

Social Media Aided Sentiment Analysis for Stock Prediction



K. Nirmala Devi, N. Krishnamoorthy, K.S. Aparna

Abstract: In the stock market research, stock prediction is a challenging task due to its dynamic characteristic very similar to wealth of a nation and opinion about a stock. It is very difficult for the investor to buy or sell their stock because of noisy, chaotic properties of the stock data. Stock prediction mostly performed depends up on the numerical data obtained with technical measures or text data provided by the data sources as sentiments. A change in the fundamental measures obtained from exchange rate, gold price and crude oil price also determines the stock value. User generated contents of sentiments available in various social media like Twitter and News sites also play an important role for deciding the price of the stock. Most of the existing work deals any one of the measures technical measures or fundamental measures or sentiment measures for predicting the price of the stock. Hence, the proposed method employs combined measures derived from technical, fundamental and sentiments. Twitter and Money Control act as a data source for providing opinion data to predict the stock price. Results of the proposed system compared with the others by using various measures such as accuracy, sensitivity, specificity, Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). It is found that the proposed methods for stock prediction outperform the existing techniques..

Keywords: Stock prediction, sentiment analysis, opinion mining, social media.

I. INTRODUCTION

Social media facilitates the user to generate the various nature of data such as text, image, audio and video, make them to use by others in remarkable ways. The exponential growth of User Generated Contents (UGC) provides an important platform to enable the user to perform various research. The variety of social media like Twitter, microblogs, discussion forums and Facebook are having precious information for identifying, detecting and predicting major public events. [1],[2]

Social media having mutual contents with collective information helps to decide the stock price in an integrated environment. The technical advancement helps the user in simple manner to create, express and share their views or comments in the social media.

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

K. Nirmala Devi*, Computer Science and Engineering Department, Kongu Engineering College, Perundurai, Erode, k_nirmal.cse@kongu.edu.

N.Krishnamoorthy, department, Computer Science and Engineering Department, Kongu Engineering College, Perundurai, Erode.

K.S. Aparna, UG Scholar, Computer Science and Engineering Department, Kongu Engineering College, Perundurai, Erode.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Collective intelligence derived from social media employed in various forecasting application to determine their upcoming performance such as forecasting stock trends [3], predicting box office sale, detecting hot topic [4], election outcome prediction and so on.

Sentiment is called as an opinion, based on belief or view. Automatic process of detecting opinions, attitudes and emotions is called as Sentiment Analysis (SA) or Opinion Mining (OM). Natural Language Processing (NLP) is an important tool to extract opinions from text or speech. SA is a part of data mining and it is used along with NLP to automatically identify, detect and understand the user generated opinions [5].

OM approaches enable to improve the value of the existing information resources in numerous ways. This play vital role in decision making. Since, decision which is affected by the opinions formed by privileged users and others. Other users' opinion plays an important role in decision making while purchasing a product in online. Before making a decision to buy a product in online the user normally begin to identify the opinions and reviews proposed by others.

The rest of this paper is organized as follows: Elaborated literature review part is available in Section II. Proposed system and its dataset is discussed in Section III. The methodology used in the proposed system is described in Section IV. The Section V covers the results and discussion and Section VI includes the conclusion part.

II. LITERATURE REVIEW

Some studies [6],[7],[8] focused on the collaborative knowledge expressed as a sentiment for predicting the returns of the stock market in desired manner. Individual decision regarding buying or selling a stock affected by a sentiment. Social media data having sentiments are useful for determining the stock price in promising manner. Most of the service providers are attracted on the basis of sentiments in text naturally. Google's search engine functions based on text sentiments to enables the user in attractive and flexible way to support the search process.

The author [9] has proposed Back Propagation (BP) and Support Vector Machine (SVM) for stock price prediction; however BP affected by a low convergence. While comparing the results of BP, SVM provides better results than BP. Stock market "bullish" and "bearish" were determined by the Tweet sentiments proposed by the author John [10]. Occurrence of financial terms and opinions considerably determine the stock returns. Asian country stock is predicted by the mixture of Back Propagation Neural Network (BPNN) and SVM.

