4th International Conference on Recent Developments in Science, Humanities & Management-2019
Institution of Engineers,Chandigarh (India) (ICRDSHM-19)
17th February 2019, www.conferenceworld.in ISBN:978-93-87793-70-5

Social network investor sentiments for predicting stock price trends

Sandeep Ranjan[1], Dr.Sumesh Sood[2]

[1]*PhD Scholar, I.K.Gujral Punjab Technical University, Kapurthala (India)*
[2]*Assistant Professor, I.K. Gujral Punjab Technical University, Kapurthala (India)*

ABSTRACT

Stock market investors constantly look for stock relates and other relavant information to make buy-sell-hold decisions on various stocks. Social networks are widely being recognized as reliable sources for application in prediction models. The presented research aims at developing a mathematical model using financial blog community sentiment scores to accurately predict stock price trends.For the research experiment, the www.mmb.moneycontrol blog was minedbetween 1st October 2017 and31st December 2017to create a blog dataset for top 10 private sector banks of India listed on the New Delhi Stock Exchange. Python libraries were used to detect community structures in the blog dataset and calculate the overall sentiment scores of the communities. A total of 25,456 blogs were mined for the 10 banks in the study duration. The model results achieved 0.84correlation with the actualstock price trends. The prediction results of the proposed mathematical model can be useful for stock market investors wishing to maximize their earnings.

Keywords: *-graphs, online word of mouth, opinion mining, sentiment analysis, social network communities, stock price trend prediction*

I. INTRODUCTION

The exponentially increasing demand and outreach of social media networks have created a number of research domains to apply the wisdom of the crowds in day to day decision making. The vast amount of content posted on social networks needs to be classified and analyzed empirically to validate its usefulness for society. The research discusses the need and methods to conduct an analysis of the segment-specific social network datasets for the benefit of stock market investors.

The decision making of individuals is affected by the opinions formed by collective reviews from others. This decision making has shifted from a few personal references to a huge number of social media user opinions throughout the world. Due to the increasing popularity of social media websites, the size of the opinion dataset has grown enormously that it can no longer be manually parsed to generate an inference leading to the evolution of sentiment analysis processes using specialized software. People create a buzz using their content which reflects in their connected user's activity. Viral content attracts a larger audience and Internet traffic which can be converted into business opportunity.

Due to the financial gains linked to stocks and stock markets, a large number of people are involved in stock trading who actively update their knowledge by keeping a watch on the latest happening about a particular

stock. Stock related knowledge can be gathered from social networks, news bulletins, websites and financial blogs where specific information is being disseminated.It is evident from figure 1 that the stock price graph of any company doesn't correspond to common geometrical patterns, but is dependent on sentiments comprising of variables like political, trade, economic and market conditions[1]. A large number of Asset Management Companies, foreign institutions and individuals participate in stock trading in a quest to earn maximum from stocks using unique prediction mechanisms and models. An efficient and reliable stock movement prediction can be a lot of help to investors.

