

Study on the Relationship between Daily News and Stock Market Performance

Atul Singh¹, Aakash Patel², Arpit Sah³, Mitanshu Khurana⁴

¹⁻⁴Computer Engineering Department of MPSTME, NMIMS, V. L. Mehta Road, Vile Parle, 400056 Mumbai, India.

Abstract - The overall purpose of the study is to determine the sentiment of daily news and its effect on the fluctuations (gain/loss) in stock market. The basic design of the study can be subdivided into three major steps. Data collection being the first step, involves collection of daily news headlines from the popular source of Times of India archives using the News API and web scraping as well. Data cleaning being the second step involved the use of Python to remove the unnecessary as well as the redundant content to obtain the appropriate content. Sentiment analysis follows as the third step involves the use of VADER which is specifically suited for the content of social media. After generating the sentiment score for each news, we move further to determine the correlation of the news sentiment and the daily Sensex change. We end up with the conclusion that daily news sentiment has a subsequent effect on the BSE Sensex and the correlation between the two is found to be a positive correlation of 12.607%. Sentiment analysis of daily news headlines can thus have said to be sufficiently correlated with the gain or loss in the stock market and the same information can be used further to make suitable decisions. Further improvements can be possible such as sentiment analysis to be more specific to financial news, extraction of data from different sources which may also make the prediction of behavior of stock market possible.

Key Words: Sentiment Analysis, VADER Lexicon, News-Headlines, SENSEX, NLP

1. INTRODUCTION

Predicting the movement of stocks in the dreams of making big money has always fascinated people. It brings together various fields of study from statistics to psychology. With the advent of data science, the task of analyzing, predicting, and mining information. It is natural for people to keep attempting to predict the stock market, no matter how complex it may seem. Sentiment of people has been said to affect several various aspects in business, and hence can have some corresponding effect on the stock market. Sentiment Analysis or Opinion Mining is the computational study of people's opinions, attitudes, and emotions toward an entity. The entity can represent individuals, events, or topics. In essence, it is the process of determining the emotional tone behind a series of words, used to gain an understanding of the attitudes, opinions and emotions expressed within an online mention. The benefits of sentiment analysis can be leveraged to determine the sentiment of daily news and put it to further use.

In this study, the polarity of daily news headline was calculated. Subsequently the sentiment of news for the whole day were calculated and the polarity score of the whole day was derived. This polarity score was the used to find the correlation between the news sentiment and the movement of the stock market.

The dataset used in the research was gathered using NewsAPI [1] this API provides the news articles of different newspapers as per their date of release. The news has been collected from Times of India. The sentiment analysis tool VADER [2] was used to analyze each news headline and calculate the overall daily news sentiment score.

The daily news sentiment score is then used to find correlation between daily news sentiment and the gain/loss in stock market. We have also calculated the lag effect i.e. how the past news affects the present Sensex change. We have considered lag effect of up to 5 days.

1.1 Literature Review

A great amount of literature is present in the field of sentiment analysis for large-scale sentiment of news and blogs [3]. Extensive Research have been conducted in predicting the effect of news on stock returns using Thomson Reuters neural network analysis [4] which proves that positive news stories have an immediate response whereas negative news tends to show a more prolonged and delayed response. Studies have addressed the issue of assessing the stock market and investor sentiment response on Earnings announcement [5].

Mutz, Diana C., and Joe Soss (1997) in their paper "Reading Public Opinion: The Influence of News Coverage on

[1] <https://newsapi.org/>

[2] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

[3] Godbole, M. Srinivasaiah, and S. Skiena, "Large scale sentiment analysis for news and blogs," in Proc. ICWSM, 2007.

[4] S. L. Heston and N. R. Sinha, "News versus sentiment: predicting stock returns from news stories," Finance and Economics Discussion Series, 2016.

[5] Y. Jiang, "Investor sentiment and stock market response to earnings announcement," in Proc. International Conference on Management and Service Science, 2011

Perceptions of Public Sentiment” [6] studied the effects of a news agenda on the actual opinions and views of the public in common by making use of a quasi-experimental design and came to a conclusion that the strategy is extremely limited, but still believe it may have some effects on citizens’ perceptions regarding issues of community as a whole.

Mark R. Joslyn (1997) [7] in his paper examines the impact of public views in individual opinion by taking public influence to be considerable yet keeping in the mind the magnitude, direction and significance of the effect on the target of public sentiment and the tendency of the perceiver. The paper makes use of a theoretical model to examine the effects of complex relationship between opinion context and individual judgement.