ANN proposed by Karymshakov [11] also used for stock forecasting. Istanbul Stock Exchange (ISE) prices have analyzed by using SVM and ANN[12]. An additional cross model consists of GA and SVM used for predicting the stock market. Face recognition was done using another model developed by PSO and SVM.

Besides historical statistic information in stock market, there could be some other factors like political movement [13] and investor’s sentiment that can affect stock market [14]. For example, some sudden events or news can have big influence on the companies’ stock prices [15]. Moreover, the stock market is related to public sentiment or mood as mood can affect the humans' decision making process.

III. PROPOSED SYSTEM

A. Data Set Description

A stock market is a place where buyers and sellers exchange trade together using the platform provided by stock exchanges. Stock markets are most important element of the world’s economy as they present a great environment for companies to raise their capitals with no trouble. Though many stock exchanges exist, the most prominent of them are the Bombay Stock Exchange (BSE) and National Stock Exchange (NSE) in India. NSE is the 2nd greatest rising stock exchange in the world and the largest stock exchange in India whereas BSE is the 10th largest in the world and largest stock exchange in South Asia. These stock exchanges are regulated and monitored by the Security Exchange Board of India (SEBI). There are two important indices in Indian stock markets: SENSitive indeX (SENSEX) and Nifty. Nifty consists of a group of 50 shares of NSE and SENSEX consists of a group of 30 shares of BSE.

The proposed system dataset focused about the every days’ closing price of various IT industries such as Infosys, Wipro, Hindustan Computer Limited (HCL), Tata Consultancy Services (TCS) and Tech Mahindra along with BSE (SENSEX and Nifty), NSE and Yahoo Finance. Fundamental measures focused about Gold Price (GP), Exchange Rate (ER) and Crude Oil Price (COP) are considered for evaluation of the proposed system. Twitter and money control are the sources for providing the stock sentiments. All the sources of data considered from January 2017 to July 2018.

IV. METHODOLOGY

Social media predictive analysis of time series permits the user to perform instant decisions to react based on sentiments [16]. Discussion forums [17] and blogs are slowly progressing to spread data, while the stock market needs the fast progressing spreadable nature of communication medium. In addition to that, the news channels and Twitter are rising in a rapid manner rather than forums and blogs.

Unstructured data is widely available in the most part of the fundamental measure. It is always a hard decision for market researchers and stock investors to buy or sell. Consequently, it is necessary to extract the information from the unstructured sources to perform the analysis of stock market as well as to make use of them in the prediction. Hence

the stock price prediction is one of the significant issues in stock market research because of its challenging character.

A. Technical Indicators (TI)

Stock price prediction needs historical prices such as High, Low, Open, Close and Volume. Technical measures are obtained from historic data. The measures of TI includes RSI, Money Flow Index (MFI), Exponential Moving Average (EMA), Stochastic Oscillator (SO) and Moving Average Convergence and Divergence (MACD) and which are used as input for determining the next days’ closing price.

B. Fundamental Indicators (FI)

A Change in the fundamental measures such as GP, ER and COP influences the stock price up or down in the Indian stock market. COP indirectly proposed to stock price and it shows decreasing stock price while COP increases. At the same time investors moved to invest yellow metals (GP) due to decreased stock price and that increases the GP. Moreover, the ER changes also create impact on the stock price fluctuations.

C. Sentiment Indicators (SI)

Investors’ sentiments towards stock determined by some trading terms such as ‘bull’, ‘slip’ etc.,. Key terms determining stock price are used in the training part of machine learning techniques. Table I represents some possible key terms associated in the financial domain are described as follows.

Table I Sample Positive and Negative Words for Stock

Positive Words	Negative Words
benefit	bear
boom	crash
bull	critical
calm	danger
certainty	decline
climb	decrease
easy	loss
safe	low
secure	pesimism
strong	slip
success	uncertain
up	unsafe
	weak

The following samples related to stocks are gathered from money control and Twitter.

