

EFFECTUAL AND PRAGMATIC APPROACH TOWARDS USER BELIEF, PROMINENCE AND SENTIMENT MINING FROM SOCIAL MEDIA

Shrawan Choudhary

M.Tech. Research Scholar

*Department of Computer Science and
Engineering*

Government Engineering College

Bikaner, Rajasthan, India

Surendra Singh Choudhary

Assistant Professor

*Department of Computer Science and
Engineering*

Government Engineering College

Bikaner, Rajasthan, India

ABSTRACT

Sentiment Analysis is a process of determining whether a piece of writing, text (blogs, user comments, review websites, community websites) is positive, negative or neutral. It's also known as opinion mining. It's aim to determine the opinion or attitude of a speaker w.r.t some topic or the overall contextual polarity of a document. Analysis of public information from social media could yield interesting results and insights into the world of public opinions about almost any product, service or personality. Social network data is one of the most effective and accurate indicators of public sentiment. Sentiment analytics is used to examine large sets of data which may contain diversity of different types of data; it can be used to find correlating previously not known variables, finding the trends in the market, checking preferences of customers and finding out data about various businesses and institutions. The main aim of sentiment analysis is to aid in better and informative decision making for the firms by taking advantage of capable data-scientists, genius model makers as well as other trained scientists to verify chunks of information that may be unused by the traditional programs.

Keywords: *Big Data Analytics, Emotions Mining, Social Media Analytics, User Belief Mining*

INTRODUCTION

Big data analysis is done by using specialized software's for over-the-top level analysis and specialized fields such as statistical-analysis, data-mining, text- analysis and predictive analysis. Modern tools such as Business Intelligence as well as Visualization is majorly used in big data analytics. Also for the regularly updating data as well as continuous data update poses a problem in the analysis process- i.e. data analytics in case of mobile apps, and real time data in airlines as well as petroleum pipes. Thus, organizations and firms famous for their data-analytics proficiency have switched to modern and more innovative technologies in the filed for e.g., Hive, NoSQL, Hadoop Tools, YARN etc. which can be used to process clustered information which are of big and sundry nature with the help of their open-source nature.

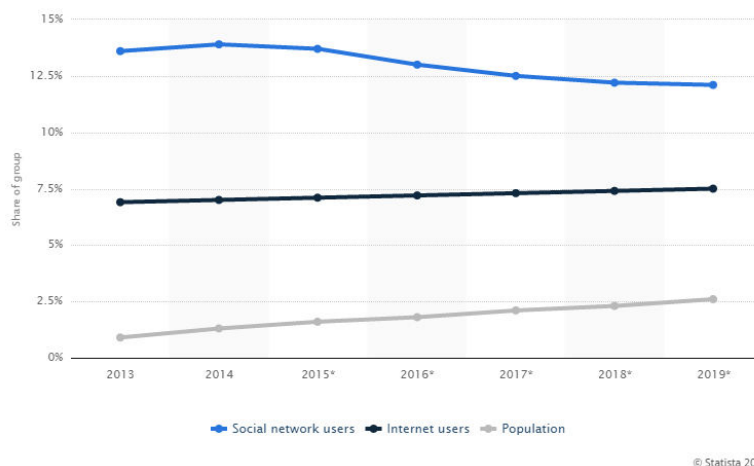


Figure 1. Statistical Description of Twitter Users in India (2012-2019)

REVIEW OF PROMINENT LITERATURE

As literature review is one of the key tasks that is performed for defend and justify the research work, following are the excerpts from assorted research manuscripts working on same direction as well as perspectives.

Poria (2015) presented that the sentiment extraction, scoring and predictive analytics is done which is now days very popular and prevalent in social as well as classical media platforms.

Guo (2016) presented a unique and effective research work in this research manuscript on the implementation of opinion sentiment data extraction on HADOOP platform.

Davis (2016) presented the implementation aspects of social media platform with their related performance factors.

Mamgain N (2016) presented in this paper a thorough effort to dive into the novel domain of performing sentiment analysis of people's opinions regarding top colleges in India.

