

Review of classification of Twitter data based upon Improved KNN

Narinder Kaur, Dr. Rakesh Kumar Principal, Prabhjeet Kaur Assistant Professor

Computer Science Engg. ,PTU Jalandhar

narinderkaur01992@gmail.com

Prabhgill59@gmail.com

Abstract— Sentiment Analysis (SA) is characterized as an intelligent strategy of removing different feelings and feeling of clients. it's one among the key fields for specialists working in dialect process. The development of net has turned out to be one of the biggest stage for clients to trade their ideas, share messages, post sees and so on. There conjointly exists a few online journals, Google+ that is increasing sensible quality as they enable people to particular their perspectives. amid this paper, the present condition of differed systems of sentiment analysis for feeling mining like machine learning and vocabulary based methodologies square measure specified. the different strategies utilized for Sentiment Analysis is broke down amid this paper to play out an analysis study and check the value of the present writing.

Keywords—

I. DATA MINING

Sentiment is a view, feeling, opinion or assessment of a person for some product, event or service [1, 2, 3].Sentiment Analysis or Opinion Mining is a challenging Text Mining and Natural Language Processing problem for automatic extraction, classification and summarization of sentiments and emotions expressed in online text [1,2].Sentiment analysis is replacing traditional and web based surveys conducted by companies for finding public opinion about entities like products and services [1]. Sentiment Analysis also assists individuals and organizations interested in knowing what other people comment about a particular product, service topic, issue and event to find an optimal

choice for which they are looking for. Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. It represents a large problem space. Sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. it has a wide arrange of applications, almost in every domain. It offers many challenging research problems, which had never been studied before. In this the opinions or sentiments are labelled as positive , negative and neutral . It is a multidisciplinary task, which exploits techniques from computational linguistics, machine learning, and natural language processing, to perform various detection tasks at different text-granularity levels.

II. SEMANTIC ANALYSIS WITH SOCIAL WEBSITES

Microblogging today has become a very popular communication tool among Internet users. Millions of messages are appearing daily in popular websites that provide services for microblogging such as Twitter, Tumblr, Facebook. In the past few years, there has been a huge growth in the use of microblogging platforms such as Twitter. Companies and media organizations are increasingly seeking ways to mine Twitter for information about what people think and feel about their products and services. Twitter contains a very large number of very short messages created by the users of this microblogging platform.

Each tweet is 140 characters in length .Tweets are frequently used to express a tweeter's emotion on a particular subject. There are firms which poll twitter for analyzing sentiment on a particular topic. The challenge is to gather all such relevant data, detect and summarize the overall sentiment on a topic. Twitter has been selected with the following purposes in mind. Twitter is an Open access social network. Twitter is an Ocean of sentiments .Twitter provides user friendly API making it easier to mine sentiments in realtime.

Twitter serves as a corpus for opinion mining due to following reasons.

- Collected corpus from twitter can be arbitrarily large since it contains an enormous number of text posts.Also it grows everyday.
- It is possible to collect text posts of users from different social and interest groups. We can collect data in different languages.

III. LEVELS OF SEMANTIC ANALYSIS

Document Level Sentiment Analysis: The basic information unit is a single document of opinionated text. In this document level classification, a single review about a single topic is considered. But in the case of forums or blogs, comparative sentences may appear. Customers may compare one product with another that has similar characteristics and hence document level analysis not desirable in forums and blogs. The challenge in the document level classification is that the entire sentence in a document may not be relevant in expressing opinion about an entity. Therefore subjectivity/objectivity classification is very important in this type of classification. The irrelevant sentences must be eliminated from the processing works. Document level sentiment classification executed on the overall sentiments expressed by authors Documents classified according to the sentiments instead of topic. It is to summarize the whole document as positive or negative polarity about any object (mobile, car, movie, and politician).