1. TCS looks good
2. Wipro is slip down to further
3. HCL Tech falls

The samples 1 is positive whereas 2, 3 and are negative samples. A word ‘good’ represents the positive indicator, where as ‘slip’, ‘falls’ represent negative indicators.

Sentiment Bullishness Index (SBI) is calculated using Equation (1).

$$\text{SentimentBullishnessIndex(SBI)} = \ln \left[\frac{1+N^{Pos}}{1+N^{Neg}} \right] \quad (1)$$



where N^{Pos} represent the total Positive sentiments and total negative sentiments are represented by N^{Neg} . Increasing stock price noted as bullish sentiment and stock price falling denoted as bearish sentiment. If Sentiment Bullishness Index(SBI) is greater than zero, it is stock price up, 0 is no change whereas less than 0 indicates the stock price down.

D. Stock Prediction using Combined Indicators (CI)

Stock markets are most important element of the world’s economy as they present a great environment for companies to raise their capital investment easily.

Fig.1 represents the stock prediction process description. The outline of stock prediction process includes sources of data, collecting data from various sources, perform preprocessing, extracting features and forecasting.

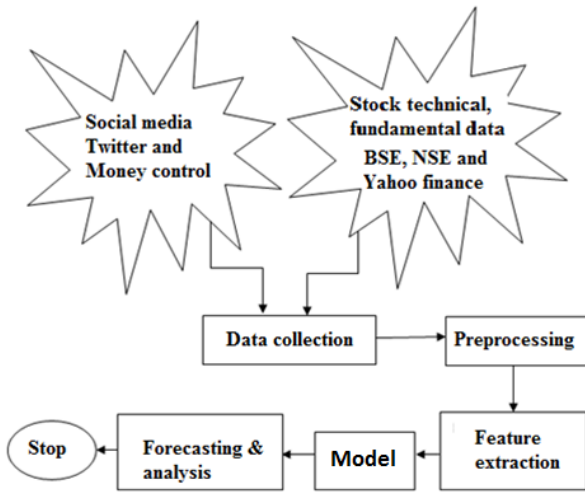


Fig.1 Stock Prediction Model

Data Collection

The proposed system uses integrated measures that are obtained from various data sources such as NSE, BSE, Yahoo finance, money control and Twitter micro blog. Historic data obtained are used to generate technical measures and sentiment data are obtained from the news source and micro blog.

Preprocessing the Stock Data

The process of removing unwanted information such as non-opinion information, stop words along with stemming is done during preprocessing phase.

Feature Extraction for Stock Data

Stock price prediction needs historical prices such as High, Low, Open, Close and Volume. Technical measures are obtained from historic data. The measures of TI includes RSI, MFI, EMA, SO and MACD. For all the stocks, Equation (2) is used to calculate the Stock Returns (SRET) as follows.

$$\text{StockReturns (SRET)} = \ln \left[\frac{CP_t - CP_{t-1}}{CP_{t-1}} \right] \tag{2}$$

where ‘t’ denotes the time, CP_t denotes the stock Closing Price (CP) at ‘t’ whereas CP_{t-1} denotes stock CP at ‘t-1’. Depending upon the predicted value of SRET the stock price may up or down. If SRET value is above zero indicates the possibility of the increasing the stock price. The SRET value is lower than zero indicates the possibility of decreasing stock

price and zero denotes no way up or down in the stock price. Trading rule guides to know about the increasing or decreasing possibility of stock trend as follows..

Trading rule

- (i) If predicted trend for the day (t+1) is up then invest else if don’t sell already available stock.
- (ii) Else if predicted trend for the day (t+1) is down then put up for sale otherwise don’t sell.
- (iii) Else if predicted trend for the day (t+1) does not have trend then don’t sell.

Forecasting and Analysis

Prediction model employs various methods like SVM, J48 and Naive Bayes (NB) to determine the stock price at the forth coming day.

