

Predicting Stock Price Movement based on Communication Network and Sentiment Analysis

Shital Shiral

Department of Computer Science & Engineering
VVPIET, Solapur University, Solapur, India

Harshal Torvi

Department of Computer Science & Engineering
VVPIET, Solapur University, Solapur, India

ABSTRACT

Stock market is responsible for trading the shares of public listed companies. Stock exchange facilitates stockbrokers to trade company stocks and other securities. Stock price prediction is one of the most challenging issue which is attracting researchers from many fields including economics, history, finance, and mathematics and computer science. It is difficult to apply simple time-series or regression techniques on stock market because of the volatile nature.

Proposed framework attempts to predict whether a stock price sometimes in the future will be higher or lower than it is on a given day. We find a little predictive ability in the short run but definite predictive ability in the long run. Using the social communication network within company among employees, the proposed algorithm can analyze the relationship between communication context and the movements (high and low) of stock price. We have also extended the system by using sentiment analysis for email content which determines whether email context is negative or positive. System gives aggregated result from number of mail exchanged and sentiment of message body.

Keywords

Share Market, Artificial Neural Network, Time series, Stock market prediction algorithm, Stock market pattern recognition

1. INTRODUCTION

Stock market prediction methods are gaining more attention, maybe due to the fact that if the trend of the market is successfully predicted the investors may be better guided. The profit gained by investing and trading in the stock market greatly depends on the predictability. If there is a system that can consistently predict the direction of the dynamic stock market will enable the users of the system to make informed decisions. Stock prediction trends of the stock market can help the stock brokers in taking corrective measures.

1.1 Stock Market

A stock market or share market is the association of buyers and sellers which makes economic transactions of stocks which is also called shares, which represent ownership claims on businesses which includes securities listed on a public stock exchange as well as those who only traded privately [1]. Examples of private share trading will be shares of private companies which are bought by investors through equity or crowd funding. Companies are important for a stock market to work. A company must be listed as a PLC (public listed company) for people to trade its shares at a stock market.

1.2 Stock Trade

Without people who trade stock (buyers and sellers) a stock market would not work. Buyers can range individuals sitting at home on their PC to huge multimillion dollar investment

funds. Trading in stock markets means the transfer for money of a stock or security from a seller to a buyer [2]. This requires these two parties to agree on a price. Equities (stocks or shares) confer an ownership interest in a particular company. Participants in the stock market range from small individual stock investors to larger trader investors, who can be based anywhere in the world, and may include banks, insurance companies, pension funds and hedge funds. Their buy or sell orders may be executed on their behalf by a stock exchange trader.

1.3 Predictions Stock

There are two approaches to carry out prediction, fundamental approach and technical approach [12]. In fundamental analysis method uses numeric information like profit ratios and earnings for prediction while technical analysis utilize modeling techniques such as graphs to predict trends in price. Even with the lack of prediction methodologies there is little success in prediction of stock price, both methodologies are rely on historical data to predict the stock market price [8][9]. Textual data can also help Stock Market prediction. Data for prediction can be extracted from company's news stories and quarterly reports. Stock market prediction using text mining is heavily depending on predefined set of keywords and some machine learning rules [17]. In this method weights are assigned to keywords with respect to the movement of a share price. Some text mining prediction tools retrieve financial news articles and blogs published in previous year and respective stock market prices for same period. Retrieved articles are then labeled as positive or negative depending on their impact on stock price movement. Classification tools like support vector machines is then trained with labeled train articles. This train model is then used to predict the future stock price of the company [11] [16].

Section I contains the introduction of stock price and stock price movement prediction and its practices, Section II contains a survey on the related work and existing frameworks, Section III contains detail description of proposed system and architecture, Section IV contains result analysis and section V concludes research work with future directions).

2. LITERATURE SURVEY

[1] This paper states that a corporate e-mail ecosystem contains meaningful information about employees' communication network and patterns. Author has only focused on the communication frequency that is e-mail exchanged between employees. Author used Enron dataset in which he identified patterns of e-mail exchange. Author believes that such patterns can reveal important information about organizations stability that may subsequently influence the company's performance in the stock market. Author used this methodology to indirectly predict company's stock performance. Finally author stated that there is a correlation

between email communication network and stock price in Enron dataset and can be accurately predicted based on the detected relationships.

