Prediction of Stock Prices Using Unstructured and Semi-structured Qualitative Data – A Neural Network Approach

Muthusamy Palani Rajakumar[1]*   Ramya Jegatheesan[1]   Rashmi Chandy[1]   Tanusshri Sampath[1]

[1]*Department of Computer Science and Engineering,
St. Joseph's College of Engineering, Chennai, India*
* Corresponding author's Email: rajranjhu@gmail.com

Abstract: As more interest by the people about stock markets is growing substantially bigger, the more their thinking is directed to a systematic method to predict stock prices that vary. There is an increasing appreciation that unstructured qualitative data contain awfully precious information, helpful for a variety of reason. It is this motivation that has fascinated awareness of researchers and specialized into the interesting and demanding area of stock market prediction using qualitative news and events. In this paper, a framework for stock price movement prediction using news, company newsletter and financial information was presented. Financial information and events were processed and multilayer feed forward network trained using back propagation was employed to empirically explore the intricate relationships between events and the stock market prices. Among the various industries we chose to focus on the Information Technology (IT) sector since it is most prevalent and applicable to the Indian markets. The stock prices and the relevant unstructured information for six IT companies were collected and filtered based on set of keywords. Our research has proven that the stock price can be accurately predicted from news articles by furnishing the empirical results that were obtained after training the multilayer neural networks.

Keywords: Unstructured qualitative data, Stock price prediction, Artificial neural network, Information technology sector, Filter process.

1. Introduction

Stock market, an icon of business world is becoming the core concern of today's business phenomenon. The state of an economy mostly relies upon the financial position of the economy wherein stock market plays an important role in this area. Stock market connects buyers and sellers as per their requirement by providing cozy atmosphere to store money for future. It spreads risks on investments among people.

Stock market is affected by many macro-economic factors such as political events, general economic conditions, firm's policies, investors' expectations, institutional investors' choices, movement of other stock markets and the psychology of investors etc some of which influence stock prices straightly and others that do so implicitly. Stock prices are driven by the conduct of human investors by using available information. News plays a vital role for investors when deciding fair stock prices. News may be positive, negative or might not have any impact at all. Accessibility to news data was impractical until the start of the most recent decade but nowadays companies publish their respective news to the public through press releases, news articles and also disclose information in their official websites. Sometimes the information conveyed in a press release is significant enough for explicit price change.

The majority of the company information in the present day does not exist in a regular relational database. An often-cited statistic is that 80% of industry data is unstructured, be it in word processor, spreadsheet and PowerPoint files, audio, video, sensor and log data, or external data such as social media feeds shown in Fig. 1.
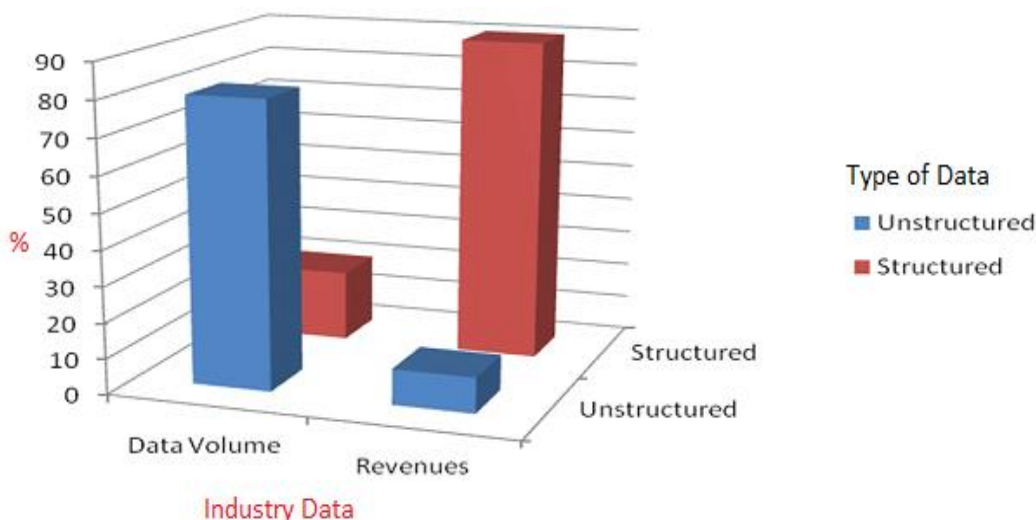
Figure. 1 Structured data vs. unstructured data

Data mining is referred to as the method of realizing important patterns in usually well-structured data with the goal of attaining economic gain. On the other hand, text mining otherwise called natural language processing is the method of scrutinizing unstructured data (word processor, spreadsheet and PowerPoint files, audio, video, sensor and log data, or peripheral data such as social media feeds), mining significant information, usually large quantities and then converting that information into ordered information that can leveraged in many ways such as integration, visualization, querying etc.

The unstructured data and structured data were a significant driver in the decision making procedure in daily buy and sell activity. A general indication of the decision making process of a trading activity starts from getting news reports and ends to building the buy, hold and sell investment strategy decision. News announcements and corporate actions are captured through different number of sources like the company database, newsletter, press, radio, television and internet. Example: news from firms, their preparations, economic information (EPS, PE), earnings potentials expectations, shares, returns, financial statement and financial comments. Agents or individual makes a selection of exciting news messages. An investigation of each message is done, taking into consideration factors such as financial knowledge, past experience, background and cognitive biases. Based on this investigation a decision is made, taking into consideration of tendency to risk and the intention of trading. Finally the results or outcome are evaluated and order is recommended to buy, hold or sell. The decision-making process is shown in Fig. 2.

Companies that expect much better rate of return in data that use more unstructured or semi structured data than do companies expecting lower rate of return. Unstructured text data has hastily materialized as a starting place of functional information ranging from news articles to individual judgments. It hoists some issues for industry that require to process and take advantage of ill-defined data to develop considerable information wisely. The study investigates prediction of stock market price using the unstructured intelligent information. This process is done by automatic analysis of these news articles.

In order to provide better forecasting, hybrid approach method is developed for forecasting the stock prices. In addition to quantitative data, financial news and news articles were analyzed for forecasting the share price using neural network. From our research, we acknowledged certain qualitative terms which had positive or negative influences on stock price. Some of these terms were frequently appearing in several different representations, which conveyed that articles containing these terms were vulnerable to price movement. The main advantage of this proposed system is that the stock prices were predicted with minimal error or deviation from the actual stock price that was fed to the neural network. We were also able to identify those new events that caused the stock price to fluctuate significantly.