V. RESULTS AND DISCUSSIONS

A. Performance Evaluation Measures

Stock markets’ increasing status, decreasing status along with prediction accuracy and error measures is evaluated with help of sensitivity, specificity, accuracy, RMSE and MAPE respectively. Proposed systems’ performance parameters are calculated by using Equations (3), (4), (5), (6) and (7).

$$\text{Accuracy} = \left[\frac{\text{No.ofRightPredictions}}{\text{TotalPredictions}} \right] * 100 \tag{3}$$

$$\text{Sensitivity} = \frac{\text{No.of Correctly classified Positive instances}}{\text{Totalinstances}} \tag{4}$$

$$\text{Specificity} = \frac{\text{No.of Correctly classified Negative instances}}{\text{Totalinstances}} \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (\text{Actual}_t - \text{Predict})^2} \tag{6}$$

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^N \left| \frac{(\text{Actual}_t - \text{Predicted}_t)}{\text{Actual}_t} \right| \tag{7}$$

SVM model is used for prediction along with technical, fundamental and combined measures. The obtained results of MAPE are shown in the Table II. Table III contains the RMSE results of the prediction model.

The lower MAPE value of proposed measure (CI) while comparing with others such as TI [18], FI [19],SI [15], TI&FI [20], TI&SI [21], FI&SI [22] shows the outstanding performance of the proposed system.

The higher value of RMSE in Table III for other systems shows the efficiency of proposed system with the low errors. It shows the better performance results for the proposed system while comparing with various systems.



Table II MAPE Results

Indicators	Infosys	HCL	TCS	Tech Mahindra	Wipro
TI	3.946	3.312	4.312	5.143	4.146
SI	2.236	3.172	4.345	6.732	4.634
FI	5.456	7.243	8.546	7.234	6.156
TI & FI	3.621	3.891	4.213	3.982	4.981
TI & SI	2.789	3.823	3.624	3.123	3.345
FI & SI	3.321	3.921	3.732	2.678	3.778
CI	1.523	2.632	3.123	2.341	2.012

Table III RMSE Results

Indicators	Infosys	HCL	TCS	Tech Mahindra	Wipro
TI	0.731	0.724	0.891	0.821	0.643
SI	0.662	0.714	0.863	0.814	0.562
FI	0.641	0.624	0.876	0.926	0.641
TI & FI	0.524	0.578	0.884	0.656	0.426
TI & SI	0.443	0.528	0.745	0.635	0.324
FI & SI	0.312	0.494	0.621	0.487	0.246
CI	0.014	0.034	0.071	0.023	0.068

In order to obtain the best prediction results, low error rate and high accuracy is needed. Low error rates (MAPE, RMSE) for the proposed system shown in the Tables II and III indicate the best performance of the system among the others.

Figures Fig.2, Fig.3, and Fig.4 show the robustness of the proposed system with other systems.

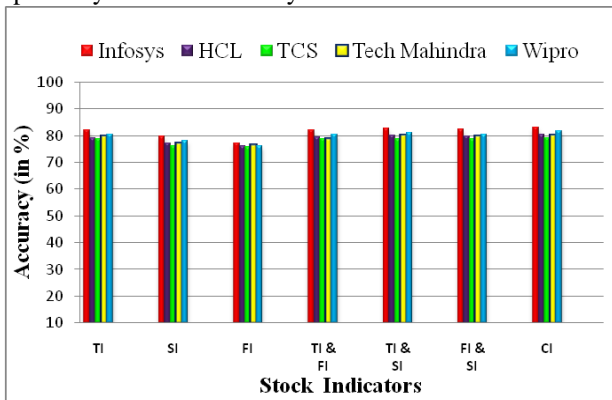


Fig.2 Accuracy of SVM for various Stock Indicators

Fig.2 represents the accuracy results of proposed model with technical, fundamental, sentiment and combined measures. The combined measure provides better accuracy results for the proposed model.

Sensitivity results of the proposed model is shown in Fig.3. This indicates the robustness of the combined measures among others.