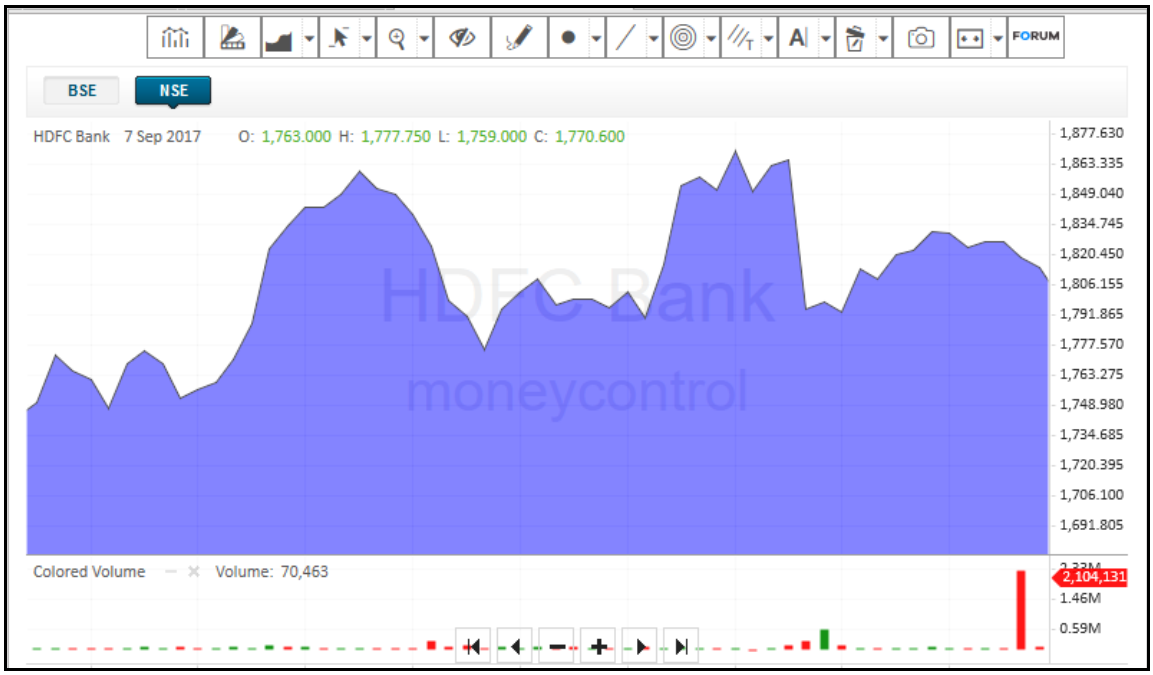

Figure 1 A snapshot of HDFC stock price from www.moneycontrol.com

Stock prediction is a complex problem involving inputs from a variety of sources like company results, government policies, international environment and investor sentiment[2]. Individual's efforts in gathering relevant and timely knowledge about a particular stock or a stock segment are limited. Investors often seek information from various online and offline channels for predicting the best time and price for buying and or selling a stock.Social networks represent the collective wisdom of crowds and hence the overall sentiment of particular communities related to stocks and stock markets can be helpful in predicting the stock price trends. There is a need to investigate the role of the general public sentiment posted on financial blog networks in predicting the stock price movement. The model proposed in the research aims at accurately predicting the stock price trends of major private sector banksof India using the overall sentiment score of blog communities.

II. RELATED WORK

Recently blogs have emerged as a self-disclosure and popular social network platform [3]. There has been a rise in the number of blogs and their associated blogger making them complex networks of individuals and information. Blogs have more specific and concentrated readership compared to other social media platforms. Blog popularity and the trust gained in the content due to correctness and relevance increases the feedback and readership for the blog. Customer support service providers, stock market investors and other consultancies are highly dependents on the blog networks.

Researchers have developed a sentiment analysis based stock price movement prediction system that combines inputs from Twitter, stock market index and Really Simple Syndication (RSS) feeds [4], [5]. Theresearchers selected Arab Bank stock listed on the Amman Stock Exchange and tested the correlation between stock price indicators and sentiments derived from RSS feed and Twitter. The models showed 14% and 20% improvement in accuracy over existing systems. The improved accuracy of results can help stock investors in making their sell, purchase and hold decisions.

Sentiment analysis of blog sentiments can provide meaningful insights to stock investors [6]. The experiment was performed on a Chinese financial blog, SinaWeibo, and the prediction results were validated againstthe Shanghai Composite Index.The study was carried out on 6.1 million blogs to predict the stock priced of stock SH000001. Filtering and cleaning of raw data are done by combining Natural Language Processing with Latent Dirichlet Allocation (LDA) model helped the researchers accurately predict the stock price. Major contributions of the research were LDA filtering of blogs and a financial lexicon.

Data from chat rooms, online forums and blogs can serve as important inputs to predict stock price and demonstrate Granger causality between the sentiments of forums and stock price predictions [7]. The authors worked on different Chinese social media platforms and employed support vector machine system using classifiers with 5 cross-validations of folds. The model had a +19.5% return in stock investments compared to -25.26% returns on conventional buy/hold strategy. Chat room sentiments revealed stock price prediction to be the best trading indicator. Timely and valuable trading hints generated by the model resulted in higher returns in the investor portfolio.

