



FEATURE EXTRACTION ENCHANCEMENT IN USERS' ATTITUDE DETECTION

S. S. Ibrahiem
Department Computer
Science, Faculty of
Computer and
Information
Sciences, Ain Shams
University,
Cairo, Egypt.
soha.elshafey@cis.asu.e
du.eg

K. A. Bahnasy
Department Information
System, Faculty of
Computer and
Information
Sciences, Ain Shams
University,
Cairo, Egypt.
khaled.bahnasy@cis.asu
.edu.eg

M. M. Morsey
Department Computer
Science, Faculty of
Computer and
Information
Sciences, Ain Shams
University,
Cairo, Egypt.
mohamed.mabrouk@liv
e.com

M. M. Aref
Department Computer
Science, Faculty of
Computer and
Information
Sciences, Ain Shams
University,
Cairo, Egypt.
mostafa.aref@cis.asu.ed
u.eg

Abstract: *The social network are the trendiest applications which are developed for sharing opinions about different topics or events e.g. Twitter. As a result, this kind of applications becomes abundant data source for NLP researchers to innovate and enhance techniques that can track users' attitudes towards target event, topic or even another person. These users' attitudes are playing a pivotal role for decision makers, so they can take an appropriate action towards users' negative or positive reactions either. This paper focuses on users' attitude detection based on new feature set and applies on different machine learning models that can monitor and enhance users' attitude identification system. Annotated emotion tweets dataset and word emotion lexicon are used in training, building, and testing classification models.*

Keywords: *Emotion classification, Natural Language processing, Sentiment analysis.*

1. Introduction & Problem Definition

Massive data elevation obstructs decision makers to now users' attitudes towards their products or events. This resulted different automatic and semi-automatic techniques that can preprocess and normalize these vast data of users' attitudes. Information extraction is a quite challenging process in some definite sequential processes to finally detect user attitude in any given text [1-3], especially tweets due to slang language and emoticons included. In this paper, a feature set extracted is discussed that can enhance users' attitude detection system. It applied on different machine learning models and applied into WASSA-2017 Shared Task on Emotion Intensity data [4,5] in addition to the NRC word emotion lexicon [6-8]. Related work discussed in section 2, Our emotion detection system design is proposed in section 3, case study is presented in section 4. Finally, conclusion in section 5.

2. Related Work

Recently, affective text emotion analysis is a subjective conscious experience characterized by psychophysiological expressions, biological reactions and mental states. Emotions are what drive audience to action. People that are unable to feel emotions, they find it extremely difficult to make decisions. Emotion can be expressed in many ways that can be seen such as facial expression and gestures, speech and by written text. Evaluating sentiments in terms of emotions such as joy, surprise, anger, or fear, to name a few instead of simply using the conventional positive and negative sentiment evaluation. This is realized through exploiting the powerful capability of the semantic web technology to provide an expressive knowledge base of a cognitive model of emotions organization, and the availability of lexical databases to measure the semantic similarity between opinion-words and emotions. Emotion Detection in text documents is essentially a content - based classification problem involving concepts from the domains of Natural Language Processing as well as Machine Learning as in the following researches [9-13]. In computational linguistics, the automatic detection of emotions in texts is becoming increasingly important from an applicative point of view. There is some ambiguity about the difference between opinion, sentiment and emotion. Opinion defined as a transitional concept that reflects attitude towards an entity. The sentiment reflects feeling while emotion reflects attitude [14,15].

Psychologists have argued that some emotions are more basic than others. However, they disagree on which emotions and how many should be classified as basic emotions. Ekman, Izard, and Pultchick lists which are shown in Table 1 [16] are the most common lists of emotions used in emotion detection methods. Emotions detection (ED) can be considered as a sentiment analysis task. ED can be implemented using machine learning or lexicon-based approaches.

Table 1 Common lists of emotions used in emotion detection methods.