The rest of the paper is organized as follows. Section 2 describes previous works that is closely related with this study. Section 3 explains prediction model. Section 4 presents the results and compares it with actual price movement. Finally the conclusions of the study are summarized in section 5 along with the hope for future work.
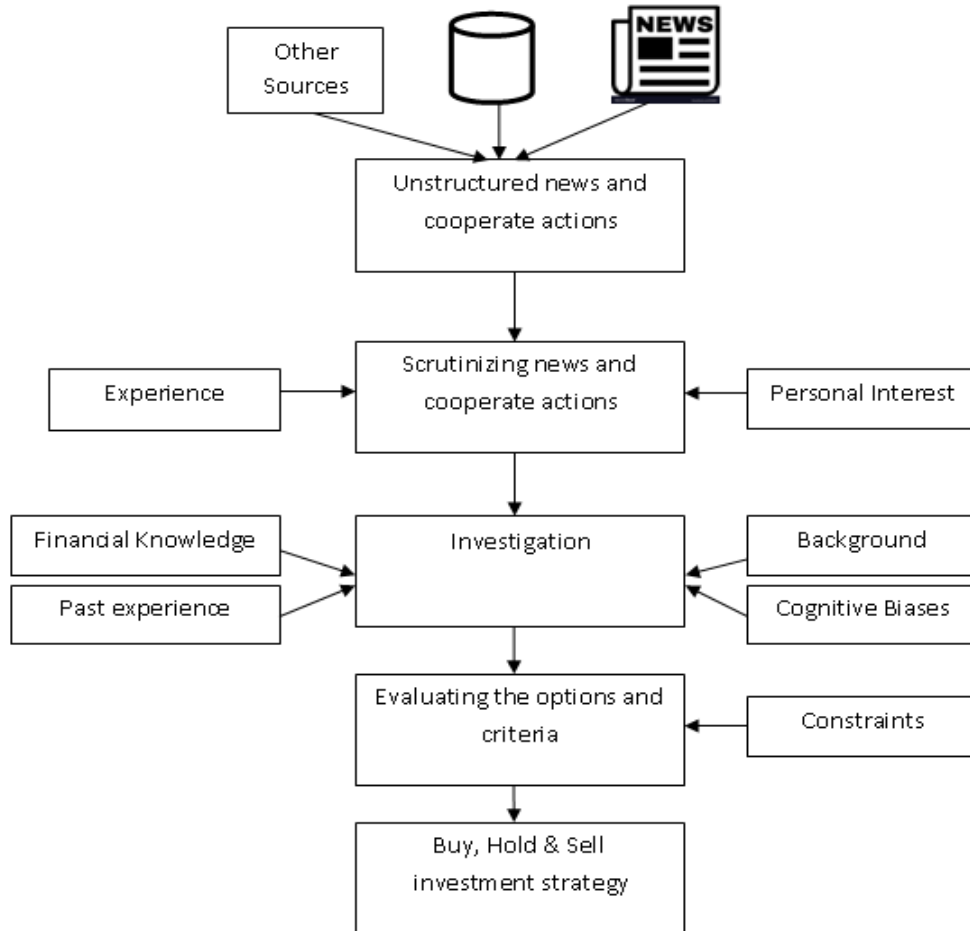
Figure. 2 Decision making process

2. Related works

A forecasting model of stock price movement by analyzing verbs in financial news and articles has built. The AZ Fin-Text system was constructed to analyze the data and implemented using the Support Vector Regression sequential minimal optimization function through weka [1]. The five parts of speech textual representations semantically meaningless stop words, noun phrases, extension of noun phrases, intermediate between named entities and noun phrases and verbs in economic news articles and their force on stock market prices were analyzed [2].

This paper proposes a novel method for forecasting the intraday stock price of Taiwan stock market, valuable for short-term investors [3]. This model is established using Support Vector Machine classifier and rough set theories after the unstructured data is converted into structured information. The experimental results declared that this model predicts the growth and decrease of a stock price within very short period after the news is distributed.

This paper determines how the Dow Jones Industrial average unstructured news information along with the historical quotes can effectively influence the closing index for a given trading day. Kea algorithm was employed to extract key phrases automatically from news articles [4]. The model used one input layer consisting of 25 neurons, one hidden layer consists of 30 neurons and one output layer.

An automated text approach is presented to aggregate news stories from diverse sources and create news corpus [5]. The news corpus is filtered down to relevant sentences and analyzing using natural language processing tools. A sentiment metric is proposed as a measure of sentiment of the overall news corpus. The study used a variety of open source packages and tools to build up the news collection and sentiment evaluation engine.

This paper investigated the relationship among the business information content on internet, sentimental value information on internet and share price [6]. Text sentiment algorithm based on HowNet sentiment dictionary was proposed to

forecast the stock price in the Chinese stock market. The work used the Support Vector Machine regression function to exploit the rapport between financial markets stock prices and financial news sentiments. The investigations indicate that in stock price forecasting the information sentimental value is more significant than information volume.

A stock price forecasting model that combines financial indicators such as price to earnings ratio, price to book value, book value per share, dividend payout ratio with the indicator obtained for the textual annual reports of the firms in the United States was presented [7]. Multilayer Perceptron Neural Network, radial basis function neural network and Support Vector Regression approach were used to forecast the stock price. The results shows that Support Vector Regression and neural network better coped with the growing complexity of the forecasting problem than linear regression model.

This paper investigates the importance of financial and textual information in the short-term investment for stock price movements up, down and stay [8]. The experimental model is trained using random forest classifier, logistic regression and multilayer perceptron using 200 trees. The result shows that using texts boosts prediction accuracy over ten percentage. This technique cannot form the basis of a viable trading strategy, transaction costs, slippage and borrowing costs.

A method to analyze Korean language news in order to predict the rises and falls on the Korea composite stock price index is introduced [9]. The method consists of carrying out the natural language processing of news article, describing the features, categorizing and extracting the sentiments and opinions and identifying the correlation between news and stock market fluctuations.

This paper explores the unique method to incorporate the sentiments in social media for the estimation of stock value trends [10]. Moreover, topics and interrelated sentiments are spontaneously take out from the texts in a message board using supervised machine learning techniques. This paper shows textual investigation and machine learning technique to a greater extent, widespread over time in recent times owing to the increasing need to handle lots of texts from firm specific news inflows and Securities and Exchange Commission filings, etc using Neural network [11].