Sentence level sentiment analysis: In the sentence level sentiment analysis, the polarity of each

sentence is calculated. The same document level classification methods can be applied to the sentence level classification problem. Objective and subjective sentences must be found out. The subjective sentences contain opinion words which help in determining the sentiment about the entity. After which the polarity classification is done into positive and negative classes. In case of simple sentences, a single sentence bears a single opinion about an entity. But there will be complex sentences also in the opinionated text. In such cases, sentence level sentiment classification is not desirable. Knowing that a sentence is positive or negative is of lesser use than knowing the polarity of a particular feature of a product. The advantage of sentence level analysis lies in the subjectivity/objectivity classification. The traditional algorithms can be used for the training processes.

Phrase level sentiment analysis: The phrase level sentiment classification is a much more pinpointed approach to opinion mining. The phrases that contain opinion words are found out and a phrase level classification is done. This can be advantageous or disadvantageous. In some cases, the exact opinion about an entity can be correctly extracted. But in some other cases, where contextual polarity also matters, the result may not be fully accurate. Negation of words can occur locally. In such cases, this level of sentiment analysis suffices. But if there are sentences with negating words which are far apart from the opinion words, phrase level analysis is not desirable. Also long range dependencies are not considered here. The words that appear very near to each other are considered to be in a phrase.

IV. RELATED STUDY

Youngsub Han and Kwangmi Ko Kim [1] presented a method to achieve accuracy in sentiment analysis on the basis of social media data. Social media became popular than ever as people are willing to share their emotions and opinions or to participate in social networking. Accordingly, the understanding of social media

usage became important. The sentiment analysis is emerged as one of useful methods to analyze emotional stats expressed in textual data including social media data.

Tanvi et. al. [2] proposed a framework for sentiment analysis using dictionary-based approach and brings out a comparative study on sentiment analysis techniques including machine learning technique and lexicon based technique. The comparisons are majorly drawn based on features such as preprocessing, technique employed, dictionary, datasets, and soft-computing approaches. An approach to sentiment analysis using dictionary-based approach incorporating fuzzy logic is proposed. The datasets used is an Amazon dataset taken from web. they apply the various pre-processing techniques like stemming , POS-Tagging and stop words removal. they have used fuzzy logic for negation handling.

Farhan et. al. [3] focuses on various primary issues like accuracy, data sparsity and sarcasm problems and presents an algorithm for twitter feeds classification based on a hybrid approach. Then proposed method includes various pre-processing steps before feeding the text to the classifier. Experimental results show that the proposed technique overcomes the previous limitations and achieves higher accuracy when compared to similar techniques. This involves pre-processing steps and a hybrid scheme of classification algorithms. Pre-processing steps include removal of URLs, hash-tags, username and special characters, spelling correction using a dictionary substitution of abbreviations and slangs with expansions, lemmatization and stop words removal. The proposed classification algorithm incorporates a hybrid scheme using an enhanced form of emoticon analysis , SentiWordNet analysis and an improved polarity classifier using list of positive/negative words.

G. et. al. [4] examined how classifiers work while doing opinion mining over Twitter data. Reducing the data size using the feature selection method produces better accuracy and increase the

computational space. The feature selection method plays a vital role in increasing the accuracy of sentiment analysis. The selected features for the research work are unigrams, negation words, emoticons, stemming and retweet count. The retweet count plays a major role in sharing others' opinion. The ranking method is used to select the top most and relevant features. The best-ranking method for the text mining is Zipfs' law and is used to rank the selected features. The proposed Senti_Classi approach is experimented with Naïve Bayes, Support Vector Machines and Maximum Entropy. The 10 cross-fold validation method is used for training and testing the classifiers. This paper presents the best machine learning approach to sentiment analysis on tweets.

Liza Wikarsa et. al. [5] developed a text mining application to detect emotion in twitter in which emotions are classified into 6 classes namely happy, sad, anger, disgust, fear and surprise. Three processes that were implemented are pre-processing in which noise will be removed, processing in which weighting and classification based upon naïve bayes classification will be implemented and in the last validation phase in which results are to be generated.

