International Conference on Computing, Communication and Automation (ICCCA2016)

Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey

Mitali Desai
Computer Engineering Department
Sarvajanik College of Engineering and Technology
Surat, India
mitalidesai17@gmail.com

Mayuri A. Mehta
Computer Engineering Department
Sarvajanik College of Engineering and Technology
Surat, India
mayuri.mehta@scet.ac.in

***Abstract*— The World Wide Web has intensely evolved a novel way for people to express their views and opinions about different topics, trends and issues. The user-generated content present on different mediums such as internet forums, discussion groups, and blogs serves a concrete and substantial base for decision making in various fields such as advertising, political polls, scientific surveys, market prediction and business intelligence. Sentiment analysis relates to the problem of mining the sentiments from online available data and categorizing the opinion expressed by an author towards a particular entity into at most three preset categories: positive, negative and neutral. In this paper, firstly we present the sentiment analysis process to classify highly unstructured data on Twitter. Secondly, we discuss various techniques to carryout sentiment analysis on Twitter data in detail. Moreover, we present the parametric comparison of the discussed techniques based on our identified parameters.**

Keywords*— *Sentiment analysis; machine learning; opinion mining; Twitter

I. INTRODUCTION

Social Computing is an innovative and growing computing exemplar for the analysis and modeling of social activities taking place on various platforms. It is used to produce intellectual and interactive applications to derive efficient results [1]. The wide availability of social media sites provides individuals to share their sentiments or opinions about a particular event, product or issue. Mining of such informal and homogeneous data is highly useful to draw conclusions in various fields. Though, the highly unstructured format of the opinion data available on web makes the mining process challenging [2].

Textual information present on web is majorly classified into either of the two categories: fact data and sentiment data [3]. Fact data are the objective terminologies concerning different entities, issues or events. Whereas sentiment data are the subjective terms, that define individual's opinions or beliefs for a particular entity, product or event. Sentiment analysis is the process of recognizing and classifying different sentiments conveyed online by the individuals to derive the writer's approach towards a specific product, topic or event is positive, negative or neutral. Sentiment analysis has three major component of study as follows: sentiment holder i.e. subject, sentiment itself i.e. belief and object i.e. the topic about which the subject has shared the sentiment. An object is an entity that represents a definite person, item, product, issue, event, topic or any organization [3-7]. Sentiment analysis is carried out at different levels ranging from coarse level to fine level. The coarse level sentiment analysis determines the sentiment of the whole manuscript or document. The fine level sentiment analysis, whereas focuses on the attributes. Sentiment analysis of Twitter data is carried out on sentence level which comes in between coarse level and fine level. In the sentiment analysis process, the sentiments present in the text are of two types: Direct and Comparative. The direct sentiments in text are independent from other objects in the same sentence [7]. For example "the picture quality of this camera is great." However, the comparative sentiments in the text denote the comparison of different objects within the same sentence. For example "car x is cheaper than car y."

The existing sentiment analysis techniques are useful in various applications such as disaster relief and humanitarian assistance, marketing and trade predictions, checking political polls, advertising market, scientific surveys, checking customer loyalty, finding job opportunities, population health care and understanding students' learning experiences [1-7].

In this paper, we present a sentiment analysis process for Twitter data. Twitter is a micro-blogging site that is rapidly growing in terms of number of users [8-9]. Moreover, Tweets are mostly public and limited to 140 characters that simplify the identification of emotions in text [9-12]. Though, the abundance of data, use of short forms, timing of different posts, and diversity of language make the sentiment analysis process difficult for Twitter data.

The rest of the paper is organized as follows: In section II, we discuss the existing work in the field of sentiment analysis. Section III describes the methodology to carryout sentiment analysis. Section IV presents numerous supervised machine learning algorithms used to conduct sentiment analysis and their comparison based on the identified parameters. Finally, Section V specifies the conclusion and future directions.