Text mining and sentiment analysis on Chinese online financial news, to predict Chinese stock values based on Support Vector Machine was proposed [12]. The Support Vector Machine technique constructed on text mining technology

displays an outstanding end result in forecasting stock market particularly when predicting the specific share value.

This paper focuses on the correlation between market value, sentiments of tweets and really simple syndication news feeds [13]. The system discovers the task of involuntary identification of news beliefs with the help of RSS news feeds and calculates the stock market measurement whether goes up or down [14].

Conventional models widely utilized to forecast stock markets include autoregressive, moving average, auto-regressive moving average, autoregressive integrated moving average, generalized auto-regressive conditional heteroscedasticity and stochastic volatility. These statistical models are capable of modeling only linear relationship between factors that influence the market and value of the market. The other problems with these models [3, 6, 12] are requires long training time, saving lots of past data points: at least the N periods used in the moving average computation, lags behind a trend and ignores complex relationships in data. In general, conventional stock market prediction methods which deals with precision, certainty and rigidity. Information regarding a stock is normally incomplete, complex, uncertain and vague.

3. Proposed system

Qualitative forecasting techniques are not merely guesswork but they are well-developed structured approaches to get hold of good forecasts. It is evident that the stock market prediction can be more clear and concrete when the information hidden in the reports are depicted in a more descriptive manner. Therefore, in this paper qualitative model was developed for forecasting share price with the help of corporate actions and publically available information.

An immense amount of original information associated to business appears persistently, with instantaneous force on stock prices. Analyzing unstructured data and converting into actionable understanding is unavoidable. Scrutinizing such information in real time is essential for both trading establishment and individual shareholder. The work mainly develops an intelligent decision support system which measures all the qualitative events that may influence the stock market. It captures many global and national events such as political actions, common economic conditions, firm's strategies, investors' potential, institutional investors'
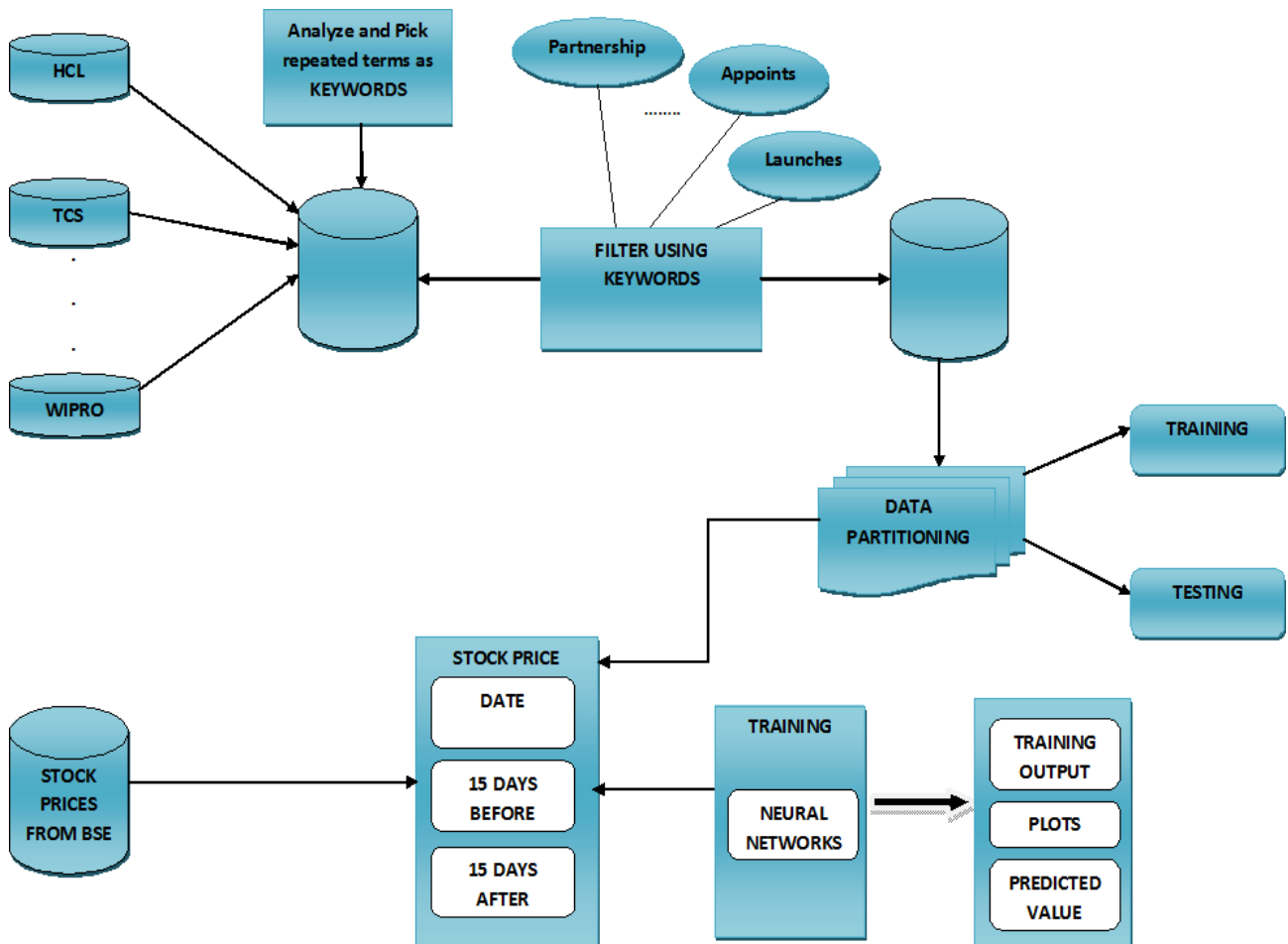
Figure. 3 System architecture for proposed work

preferences, progress of other stock markets and the mind and behavior of investors.

The system for supporting decision making to invest, whereby substantial news articles, unstructured big data are collected, parsed, tagged, analyzed and converted to opinions appropriate for making prediction. The system framework consists of gathering unstructured information, classification of news message, locating the key instances from the qualitative database, normalization, network training and validation of qualitative news shown in Fig. 3.

The various forecasting models focused the various qualitative information such as historical quotes, financial verbs, sentimental values, firm specific news. The forecasting model [1, 8, 9] discussed the verbs in financial news, [2] referred five parts of speech textual representation, [3] referred the publication of press releases, [4] stated historical quotes, [5, 6, 10-14] discussed sentimental value information, [7] referred financial indicators for the annual reports and model [11] referred firms specific news. Our forecasting model concentrates the most frequently occurring words in information technology related companies.

