ISSN: 0975-833X

*Available online at http://www.journalcra.com*

*International Journal of Current Research*
*Vol. 10, Issue, 09, pp.73661-73666, September, 2018*

**DOI: https://doi.org/10.24941/ijcr.32244.09.2018**

**INTERNATIONAL JOURNAL OF CURRENT RESEARCH**

## RESEARCH ARTICLE

# A KNN BASED TECHNIQUE ON OPINION MINING USING SEMANTIC ANALYSIS

***Gaganpreet Singh and Rekha Bhatia**

Punjab University Regional Centre for Information Technology and Management

**ARTICLE INFO**

*Article History:*
Received 27th June, 2018
Received in revised form
20th July, 2018
Accepted 19th August, 2018
Published online 30th September, 2018

*Key Words:*

kNN, Sentiment Analysis,
Opinion Mining.

**ABSTRACT**

Due to the sheer volume of opinion rich web resources such as discussion forum, review sites, blogs and news corpora available in digital form, much of the current research is focusing on the area of sentiment analysis. People are intended to develop a system that can identify and classify opinion or sentiment as represented in an electronic text. Therefore product features or aspects have got significant role in sentiment analysis. In addition to sufficient work being performed in text analytics, feature extraction in sentiment analysis is now becoming an active area of research. This review paper discusses existing techniques and approaches for feature extraction in sentiment analysis and opinion mining.



**Citation:** ***Gaganpreet Singh and Rekha Bhatia*****, 2018.** "A knn based technique on opinion mining using semantic analysis", *International Journal of Current Research*, 10, (09), 73661-73666.

## INTRODUCTION

Sentiment is a view, feeling, opinion or assessment of a person for some product, event or service (Au, 2010; Chen, 2006 and Ziqiong Zhang, 2011). Sentiment Analysis or Opinion Mining is a challenging Text Mining and Natural Language Processing problemfor automatic extraction, classification and summarization of sentiments and emotions expressed in online text (Au, 2010 and Chen, 2006). Sentiment analysis is replacing traditional and web based surveys conducted by companies for finding public opinion about entities like products and services (Ziqiong Zhang, 2011). 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.

**Corresponding author:* **Gaganpreet Singh**
Punjab University Regional Centre for Information Technology and Management.

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.

### 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 web-sites 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 analysing 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.

**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.

**Opinion Mining:** Opinion Mining (OM), a promising discipline, defined as combination of information retrieval and computational linguistic techniques deals with the opinions expressed in a document (Pang, 2002). The field aims at solving the problems related to opinions about products, politics in newsgroup posts, review sites, etc (Ku). There are different techniques for summarizing customer reviews like Data Mining, Information Retrieval, Text Classification and Text Summarization (Melville, 2009). Opinion Mining or Sentiment Analysis is the field to extract the opinionated text datasets and summarize in understandable form for end user (Titov, 2008). 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 (Kaiquan, 2011 and Ziqiong Zhang, 2011). Web 2.0 (Melville, 2009), 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 (Titov, 2008). 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 (Ku, ? and Titov, 2008). 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 (Au, 2010).

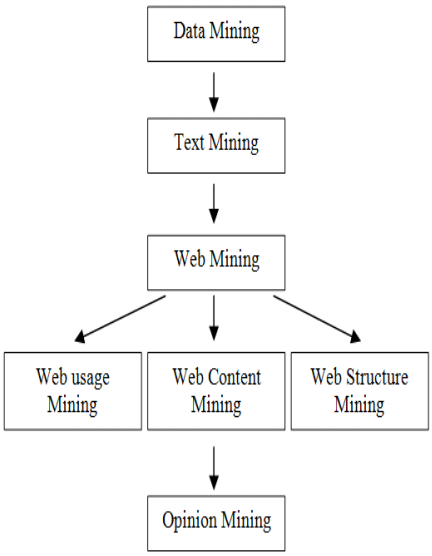



**Fig. 1. Opinion Mining Overview**

**Applications of Opinion Mining**

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

- 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.

- 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.
- 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).

**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.

**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 (Padmaja, 2013), and opinion summarization are key issues. Many researchers worked on summarization product reviews (Kaiquan, 2011). 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 = {d1..d1} and predefined categories set C = {positive, negative}, sentiment classification is to classify each di 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.

**Proposed Work**

In the proposed algorithm first of all twitter data is downloaded and then analyzed to pre-process data in which the username and time stamp are removed. After that the features are extracted from the pre-processed data. If the processed data is high dimensional data i.e. non–linear data then Enhanced Decision Tree is implemented to check t=out the root node and then recommendation system is applied that is ISTS2 and the results are generated from the imported pre-processed twitter data.