Lists of Basic Emotions	
Ekman	anger, disgust, fear, joy, sadness, and surprise
Izard	anger, Contempt, disgust, distress, fear, guilt, interest, joy, shame, and surprise
Pultchick	anger, anticipation, disgust, fear, joy, sadness, surprise, and trust

Feature extraction and selection is a basic process for text mining and information retrieval. It directly leverages the accuracy of text classification system. In emotion classification, text preprocessing is an indispensable task in mining large amount of data. Pre-processing is the first step in text classification and choosing the right pre-processing techniques can improve classification effectiveness. Major challenges are to tame the data in terms of noise, relevance, emoticons, folksonomies and slangs. From the previous researches, challenges can be met in preprocessing are text normalization and feature extraction stages [17-23].

3. Emotion Detection Workflow

This research proposes an approach to the system design of intelligent sentiment mining systems that can handle semantic knowledge, learn new affective knowledge and detect perceive and feel these four emotions from a written text which are anger, sad, fear, and joy. It develops a probabilistic model that recommends user attitude from the data extracted from real online social media application (e.g. twitter) and measure its results with annotated emotion tweets and word emotion lexicon.

Twitter is a social networking service that allows broadcasting of short messages called tweets. These tweets from millions of active users all over the world are considered as an information treasure, that not only attracts academics attention to know what users' interests are but also organizations as well [18-21]. Figure 1 shows the proposed users' attitude detection workflow that describes phases that detects emotion of a given users' tweet in specific topic.

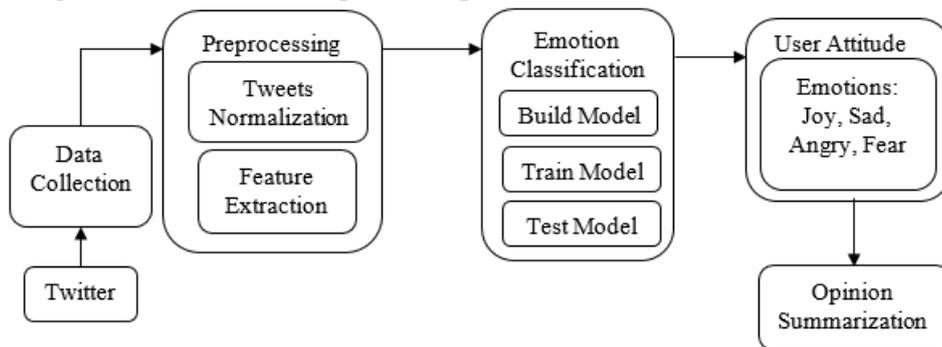


Figure. 1: Emotion detection workflow

3.1 Tweets Normalization

This section discusses different types of processes to denoise and filter important words or synonyms from tweets that includes emotion or sentiment, that makes classifier result is better and more accurate, such as [17-21]:

- Username. There are usernames like “e.g.: @hany”, that starts with symbol “@”. The usernames indicate who is the information pointing to the target [17].
- Hashtags. Hashtags, marked by symbol “#”, mean that the tweets are associated with the particular topics [17].
- Emotions: Users express sentiment with emotions, e.g. “:)” means happy or other positive affections. These emotions, in order to be interpreted correctly, have to be replaced to impute their meaning [17].
- Numbers. It is a common tactic to remove numbers from text, because they do not contain any sentiment [17].
- Handling Capitalized Words. Capitalized words may imply intense emotion [17].
- Lowercasing. One of the most common pre-processing techniques is to lowercase all words. By doing so, many words are merged and the dimensionality of the problem is reduced [17].
- Replace emoticons. Social media users usually write emoticons in their texts, in order to be interpreted correctly, have to be replaced to impute their meaning [17].
- Replace Elongated Words. Elongated is a word when it contains a character that is repeating more than two times, like the word ‘greeeeat’. It is important to replace words like this with their source

words, so they can be merged. Otherwise, the classifier will treat them as different words, and probably the elongated ones will be ignored because of their low frequency of occurrence [17].