Ahsan U (2016) proposed to capture sentiment information of such social event images leveraging their visual content.

Li M (2016) presented a methodological framework to collect, pre-process, analyse and map citizen sentiment from Twitter in helping the Governments monitor their citizens' moods is proposed based on the prior works.

Desai M (2016) in this paper, firstly present the sentiment analysis process to classify highly unstructured data on Twitter.

Peng Y (2016) in this paper a simple, effective computation pipeline is proposed, which uses simple drug-related classification and sentiment analysis to extract ADEs on Twitter.

Yadav P (2016) focused more on how emoticons play an important role in sentiment analysis. Various factors that affect sentiment analysis are discussed briefly in this paper. Also various issues like sarcasm detection, multilingualism, handling acronyms and slang language, lexical variation and dynamic dictionary handling are discussed.

Kumar P (2016) perform an assessment examination of general's conclusions mined from the well known smaller scale blogging site Twitter.

Pandhare KR (2016) employed logistic regression and SVM classifier. The work also used some text mining, machine learning, statistics and natural language processing techniques to extract meaningful information from raw tweets.

Dai X (2016) summarized the recent classification methods using social media in public health. These methods rely on bag-of-words (BOW) model and have difficulty grasping the semantic meaning of texts.

Sehgal D (2016) used the twitter data for the business purpose and industrial or social purpose according to our data requirement and processing the data.

Ramteke J (2016) proposed a two stage framework which can be used to create a training data from the mined Twitter data without compromising on features and contextual relevance.

Siddiqua UA (2016) introduced a set of rules for the rule-based classifier based on the occurrences of emoticons and sentiment-bearing words, whereas several sentiment lexicons are applied to train the Naive-Bayes classifier.

Qaisi LM (2016) worked on the twitter sentiment analysis for cloud providers with case study of Azure and AWS.

Kulcu S (2016) use three of the most used machine learning methods, namely Naive Bayes, Complementary Naive Bayes, and Logistic Regression, and for linking tweets to news items, Natural Language Processing (NLP) techniques are used, including Zemberek NLP library for stemming and morphological analysis and then bag-of words method for mapping.

Li X (2016) obtain the results for NGrams term of tokens. Finally, human has labelled the data and this may involve some mistakes in the labelling process.

Qian S (2016) propose a novel multi-modal event topic model (mmETM), which can effectively model social media documents, including long text with related images, and learn the correlations between textual and visual modalities to separate the visual-representative topics and non-visual-representative topics.

Buettner R (2016) presented the systematic literature review of twitter research from a socio-political revolution perspective.

Al-Rubaiee H (2016) Firstly, document's Pre-processing are explored on the dataset. Secondly, Naive Bayes and Support Vector Machines (SVMs) are applied with different feature selection schemes like TF-IDF (Term Frequency–Inverse Document Frequency) and BTO (Binary-Term Occurrence).

Following are the research aspects and methodology involved in the present research work:

1. Extraction of Live Tweets from Twitter Social Media

2. Refining and Cleaning of Data Set
3. Training of the Dataset and Machine Learning Approach
4. Analytics and Matching with the Emotional Words in English Dictionary using metadata and data dictionary
5. Allocation of Sentiment and User Belief Score
6. Analytics of the Results and Allocation of User Belief of Sentiment Score
7. Extraction and Log of the Results with Predictive Mining

