Stock Market Prediction System

Ritvik Sanjeev Patil
ritvikpat@gmail.com

H Gururaghavendra
hguru001@gmail.com

Yashas B Prasad
yashasbprasad@gmail.com

Department of Computer Science and Engineering
PES University
Bangalore, India

1. Introduction and background:

Stock Market Analysis and prediction is a project for technical analysis, visualization, and estimation using Yahoo Financial data. Seeing data from the market, especially some general and other software columns. Pandas used to take stock of the information, looked at different aspects of it, and finally looked at it in some way to assess the risk of a stock based on its recent performance history.

Stock exchange analysis is only intended for the analysis of stock company data for various organizations. Using this method of data analysis, any organization can easily extract relevant information

2. Aim of the Project :

The main goal of our project is to analyze the data of some of the top financial assets in the world , predict the future price of the asset and also to find some trading and investment opportunity.

3. Problem Statement:

Analysing the past data of the given assets and predicting the future value of that asset.

3.1 Importance

1) Performing analysis before making an investment is a must. It is only after a thorough research that you can make some assumptions into the value and future performance of an investment.
2) The accurate prediction of stock price movement will lead to more profit investors can make.
3) Stock market is an important part of the economy of a country and it can be used as an index to view the development of the country

3.2 Specific problem

1) Performing basic data analysis
    a) Getting the past 5 years data which indicate all the major attributes like opening price, closing

price, high price, low price and volume.

2) Calculating the daily and cumulative returns.

3) Calculating the volatility of the stock based on daily returns.

4) Finding the short term and long term moving averages.

5) Price analysis based on volume of transactions

6) Finding the correlation between the stocks

7) Predicting the future stock price by using the LSTM machine learning model.

8) Giving users buy and sell signals using SMA crossovers and bollinger bands strategy

4. Previous works:

In the past couple of years, with the exponential development in the field of data analytics, we have come across a wide number of projects that have done extensive research and analysis on stock markets.

The projects we have come across so far analyse the past data and provide insights as to where an individual could have invested during this timeline(past). When it came to predicting future stock prices and providing accurate buying and selling signals, these projects were lacking in these domains.

The above mentioned shortcomings have been addressed in our project. Our project not only analyses and learns from the past data, but also predicts the future stock prices by making use of machine learning algorithms.

We have also added advanced trading and investment strategies that help a novice/professional investor to invest by providing accurate buying and selling signals.

(refer (10) under references)

4.1 Assumptions

The stock market is influenced by various external factors such as weather, people sentiments, economy of the country, etc., which means there is randomness involved.

Since the future is highly uncertain, no model can make accurate predictions. The probability that a prediction is correct cannot exceed 60-70% because of the above mentioned factors.

Another assumption our model makes is that the investor is following an asymmetrical risk reward ratio to be profitable in the long run.

(refer (11) under references)

5. Proposed solutions

We humans learn from our past experiences. Similarly for any model to make a prediction, it needs to learn from data(past data).

The process of making the computer learn from data involves the following steps. (refer (4) under references)

5.1 Preprocessing

For a computer to make accurate predictions it has to learn from

clean data. Hence the first step involves cleaning the dataset.

First we tried to check for empty values in the dataset. This can be done using the python function :

```
dataset.isna().any()
```

Since our dataset was obtained from a reliable source, it was already clean.

Another prerequisite for the computer to learn from data, the values of the input data must be of the same scale. To scale the data we have made use of MinMaxScaler().

5.2 Building a model

To train our model, we split the input dataset into a training and testing set of which the training data consisted of 80% of the original dataset and the remaining 20% was used for testing.

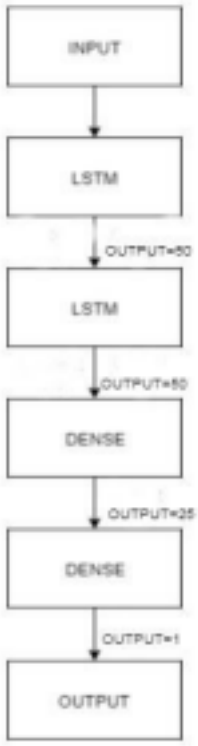
For training the model, we have made use of LSTM recurrent neural network architecture.

The architecture includes 4 hidden layers. The first two layers are made up of 50 neurons each followed by a 25 neurons third layer and a single neuron fourth layer.

We have used a single epoch for training our model. This can be justified by the following reasons:

- Every epoch consumes a lot of time, we are faced with time constraint.
- As the number of epochs increase, there seems to be a gradual decline in the accuracy followed by an increase in the long run.

