Sentiment Analysis of Stock Blog Network Communities for Prediction of Stock Price Trends

** Sandeep Ranjan*
*** Inderpal Singh*
**** Sonu Dua*
***** Sumesh Sood*

Abstract

It has been a challenge to develop a successful model for accurate stock price trend prediction. This paper aimed to develop an accurate model based on semantic analysis of social network communities for predicting stock day end closing prices using the wisdom of crowds ; www.mmb.moneycontrol.com, a financial blog covers all the companies listed on the National Stock Exchange of India. Influential and accurate opinions expressed in the blogs lead to community formation. The study focused on detection of such communities using betweenness centrality measure and performed a semantic analysis of their content to develop a prediction model based on correlation between blog sentiments and stock day end closing prices for predicting the stock trends. Thirty nine Indian banks were selected for the study during the period from October 1, 2017 to December 31, 2017 and the experimental results of the number of correct predictions of upside and downside movement of day end stock price were validated against the actual values. The model achieved a prediction accuracy of 84% and correlation of the model was within the significant limits of *Z* - test and Pearson's coefficient of 0.8.

Keywords : social network communities, sentiment analysis, stock prediction, wisdom of crowds

JEL Classification : C4, O2, O3

Paper Submission Date : June 26, 2018 ; Paper sent back for Revision : November 20, 2018 ; Paper Acceptance Date : November 23, 2018

Stock prediction is a complex problem involving inputs from a variety of sources like company results, government policies, international environment, and investor sentiment. Individual's efforts in gathering relevant and timely knowledge about a particular stock or a stock segment is limited. Investors often seek information from various online and offline channels for predicting best time and price for buying and or selling a stock. In the last few years, the acceptance of the Internet and online sources as a trustworthy source has increased manifolds (Patino, Pitta, & Quinones, 2012). With the emergence of the Web 2.0 era of the Internet, the amount of

* *Research Scholar,* Inder Kumar Gujral Punjab Technical University, Jalandhar - Kapurthala Highway, Kapurthala -144 603, Punjab and **Assistant Professor,* Lyallpur Khalsa College of Engineering, Jalandhar -144 001, Punjab. E-mail: ersandeepranjan@yahoo.com

***Associate Professor & Dean Research*, Khalsa College Lyallpur Institute of Management & Technology, Jalandhar -144 001, Punjab. E-mail: arorainderpal8086@gmail.com

****Assistant Professor*, Khalsa College Lyallpur Institute of Management & Technology, Jalandhar -144 001, Punjab. E-mail: sonudua3778@gmail.com

*****Assistant Professor,* Inder Kumar Gujral Punjab Technical University, Jalandhar - Kapurthala Highway, Kapurthala -144 603, Punjab. E-mail: sumesh64@gmail.com

the general public posted content has been rising exponentially, leading to the evolution of blogs and other social media. Social media's emergence and ability to affect users has changed the marketing and brand management basics, giving more importance to user reviews compared to other channels (Bollen, Mao, & Zeng, 2011).

Blogs have users from a variety of age groups, nationalities, and backgrounds who express their views on common topics of interest. A herd behavior can be seen in case of relevant or attractive topics posted for discussion (Wu & Lu, 2017). Evidently or factually correct appearing data attracts more attention and forms a chain of opinion sharing, representing the online word of mouth (OWOM).

Traditionally, fundamental analysis and technical analysis are used to predict stock price movement. Daily price movement, various stock trend indicators, and day's lowest and highest price values are used for the technical analysis. In fundamental analysis, financial data and reports are studied. The world of finance has seen significant changes due to social media and information technology revolution (Swain & Dash, 2017). These days, stock analysts share their opinions on various online blogs. Their opinions and views impact a large number of social media users interested in predicting stock price movement for maximizing their profits. Users respond more actively to those posts which have been accurately predicting stock price movements in the past as compared to other posts. Emotional attachment, herd behaviour, imitation, and information streams have a significant effect on user decisions (Rout & Das, 2015). The blogs, just like other social media platforms, can be represented through graphs where the nodes are the users, and the edges are the relations that develop when they interact with each other. When users share, like, or respond to the threads in blog networks, communities start evolving in the networks. Communities represent the essence of the network and can be harnessed for calculating the overall sentiment of big networks. The paper proposes to investigate the role of general public sentiment posted on financial blog networks in predicting the stock price movement.

This research article emphasizes on the collective wisdom of the crowds contained in the specific blogs such as a financial blog where serious investors share their valuable opinions. The blogs are relevant to the situations of a company, market segment, the entire market, political environment, or global indices at a given time. A new model has been proposed to identify the communities in the blog network of www.mmb.moneycontrol.com - a popular stock related blog hosted on www.moneycontrol.com and semantic analysis was performed on the communities to establish a correlation between community behavior and stock price prediction. The correlation results have been validated statistically.

Literature Review

A popular hypothesis, known as the wisdom of the crowds, states that predictions made by a random and diverse crowd are more precise than the predictions made by an individual or a group of a few individuals (Chen, De, Hu, & Hwang, 2014; Qiu, Rui, & Whinston, 2013). The researchers investigated the social media based community sentiments for accurately predicting stock movements. The model's results lacked consistency in accurately predicting the stock trends. To improve consistency, a word of mouth based chain of sentiments is to be trapped in the form of large communities to be used for accurate predictions.