PROPOSED FLOW OF WORK AND RESULTS

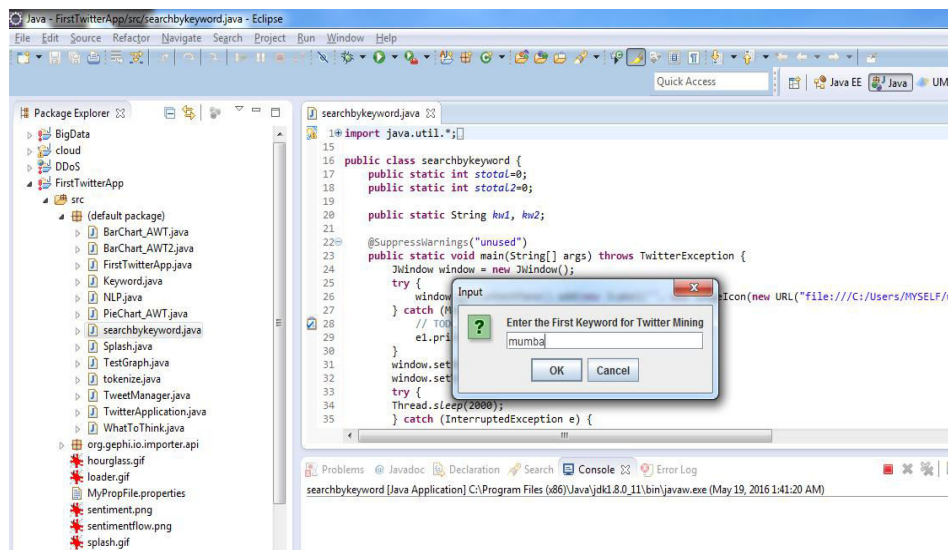


Figure 2 – Searching Panel

It depicts the searching panel of keywords.