Namita et. al. [6] proposed a three stage hierarchical model for sentiment extraction, first labeling with emoticons is done, then tweets are labeled using pre-defined lists of words with strong positive or negative sentiments and finally tokens are weighted based on subjectivity lexicon and proposed probability based method. Further, various cascading and hybrid methods are proposed based on subjectivity lexicon and Probability based method. In addition to this, effect of discourse relations is also investigated at the pre-processing step. Experimental results show the effectiveness of the proposed hybrid approach for sentiment classification of tweets.

Pedro et. al. [7] adopted a hybrid classification process that uses three classification approaches: rule-based, lexicon-based and machine learning approaches. They suggest a pipeline architecture that extracts the best characteristics from each

classifier. Our system achieved an Fscore of 56.31% in the Twitter message-level subtask. They used POS-Tagger and SVM machine learning algorithm. They performed normalization which is responsible to correct and normalize the text before the classifiers use it. This architecture improves the classification process because it takes advantage of the multiple approaches.

Amit et. al. [8] introduced a novel approach for automatically classifying the sentiment of “tweets” into positive, negative and neutral sentiment. Experimental evaluations show that proposed techniques are efficient and perform better than previously proposed methods. They worked with English language however the proposed technique can be used with any other language. The Techniques used for feature selection is PMI and Chi Square. They used the three dictionaries for pre-processing the data and are Stop word Dictionary , Emoticon Dictionary , Acronym Dictionary. They presented a method that automatically collects tweets from twitter using Twitter API.

Shoushan et. al. [9] proposed a machine learning approach to incorporate polarity shifting information into a document-level sentiment classification system. First, a feature selection method is adopted to automatically generate the training data for a binary classifier on polarity shifting detection of sentences. Then, by using the obtained binary classifier, each document in the original polarity classification training data is split into two partitions, polarity-shifted and polarity-unshifted, which are used to train two base classifiers respectively for further classifier combination. The experimental results across four different domains demonstrate the effectiveness of our approach.

K. Revathy et. al. [10] presented more significant approach towards the contextual information in the document which is one of the drawbacks of the systems which are available for determining contextual information. The first model uses rule-based classification based on compositional semantic rules that identifies expression level

polarity. The second one performs sense-based classification based on WordNet senses as features to Support Vector Machine classifier. Further to provide a meaningful classification, semantics are incorporated as additional feature into the training data by the interpolation method. Thus, the third model performs entity-level analysis based on concepts obtained. The outputs of three models are handled by knowledge inference system to predict the polarity of sentence. This system is expected to produce better results when compared to the baseline system performance. The system aims to predict consumer moods and the attitude in real-time which can be efficiently utilized by the firms to increase productivity and revenue.

S. No	Author	Title	Contribution
1	Liza Wikarsa et. al.[21]	A Text Mining Application of Emotion Classification s of Twitter’s Users Using Naïve Bayes Method	developed a text mining application to detect emotion in twitter in which emotions are classified into 6 classes namely happy, sad, anger, disgust, fear and surprise. Three processes that were implemented are pre-processing in which noise will be removed, processing in which weighting and classification based upon naïve bayes classification will be implemented and in the last validation phase