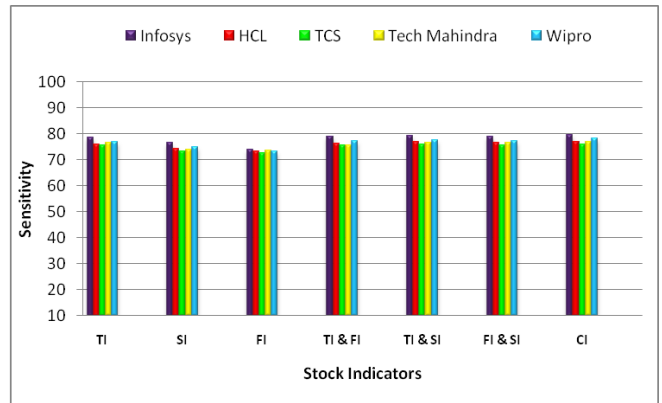


Fig.3 Sensitivity of SVM for various Stock Indicators

Results of specificity showed in Fig.4, indicate the better results of proposed system with others.

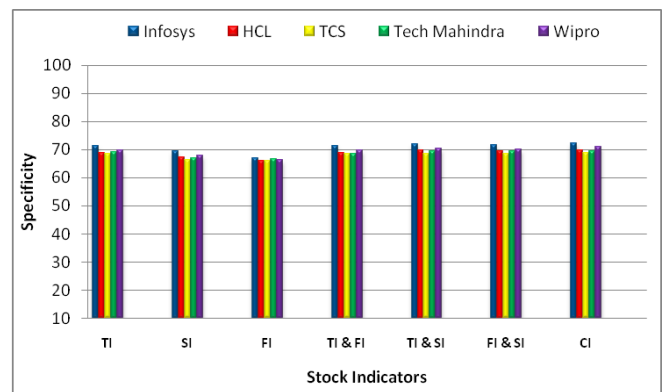


Fig.4 Specificity of SVM for various Stock Indicators

The obtained results guarantee that combined measure provides better prediction while comparing to other measures. When incorporating sentiments the prediction accuracy will get boosted in easy manner. There is considerable accuracy prediction results reveals the robustness of the proposed system with other baselines.

VI. CONCLUSION

This paper proposed combined indicators for stock prediction and also has the detailed comparison results of the various measures for evaluating the proposed system with others. Combined measure predicts the stock price with expected level in more accurate form.

The proposed research enables the user to make buy or sell their stock depending upon the other user’s sentiment regarding the stock trend. It is confirmed that sentiments in the news and micro blog determine the stock price up or down. Hence, it is proved that social media aided sentiments provides higher accuracy results in the stock prediction.

Decision makers in financial company and other fields gain the appropriate guidelines for effective fund management to provide gainful environment with positive returns. Besides various trading measures, sentiments play an important role for making good investment decisions and effective portfolio management.



Furthermore the proposed system can be extended to identify the valuable and relevant features for improved decision making.