In the era of information technology, investors find it easy to gather other fellow investor's opinions through financial blogs and social networks. Specific blogs like financial or stock market-related blogs are used by investors before making a hold, sell, or a purchase decision (Bharathi, Geetha, & Sathiynarayanan, 2017). The studied hypothesis was based on the sentiments of investors gathered from really simple syndication (RSS) feeds as well from Twitter users, who could be amateur in stock fundamentals. To get better results, this study relies on a single, more concentrated and specific data source ,that is, a financial blog.

News articles capture the sentiments, and the polarity of news can cause the stock prices to exhibit an upward trend or a downward trend. A weighted network of sentiments of users obtained from www.mmb.moneycontrol.com was analyzed for underlying sentiments using tools like Gephi to model the financial market and predict the price movements (Saumya, Singh, & Kumar, 2016). The study concluded that the combined wisdom of crowds has better performance than smaller groups or individuals. One can benefit from the experiences of others who have gone through similar situations. Stock price trend was predicted for a bank stock and a telecom stock. The major limitation of the work was a small dataset studied in the research, and the results were not validated by any methods. To give authenticity to the model, the dataset should be large and the results need to be statistically validated.

Microblog sentiment analysis based prediction systems can yield higher returns to investors (Zhao, He, Yuan, & Huang, 2016). The study was conducted by the authors' study on Sina Weibo, a popular Chinese financial microblog and used Shanghai Composite Index as the benchmark for their prediction results. With 6.1 million microblogs, the model yielded only 69.09% accuracy. A higher accuracy is needed as every inaccurate prediction leads to financial loss for the investor. The present study aims at developing a model which provides accurate prediction with relatively small input size.

Researchers have used hybrid neural networks based blog sentiment analysis algorithms for stock prediction models (Wang, 2017). The algorithm divided micro-blogs into sentence fragments and analyzed them with parsing algorithm, generating a sentiment weights series, which was tested in predicting stock prices in the Shanghai Stock Exchange Composite Index. The weighted sentiment series did not include the trading volumes which is one of the major factors motivating the investors in their decisions. The model can be improved by incorporating trading volumes and current stock price as sentiments affect the stock price movements.

Market live server based research was conducted on the two Indian stock exchange indices, the Bombay Stock Exchange SENSEX and the National Stock Exchange NIFTY (Bhardwaj, Narayan, Vanraj, Pawan, & Dutta, 2015). These indices are dependent on the stock values of selected company stocks listed on the respective exchanges and reflect the overall market mood at a given time. The study focused on the web pages related to the two exchanges, and the findings suggested that indices values at regular intervals can help investors in predicting the stock prices. This can be improved by including the sentiments of the investors.

Media pessimism was studied in the *Wall Street Journal* newspaper's daily columns to establish a quantitative

Table 1. Showing Gaps Identified in Previous Models

S. No	Author Name	Paper Title	Gaps Identified
1	Chen, De, Hu, & Hwang (2014)	"Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media"	Community online word of mouth was not used in the model.
2	Bharathi, Geetha, & Sathiynarayanan (2017)	"Sentiment Analysis of Twitter and RSS News Feeds and its Impact on Stock Market Prediction"	Only RSS feeds were used which don't represent the sentiments completely.
3	Saumya, Singh, & Kumar (2016)	"Predicting Stock Movements Using Social Network"	Small dataset was analysed, sentiments were not found to represent the overall investor sentiment.
4	Zhao, He, Yuan, & Huang (2016)	"Stock Market Prediction Exploiting Microblog Sentiment Analysis"	Small input size of sentiments did not represent the overall investor sentiment.
5	Wang (2017)	"Stock Market Forecasting with Financial Micro-Blog Based on Sentiment and Time Series Analysis"	Online word of mouth was not effectively utilized in the model.
6	Bhardwaj, Narayan, Vanraj, Pawan, & Dutta (2015)	"Sentiment Analysis for Indian Stock Market Prediction using Sensex and Nifty"	Investor sentiment was not given due weightage in predicting market sentiment.
7	Tetlock (2007)	"Giving Content to Investor Sentiment : The Role of Media in the Stock Market"	Large communities of networks have not been considered in the model.

relationship between the media sentiments and stock market (Tetlock, 2007). The author concluded that a high amount of pessimism by media is a prediction of a downward trend in the stock market and a low pessimism predicts high trading volume. General public's opinion differs with the different trading volumes and stock prices and communities evolve in the network when predictions are accurate. The results can be improved by taking into consideration the large communities formed as a result of correct predictions.