[2] In this paper author proposed an intelligent time series prediction system. To construct a stock price forecasting expert system author used meta heuristic firefly algorithm and least squares support vector regression (MetaFA-LSSVR). System uses sliding-window meta heuristic optimization and is a graphical user interface that runs as a stand-alone application. As compared to other prediction methodologies proposed methodology is very simple and involves very few computations. To generalize the application of proposed system author used stock price market in Taiwan and can be extended to analyze multivariate time series data and estimate other stock markets such as China, Korea, and Vietnam etc.

[3] To build prediction module author proposed LSTM (Term Memory Long-Short) recurrent neural networks. To analyze the stock data author has first filtered the data extracted feature value. The main work of LSTM recurrent neural networks is to extract feature value and analyze the stock data. Author stated that the proposed model can play a better forecasting effect, even though the accuracy is not very high as per the experimental results. Results are calculated on dataset of JoinQuant platform. System achieved 72% of accuracy for the short period of data. Author believes that the model has a lot of space to improve its accuracy and if the more features to train the model are provided then system can achieve desirable accuracy.

[4] In this paper before developing the model author first studied various methods to find the most accurate model for prediction of prices of the stock. To get better results author increased nodes of neural network as stated by author of paper “Stock Price Prediction and Trend Prediction Using Neural Networks”. Author found that Feed Forward Neural network provides the highest accuracy for the stock prediction. Finally author calculated error of results in which least amount of mean absolute percentage error is 1.81598342% for feed forward neural network and the maximum error is 11.32847594% which is obtained using linear model with polynomial trend.

The result obtained was the opening price of the stock and that too was average for a full month. So an improvement in this system can be achieved by forecasting the opening price of each day.

[5] In this paper author aims to prove that artificial neural network is an effective tool to predict market stock. Paper provides comparative study between Long Short-Term Memory (LSTM) model and Multilayer Perceptron (MLP) model. Proposed heuristic method of analysis and study followed in this paper are optimal for stocks which are highly frequently traded and can outperform (EMH) Efficient Market Hypothesis theory. 10 unique stocks recorded on New York Stock Exchange are used to calculate results. Calculated results prove that MLP has outperformed LSTM model, in predicting short term stock prices.

[6] In this paper author proposed a stock recommendation system using J48 and random forest algorithm. Author firstly optimized the stock price prediction only for short term, author found that using specific pre-processing tools and machine learning framework it is possible to improve accuracy of short term stock prediction but these technical indicators are not that helpful to predict long term stock trend but some other factors like historical data that is net profit of a

company of previous years, PE ratio and promoters holding can help to predict long term stock price.

3. PROPOSED METHODOLOGY

Using the social communication network within company among employees, the proposed algorithm can analyze the relationship between communication context and the movements (high and low) of stock price. We assume that the values of movements of stock price are related to the values of frequency of communication in the previous time point.

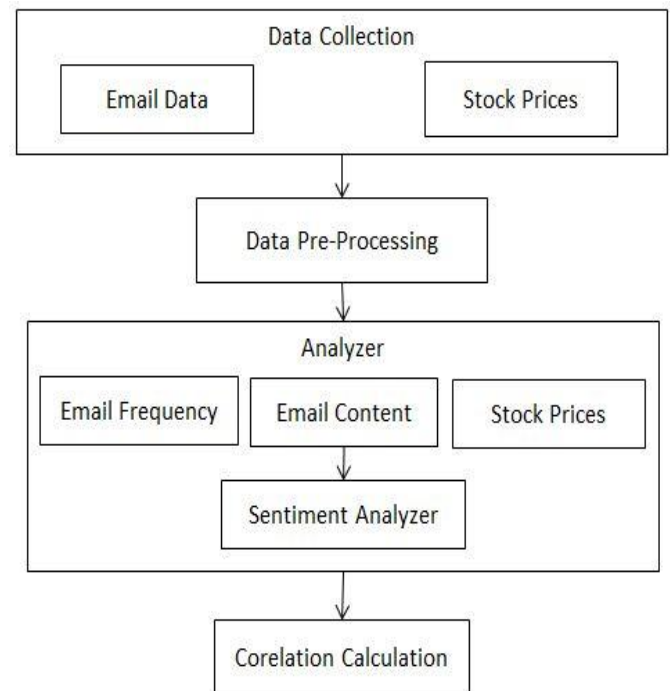


Fig. 2 System design for proposed system

The task of our proposed algorithm is to uncover these temporal relationships and predict the movements of stock price using communication network. The proposed algorithm includes the following steps:

1. Discretizing the weight matrices and the value of movement
2. Discovering patterns for describing the relationship between communication frequency and stock price
3. Constructing prediction rules based on patterns
4. Predicting the movement of stock price using prediction rules

The purposed method for developing the system consists of mainly three main steps. Firstly, data is collected and sorted for relevancy. Secondly, analysis is carried out on the collected data by examining the email flow, email content after which the data is represented and scored accordingly. At last, an ANN is designed and a suitable algorithm yielding best accuracy is chosen to predict the stock value

3.1 Modules Description

3.1.1 Data Collection

In this module data is collected as per the time point from dataset. Here we are using Enron dataset. When Enron collapsed in 2001, about 500,000 internal emails were made public. We are using cleaned Enron data and that is in

MSQL format.

3.1.2 Data Pre-Processing

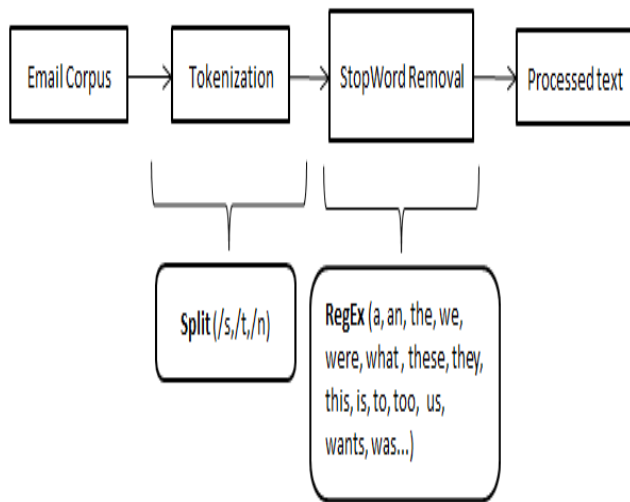


Fig. 3 Data Pre-Processing

3.1.2.1 Tokenization:

Email body is split into individual words based on the space and irrelevant symbols are removed. We form a list of individual words for each email body.

In fig. 3.2 we are using tokenization, we are using Split function with some parameters. Parameters that we are providing are “/s”, “/t”, “/n”

Table 1 Regular expression

Regular Expression	Description
\s	stands for "whitespace character"
\n	Refers to "new line" (aka one tab vertical space)
\t	Refers to "four horizontal spaces"

3.1.2.2 Stop word Removal:

Words that do not express any emotion are called stop words. Words like a, is, the, with etc. are removed from the list of words. To remove stop words we are using Regex function. We are providing dictionary of stop words to Regex function to match those words and remove them from statements.

3.1.3 Analyzer

3.1.3.1 Sentiment Analysis:

Emails are classified as positive, negative and neutral based on the sentiment present. 70% email body out of the total emails are examining manually and annotated as 1 for Positive, 0 for Neutral and 2 for Negative emotions. For classification of nonhuman annotated emails a machine learning model is trained whose features are extracted from the human annotated emails. The extracted features are fed to the classifier and trained using LibSVM algorithm.

3.1.4 communication frequency calculation

This module calculates the communication frequency that is mail exchange count between two nodes within given time interval.

The proposed algorithm includes the following steps:

1. Discretizing the weight matrices and the value of movement;
2. Discovering patterns for describing the relationship between communication frequency and stock price;
3. Constructing predication rules based on patterns;
4. Predicting the movement of stock price using prediction rules

3.1.5 Correlation calculations

This is a final stage of proposed system. All features extracted from analyzer module (sentiment result value, communication frequency value) is then provided to LibSVM. When the model with LibSVM is trained with 90 percent of data, it is expected to give a result of 80%. These results give a significant edge to the investors and they show good correlation between stock market movements and the communication between company networks.

