

Sentiment Analysis: Gauging the Effect of News on Stock Prices in Indian Stock Market

Nikhil Vijay, Snimarpal Singh, and Gyanesh Malhotra

Abstract—This paper analyses the relationship between the financial news reported in the mass media and its effects on the stock market returns. We have devised an algorithm that successfully determines polarity in financial texts using web scrapping and machine learning techniques in python to gather news texts from major financial news websites like Economic Times, Money Control, Reuters, The Hindu etc. The data collected is divided into 3 sentiments namely positive, negative and compound. By correlating the closing prices of leading players of the major sectors (like ITC-FMCG, TCS-IT, HDFC-Banking, ONGC-Energy, SunPharma-Pharmaceuticals, Maruti Suzuki-Automobile and SBI-Banking) with the sentiment generated from our algorithm we were able to conclude that indeed stock market returns are largely affected with the news and any significant news or shocks about a company can affect its stock price in the short run even though the news doesn't have any bearing on the company's fundamentals. Algorithms like the one used in this paper can be used to predict stock market returns in the event of any news by extrapolating the results of market behavior from the past.

Index Terms—Financial markets, python, web scrapping, sentiment.

I. INTRODUCTION

Researchers have been always keen on understanding the importance of sentiment in financial markets. The big question that arises is that is if there exists an algorithm that can classify the texts as negative or positive and understand the emotions attached to it. The seminal work in sentiment analysis had used movie reviews in training an algorithm to detect sentiment in text [1]. However, it is easier to train algorithm on movie reviews as they are numerically rated and express opinions explicitly.

This paper focuses on another set of data i.e. financial news. Investors receive a variety of news positive or negative regarding various companies on an everyday basis. The information conveyed in such news affect the decisions of various investors differently. These decisions reflect their trading patterns via changes in stock price in the financial markets.

It is via this process that a subjective and a verbal form of information gets converted to a numerical form which can then, in turn, be easily summarized, evaluated and compared.

Different people may have varied opinions regarding news but the investor opinions move in tandem on the question of markets returns. So, analyzing the effect of a news text on the stock market returns of the concerned entity addressed in the news allows one to numerically categorize the news texts as good or bad.

However, there is a problem associated with classifying financial news, the data associated with it is much noisier making it harder to classify as compared to that of movie reviews. In case of movie reviews, the critic has full control over the classification - be it positive or negative, whereas that is not the case for financial news wherein the news is reported by a reporter which in turn is interpreted by different individuals based on their own rationale. The impact of financial news texts depends upon both the individual behavior as well as the overall behavior of the market (herd mentality) thus making it difficult to predict.

The core aim of this paper is to determine how can we still use these financial news texts to train algorithms in order to predict market sentiment and also to determine the sensitivity of each major sector to the online published news.

The further scope of this research could be to determine whether the same algorithm used to classify text, be used to predict future stock returns based on sentiment generated by similar news in previous years. Previous works has already shown the possibility of using such sentiment detection algorithm to generate positive returns provided one trades before the news are announced [2].

However, practically it becomes unfeasible after accounting for trading costs and learnings from the efficient market hypothesis.

II. LITERATURE REVIEW

There have been various studies in the field of sentiment analysis that analyzed large-scale sentiment for news and blogs [3] and have also developed algorithms to classify sentiment as positive and negative. There has been extensive research in predicting the effect of news on stock returns using Thompson Reuters neural network analysis [4] wherein it was demonstrated that positive news like good earnings announcement have an immediate response whereas negative news tends to show a more prolonged and delayed response. Some seminal papers addressed the issue of assessing the stock market and investor sentiment response on Earnings announcement [5]. This research not only helped us to understand the rationality of stock market response but also the mispricing in the stock markets led by investor-driven sentiment.

Important sentiment analysis models like VADER have

Manuscript received June 18, 2018; revised August 7, 2018.

