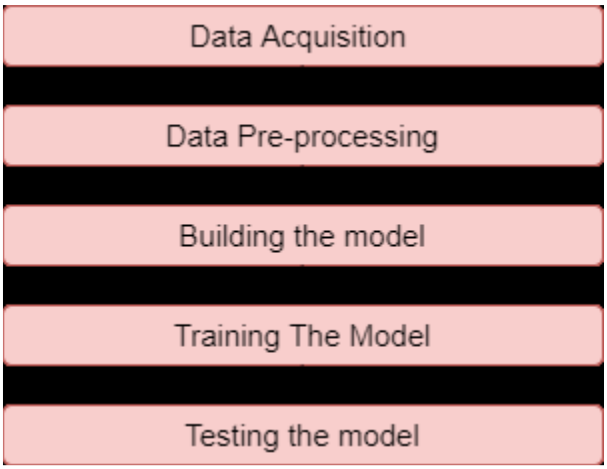
Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture[1] used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can process not only single data points (such as images), but also entire sequences of data (such as speech or video). For example, LSTM is applicable to tasks such as unsegmented, connected handwriting recognition,[2] speech recognition[3][4] and anomaly detection in network traffic or IDSs (intrusion detection systems).A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell.LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. LSTMs were developed to deal with the vanishing gradient problem that can be encountered when training traditional RNNs. Relative insensitivity to gap length is an advantage of LSTM over RNNs, hidden Markov models and other sequence learning methods in numerous applications.



The Long Short-Term Memory (LSTM) cell can process data sequentially and keep its hidden state through time.

3 Methodoloy

This chapter discusses the methodology. The following methods are utilized, considering research trends in topics about machine learning and artificial intelligence, particularly, machine learning. The over-all framework of this project is shown below



3.1 Data Processing

Historical data for last 10 years is collected from Yahoo Finance (yfinance).It has data like Date, Open Price, Close Price, Adjusted Price, Volume, High, Low.For prediction we have used Open, Close, High, Low, Adjusted Low from this dataframe.New variables like “7 days Moving Average” and “14 days Moving Average” are introduced for each data point which i believe have a certain impact on Stock Price and will be useful for predictions.Re-formatted data to adjust the 7MA and 14MA.

	Open	High	Low	Close	Adj Close	7MA	14MA
0	227.500000	230.000000	227.500000	229.119995	202.633774	227.691428	221.631786
1	230.800003	231.000000	228.009995	229.205002	202.708954	227.691428	221.631786
2	229.699997	231.500000	228.500000	230.580002	203.925018	227.691428	221.631786
3	230.600006	231.000000	228.110001	229.289993	202.784119	227.691428	221.631786
4	229.399994	230.389999	228.009995	228.604996	202.178329	227.691428	221.631786
...	...	...	...	...	...	...	...
2932	485.250000	495.399994	480.000000	493.049988	493.049988	499.199995	511.471433
2933	495.799988	500.000000	489.049988	493.149994	493.149994	496.599993	509.432146
2934	490.000000	495.450012	486.299988	490.549988	490.549988	494.314279	506.782144
2935	486.250000	487.899994	467.100006	470.500000	470.500000	490.799992	502.500000
2936	470.000000	477.750000	454.299988	465.100006	465.100006	486.078565	498.346429

2937 rows × 7 columns

3.2 Building the Model

The Model Consists of 6 Layers :

