A Study of Stock Market Price using Sentimental Analysis on Banking Sector

Kalyan Sengupta
Professor in MarketinAnalytics
IFIM Busines
School,Bangalore
. kalyan.sengupta@ifim.edu.in

Aman Abhishek Kisku
PGDM Second Year
IFIM Business
School,Bangalore
aman.kisku@ifimbschool.com

Srujana R, Prakhar Singhal
PGDM Second Year
IFIM Business School, Bangalore
srujana.r@ifimbschool.com,
prakhar.singhal@ifimbschool.com

***Abstract*-Interpreting stock market has been an area of interest for some time now. Social media now a days represents the sentiment of public and opinion about ongoing events. Stock market interpretation using twitter on basis of sentiment of the public is an attractive research field. Sentiment analysis includes the process of extracting useful information from various books and social media platforms like Twitter, Facebook to understand about the emotions of the people about an event. The sentiment of the collected data can be analysed using various languages and tools like Python, R, KNIME, Orange. For any investor it is important to invest in a stock which would give decent return and that would perform consistently withstanding the market tension. Nowadays, with the availability of Data Science and interpretation techniques, this task can be automatically done, which will help the investors to make better investment decision. Many researchers have used Python to interpret the sentiment of the data. In this paper, we have used Orange to obtain the polarity scores which will used to interpret the sentiment of the data.**

Keywords - Sentiment, Twitter, Orange, Stocks, Analysis

1. INTRODUCTION

The number of social media users goes on increasing and due which the amount of data available has been increasing tremendously. Sentiment analysis or opinion mining can be used to interpret the mood of individuals, that may have an impact on stock costs, so it will facilitate in interpretation of actual stock movement. Most of those data are in the form of text from mail, chats, survey, articles, documents and social media. Going through all these data manually will be expensive and impossible to do within a short duration of time and there will be some errors. Sentiment Analysis will provide consistent and real time analysis within a short period of time. The data is being used in many data driven decision making like election, disaster detection (Sakaki,2010), opinion mining (Kouloumpis, Wilson & Moore, 2011) (Gokulakrishnan,Priyanthan, Ragavan & Prasath, 2012) (Khan, Bashir & Qamar, 2014). Share markets works on the principle of supply and demand, if the demand increases the share price also increases and vice versa. There are few theories which explain the movement of the stock market. Chartist theory (Cowles, 1933) is based on assumption that pattern in stock market is repetitive. The future stock returns can be predicted by using historical data. Random Walk theory (Fama,1995) considers that the variables that affect the stock price are identical and they are independent. The assumptions of these theories may be different, but many existing models use historical stock data like close price, open price, volume traded every day to predict the future stock price. Apart from historical stock prices, investor decisions are also based on current news and trends. It also depends on the media (Tetlock, 2007) (Fang & Peress, 2009) i.e., depends on the type in which the organization tries to project itself to the outside world. Corporates and stockholders are very much interested in knowing whether the stock prices will go up or down in a particular time frame. The traditional analysis of data takes weeks or sometimes even month to finish, but this analysis can be done within a very short duration.