II. RELATED WORK

In current years, a voluminous amount of research has been conducted in the sentiment analysis domain. In [7],

ISBN: 978-1-5090-1666-2/16/$31.00 ©2016 IEEE

authors have proposed a technique to classify students' data generated on Twitter into various categories to encounter students' various problems. In [13], authors have presented the logical approach to mine the sentiments shared on different social media platforms. They have analysed the sentiments of the text using combinatory categorical grammar, annotation, lexicon acquisition and semantic networks. The basic techniques of sentiment classification and the methods for data collection are presented in [14]. The accuracy of classification process with selected feature vector is verified for the electronic products domain using various classifiers such as Nave Bayes, Maximum Entropy, Support Vector Machine, and Ensemble classifiers in [15]. In [16], authors have introduced a hybrid method that is a combination of the usage of sentiment lexicons with a machine learning classifier for polarity detection of subjective texts in the consumer-products domain. In [17], authors have proposed a batch of machine learning methods with semantic analysis to classify the sentence and reviews of different products based on twitter data using WordNet for better accuracy. In [18], authors have examined the performance of different classifiers such as Naïve Bayes, SMO, SVM and Random Forest to classify Twitter data. In [19], authors have presented a technique to normalize the noisy or irrelevant tweets and classify them according to the polarity i.e. positive or negative. Moreover, they have employed a mixture model approach to generate different sentimental words. The generated words were later used as feature indicators in the classification model. Authors have introduced a novel method to predict sentiments about stocks using various monetary communication boards and performed an automatic prediction for the stock market using web sentiments in [20]. In [21], authors have examined the performance of sentiment analysis in e-learning domain using various methods of feature selection i.e. CHI statistics, Mutual Information (MI) and Information Gain (IG). In [22], authors have proposed an automatic sentiment classifier to classify reviews of Brazilian TV shows into positive or negative category and possessed 90% of accuracy. Authors have demonstrated a system to extract the Tweets and classify them using domain oriented seed based enrichment technique to reduce the information loss in the knowledge domain in [23]. In [24], authors have investigated numerous combinations of different preprocessing levels, machine learning techniques and features combining with neutral class to analyze real-time students' feedback. In [25], authors have developed an enhanced sentiment classification method that can detect and remove anomalies from Twitter data in addition to the classification.

III. METHODOLOGY FOR SENTIMENT ANALYSIS

The sentiment analysis of Twitter data is an emerging field that needs much more attention. Fig. 1 shows the steps to carry out the process of sentiment analysis on Twitter data.

Firstly, the collected Twitter data is pre-processed to perform the data cleaning. Secondly, the important features are extracted from the clean text, applying any of the feature selection methods. Thirdly, the portion of the data is manually labeled as positive or negative Tweets to prepare a training set. Finally, the extracted features and the labeled training set are provided as an input to the built classifier to classify the remaining data i.e. test set. Each of the processing steps is discussed thoroughly in the following sub-sections.

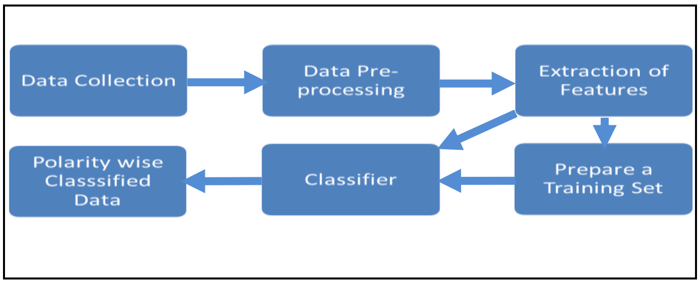

Fig. 1. Sentiment analysis process of Twitter data

A. *Data Sources*

Selection of data source to conduct the sentiment analysis plays a significant role. Social media platforms as the data sources are broadly categorized into three general categories: blogs, micro-blogging sites, and review site [13-16]. Among all categories, a micro-blogging site such as Twitter has gained higher popularity due to its limited strength of the content and publically availability of data. From the following statistics of the Twitter growth rate, it's evident to use Twitter as the data source for sentiment analysis.

- Twitter Growth Rate Statistics

Approximately 6,000 tweets are tweeted on Twitter on per second basis. It resembles to 350,000 tweets sent per minute and 500 million tweets per day. That makes it around 200 billion tweets per year. In Twitter's history, the number of Tweets increased from 5,000 tweets per day in 2007 [8] to 500,000,000 tweets per day in 2013, that is approximately a six orders of magnitude [8]. At the intermediate stages it has the statistics of 300,000 tweets per day in 2008 [9], 2.5 million tweets per day in 2009 [9], 35 million tweets per day in 2010 [8], 200 million tweets per day in 2011 [10]. And 340 million tweets per day six years after the emergence of Twitter i.e. on March 21, 2012 [12]. This statistics conclude the use of Twitter for our research.