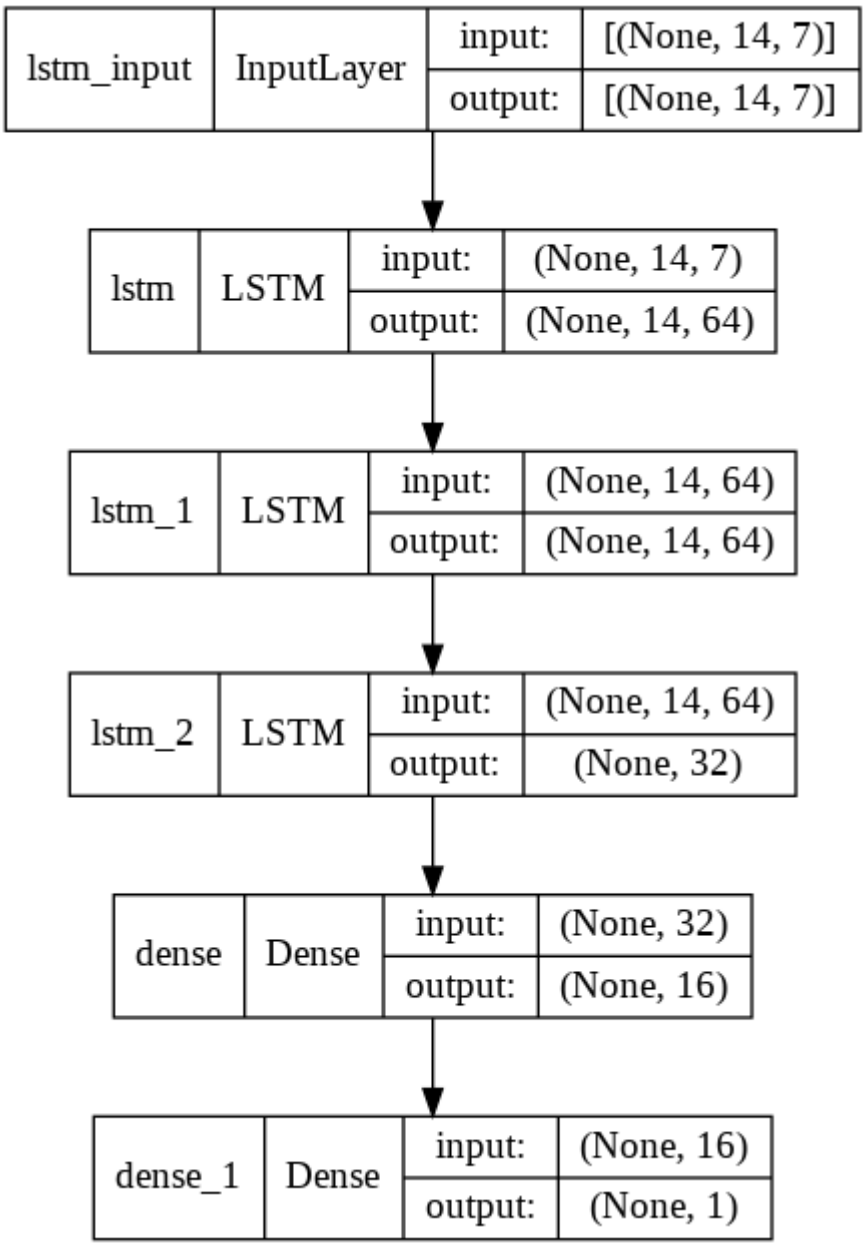
• 4 stacked LSTM layers.

• 2 Dense Layers.

Total parameters: 64,417
Trainable parameters: 64,417
Non-trainable paramaters: 0

The input layer would take 7 features as mentioned before.Output layer would give output for

Predicted Open Prices



3.3 Training the Model

Above model is trained with following configurations :