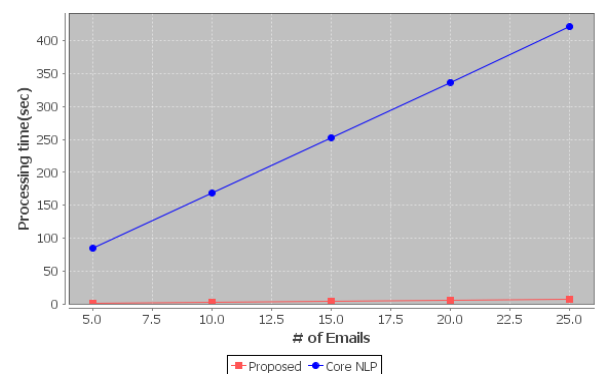
4. RESULT ANALYSIS

This project attempts to predict the stock value with respect to the stock's previous value and trends. It requires historic data of stock market as the project also emphasizes on data mining techniques. So, it is necessary to have a trusted source having relevant and necessary data required for the prediction. We will be Enron Email Dataset as the primary source of data. Dataset contains all the details such as: Opening value, Closing value, Highest value, Lowest value, Date wise sent and received emails.

Dataset URL: <https://www.cs.cmu.edu/~enron/>

Table 1. Number of email v/s Time for sentiment extraction

Number of Emails	Proposed System Time (Sec)	Core NLP Time (Sec)
5	1.419	84.24
10	2.81	168.48
15	4.212	252.72
20	5.632	336.96
25	7.027	421.2



Graph 1. Number of email v/s Time for sentiment extraction

Accuracy: It is the degree to which the result of a measurement, calculation, or specification conforms to the correct value (true). The proposed system gives maximum

83% Accuracy. To calculate the accuracy in percentage we can multiply by 100 to the result.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where,

TP (True Positive) – Correctly Identified.

FP (False Positive) – Incorrectly Identified.

TN (True Negative) – Correctly Rejected.

FN (False Negative) – Incorrectly Rejected.

Below graph 2 shows the sentiment extraction accuracy of a proposed system. For calculating accuracy result we took 18 emails 1032 length. From 18 emails system gave 12 emails as true positive 3 as true negative 2 emails as false positive and 1 as false negative.

TP – 12, FP – 2, TN – 3, FN – 1

$$\text{Accuracy} = \frac{12 + 3}{12 + 3 + 2 + 1} = 0.83$$

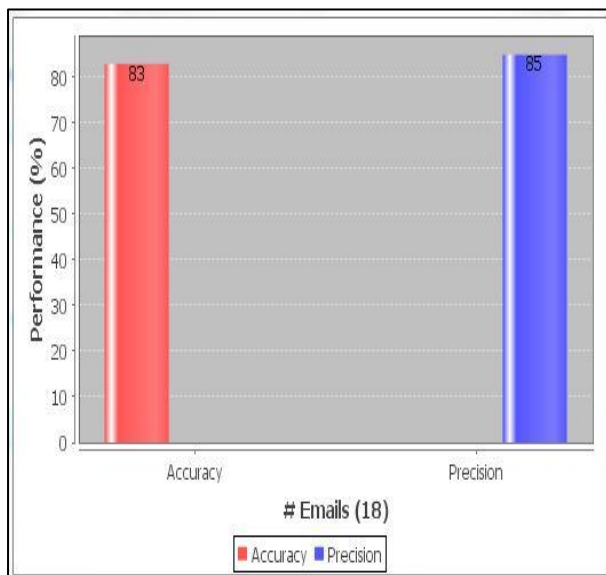
Precision: It is degree to which repeated measurements under unchanged condition show the same results.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Precision} = \frac{12}{12 + 2} = 0.85$$