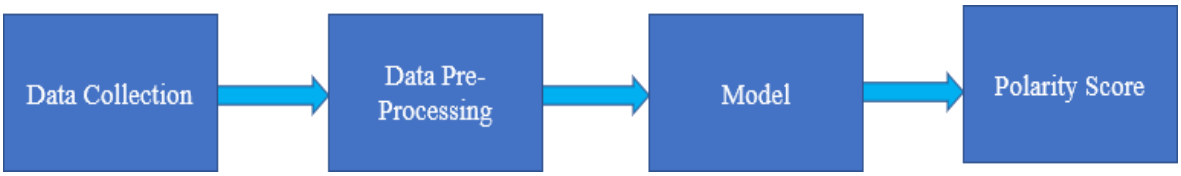

Figure 1. Analytical model of sentiment interpretation

Earlier there was no significant way to analyze the stock as it is non-linear, dynamic and doesn't follow specific rules, but now analysis of tweets from the user to ascertain the price of the stock is a new way of using machine learning to get the opinion of the people as they are very efficient. Many methods like technical analysis, basic analysis, statistical analysis and applied mathematics analysis are employed in a shot to investigate the shared trends within the market however up to this point none of those tools have proved to be a universal approach for acceptance as an interpretation tool. As Refs. (Tetlock, 2007) (Luo, Zhang & Duan, W., 2013) (Chan, 2012) indicated, the emotions or the sentiment of the investor will be reflected by the movement of the stock market. Pessimism or Negative Sentiment on social media might be reason for the fall in price. Positive sentiment will result in upward movement of the stock price. The tools which are used the analyse the sentiment can be divided into two categories: word count based analysis strategy and machine learning. As shown in figure 1 the real time twitter data is collected and processed to remove external URL's, stop word which will not add value to the analysis. In this model we have used VADER sentiment analysis (Valence Aware Dictionary and Sentiment Reasoner). It is a rule-based analysis tool which is mainly used to analysis sentiments which are expressed in social media. When the pre-processed data is given as input to this model it will process and give polarity scores as output.The rest of the paper is organized as follows. In section 2, we have described about various theories that are used to interpret the movement of stock market. In section 3 we have explained about the data collection method and how we estimated the polarity score. In section 5, we have stated the conclusion which we came up after the analysis.

2. LITERATURE REVIEW

Sentiment analysis has been so far used as method to handle the Natural Language's processing task at several levels of granular data. Recently, microblog like Twitter, has become the easier and faster medium for the users to post real time reactions and to spread their opinions regarding "everything". They use emoticons to express their views. There are few tweets where users use ":)" ": -)" to describe about a positive emotion about a particular stock or event and ":(" ":-(" to describe about a negative emotion and view about a stock or event.(Lie ,2012) Tweets consists not only texts but also too many acronyms, emoticons and inessential information like photos and URL's. Therefore, tweets should be pre-processed to represent correct emotions of public. For pre-processing of tweets there are three stages of filtering process. They are Tokenization, Stop words removal and regex matching for removing special characters. (Nagarajan,2010) Emotions are measure of subjective feelings and thoughts expressed by us. Significant study has been done on these emotions and studied in several fields like, psychology biology, philosophy, science etc. Still it's found by researchers don't seem to have consensus on basic set of emotions. Mainly, there exists six sets of emotions, namely anger, joy, surprise, love, unhappiness and concern. (Herer,2005) These emotions additionally can be further classified into their own subsets. Although emotions influence the strength of opinion and exhibit relationship with opinion, there's no equivalence between opinion and feeling. Therefore, once emotions or an opinion is mentioned, it's necessary to differentiate between these. As mentioned earlier, there are solely six styles of primary emotions; however, there are several other expressions of the language that may be used to convey emotions. In same way, there also are several opinion expressions used to justify positive sentiments or negative sentiments by users. Therefore, primarily sentiment analysis tries to conclude individual sentiments supported by expressions supplied with the assistance of any language used to portray their emotions. (Agarwal, Xie, Vovsha, Rambow & Passonneau,2011). Analysis of finances like, stock movement interpretation preponderantly depends on the thought of client which means that the rise or fall of the stock costs of a corporation principally depends on the impression of it on the general public mind.(Chen & Lazer, 2011) Unhealthy reviews usually decrease the recognition of the corporation name and a decent review brings an improved impact. (Bing,Li, Keith CC Chan & Carol,2014)(Mishne, 2006) An example to such hypothesis is that of Tesla, in which a positive review relating to launch of an advanced motor, its exchange price grew multiple fold. Multitude of researches have already proven that this knowledge has helped to make informed decision upon varied fields like polling results prediction etc. (Weiming Hu, Tieniu Tan, Liang Wang & Steve Maybank,2004) (Eiji Aramaki,Sachiko Maskawa & Mizuki Morita,2011). The fundamental idea behind this technique is to discover the mood of the customers and then predict on the parameter in subject. (Eiji Aramaki Sachiko Maskawa & Mizuki Morita, 2015).

