



SENTIMENTAL ANALYSIS OF STOCK MARKET VOLATILITY BY MACHINE LEARNING

Aishwarya V Kambar¹ Anjali Mathew² Ayesha Siddika³ Kalyan S⁴ ,Ranjith J⁵

SJB Institute of Technology, Bengaluru¹⁻⁴

Abstract: Stock Market Prediction is a platform where an enormous amount of data exists and constantly needs to be scrutinized. Traditional stock market prediction models are based on analysis of historical market data. Stock Market Prediction on the basis of public sentiments based on social media – an intriguing field of research. Positive tweets and news in social media encourage people to invest in the stocks of the company, thereby increasing the stock price of the company. Data sets are so large that it is difficult to analyze them using traditional data processing applications. We can extract the useful information to an understandable structure using machine learning algorithms, and Sentiment Analysis as the approaches, we implement and simulate a brokerage system and analyze the stock market.

Key Words: Stock Market, Sentimental analysis, TextBlob, Machine Learning Algorithms, Natural language Processing, WebScraping

I. INTRODUCTION

Nowadays the stock market has been playing a vital role in people's lives. There are companies who have lost their market price. Ever since the pandemic has hit us, the economy of our country has dramatically reduced. It's not only made the lives of people difficult but also has reduced the stock price of various companies. To predict the stock market, we used various machine learning techniques like naive bayes algorithm, Random forest[1], Trend Predictive Model[2], Time series analysis[4] and so on. The popularity of the online media-driven social network relation is proven in today's digital era. In our paper we predict the stock market volatility using sentimental analysis. We consider the people's opinion, their emotions and their personal feelings to determine what they think about a particular company. We use the method of web scraping for collecting data from twitter and with the help of natural language processing and Textblob library, we determine the sentiment of the quotes. Finally we predict the percentage of positive, negative and neutral sentiments.

II. LITERATURE SURVEY

The stock market prediction approaches have been built upon technical and fundamental analyses of the stocks. In recent studies, it has been evident that there is a strong correlation between news articles related to a company and its stock price movements.

Saloni Mohan, Sahitya Mullapudi[2], gathered data includes daily stock prices for S&P500 companies for five years, along with more than 265,000 financial news articles related to these companies. Jaeyoon Kim, Jangwon Seo[3] news articles were downloaded from a website called 'bigkinds' for two years articles from 2016 March 8 to 2018 December

31. The total number of news articles is 67,968. Venkata Sasank Pagolu, Kamal Nayan Reddy Challa[5] collected a total of 2,50,000 tweets over a period of August 31st, 2015 to August 25th, 2016 on Microsoft are extracted from twitter API.

Collected articles have been processed in different ways to extract features. Saloni Mohan, Sahitya Mullapudi[2] used cloud computing as an invaluable resource for training prediction models and performing inference for a given stock in real time. They used the RNN LSTM with Stock Prices and Textual. They trained the model based on current polarity and previous stock prices. Jaeyoon Kim, Jangwon Seo[3] the method they used for the emotional analysis is to construct a new sentiment dictionary based on the news articles. This helped them obtain the correlation between the positive index for each date (DP) and the return value of the next day's stock price (SPR). Ganapati Panda, Babita Majhi[5], used two different textual representations, Word2vec and N[1]gram, for analyzing the public sentiments in



tweets. They developed a sentiment analyser that could determine the sentiment of the tweet.

Finally the outcome of their experiments showed different accuracies and results. The RNN-pp model[2] performed well for companies for which we had more textual data when compared to other models described in their paper. The models did not perform well in cases where stock prices are low or highly volatile. Based on the experiment[3], approximately 0.3034 correlation value was checked. Due to the nature of Korean language, it was relatively difficult to derive its high accuracy by using natural language processing methods. The model with LibSVM[5] is trained with 90 percent of data, it gave a result of 71.82% and showed a good correlation between stock market movements and the sentiments of the public expressed in twitter.

III. PROPOSED SYSTEM

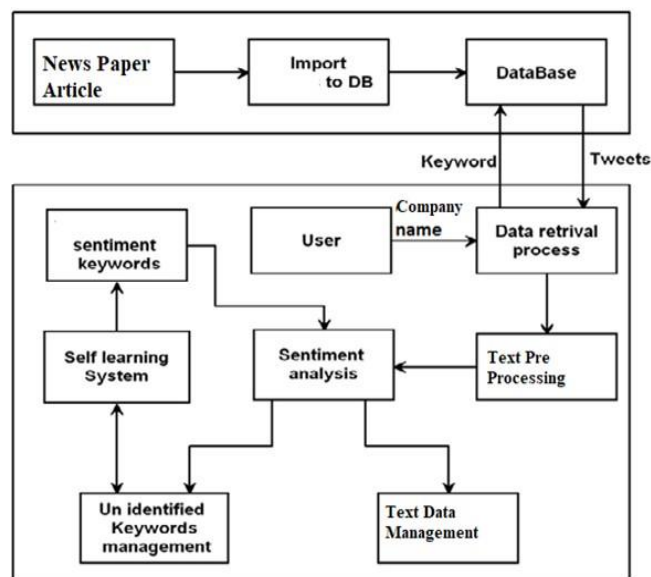


Figure 1.1 Proposed System Architecture

The figure 1.1 shows the architecture of our system. In our proposed system, we have divided the system into two main parts. One part is for data collection and the other part is for data processing and analyzing the data.