			in which results are to be generated.			Accuracy of Sentiment Analysis	data. Reducing the data size using the feature selection method produces better accuracy and increase the computational space. The feature selection method plays a vital role in increasing the accuracy of sentiment analysis.
2	Tanvi et. al. [22]	An Approach to Sentiment Analysis Using Lexicons With Comparative Analysis of Different Techniques	proposed operations on datasets that is taken from an Amazon dataset taken from web. They execute the another pre-processing techniques like stemming, POS-Tagging and stop words removal. They endeavour used fuzzy logic for negation handling.				
3	Farhan et. al. [23]	TOM: Twitter Opinion Mining Framework using Hybrid Classification Scheme, Decision Support Systems	focused on various primary issues like accuracy, data sparsity and sarcasm problems and presents an algorithm for twitter feeds classification based on a hybrid approach. Then proposed method includes various pre-processing steps before feeding the text to the classifier.	5	Namit a et. al. [25]	A Hybrid Approach for Twitter Sentiment Analysis	proposed a three stage hierarchical model for sentiment extraction, first labeling with emoticons is done, then tweets are labeled using pre-defined lists of words with strong positive or negative sentiments and finally tokens are weighted based on subjectivity lexicon and proposed probability based method.
4	G. et. al. [24]	Machine Learning Based Approach to Enhance the	examined how classifiers work while doing opinion mining over Twitter	6	Pedro et. al. [26]	NILC_USP : A hybrid system for sentiment analysis in Twitter	adopted a hybrid classification process that uses three classification

		Messages	approaches: rulebased, lexicon-based and machine learning approaches. They suggest a pipeline architecture that extracts the best characteristics from each classifier.				the systems which are available for determining contextual information.
7	Amit et. al. [27]	Feature Extraction for Sentiment Classification on Twitter Data	introduced a novel approach for automatically classifying the sentiment of "tweets" into positive, negative and neutral sentiment. Experimental evaluations show that proposed techniques are efficient and perform better than previously proposed methods.	10	Yu-Long Qiao et al. [30]	Improved K Nearest Neighbor Classification Algorithm	proposed a technique to reduce the complexity of K-NN classification by using approximation coefficient of a fully decomposed feature vector with Haar wavelet and the variance of the corresponding untransformed vector, to produce two efficient test conditions.
8	Shoushan et. al. [28]	Sentiment Classification and Polarity Shifting		V. OPINION MINING			
9	K. Revathy et. al. [29]	A Hybrid Approach for Supervised Twitter Sentiment Classification	presented more significant approach towards the contextual information in the document which is one of the drawbacks of	Opinion Mining (OM), a promising discipline, defined as combination of information retrieval and computational linguistic techniques deals with the opinions expressed in a document [12]. The field aims at solving the problems related to opinions about products, politics in newsgroup posts, review sites, etc [13]. There are different techniques for summarizing customer reviews like Data Mining, Information Retrieval, Text Classification and Text Summarization [13]. Opinion Mining or Sentiment Analysis is the field to extract the opinionated text datasets and summarize in understandable form for end user [15]. Opinion mining is to extract the positive, negative or neutral opinion summary from unstructured data. World Wide Web users asked			

the opinions of his family and friends to purchase the product. In the same way when organizations needed to take the decision about their products they had to conduct the surveys to the focused groups or they had to hire the external consultants [5, 13]. Web 2.0 [14], facilitate the customers to take decision to purchase the product by reviewing the posted comments. Customers can post reviews on web communities, blogs, discussion forums, twitters, product's web site these comments are called user generated contents [12]. Web2.0 is playing a vital role in data extracting source in opinion mining. It facilitates users to know about the product from other customer's reviews that have already used it instead of asking friends and families. Companies, instead of conducting surveys and hiring the external consultants to know about the consumers opinions, extract opinionated text from product web site [13, 15]. An automated opinion summarization model is needed to perform these tasks. It is the sub-discipline of web content mining, involves Natural Language Processing and opinion extraction task to find out the polarity of any product consumers feedback [1].