Company fundamental theory states that the simplest method to create a call concerning investment within the securities market is to try and do analysis on the corporate itself.(Cootner, 1964) Even though there's no strict law of determination, this theory holds that an organization that's winning in its operations are going to be good investment in. The indicators like profit don't directly confirm stock worth, they will be a decent sign of that approach the stock goes i.e., high profit earning need not necessarily have high stock value.(Fama, 1995). Investor Psychology theory goes against the normal concept, investors know before time about the performance of the stock, they are informed and given full details about the performance of the stock and the company, and they behave more rationally.

Psychological theories (Daniel,Hirshleifer & Subrahmanyam,1998) of investment chalk up market movements in addition with interaction of psychological and social drives at play in every capitalist within the finance public generally. However, this theory doesn't discount any chance for predicting market behaviours. On the contrary, it merely says the market does not essentially behave normally in a manner that creates logical sense. This can be the sort of viewpoint that observes investors behaving with herd mentality. Stock costs will skyrocket or plummet with very little prompting aside from the perception of what others are doing. (Loewengstein, 2000)

.

3. METHODOLOGY

For analysing financial data, which consists of huge amount of streaming data in the form of event data (tweets in this case) the data is collected through Vicintas. Since the rate of incoming data varies with every tick of the clock, we collected data for a fixed time interval.Time-series information transverses over numerous areas, including financial markets, seismology, medicinal services, meteorology, and space science. Information visualization (InfoVis) and Visual Analytics (VA) methods are used to separate such time-arrangement datasets. They help us in better utilization of the human ability to expand understanding from discernments, including recognizing the basic examples, anomalies, and correlation, through a smart and undirected request. This interpreted information has been fused into time-arrangement depictions. Conventional approaches may not totally reinforce visual expectation in examples of complex multivariate datasets, for example securities trade, climate, and human services, since it is without considering the in between relationships. To help such examination, tweets from 25 Sep 2018 to 22 July 2019 on ICICI and Canara Bank are extracted from twitter. Each individual data collected represents the state or mood of the user regarding a theme. This can be fetched through an elemental HTTP authentication and a twitter account. Figure 2 shows the twitter authentication process and pre-processing of tweets is shown in figure 3.