For the data collection phase, the user has to enter a company name. The most recent tweets from twitter which have the company name are collected. All the tweets are collected in the form of quotes and they are stored in a text document. These text documents are stored in a folder with the company name. This folder becomes our database.

The text document is now uploaded to the system for preprocessing. The system cannot understand human language/words. To make the system understand these words, we use the Natural Language Processing (NLP) to process the data. Once the data is preprocessed we use the sentiment function from the TextBlob library in Python to determine the sentiment of the text. The sentiment is categorized into three, one is positive, two is negative and the third is neutral.

There may be certain words in the text that are unidentified by the system. In such cases, we use the self learning system to identify these words and again send them for analysis. After the complete analysis, the results are shown as a graph which give the percentage of each sentiment.

IV. IMPLEMENTATION

In our project, we have written the program using Python language. Python language provides constructs intended to enable writing clear programs for developing complex scientific and numeric applications on both a small and large scale. There are several methods on how the data can be collected, one of which is downloading a dataset called 'bigkinds'[5]. Here we create our own dataset by collecting data from social media sites like google, twitter by using the twitter API through web scraping. These data collected are stored in a text document.



The collected data is in the form of quotes and we need to break down these words in order to analyse the sentiment. We use the Natural

Language Toolkit (NLTK) API for Natural Language Processing (NLP). NLTK is a leading platform for building Python programs to work with human language data. NLTK helps with the preprocessing of the collected data. Another dataset is constructed which contains the testing data, that is the data we use to analyze our input data.

Now the user is allowed to enter a company name and all the tweets with the company name are retrieved from twitter and stored in the form of a text document. This text document is then uploaded to the system by the user for data processing. The stored data is preprocessed and sent for analysis. We use the Natural Language Processing used for processing textual data. The sentiment function of the TextBlob library analyze the sentiment of the text and returns two properties, the polarity (positive,1 or negative -1)and the subjectivity, where the value is between0 and 1. Once the sentimental analysis is done, the number of positive sentiment and the number of negative sentiment is determined. With the help of these values we plot a graph in terms of percentages.

V. EXPERIMENTS AND RESULTS

The tweets of two companies Microsoft and Capgemini respectively stored into a folder after being entered by the user

Here is Sentimental Analysis of the company you requested	
Positive Sentiment	10
Negative Sentiment	1
Neutral Sentiment	189

Fig 1.2: Result Page

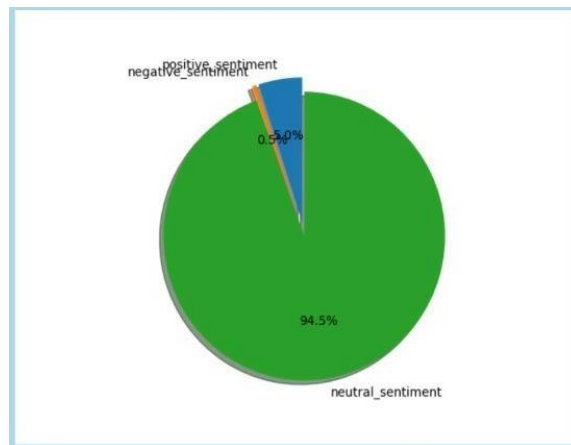


Fig 1.3: Pie-chart

The figures 1.3 and 1.4 show the sentimental analysis of Capgemini obtained by uploading the company name and the user will get the output inthe form of pie-chart.

Positive Sentiment	25
Negative Sentiment	0
Neutral Sentiment	175

Fig 1.4: Result Page

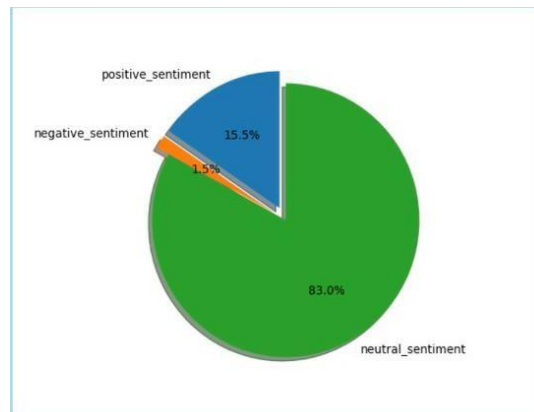


Fig 1.5: Pie-chart

The above figures show the sentimental analysis of Microsoft obtained by uploading the company name and the user will get the output in the form of pie-chart.