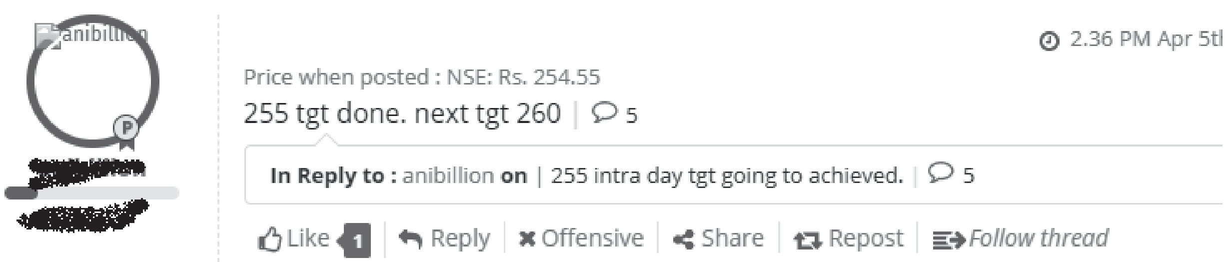
Personal behaviour and feelings are major driving forces in the decision making process of an investor (Mangala & Sharma, 2014). Sentiment analysis has even been used in developing countries for the prediction of stock markets where the popularity of social media is a still way behind as compared to its popularity in developed countries. Perez - Liston and Huerta (2012) tried to fill the gap between the sentiments of developed and developing countries using the Mexican Stock Exchange data and indicated that a herd behaviour can be detected in case of accurate predictions adding to the weight of the initial opinion when it passes the test of time. Results were consistent in both large-cap and small-cap portfolios where investors followed accurate predictions conveyed in the sentiments of conversation initiators. Suresh and George (2016) analyzed the effect of sentiment on returns of equities in Indian markets. The model achieved fair accuracy with respect to the dataset studied. The research gaps identified in the various studies discussed in the literature review have been summed up in the Table 1.

Data and Method

(1) Dataset Creation : This section discusses the dataset creation process and prediction model used in the research. The blog www.mmb.moneycontrol.com was mined for 39 banks listed on the National Stock Exchange of India from October 1, 2017 to December 31, 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. Banks and bank stocks have been popular in India due to the on-going economic reforms in the banking sector and their effect on the Indian economy. They attract a large number of investors and were chosen as an ideal stock segment for the present study due to this reason.

Bloggers reply to the posts they find appealing by expressing their support or opposition in the replies. The data scraping technique was used to mine the human - readable content from the blog website. The blog contains information about users such as their usernames, total number of messages posted by a particular user, stock about which the message has been posted, stock price at the moment of posting of the opinion, and the number of replies posted by other users in response to that post. Some of the posts were about the marketing of "demat" service providers, insurance companies, and or some promotional events. Such irrelevant posts were discarded from the dataset in the pre - processing phase.

Figure 1. A Snapshot of mmb.moneycontrol.com Post



Source: www.mmb.moneycontrol.com

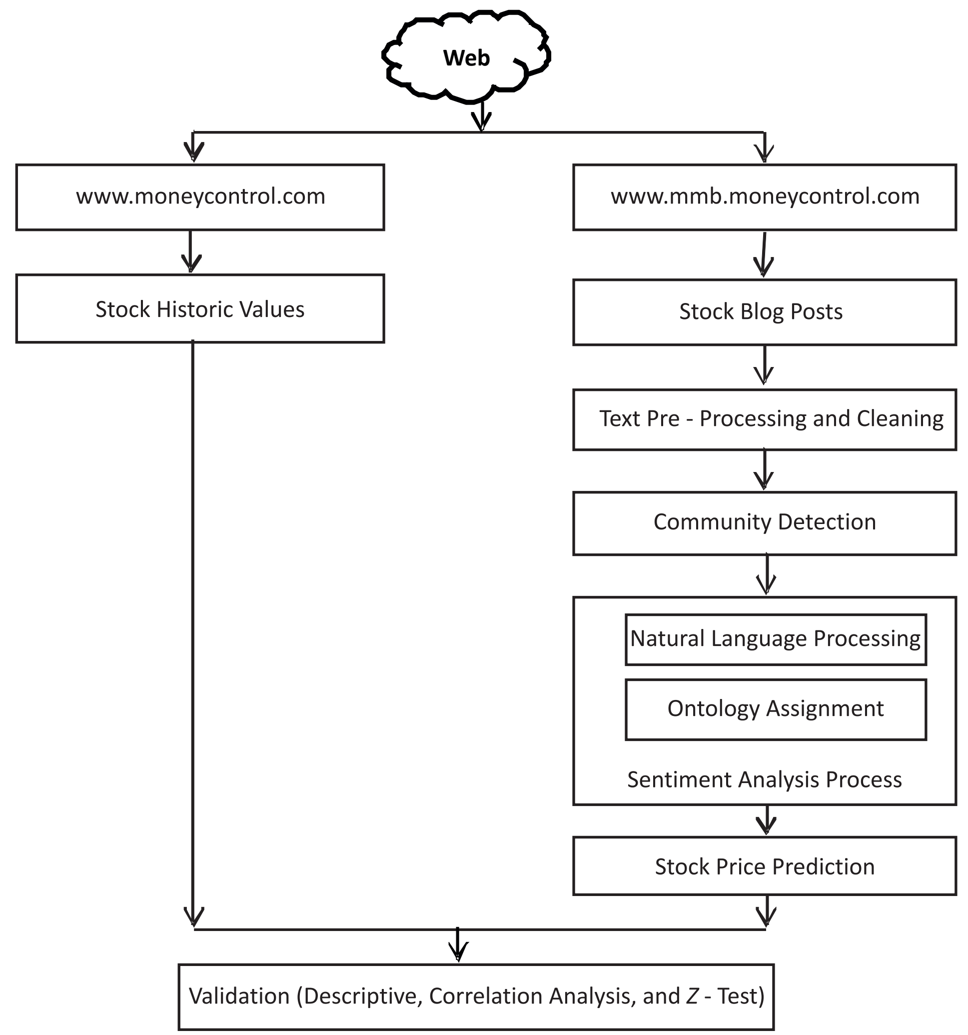
Table 2. User Blog Field Details

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.