Since we are working under time constraint we have stuck to a single epoch.



LSTM FLOWCHART

5.3 Evaluation

After predicting the future stock price, we check its root mean square error value, thereby, verifying its accuracy.

We have also evaluated the model by visualizing the predicted data and the actual data in the same graph.

Given the training dataset, the model can predict the next day's closing price which can be cross verified with the actual outcome on the next day.

5.4 Component Solutions

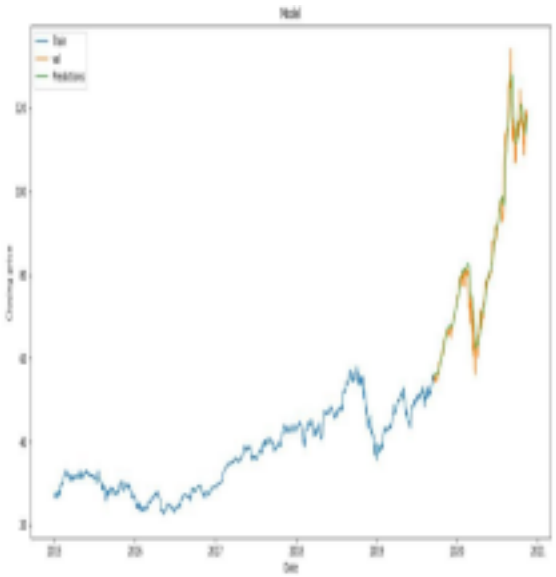
Our project can be broken down into

three basic components.

1) Machine learning model
2) Visualization
3) Trading and investment strategies

5.4.1 Machine learning model

This component happens to be the main component of our project that is responsible for learning from the data and predicting the future stock prices. The model that we have designed has been explained in detail in the previous sections.



PREDICTIONS BY MODEL

5.4.2 Visualization

This component is useful for the user so that he/she develops an understanding of the dataset the model is being trained on. Visualization is also a great tool to evaluate the outputs of the model. The visualization plots that we have used to gain an understanding of the dataset has also helped us to visualize the predictions of our model and verify its correctness. An in detail explanation of the insights we have gathered from the plotted graphs will be explained in the upcoming section.

Predictions
55.495468
55.732483
55.967590
56.130608
56.157627
...
115.341682
116.083679
116.626312
117.409363
118.152573

PREDICTED PRICE ATTRIBUTE

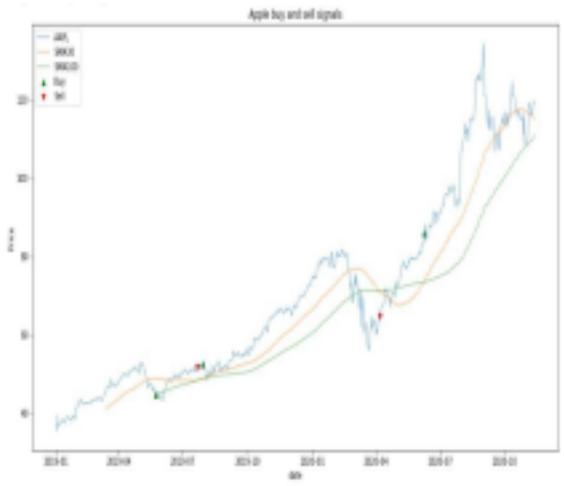
5.4.3 Trading and investment

strategies

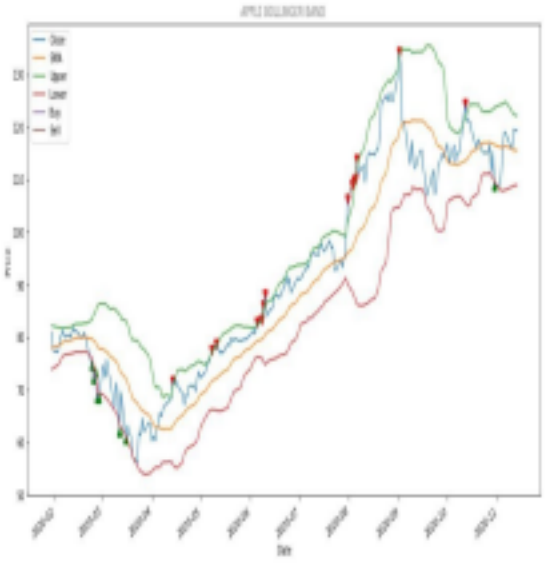
After looking at the prediction of our model, we advise the investor on the trading strategies he can utilize to earn profits.