- **Replace Contractions and Negation handling.** One technique that can be used in pre-process is the replacement of contractions, i.e. words like ‘won’t’ and ‘don’t’, that will be replaced with ‘will not’ and ‘do not’, respectively [17].
- **Remove stop words.** stop words are function words with high frequency of presence across all sentences. It is considered needless to analyze them, because they do not contain much useful information [17].
- **Stemming.** It is the process of removing the endings of the words in order to detect their root form. By doing so, many words are merged and the dimensionality is reduced. It is a widely used method that generally provides good results [17].
- **Remove URLs and User Mentions.** In Twitter texts, almost every sentence contains a URL and a user mention. Their presence does not contain any sentiment and one approach is to remove them in pre-processing [17].
- **Spelling Correction.** It is very common in informal texts for users to make spelling errors that might make classification harder. By using tools that automatically correct these errors, it is possible to improve classification effectiveness. While no corrector is perfect, they have some —usually high— accuracy of success [17].
- **Remove Punctuation.** In many works, it is common to remove punctuation signs in pre-processing. However, many times the presence of punctuation marks denotes the existence of some sentiment. For example, an exclamation mark may mean an intense positive or negative sentiment. So, if we remove them, we might decrease the accuracy of classification [17].

Figure 2 illustrates the preprocessing data flow diagram applied on the annotated emotion tweets data sets.

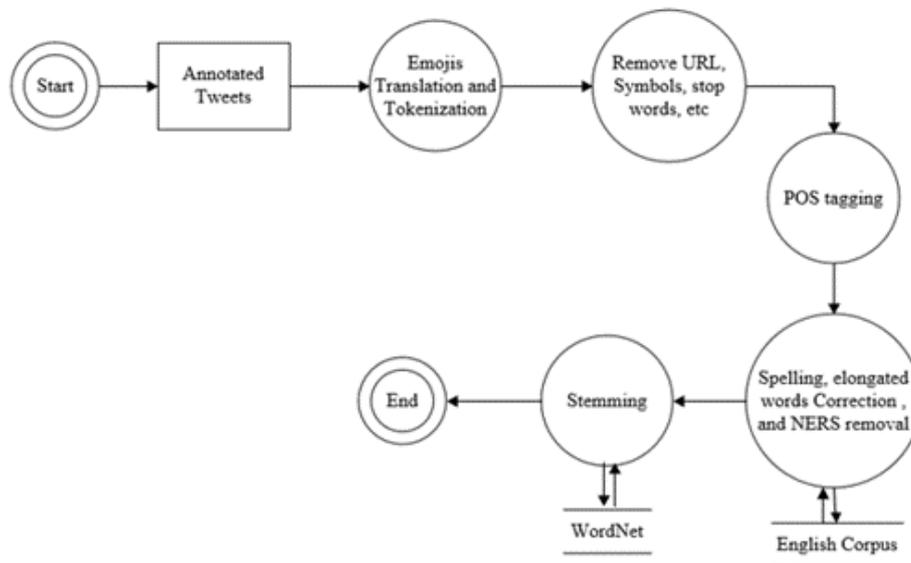


Figure 2: Tweets Preprocessing Data Flow Diagram (DFD)

3.2 Feature Extraction

The preprocessed tweets have many distinctive properties. In the feature extraction process, different aspects have been extracted. Later these aspects are used to compute the emotion of user's tweet either it implies joy, sad, angry or fear emotion. These aspects are trained on different machine learning model. Some features that have been reported in literature are [21, 24-28]:

1. Words and Their Frequencies: Unigrams, bigrams and n-gram models with their frequency counts are considered as features. There has been more research on using word presence rather than frequencies to better describe this feature [29, 30].
2. Parts of Speech Tags: Parts of speech like adjectives, adverbs and some groups of verbs and nouns are good indicators of subjectivity and sentiment. A syntactic dependency patterns can be generated by parsing or dependency trees [21].
3. Opinion Words and Phrases: Apart from specific words, some phrases and idioms which convey sentiments can be used as features [21].
4. Position of Terms: The position of a term with in a text can effect on how much the term makes difference in overall sentiment of the text [21].
5. Negation is an important but difficult feature to interpret. The presence of a negation usually changes the polarity of the opinion [21].
6. Syntax: Syntactic patterns like collocations are used as features to learn subjectivity patterns by many of the researchers [21].