Source: Compiled from Blog structure used by mmb.moneycontrol.com

Figure 1 is a screenshot of an actual post on the www.mmb.moneycontrol.com and Table 2 describes the various fields in the posts. The price of stock at the time of post is also shown in the screenshot. The Figure 2 shows the

Figure 2. Architecture of the Proposed Model



architecture of the proposed model. User posts are fetched from the blog website and stock prices were fetched from www.moneycontrol.com website. Major tasks are to identify community structures in the datasets and perform semantic analysis of the community posts. The overall sentiment score is used to predict day end stock price of each stock.

(2) Pre - Processing : This process discards the data which is incomplete, incorrect, or irrelevant. Many blogs contain irrelevant data like advertisement links of different companies involved in paid market analysis and stock trading software. Sometimes, bloggers post non-English language text, which cannot be used in the natural language processing. Such data needs to be removed from the dataset. This module identifies and removes the dirty data, and the cleaned data are passed to the next module.

(3) Community Detection : The advancement in data mining methods has made it easier to create datasets of interrelated data from different sources. These datasets, when viewed as complex graph, reveal the constituent nodes, edges, and communities. Many existing community detection algorithms are specific to problem sets. The results of the framework tested on simulation model show that in the future, the enhanced performance compared to individual algorithms can be used for real-life datasets also.

Communities evolve in a graph when there is a multiplicity of edges amongst the nodes. Such multiple edges may be due to the popularity of the topic being discussed as is the case of a financial blog. When a user posts something about a stock, other users may reply to it if they feel the sentiment expressed in the said post is relevant (whether positive or negative). Communities play a major role in the spread of the online word of mouth as they are composed of like-minded users who discuss common situations and issues (Ren, Yeoh, Shan Ee, & Popovič, 2017). There are retweets, sharing, and linking of common interest posts, which adds on to the weight of the content being shared amongst community members. Due to the increasing popularity of social media as an inexpensive mode of communication, there is a motivation to detect communities which can help to identify participants and audience of the content (Gialampoukidis, Kalpakis, Tsikrika, Papadopoulos, Vrochidis, & Kompatsiaris, 2017).

The task of community detection can be defined as a process of automatically extracting a simplified view of the network (Peel, Larremore, & Clauset, 2016). Community detection also provides the coarse grain view of how network constituents interact with each other. Python libraries and packages were used to identify communities in the stock datasets based on clustering. Posts with sentiments reflecting the stock price movement and prediction had more following relation with other posts resulting communities. The largest community detected in this module is chosen for further processing. Consider a graph with the following parameters :

(i) An undirected, unweighted, connected graph $G = < V, E >$.

(ii) $\sigma(s, t)$, the number of shortest paths existing between nodes s and t.

(iii) $\sigma(s, t \mid v)$, the number of shortest paths existing between nodes s and t passing through node v.

The betweenness for each node is calculated as the number of edges connected to that node and is defined in terms of shortest paths that go through v (Ranjan & Sood, 2018). The betweenness centrality of v, $BC(v)$ is defined as :

$$CB(v) = \sum_{s,t \in v} \frac{\sigma(s, t \mid v)}{\sigma(s, t)}$$

Table 3. Polarity and Weights

Polarity	N+	N	NEU	P	P+
Weight	-2	-1	0	1	2

Source: Polarity and weights assignments used in the proposed model.

(4) Sentiment Analysis : Sentiment analysis is done to associate polarity with each post based on the opinion expressed in the text of the post. Social network sentiment analysis is finding crucial applications in recent developments. Researchers have used Twitter sentiment analysis in the on-going Syrian refugee crisis (Öztürk & Ayvaz, 2018). Social media offers a free platform to people to share their feelings with others. Various techniques of natural language processing are used to tokenize, parse, and generate meaning of the phrases (Xiaomei, Jing, Jianpei, & Hongyu, 2018). Online customers are more accessible and active, and can provide cultural and marketing information that enables them to become a driving force.

The study used the textblob & natural language toolkit of Python to perform the sentiment analysis of the datasets. The positive sentiment post is assigned +1 or +2 weight as per the intensity of the sentiments ; similarly, negative posts are awarded -1 or -2 weight as shown in the Table 3. Sentiment analysis was not done on individual posts, which were not a part of any community. Words containing sentiments such as buy, sell, hold, moving up, scrap, debt, shoot, plunges, etc. were added to the dictionary for grasping the sentiment of the data.

Proposed Model

Stock price prediction is used to predict the day end stock price. If it is higher than the day end value of the previous working day, the stock is said to show a positive movement and vice versa. The overall sentiment score of a particular community is calculated by summing the scores of the individual posts (Ravi & Ravi, 2015). The overall sentiment score of a stock over a day is used to predict the stock day end price. If the overall score is net positive, the stock price will move up and if the overall score is negative, the stock price will go down. 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 the day high and low price of stocks so as to decide upon the buy - sell - hold call for a 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.

Correct predictions of low, high, and close are essential for profitable stock trading. The study focuses on predicting the closing price movement with respect to the opening price for 39 banking stocks listed on the National Stock Exchange of India on a day to day basis for the said period. The stock prices of the banks were mined from the website www.moneycontrol.com. The Table 3 shows a snapshot of State Bank of India (SBI) stock data for the month of October 2017. The values have been taken from www.moneycontrol.com website.

