# A Survey On Sentiment Analysis

## M. Janaki  M.C.A, M.Phil, NET, SLET

Assistant Professor, Department of Computer Science, Sacred Heart Girls First Grade College, Bengaluru,
Karnataka, India

## ABSTRACT

Sentiment analysis is a technique to analyze people's opinion on given topics such as political, social, and economical or review on product etc.The techniques for sentiment analysis include machine learning (supervised and unsupervised), and lexical-based approaches. The most important focus of the realm of Sentiment analysis lies find the emotions indicate within the texts. Sentiment analysis allows us to extract reviews and present the summary which could be beneficial for market research and product enhancement.It helps business and organization because it's easy for them to know how people feel about their product or services so that they can make better decision or improve their services. For that purpose we have different sentiment analysis techniques like Naïve Bayes, Maximum Entropy, and Support Vector Machine which gives correctness of information or provides us accuracy. For sentiment we use machine learning because it train the computer to recognize the emoticon behind the sentence.

Keywords : Sentiment Analysis, Polarity classification, Techniques, Applications, Challenges.

## I. INTRODUCTION

Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product or topic. Sentiment Analysis has many names. It's often referred to as subjectivity analysis, opinion mining, and appraisal extraction, The purpose of sentiment analysis is to automatically determine the expressive direction of user reviews [6].Sentiment analysis, which is also called opinion mining, involves in building a system to collect and examine opinions about the product made in blog posts, comments, reviews or tweets. Two types of sentiment analysis are Subjectivity/Objectivity Identification and Feature /Aspect Based Sentiment Analysis.The users are now more interested to share their opinion on the internet using ratings, reviews, and a suggestion with diversified forms of user's expression.

The aim of Sentiment Analysis is to with strain this data in order to obtain critical information regarding public opinion, an emotions that help to make smarter business decisions, political campaigns and better product consumption.

To sentiment the data, here are some methodology like extract data from any site like twitter, amazon etc. Next step is cleaning process where irrelevant data is removed then on the pre-processed data apply feature selection which extract useful data from bag of words then give training and testing to it. Finally apply classifier algorithm which gives accuracy of that classifier.

Information available in textual format can be classified into two main things: Facts and Opinions. An objective expression made by user regarding

certain objects, entities or events and their attributes is known as facts. In the similar way, a subjective expression which describes emotions of a person, her sentiments and performance assessment about objects, entities and events and their characteristics is known as opinion.
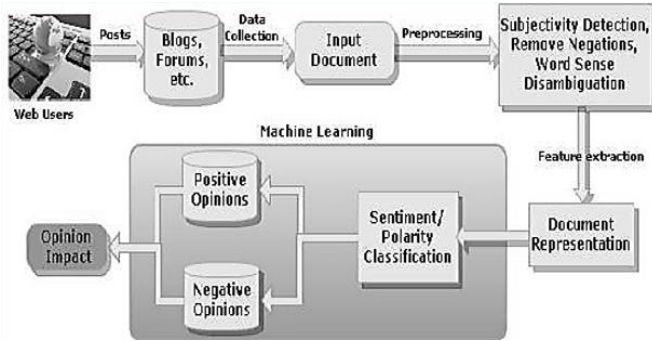


**Fig 1 :** Systematic Flow of Sentiment Analysis

## 1.1 Explicit Opinion and Implicit Opinion

Sentiment that appears in text comes in two flavours: explicit where the subjective sentence directly expresses an opinion ("It's a beautiful day"), and implicit where the text implies an opinion ("The earphone broke in two parts").Most of the work done so far focuses on the first kind of sentiment, since it is the easier one to analyse.

## 1.2 Feature Selection in Sentiment Classification

Feature extraction phase deals with feature types (which identifies the type of features used for opinion mining), feature selection (used to select good features for opinion classification), feature weighting mechanism (weights each feature for good recommendation) reduction mechanisms (features for optimizing the classification process). [11]

**Types of features used for Sentiment analysis could be:**

1) Term frequency (The presence of the term in a document carries a weight age).
2) Term co-occurrence (features which occurs together like uni-gram, bi-gram or n-gram),
3) Part of speech information (POS tagger is used to separate POS tokens).
4) Opinion words (Opinion words are words which express positive (good) or negative (bad) emotions)
5) Negations (Negation words (not, not only)
6) Syntactic dependency (It is represented as a parse tree and it contains word dependency based features)

**Term (T) and Term Frequency (TF)** The feature considered as individual word or word n-grams is called as term and its occurrence count in the document is known as term frequency. [8]

**Part of Speech Tags (POS)** It is the method used to assign a Part-of- Speech to every word present in the sentence. Every word in the sentence is assigned a tag like, verb, noun, prepositions and adjective etc. In English language, mainly adjectives are used to identify subjectivity and opinions. So, the earlier researchers used these adjectives as significant indicators of either subjectivities or opinions and counted these adjectives as the special features in the field of opinion mining.