- Twitter Studies

As per the recent work, the studies carry out on Twitter data are in the field of health care, marketing, politics, advertising market, athletics etc. Analysis techniques used in these studies include qualitative content analysis, network or graph analysis, linguistic or psycholinguistic analysis, word clouds and histograms [5]. In addition, Twitter has been voted as the most promising source for the studies such as community or influence detection, topic discovery, market and business predictions, recommendation systems and tweet classification.

- Tweets

The message posted on Twitter is called Tweet, which is limited to 140 characters. Tweets are generally composed of one of the followings [10] [13] [14]: text, links, emoticons, and images. A six seconds video is even added as a Tweet component in 2012 [8-12]. Based on these components the mining is applied to classify text, links, images, emoji or emoticons and even videos. The Tweets contains three notations including hashtags (#), retweets (RT) and account Id (@).

B. *Twitter Data Collection Methods*

The three possible ways to collect Tweets for research are as follows [11]:

- Data repositories such as UCI, Friendster, Kdnuggets, and SNAP
- APIs: Twitter provides two types of APIs such as search API and stream API. Search API is used to collect Twitter data on the basis of hashtags and stream API is used to stream real time data from Twitter
- Automated tools that are further classified into premium tools such as Radian6 [18], Sysmos, Simplify360, Lithium and non-premium tools such as Keyhole, Topsy, Tagboard and SocialMention

C. *Data Preprocessing*

Mining of Twitter data is a challenging task. The collected data is raw data. In order to apply classifier, it is essential to pre-process or clean the raw data. The pre-processing task involves uniform casing, removal of hashtags and other Twitter notations (@, RT), emoticons, URLs, stop words, decompression of slang words and compression of elongated words. The following steps show the pre-processing procedure.

- Remove the Twitter notations such as hashtags (#), retweets (RT), and account Id (@).
- Remove the URLs, hyperlinks and emoticon. It is necessary to remove non letter data and symbols as we are dealing with only text data.
- Remove the stop words such as are, is, am etc. The stop words do not emphasize on any emotions, it is intended to remove them to compress the dataset.
- Compress the elongated words such as happyyy into happy.
- Decompress the slag words such as g8, f9. Generally slang words are adjectives or nouns and they contain the extreme level of sentiments. So it is necessary to decompress them.

D. *Feature Extraction*

The pre-processed dataset has various discrete properties. In feature extraction methods, we extract different aspects such as adjectives, verbs and nouns and later these aspects are identified as positive or negative to detect the polarity of the whole sentence. Followings are the widely used Feature Extraction methods.

- Terms Frequency and Term Presence: These features denote individual and distinct words and their occurrence counts.
- Negative Phrases: The presence of negative words can change the meaning or orientation of the opinion. So it is evident to take negative word orientation in account.
- Parts Of Speech (POS): Finding nouns, verbs, adjectives etc. as they are significant gauges of opinions.

E. *Sentiment Classification Techniques*

There are typically two techniques to identify sentiment of the text [7] [13] [26-32]: knowledge based technique and machine learning techniques.

Knowledge based technique is also called Lexicon based technique. The lexicon-based technique focuses on deriving the opinion based lexicons from the text and then identifying the polarity of those lexicons. Lexicons are the collection of known and precompiled sentiment terms. This approach is further classified into Dictionary-based approach and Corpus-based approach. In the Dictionary-based approach, we find the opinion oriented words, and then examine the dictionary to collect their synonyms and antonyms. Whereas in the Corpus-based approach, we create a list of opinion words and then based on their context specific orientations, we find additional related opinion words in a vast corpus. To conduct lexicon approach, a trivial set of words describing opinions is collected manually with their known orientations as a mean of pre-processing task. The set is then grown gradually by searching in the distinguished and widely used lexicon dictionary tool such as WordNet or Sentiful for their synonyms and antonyms [17-18].

Whereas the main objective of machine learning techniques is to develop the algorithm that optimizes the performance of the system using training data such as examples and/or past knowledge and experiences. The machine learning provides a solution of the sentiment classification problem in two sequential steps:

1) Develop and train the model using training set data i.e. already labeled data.
2) Classifying the unlabeled or unclassified data based on the trained or skilled model.