3.3 Feature Set

The proposed feature set used to classify and detect users' attitude in input tweets are:

Words and Their Frequencies: Unigrams models with their frequency counts are considered as features [21,29].

Words' implied attitude from Emotion lexicon: Apart from specific words which convey sentiments can be used as features. The NRC Emotion Lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The annotations were manually done by crowdsourcing [7]. In the proposed system design, this lexicon will be used to detect the sentiment and polarity of the featured keywords detected of the input tweet from the preprocessed phase for four basic emotions (sad, anger, fear, & joy).

Word to vector similarity to emotions: it groups the vectors of similar words together in vector space. It detects similarities mathematically. Word2vec creates vectors that are distributed numerical representations of word features, features such as the context of individual words. Many researchers used Word2Vec as the features for text classification tasks such as emotion classification. Its ability to model high quality distributional semantics among words has contributed to its success in many of the tasks. It turns text into a numerical form that machine learning models can understand. The purpose and usefulness of Word2vec is to group the vectors of similar words together in vector space. That is, it detects similarities mathematically. Word2vec creates vectors that are distributed numerical representations of word features, features such as the context of individual words. It does so without human intervention. Given enough data, usage and contexts, Word2vec can make highly accurate guesses about a word's meaning based on past appearances. Those guesses can be used to establish a

word’s association with other words (e.g. “man” is to “boy” what “woman” is to “girl”). Word2vec is similar to an autoencoder, encoding each word in a vector [30-34].

3.4 Learning Model

Annotated datasets are preprocessed and normalized. A tweet vector is proposed to represent each tweet from the mentioned feature set and trained it on Naïve Bayes, simple feedforward neural network and deep neural network models in order to get a trained users’ attitude model. Table 2 and Table 3 represent evaluation on Naïve Bayes model on training and testing datasets respectively.

Table 2: Naïve Bayes output model evaluation on training dataset

Class	Precision	Recall	F-Measure
Anger	0.927	0.855	0.890
Fear	0.883	0.884	0.884
Joy	0.809	0.899	0.852
Sad	0.826	0.798	0.812
Correctly Classified Instances	3114, 86.1888 %		
Incorrectly Classified Instance	499, 13.8112 %		
Total Number of Instances	3613		

Table 3: Naïve Bayes output model evaluation on testing dataset

Class	Precision	Recall	F-Measure
Anger	0.595	0.603	0.599
Fear	0.644	0.537	0.586
Joy	0.681	0.728	0.704
Sad	0.526	0.609	0.565
Correctly Classified Instances	1922, 61.1712 %		
Incorrectly Classified Instance	1220, 38.8288 %		
Total Number of Instances	3142		

4. Case Study

In this section, a case study is applied on the three-machine learning model to detect users’ attitude. The input tweet from twitter: "We have lost one of the greatest artists of all times. Thank u Shadia for the beautiful art you have given us ". User attitude detected from this sentence is sadness.

Preprocessing Phase: In this phase, the input tweet is taken and preprocessed for further processes in opinion mining. It includes tokenization, POS tagging, name entity removal, stop word removal, case normalization, and stemming. POS tagging is an essential process to conclude effective words that

implies user's attitude (see figure 3), Table 4 expresses important tokens that deduced by the system that classification process works on it later on.