Each of the 39 bank stocks were monitored on a daily basis (on working days of the National Stock Exchange during the study period) and predictions were made based on the overall sentiment score of the communities in the dataset of stock related blogs. A positive score representing positive sentiment predicts the day end price to be higher than the opening price (an upside closing trend or value), while a negative score representing negative sentiment predicts the day end price to be lower than the opening price (a downside closing trend or value).

Analysis and Results

(1) Experimental Results : Using the proposed model based on the overall sentiment score of communities in the

Table 4. A Snapshot of the SBI Stock Price : Actual and Predicted Trend

4 (a)

Date	Opening Price	Closing Price
03-10-17	253.85	251.3
04-10-17	251.65	253.2
05-10-17	253.7	251.6
06-10-17	252.15	256.75
09-10-17	256.8	256.85
10-10-17	256.85	256.95
11-10-17	257.15	251.75
12-10-17	252	251.2
13-10-17	251.1	252.1
16-10-17	254.6	252.05
17-10-17	252.05	251.15
18-10-17	246	243.75
19-10-17	243.8	242.75
23-10-17	242.75	245.95
24-10-17	246	254.45
25-10-17	279.85	324.9
26-10-17	330	320.5
27-10-17	320.25	311.05
30-10-17	312	312

4 (b)

Date	Overall Sentiment Score	Trend Prediction	Actual Trend	Prediction Success
03-10-17	-10	Downside	Downside	YES
04-10-17	15	Upside	Upside	YES
05-10-17	-8	Downside	Downside	YES
06-10-17	18	Upside	Upside	YES
09-10-17	6	Upside	Upside	YES
10-10-17	8	Upside	Upside	YES
11-10-17	-15	Downside	Downside	YES
12-10-17	6	Upside	Downside	NO
13-10-17	13	Upside	Upside	YES
16-10-17	-11	Downside	Downside	YES
17-10-17	-6	Downside	Downside	YES
18-10-17	-9	Downside	Downside	YES
19-10-17	-7	Downside	Downside	YES
23-10-17	15	Upside	Upside	YES
24-10-17	23	Upside	Upside	YES
25-10-17	53	Upside	Upside	YES
26-10-17	-14	Downside	Downside	YES
27-10-17	-14	Downside	Downside	YES
30-10-17	0	Stagnant	Stagnant	YES

Source: Table 4(a) and Table 4(b) compiled from Moneycontrol (n.d.b.) : https://www.moneycontrol.com/india/stockpricequote/banks-public-sector/statebankindia/SBI

Table 5. Comparison of Predicted Values of Day End Closing Price Movement with Actual Values

S. No	Stock Name	Predicted Values			Real 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	Allahabad Bank	34	26	2	27	34	1	52	83.9
2	Andhra Bank	30	32	0	24	37	1	57	91.9
3	Axis Bank	29	32	1	30	32	0	59	95.2
4	Bank of Baroda	33	28	1	23	39	0	48	77.4
5	Bank of India	27	35	0	22	40	0	49	79.0
6	Bank of Maharashtra	22	34	6	20	37	5	55	88.7
7	Canara Bank	31	31	0	26	36	0	47	75.8
8	Central Bank	21	41	0	18	44	0	50	80.6
9	City Union Bank	21	40	1	23	38	1	56	90.3
10	Corporation Bank	25	35	2	22	38	2	53	85.5
11	DCB Bank	32	29	1	31	31	0	57	91.9
12	Dena Bank	20	40	2	17	45	0	55	88.7
13	Dhanlaxmi Bank	18	40	4	12	50	0	47	75.8
14	Federal Bank	25	37	1	24	38	0	50	80.6

15	HDFC Bank	38	24	0	35	27	0	57	91.9
16	ICICI Bank	34	28	0	32	30	0	57	91.9
17	IDBI Bank	23	38	1	27	34	1	54	87.1
18	IDFC Bank	19	38	5	15	45	2	51	82.3
19	Indian Bank	30	32	0	29	33	0	55	88.7
20	Indian Overseas Bank	22	35	5	19	36	7	50	80.6
21	IndusInd Bank	29	31	2	24	38	0	50	80.6
22	JK Bank	14	44	4	16	44	2	49	79.0
23	Karnataka Bank	31	31	0	26	36	0	51	82.3
24	Karur Vysya	30	31	1	24	38	0	53	85.5
25	Kotak Mahindra	26	33	3	28	34	0	57	91.9
26	Lakshmi Vilas	34	27	1	30	32	0	50	80.6
27	Oriental Bank	24	38	0	23	39	0	49	79.0
28	Punjab & Sind	22	37	3	25	36	1	48	77.4
29	Punjab National Bank	29	30	3	26	36	0	47	75.8
30	RBL Bank	29	33	0	24	38	0	49	79.0
31	South Indian Bank	25	36	1	21	37	4	54	87.1
32	State Bank of India	20	38	4	25	37	0	52	83.9
33	Standard Chartered Bank	23	36	3	25	32	5	57	91.9
34	Syndicate Bank	31	30	1	27	35	0	51	82.3
35	UCO Bank	27	32	3	19	42	1	47	75.8
36	Union Bank of India	28	33	1	23	38	1	48	77.4
37	United Bank	18	42	2	15	43	4	48	77.4
38	Vijaya Bank	26	35	1	22	38	2	54	87.1
39	Yes Bank	25	36	1	22	39	1	58	93.5
	Total	**1025**	**1328**	**66**	**921**	**1456**	**41**	**2022**	**84.0**

Source: Predicted values generated from the proposed model and actual values reproduced from www.mmb.moneycontrol.com

dataset, predictions were made on the day closing price compared to the day opening price. An upside prediction arising from positive sentiment means that the day end closing price will be higher than the opening price of the stock and a downside prediction arising from negative sentiment means that the day end closing price will be higher than the opening price of the stock.