Investors often converse with each other to seek help in predicting stock price [8]. The researchers studied the impact of general public sentiment on the closing prices of stocks listed on the Philippine Stock Exchange. The research used Granger causality on the past tweet sentiment values if they have been accurate in predicting the future stock closing trends. The tweets fetched in the experiment were specifically for the geographic locations within the Philippine. The dataset created over a period of three months included about 80,000 tweets.

Researchers fetched tweets from the accounts of Bloomberg and Reuters for analysis [9]. Bloomberg and Reuters are amongst the biggest and most popular media houses in the United States. The financial news covered by these two organizations is considered to be the most trusted by the investors. News articles such as emerging markets, investment policies, market indices, mergers etc had the highest impact on the investors and their stock trading behavior.

An investigation was carried out to ascertain the relation between stock closing prices and online media content related to them [10]. The researchers experimented on the stocks of Shell, ING and Phillips from 2014 to 2015. Corporate announcements, predictions by analysts, financial news and investor opinions play a major role in the volatility of stock prices. Various forms of media including blogs disburse information that has an impact on investors and their sell-hold-buy decision making. The researchers chose the websites of "de Volkskrant", "De Telegraaf" and "NRC Handelsblad" newspapers for financial news coverage and Yahoo Finance for the stock closing price values. The paper concludes that the volume of trading (number of shares bought/sold) has a positive relationship with the media coverage volume. The sentiment covered in media coverage is more significant than the volume of coverage.

III. DATASET CREATION

The blog www.mmb.moneycontrol.com was mined for the top 5 privatesector Indian banks listed on the National Stock Exchange of India from 1st October 2017 to 31st December 2017 to create the dataset of user blogs. The National Stock Exchange of India (NSE) operates from Monday to Friday and observes some pre-declared holidays. Data scraping was used to create the dataset from the blog. Figure 2 shows a snapshot of a post from the blog hosted on www.mmb.moneycontrol.com[11].

Figure 2 A snapshot of mmb.moneycontrol.com post

Table 1.User blog field details

S. No	Field	Description
1	Username	The username of the blog user
2	Num_of_messages	Total number of messages posted by the user
3	Followers	The number of other fellow users following this user
4	Stock_name	The stock mentioned in the post
5	Time	The date and time of posting the message
6	Stock_price	The stock price when the message was posted
7	Num_of_replies	The number of replies to the concerned post

Day's highest and lowest prices are termed as "High" and "Low" price and the "Close" price is the price at the end of business hours of the National Exchange of India. Investors focus on predicting day High and

Low price of stocks so as to decide upon buy-sell-hold call for particular company stock. If the prediction points to a lower price, the investors may choose between hold or buy option, depending upon the available funds with them. Similarly, a high price prediction may encourage investors to sell the stock at a higher price to reap profits from trading.Table 2 shows the top 10 private sector banks in India with the number of blogs collected in the study duration.

Table 5.6 Description of the bank stock price blog dataset

S. No	Bank name	Number of blogs	S. No	Bank name	Number of blogs
1	Axis Bank	3688	6	IDFC Bank	1887
2	City Union Bank	768	7	IndusInd Bank	549
3	Federal Bank	2387	8	Kotak Mahindra	1688
4	HDFC Bank	4100	9	RBL Bank	957
5	ICICI Bank	7845	10	Yes Bank	1587

Table 3 shows a snapshot of the HDFC bank stock data for the month of October 2017[12]. The values have been fetched from www.moneycontrol.com website.

Table 3. HDFC bank stock prices for October 2017

Date	Opening price	Closing price	Trend
03-10-17	1808	1808.85	upside
04-10-17	1807.9	1797	downside
05-10-17	1795	1798.6	upside
06-10-17	1798	1800.1	upside
09-10-17	1798.8	1795.5	downside
10-10-17	1799	1802.7	upside
11-10-17	1803.25	1790.15	downside
12-10-17	1795.1	1818.8	upside
13-10-17	1826.7	1850.8	upside
16-10-17	1860	1857.15	downside
17-10-17	1850	1851.25	upside
18-10-17	1841	1868.5	upside
19-10-17	1861.8	1848.4	downside
23-10-17	1859.9	1863.3	upside
24-10-17	1863	1867.1	upside
25-10-17	1860	1795.1	downside
26-10-17	1768	1795.35	upside
27-10-17	1804.4	1791.05	downside
30-10-17	1798.25	1815	upside
31-10-17	1811.5	1808.5	downside