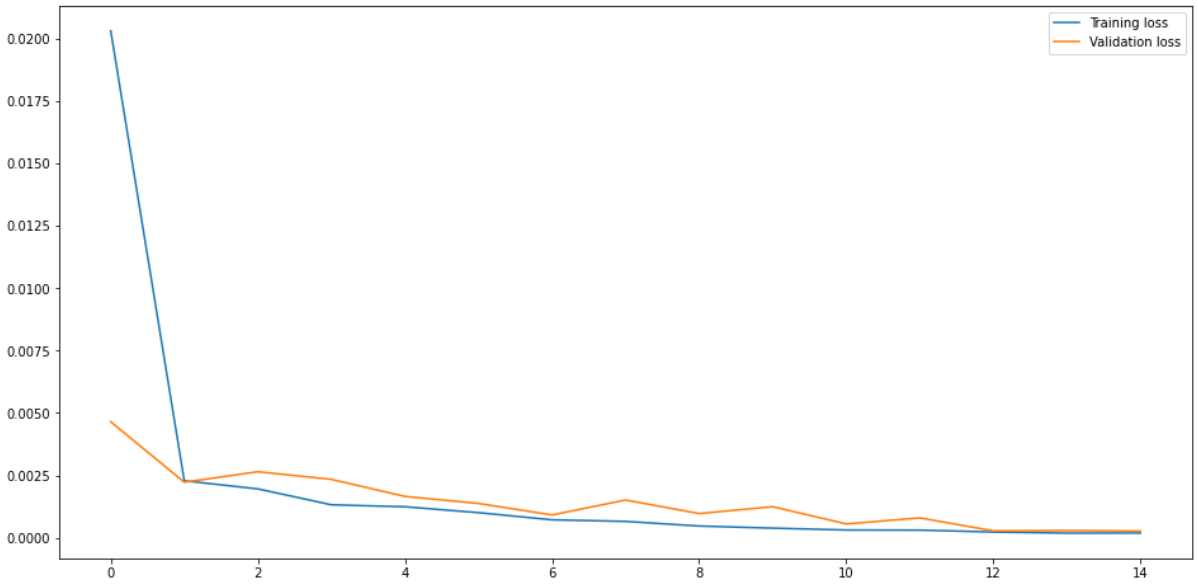
- Input Shape = n_samples x timesteps x n_features
- Training Data = 80%
  - 2349x 14 x 7
- Testing data = 20%
  - 574 x 14 x 7
- Here we are considering timesteps as 14 as we assume that past 14 days trend will decide the prediction ie. LSTM will learn from last 14 days and predict the 15th day prices.
- Number of Epochs = 15
- Validation set = 20%
- Batch size = 16.
- The input layer would take 7 features as mentioned before.
- Output layer would give output for Predicted Open Prices.
- Optimizer : Adam
- Loss function : MSE
- Metrics : Accuracy
- Activation function: ReLU

After model is created it is saved for future use. A small server is created for user interaction where user can enter the stock name and get output in Web Browser. Outupts are shown in Results section.