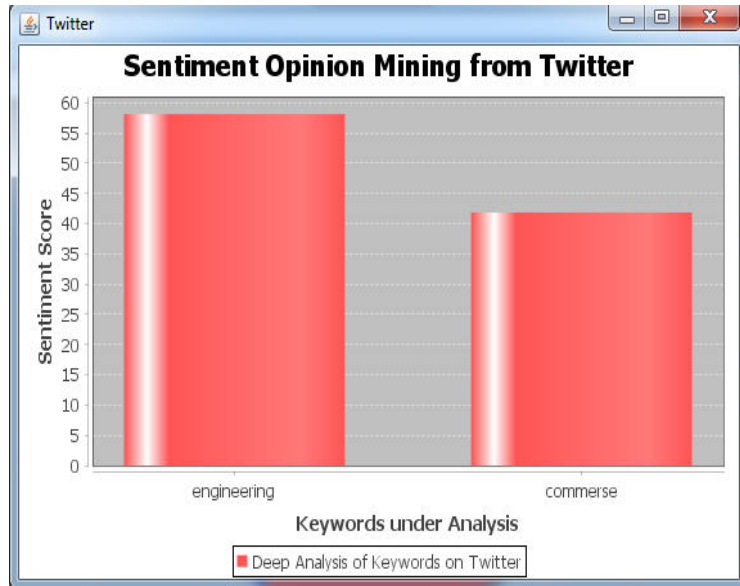


Figure 3: Result Analysis

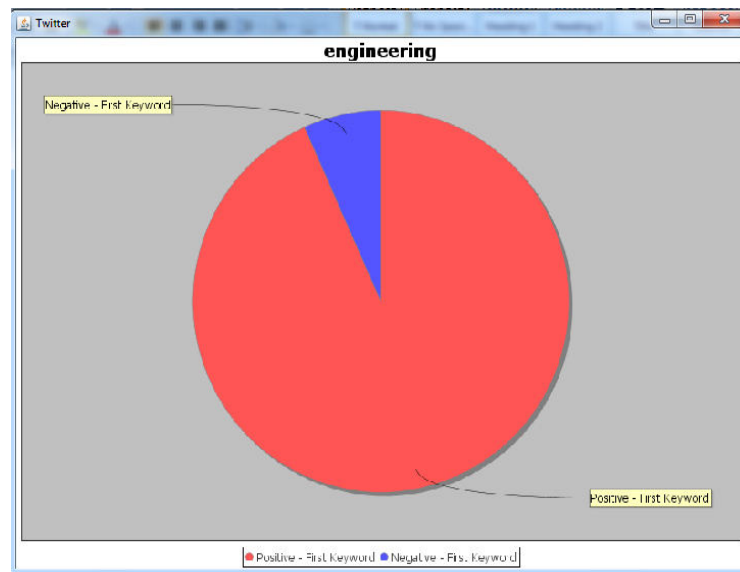


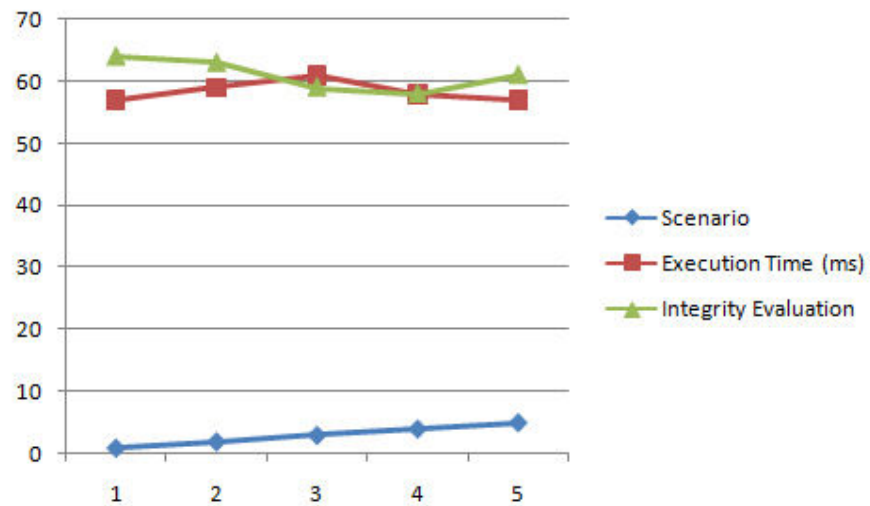
Figure 4: Result Analysis

To perform basic frequency analysis, tweet-retweet analysis and sentiment analysis on twitter data. Using the above results, the comparative analysis in multiple persons or objects can be done.

To perform basic frequency analysis, tweet-retweet analysis and sentiment analysis on twitter data. Using the above results, the comparative analysis in multiple persons or objects can be done.

Table 1. Integrity Analysis under different scenarios

Scenario	Execution Time (ms)	Integrity Evaluation
1	57	64
2	59	63
3	61	59
4	58	58
5	57	61



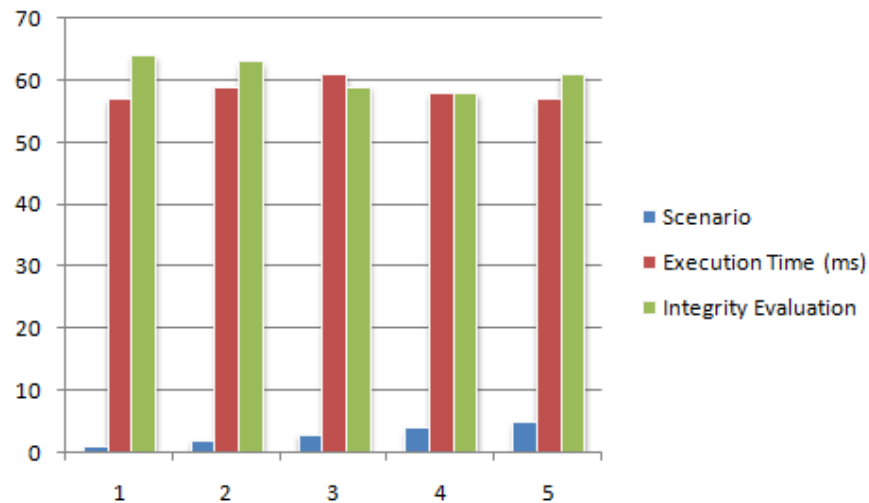


Figure 5: Integrity and Consistency Analysis

Figure 5 shows the line graph of consistency and integrity of the results with different keywords. The results show that the execution time is integrity aware and consistent without any ambiguity.

Conclusion

Sentiment analysis, also referred to as Opinion Mining, implies extracting opinions, emotions and sentiments in text. One of the most common applications of sentiment analysis is to track attitudes and feelings on the web, especially for tracking products, services, brands or even people. The main idea is to determine whether they are viewed positively or negatively by the viewers or users on the social media. Twitter is a popular micro blogging service where users create status messages.