REFERENCES

1. P.S Earle,D.C Bowden and M.Guy,"Twitter Earthquake Detection: Earthquake Monitoring in a Social World", Annals Geophysics,2011,vol. 54, no. 6.
2. Mark Dredze,"How Social Media Will Change Public Health", IEEE Intelligent Systems, 2012, vol.27, no.04, pp. 81-84.
3. PSM. Nizer and J.C.Nievola,"Predicting published news effect in the Brazilian stock market", Expert Systems with Applications, 2012,vol.39, pp.10674–10680.
4. Kuan-Yu Chen, Luesak Luesukprasert and Seng-cho Chou,"Hot Topic Extraction Based on Timeline Analysis and Multidimensional Sentence Modeling", IEEE Transactions on Knowledge and Data Engineering, 2007, vol.19, no. 8, pp.1016-1025
5. Bing Liu, "Sentiment Analysis and Opinion Mining", Morgan & Claypool Publishers, 2012.
6. E.Cambria, B.Schuller,X.Yunqing and c.Havasi, "New avenues in opinion mining and sentiment analysis", IEEE Intelligent Systems, 2013, vol.28, pp. 15–21
7. Efthimos Bothos, Dimitris Apostolou and Gregoris Mentzas, "Using Social Media to Predict Future Events with Agent-Based Markets", IEEE Intelligent Systems, 2010, vol.25, no.6, pp.50 – 58
8. Wu.Desheng Dash, Lijuan Zheng, L.David and Olson,"A Decision Support Approach for Online Stock Forum Sentiment Analysis", IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2014, vol.44, no. 8, pp.1077-1087
9. Ming-Chi Lee, "Using support vector machine with a hybrid feature selection method to the stock trend prediction", Expert Systems with Applications 2009, vol.36, pp. 10896–10904
10. Johan Bollen, Huina Mao, and Xiao-Jun Zeng, "Twitter mood predicts the stock market", IEEE Computer, 2011, vol. 44, no.10, pp. 91–94.
11. K. Karymshakov, and Y.Abdykaparov, "Forecasting stock index movement with artificial neural networks : The case of Istanbul stock exchange", Trakya Universitesi Sosyal Bilimler Dergisi, 2012, vol.14,no.2, pp. 231-242
12. Y. Kara, M.A.Boyacioglu and O.K.Baykan,"Predicting Direction of Stock Price Movement Using Artificial Neural Networks and Support Vector Machines: The Sample of the Istanbul Stock Exchange", Expert Systems with Applications, 2011, vol.38, no.5, pp. 5311-5319.
13. Pawel Sobkowicz , Michael Kaschesky and Guillaume Bouchard,"Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web", Government Information Quarterly, 2012 vol. 29 ,pp. 470–479.
14. L.J Cao and F.E.H Tay, "Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting", J. IEEE Tran. on Neural Network, 2003, vol. 14, pp.1506–1518
15. R.P. Schumaker, Y.Zhang, C.N.Huang and H.Chen,"Evaluating sentiment in financial news articles", Decision Support Systems, 2012, vol. 53, pp.458–464.
16. T.Le, Nguyen, Wu. Pang, William Chan, Wei Peng and Ying Zhang, "Predicting Collective Sentiment Dynamics from Time-series Social Media", Proceedings of the ACM - WISDOM '12, 2012.
17. Nan Li and Desheng Dash Wu, "Using text mining and sentiment analysis for online forums hotspot detection and forecast", Decision Support Systems, 2010, vol.48, pp.354–368.
18. C.Chun, M.Qindhua, and L.Shuqiang, "Research on Support Vector Regression in the Stock Market Forecasting", Springer– Advances in Intelligent and soft Computing ,2012, vol.148, pp. 607-612.
19. Gagan Deep Sharma and Mandeep Mahendra , "Relationships among Oil Price, Gold Price, Exchange Rate and International Stock Markets", International Research Journal of Finance and Economics, 2010, vol.47, pp.80-89.
20. J.A.Macedo, L.T.Camargo, H.C.B.Oliveira,L.E.Silva and R.M Salgado,"An Intelligent Decision Support System to Investment in the Stock Market", IEEE Latin America Transactions, 2013,vol. 11, no. 2,pp. 812-819
21. Yuzheng Zhai, Arthur Hsu and Saman Halgamuge, "Combining News and Technical Indicators in Daily Stock Price Trends Prediction", 2007, pp. 1087-1096, Springer-Verlag, Berlin.
22. Yuh-Jen Chen and Yuh-Min Chen, "A fundamental Analysis-based Method for Stock Market Forecasting", Proceedings of ICICIP, China, 2013.

AUTHORS PROFILE



Dr.K.Nirmala Devi is associated with the Department of Computer Science and Engineering as Associate Professor at Kongu Engineering College, Tamil Nadu, India. She has presented more than 20 papers in national and international conferences and published more than 20 papers in national and international journals. Her Area of Interest is on Big data, Machine Learning and Sentiment Analysis.



Dr.N.Krishnamoorthy is associated with the Department of Computer Science and Engineering as Associate Professor at Kongu Engineering College, Tamil Nadu, India. He has presented more than 20 papers in national and international conferences and published more than 20 papers in national and international journals. His Area of Interest is on Operating Systems, Grid Computing and Big Data Analytics.



Ms. K.S.Aparna is associated with the Department of Computer Science and Engineering as UG Scholar at Kongu Engineering College, Tamil Nadu, India. She has presented more than 5 papers in national and international conferences. Her Area of Interest is on Machine Learning and Web mining.