3.1 Data collection

In our study, we considered news articles for each of the six IT companies namely, Tata Consultancy Services (TCS), Wipro, Infosys, Mindtree, Mphasis and Tech Mahindra for duration of about ten years from December 2, 2006 to December 2, 2016. We stored the news information, such as when it was published, the headline, and the content of the article in a database. Similarly, Ten years stock prices of the six IT Companies were fetched from Bombay Stock Exchange (BSE). Another database was created and the stock prices were stored in it for all the six IT companies. The study divided the data into nine years of news articles from December 2, 2006 to December 2, 2015 as the training data set and into one year of news articles from December 3, 2015 to December 2, 2016 as a testing data set. Testing data set enables to verify the findings obtained from the training data set. The second database lists the company name, date and its information. The sample is shown below.

Table 1. News information of the Company TCS

TCS	15/03/2016	TCS launches peer-to-peer mobile payment solution with DNB, a leading bank in Norway
TCS	03/09/2016	TCS recognized as a Leader in Manufacturing PLM Strategic Consulting by IDC MarketScape
TCS	03/01/2016	TCS and GE Enter Strategic agreement to Expand the Industrial Internet through Digital Reimagination
TCS	15/02/2016	TCS launches Voice over LTE solution
TCS	02/02/2016	TCS recognized as the world's most powerful brand in IT Services
TCS	21/12/2015	TCS recognized as a Leader in Application Testing Services for Second Consecutive Year by Gartner
TCS	23/11/2015	TCS and Gfi Informatique Sign a Strategic agreement to Accelerate Smart Cities Programs in France
TCS	17/11/2015	TCS launches Center of Excellence (CoE) for IoT Solutions in Hyderabad in collaboration with Intel in India
TCS	20/10/2015	TCS recognized as a Star Performer and Leader in Global Capital Markets Application Outsourcing by Everest Group
TCS	29/09/2015	Nielsen appoints ignio™–TCS' neural automation platform to transform enterprise technology operations
TCS	31/08/2015	TCS launches Japan-centric Delivery Center in Pune
TCS	15/06/2015	TCS launches Center of Excellence (CoE) for Next Gen Technology Solutions
TCS	27/05/2015	TCS partnership with The Royal College of Art to establish UK Design Innovation Lab
TCS	21/04/2015	TCS and Solace Systems sign agreement to Deliver Digital Reimagination™ Solutions and Services

3.2 Data analysis

The data collected is huge. In order to extract useful information the dataset is analyzed to find the frequently occurring words which are taken as keywords listed in Table 2.

These keywords were chosen as they had a significant impact on the stock market prices. When the news article containing the keyword occurred the stock prices were found to fluctuate which was a found to be applicable across the six Information Technology companies. For example, consider the following news articles of Mindtree Company as shown in Table 3.

The stock price 15 days before and 15 days after the date the news event occurred was examined as shown in Fig. 4. It was seen that the stock price had either increased or decreased which is evident from the bar chart below. These fluctuations where also

observed in the other IT companied hence these nine words were considered as keywords for data collection process.

Table 2. List of keywords

ID	Keyword
K1	Recognized
K2	Partnership
K3	Acquires
K4	Appoints
K5	Won
K6	Launches
K7	Agreement
K8	Awarded
K9	Certified

Table 3. List of news for Mindtree company

|	
10/08/2015	Mindtree **recognized** amongst the Top 10 Outsourcing Service Providers in USA and EMEA by ISG
16/09/2015	Flydubai is in **partnership** with Mindtree as a strategic technology partner
18/01/2016	Mindtree **acquires** Magnet 360, a Salesforce Platinum Consulting Partner
21/05/2014	Mindtree **appoints** Manisha Girotra, CEO of Moelis & Company, India to its Board of Directors; elevates Rostow Ravanan to Executive Directo
14/01/2014	Mindtree **won** the 'Most Promising Company of the Year' Award from CNBC TV18
13/01/2014	Mindtree **launches** two new solutions at NRF's "Big Show"
08/02/2016	Mindtree signs service integration capabilities **agreement** with Sandvik
06/03/2014	Mindtree **awarded** #1 by Euromoney in the 'Best Managed Company for 2013' under two award categories
26/09/2013	Mindtree **certified** among the top global product engineering service providers by Zinnov research and analysis
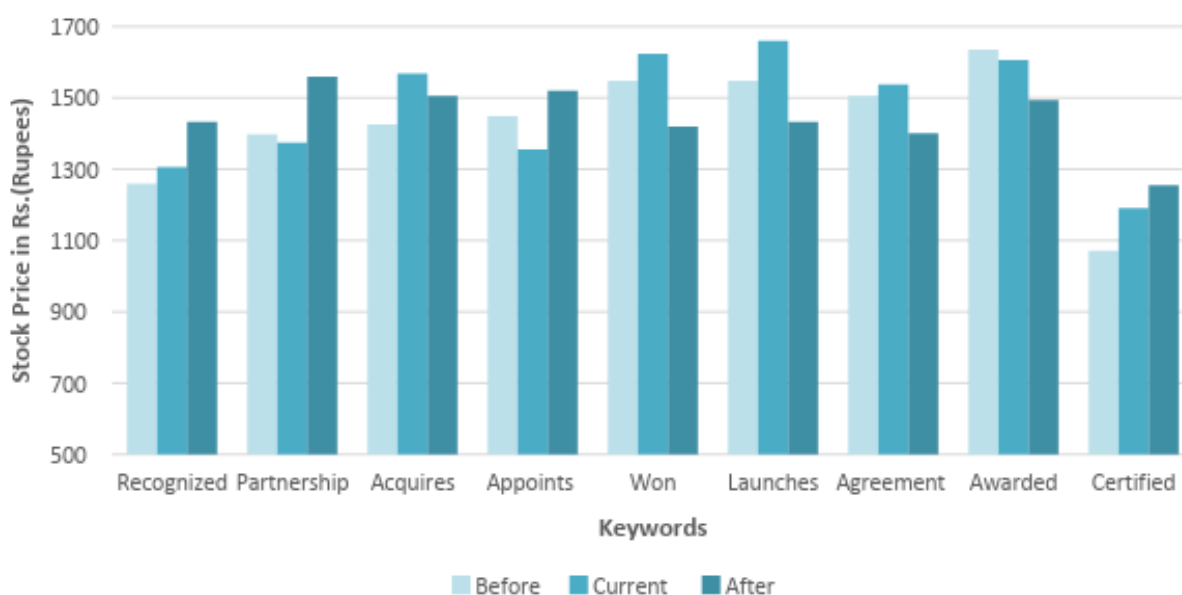
Figure. 4 The fluctuation of Mindtree's stock prices seen for the news events containing the keywords