We/PRP have/VBP lost/VBN one/CD of/IN the/DT greatest/JJS artists/NNS of/IN all/DT times/NNS./
Thank/VB u/NNP Shadia/NNP for/IN the/DT beautiful/JJ art/NN you/PRP have/VBP given/VBN us/PRP

Figure 3: Stanford online POS Tagging output for the sample tweet

Table 4: Case study preprocessed keywords POS

Word	POS
Lose	Verb
Great	Adjective
Artist	Noun
Time	Noun
Thank	Noun
Beautiful	Adjective
Art	Noun
Give	Verb
Purple	Adjective
Heart	Noun

Emotion Classification Phase: A set of annotated tweets [1, 5] T that are predefined by emotions categories $E = \{\text{anger, fear, sad, joy}\}$ are used, each tweet $t \in T$ expressed by $e \in E$. Emotion classification techniques varies, there are machine learning supervised, unsupervised, lexicon based and hybrid approaches. In this case study, the NRC emotion lexicon [错误!未找到引用源。-7] and the annotated tweets dataset [1, 5] are trained and combined together to deduce user's tweet attitude. Table 5 expresses the input tweet preprocessed words' occurrence probability in each emotion labelled tweets' [1, 5]. The annotated tweets dataset is searched to calculate the total occurrence of each preprocessed word in each emotion tweets' category and its probability or average to all preprocessed words in each emotion tweets'.

Table 5: Word occurrences & probabilities in tweets datasets

Word	Anger	Fear	Sad	Joy
Lose	10	11	61	9
Great	2	22	7	24
Artist	0	4	2	0
Time	23	34	42	29
Thank	2	6	3	2
Beautiful	0	1	1	10
Art	1	0	4	0
Give	20	22	12	25
Purple	0	0	2	1
Heart	13	32	18	42

Total	71	132	152	142
Average	0.0113654	0.01564166	0.02516139	0.02349048

Word to vector is applied on each emotion labelled tweets', it produces words embedding vector for all vocabularies mentioned on each emotion category. Word2vec was used to measure the similarity of each emotion and input preprocessed tweet words' and showed in Table 6. The average word to vector of preprocessed tokens of the input tweet is calculated to each emotion label and is shown below. Table 7 shows preprocessed words occurrences on NRC emotion lexicon [7] and their total average to each emotion label. Table 8 shows the results of the proposed feature set to three different machine learning model, Na ĩve Bayes, single feedforward neural network, and deep neural network.

Table 6: Word to vector similarity between each word and emotion label of trained tweets

Word	Anger	Fear	Sad	Joy
Lose	0.9502305	0.9928561	0.96704245	0.9548956
Great	0.5485947	0.9980906	0.88565034	0.9783221
Artist	0	0.98108965	0.6075043	0
Time	0.98625034	0.9981958	0.9681902	0.9754449
Thank	0.6957974	0.9839922	0.73311645	0.8331418
Beautiful	0	0.8595918	0.38121778	0.94535273
Art	0.5443038	0	0.83291364	0
Give	0.9693511	0.9974107	0.942467	0.9746235
Purple	0	0	0.7105429	0.2577447
Heart	0.9776024	0.995376	0.96545243	0.97279197
Average	0.567213024	0.780660285	0.799409749	0.68923173

Table 7: Word emotion presence in NRC Lexicon

Word	Angry	Fear	Sad	Joy
Lose	1	1	1	0
Great	0	0	0	0
Artist	0	0	0	0
Time	0	0	0	0
Thank	0	0	0	0
Beautiful	0	0	0	1
Art	0	0	1	1
Give	0	0	0	0
Purple	0	0	0	0
Heart	0	0	0	0
Total	1	1	2	1
Average	0.1	0.1	0.2	0.2

Table 8: Machine learning models output of the input tweet

Models	Angry	Fear	Sad	Joy
Naïve Bayes Model	0	0	1	0
Single NN	0	0	1	0
Deep NN	0	0	1	0

Finally, the three different machine learning models tests the proposed feature set: word to vector similarity, frequency distribution on the emotion-labelled tweets dataset, and the emotion word emotion lexicon. The deduced user's attitude from the input tweet is sad feeling.