Machine learning techniques are further classified into supervised and unsupervised techniques [13] [15] [26-30]. To carry out sentiment analysis, typically the supervised machine learning techniques are used as we are dealing with subjective data. Supervised machine learning techniques highly depend on training data which are already labeled data unlike in the case of unsupervised machine learning techniques. Based on the provided training data, the classifier will classify the rest data i.e. test data. A large number of supervised machine learning algorithms such as Logistic Regression, Naïve Bayes, Decision Tree, Support Vector Machine (SVM), Random Forest, Maximum Entropy, and Bayesian Network are used for sentiment analysis [7] [13-26]. Choice of an appropriate algorithm for selected data and domain is a crucial step.

IV. SUPERVISED MACHINE LEARNING ALGORITHMS FOR SENTIMENT ANALYSIS

From out in-depth study of the supervised machine learning algorithms, it has been observed that the following machine learning algorithms are widely used and give average accuracy in majority of domains as well as with different types of data. Moreover, they provide consistent average speed of classification process irrespective of the size of input data while handling the outliers.

A. Naïve Bayes (NB) Approach

Naïve Bayes classifier [7] [17] [21] [26-30] is a simple probabilistic classifier that uses the concept of mixture models to perform classification. The mixture model relies on the assumption that each of the predefined classes is one of the components of the mixture itself. The components of the mixture model denote the probability of belongingness of any term to the particular component. Thus, they are also known as generative classifiers. Naïve Bayes classifier is a probabilistic classifier that uses the concept of Bayes Theorem and finds maximum prospect of probability of any word fitting to a particular given or predefined class. The probability P is defined as follows:

$$P\ (Xi \mid c) = \frac{Count\ of\ Xi\ in\ document\ of\ class\ c}{Total\ no\ of\ words\ in\ document\ of\ class\ c} \qquad (1)$$

Where X_i is a given term and c is a predefined class label. During the training phase, the incidence counts of the words are collected and stored in the hash tables. NB approach suffers from an assumption that the features are independent in the feature space.

As per the definition of probability, the document d is classified into class c using following equation:

$$c *= argmax\ P\ (c \mid d) \qquad (2)$$

B. Maximum Entropy

The maximum entropy relies on probability distribution estimation technique to perform classification. In this technique, firstly the categorized feature sets are converted into definite vectors using any of the encoding schemes. Secondly, this encoded vector is used to compute weights for each of the extracted features that can collectively support in determining the most prospective label for a feature set. It is used for various natural language processing tasks such as text classification. It depends on the probabilistic approach like Naïve Bayes [26-30]. The fundamental concept of maximum entropy is that if much information regarding the data is not known, the distribution should be extremely uniform. This constraint eliminates the probability of non-uniform distribution. The probability is derived from the categorized training data and denoted as expected values of extracted features as follows:

$$P\ (c \mid d) = \frac{1}{Z\ (d)\{\exp(\sum \lambda i\ fi\ (c, d))\}} \qquad (3)$$

Where $f_i(c, d)$ is a feature, λ_i is a parameter to be predicted and Z(d) is a normalization function. Unlike NB, maximum entropy doesn't make any assumption regarding the feature independency.

The motivating idea behind maximum entropy is building a uniform model that satisfies all the given constrains. For example, consider a four-way text classification problem with a constraint given as: on average 40% of documents having a word "professor" is labeled as faculty class. Innately, when given a document with "professor" word within, we assume that it has a 40% chance to be labeled as a faculty class, and a 20% chance to be labeled as each of the other three classes. If a document does not have "professor" word within then as per the law of uniform class distribution, we assume the probability of that document to be in each class is 25%. This model is precisely the maximum entropy model that conforms to each given or known constraint [30].

C. Support Vector Machine (SVM)

Support vector machine (SVM) solves the traditional text categorization problem effectively; generally outperforming Naïve Bayes as it supports the concept of maximum margin. The main principle of SVMs is to determine a linear separator that separates different classes in the search space with maximum distance i.e. with maximum margin [13-17]. If we represent the tweet using t, the hyper plane using h, and classes using a set Cj € {l, -1} into which the tweet has to be classified, the solution is written as follows equivalent to the sentiment of the tweet.

$$\vec{h} = \sum i\ ai\ Ci\ \vec{tj}, \qquad ai \geq 0 \qquad (4)$$

The idea of SVM is to determine a boundary or boundaries that separate distinct clusters or groups of data. SVM performs this task constructing a set of points and separating those points using mathematical formulas. Fig. 2 illustrates the data flow of SVM.