Table 4. Filtered news for the various keyword

keyword	Company	Date	News
recognized	Infosys	08/11/2016	Independent Research Firm recognized Infosys as a Leader among Services Providers for Next-Generation Oracle Application Projects
	TCS	20/9/2016	TCS recognized for leading CSR practices
partnership	Tech Mahindra	05/06/2014	Informatica and Tech Mahindra Announce Global partnership to Transform Data Management Services
	TCS	3/9/2009	TCS enters into a partnership with Infineon Technologies AG.
acquires	Wipro	30/4/2012	Wipro acquires Analytics Co, Promax Applications Group
	Mindtree	22/6/2016	Tech Mahindra acquires leading Digital Change Agents

3.3 Data filtration

Data filtration is the process of focusing on data according to the user's requirements and ignoring the unwanted information. Redundant or impartial pieces of data can confuse or disorient a user. Filtering data can also make results more efficient. Thus in our study filtering is done based on the set of keywords for each of the company shown in Table 4.

3.4 Training

Training a network involves presenting input patterns in a way so that the system minimizes its error and improves its performance. The most prevalent network architecture for artificial neural networks is a multilayer feed-forward network trained using back-propagation.

The Pseudo code for the training module is given below
1. Create two variables namely input and target in the workspace.
2. Load the stock price values for a particular news event from 15 days before to 15 days after the occurrence of the news into the input and target values.
3. Transpose the input value IV: = (IV)-1 and target values as TV: = (TV)-1.
4. Open Network Data Manager.
    4.1 Import input and target variables from the source workspace.
    4.2 Map the input variable as Input data and target variable as Target Data respectively.
5   Create network
    5.1 Decide the number of layers needed and number of neurons for each layer.
    5.2 Change training parameters such as Training function and Adaptation learning function.
6   Train the network and Simulate network outputs.
7   Repeat the step 5 by varying the number of layers, neurons in each layer and training parameters until best performance is found.
8   Save the trained network and then use the predicted values to calculate the error indexes.

3.5 Training results

Furthermore, training has been done for all the six companies and results for the TCS Company have been tabulated in Table 5. The result which gave the best performance have been highlighted below for each keyword.

There are three plots that can be evaluated at the end of training the dataset namely

- Performance plot
- Training plot
- Regression plot.

The Fig. 5 shows the Mean Square Error (MSE) equals 1.2077×10^{-6} at the end of the training phase, which is very low. This means that the neural network is well trained by which target output and ANN's output for the training set are very close to each other. The validation and test curves are very similar. The test curve has not increased significantly before the validation curve increased, therefore over fitting has not occurred.

The Figs. 6 (a), (b), and (c) show the variation of the gradient coefficient with reference to the number epochs. The training process was made more accurate by either increasing or decreasing the number of neurons in each layer and testing the network for best performance. The final value of the gradient coefficient is 0.0019573 which is approximately near to zero. The minimum value of the gradient coefficient shows that the network has successfully predicted values close to the target values.

The Fig. 7 is a regression plot that is obtained as a result of training the neural network; it is used to validate the network performance. Figs. 7 (a), (b), and (c) represent the training, validation, and testing data respectively while 7 (d) shows the combined dataset. The above regression plot displays the network outputs with respect to targets for training, validation and test sets. The data falls along a 45 degree line which shows that the network outputs are equal to the targets. Here R value is the correlation coefficient which is greater than 0, it indicated positive association. When the R value is equivalent to 1 it shows that all the data points are included in the line of best fit.

4. Performance analysis

To forecast the market movement, new unstructured data like news messages will be transformed into the representation selected previously for training and then matched to existing representations of news in the teaching data. This process decreases the multifaceted problem into a much simpler classification problem is shown Fig. 8.

Consider an event for the company TCS and keyword "partnership". The date the event occurred on is 21/07/2014. The stock price on the date of the

Table 5. Training results for the various keyword (company: TCS)

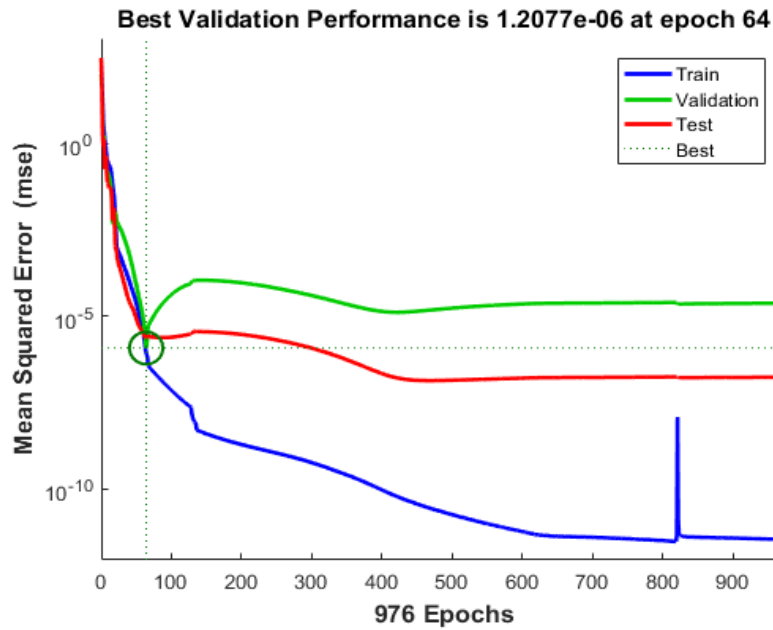| Keyword	No of Layers	Hidden Layers	No of neurons	Training Function	Transfer Function	Performance
Recognized	3	1	7	TRAINCGB	TANSIG	0.0017
		2	6			
	3	1	8	TRAINCGP	TANSIG	0.00978
		2	7			
	3	**1**	**8**	**TRAINRP**	**TANSIG**	**0.00022**
		2	**7**			
Partnership	**3**	**1**	**7**	**TRAINLM**	**TANSIG**	**8.7012×10⁻¹⁰**
		2	**8**			
	3	1	8	TRAINRP	TANSIG	0.0052
		2	7			
	3	1	7	TRAINBR	TANSIG	0.000569
		2	6			
Acquires	3	1	10	TRAINR	TANSIG	0.00293
		2	9			
	3	1	8	TRAINGDX	TANSIG	0.000901
		2	7			
	4	**1**	**7**	**TRAINCGF**	**TANSIG**	**0.000239**
		2	**6**			
		3	**5**			
Appoints	**3**	**1**	**11**	**TRAINBR**	**TANSIG**	**0.00014**
		2	**10**			
	3	1	11	TRAINCGB	TANSIG	0.00046
		2	10			
	3	1	10	TRAINLM	TANSIG	0.000722
		2	9			
Won	**3**	**1**	**6**	**TRAINBR**	**TANSIG**	**2.4875×10⁻⁹**
		2	**5**			
	3	1	11	TRAINLM	TANSIG	0.0099
		2	10			
	3	**1**	**8**	TRAINCGB	TANSIG	0.0018
		2	**7**			
Launches	3	1	6	TRAINCGF	TANSIG	0.00592
		2	5			
	3	**1**	**8**	**TRAINBR**	**TANSIG**	**1.2825×10⁻¹²**
		2	**7**			
		3	**6**			
	3	1	8	TRAINCGP	TANSIG	0.000632
		2	7			
		3	6			
Agreement	3	1	7	TRAINLM	TANSIG	0.0022
		2	6			
	4	**1**	**7**	**TRAINLM**	**TANSIG**	**0.00037**
		2	**6**			
		3	**5**			
Awarded	**3**	**1**	**9**	**TRAINBR**	**TANSIG**	**1.8199×10⁻⁸**
		2	**8**			
	4	1	9	TRAINLM	TANSIG	0.0026
		2	**8**			
		3	**7**			
Certified	**4**	**1**	**9**	**TRAINLM**	**TANSIG**	**7.522×10⁻⁷**
		2	**8**			
		3	**7**			
	4	1	9	TRAINGDA	TANSIG	0.0001
		2	**8**			
		3	**7**			