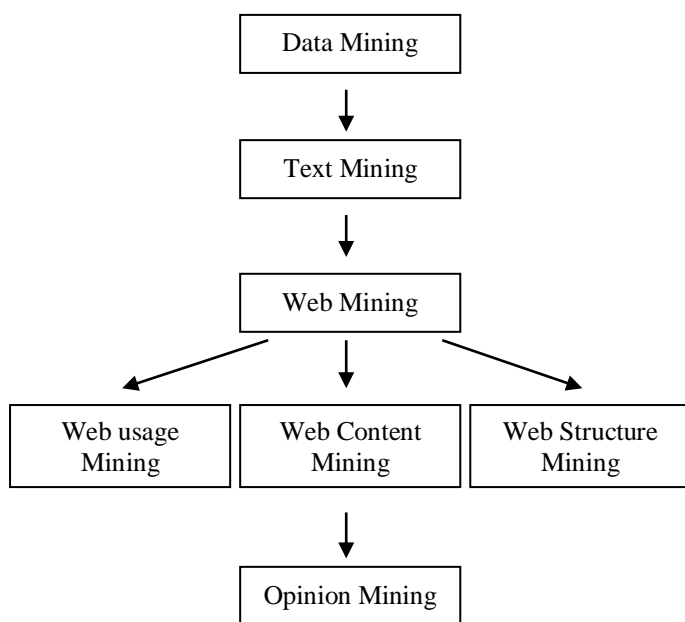


Fig 1: Opinion Mining Overview

VI. APPLICATIONS OF OPINION MINING

Opinion mining and sentiment analysis cover a wide range of applications.

1. Argument mapping software helps organising in a logical way these policy statements, by explicitating the logical links between them. Under the research field of Online Deliberation, tools like Compendium, Debatepedia, Cohere, Debategraph have been developed to give a logical structure to a number of policy statement, and to link arguments with the evidence to back it up.
2. Voting Advise Applications help voters understanding which political party (or other voters) have closer positions to theirs. For instance, SmartVote.ch asks the voter to declare its degree of agreement with a number of policy statements, then matches its position with the political parties.
3. Automated content analysis helps processing large amount of qualitative data. There are today on the market many tools that combine statistical algorithm with semantics and ontologies, as well as machine learning with human supervision. These solutions are able to identify relevant comments and assign positive or negative connotations to it (the so-called sentiment).

VII. TECHNIQUE USED IN OPINION MINING

Database contains the important hidden information used for decision making. Different databases like relational, object oriented, transactional and spatial databases consist on the complex dataset. Major data mining techniques used to extract the knowledge and information are: generalization, classification, clustering, association rule mining, data visualization, neural networks, fuzzy logic, Bayesian networks, genetic algorithm, decision tree, multi agent systems, CRISP-DM model, churn prediction, Case Based Reasoning and many more.

Rapidgrowth in databases has created the need to develop such technologies to extract the suggest of knowledge and information intelligently. Data mining techniques are most suitable for this purpose, these techniques directly refers Artificial Intelligence. Major data rule mining, data visualization, neural networks, fuzzy logic, Bayesian

networks, genetic algorithm, mining techniques used to extract the knowledge and information are: generalization, classification, clustering, association decision tree, multi agent systems, churn prediction, Case Based Reasoning, techniques, association and many more.

Supervised Machine Learning Classification is most frequently used and popular data mining technique. Classification used to predict the possible outcome from given data set on the basis of defined set of attributes and a given predictive attributes. The given dataset is called training dataset consist on independent variables (dataset related properties) and a dependent attribute (predicted attribute). A training dataset created model test on test corpora contains the same attributes but no predicted attribute. Accuracy of model checked that how accurate it is to make prediction. Classification is a supervised learning used to find the relationship among attributes.

Unsupervised Learning In contrast of supervised learning, unsupervised learning has no explicit targeted output associated with input. Class label for any instance is unknown so unsupervised learning is about to learn by observation instead of learn by example. Clustering is a technique used in unsupervised learning. The process of gathering objects of similar characteristics into a group is called clustering. Objects in one cluster are dissimilar to the objects in other clusters.