IV. TEXT PREPROCESSING AND SENTIMENT ANALYSIS

The mined blog datasets were preprocessed to remove emojis, uniform resource locators (URL) and non-English characters using various Python libraries and KU tools for MS Excel. Sentiment analysis aims at digging out opinions from data sources[10], [13], [14]. As Web has become a huge repository of opinions, processing of opinions is required to be used as input in marketing or other modeling tasks. The domain sentiment analysis research attracts researchers from computer science, management science and social science owing to its significance in an economic and social framework.

For the research presented, Python NLTK and TextBlob libraries were employed to perform sentiment analysis on the blog dataset. The "sentiment" function returns a tuple of the form (polarity, subjectivity) where polarity is a float value within the range [-1.0, 1.0] where -1.0 is a negative sentiment, 0 is a neutral sentiment and 1.0 is a positive sentiment; and subjectivity is a float value within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective.

Polarity and subjectivity values are computed for the individual blog (each row of the CSV dataset file). The negative polarity values or the polarity values less than zero represent a negative sentiment and the positive polarity values or the polarity values greater than zero represent a positive sentiment. The final polarity of a tweet/ blog is calculated as the product of polarity and subjectivity value as shown in the equation.

$$Polarity\_Final = Polarity * Subjectivity$$

Based on the final polarity value i.e. Polarity_Final, highly positive sentiment tweets or blogs were labeled as P+, slightly positive sentiment tweets or blogs were labeled as P, highly negative sentiment tweets or blogs were labeled as N+, slightly negative sentiment tweets or blogs were labeled as N. Weights were assigned to the tweets as shown in table 2. Overall sentiment score of a bankstockwas calculated as the product of the number of blogs andthe polarity weight.

Table 4. Sentiment polarity and weights

Polarity_Final	**Range**				
	-1.0 to -0.5	-0.5 to -0.1	-0.1 to 0.1	0.1 to 0.5	0.5 to -1
Polarity	N+	N	NEU	P	P+
Weight	-2	-1	0	1	2

V. PREDICTION MODEL

Figure 8.3 shows the proposed mathematical model used for predicting bank stock price day end value. Predicted trends were calculated for the top 10private sector listed banks in India for the experiment period