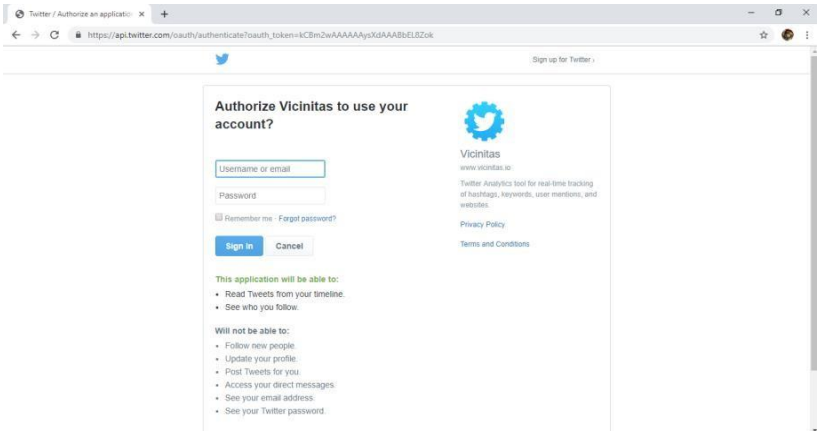

Figure 2. Twitter Authentication Process

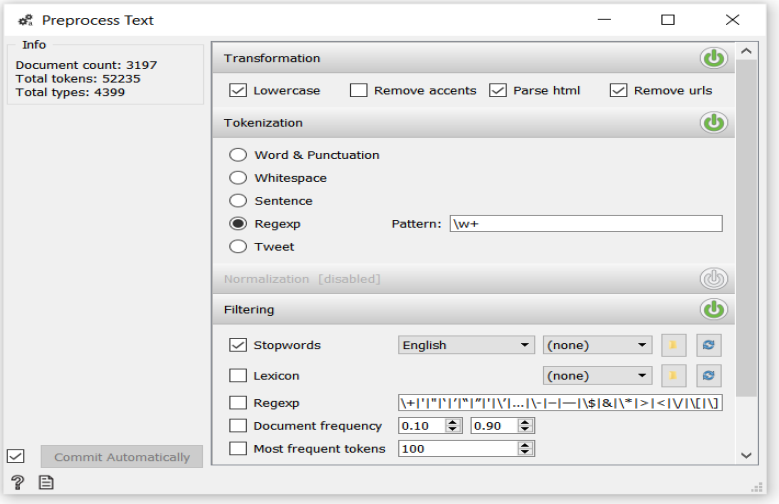

Figure 3. Pre-Processing of Tweets

The tweets were collected and filtered using keywords like Canara bank tweets, Icicibank. It consists of tweets from the bank side about products and services offered by them which would have a significant impact and are worth studying. Stock opening and closing prices of Canara Bank and ICICI Bank are obtained from the BSE website which will compared with the polarity score. Here, in this paper, we are focusing about the financial data analysis, a website known as Vincitas has been used to collect tweets. Each tweet represents the new service and the product about the bank and information about any vital investments or collaborations. After retrieving all the data for every tick of data, analysis of the sentiments related to each tweet is initiated and thus the mood which has a direct implication on the status of the stock is being tracked. Sentiment analysis is basically a classification problem in which the input data having a positive or a negative sentiment are classified. The collected tweets have time stamp attached to it, it also has information about the number of retweets, number of favourites. After the data collection it is processed using Orange to remove irrelevant information like http links from the tweets. Orange is an open-source, fault tolerant model that can be scaled up dependent on necessity with a guarantee of lower latency and robustness.

The pre-processed data is then used to analyse the sentiment of the stock which has a direct effect on the stock status which is being interpreted. The pre-processed data is then given to model and the polarity score is obtained. The polarity score that is obtained is day wise to make the analysis easier weak wise polarity score is calculated. The model gives us the positive, negative and the neutral score for the text. The score that we have initially obtained in tweet wise score. Figure 4 shows the polarity score estimation.

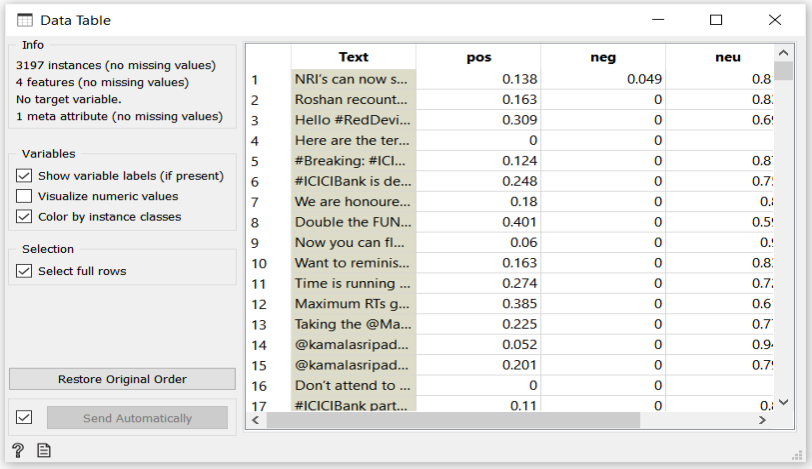

Figure 4. Polarity Score Estimation

The next process is it to find the weekly average polarity score for the tweets. After the average score is obtained it is compared with the closing price to find if there is any relationship between them. Table 1 represents the correlation between the sentiment and the stock price with lag effect. The lag effect is estimated to see if there is any evidential relationship between the stock price and the estimated sentiment scores.

