Stock Analysis using Sentiment Analysis and Machine Learning

Mihir Athale[1], Sumedh Nakod[2], Anmol Kumar[3]
[1,2,3]*Student, B.Tech Computer Science Engineering, Narsee Monjee Institute of Management Studies, Mukesh Patel School of Technology, Management and Engineering, Shirpur, M.H, India*

***Abstract*- Public image of a company plays a very major role on its position in the market. It may also affect the rise or fall in the stock price of the company. Public image is the result of the news articles, tweets etc posted over the internet or through any other form of media. Hence, this paper aims to analyze the relationship between public sentiment and stocks with the help of sentiment analysis and machine learning approach.**

I.INTRODUCTION

Imagine one day a person is reading an article about a company and something negative is written about the company in it. So, the person would form an image of the company in his mind that "this company is not right for investment as there are certain negative points about it in the media". Now this image is based on a single article in a single persons mind.

There are numerous articles or tweets on twitter over the internet, which are read by the people, hence forming certain kind of image in their mind. But it is impossible for a human to go through a huge amount of text and analyze them all. Hence, we propose a solution to this problem with the help of computer technologies like NLP, Sentiment Analysis and Machine Learning.

Computer technology can enable us to analyze a huge amount of data and get an insight over it easily and efficiently. Hence, we will use these technologies to analyze the effect of sentiments or public opinion on the stocks.

II. REVIEW

As financial sector is a very huge sector that deals with huge amounts of money, technology has already been applied in the sector to get useful results from it. Stock Analysis and Prediction is one of the field in which a lot of research has been done.

People use techniques like regression on the historical stock price data to predict the future values and help people invest accordingly in the sector. This technique uses various machine learning and regression algorithms to produce numeric results containing the future stock prices. Though this method is not very accurate, it is a very useful to get the insight about the future stock prices. Sentiment analysis approach has also been used in stocks to get the insight whether the stocks will fall or will they rise based on the sentiment or public opinions available openly.

III. PROCESS

We have adopted the following procedure to complete our study. It involves various steps which have been explained below. The whole coding process was performed using python as the programming language.

A. Data Gathering

We started by gathering data from twitter. The data was gathered in the form of tweets. The twitter API was used to fetch the tweets from the web. The python module that was used to fetch the tweets was tweepy. In the tweet search module the query was passed as 'tesla'. So through this all the tweets between the specified time periods were fetched with the API. For the task we took the data for five days. The tweets were fetched and they were stored in a json file with their respective dates as the key for each tweet. The json files helped us in managing all the tweets date wise. On an average we gathered about 18,000 tweets for each day. Apart from the tweets we also downloaded the stock prices from yahoo finance for the company 'Tesla'. This was in the form of a CSV file with various data parameters.

B. Pre-Processing

Next step was the processing of all the data that was gathered in the previous step. The tweets on the twitter website are posted by the public. These tweets are basically in the form of textual sentences but also have some special symbols like the emoticons. Along with these emoticons there is also data like the username of the person who posted the tweet, the username of the person who the user mentioned in their tweet, name of some website, images etc. This type of data cannot be used for the task like sentiment analysis. Sentiment analysis is performed on textual data to obtain the user sentiment from it, data like emojis, usernames, websites etc are not the kind of attributes from which we can obtain some sentiment score. So it is necessary to remove the useless data from the fetched tweets. We used regular expressions to filter out this data from the tweets. This process is known as cleaning of the data. The tweets which were stored in the form of json file were imported with the help of pandas library into a pandas data frame. This data frame was used to organize the tweets properly according to their dates in a tabular form. Then the process of cleaning was applied to each tweet that was stored in the dataframe. After the process of cleaning we got the text which was clean and suitable for sentiment analysis.

C. Sentimental Analysis

Sentiment analysis is a process in which natural language processing, computational linguistics and text analysis is used to analyze the subjective information present in a piece of text. In sentiment analysis the algorithm tries to count the number of good words, bad words and neutral words in the text and then generate a final sentiment score. This sentiment score represents the overall sentiment of the piece of text that was analyzed. In today's world, with the increase of technology, huge data is generated on a daily basis. This data is analyzed with the help of sentiment analysis to provide an insight for companies to develop their future strategies.