The authors are with Birla Institute of Technology and Science, Pilani, Goa, 403726, Indian (e-mail: f20130606@goa.bits-pilani.ac.in, f20140646@goa.bits-pilani.ac.in, f20140495@goa.bits-pilani.ac.in).

been used in this paper to access the sentiments of financial texts. The model has already been proven to be viable in previous studies wherein they assessed the sentiment of tweets with VADER outperforming individual human raters, as well as its generality across various contexts is also emphasized [6].

This paper also implies the use of Natural Language Processing methodology which previously had been successfully implemented in mining text opinions and unstructured big data to predict the rise and fall of KOSPI (Korean Composite Stock Price Index) [7].

Similarly, various other researches depicted that the method of combining various classifiers using supervised learning and machine learning could create hybrid classifiers which are better in terms of both precision and effectiveness in predicting stock market returns based on market sentiment [8]. Other seminal works on market sentiment emphasized that media coverage can exacerbate investor biases leading high return predictability for the firms in the spotlight of public attention [9], thereby lending more credibility to the further research conducted in this paper.

There are two important works that have influenced this study. First one in the field of natural language processing by [10], [11] where decision trees and support vector machines have been used to map the words in a text to its classification. The second work includes two seminal papers which highlighted the importance of sentiment in predicting stock market returns [12].

The main original contribution of this paper is to combine these two approaches in a single approach. This paper expands on the work done previously in the context of U.S. markets to Indian Financial Markets. In particular, we have focused on stock market returns from some major players within the various sectors like FMCG, Pharma, Energy, IT, Banking etc. Using the market leaders, in each sector as a proxy for the whole sector we have also tried to conclude about the sensitivity of each sector to the news.

III. OBJECTIVE

- To analyze the effect of news on stock market returns.
- To identify which sector is most affected and which is least affected by the news, by correlating the sentiment (both positive and negative) with log of closing prices and further suggesting trading strategies accordingly.

IV. DATA COLLECTION

Data collection was done for industry leaders of major sectors like FMCG, automobiles, pharmaceuticals, banking, etc. The data was collected from six websites namely *economictimes.com*, *thehindubusinessline.com*, *moneycontrol.com*, *the hindu.com*, *ndtv.com*, *reuters.com* for a period of six years from 2016-2011. Although the data for the years 2011 and 2012 was very sparse, enough data was collected to generate a sentiment score. The data were collected from the archives of these websites as well from Google search based on their relevance. Stock prices data was collected from yahoo finance for the same period.

V. METHODOLOGY

For the purpose of collecting data, various scrapping tools like *xpath*, *beautiful Soup*, *selenium webdriver*, *scrapy* etc. were used to get data from Google search results as well as archives of aforementioned websites. All these are libraries available in python (a programming language). Specifically, *python3* was used for scraping data. Daily news data was collected and filtered using various techniques to remove any unrelated news from the final data.

The final sentiment calculation was done using *vader*, which is a part of *python nltk*- a natural language processing library. *Vader* calculates sentiment on the basis of empirically derived values. Graphical analysis was done by plotting average daily sentiment (positive and negative) with daily closing stock prices.

VI. CALCULATIONS

A. HDFC Sentiment Plots

HDFC is one of the largest private sector banks in India and has the ability to influence the markets for the banking industry.

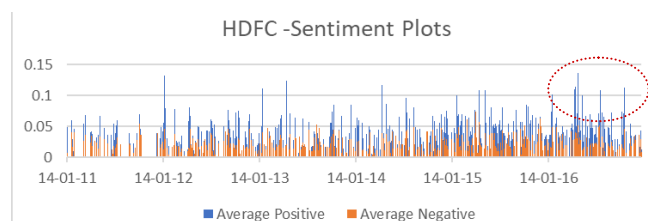


Fig. 1. HDFC Sentiment Plots: The highlighted area shows a major spike of positive sentiment. The spike in later half of 2016 is due to demonetization by Indian government and subsequent money inflow into the Indian banks which would help to smoothen their bad debt burden.