In all, the National Stock Exchange was operational for 62 days and predictions for the day end price were made for each of the 62 days. Stagnant closing value means the day end closing price of the stock was equal to the opening price. An overall neutral sentiment of the community predicts a stagnant price of the stock. The stock price movement predictions obtained from the prediction model for all 39 bank stocks were compared with the actual stock day end price figures mined from www.moneycontrol.com as shown in the Table 5 (Moneycontrol, n.d.a. ; State Bank of India, n.d.).

It can be seen from the Table 4 that the model made 1025 positive movement predictions, 1328 negative movement predictions, and 66 stagnant predictions of stock closing prices for the study period. Out of the total 2418 predictions (39 predictions per day for 62 working days of National Stock Exchange of India), 2022 predictions were correct in forecasting the closing price movement with respect to the opening price of stocks.

Table 6. Upside and Downside Closing Trends : Descriptives

6(a)

Variable	Mean	Std. Deviation	Number of Banks
Predicted Upside Closing Days	26.28	5.380	39
Actual Upside Closing Days	23.62	4.956	39

6(b)

Variable	Mean	Std. Deviation	Number of banks
Predicted Upside Closing Days	34.05	4.634	39
Actual Upside Closing Days	37.33	4.567	39

Table 7. Upside and Downside Closing Trends : Correlation Analysis

7(a)

		Predicted Upside Closing Days	Actual Upside Closing Days
Predicted Upside Closing Days	Pearson Correlation	1	.80
	Significant value		.001
	Number of banks	39	39
Actual Upside Closing Days	Pearson Correlation	.80	1
	Significant value	.0001	
	Number of banks	39	39

7(b)

		Predicted Upside Closing Days	Actual Upside Closing Days
Predicted Downside Closing Days	Pearson Correlation	1	.79
	Significant value		.001
	Number of banks	39	39
Actual Downside Closing Days	Pearson Correlation	.79	1
	Significant value	.001	
	Number of banks	39	39

Table 8. Upside and Downside Closing Trends : *Z* - Test

8(a)

	t	Degree of Freedom	Significant Value
Predicted Upside Closing Days	30.508	38	.0001
Actual Upside Closing Days	29.759	38	.0001

8(b)

	t	Degree of Freedom	Significant Value
Predicted Downside Closing Days	45.892	38	.0001
Actual Downside Closing Days	51.048	38	.0001

This gives the model a success percentage of 84%, which can be of great help to investors seeking to maximize their profits from stock trading.

(2) Validation of Results : As per the requirements of the proposed model, it is necessary to validate the results with the help of various statistical techniques.

➭ **Correlation Analysis :** To test the proposed model ; firstly, correlation analysis was applied on prediction value and actual value of the stock price movement. This is shown in the Table 6. As per the Table 6, it can be concluded that there is no significant difference between the mean and standard deviations of the variables. As per the results of the correlation (Table 7), the correlation between the two variables (0.80) is within the significant limit. Finally, *Z* - test was applied to check the significance between the two variables, and the resultant value was within the permissible limit, and it can be concluded from the Table 8 that the predicted upside closing trend is almost similar to the actual upside closing values. The results have been obtained using IBM SPSS 20.

Discussion

Coyne, Madiraju, and Coelho (2017) created a Twitter dataset and labelled the individual tweets as bullish, neutral, or bearish and assigned the weights according to them. The method proposed in the present study labels the blogs into five categories instead of three to get more insights of the opinions.

Pandarachalil, Sendhilkumar, and Mahalakshmi (2015) worked on assigning polarity to Twitter datasets using SentiWordNet, SentislangNet, and SenticNet. The results showed fair results in polarity assignment to the datasets of 100,000 tweets, but they are not being used accurately in decision making of real - life applications due to lack of confidence in results. The method proposed in the present study achieves high accuracy in polarity assignment, which can be used for decision making involving high investment with higher confidence in the investments.

Alajeku, Okoro, Obialor, and Ibenta (2017) worked on the stock returns of the Nigerian Stock Exchange using investor sentiment analysis. Their results suggested a weak relationship between the investor sentiment and stock returns with low level of confidence in sentiment based information segregation. The model presented in this study provides 84% accuracy in predicting the day end closing prices of stocks, showing an improvement in model design and implementation.

The Table 9 shows the comparison of results of the proposed model with other models. As per the Table 8, it can be concluded that the proposed model has the highest accuracy rate as compared to the studied models, so this model can be applied by investors in maximizing their returns in equity trading. This model can also be applied in different sectors to get better prediction results.