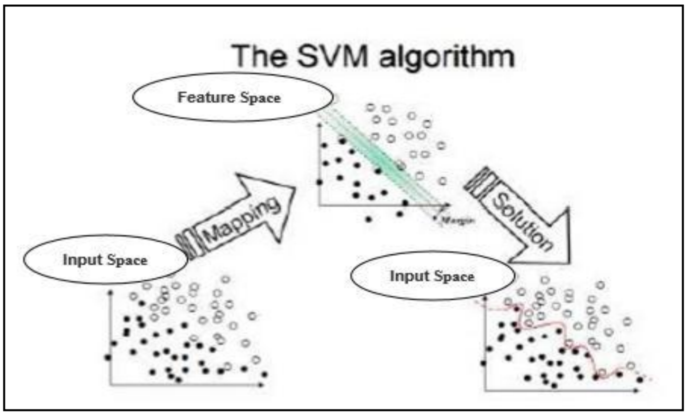

Fig. 2. Support Vector Machine (SVM) workflow [30]

D. *Random Forest*

Random Forest classifier is a tree-based classifier. It consists of numerous classification trees that can be used to predict the class label for a given data point based on the categorical dependent variable [19]. For a given data point, each tree votes for a particular class label and the class label gaining the maximum votes will be assigned to that data point. The error rate of this classifier depends on the correlation or association among any two trees in the forest in addition to the strength of definite or individual tree in the forest. In order to minimize the error rate, the trees should be strong and the degree of associativity should be as less as possible.

In the classifier tree, the internal nodes are represented as the features, the edges leaving a node are represented as tests on the feature's weight, and the leaves are represented as class categories. It performs classification preliminary from the root node and moves incrementally downward until a leaf node is detected. The document is then classified in the category that labels the leaf node. This algorithm is used in many applications of speech and language processing.

E. *Evaluation of Supervised Machine Learning Algorithms*

From our in-depth study of the above supervised machine learning algorithms used to perform sentiment analysis, we have identified several parameters such as understanding complexity, theoretical accuracy, theoretical training speed, performance with small number of observations and type of the classifier.

Understanding complexity refers to the technical difficulties to understand the algorithm. Theoretical accuracy is the theoretical measure of how accurately the algorithm can classify the test set data according to the provided training data. Theoretical training speed refers to how fast the data can be trained. Performance is related to the accuracy of the algorithm. In general, accurate algorithms have good performance. Classifier refers to the type of classifier the algorithm belongs. There are different types of classifiers such as linear classifiers, probabilistic classifiers, decision based classifier [7] [13] [26-32].

TABLE I. PARAMETRIC COMPARISON OF THE SUPERVISED MACHINE LEARNING ALGORITHMS

Algorithm	NB	SVM	Maximum Entropy	Random Forest
Understanding complexity	Very less	High	Moderate	Moderate
Theoretical accuracy	Low	High	Moderate	High
Theoretical Training Speed	High	High	Moderate	Low
Performance with small no. of Observations	High	Low	Low	Low
Classifier	Probabilistic	Linear	Probabilistic	Tree Based

From the parametric comparison shown in Table I, we can conclude that Naïve Bayes algorithms are the simplest and easiest to understand and implement compare to Support Vector Machine and Maximum Entropy. However, it suffers from lower accuracy due to its simple Bayesian probability assumption. Whereas Support Vector Machine provides the better accuracy but it doesn't support the automatic learning of features. Maximum Entropy provides the moderated accuracy but supports the automatic learning of features. Random Forest is based on decision tree method, which gives high accuracy with automatic feature learning.

Though, the implementation accuracy of all these algorithms highly depends on the numerous factors such as domain chosen, data source, amount of data and preprocessing method applied on the data.