Table 2. Accuracy and Precision of proposed system

# Emails	Accuracy (%)	Precision (%)
18	83	85

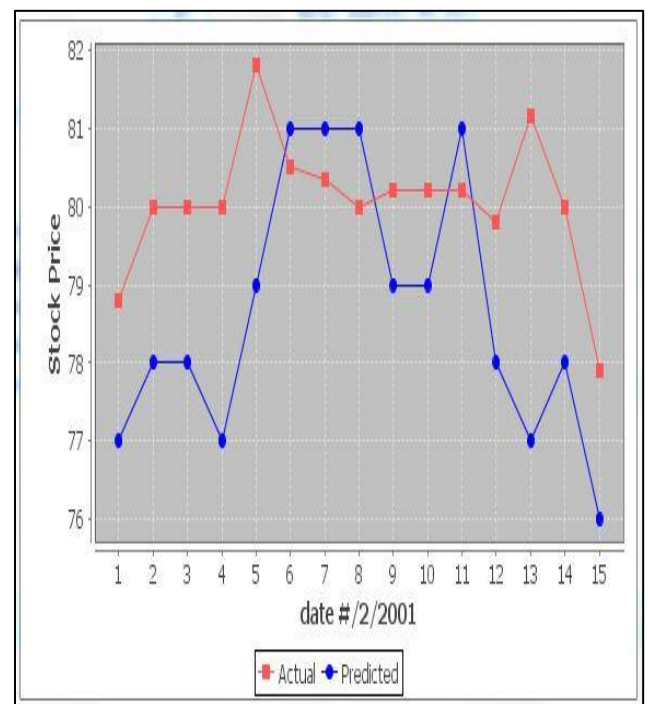


Graph 2. Accuracy and Precision of proposed system

Table 3 shows the difference between actual values and predicted values evaluated by proposed system. As the table shows the system predicted values are quite similar to actual values the average difference between actual value and predicted value is 1.

Table 3. Actual values v/s Predicted Values

Date	Actual Values	Predicted Values
1/2/2001	78.79	77
2/2/2001	79.98	78
3/2/2001	79.98	78
4/2/2001	79.98	77
5/2/2001	81.81	79
6/2/2001	80.51	81
7/2/2001	80.35	81
8/2/2001	80.00	81
9/2/2001	80.20	79
10/2/2001	80.20	79
11/2/2001	80.20	81
12/2/2001	79.80	78
13/2/2001	81.15	77
14/2/2001	80.00	78
15/2/2001	77.90	76

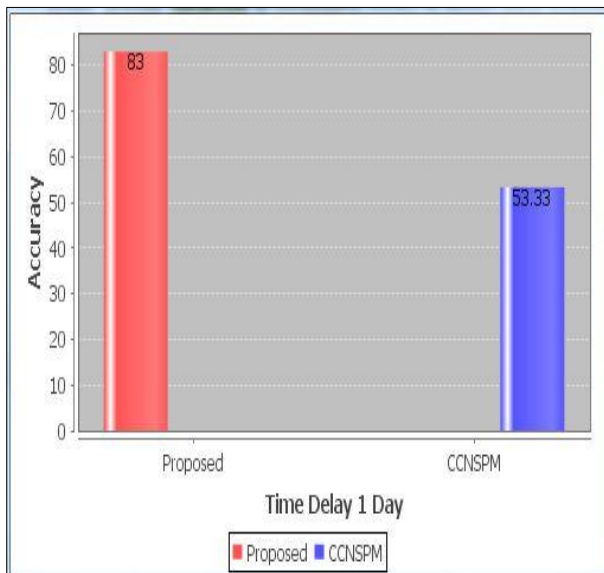


Graph 3. Actual values v/s Predicted Values

Table 4 shows accuracy comparison between proposed system and paper Corporate Communication Network and Stock Price Movements [1]. It also shows the idea of the days of look-back which means discovering the patterns of communication frequency and stock price of previous year. Here we considered 1 day. The time delay can be any number of days. As the result shows proposed system gives better accuracy result than base system Corporate Communication Network and Stock Price Movements.

Table 4. Accuracy of proposed system v/s CCNSPM System

Time Delay	Proposed System Accuracy (%)	CCNSPM Accuracy (%)
1	83	53.33



Graph 4. Accuracy of proposed system v/s CCNSPM System

5. CONCLUSION

We have implemented the stock market prediction framework using the social communication network within company among employees; the proposed algorithm can analyze the relationship between communication context and the movements (high and low) of stock price. We have also extended the system by using sentiment analysis for email content which determines whether email context is negative or positive. System gives aggregated result from number of mail exchanged and sentiment of message body. The predicted stock prices help investors make smart investment decisions as well as help analysts to predict and study trends market stocks

Our initial analysis show significant correlation between email exchange frequency and stock price movement also correlation between email body sentiment and stock price movement.