The first strategy is that of a moving average crossover. It gives a buying signal whenever a short term moving average crosses over the long term moving average and it gives a selling signal whenever a short term moving average crosses below the long term moving average.

The second strategy makes use of bollinger bands which depends on simple moving average and standard deviation to calculate upper and lower bands. This strategy gives a buying signal whenever the price goes below the lower band and gives a selling signal whenever the price goes above the upper band.



Output of SMA Trading Strategy



Output of Bollinger band Strategy

6. Experimental results

In this section we will be addressing the specific problems(components) that were mentioned in the problem statement.

We have analysed the various attributes of the dataset by visualizing them. 1) To analyse which stock is performing well, we have plotted the prices of all stocks on the same graph in which the x axis is the investment horizon and the y axis is the price.Using this graph we can conclude which stock would have been a good investment.

2) To analyse if there was any major event that happened in the company, we have plotted a graph in which the x axis is the investment horizon and the y axis

is the volume.Using this graph we can conclude if it would be a good decision to hold our investments.

3) To find out the major trend of the stock, we have plotted the short term moving average and the long term moving average since it is very important to know if our investments are going to grow in the future or not.

4) To find out the correlation between stocks, we have plotted a scatter plot. This graph is important because if any one of the businesses get affected because of any external reasons, there is a high probability of another highly correlated stock being affected which might have a negative impact on our investment.

5) We have plotted a histogram and a kernel density estimation graph to find out the volatility of our investments. This tells us that on any given day, how much impact we will have on our investments.

6) We have plotted a box plot to find out the extreme outliers which we should consider before investing to prepare ourselves for the extremes.

7) We have also plotted a graph for visualizing the cumulative returns. This graph is important because it shows how much our investments would be worth currently and the percentage of gains we have achieved.

6.1 Scenario under which our model works the best

Our model works well:

1) When the market is stable.
2) When there is no panic among the investors.
3) When people are not scared to hold their investments.
4) When the market is not manipulated.
5) When the stock market is strictly monitored from illegal activities by the respective boards

6.2 Scenario under which our model fails

Our model fails:

1) When there is panic among investors.
2) When there is huge selling. 3) When a country enters recession. 4) When there is irregular supply and demand of stocks in the market
5) When there is positive/negative news about a stock in the market.

7 Conclusion

At the conclusion of this project we will have experienced a piece of the business world, investing money and presenting data and different graphical representations. After we have completed the project, we think about what we might have done differently to get different results when investing

money.

Investing is risky and, as some of us saw, we can lose money just as quickly as we can make it.

The Technique we have used in this paper is LSTM on the Yahoo finance dataset. (refer (9) under references) The techniques have shown an improvement in the accuracy of predictions, thereby yielding positive results. Use of recently introduced machine learning techniques in the prediction of stocks have yielded promising results and thereby marked the use of them in profitable exchange schemes. It has led to the conclusion that it is possible to predict stock market with more accuracy and efficiency using machine learning techniques.

In the future, the stock market prediction system can be further improved by utilizing a much bigger dataset than the one being utilized currently. This would help to increase the accuracy of our prediction models. Furthermore, other models of Machine Learning could also be studied to check for the accuracy rate resulting from them.

8 References

1) Business Analytics, The Science of Data-Driven Decision Making, U Dinesh Kumar, Wiley 2017.

2) Data Mining: Concept and Techniques by Jiawei Han, Micheline Kamber and Jian Pei, The Morgan Kaufmann Series in Data Management System, 3rd Edition

3) The Elements of Statistical Learning Trevor Friedman, Robert Tibshirani and Jerome Hastie, Data Mining, Inference and Prediction, Springer 2001

4) https://towardsdatascience.com/how-to-use-data-science-on-the-stock-market-9d97a4d348d6

5) https://www.kaggle.com/

6) https://www.datacamp.com/projects

7) https://www.ml.cmu.edu/research/data-analysis-projects.html

8) https://www.edureka.co/

9) https://www.researchgate.net/publication/331345883_Stock_Market_Prediction_Using_Machine_Learning

10) https://www.analyticsinsight.net/data-analytics-and-stock-trading-how-to-use-data-science-in-stock market-analysis/

11) https://www.computersinthecity.co.uk/how-technology-has-influenced-the-stock-market/#:~:text=Today%20the%20high%20speed%20computers,stocks%20to%20buy%20and%20sell.&text=Along%20with%20the%20benefits%20there,manipulating%20the%20price%2 0of%20stocks.