Table 1. Lag Effect

Lag	ICICI	SYNDICATE	HDFC	CANARA	UNION
No	0.299223	0.055873	-0.14538	-0.17146	-0.06502
1	0.264975	0.195351	-0.1612	-0.06617	0.05138
2	0.29655	0.121034	-0.16886	0.052711	0.01458
3	0.320188	-.0.00371	-0.135	0.089993	0.03146
4	0.331622	-0.04716	-0.14593	0.18017	0.00627
5	0.346168	-0.017042	-0.27599	0.335792	-0.01287

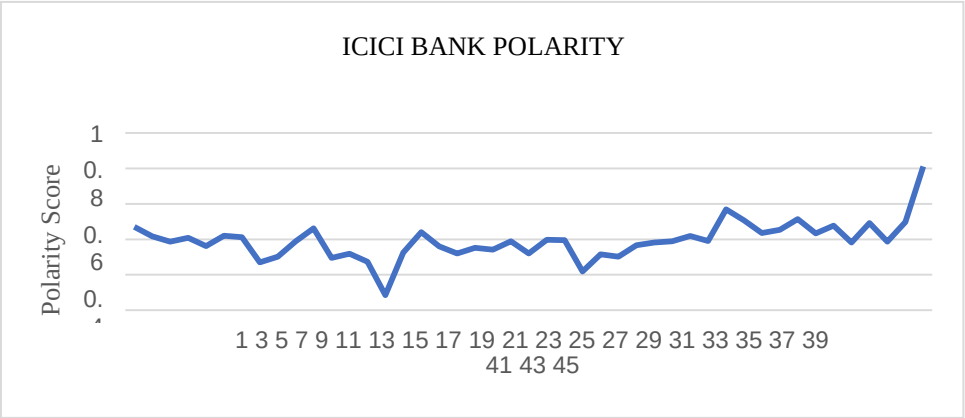

Figure 5. Average Polarity Score of ICICI

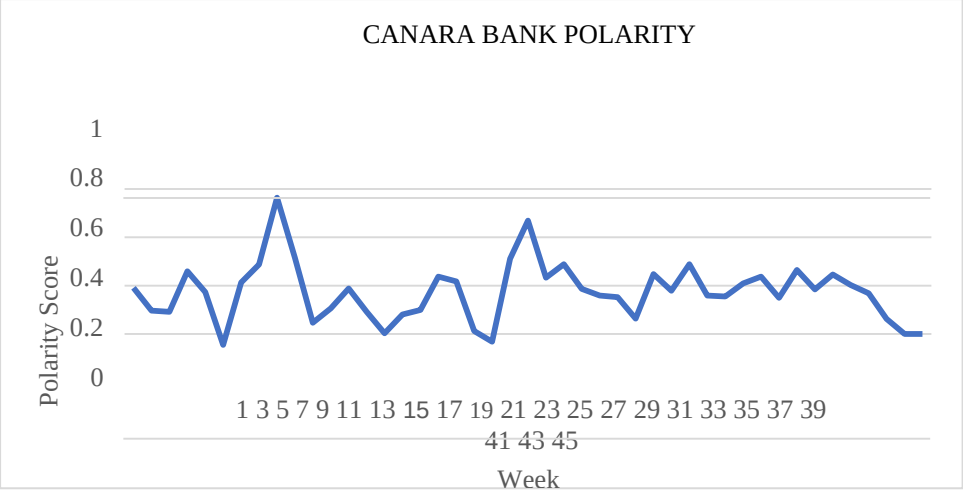