As clearly shown in the Fig. 1 the average sentiment for HDFC has been more towards the positive side rather than negative and the same has been clearly evident from the increasing trend in the plot shown below (Fig. 2) depicting the closing price of its Stock for the past 6 years.

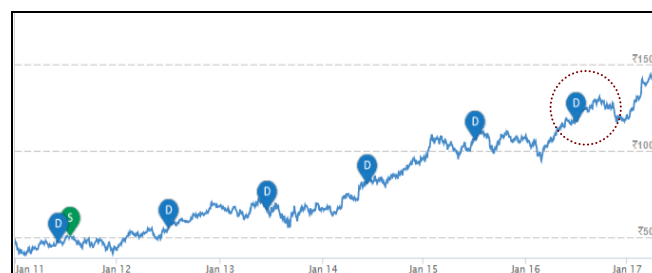


Fig. 2. Showing Closing prices of HDFC from 2011-2017. The highlighted portion shows Spike in Bank stock prices due to positive sentiment related to demonetization by Indian government.

Graph source: *moneycontrol.com*

B. ITC Sentiment Plots

ITC Limited is a major FMCG player in India with a strong market presence. The company is also one of the major tobacco manufacturers for Indian markets.

As clearly shown in the Fig. 3 the average sentiment for ITC has been more towards the positive side rather than negative and the same has been clearly evident from the

increasing trend in Fig.4 shown below depicting the closing price of its Stock for the past 6 years.

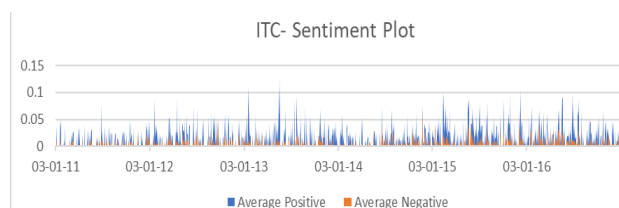


Fig. 3. ITC sentiment plots.



Fig. 4. Showing Closing prices of ITC from 2011-2017.
Graph source: moneycontrol.com

The dips in the closing prices during Jan 2012 and Jan 2016 can be clearly attributed to an increase in the negative sentiment during the same period as shown in the Fig. 3.

C. Maruti Suzuki Sentiment Plots

Maruti Suzuki is one of the largest automobile manufacturers in India with 51% of Indian passenger car market.

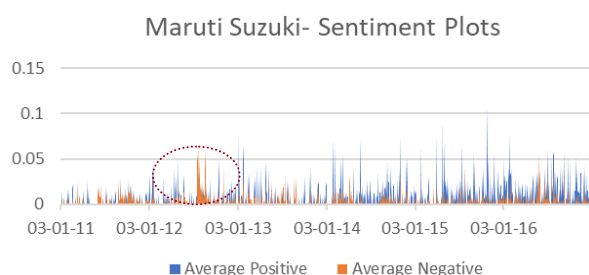


Fig. 5. Maruti Suzuki Sentiment Plots: Large negative sentiment generated due to labor disputes, high interest rates on automobile loans, and costlier fuel. This negative sentiment is clearly reflected in company's stock prices as on Dec 2011 and July 2012.

As indicated in Fig. 5 the average sentiment for Maruti Suzuki has been positive over the span of 6 years. But there were some periods like at the end of 2011, July 2012, September 2013 when there was a spike in the negative sentiment which dragged the market prices of the stock along with it.



Fig. 6. Showing Closing prices of Maruti Suzuki from 2011-2017.
Graph source: moneycontrol.com

The spurt of the negative news by the end of 2011 which

continued till the end of 2012, like higher interest rates on Automobile loans, costlier fuel, crippling labor disputes at Maruti Suzuki's Manesar Plant in India caused the stock prices to plunge at a six year low of Rs 931.55 as clearly indicated in the Fig. 6.