Keith Cortis et al. (2017) in “Fine-Grained Sentiment Analysis on Financial Microblogs and News” [8] studied how sentiments contained in financial microblogs and news could be used to predict the sentiment score of stock/companies. It used various techniques that could be used and compared all these techniques and algorithms. The scores were between -1 and 1 and mentioned that Deep Learning and traditional ML techniques provided the maximum contributions. Review of all the techniques used was done and one of the best methods was the hybrid (DL, Lexicon) technique.

Max Sorto et al. (2017) in their paper “Feeling The Stock Market: A Study in the Prediction of Financial Markets Based on News Sentiment” [9] used the news articles from Wall Street Journal and financial market data from the NASDAQ to find the polarity of news. The sentiment analysis system used a summarization algorithm and the SenticNet 4.0 API. A modified version of the system was also tested using only news headlines, which omitted the summarization portion of the original system and found that both news headline and the news article both give.

Gang Li-Fei Liu (2012) in “Application of a clustering method on sentiment analysis” [10] found that traditional clustering methods can also be effective in sentiment analysis. It used tf-

idf weighting method, voting mechanism, and importing term scores to improve the accuracy for the clustering. It has proved to be an efficient and non-human intervening way to solve sentiment analysis problem.

Sheikh Shaugat Abdullah et al. (2013) in their paper “Analysis of stock market using text mining and natural language processing.” [11] used a framework that uses our text parser and analyzer algorithm with an open source natural language processing tool to analyze, retrieve, and forecast investment decisions from any text data source on stock market. They used the data of Dhaka Stock Exchange (DSE), capital market of Bangladesh for their study.

Y. Kim et al. (2014) in their paper “Text opinion mining to analyze news for stock market prediction” [12] have successfully implemented Natural Language Processing methodology in mining text opinions and unstructured big data to predict the rise and fall of KOSPI (Korean Composite Stock Price Index). This paper also proposes the use of Natural Language Processing methodology to infer the polarity of news articles.

Filipe Ribeiro et al. (2016) in their paper “SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods” [13] evaluated twenty-four popular sentiment analysis methods on a benchmark of eighteen labelled datasets, covering messages posted on social networks, movie and product reviews, as well as opinions and comments in news articles.

They found that the selection of sentiment analysis tool depends on the dataset that we are work on. Another thing that comes into play while deciding the tool is a good trade-off between prediction performance and coverage.

Diana C. Mutz and Joe Soss (1997) in his paper “Reading Public Opinion: The Influence of News Coverage on Perceptions of Public Sentiment” [14] finds that news coverage does affect the public sentiment.

Influenced by the study of Diana C. Mutz and Joe Soss in this paper we will find out the relationship between daily news headlines and the stock market performance, this paper concludes that both news headlines and news article will give similar results. Keeping study of Filipe Ribeiro et al. (2016) [11] into consideration we will be using VADER [15] for finding

[6] Mutz, Diana C., and Joe Soss. “Reading Public Opinion: The Influence of News Coverage on Perceptions of Public Sentiment.” *The Public Opinion Quarterly*, vol. 61, no. 3, 1997

[7] Joslyn, M.R. *Political Behavior* (1997) 19: 337. <https://doi.org/10.1023/A:1024847805976>

[8] “Fine-Grained Sentiment Analysis on Financial Microblogs and News” Keith Cortis-André Freitas-Tobias Daudert-Manuela Huerlimann-Manel Zarrouk-Siegfried Handschuh-Brian Davis - Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) – 2017

[9] Sorto, Max; Aasheim, Cheryl; and Wimmer, Hayden, “Feeling The Stock Market: A Study in the Prediction of Financial Markets Based on News Sentiment” (2017). SAIS 2017 Proceedings. 30. <https://aisel.aisnet.org/sais2017/30>

[10] “Application of a clustering method on sentiment analysis” Gang Li-Fei Liu - *Journal of Information Science* – 2012

[11] Shaugat Abdullah, Sheikh & Rahaman, Mohammad & Rahman, Mohammad. (2013). Analysis of stock market using text mining and natural language processing. 1-6. 10.1109/ICIEV.2013.6572673.

[12] Y. Kim, S. R. Jeong, and I. Ghani, 2014, “Text opinion mining to analyze news for stock market prediction,” *Int. J. Advance. Soft Comput. Appl.*, vol. 6, no. 1, March 2014.

[13] “SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods” Filipe Ribeiro-Matheus Araújo-Pollyanna Gonçalves-Marcos Gonçalves-Fabrcio Benevenuto - *EPJ Data Science* – 2016

[14] Diana C. Mutz and Joe Soss *The Public Opinion Quarterly* Vol. 61, No. 3 (Autumn, 1997), pp. 431-451

[15] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media

out the sentiment polarity of news. Using the polarity score to measure the correlation with the stock market performance.