Figure 6. Average Polarity Score of CANARA

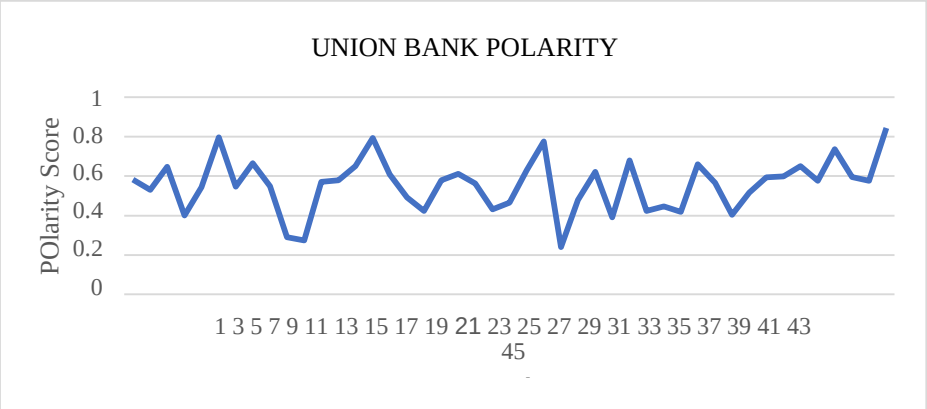

Figure 7. Average Polarity Score of UNION

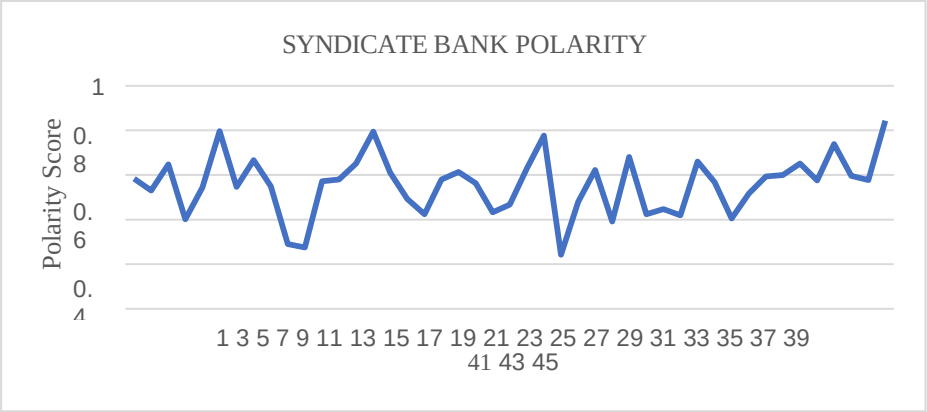

Figure 8. Average Polarity Score of SYNDICATE

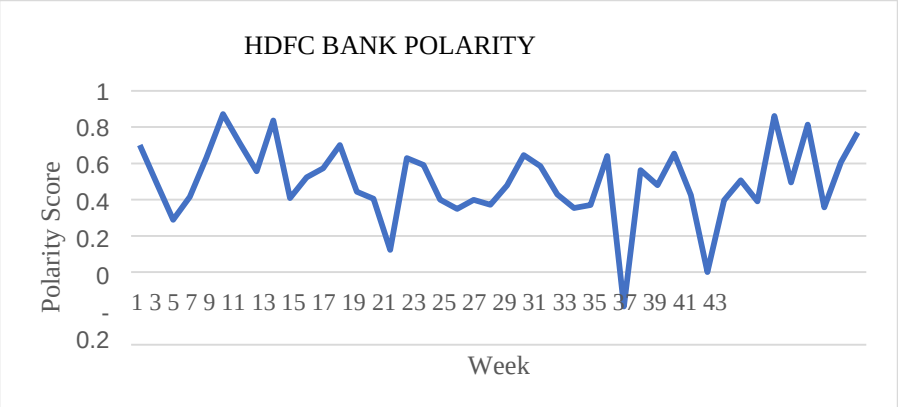

Figure 9. Average Sentiment Score of HDFC

Figure 5,6,7,8 and 9 represents the average polarity score of that week for different banks. This clearly represents that there is change in sentiment based on the tweet by the organization.