Table 9. Comparison of Research Models with the Proposed Model

S. No	Author Name	Paper title	Results
1	Coyne, Madiraju, & Coelho (2017)	"Forecasting Stock Prices Using Social Media Analysis"	65% accuracy
2	Pandarachalil, Sendhilkumar, & Mahalakshmi (2015)	"Twitter Sentiment Analysis for Large-Scale Data : An Unsupervised Approach"	Only 1 polarity score instead of 3 or 5, low accuracy
3	Alakaju, Okoro, Obialor, & Ibenta (2017)	"Effect of Investor Sentiment on Future Returns in the Nigerian Stock Market"	33.05% accuracy

Conclusion and Implications

The research model aims to predict the day end closing price movement of 39 banking stocks by analyzing the overall community sentiment score from the dataset of user posts posted on www.mmb.moneycontrol.com. The model made 2022 successful predictions out of the total 2418 predictions with a success percentage of 84% and the predictions done by the model are well within the significance limits. The results can be further improved by extending the dataset by considering the sentiment score of the RSS feeds and news articles from relevant sources in addition to the participant's posts. The results of the proposed model can be used by investors in deciding if a particular stock price will increase or decrease so as to decide if the stock is to be held, sold, or more stocks are to be purchased. Correct predictions can help in accurate decisions and can maximize investors' profits.

The research experiment provides insights for researchers, academicians, and corporates. It addresses a vital question regarding how can investors and traders benefit from the vast pools of sentiments stored on blogs and other forms of information hosted on the Internet Researchers and academicians can benefit from this experiment and can build their own models for testing on the upcoming automatic classification and sentiment analysis techniques.

Limitations of the Study and Scope for Further Research

The experiment considered blogs as the only input source of sentiments. Currently, other forms of social media like Twitter, Facebook, news bulletins, etc. have grown exponentially. Application programming interface (APIs) can be embedded in the experiment to create a bigger and varied dataset of opinions covering more population. Blogs have a limited participation and audience, but targeting other social media websites can bring more useful inputs into the experiment.

The proposed research model can be applied to predict the day end closing values of stocks of other equity segments like fast moving consumer goods (FMCG), infrastructure, automobile, etc. to analyze the effect of blog sentiments on stock prices.

References

Alajeku, U. B., Okoro, C.O., Obialor, M. C., & Ibenta, N. S. (2017). Effect of investor sentiment on future returns in the Nigerian stock market. *International Journal of Trend in Scientific Research and Development, 1*(5), 103 - 126.

Bharathi, S., Geetha, A., & Sathiynarayanan, R. (2017). Sentiment analysis of Twitter and RSS news feeds and its impact on stock market prediction. *International Journal of Intelligent Engineering and Systems, 10* (6), 68 - 77. doi : https://doi.org/10.22266/ijies2017.1231.08

Bhardwaj, A., Narayan, Y., Vanraj, P., & Dutta, M. (2015). Sentiment analysis for Indian stock market prediction using Sensex and Nifty. *Procedia Computer Science, 70,* 85 - 91. DOI : https://doi.org/10.1016/j.procs.2015.10.043

Bollen, J., Mao, H., & Zeng, X. - J. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2* (1), 1 - 8. doi : https://doi.org/10.1016/j.jocs.2010.12.007

Chen, H., De, P., Hu, Y., & Hwang, B. - H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies, 27*(5), 1367-1403. DOI : https://doi.org/10.1093/rfs/hhu001

Coyne, S., Madiraju, P., & Coelho, J. (2017). *Forecasting stock prices using social media analysis*. 2017 IEEE 15th International Conference on Dependable, Autonomic and Secure Computing, 15th Intl Conference on Pervasive Intelligence and Computing, 3rd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), 1031-1038. DOI: 10.1109/DASC-PICom-DataCom-CyberSciTec.2017.169

Gialampoukidis, I., Kalpakis, G., Tsikrika, T., Papadopoulos, S., Vrochidis, S., & Kompatsiaris, I. (2017). Detection of terrorism-related twitter communities using centrality scores. *Proceedings of the 2nd International Workshop on Multimedia Forensics and Security- MFSec, 17,* 21 - 25. DOI : https://doi.org/10.1145/3078897.3080534

Mangala, D., & Sharma, M. (2014). A brief mapping of theory and evidence of investors' behavioural biases. *Indian Journal of Finance, 8* (8), 44 - 56.

Moneycontrol. (n.d.a.). *SBI stock discussion forum online India*. Retrieved from https://mmb.moneycontrol.com/forum-topics/stocks/sbi-406.html

Moneycontrol. (n.d.b.). *Moneycontrol - State Bank of India*. Retrieved from https://www.moneycontrol.com/india/stockpricequote/banks-public-sector/statebankindia/SBI

Öztürk, N., & Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telematics and Informatics, 35*(1), 136 - 147. DOI : https://doi.org/10.1016/j.tele.2017.10.006

Pandarachalil, R., Sendhilkumar, S., & Mahalakshmi, G.S. (2015).Twitter sentiment analysis for large-scale data : An unsupervised approach. *Cognitive Computation, 7* (2), 254 - 262.

Patino, A., Pitta, D. A., & Quinones, R. (2012). Social media's emerging importance in market research. *Journal of Consumer Marketing, 29* (3), 233 - 237. DOI : https://doi.org/10.1108/07363761211221800

Peel, L., Larremore, D. B., & Clauset, A. (2016). The ground truth about metadata and community detection in networks. *Science Advances, 3* (5). DOI : https://doi.org/10.1126/sciadv.1602548

Perez - Liston, D., & Huerta, D. (2012). Does investor sentiment affect Mexican stock market returns and volatility ? *The Global Journal of Finance and Economics, 9* (2), 121-132.