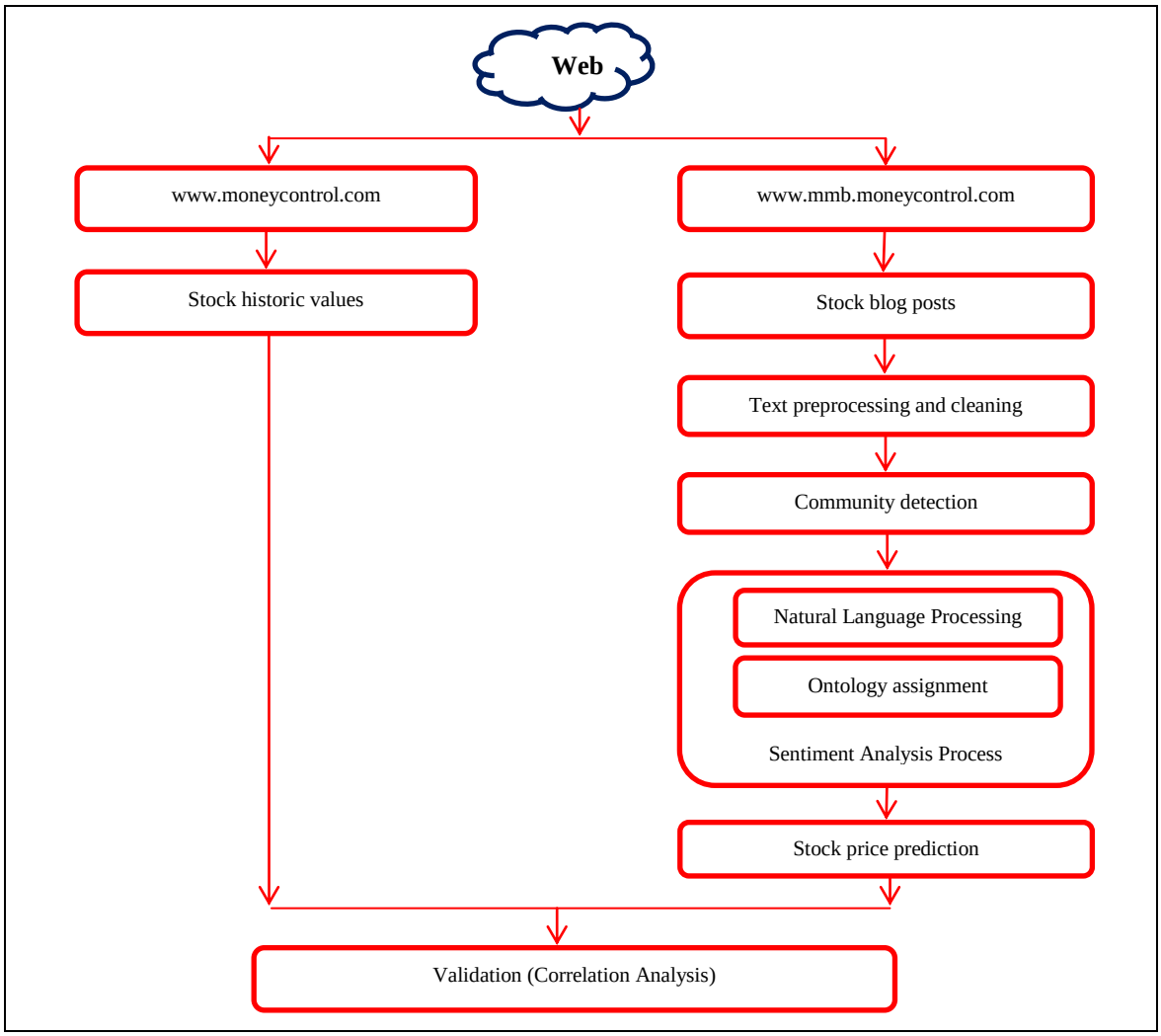

Figure 1. Community sentiment analysis based prediction model

Table 5. Model prediction results for top 10 private bank stocks

S. No	Stock Name	Predicted Values			Actual Values			Number of correct predictions	% of correct predictions
		Upside closing price days	Downside closing price days	Stagnant closing price days	Upside closing price days	Downside closing price days	Stagnant closing price days		
1	Axis Bank	29	32	1	30	32	0	59	95.2
2	City Union Bank	21	40	1	23	38	1	56	90.3
3	Federal Bank	25	37	1	24	38	0	50	80.6
4	HDFC Bank	38	24	0	35	27	0	57	91.9

5	ICICI Bank	34	28	0	32	30	0	57	91.9
6	IDFC Bank	19	38	5	15	45	2	51	82.3
7	IndusInd Bank	29	31	2	24	38	0	50	80.6
8	Kotak Mahindra	26	33	3	28	34	0	57	91.9
9	RBL Bank	29	33	0	24	38	0	49	79
10	Yes Bank	25	36	1	22	39	1	58	93.5
	Total	275	332	14	257	359	4	544	87.74

VI. VALIDATION OF RESULTS

Statistical correlation tests were used to validate the results of the study analyzing the impact of investor sentiment on stock prices [15]. IBM SPSS 21 was used to validate the model's prediction trends against the actual trends obtained from www.moneycontrol.com. Table 6 shows the results of Bivariate Correlation Analysis test performed on the predicted and actual values from table 5. A high correlation of 0.84 has been achieved for the proposed mathematical model results. This validates the results of the mathematical model.