Opinion Words and Opinion Phrases In the opinionated text positive or negative sentiments which are commonly used to express emotions of the opinion holder are called as opinion words. We can consider the beautiful, good, amazing, etc. as positive opinion words and negative opinion words like, bad, weak, poor, etc. Moreover, instead of such individual opinion words, there are also idioms and phrases which can be used to indicate opinions. Consider an example, "These opera tickets cost us an arm and a leg". Here, 'cost someone an arm and a leg' is a phrase which means having a negative impact of something. Therefore, opinion phrases and opinion words play a vital role in performing sentiment analysis.

1.3 Preprocessing Task into Several Sub phases:
Pre-processing the text is a process of cleaning the text and prepare them for text classification. Usually the online product reviews contain some noisy and

irrelevant information such as tags, scripts and advertisements.

There are some predefined steps used for pre-processing of texts [19]. They are online text cleaning, White space removal, Abbreviation expansion, Stemming, Removal of Stop words, Negation handling and finally feature selection. Fig. 1 illustrates the steps involved in processing of sentiment analysis. Online text cleaning involves in clean up line breaks, HTML tags and word formatting. Removing the empty spaces from the document is called white

space removal. In abbreviation expansion, the abbreviations such as TV, AC are expanded as Television and Air Conditioner by using pattern recognition and regular expression techniques. Stemming attempts to remove the inflected forms of a word, in order to reduce each word to its root form. Stop words such as 'the', 'of', 'a' are removed in next step [20]. But it should be deliberated for the opinion words which are expressed as phrasal words. In negation handling the negative terms are considered for polarity. For feature selection(remove noisy features), Point- wise Mutual Information (PMI), Chi- Square and Latent Semantic Indexing (LSI) methods are used [21].
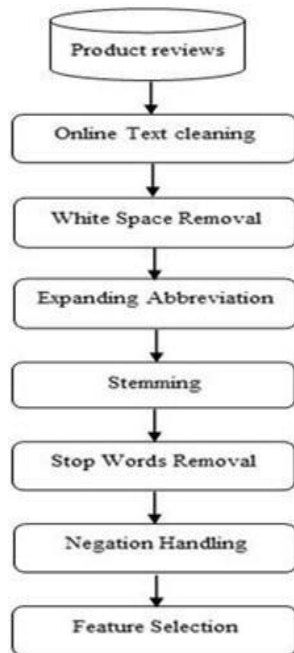


**Figure 1.** Processing steps in Sentiment Analysis

## DIFFERENT LEVELS OF SENTIMENT ANALYSIS

Different three levels in sentiment analysis which is document level, sentence level and aspect level. In document level i.e. identified that is the review is positive or negative. In sentence level i.e., identified every sentence is positive or negative and in aspect level entities and their features/aspects sentiments is positive and negative. [2]

### Document level

In Document level analysis task is characterize whether an entire opinion of document level communicates a positive or negative supposition For instance, given thing audit, the framework figures out if the survey communicates a general positive or negative decision about anything. This undertaking is regularly known as document level sentiment classification

### Sentence level

In Sentence level the fundamental undertaking is goes to the Sentence and makes sense of if every sentence communicated a positive, negative, or neutral sentiment. Neutral means no opinion about any sentence. This level of investigation is immovably related to the subjectivity arrangement. Which is recognizes sentences (called target sentences) [2] that is express genuine information from the sentences (called subjective sentences) that express subjective perspectives and opinions.

### Aspect level

In Aspect Level both the document level and the sentence level analyses do not discover what exactly people liked and didn't like. Aspect level performs better- grained investigation. Aspect level is directly looks at the opinion itself. In the Aspect level is depend on the possibility that an opinion consists of a sentiment positive, negative or neutral or an objective of sentiment

For e.g. Sentence is "The Redmi phone's call quality is amazing, yet its battery life is short" assesses two focuses first is call quality second is battery life, of Redmi (component). The conclusion on Redmi's call quality is certain in sentence however the opinion on its battery life is negative. Redmi phone's call quality and battery life of Phone are the feeling targets. In this level of investigation, an organized of assessments about elements and their viewpoints can be created, which turns unstructured content to organized.