In our project the sentiment analysis process was applied on the tweets that we obtained after the cleaning process. The sentiment analysis module was imported from a library known as textblob. It is a python library which is used for natural language processing applications. So, for each tweet stored in the pandas data frame according to the dates, the polarity function of the textblob module was applied. This function returns as score based on the sentiment of the text. This score can be positive negative or zero. Based on positivity and negativity of the score we labeled the tweets as positive, negative or neutral. New columns were added to the data frame to store this data. All the tweets for each day were labeled as positive, negative or neutral.

The next task was to calculate the overall sentiment score for a particular day. All the tweets for each day were analyzed one by one. The number of positive, negative and neutral tweets were counted and the final sentiment score was generated by the formula = (Pos - Neg)/ (Pos + Neg).

	Date	Sentiment Score	Sentiment
0	2020-03-19	0.421287	Pos
1	2020-03-17	0.379327	Pos
2	2020-03-20	0.382173	Pos
3	2020-03-18	0.339758	Pos
4	2020-03-15	0.531697	Pos

Neutral sentiments weren't considered as they hardly contribute to the polarity of the statements. The generated overall average sentiment score was stored in a dataframe date wise. This dataframe was saved in a CSV file. Hence, this sentiment score will be used to analyze the effect of sentiment on the stock prices.

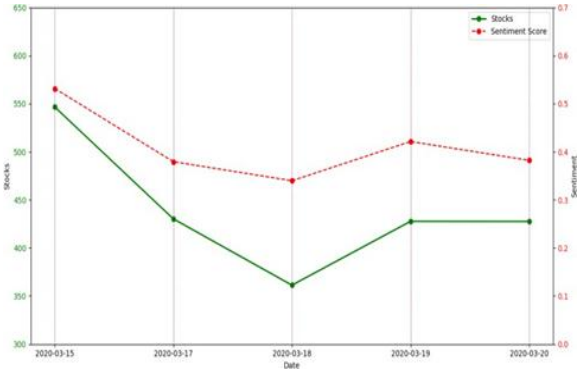
D. Analysis

For the analysis phase, we selected the stock data from the downloaded stock data for Tesla for the same five dates for which the sentiment score was calculated. We only selected the closing price of the stock for our analysis. This stock data was plotted on a twinx plot along with the sentiment scores for each respective day. Also we calculated the percentage of the positive, negative and neutral tweets we obtained per day and plotted the data in a pie chart. Along with this we also made a wordcloud to analyze the most occurred word in the tweets. All the three plots were used to analyze the result. The plots were created using the matplotlib library in python.

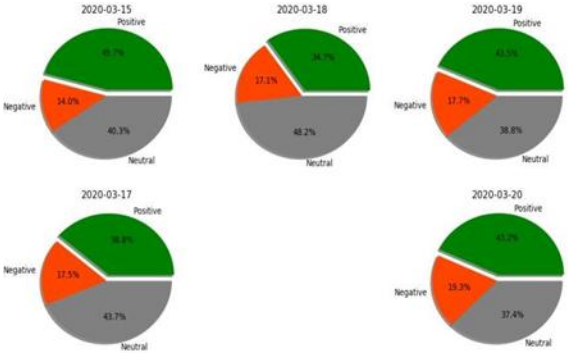
IV. RESULT

The following graph represents the relation between the overall sentiment and the stock closing price. We can see that the red dotted line represents the sentiment score and the green line represents the closing price. We have a separate y- axis for each