Table 6. Bivariate Correlation Analysis validation results

		Predicted verdict	Actual verdict
Predicted verdict	Pearson Correlation	1	.84
	Sig. (2-tailed)		.000
	Number of banks	10	10
Actual verdict	Pearson Correlation	.84**	1
	Sig. (2-tailed)	.000	
	Number of banks	10	10

VII. CONCLUSION

Social networks like blogs contain an online word of mouth in the form of dense communities.These network communities evolve and flourish when valuable, timely and correct information is sharedby the participating nodes. This OWOM can be used to develop mathematical models and applications for guiding stock market investors who are constantly looking for authentic information for their buy-sell-hold decisions. The research presented proposed a mathematical model on the overall sentiment scores of the financial blog community sentiments. The results have been validated statistically with a high accuracy of 87.74% and showing a correlation of 0.84 with the actual stock price trends.

REFERENCES

[1] "HDFC Bank Ltd. Stock Price," 2017. [Online]. Available: https://www.moneycontrol.com/india/stockpricequote/ banks-private-sector/hdfcbank/HDF01. [Accessed: 16-Jan-2019].

[2] S. Ranjan, I. Singh, S. Dua, and S. Sood, Sentiment Analysis of Stock Blog Network Communities for Prediction of Stock Price Trends, Indian Journal of Finance, 12(12), 2018, 7-21.

[3] L. S. Huang, Trust in product review blogs: The influence of self-disclosure and popularity, Behaviour and Information Technology, 34(1), 2015, 33-44.

[4] S. Bharathi and A. Geetha, Sentiment Analysis for Effective Stock Market Prediction, International Journal of Intelligent Engineering and Systems, 10(3), 2017, 146-153.

[5] S. Bharathi, A. Geetha, and R. Sathiynarayanan, Sentiment Analysis of Twitter and RSS News Feeds and Its Impact on Stock Market Prediction, International Journal of Intelligent Engineering and Systems, 10(6), 2017, 68-77.

[6] B. Zhao, Y. He, C. Yuan, and Y. Huang, Stock Market Prediction Exploiting Microblog Sentiment Analysis, in 2016 International Joint Conference on Neural Networks (IJCNN) IEEE, 2016, pp. 4482-4488.

[7] T. Sun, J. Wang, P. Zhang, Y. Cao, B. Liu, and D. Wang, Predicting Stock Price Returns Using Microblog Sentiment for Chinese Stock Market, in Proceedings - 2017 3rd International Conference on Big Data Computing and Communications, BigCom, 2017, 87-96.

[8] A. R. Caliñgo, A. M. Sison, and B. T. Tanguilig III, Prediction Model of the Stock Market Index Using Twitter Sentiment Analysis, International Journal of Information Technology and Computer Science, 8(10), 2016, 11-21.

[9] N. Strauß, R. Vliegenthart, and P. Verhoeven, Intraday News Trading: The Reciprocal Relationships Between the Stock Market and Economic News, Communication Research, 45(7), 2017, 1054-1077.

[10] J. Strycharz, N. Strauss, and D. Trilling, The role of media coverage in explaining stock market fluctuations: Insights for strategic financial communication, International Journal of Strategic Communication, 12(1), 2018, 67-85.

[11] Online Business Forum, Economics; Stocks Discussion, Share Market Ideas, Questions and Answers India - MoneyControl, 2019. [Online]. Available: https://mmb.moneycontrol.com/. [Accessed: 12-Jan-2019].

[12] Historic prices of stocks and indices, 2017. [Online]. Available: https://www.moneycontrol.com/stocks/hist_stock_result.php?ex=N&sc_id=HDF01&mycomp=HDFC Bank. [Accessed: 01-Oct-2017].

[13] A. F. Alsaqer and S. Sasi, Movie review summarization and sentiment analysis using rapidminer, in 2017 International Conference on Networks and Advances in Computational Technologies, NetACT 2017, 2017, 329-335.

[14] S. K. Janane, M. S. Keerthana, and B. Subbulakshmi, Hybrid Classification for Sentiment Analysis of Movie Reviews, International Journal of Engineering Sciences and Research, 7(4), 2018, 724-728.

[15] B. Dickinson and W. Hu, Sentiment Analysis of Investor Opinions on Twitter, Scientific Research Publishing, 4(3), 2015, 62-71.