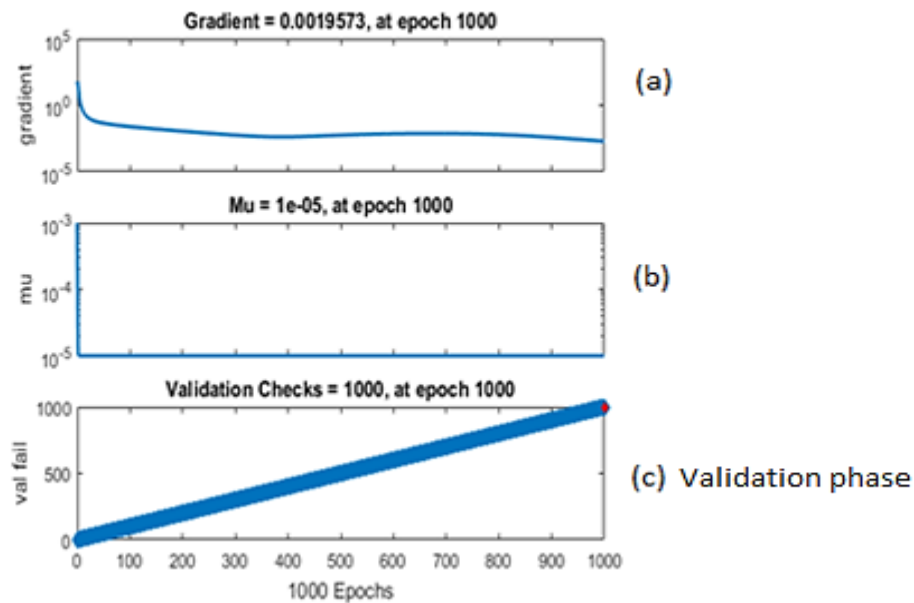Figure. 5 Performance plot for Mphasis keyword "awarded"

Figure. 6 Training plot for TCS keyword "partnership"

event was considered along with the stock price for the company 15 days before and 15 days after. The data was fed into the neural network as input and target data. The training and transfer functions for the layers are set and the input is trained. The network learns and adapts to the dataset to give accurate predictions. This process is repeated till we get a result with best performance.

The training output of the network with best performance is noted down and various performance metrics such as MAPE, MAE, MSE, and RMSE are applied. The Fig. 9 shows the actual and predicted

output for the company TCS's event with keywords as partnership.

For any modeling method, it is essential to illustrate the performance criteria, and to reveal how performance of the model will be measured. The performance measurement traditionally involves calculation of errors between the actual and the predicted result. Some of the traditional performance measurements are Mean Absolute Percent Error (MAPE), Mean Squared Error (MSE), Root Mean Square (RMS), and Mean Absolute Error (MAE) can be used to calculate the difference
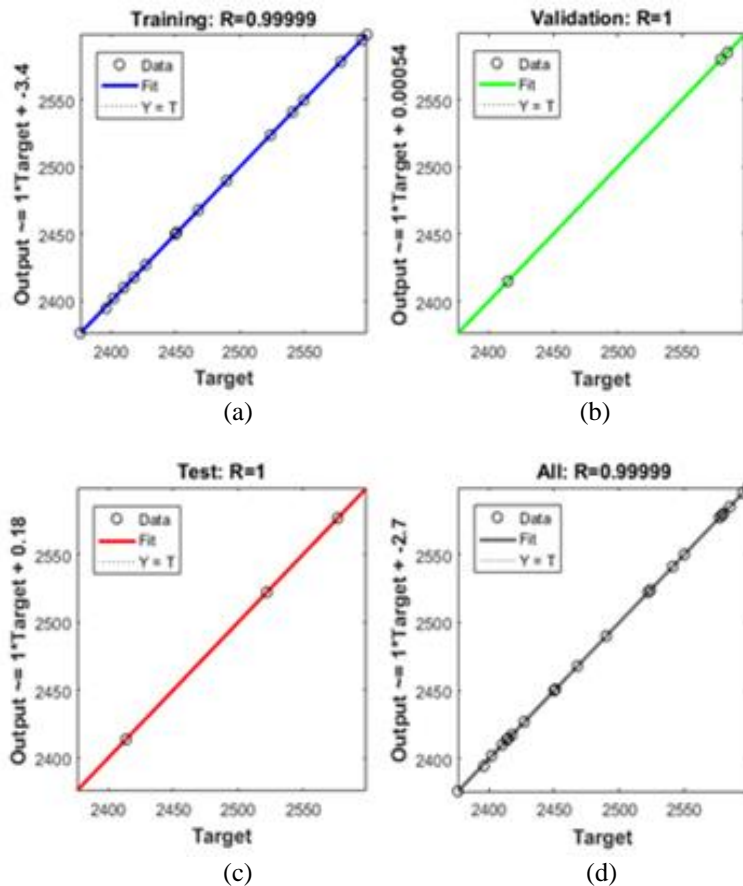
(a)   (b)