Case Based Reasoning Case based reasoning is an emerging Artificial Intelligence supervised technique used to find the solution of a new problem on the basis of past similar problems. CBR is a powerful tool of computer reasoning and solve the problems (cases) in such a way which is closest to real time scenario. It is a recent problem solving technique in which knowledge is personified as past cases in library and it does not depend on classical rules. The past problem's solutions are stored in CBR repository called Knowledge base or Case base. Instead of solving the new problem by "first principal" reasoning, CBR use the knowledge base to reuse the The solution of past similar problem if needed to the In case base repository as

a new solution instance in CBR cycle consists of four R's. Nowadays it is the most emerging technique used in opinion mining systems. Statistical methods are combined with knowledge extracting techniques in to enhance case searching, browsing and Reuse it for the problem solving methods semantic analysis of a sentence in natural language that can be easily used and manipulated in a textual data mining process. This sentence analysis uses and depends on several types of knowledge that are: a lexicon, a case base and hierarchy of index. In this methodology a case based reasoning model is adopted that is based on the classification rules and course of similarity for the assurance of the compliance.

VIII. ISSUES IN SENTIMENT ANALYSIS

Sentiment Analysis is a Natural Language Processing and Information Extraction task that aims to obtain writer's feelings expressed in positive or negative comments, questions and requests, by analyzing a large numbers of documents. Generally speaking, Sentiment analysis with respect to the subject of speaker or an author's point of view or a document to determine the overall functionality.

Sentiment Analysis uses various classification techniques to identify the tone of a given piece of text. It indicates whether the text is positive, negative or neutral. This analysis can be aggregated over large sets of data and the resulting information can be helpful in different contexts.

- High complexity
- Higher execution time
- Low Accuracy

IX. OPINION MINING AND SUMMARIZATION PROCESS

Opinion Mining also called sentiment analysis is a process of finding user's opinion towards a topic. Opinion mining concludes whether user's view is positive, negative, or neutral about product, topic, event etc. Opinion mining involves analyzing user's opinion, attitude, and emotion towards particular topic. This consists of first categories text into

subjective and objective information, and then finding polarity in subjective text. Opinion mining can be performed word, sentence or document level. Opinion mining and summarization process involve three main steps, first is Opinion Retrieval, Opinion Classification and Opinion Summarization. Summarization of opinions is a major part in opinion mining process. Summary of reviews provided should be based on features or subtopics that are mentioned in reviews. Therefore, feature extraction [4] and opinion summarization are key issues. Many researchers worked on summarization product reviews [2]. The opinion summarization process mainly involve following two approaches. One is Feature based summarization another one is Term Frequency based summarization.

Opinion Retrieval: Opinion retrieval is a process of collecting reviews text from review websites. Different review websites contain reviews for products, movies, hotels, news etc. Information retrieval techniques such as web crawler can be applied to collect review text data from many sources and store them in database. This step involves retrieval of reviews, microblogs, comments etc of user. We should only consider the data which contain subjective data but not the objective data. Reviews are retrieved by query based information retrieval techniques.

Opinion Classification: Primary step in sentiment analysis is classification of review text. Given a review document $D = \{d_1..d_n\}$ and predefined categories set $C = \{\text{positive, negative}\}$, sentiment classification is to classify each d_i in D , with expressed in C . There are many approaches for sentiment classification in opinion text. Machine learning and lexicon based approach is more popular.

Machine learning approach for opinion classification: The machine learning approach uses supervised learning method for classification of review text. The first step is to train a classifier using sample of reviews with its class (positive/negative). Then the built model of trained classifier is used to predict category of new text

reviews. Popular machine learning classifiers for text categorization are Support Vector Machines (SVM) and Naive Bayes(NB).

Lexicon based approach for opinion classification: The lexicon Approach predicts sentiment of review text using databases which contain word polarity values e.g. SentiWordNet [10]. Review text is classified by calculating and averaging polarity score of individual words in sentences. Many factors such as word position, word relationships, negation handling should be considered while sentiment classification using lexicon based approach.