2. Data Collection, Methodology and Analysis

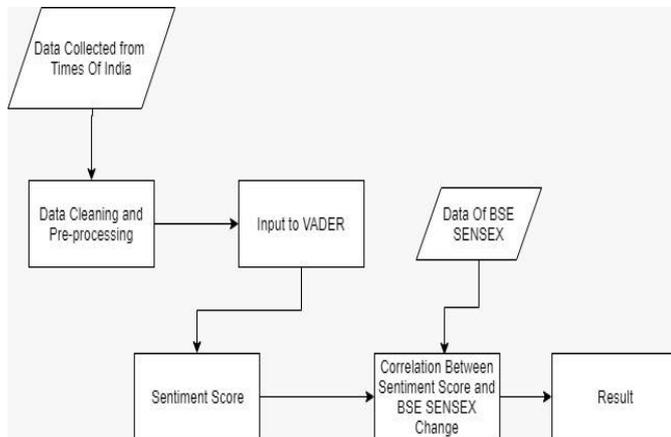


Fig -1: Flowchart of the Research

The research can be broken down into three major parts – Data Collection and Cleaning, Sentiment Analysis and Analysis of sentiment with the Sensex.

For Data Collection we have collected daily news headlines of The Times Of India from News API, this API crawls the websites of those newspapers and gives us all the details of the news like date, title, description etc. in JSON format. The output was then cleaned using Python which involves removing of unrequired words duplicate news which gets along with the scraped data. After collection and cleaning we get the required news title on which we are going to apply the sentiment analysis.

We have used VADER for finding the sentiment of any news headlines, which is specifically attuned to sentiments expressed in social media. It is a lexicon and mainly has a dictionary of good and bad words and the rest of the words which are out of its dictionary are classified as neutral.

We found that there were many missing words in VADER's dictionary especially proper nouns which carried a sentiment with themselves. These words should be considered when calculating the sentiment score. To tackle the problem we executed a query on Twitter API to find out the tweets relating to the mentioned entity for the past week. The tweets carry public sentiments of that entity. Each tweet was passed through VADER as a parameter and an average sentiment score was calculated. The sentiment score was added to VADER's dictionary.

By leveraging Twitter API, we were able to calculate the dynamic sentiment of the entities. This improved the sentiment that we got from the news made the sentiment score more accurate.

After generating the sentiment score for each news, average sentiment for the total day was calculated.

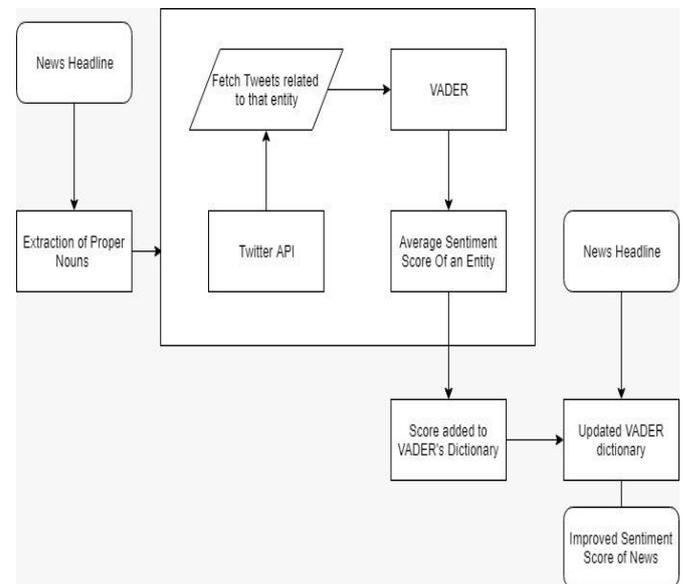


Fig -2: Process of finding the Sentiment Scores.

After getting the sentiment scores, the next step was to find the correlation between the gain percentage in Sensex and news sentiment. For this, we got the Sensex data from 11-Dec-2018 to 19-Mar-2019 from BSE website and calculated the gain percentage of each day using Excel. We also had to consider the lag effect of news on the gain or loss in Sensex. So, we decided that we will consider lag effects of news, up to 5 days.

But before proceeding to any kind of analysis, we were required to check whether the data we got now is stationary or not. So, we plotted the line chart for the Sensex gain percentage for the given dates in Excel and found out that our data is stationary.

To back our claim, we also checked the ACF and PACF graphs in SPSS, and again we could see that our data is stationary.

The next thing now we did was to find the correlation, which was done using Excel.