4. CONCLUSION

In this paper we have analyzed the tweets which were posted by the organization. From the analysis we were able to find that there is no significant relationship between the sentiment of the tweets posted by the bank and their stock price. As correlation is small, effect is minimal, but it is observed that correlation between sentiment value and closing price is increasing when increasing the number of lags in case of ICICI and Canara Bank whereas as no correlation is found between Syndicate, HDFC and Union Banks as the correlation is moving randomly.

This make us believe that the effect of sentiments on ICICI and Canara bank is higher than that of HDFC, Syndicate and Union Bank. The reason behind this could be that ICICI and Canara banks investors are directly affected by twitter news and speculations in the market as most of them are from high income group and are active on social media. People investing in HDFC, Syndicate and Union Bank are different than that of ICICI Bank, as many of these Banks investors don't bother about what's happening on social media or Twitter as the correlation for them is moving in Zig -Zag motion without certifying any correlation between Sentiment value and stock prices. From this, we can say that these Banks (HDFC, Syndicate and Union) does not have a good customer base of high-income groups or people who are active on social media. Thus, the revenue generation is from low income groups which do not increase the bottom line of these banks and hence the growth is slow.

Canara

2-day strike from 26 September 2018 leading to sentiment downfall. (week 40). On 23rd December Canara PO results came which increased sentiments (week 44). Nationwide banks strike from Jan 8 to 9 leading to sentiment downfall in Jan (week 2). Canara Bank reported a 152.5 per cent surge in quarterly net profit (week 21) which increased sentiments value. Canara Bank consolidated profit rose by 23.43% in June Quarter. (week 32) which increased sentiments.

ICICI

Net NPA ratio decreased from 3.65% at September 30, 2018 to 2.58% at December 31, 2018 - the lowest in the last 12 quarters that's why, there is continuous downfall in sentiment value from week 1 to 7. Core operating profit (profit before provisions and tax, excluding treasury income) grew by 14% year-on-year to ₹ 5,667 crore (US$ 812 million) in the quarter ended December 31, 2018 (Q3-2019) therefore there is a increase in sentiment value in week 17. Net NPA ratio decreased from 4.19% in June 2018 to 1.77% in June 2019, the lowest in 14 quarters reflecting the buildup of provisions by the bank over the last year and a slower addition to NPAs. Reflecting increase in sentiment scores from week 42 to 44.

Syndicate Bank

A monetary penalty of Rs 75 lakh has been imposed on Syndicate Bank for non-compliance with the directions issued by RBI on frauds classification and reporting; innovative housing loan products wherein upfront disbursal of loans is done leading to sentiment score going down in December (week 11). Syndicate Bank reported a profit of Rs 128.02 crore for the March quarter therefore high sentiment values are visible from week 14 to 24. Syndicate Bank had posted a net loss of Rs 1,2821.77 crore in the same April-June Quarter of the previous fiscal ended March 2019 thus taking sentiment score down in week 26. On August 30, *2019*, the Government of India announced the bank would be merged into Canara Bank. Thus, increasing sentiment score in weeks from 43 to 45.

HDFC Bank

HDFC Bank hit a 52-week low of Rs 1884.40 on 31 Oct 2018 leading to lower sentiment value in weeks 6 to 7. HDFC Bank hit a 52-week high of Rs 2502.90 on 3rd July 2019 leading to higher sentiment scores in week 38. Net profit of HDFC Bank rose 22.6% to Rs 5,885.12 crore on a 22.1% increase in total income to Rs 31,204.46 crore in Q4 March 2019 over Q4 March 2018 thus leading to higher sentiment score in week 29.

For future work, analysis of the tweets from the users about the organization can be collected and the analysis can be done to check if there is any correlation between those tweets and the stock price.