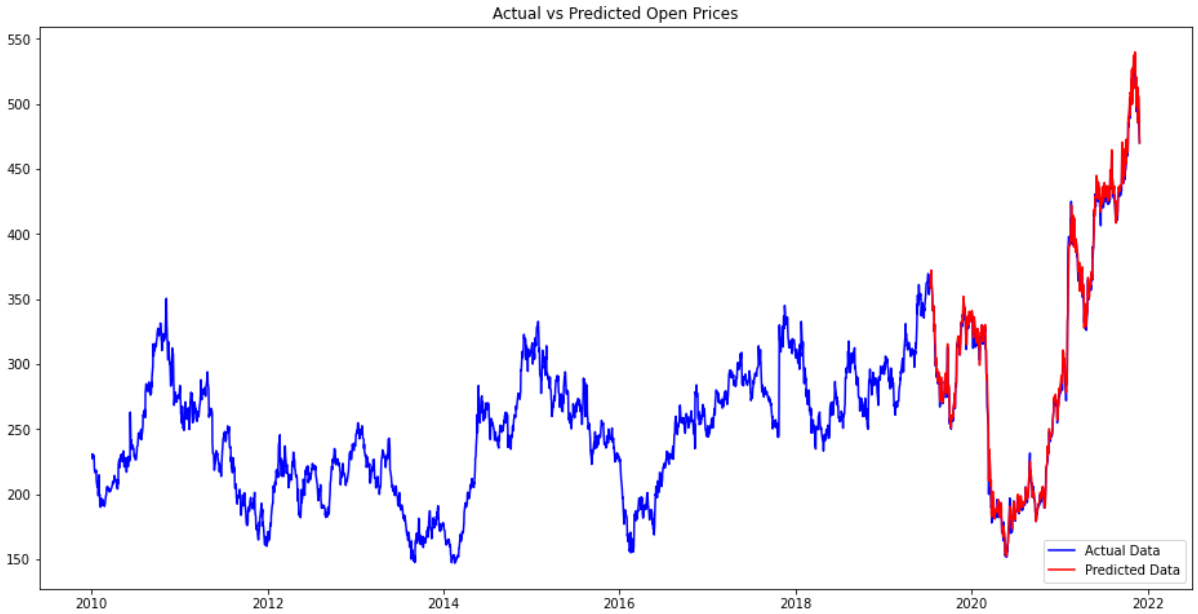
4 Results

Now that we have our model ready, we can use it to forecast the Open Price of the SBI stock by using a model trained using the LSTM network on the test set. This is accomplished by employing the simple predict function on the LSTM model that has been created. Finally, now that we've projected the values for the test set, we can display the graph to compare Open Prices and Open Prices predicted value using the LSTM Machine Learning model. The model was evaluated on validation split of 0.2% and after hyperparameter tuning tit was test on testing data. Following results are obtained.

4.1 Train v/s Validation Loss

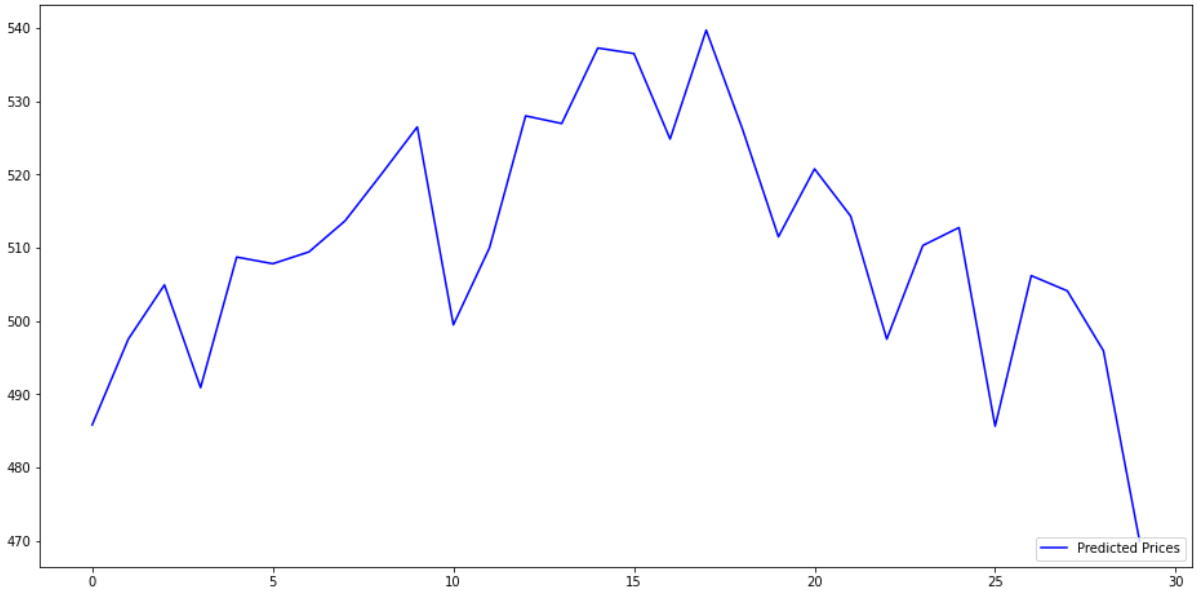


4.2 Actual v/s Predicted Prices



The graph above demonstrates that the extremely basic single LSTM network model created above detects some patterns. We may get a more accurate depiction of every specific company's stock value by fine-tuning many parameters and adding more LSTM layers to the model.