6. ACKNOWLEDGMENTS

We would like to express our gratitude to all those who helped us to complete this work. We would like to thank our colleagues who helped us time to time from preparing report and giving good suggestions. We also extend sincere thanks to all the staff members of Department for helping us in various aspects.

7. REFERENCES

[1] Pei-Yuan Zhou , Keith C.C. Chan, Member, IEEE, and Carol Xiaojuan Ou, “Corporate Communication Network and Stock Price Movements: Insights From Data Mining”, IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, 2018.
[2] Jui-Sheng Chou and Thi-Kha Nguyen “Forward Forecast

of Stock Price Using Sliding-window Metaheuristic-optimized Machine Learning Regression”. IEEE 2018

[4] Siyuan Liu¹ , Guangzhong Liao¹ ,Yifan Ding¹Hubei Key Laboratory of Intelligent Information Processing and Real-time Industrial System Stock Transaction Prediction Modeling and Analysis Based on LSTM IEEE 2018.
[5] Shashank Tiwari Akshay Bharadwaj Dr. Sudha Gupta “Stock Price Prediction Using Data Analytics. 978-1-5386-3852-1/17/\$31.00 2017 IEEE”
[6] Kaustubh Khare, Omkar Darekar, Prafull Gupta, Dr. V. Z. Attar Department of Computer En-gineering and Information Technology, College of Engineering, Pune.“Short Term Stock Price Prediction Using Deep Learning IEEE 2017”
[7] Mr.Rupesh A. Kamble Computer science and engineering Government College of Engineering Aurangabad, India Kamblerupesh9@gmail.com “Short and Long Term Stock Trend Prediction using Decision Tree” ICICCS IEEE 2017.
[8] David M. Q. Nelson, Adriano C. M. Pereira, Renato A. de Oliveira ”Stock Market’s Price Movement Prediction With LSTM Neural Networks” IEEE 2017.
[9] R. Yamini Nivetha, Dr. C. Dhaya Department of Computer Science and Engineering. Adhiparasakthi Engineering College Melmaruvathur, Kanchipuram District “Developing a Prediction Model for Stock Analysis” IEEE 2017.
[10] ASHISH SHARMA, DINESH BHURIYA, UPENDRA SINGH ”Survey of Stock Market Prediction Using Machine Learning Approach” IEEE 2017.
[11] Debadrita Banerjee Student, Department of Statistics St.Xavier’s College Kolkata, India deba.gb@gmail.com “Forecasting of Indian Stock Market using Time-series ARIMA Model” IEEE 2014
[12] Girija V Attigeri, Manohara Pai M M, Radhika M Pai, Aparna Nayak Manipal Institute of Technology, Manipal-576104, India “Stock Market Prediction: A Big Data Approach”IEEE 2014
[13] Aparna Anant Bhat, Sowmya Kamath S. Department of Information TechnologyNational Institute of Technology, KarnatakaSurathkal, India “Automated Stock Price Prediction and TradingFramework for Nifty Intraday Trading” IEEE 2013
[14] Radu IacominFaculty of Automatic Control and Computers University POLITEHNICA of Bucharest Bucharest, Romania Email: radiuacomin@yahoo.com “Stock Market Prediction”IEEE 2015
[15] Tao Xing, Yuan Sun, Qian Wang, Guo Yu International School of Software Wuhan University Wuhan, China e-mail:xingtao@whu.edu.cn, cjhsun@hotmail.com “The Analysis and Prediction of Stock Price”IEEE 2013
[16] Zhao,LeiBaylorUniversityEmail:cxhdy@foxmail.comW ang, LinJapan Advanced Institute of Science and TechnologyEmail: linwang@jaist.ac.jp “Price Trend Prediction of Stock Market UsingOutlier Data Mining Algorithm” IEEE 2015
[17] Vishal S. Shirsat,Rajkumar S. Jagdale,S. N. Deshmukh, Department of Computer Science and IT Dr. B. A. M.