F. *Evaluation Parameters*

In common, the performance of sentiment classification techniques is estimated using four indicators as follows: Accuracy, Precision, Recall and F1-score [13-32]. These indicators are computed using the confusion matrix given in Table II:

TABLE II. THE CONFUSION MATRIX

#	Predicted Positives	Predicted Negatives
Actual Positive Cases	Number of True Positive Cases (TP)	Number of False Negative Cases (FN)
Actual Negative Cases	Number of False Positive Cases (FP)	Number of True Negative Cases (TN)

These indicators are defined by the following equations:

$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (8)$$

Accuracy is defined as all true predicted cases against all predicted cases. If we receive 100% accuracy, it denotes that the predicted cases are precisely the same as the actual cases. Precision is defined as the true positive predicted cases against all positive predicted cases. Recall is defined as the true positive predicted cases against all actual positive cases. F1 is a harmonic average of the precision and the recall.

V. CONCLUSION AND FUTURE WORK

In this paper, we have firstly presented the detailed procedure to carryout sentiment analysis process to classify highly unstructured data of Twitter into positive or negative categories. Secondly, we have discussed various techniques to carryout sentiment analysis on Twitter data including knowledge based technique and machine learning techniques.

Moreover, we presented the parametric comparison of the discussed supervised machine learning techniques based on our identified parameters. It has been found that various techniques applied for sentiment analysis are domain specific and language specific.

Hence, the future opportunities in the domain of sentiment analysis include developing a technique to perform sentiment classification that can be applicable to any data regardless of domain. In addition, language diversity in social media data is a key issue which is required to be eliminated in future. Moreover, some of the more crucial challenges of Natural Language Processing (NLP) can also be used as further developments in sentiment analysis, such as hidden or veiled sentiment detection, satire detection, comparison or association handling and emoticon detection.

REFERENCES

[1] I. King, J. Li and K. T. Chan, “A Brief Survey of Computational Approaches in Social Computing”, in *Proc. of Int. Joint Conf. on Neural Network*, 2009, pp. 2699-2706.

[2] S. R. Barahate and V. M. Shelake, “A Survey and Future Vision of Data mining in Educational Field”, in *Proc. 2nd Int. Conf. on Advanced Computing & Communication Technology*, 2012, pp. 96-100.

[3] Bing Liu, N. Indurkhya and F. J. Damerau, Handbook of Natural Language Processing, Second Edition, 2010, pp. 1-3860-68.

[4] M. Dredze , “How Social Media Will Change Public Health”, *IEEE Intelligent Systems*, 2012, pp. 1541-1672.

[5] G. Siemens and P. Long, “Penetrating the fog: Analytics in learning and education”, *Educause Review*, 2011, vol. 46, no. 5, pp. 30-32.

[6] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," in *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions*, 2010, vol. 40, no.6, pp. 601-618.

[7] X. Chen, M. Vorvoreanu and K. Madhavan, “Mining Social Media Data to Understand Students’ Learning Experiences”, *IEEE Transaction*, 2014, vol. 7, no. 3, pp. 246-259.

[8] Weil, Kevin (VP of Product for Revenue and former Big Data engineer, Twitter Inc.), "Measuring Tweets." Twitter Official Blog, February 22, 2010. [Online]. Available: http://www.internetlivestats.com/twitter-statistics. [Accessed: 19-Oct-2015].

[9] Krikorian, Raffi (VP, Platform Engineering, Twitter Inc.), "New Tweets per second record, and how!" Twitter Official Blog. August 16, 2013.[Online]. Available: https:// blog.twitter.com/ 2013/ new-tweets-per- second-record-and-how. [Accessed: 19-Oct-2015].

[10] Twitter Engineering, "200 million Tweets per day." Twitter Official Blog. June 30, 2011. [Online]. Available: https://blog.twitter.com/2011/200-million-tweets-per-day. [Accessed: 19-Oct-2015].

[11] “Three Cool and Inexpensive Tools to Track Twitter Hashtags”, June 11, 2013. [Online]. Available http://dannybrown.me/2013/06/11/three-cool-toolstwitterhashtags/ [Accessed: 19-Oct-2015].

[12] "Twitter turns six." Twitter Official Blog. March 21, 2012. [Online].

Available: https://blog.twitter.com/2012/twitter-turns-six. [Accessed: 19-Oct-2015].

[13] N. Kasture and P. Bhilare, “An Approach for Sentiment analysis on social networking sites”, *Computing Communication Control and Automation (ICCUBEA),* 2015, pp. 390-395.