The recommendation associated with this work is towards the predictive mining on assorted events, celebrities or popularity factors in real time domain. The predictions associated with business including stock market can be done effectually with this implementation.

For future scope of the work, following techniques can be used in hybrid approach to better and efficient results –

- Particle Swarm Optimization
- HoneyBee Algorithm
- Simulated Annealing
- Firefly algorithm
- Bat Algorithm
- River Formation Dynamics

REFERENCES

- [1] Poria, S. Cambria, E., Fu, J., Bisio, F., & (2015). AffectiveSpace 2: Enabling Affective Intuition for Concept-Level Sentiment Analysis. In AAAI (pp. 508-514).
- [2] Guo, Y., Rao, J., Cheng, D., & Zhou, X. (2016). ishuffle: Improving hadoop performance with shuffle-on-write. IEEE Transactions on Parallel and Distributed Systems.
- [3] Davis, B., Hürlimann, M., Cortis, K., Freitas, A., Handschuh, S., & Fernández, S. (2016, September). A Twitter Sentiment Gold Standard for the Brexit Referendum. In Proceedings of the 12th International Conference on Semantic Systems (pp. 193-196). ACM.
- [4] Zimbra D, Ghiassi M, Lee S. Brand-related Twitter sentiment analysis using feature engineering and the dynamic architecture for artificial neural networks. In System Sciences (HICSS), 2016 49th Hawaii International Conference on 2016 Jan 5 (pp. 1930-1938). IEEE.
- [5] Wehrmann J, Becker W, Cagnini HE, Barros RC. A character-based convolutional neural network for language-agnostic Twitter sentiment analysis. In Neural Networks (IJCNN), 2017 International Joint Conference on 2017 May 14 (pp. 2384-2391). IEEE.
- [6] Cambria E. Affective computing and sentiment analysis. IEEE Intelligent Systems. 2016 Mar;31(2):102-7.
- [7] Zimbra D, Ghiassi M, Lee S. Brand-related Twitter sentiment analysis using feature engineering and the dynamic architecture for artificial neural networks. In System

- Sciences (HICSS), 2016 49th Hawaii International Conference on 2016 Jan 5 (pp. 1930-1938). IEEE.
- [8] Jimenez-Zafra SM, Valdivia MT, Camara EM, Urena-Lopez LA. Studying the Scope of Negation for Spanish Sentiment Analysis on Twitter. IEEE Transactions on Affective Computing. 2017 Apr 12.
- [9] Subrahmanian VS, Azaria A, Durst S, Kagan V, Galstyan A, Lerman K, Zhu L, Ferrara E, Flammini A, Menczer F. The DARPA Twitter bot challenge. Computer. 2016 Jun;49(6):38-46.
- [10] Furini M, Montangero M. TSentiment: On gamifying Twitter sentiment analysis. InComputers and Communication (ISCC), 2016 IEEE Symposium on 2016 Jun 27 (pp. 91-96). IEEE.
- [11] Tang D, Wei F, Qin B, Yang N, Liu T, Zhou M. Sentiment embeddings with applications to sentiment analysis. IEEE Transactions on Knowledge and Data Engineering. 2016 Feb 1;28(2):496-509.
- [12] Mamgain N, Mehta E, Mittal A, Bhatt G. Sentiment analysis of top colleges in India using Twitter data. InComputational Techniques in Information and Communication Technologies (ICCTICT), 2016 International Conference on 2016 Mar 11 (pp. 525-530). IEEE.
- [13] Ahsan U, De Choudhury M, Essa I. Towards using visual attributes to infer image sentiment of social events. InNeural Networks (IJCNN), 2017 International Joint Conference on 2017 May 14 (pp. 1372-1379). IEEE.
- [14] Li M, Ch'ng E, Chong A, See S. The new eye of smart city: Novel citizen Sentiment Analysis in Twitter. InAudio, Language and Image Processing (ICALIP), 2016 International Conference on 2016 Jul 11 (pp. 557-562). IEEE.
- [15] Desai M, Mehta MA. Techniques for sentiment analysis of Twitter data: A comprehensive survey. InComputing, Communication and Automation (ICCCA), 2016 International Conference on 2016 Apr 29 (pp. 149-154). IEEE.
- [16] Peng Y, Moh M, Moh TS. Efficient adverse drug event extraction using Twitter sentiment analysis. InAdvances in Social Networks Analysis and Mining