- University, Aurangabad, India “Document Level Sentiment Analysis from News Articles” 978-1-5386-4008-1/17/\$31.00 ©2017 IEEE
- [18] Xiaobo Zhang, Qingsong Yu School of Computer Science and Software Engineering East China Normal University Shanghai, China e-mail: Bobe_24@126.com, qsyu@cc.ecnu.edu.cn “Hotel Reviews Sentiment Analysis Based on Word Vector Clustering” 2017 2nd IEEE International Conference on Computational Intelligence and Applications
- [19] Snehal Kale, Vijaya Padmadas ,Department of Computer Science & Engg. Thodomal Shahani Engineering College Mumbai, India “Sentiment Analysis of Tweets Using Semantic Analysis” 978-1-5386-4008-1/17/\$31.00 ©2017 IEEE
- [20] M.Trupthi, Suresh Pabboju, G.Narasimha, Computer Science Department, JNTUH, Jagital, Telangana State, India “SENTIMENT ANALYSIS ON TWITTER USING STREAMING API” 2017 IEEE 7th International Advance Computing Conference
- [21] Ana Valdivia, M. Victoria Luzón, and Francisco Herrera, University of Granada “Sentiment Analysis in TripAdvisor” 1541-1672/17/\$33.00 © 2017 IEEE
- [22] Zeenia Singla, Sukhchandan Randhawa, and Sushma Jain,Department of Computer Science and Engineering, Thapar University, Patiala “STATISTICAL AND SENTIMENT ANALYSISOF CONSUMER PRODUCT REVIEWS” IEEE 8th ICCCNT 2017
- [23] Mauro Dragoni,Fondazione Bruno KesslerGiulio Petrucci,Fondazione Bruno Kessler, University of Trento E-mail: [dragoni.petrucci]@fbk.eu “A Neural Word Embeddings Approach ForMulti-Domain Sentiment Analysis” OI 10.1109/TAFFC.2017.2717879, IEEE Transactions on Affective Computing
- [24] Chloé Clavel and Zoraida Callejas “Sentiment analysis: from opinion mining to human-agent interaction” DOI 10.1109/TAFFC.2015.2444846, IEEE Transactions on Affective Computing
- [25] Shenghua Liu, Xueqi Cheng, Fuxin Li, and Fangtao Li “TASC:Topic-Adaptive Sentiment Classificationon Dynamic Tweets” DOI10.1109/TKDE.2014.2382600, IEEE Transactions on Knowledge and Data Engineering
- [26] Wei Zhao, Ziyu Guan*, Long Chen, Xiaofei He,Fellow, IAPR,Deng Cai, Beidou Wang and Quan Wang “Weakly-supervised Deep Embedding for Product Review Sentiment Analysis” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. ?, NO. ?, AUGUST 2017
- [27] Abdullah Alfarrarjeh 1, Sumeet Agrawal 2, Seon Ho Kim 3, Cyrus Shahabi 4Integrated Media Systems Center, University of Southern California, Los Angeles, CA 90089, USA “Geo-spatial Multimedia Sentiment Analysis in Disasters” IEEE 2017 International Conference on Data Science and Advanced Analytics
- [28] Nishant Suman, P K Gupta, Pankaj Sharma Department of Computer Science and Engineering Jaypee University of Information Technology, Solan, “Analysis of Stock Price Flow Based on Social Media Sentiments” IEEE 2017 International Conference on Next Generation Computing and Information Systems
- [29] Rushlene kaur bakshi, Navneet Kaur, Ravneet kaur, Gurpreet Kaur “Opinion mining and sentiment analysis” 978-9-3805-4421-2/16/\$31.00©2016 IEEE
- [30] Aliza Sarlan1, Chayanit Nadam2, Shuib Basri3Computer Information Science University Teknologi PETRONAS, Perak, Malaysia,aliza_sarlan@petronas.com.my; hayanit171@gmail.com; shuib_basri@petronas.com. “Twitter Sentiment Analysis” IEEE 2014 International Conference on Information Technology and Multimedia