Qiu, L., Rui, H., & Whinston, A. (2013). Social network - embedded prediction markets : The effects of information acquisition and communication on predictions. *Decision Support Systems, 55* (4), 978 - 987. DOI : https://doi.org/10.1016/j.dss.2013.01.007

Ranjan, S., & Sood, S. (2018). Analyzing social media community sentiment score for prediction of success of bollywood movies. *International Journal of Latest Engineering and Management Research, 3* (2), 80 - 88.

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge - Based Systems, 89* (6), 14 - 46. DOI : https://doi.org/10.1016/j.knosys.2015.06.015

Ren, J., Yeoh, W., Shan Ee, M., & Popovič, A. (2017). Online consumer reviews and sales: Examining the chicken-egg relationships. *Journal of the Association for Information Science and Technology, 69* (3), 449 - 460. DOI : https://doi.org/10.1002/asi.23967

Rout, R. K., & Das, N. (2015). Behavioral prospects of individual investor decision making process : A review. *Indian Journal of Finance, 9* (4), 44 - 55.

Saumya, S., Singh, J. P., & Kumar, P. (2016). *Predicting stock movements using social network*. 15th Conference on e-Business, e-Services and e-Society (I3E), Sep 2016, Swansea, United Kingdom (pp. 567 - 572). Springer International Publishing, Lecture Notes in Computer Science, LNCS - 9844.

State Bank of India. (n.d.). *Stock price, share price, live BSE/NSE*. Retrieved from https://www.moneycontrol.com/india/stockpricequote/banks-public-sector/statebankindia/SBI

Suresh, P. S., & George, S. (2016). Market sentiment dynamics and return volatility in the Indian equity market. *Indian Journal of Finance, 10* (6), 7 - 23.

Swain, P. K., & Dash, M. (2017). A literature review on investors' perception towards mutual funds with reference to performance, risk-return, and awareness. *Indian Journal of Finance, 11*(12), 60 - 70.

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

Wang, Y. (2017). Stock market forecasting with financial micro-blog based on sentiment and time series analysis. *Journal of Shanghai Jiaotong University (Science), 22* (2), 173-179.DOI : https://doi.org/10.1007/s12204-017-1818-4

Wu, Z., & Lu, Y. (2017). A study on micro-blog sentiment analysis of public emergencies under the environment of big data. In *Proceedings of the 29th Chinese Control and Decision Conference, CCDC 2017* (pp. 4435-4438). DOI : https://doi.org/10.1109/CCDC.2017.7979279

Xiaomei, Z., Jing, Y., Jianpei, Z., & Hongyu, H. (2018). Microblog sentiment analysis with weak dependency connections. *Knowledge-Based Systems, 142,* 170 - 180. doi : https://doi.org/10.1016/j.knosys.2017.11.035

Zhao, B., He, Y., Yuan, C., & Huang, Y. (2016). *Stock market prediction exploiting microblog sentiment analysis.* In 2016 International Joint Conference on Neural Networks (IJCNN), pp. 4482 - 4488. DOI: 10.1109/IJCNN.2016.7727786

Note

The contribution made by the authors to this paper is as follows :

Mr. Sandeep Ranjan : 30%
Dr. Inderpal Singh : 30%
Dr. Sonu Dua : 30%
Dr. Sumesh Sood : 10%

About the Authors

Sandeep Ranjan is registered as a Research Scholar in I K G Punjab Technical University, Kapurthala and is working as an Assistant Professor in Lyallpur Khalsa College of Engineering, Jalandhar, Punjab. He has B. Tech in Computer Science & Engineering and M. Tech in Information Technology (distinction and AIR 8th) and has been an International Doctoral Visiting Fellow in Europe. He has a total experience of 13 years and has 10 publications in international journals and conferences.

Dr. Inderpal Singh is presently working as an Associate Professor & Dean Research in the esteemed Institution, Khalsa College Lyallpur Institute of Management & Technology, Jalandhar, Punjab since July 2010. He has to his credit more than 20 Books published in the field of commerce and management, bio-technology, & research. He has published more than 60 research papers presented in national & international conferences and has published 40 papers in national & international journals including Scopus indexed journals. He has published six papers in Scopus indexed journals. He was a gold medalist in M.Phil & has been awarded the "Exemplary Educationist Award" by Indo Global Chamber of Commerce, Industries, & Agriculture and "Best Faculty Award" by Indian Society of Technical Education (ISTE) for academic excellence and Outstanding Educator Award from I2OR Australia.

Dr. Sonu Dua is presently working as an Assistant Professor in the esteemed institution Khalsa College Lyallpur Institute of Management & Technology, Jalandhar, Punjab. He has completed his doctoral degree in the field of management. He has 11 years of experience (both academics and industry). He has to his credit 10 national & international publications including Scopus indexed journals, one book, and 11 chapters.

Dr. Sumesh Sood is presently working as an Assistant Professor in I K G Punjab Technical University, Dinanangar Campus, Kapurthala, Punjab. He has done his Ph.D. in Computer Science from Himachal Pradesh University, Shimla in the field of Software Engineering. He has more than 18 years of experience in teaching and administration and has published 35 research papers in international journals and conferences. He has successfully supervised two Ph.D. candidates.