5. Conclusion

In this paper, a feature set was proposed and applied on annotated tweets to build three different machine learning models that detect users' attitude from the input tweets. A case study was presented that an input tweet is applied, and the feature set was extracted, followed by testing it on the three machine learning models and the user's attitude was detected correctly. In the future work, a more complicated and different implied emotions' tweets will be applied to detect more than one implied emotion in single tweet, and enhancing the accuracy of classification models.

References

1. Bing Liu, Lei Zhang, "Sentiment Analysis and Opinion Mining", Morgan & Claypool Publishers, Chapters [1,2], pages 7-29, May 2012.
2. Dudhat Ankitkumar, R. R. Badre, Mayura Kinikar, " A Survey on Sentiment Analysis and Opinion Mining", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 11, November 2014.
3. Federico Alberto Pozzi, Elisabetta Fersini, Enza Messina, Bing Liu, "Sentiment Analysis in Social Networks", 1st Edition, Published 15th September 2016, Chapters [1,2], ISBN: 9780128044384, Morgan Kaufmann.
4. Saif M. Mohammad and Felipe Bravo-Marquez. In Proceedings of the EMNLP 2017 Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media (WASSA), pages 34–49, September 2017, Copenhagen, Denmark.
5. Saif M. Mohammad and Felipe Bravo-Marquez. "Emotion Intensities in Tweets", In Proceedings of the Sixth Joint Conference on Lexical and Computational Semantics (*Sem), August 2017, pages 65–77, Vancouver, Canada.
6. Saif Mohammad and Peter Turney, "Crowdsourcing a Word-Emotion Association Lexicon", Computational Intelligence, 29 (3), p 436-465, 2013.
7. Saif Mohammad and Peter Turney, "Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon", Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, pages 26–34, Los Angeles, California, June 2010.

8. Soujanya Poria, Erik Cambria, Grégoire Winterstein, Guang-Bin Huang, “Sentic patterns: Dependency-based rules for concept-level sentiment”, *Knowledge-Based Systems* 69(2014) p45-63.
9. Soujanya Poria, Erik Cambria, Rajiv Bajpaib, Amir Hussaina. “A review of affective computing: From unimodal analysis to multimodal fusion”. *Information Fusion*, Volume 37, Pages 98–125, September 2017.
10. Asha .K ,Devi.T, ”Semantic Affect Sensing in User Generated Contents: An intelligent machine informatics to Audit Human Sentiments “, *Journal of emerging Trends in Computing and Information Sciences*, Vol 3, pages 605-611, No 4,2012.
11. Weichselbraun, Albert and Gindl, Stefan and Fischer, Fabian and Vakulenko, Svitlana and Scharl, Arno, “Aspect-Based Extraction and Analysis of Affective Knowledge from Social Media Streams”. *IEEE Intelligent Systems*, 2016.
12. Mohammad, Saif M. "Sentiment analysis: Detecting valence, emotions, and other affectual states from text." *Emotion Measurement* (2015): 201-238.
13. Erik Cambria, “Affective Computing & Sentiment Analysis”, *IEEE Intelligent Systems* 31(2), pp. 102-107 (2016).
14. Mikalai Tsytsarau, Themis Palpanas,” Survey on mining subjective data on the web *Data Mining Knowledge Discovery*, 24 (2012), pp. 478-514
15. Walaa Medhat, Ahmed Hassan, Hoda Korashy, " Sentiment analysis algorithms and applications: A survey", *Ain Shams Engineering Journal*, Volume 5, Issue 4, December 2014, Pages 1093-1113, May 2014.
16. Mohamed Haggag, Samar Fathy, Nahla Elhaggar, “Ontology-Based Textual Emotion Detection”, (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, Vol. 6, No. 9, pages 239- 246, 2015.
17. Dimitrios Effrosynidis, Symeon Symeonidis, and Avi Arampatzis, “A Comparison of Pre-processing Techniques for Twitter Sentiment Analysis”, *21st International Conference on Theory and Practice of Digital Libraries (TPDL) 2017*, *Proceedings, Lecture Notes in Computer Science (LNCS)* 10450, pp. 394–406, 2017.
18. Tajinder Singh and Madhu Kumari, “Role of Text Pre-Processing in Twitter Sentiment Analysis”, *Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)*, *Procedia Computer Science* 89 (2016) 549 – 554.
19. Yanwei Bao, Changqin Quan, Lijuan Wang, and Fuji Ren, “The Role of Pre-processing in Twitter Sentiment Analysis”, *ICIC 2014*, *LNAI* 8589, pp. 615–624, 2014.
20. Mr.SanketPatil, Prof.VarshaWangikar, Prof. K. Jayamalini, “Data Preprocessing, Sentiment Analysis & NER On Twitter Data.”, *IOSR Journal of Computer Engineering (IOSR-JCE)*, e-ISSN: 2278-0661,p-ISSN: 2278-8727, PP 73-79.
21. Vishal A. Kharde, S.S. Sonawane, “Sentiment Analysis of Twitter Data: A Survey of Techniques”, *International Journal of Computer Applications* (0975 – 8887), Volume 139 – No.11, Pages 5-15, April 2016.
22. Hong Liang, XiaoSun, YunlieSun, and Yuan Gao, “Text feature extraction based on deep learning: a review”, *EURASIP Journal on Wireless Communications and Networking*, Volume 2017 No(1), page 211, December 2017.
23. Muhammad Zubair Asghar, Aurangzeb Khan, Shakeel Ahmad, Fazal Masud Kundi, “A Review of Feature Extraction in Sentiment Analysis”, *Journal of Basic and Applied Scientific Research*, J. Basic. Appl. Sci. Res., 4(3)181-186, 2014.