**Algorithm**

1. o = opinion word
2. initial_opinion_orientation = opinion's orientation
3. done = FALSE
4. distance = 1
5. WHILE (distance <= 5 AND !done ) DO
6. IF (o's position – distance) is coordinating conjunction
7. done = TRUE
8. END IF
9. IF (o's position – distance) is aspect
10. done = TRUE
11. END IF
12. IF (o's position – distance) is opinion
13. done = TRUE
14. END IF
15. IF (o's position – distance) is negation word
16. done = TRUE
17. orientation = opposite (initial_opinion_orientation)
18. END IF
19. distance++
20. END WHILE

**Similarity Measure**

**Tanimoto coefficient**: It is comparability between two sets. It is a proportion of convergences. Accept that set X is {B,C, D} and set Y is {C, D, E}. The Tanimoto coefficient T of two set An and B is 0.5. This metric doesn't consider the client rating yet the instance of an extremely scanty informational index is productive.

**Cosine likeness:** The Cosine similitude is known as the Vector comparability or Cosine coefficient. This metric expect that regular rating things of two clients are two focuses in a vector space model, and afterward computes cosə between the two focuses.

**Individual's Correlation:** In Equation, SU1 is the standard deviation of client U1. The Pearson Correlation measures the quality of the straight connection between two factors. It is generally connoted by r, and has values in the range [-1.0,1.0]. Where - 1.0 is an impeccable negative connection, 0.0 is no relationship, and 1.0 is a flawless positive relationship.

**Spearman's Rank Correlation:** The Spearman Rank Correlation additionally measures the quality of the straight connection between two factors. Not at all like the Pearson Correlation, this metric considers rank of scores. So this comparability measure has more broad pertinence than the Pearson Correlation, which isn't appropriate outside a standardized inclination run. Since the scope of inclination scores for CF is standardized, the Spearman Rank Correlation in the CF field indicates practically identical execution to the Pearson Correlation.

**Evaluation Metrics**

**Accuracy:** It is a description of systematic errors, a measure of statistical bias; as these cause a difference between a result and a "true" value, ISO calls this trueness.

Accuracy = (TP+TN)/(TP+TN+FP+FN)

**Precision:** It is a description of random errors, a measure of statistical variability.

Precision(P) = TP / (TP + FP)

**Recall:** Recall (also known as sensitivity) is the fraction of relevant instances that have been retrieved over total relevant

instances in the image. Both precision and recall are therefore based on an understanding and measure of relevance.

Recall (R) = TP / (TP + FN)

True positive (TP) = the number of cases correctly identified as true
False positive (FP) = the number of cases incorrectly identified as true
True negative (TN) = the number of cases correctly identified as false
False negative (FN) = the number of cases incorrectly identified as false

**F-Measure:** In statistical examination of parallel arrangement, the F1 score (likewise F-score or F-measure) is a measure of a test's precision. It considers both the accuracy p and the review r of the test to process the score: p is the quantity of right positive outcomes partitioned by the quantity of every single positive outcome, and r is the quantity of right positive outcomes isolated by the quantity of positive outcomes that ought to have been returned. The F1 score can be deciphered as a weighted normal of the exactness and review, where a F1 score achieves its best an incentive even from a pessimistic standpoint at 0

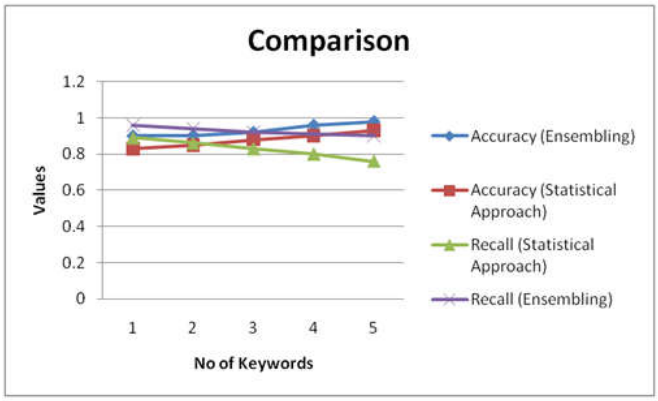
F = 1/((a)(1/P) + (1-a)(1/R))