D. ONGC Sentiment Plots

Oil and Natural Gas Corporation is an Indian multinational Oil and natural gas company producing around 70% of India's crude oil and around 62% of its natural gas.

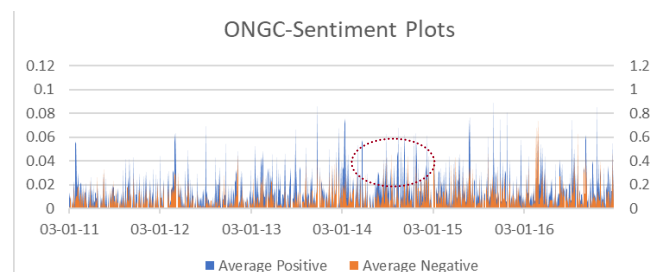


Fig. 7. ONGC Sentiment Plots: 6 June 2014 (indicated by highlighted portion), Large spikes in positive sentiment after it was published that Modi Government may hike the gas prices in July 2014. The same is reflected in Fig. 8 with ONGC hitting its 52 week high.

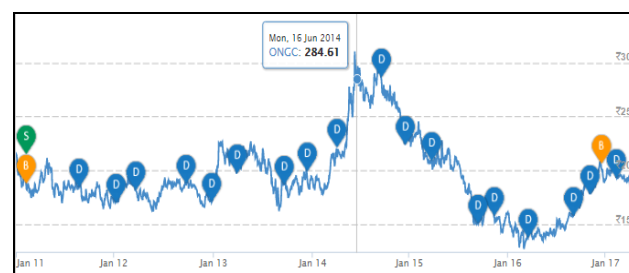


Fig. 8. Showing closing prices of ONGC 2011-2017.
Graph source: moneycontrol.com

The Fig. 7 clearly denotes the important role sentiment plays in the stock market because a rise in the positive sentiment of ONGC due to speculation of oil and gas price hikes led to a spike in its stock prices on June 2014 as clearly indicated in the Fig. 8 above.

E. SBI Sentiment Plots

State bank of India (SBI) is a public-sector banking and financial services company and is one of the largest banks with 23% of market share in assets and one-fourth of the total loan and deposit market.

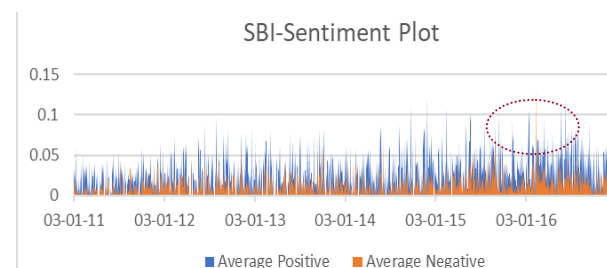


Fig. 9. SBI Sentiment Plots : As indicated by highlighted portion a large spike in the negative sentiment against SBI was noticed, after the company announced that bad loans are expected to hit its future profitability.

The overall sentiment has been highly volatile (as indicated in Fig. 9) with sharp increase in negative sentiment like in July 2011, Jan 2014 and Feb 2016 wherein negative news lead to

widespread decline in the market prices (as indicated by Fig. 10) for its stock along with some positive vibes shown like during Jan 2015 post Indian Prime Ministerial elections and overall formation of Strong development-oriented government by Narendra Modi.



Fig. 10. Showing closing prices for SBI -2011-2017. Shares of SBI plunged nearly 7 per cent, wiping out Rs 8,422 crore from its market valuation on account of announcement of decrease in profit due to bad debts (indicated by highlighted portion).

Graph source: moneycontrol.com

F. Sun-Pharma Sentiment Plots

Sun Pharmaceuticals is a major player operating primarily in Indian and US markets, selling pharmaceutical formulations and active pharmaceutical ingredients.