24. Mr.SanketPatil, Prof.VarshaWangikar, Prof. K. Jayamalini, "Data Preprocessing, Sentiment Analysis & NER On Twitter Data.", IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661,p-ISSN: 2278-8727, PP 73-79.
25. Hong Liang, XiaoSun, YunlieSun, and Yuan Gao, "Text feature extraction based on deep learning: a review", EURASIP Journal on Wireless Communications and Networking, 2017.
26. Soo-Min Kim, Eduard Hovy, "Determining the Sentiment of Opinions", 20th Proceedings of the COLING conference, Geneva, Volume 1200, 2004.
27. Janyce Wiebe, Ellen Riloff, "Creating Subjective and Objective Sentence Classifiers from Unannotated Texts", Proceedings of International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'05), pages 486-497, Mexico City, Mexico, 2005.
28. Bing Liu, "Sentiment Analysis and Subjectivity", Handbook of Natural Language Processing, Second Edition. Taylor and Francis Group, Boca, 2010.
29. Pang, B.and Lee, L. "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts". 42nd Meeting of the Association for Computational Linguistics[C] (ACL-04). 2004, 271-278.
30. Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng and Christopher Potts "Recursive deep models for semantic compositionality over a sentiment Treebank." Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2013.
31. Maynard, D.G. and Bontcheva, K., "Challenges of Evaluating Sentiment Analysis Tools on Social Media". In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). 10th edition of the Language Resources and Evaluation Conference, 23-28 May 2016, Portorož. LREC , pp. 1142-1148.
32. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," Advances in Neural Information Processing Systems 26 (NIPS 2013), pages 3111-3119,2013.
33. T. Mikolov, G. Corrado, K. Chen, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Proceedings of the International Conference on Learning Representations (ICLR 2013), pp. 1–12, 2013.
34. Eissa M.Alshari, Azreen Azman, Shyamala Doraisamy,Norwati Mustapha and Mustafa Alkeshr, " Improvement of Sentiment Analysis based on Clustering of Word2Vec Features", 28th International Workshop on Database and Expert Systems Applications, pages 123-126, Lyon, France, August 2017.