VI. CONCLUSION

In this paper, we are able to predict the sentiments of individuals through the tweets and news to predict the behavior of the stock market. Rise or fall in stock of a company to the public opinions or emotions about that company expressed on twitter through tweets. Sentiments of the news and tweets are calculated for the prediction of stock news. The sentiment of the tweets are classified into three categories: positive, negative and neutral. If the sentiment is positive, then we can state that this tweet's impact is good in the market, so more chances of the stock market to go high in trend. And if the sentiment is negative, then it may impact the stock market to go down in trend. Moreover it is intended to incorporate social contents from twitter for effective prediction. By this, we are able to bring profit to the investment made on a stock. Stock market is unpredictable, many factors affect the price of the stock so our project is able to predict with only a definite number of accuracy.

REFERENCES

- [1] "Stock Market Prediction Using Machine Learning Algorithms", K.Hiba Sadia, Aditya Sharma, Adarrsh Paul, Sarmistha Padhi, Saurav Sanya, April 2019
- [2] "Stock Price Prediction Using News Sentiment Analysis", Saloni Mohan, Sahitya Mullapudi, Sudheer Sammeta, Parag Vijayvergia and David C. Anastasiu, 2019
- [3] "Stock Price Prediction Through the Sentimental Analysis of News Articles", Jaeyoon Kim, Jangwon Seo, Minhyeok Lee, Junhee Seok
- [4] "Visual Interpretation of Text-based Deep Stock Prediction", Lei Shi, Zhiyang Teng, Le Wang, Yue Zhang, and Alexander Binde, June 2019
- [5] "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements", Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Pand and Babita Majhi, October 2016
- [6] "A Prediction Approach for Stock Market Volatility Based on Time Series Data.", Sheik Mohammad Idrees, M. Afsar Alam, and Parul Agarwal
- [7] "Using Sentimental Analysis in Prediction of Stock Market Investment", Sunil Kumar Khatri, Ayush Srivastava, (2016).
- [8] "Capital Market Forecasting By Using Sentimental Analysis.", Rohan Srivastava, Soumil Agarwal, Dhruv Garg, Jagdish Chandra Patni, (2016).
- [9] "Using Social Media Mining Technology to Assist in Price Prediction of Stock Market Sentimental analysis of Twitter data.", Yaojun Wang, Yaoqing Wang, 2016
- [10] "Stock Market Prediction Exploiting Microblog Sentiment Analysis", Bo Zhao, Yongji He, Chunfeng Yuan, and Yihua Huang
- [11] J. Kim, J. Seo, M. Lee and J. Seok, "Stock Price Prediction Through the Sentimental Analysis of News Articles," 2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), 2019, pp. 700-702, doi: 10.1109/ICUFN.2019.8806182.
- [12] Chao Liang, Linchun Tang, Yan Li, Yu Wei, Which sentiment index is more informative to forecast stock market volatility? Evidence from China, International Review of Financial Analysis, Volume 71, 2020, 101552, ISSN 1057-5219, <https://doi.org/10.1016/j.irfa.2020.101552>.
- [13] Francesco Audrino, Fabio Sigris, Daniele Ballinari, The impact of sentiment and attention measures on stock market volatility, International



Journal of Forecasting, Volume 36, Issue 2, 2020, Pages 334-357, ISSN 0169-2070, <https://doi.org/10.1016/j.ijforecast.2019.05.010>.

[14] Jiahong Li, Hui Bu and Junjie Wu, "Sentiment-aware stock market prediction: A deep learning method," 2017 International Conference on Service Systems and Service Management, 2017, pp. 1-6, doi: 10.1109/ICSSSM.2017.7996306.

[15] Nuno Oliveira, Paulo Cortez, Nelson Areal, The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices, Expert Systems with Applications, Volume 73, 2017, Pages 125-144, ISSN 09574174, <https://doi.org/10.1016/j.eswa.2016.12.036>.

[16] John Griffith, Mohammad Najand & Jiancheng Shen (2020) Emotions in the Stock Market, Journal of Behavioral Finance, 21:1, 42- 56, DOI: 10.1080/15427560.2019.1588275

[17] Xi Zhang, Yunjia Zhang, Senzhang Wang, Yuntao Yao, Binxing Fang, Philip S. Yu, Improving stock market prediction via heterogeneous information fusion, Knowledge-Based Systems, Volume 143, 2018, Pages 236- 247, ISSN 09507051, <https://doi.org/10.1016/j.knosys.2017.12.025>.