**Performance Evaluation on the basis of parametres for an Existing Technique**

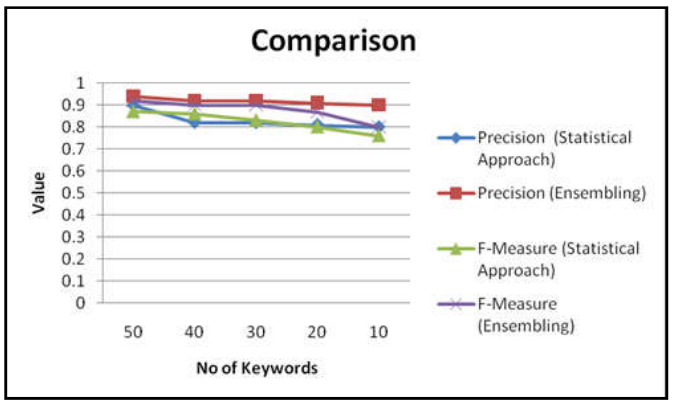
| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area |
|---|---|---|---|---|---|
| 0.45 | 0 | 1 | 0.45 | 0.621 | 0.888 |
| 1 | 0.55 | 9.11 | 1 | 0.954 | 0.888 |
| 0.917 | 0.467 | 9.25 | 0.917 | 0.904 | 0.888 |

**Performance Evaluation on the basis of parametres for an Proposed Technique**

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area |
|---|---|---|---|---|---|
| 0.5 | 0.007 | 0.923 | 0.5 | 0.649 | 0.904 |
| 0.993 | 0.5 | 0.921 | 0.993 | 0.966 | 0.904 |
| 0.921 | 0.428 | 0.921 | 0.921 | 0.911 | 0.904 |




Proposed approach described how to adapt discriminative re ranking to improve the performance of the generative models for grounded language learning. Specifically, we delve into the problem of navigational instruction following discussed in last chapter and aid two PCFG models described earlier with the framework of discriminative re-ranking. Conventional methods of discriminative re-ranking require gold-standard references in order to evaluate candidates and update the model parameters in the training phase of re-ranking. However, grounded language learning problems do not have gold-standard references naturally available; therefore, direct application of conventional re-ranking approaches do not work. Instead, we show how the weak supervision of response feedback (e.g., successful task completion in the navigational task) can be used as an alternative, experimentally demonstrating that its performance is comparable and even more effective compared to training on gold-standard parse trees. Modified Re-ranking Algorithm for Grounded Language Learning. In re-ranking, a baseline generative model is first trained and it generates a set of candidate outputs for each training example.




**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.

## REFERENCES

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, 9-11October 2010.AssociationforComputationalLinguistics.

Au, N., R. Law, and D. Buhalis. The impact of culture on ecomplaints: Evidence from the chinese consumers in hospitality organization. In U. Gretzel, R. Law, and M. Fuchs, editors, Information and Communi- cation Technologies in Tourism 2010, pages 285–296. Springer Verlag Wien, 2010.

Bing Liu. Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012.

Chen, C. F. Ibekwe-SanJuan, E. SanJuan, and C. Weaver. Visual analysis of conflicting opinions. In IEEE Symposium n Visual Analytics Science And Technology, pages 35 – 42, 2006.

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.

Christopher C. Yang and TobunDorbin 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, novem BER 2011.

Hanhoon Kang, SeongJoonYoo, Dongil Han," Sentilexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews". Expert Systems with Applications 39 (2012) 6000–6010.

Jaganadh, G. 2012. Opinion mining and Sentiment analysis CSI communication.

KaiquanXu , Stephen Shaoyi Liao , Jiexun Li, Yuxia Song, "Mining comparative opinions from customer reviews for Competitive Intelligence", Decision Support Systems 50 ,743–754, (2011).

Ku, L.-W., Liang, Y.-T., & Chen, H, "Opinion extraction, summarization and tracking in news and blog corpora". In AAAI-CAAW‟06.

Liu, C.L. W.H. Hsaio, C.H. Lee, G.C. Lu and E. Jou, "Movie Rating and Review Summarization in Mobile Environment", IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, Vol. 42, No. 3, pp. 397-407, 2012.

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.

Nilesh M. Shelke, Shriniwas 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.

Padmaja. S and Sameen Fatima, "Opinion Mining and Sentiment Analysis –An Assessment of Peoples‟ Belief: A Survey", International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC) Vol.4, No.1, February 2013

Pang, B. L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques,"inProc.ACL-02Conf.Empirical Methods Natural Lang. Process., 2002, pp. 79–86.

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.

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).

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.

Yingcai Wu, Furu Wei, Shixia Liu, Norman Au, Weiwei Cui, Hong Zhou, and Huamin Qu, Member, IEEE "Opinion Seer: Interactive Visualization of Hotel Customer Feedback," IEEE transactionson visualization and computer graphics,vol.16,no.6, november/December 2010.

Ziqiong Zhang, Qiang Ye, Zili Zhang, Yijun Li, "Sentiment classification of Internet restaurant reviews written in Cantonese", Expert System with applications,2011.

*******