REFERENCES

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011, June). Sentiment analysis of twitter data. In Proceedings of the Workshop on Language in Social Media (LSM 2011) (pp. 30-38).

Aibek Makazhanov, Davood Rafiei, and Muhammad Waqar. Predicting political preference of twitter users. Social Network Analysis and Mining, 4(1):193, 2014

Bing, Li, Keith CC Chan, and Carol Ou. "Public sentiment analysis in Twitter data for prediction of a company's stock price movements." e- Business Engineering (ICEBE), 2014 IEEE 11th International Conference on. IEEE, 2014.

Carl Sabottke, Octavian Suciu, and Tudor Dumitras. Vulnerability disclosure in the age of social media: Exploiting twitter for predicting real-world exploits. In USENIX Security Symposium, pages 1041–1056, 2015

Chen, R., Lazer, M., Sentiment Analysis of Twitter Feeds for the Prediction of Stock Market Movement, Cs 229, pp. 15, 2011.

Cootner, P. H., "The Random Character of Stock Market," The Journal of Business. (1964)

Cowles 3rd, A., "Can stock market forecasters forecast?" Econom. J. Econom. Soc. 1 (1933) 309-324.

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. the Journal of Finance, 53(6), 1839-1885.

De Rijke M. Mishne, G. Capturing global mood levels using blog posts. AAAI spring symposium:computational approaches to analyzing weblogs, 6:145–152, 2006.

Eiji Aramaki, Sachiko Maskawa, and Mizuki Morita. Twitter catches the flu: detecting influenza epidemics using twitter. In Proceedings of the conference on empirical methods in natural language processing, pages 1568–1576. Association for Computational Linguistics, 2011

Fama, E.F., Random walks in stock market prices, Finan. Anal. J. 51 (1) (1995) 75–80

Fang, L., Peress, J., Media coverage and the cross-section of stock returns, J. Finance 64 (5) (2009) 2023–2052.

Gokulakrishnan, B., Priyanthan, P., Ragavan, T., Prasath, N., Perera, A., Opinion mining and sentiment analysis on a Twitter data stream, in: Proceedings of the International Conference on Advances in ICT for Emerging Regions (ICTer), IEEE, 2012, pp. 182–188.

Kouloumpis, E., Wilson, T., Moore, J.D., Twitter sentiment analysis: the good the bad and the omg! 11 (538– 541) (2011) 164.

Khan, F.H., Bashir, S., Qamar, U., TOM: Twitter opinion mining framework using hybrid classification scheme, Decis. Support Syst. 57 (2014) 245–257.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. American economic review, 90(2), 426-432.

Luo, X., Zhang, J., Duan, W., Social media and firm equity value, Inf. Syst. Res. 24 (1) (2013) 146–163.

Nagarajan, B. M. (2010). Understanding user-generated content on social media (Doctoral dissertation, Wright State University).

Sakaki, T., Okazaki, M., Matsuo, Y., Earthquake shakes Twitter users: real-time event detection by social sensors, in: Proceedings of the Nineteenth International Conference on World Wide Web, ACM, 2010, pp. 851–860.

Scherer, K. R. (2005). What are emotions? And how can they be measured? Social science information, 44(4), 695-729.

Tetlock, P.C., Giving content to investor sentiment: the role of media in the stock market, J. Finance 62 (3) (2007) 1139–1168.

Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M., Predicting elections with Twitter: what 140 characters reveal about political sentiment 10 (1) (2010) 178–185.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., Narayanan, S., A system for real-time Twitter sentiment analysis of 2012 us presidential election cycle, in: Proceedings of the ACL System Demonstrations, Association for Computational Linguistics, 2012, pp. 115–120.

Weiming Hu, Tieniu Tan, Liang Wang, and Steve Maybank. A survey on visual surveillance of object motion and behaviors. IEEE Transac-tions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34(3):334–352, 2004.