- (ASONAM), 2016 IEEE/ACM International Conference on 2016 Aug 18 (pp. 1011-1018). IEEE.
- [17] Yadav P, Pandya D. SentiReview: Sentiment analysis based on text and emoticons. In Innovative Mechanisms for Industry Applications (ICIMIA), 2017 International Conference on 2017 Feb 21 (pp. 467-472). IEEE.
- [18] Kumar P, Choudhury T, Rawat S, Jayaraman S. Analysis of Various Machine Learning Algorithms for Enhanced Opinion Mining Using Twitter Data Streams. In Micro-Electronics and Telecommunication Engineering (ICMETE), 2016 International Conference on 2016 Sep 22 (pp. 265-270). IEEE.
- [19] Pandhare KR, Shah MA. Real time road traffic event detection using Twitter and spark. In Inventive Communication and Computational Technologies (ICICCT), 2017 International Conference on 2017 Mar 10 (pp. 445-449). IEEE.
- [20] Dai X, Bikdash M, Meyer B. From social media to public health surveillance: Word embedding based clustering method for twitter classification. In SoutheastCon, 2017 2017 Mar 30 (pp. 1-7). IEEE.
- [21] Sehgal D, Agarwal AK. Sentiment analysis of big data applications using Twitter Data with the help of HADOOP framework. In System Modeling & Advancement in Research Trends (SMART), International Conference 2016 Nov 25 (pp. 251-255). IEEE.
- [22] Ramteke J, Shah S, Godhia D, Shaikh A. Election result prediction using Twitter sentiment analysis. In Inventive Computation Technologies (ICICT), International Conference on 2016 Aug 26 (Vol. 1, pp. 1-5). IEEE.
- [23] Siddiqua UA, Ahsan T, Chy AN. Combining a rule-based classifier with weakly supervised learning for twitter sentiment analysis. In Innovations in Science, Engineering and Technology (ICISSET), International Conference on 2016 Oct 28 (pp. 1-4). IEEE.
- [24] Qaisi LM, Aljarah I. A twitter sentiment analysis for cloud providers: a case study of Azure vs. AWS. In Computer Science and Information Technology (CSIT), 2016 7th International Conference on 2016 Jul 13 (pp. 1-6). IEEE.

- [25] Kulcu S, Dogdu E. A scalable approach for sentiment analysis of turkish tweets and linking tweets to news. InSemantic Computing (ICSC), 2016 IEEE Tenth International Conference on 2016 Feb 4 (pp. 471-476). IEEE.
- [26] Li X, Xie H, Rao Y, Chen Y, Liu X, Huang H, Wang FL. Weighted multi-label classification model for sentiment analysis of online news. InBig Data and Smart Computing (BigComp), 2016 International Conference on 2016 Jan 18 (pp. 215-222). IEEE.
- [27] Qian S, Zhang T, Xu C, Shao J. Multi-modal event topic model for social event analysis. IEEE Transactions on Multimedia. 2016 Feb;18(2):233-46.
- [28] Buettner R, Buettner K. A systematic literature review of twitter research from a socio-political revolution perspective. InSystem Sciences (HICSS), 2016 49th Hawaii International Conference on 2016 Jan 5 (pp. 2206-2215). IEEE.
- [29] Al-Rubaiee H, Qiu R, Li D. Identifying Mubasher software products through sentiment analysis of Arabic tweets. InIndustrial Informatics and Computer Systems (CIICS), 2016 International Conference on 2016 Mar 13 (pp. 1-6). IEEE.