Opinion Summarization Summarization of opinions is a major part in opinion mining process. Summary of reviews provided should be based on features or subtopics that are mentioned in reviews. Therefore, feature extraction and opinion summarization are key issues. Many researchers worked on summarization product reviews. The opinion summarization process mainly involve following two approaches. 1)Feature based summarization: This type summarization involve finding of frequent terms (features) that are appearing in many reviews. The summary is presented by selecting sentences that contain particular feature information. Features present in review text can be identified using Latent Semantic Analysis (LSA) method. For a short summary of product reviews, product features and opinion words associated. 2)Term Frequency based summarization: Term frequency is count of term occurrences in a document. If a term has higher frequency it means that term is more import for summary presentation. In many product reviews certain product features appear frequently and associated with user opinions about it. In this method sentences are scored by term frequency. The summary is presented by selecting sentences that are relevant and which contain highest frequency terms. Opinion Summarization process is shown in Fig.5 It shows review text is preprocessed which involve sentence segmentation and tokenization of sentence in terms. After calculating term frequency of each term, each sentence score and relevance is

calculated. As per the compression rate highest scoring and relevant sentences are presented in summary.

X. CONCLUSION

Sentiment detection has a wide variety of applications in information systems, including classifying reviews, summarizing review and other real time applications. There are likely to be many other applications that is not discussed. It is found that sentiment classifiers are severely dependent on domains or topics. From the above work it is evident that neither classification model consistently outperforms the other, different types of features have distinct distributions. It is also found that different types of features and classification algorithms are combined in an efficient way in order to overcome their individual drawbacks and benefit from each other's merits, and finally enhance the sentiment classification performance. Due to web and social network, large amount of data are generated on Internet every day. This web data can be mined and useful knowledge information can be fetched through opinion mining process. This paper discussed different opinion classification and summarization approaches, and their outcomes along with semantic analysis. This study shows that machine learning approach works well for sentiment analysis of data in particular domain such as movie, product, hotel etc., while lexicon based approach is suitable for short text in micro-blogs, tweets, and comments data on web. Due to applications of opinion detection in various domains such as product, travel, movie etc, it is emerged as a popular topic in web mining.

XI. REFERENCES

- [1]. Youngsub Han, Kwangmi Ko Kim, "Sentiment Analysis on Social Media Using Morphological Sentence Pattern Model", IEEE, 2017, pp: 79-84
- [2]. Tanvi Hardeniya , D. A. Borikar , " An Approach to Sentiment Analysis Using Lexicons With Comparative Analysis of Different Techniques" IOSR Journal of computer engineering, Vol. 8, Issue 3, 2016.
- [3]. Farhan Hasan Khan "TOM: Twitter Opinion Mining Framework using Hybrid Classification Scheme, Decision Support Systems", 2016
- [4]. G. Vaitheeswaran , L. Arockiam " Machine Learning Based Approach to Enhance the Accuracy of Sentiment Analysis " International Journal of Computer Science and Management Studies , Vol. 4 , Issue 5 , 2016.
- [5]. Liza Wikarsa, Sherly NoviantiThahir, "A Text Mining Application of Emotion Classifications of Twitter's Users Using Naïve Bayes Method", IEEE, ISBN: 978-1-4673-8434-6, 2015
- [6]. Namita Mittal , Basant Agarwal "A Hybrid Approach for Twitter Sentiment Analysis", 2015
- [7]. Pedro P. B. Filho ,Thoago A. S. Pardo "NILC_USP : A hybrid system for sentiment analysis in Twitter Messages " Internatonal workshop on Semantic Evaluation,2014.
- [8]. Amit G. Shirbhate ,Sachin N. Deshmukh "Feature Extraction for Sentiment Classification on Twitter Data" International Journal of Science and Research.
- [9]. Shoushan Li, Sophia Yat Mei Lee, Ying Chen , Chu-Ren Huang , Guodong Zhou" Sentiment Classification and Polarity Shifting".
- [10]. K. Revathy, B. Sathiyabhama " A Hybrid Approach for Supervised Twitter Sentiment Classification" International Journal of Computer Science and Business Informatics.
- [11]. Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou, "Movie Rating and Review