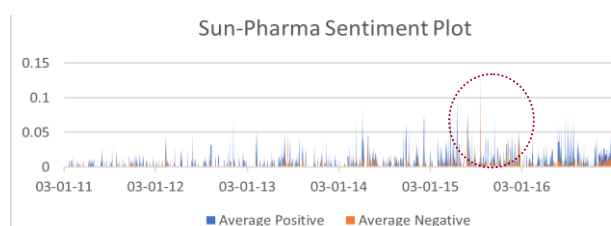


Fig. 11. Sun-Pharma Sentiment Plots: As shown by the highlighted portion a sharp rise in the negative sentiment on 21 July 2015 was observed due to company announcement of a decline in future consolidated profitability due to Ranbaxy's (a rival pharma company) acquisition.

Overall sentiment for Sun-Pharma is more towards the positive side (as shown in Fig. 11), with a few negative sentiments associated with USFDA issuing sanctions against Sun pharma during 2014 and announcement of a decline in profitability due to the acquisition of Ranbaxy in 2015. These negative sentiments are clearly reflected by the market prices of Sun Pharma as shown in the Fig. 12 below.

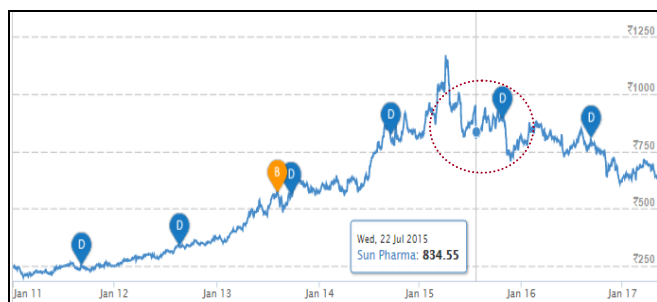


Fig. 12. Showing closing prices of Sun-Pharma 2011-2017. The region highlighted shows the 15% decline in the prices of Sun-Pharma's shares due to negative surprises.

Graph source: moneycontrol.com

G. TCS Sentiment Plots

Tata Consultancy Services is the leading information technology service, consulting and business solutions

company in India.

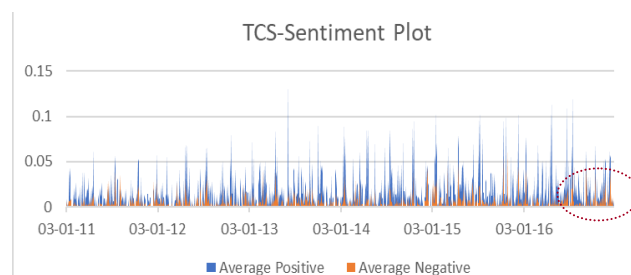


Fig. 13. TCS Sentiment Plots : A rise in negative sentiment can be noticed (highlighted portion) due to expectations of stricter visa regulations in USA due to victory of Donald Trump leading to prospects of decline in future profitability.

As shown in Fig. 13 overall TCS has shown tremendous positive sentiment over the period of 6 years and some of the major factors contributing to it are great management, exceeding the profit margins set each quarter and great cost control and efficiency of operations. It is only recently that the IT sector and companies like TCS and Infosys have come under Fire due to H-1 B visa policy regulations and accusations against the Indian IT companies misusing the visa facility. The Trump administration coming to power in November 2016 and thereby increasing the chances of a more stringent visa policy, generated a significant amount of negative sentiment due to prospects of loss in future profitability of such IT companies which would then incur higher overseas labor costs. This is clearly depicted in the graph below (Fig. 14).



Fig. 14: Showing Closing price of TCS 2011-2017. The highlighted portion indicates decline in stock prices due to negative sentiment due to H-1B visa misuse allegations and future prospects of decline in profitability.

Graph source: moneycontrol.com

H. Correlations Calculations

TABLE I: THE CORRELATIONS BETWEEN LOG CLOSING PRICES AND POSITIVE AND NEGATIVE SENTIMENT

Company	Correlation between log closing prices and Avg. positive sentiment	Correlation between log closing prices and Avg. negative sentiment
HDFC	0.0882	-0.1466
ITC	0.0224	-0.0015
SBI	0.2679	-0.1839
ONGC	0.1593	-0.0305
TCS	0.1786	-0.0569
MARUTI SUZUKI	0.0736	-0.3349

Shown below are correlations between closing prices and positive and negative sentiment. The results generated serve

as a measure of a sector's sensitivity with market sentiments.