3. CONCLUSIONS

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.2320							
R Square	0.0538							
Adjusted R Square	0.0393							
Standard Error	0.6958							
Observations	67							
<i>ANOVA</i>								
	df	SS	MS	F	Significance F			
Regression	1	1.79053	1.79053	3.69838	0.05885			
Residual	65	31.46916	0.48414					
Total	66	33.25969						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 94.0%	Upper 94.0%
Intercept	0.24644	0.10334	2.38482	0.02001	0.04006	0.45282	0.04863	0.44425
X Variable 1	2.48648	1.29294	1.92312	0.05885	-0.09571	5.06866	0.01154	4.96142

Fig -3: Statistical Tests Outcomes

Considering the Sensex change in % from 11-Dec-2018 to 19-Mar-2019, we found that Sensex gain and the daily news sentiment are positively correlated with a correlation coefficient of 0.2320 with 94% significance level, and hence confirms that there is a significant linear relationship between the two variables.

Considering the lag effect of one day, Sensex gain and daily news sentiment are negatively correlated with a correlation of 0.11921. Considering the lag effect of two days, Sensex gain and daily news sentiment are negatively correlated with a correlation of 0.09284. Considering the lag effect of three days, Sensex gain and daily news sentiment are positively correlated with a correlation of 0.08416. Considering the lag effect of four days, Sensex gain and daily news sentiment are positively correlated with a correlation of 0.12558. Considering the lag effect of five days, Sensex gain and daily news sentiment are positively correlated with a correlation of 0.09378. We used VADER as a tool for quantifying the sentiment of daily news. By leveraging Twitter API, we were able to find the dynamic sentiment of any entity (proper nouns) in each news headlines. Using the new compound sentiment score of the sentence, a correlation between the news sentiment and BSE SENSEX was calculated which came out to be a positive correlation of 0.2320 with 94% significance level. We also found the lag effect of the news and we found that there exists a correlation between past news and present Sensex gain. Hence, we can conclude that, the news headlines were affecting BSE SENSEX. We can also conclude that these results may be more accurate if data is cleaner. Also, data could be collected from many sources, instead of just one. We can do this analysis on just the financial news headlines, instead of general news headlines to see improvement in results.

REFERENCES

- [1] N. Godbole, M. Srinivasaiah, and S. Skiena, "Large scale sentiment analysis for news and blogs," in Proc. ICWSM, 2007.
- [2] S. L. Heston and N. R. Sinha, "News versus sentiment: predicting stock returns from news stories," Finance and Economics Discussion Series, 2016.
- [3] Y. Jiang, "Investor sentiment and stock market response to earnings announcement," in Proc. International Conference on Management and Service Science, 2011, pp. 1-4
- [4] Mutz, Diana C., and Joe Soss. "Reading Public Opinion: The Influence of News Coverage on Perceptions of Public Sentiment." The Public Opinion Quarterly, vol. 61, no. 3, 1997, pp. 431-451.
- [5] Joslyn, M.R. Political Behavior (1997) 19: 337. <https://doi.org/10.1023/A:1024847805976>
- [6] "Fine-Grained Sentiment Analysis on Financial Microblogs and News" Keith Cortis-André Freitas-Tobias Daudert-Manuela Huerlimann-Manel Zarrouk-Siegfried Handschuh-Brian Davis - Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) - 2017
- [7] Sorto, Max; Aasheim, Cheryl; and Wimmer, Hayden, "Feeling the Stock Market: A Study in the Prediction of Financial Markets Based on News Sentiment" (2017). SAIS 2017 Proceedings. 30. <https://aisel.aisnet.org/sais2017/30>
- [8] "Application of a clustering method on sentiment analysis" Gang Li-Fei Liu - Journal of Information Science - 2012
- [9] Shaugat Abdullah, Sheikh & Rahaman, Mohammad & Rahman, Mohammad. (2013). Analysis of stock market using text mining and natural language processing. 1-6. 10.1109/ICIEV.2013.6572673.
- [10] Y. Kim, S. R. Jeong, and I. Ghani, 2014, "Text opinion mining to analyze news for stock market prediction," Int. J. Advance. Soft Comput. Appl., vol. 6, no. 1, March 2014.
- [11] "SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods" Filipe Ribeiro, Matheus Araújo, Pollyanna Gonçalves, Marcos Gonçalves, Fabrício Benevenuto - EPJ Data Science - 2016
- [12] E. Cambria, S. Poria, R. Bajpai, and B. Schuller. SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. In: COLING, pp. 2666-2677, Osaka (2016)
- [13] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.
- [14] <https://textblob.readthedocs.io/en/dev/index.html>