- Summarization in Mobile Environment”, IEEE VOL. 42, NO. 3, MAY 2012.
- [12]. Pang, B., L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2002), 2002.
- [13]. Ku, L.-W., Liang, Y.-T., & Chen, H, “Opinion extraction, summarization and tracking in news and blog corpora”. In AAAI-CAAW’06.
- [14]. Melville, Wojciech Gryc, “Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification”, KDD09, June 28–July 1, 2009, Paris, France. Copyright 2009 ACM 978-1-60558-495-9/09/06.
- [15]. Titov, I., McDonald, R.: A Joint Model of Text and Aspect Ratings for Sentiment Summarization. In: Proceedings of ACL-2008: HLT, pp. 308–316 (2008).
- [16]. Nilesh M. Shelke, Shrinivas Deshpande, PhD. and Vilas Thakre, PhD., Survey of Techniques for Opinion Mining, International Journal of Computer Applications (0975 – 8887) Volume 57– No.13, November 2012.
- [17]. Xiaohui Yu, Member, IEEE, Yang Liu, Member, IEEE, Jimmy Xiangji Huang, Member, IEEE, and Aijun An, Member, IEEE, “Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain,” IEEE Transactions on Knowledge and Data Engineering, Vol. 24, NO. 4, APRIL 2012.
- [18]. Christopher C. Yang and Tobun Dorbin Ng, Member, IEEE, “Analyzing and Visualizing Web Opinion Development and Social Interactions With Density Based Clustering,” IEEE Transactions on Systems, man, and cybernetics—part a: systems and humans, vol. 41, no. 6, november 2011.
- [19]. Ainur Yessenalina, Yisong Yue, Claire Cardie, Multi-level Structured Models for Document-level Sentiment Classification, Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pages: 1046–1056, MIT, Massachusetts, USA, 9-11 October 2010. Association for Computational Linguistics.
- [20]. Hanhoon Kang, SeongJoon Yoo, Dongil Han, “Sentilexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews”. Expert Systems with Applications 39 (2012) 6000–6010.
- [21]. Liza Wikarsa, Sherly Novianti Thahir, “A Text Mining Application of Emotion Classifications of Twitter’s Users Using Naïve Bayes Method”, IEEE, ISBN: 978-1-4673-8434-6, 2015
- [22]. Tanvi Hardeniya, D. A. Borikar, “An Approach to Sentiment Analysis Using Lexicons With Comparative Analysis of Different Techniques” IOSR Journal of computer engineering, Vol. 8, Issue 3, 2016.
- [23]. Farhan Hasan Khan “TOM: Twitter Opinion Mining Framework using Hybrid Classification Scheme, Decision Support Systems”.
- [24]. G. Vaitheeswaran, L. Arockiam “Machine Learning Based Approach to Enhance the Accuracy of Sentiment Analysis” International Journal of Computer Science and Management Studies, Vol. 4, Issue 5, 2016.
- [25]. Namita Mittal, Basant Agarwal “A Hybrid Approach for Twitter Sentiment Analysis”.
- [26]. Pedro P. B. Filho, Thoago A. S. Pardo “NILC_USP : A hybrid system for

- sentiment analysis in Twitter Messages " International workshop on Semantic Evaluation,2014.
- [27]. Amit G. Shirbhate ,Sachin N. Deshmukh "Feature Extraction for Sentiment Classification on Twitter Data" International Journal of Science and Research.
- [28]. Shoushan Li, Sophia Yat Mei Lee, Ying Chen , Chu-Ren Huang , Guodong Zhou" Sentiment Classification and Polarity Shifting".
- [29]. K. Revathy, B. Sathiyabhama " A Hybrid Approach for Supervised Twitter Sentiment Classification" International Journal of Computer Science and Business Informatics.
- [30]. M. Govindarajan"Sentiment Analysis of Movie Reviews using Hybrid Method of Naive Bayes and Genetic Algorithm "International Journal of Advanced Computer Research, Vol.3, Issue-13, December-2013