[14] S. Bhuta, A. Doshi, U. Doshi and M. Narvekar, “A review of techniques for sentiment analysis Of Twitter data”, *Issues and Challenges in Intelligent Computing Techniques (ICICT)*, 2014, pp. 583-591.

[15] M. S. Neethu and R. Rajasree, “Sentiment Analysis in Twitter using Machine Learning Techniques”, in *4th Int. Conf. of Computing, Communications and Networking Technologies (ICCCNT)*, 2013, pp. 1-5.

[16] S. Bahrainian and A. Dangel, “Sentiment Analysis using Sentiment Features”, in *Int. joint Conf. of Web Intelligence and Intelligent Agent Technologies*, 2013, pp. 26-29.

[17] G. Gautam and D. Yadav, “Sentiment analysis of twitter data using machine learning approaches and semantic analysis”, in *7th Int. Conf. on Contemporary Computing*, 2014, pp. 437-442.

[18] B. Gokulakrishnan, P. Plavnathan, R. Thiruchittampalam, A. Perera and N. Prasath, “Opinion Mining and Sentiment Analysis on aTwitter Data Stream”, in *Int. Conf. on Advances in ICT for Engineering Regions*, 2012, pp. 182-188.

[19] A. Celikyilmaz, D. Hakkani-Tur and Junlan Feng, “Probabilistic model-based sentiment analysis of twitter messages”, *IEEE Spoken Language Technology Workshop (SLT),* 2010, pp. 79-84.

[20] V. Sehgal and C. Song, “SOPS: Stock Prediction Using Web Sentiment”, in *7th IEEE Int. Conf. on Data Mining Workshop*, 2007, pp. 21-26.

[21] Z. Kechaou, B. M. Ammar and A. M. Alimi, “Improving e-learning with sentiment analysis of users' opinions”, in *Global Engineering Education Conference (EDUCON),* 2011, pp. 1032-1038.

[22] A.C.E.S Lima. and L.N. de Castro, “Automatic sentiment analysis of Twitter messages”, in *4th Int. Conf. on Computational Aspects of Social Networks (CASoN)*, 2012, pp. 52-57.

[23] R. Batool, A. M. Khattak, J. Maqbool and S. Lee, “Precise tweet classification and sentiment analysis”, in *12th Int. Conf. on Computer and Information Science (ICIS)*, 2013, pp. 461-466.

[24] N. Altrabsheh, M. Cocea and S. Fallahkhair, “Sentiment analysis: towards a tool for analysing real-time students feedback”, in *26th International Conference on Tools with Artificial Intelligence,* 2014, pp. 420-423.

[25] Z. WANG, V. J. Chuan TONG, X. XIN and H. C. CHIN, “Anomaly Detection through Enhanced Sentiment Analysis on Social Media Data”, in *6th International Conference on Cloud Computing Technology and Science*, 2014, pp. 918-922.

[26] V. Singh and S. K. Dubey, “Opinion mining and analysis: A literature review” , in *5th Int. Conf. on Confluence The Next Generation Information Technology Summit (Confluence),* 2014, pp. 232-239.

[27] K. Khan, B. Baharudin, A. Khan and F. Malik, “Mining Opinion from Text Documents: A Survey”, *Digital Ecosystems and Technologies*, 2009, pp. 217-222.

[28] K. Ghag and K. Shah, “Comparative analysis of the techniques for Sentiment Analysis”, in *Int. Conf. on Advances in Technology and Engineering*, 2013, pp. 1-7.

[29] W. Medhat, A. Hassan and H. Korashy, "Sentiment analysis algorithms and applications: A survey", *Ain Shams Engineering Journal,* vol. 5, no. 4, 2014, pp. 1093-1113.

[30] J. Khairnar and M. Kinikar, “Machine Learning Algorithms for Opinion Mining and Sentiment Classification”, in *International Journal of Scientific and Research Publications*, vol. 3, no. 6, June 2013.

[31] A. Sarlan, C. Nadam and S. Basri, “Twitter Sentiment Analysis”, in *Int. Conf. on Information Technology and Multimedia*, 2014, pp. 213-216.

[32] P. Saloun, M. Hruzik and I. Zelinka, “Sentiment Analysis – e-Bussines and e-Learning Common Issue”, in *11th IEEE Int. Conf. on Emerging eLearning Technologies and Applications*, 2013, pp. 339-34.