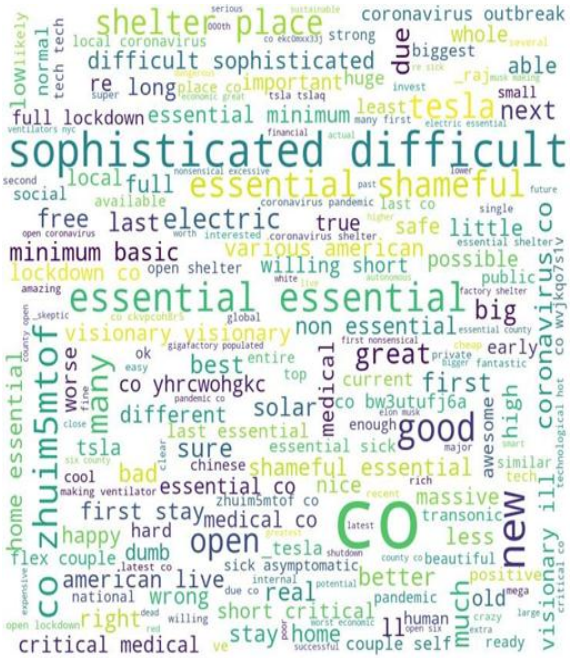
respectively. We can see that the relation is of dependence between the closing price and the sentiment score. We can see that when the sentiment score decreases from a day to another day the stock price also falls. This decrease in sentiment score can be seen as a decrease in the ratio of no. of positive tweets to the no. of negative tweets. We can also see that, even though the overall sentiment remains positive the decrease in the ratio of the sentiments also plays a major role in this case. The closing price decreases as the sentiment score does and it also increases when the sentiment score increases. Hence, we can truly say that sentiment score affects the stock price to some extent.



The graph given below is a representation of the percentage of the number of positive, negative and neutral tweets per day. This chart was used to analyze the day to day change in the tweets. The orange slice represents negative, green represents positive and grey represents neutral tweets. We can also relate this with the above chart that when the percentage of the negative tweets increases the sentiment score falls and vice versa. Hence, we can say that not the exact number but the percentage too affects the sentiment score.



The next plot is a wordcloud, wordcloud is just a collection of words in a pictorial representation. It helps us represent the most frequently occurring words in an image for a better analysis. In our case we can see the positive and negative words in the cloud according to their occurrence in the tweets.



V. FUTURE WORK

From our study we have seen that there is a dependent relation between the stocks and the public sentiment. So, this study can be further developed to:-

- Predict stock prices in the future for trading applications.
- Be used by people to predict whether a stock will rise or fall (with some probability) in the future and then invest accordingly.
- Also, companies may use sentiment analysis to analyse what is being published about them on the internet and then they may be able to control the wrong information.
- Hence, this application may prove to be useful for market trading in the future.

VI. CONCLUSION

Hence, from this paper we can conclude that the public sentiment plays a crucial role in the stock market. From our results we can see that there is a effect of the sentiment score on the closing price of the Tesla's stocks. The positivity in the public

sentiment leads to increase in the closing price and negativity in the public sentiment leads to the decrease in the closing price. So, we can finally say that the relation between the closing price and the public sentiment is of dependency. This result may prove to be very useful for the public and in the investment sector too. People can get insight from the sentiment score and invest accordingly in the stock market. Also, it may help the companies to study the impact of the public sentiment about their company on their stocks.

REFERENCES

[1] Mittal, Anshul, and Arpit Goel. "Stock prediction using twitter sentiment analysis." Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarket PredictionUsingTwitterSentimentAnalysis. pdf) 15 (2012).

[2] Xiaodong Li, Haoran Xie, Li Chen, Jianping Wang, Xiaotie Deng, News impact on stock price return via sentiment analysis, Knowledge-Based Systems, Volume 69,2014,Pages 14-23, ISSN 0950-7051

[3] Tushar Rao and Saket Srivastava. 2012. Analyzing Stock Market Movements Using Twitter Sentiment Analysis. In Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012) (ASONAM '12). IEEE Computer Society, Washington, DC, USA, 119-123.

[4] Nguyen, Thien Hai, Kiyoaki Shirai, and Julien Velcin. "Sentiment analysis on social media for stock movement prediction." Expert Systems with Applications 42.24 (2015): 9603-9611.

[5] Thien Hai Nguyen, Kiyoaki Shirai, Julien Velcin, Sentiment analysis on social media for stock movement prediction, Expert Systems with Applications, Volume 42, Issue 24, 2015, Pages 9603-9611, ISSN 0957-4174

[6] Sentiment Analysis for Effective Stock Market Prediction Shri Bharathi1* Angelina Geetha2 Department of Computer Science and Engineering, B.S.Abdur Rahman University, Vandalur, Chennai-600 048, Tamil Nadu, India * Corresponding author's