VII. CONCLUSIONS

As clearly depicted in the Table I above banking sector as indicated by its proxies SBI and HDFC is the most sensitive sector to news as clearly indicated with high correlation of sentiment generated by news with that of the Log (closing prices). Whereas market prices of companies in FMCG sector as indicated by its proxy ITC is least correlated with that of sentiment generated by news. This important result can help in forming various trading strategies on these sectors. For instance:

For Banking sector: Due to high volatility and high susceptibility to market sentiment it is better to bet using options so that we can take advantage of high volatility.

For FMCG Sector: This sector is less volatile and mostly shows long-term trends without any significant short-term volatility, as a result, it is more beneficial to trade in fixed income debt instruments or equity-based instruments.

This study leaves a large scope for further research in devising trading strategies and classification of the sentiment using various advanced techniques of Big data, machine learning, and algorithmic trading which can be expanded further in future research works.

ACKNOWLEDGMENT

We would like to thank Dr. Anoop Kumar for his guidance and motivation to undertake this research. We would also like to extend our gratitude to the Economics department of Birla Institute of Technology and Science, Pilani, Goa for providing an encouraging and collaborative environment to undertake this study.

REFERENCES

- [1] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," *EMNLP*, 2002.
- [2] P. D. Azar, *Sentiment Analysis in Financial News*, Cambridge: Harvard College, Cambridge, 2009.
- [3] N. Godbole, M. Srinivasiah, 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, pp. 1-4.
- [6] C. G. Hutto and E. G. Vader, "A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. International AAAI Conference on Web and Social Media, North America*, 2014.
- [7] 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.
- [8] R. Prabowo and M. Thelwall, "Sentiment Analysis: A Combined Approach," *Journal of Informetrics*, vol. 3, issue 2, pp. 143-157, 2009.
- [9] A. Hillert, H. Jacobs, and S. Muller, "Media makes momentum," in *Proc. European Retail Investment Conference*, 2013.
- [10] M. Koppel and I. Shtrimerberg, *Computing Attitude and Affect in Text: Theory and Applications*, New York: Springer, chapt. 22, 2006.
- [11] M. Genereux, T. Poibeau, and M. Koppel, "Sentiment analysis using automatically labeled financial news," in *Proc. LREC Workshop on Sentiment Analysis*, 2008.
- [12] P. C. Tetlock, M. S. Tsechansky, and S. Macskassy, "More than words: quantifying language to measure firms' fundamentals," *Journal of Finance*, vol. 63, issue 3, pp. 1437-1467, May 9, 2008.



Nikhil Vijay was born on March 13, 1994 in Rajasthan, India. He is a final year student, who is learning the masters of Sciences in economics at Birla Institute of Technology and Science, Pilani, India.

He is currently a visiting researcher at Scuola Superiore Sant'Anna, Pisa, Italy. He also has the experience of interning in various global investment banks like J.P.Morgan etc. His research interests

include behavioral and quantitative finance.



Gyanesh Malhotra was born on June 5, 1995, in Varanasi, India. He is a pre-final year student, who is learning the masters in economics with bachelors in computer science at Birla Institute of Technology and Science, Pilani, Goa, India.

He has experience in python and has interned previously at Bharat Heavy Electricals Limited. He also has interest in NLP and CV.



Snimarpal Singh was born on August 23, 1994 in Punjab, India, he is a pre final year student, who is learning the masters of sciences in economics and bachelors in comp. sci. at Birla Institute of Technology and Science, Pilani, India.

He works on web technologies and data science paradigms like data mining. He is also interested in distributed ledger technologies like the block chain.