(c)   (d)

Figure. 7 Regression plot for TCS keyword "partnership": (a) training, (b) validation plot, (c) test plot, and (d) combined plot
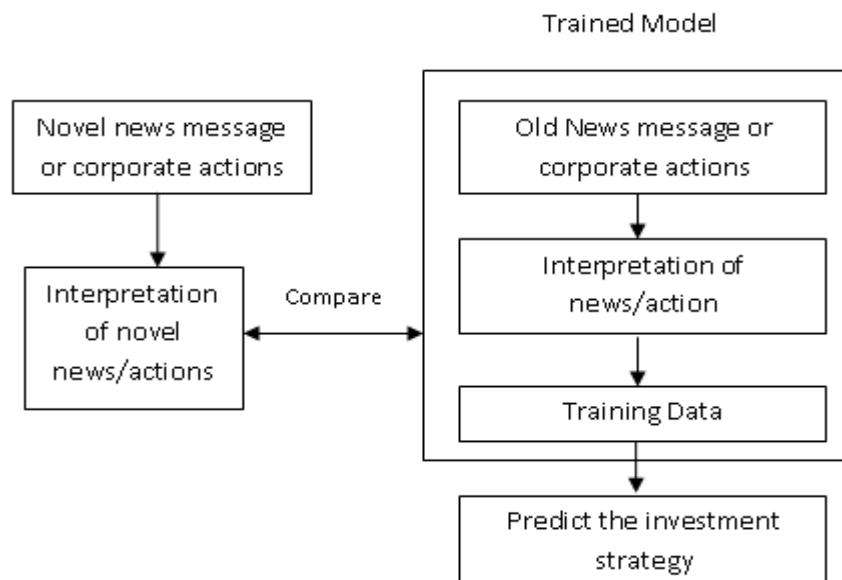

Figure. 8 Prediction model

in the actual and predicted values. The actual result is denoted by s_i, model result or predicted result is denoted by o_i and n is the total number of data points. The error index calculation between the actual and predicted values for keywords in Table 2 is shown in Table 6. It also noticeable that error metrics for the keyword "agreement" (K7) of the company Mphasis is empty because there were no news articles found for the given duration with that keyword.
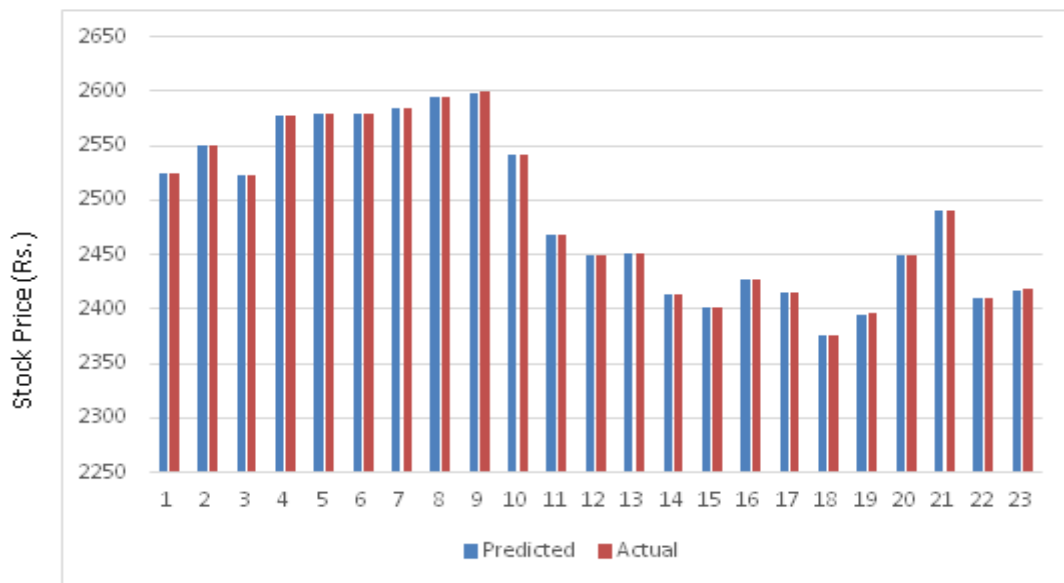
Figure. 9 Graphical representation of actual vs. predicted values of keyword "partnership"

Table 6. Error metrics for various keywords

Error Index	ID	WIPRO	TCS	MPHASIS	MINDTREE	TECH MAHINDRA	INFOSYS
MAPE	K1	0.000130793	9.82487×10^{-5}	0.003205383	0.000404	0.001598295	3.36691×10^{-5}
	K2	0.00010266	3.38774×10^{-5}	0.000565689	0.000160932	1.38451×10^{-5}	0.00064262
	K3	1.0115×10^{-5}	0.000523346	0.00017272	6.17586×10^{-7}	0.000250765	6.04793×10^{-5}
	K4	0.000194776	1.04605×10^{-6}	0.000468779	0.000545175	0.000102133	0.000144778
	K5	0.000602	0.000132	0.000263	5.69×10^{-5}	0.003772	3.6729×10^{-6}
	K6	0.002934	0.000146	0.000188	1×10^{-5}	0.000177	3.43099×10^{-5}
	K7	0.002934	0.000146	——	1×10^{-5}	0.000177	3.43099×10^{-5}
	K8	6.74319×10^{-5}	5.91748×10^{-5}	0.000686382	0.000187572	0.002547048	0.000121071
	K9	0.000505	0.000338	0.000235	0.000326	9.96291×10^{-5}	4.63075×10^{-5}
MSE	K1	0.047196	0.312679	35.0367	2.649913	135.8879366	0.058811043
	K2	0.00554909	0.104655059	0.909300615	1.302966411	0.001601464	95.17257699
	K3	0.00025781	0.515959999	0.043476398	1.43696×10^{-6}	4.51821273	0.068734974
	K4	0.289685813	3.81686×10^{-5}	0.22168555	2.3553258	0.123962924	0.460262924
	K5	1.75925	0.36083	0.136938	0.051384	784.3311	0.002753
	K6	7.521186	1.395626	0.054369	0.000254	0.27711	0.065533
	K7	0.196049	0.691202	——	0.002717	4.357271	2.807872