4.3 Monthly Prediction



Above graph represents the next month prediction for SBI Stock.

4.4 Performance Metrics

- **Mean Absolute Error:** The numerical Absolute difference between the predicted value and the actual value.
  - **Observed Value:** 3.64

- **Root Mean Squared Error:** Standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are.RMSE is a measure of how spread out these residuals are.
  - **Observed Value:** 5.08

4.5 Trading Strategy and generating CALLs

- When the 7MA crosses above the 14MA, it's a BUY signal, as it indicates that the trend is shifting up. This is known as a "golden cross."
- When the 7MA crosses below the 14 MA, it's a SELL signal, as it indicates that the trend is shifting down. This is known as a "dead/death cross."

```
2019-09-16 00:00:00 Buy at Price : 293.122802734375
2019-09-17 00:00:00 Buy at Price : 288.63946533203125
2019-09-18 00:00:00 Buy at Price : 278.6844482421875
2019-09-20 00:00:00 Buy at Price : 277.1058654785156
2019-10-01 00:00:00 Buy at Price : 274.9959106445312
2019-10-03 00:00:00 Buy at Price : 259.508056640625
2019-12-10 00:00:00 Sell at Price : 319.8402099609375
2019-12-13 00:00:00 Sell at Price : 324.9873962402344
2019-12-16 00:00:00 Sell at Price : 336.67626953125
2019-12-17 00:00:00 Sell at Price : 336.79827880859375
2020-04-23 00:00:00 Buy at Price : 188.989013671875
2020-04-24 00:00:00 Buy at Price : 188.44374084472656
2020-04-27 00:00:00 Buy at Price : 183.0615692138672
2020-04-28 00:00:00 Buy at Price : 182.34320068359375
2020-06-04 00:00:00 Buy at Price : 177.18685913085938
2020-06-16 00:00:00 Buy at Price : 175.4935760498047
2020-06-17 00:00:00 Buy at Price : 174.22682189941406
2020-06-18 00:00:00 Buy at Price : 173.58595275878906
2021-03-02 00:00:00 Sell at Price : 401.182373046875
2021-03-04 00:00:00 Sell at Price : 411.87884521484375
```

5 Conclusion and Future Work

However, with the introduction of Machine Learning and its strong algorithms, the most recent market research and Stock Market Prediction advancements have begun to include such approaches in analyzing stock market data. The Opening Value of the stock, the Highest and Lowest values of that stock on the same days, as well as the Closing Value at the end of the day, are all indicated for each date. Furthermore, the total volume of the stocks in the market is provided, With this information, it is up to the job of a Machine LearningData Scientist to look at the data and develop different algorithms that may help in finding appropriate stocks values. Predicting the stock market was a time-consuming and laborious procedure a few years or even a decade ago. However, with the application of machine learning for stock market forecasts, the procedure has become much simpler. Machine learning not only saves time and resources but also outperforms people in terms of performance. it will always prefer to use a trained computer algorithm since it will advise you based only on facts, numbers, and data and will not factor in emotions or prejudice.

The prediction can be improved by integrating sentiment analysis Tweets , News articles, market sentiment can be used for this.This can be done in future plans for increasing the accuracy.

6 REFERENCES

1)https://reader.elsevier.com/reader/sd/pii/S1877050920307924?token=945E023FD8ECEFB9BFD724BF9BAC91D7BD8CA63CDBAA612B9CB154F565CBD21D25CFE4CD53F8D9168138C1C9C799DDF8&originRegion=eu-west-1&originCreation=20211108053320

2) https://reader.elsevier.com/reader/sd/pii/S1877050916311619?token=F574E431DFBCBA9AC25BF3AEBC72651DD70749E0964D295617F021B6550FF0A79274DD5E6335AC13B71BB18E27A56382&originRegion=eu-west-1&originCreation=20211108050808