	K8	0.013684	0.047705786	0.308055	0.98995427	86.39131082	3.61567847
	K9	0.479918	5.817229	0.263343	0.668223	0.558187	0.026908
RMS	K1	0.217246	0.559177	5.919181	1.627855	11.6571	0.24251
	K2	0.074492211	0.323504342	0.953572553	1.141475541	0.0400183	9.755643341
	K3	0.016056463	0.718303556	0.208509946	0.001198731	2.125608791	0.262173557
	K4	0.538224686	0.006178077	0.47083495	1.534707073	0.352083689	0.678426801
	K5	1.326367	0.600691	0.370051	0.22668	28.00591	0.052469
	K6	2.742478	1.181366	0.233171	0.015937	0.526412	0.255995
	K7	0.442774	0.831386	_____	0.052122	2.087408	1.675671
	K8	0.116977	0.218416542	0.555027	0.994964457	9.294692616	1.901493747
	K9	0.692761	2.411893	0.51317	0.817449	0.747119	0.164037
MAE	K1	0.069917	0.229974	1.3026	0.520696	3.927874	0.099891
	K2	0.038313043	0.081808696	0.257356522	0.25077619	0.013530435	2.127147826
	K3	0.004812	0.2627	0.076531818	0.000917391	0.670365217	0.068020833
	K4	0.124430769	0.002595455	0.199026087	0.559395652	0.101486957	0.142334783
	K5	0.360919	0.173509	0.102295	0.083118	6.891128	0.011862
	K6	1.666657	0.30833	0.073932	0.006564	0.128271	0.107942
	K7	0.145084	0.200346	_____	0.018177	0.763243	0.369038
	K8	0.037887	0.0579	0.227368	0.282615789	3.53783913	0.381166667
	K9	0.283143	0.89124	0.11303	0.386084	0.232962	0.049682

5. Conclusion

Stock market is unstable in nature and is very challenging to predict with traditional techniques. Many studies acknowledged that the news influences stock prices. A framework for stock movement prediction using new news messages and corporate actions was presented by rapidly analyzing remarkably increasing information of recent times. Financial information and events were processed using filter process and multilayer feed forward network with backward propagation was employed to empirically explore the intricate relationships between events and the stock market. Experimental results showed that there is a significant correlation between the unregulated news, events and the stock price. A combination of fundamental, technical and unstructured information provides optimum gain in determining the share price. From our investigation, we identified certain terms which had positive or negative influences on stock price. Some of these terms were frequently appearing in several different representations, which conveyed that articles containing these terms were vulnerable to price movement. The dataset considered was from a compressed time span and the results obtained are indicative of the patterns observed within this dataset. Hence, it suggested a systematic and scientific method to predict stock prices by rapidly analyzing enormously increasing information of recent era. The predictions were made in such a way that the actual and the predicted value should have an error less than 0.1%. Using

this technique we can get accurate and reliable prediction which gives customer better solution for where to invest their valuable money. Therefore, it is required for future research to study the methods that can increase the accuracy and credibility further by combining other information that can make up the existing news and social media data as well as apply this to various other sectors apart from the IT industry as shown through this study.

References

[1] R. P. Schumaker, "Analysis of Verbs in Financial Articles and Their Impact on Stock Price", In: *Proc. of the NAACL HLT Workshop on Computational Linguistics in a World of Social Media Conf.*, pp. 3-4, 2010.

[2] R. P. Schumaker, "Analyzing Parts of Speech and Their Impact on Stock Price", *Communications of IIMA*, Vol. 10, No. 3, pp. 1-10, 2010.

[3] S. H. Cheng, "Forecasting the Change of Intraday Stock Price by Using Text Mining News of Stock", In: *Proc. of the International Conference on Machine Learning and Cybernetics*, pp. 2605-2609, 2010.

[4] M. V. Pinto and K. Asnani, "Stock Price Prediction Using Quotes and Financial News", *International Journal of Soft Computing and Engineering*, Vol. 1, No. 5, pp. 266-269, 2011.

[5] A. Nagar and M. Hahsler, "Using Text and Data Mining Techniques to Extract Stock Market Sentiment from Live News Streams", In: *Proc. of the International Conference on Computer Technology and Science,* pp. 1-5, 2012.

[6] R. Cao, X. Liang, and Z. Ni ,"Stock Price Forecasting with Support Vector Machines Based on Web Financial Information Sentiment Analysis", In: *Proc. of the International Conference on Advanced Data Mining and Applications*, pp. 527-538, 2012.

[7] P. Hajek, V. Olej, and R. Myskova, "Forecasting Stock Prices using Sentiment Information in Annual Reports - A Neural Network and Support Vector Regression Approach", *WSEAS Transactions on Business and Economics*, Vol. 10, No. 4, pp. 293-305, 2013.

[8] H. Lee, M. Surdeanu, B. MacCartney, and D. Jurafsky, "On the Importance of Text Analysis for Stock Price Prediction", In: *Proc. of the International Conference on Language Resources and Evaluation*, pp. 1170-1175, 2014.

[9] Y. Kim, S. R. Jeong, and I. Ghani, "Text Opinion Mining to Analyze News for Stock Market Prediction", *International Journal of Advances in Soft Computing and its Applications*, Vol.6, No.1, pp.1-13, 2014.

[10] T. H. Nguyen, K. Shirai, and J. Velcin, "Sentiment Analysis on Social Media for Stock Movement Prediction", *Expert Systems with Applications*, Vol.42, No.24, pp.9603-9611, 2015.

[11] L. Guo, F. Shi, and J. Tu, "Textual Analysis and Machine Leaning: Crack Unstructured Data in Finance and Accounting", *The Journal of Finance and Data Science*, Vol. 2, No.3, pp. 153-170, 2016.

[12] Y. Xie and H. Jiang, "Stock Market Forecasting Based on Text Mining Technology: A Support Vector Machine Method", *Journal of Computers*, Vol. 12, No.6, pp. 500-510, 2017.

[13] 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*, Vol.10, No.6, pp.62-77, 2017.

[14] S. Bharathi and A. Geetha, "Sentiment Analysis for Effective Stock Market Prediction", *International Journal of Intelligent Engineering and Systems*, Vol.10, No.3, pp.146-154, 2017.