## NEGATION:

Negation is a very common linguistic construction that affects polarity and, therefore, needs to be taken into consideration in sentiment analysis. When treating negation, one must be able to correctly determine what part of the meaning expressed is modified by the presence of the negation. Most of the times, its expression is far from being simple , and does not only contain obvious negation words, such as not, neither or nor. Research in the field has shown that there are many other words that invert the polarity of an opinion expressed [16],

## SENTIMENT ANALYSIS TECHNIQUES:

Sentiment Classification techniques are separated into two different techniques which is ML and Lexicon based Approaches. [3]

Machine learning methods are based on training an algorithm. It mostly based on classification on a set of selected features for a specific purpose and then test on another set whether it is able to detect the right features and give the right classification. A lexicon based method depends on a predefined list or corpus of words with a certain polarity. There are wide variety of machine learning methods such as Naïve Bayes Classifier, Support Vector Machine and Maximum Entropy Classifier .Various lexicon based sentiment methods such as Senticnet.
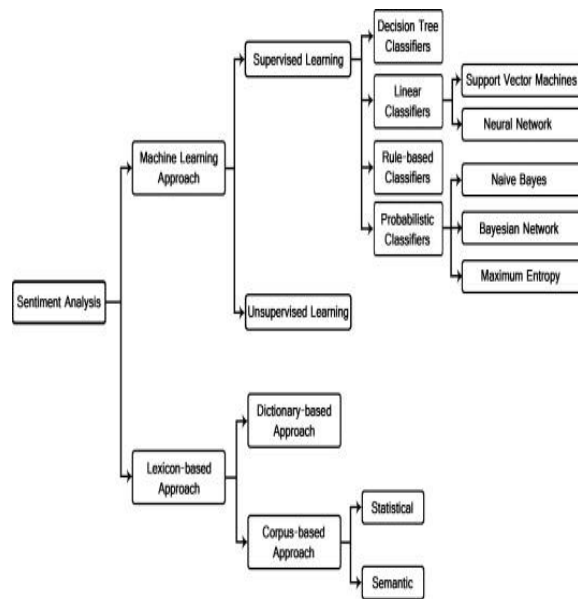


## II. LITERATURE REVIEW

In the G. Vinodhini[12] research paper, Naïve Bayes, a commonly used algorithm for document categorization is used to compute the probabilities by using the collective probabilities of topics and words. Support Vector Machine is a text categorization which outperforms the Naïve Bayes technique.

It searches for a decision surface to split the training data points into two categories and makes decisions based on the given support vectors.

It is observed in Liu [15] that, the opinion contents which are available online on internet as well as off line are containing mostly textual information used by the customer to provide relevant product feedback. The information available in textual format can be generally classified as either facts or opinions.

The research paper the survey gives an overview of the efficient techniques, recent advancements and the future research directions in the field of Sentiment Analysis[11].This research paper describes some of the considerable challenges in sentiment analysis and the techniques use to analyse, the main challenge in the opinion mining is to identify the sentiment expressed

by the text and the significant approaches of enhancing the performance of sentiment analysis are through i) N-Gram model,

ii) subjective lexicon, and iii) machine learning[5]

This study ensures an overall survey about sentiment analysis related to product reviews, and classification algorithms used for sentiment classification. It is a system thatidentifies and classifies opinion/sentiment as represented in electronic text [1].Sentiment Analysis is also investigated on Indian Language, Chinese language, Arabic Language [22] apart from the English language. At present, existing techniques towards sentiment analysis is focused on using lexicon generation in text-based processing [23], subjectivity detection [24], sentiment polarity detection [25], sentiment structurization [26], summarization-visualization tracking [27], etc. Apart from this, adopt of sentiment analysis is very much frequent for analyzing social network data from Facebook, Twitter, and Google [28].

## Current Trends and Techniques some novel approaches:

1. Document level sentiment classification: This technique, identifies whether the given document contains positive or negative sentiment about any topic. Generally classification techniques are used to solve these issues. The general features used in these techniques are: (1) terms and their occurrence frequency (for example the use of Tf- Idf), (2) POS taggers, (3) Opinion words and phrases, (4) Syntactic dependencies and (5) negative & Positive words.

2. Using unsupervised learning: For example, the use of POS tagger to identify two word phrases. It estimates the orientation of the extracted phrases using the Pointwise mutual information (PMI).

3. Sentiment analysis at sentence level: Techniques using this approach, considers the sentences as the source of single opinion. For a given a sentence s, it applies two sub-tasks: (a) Subjectivity classification: Determine whether s is a subjective sentence or an objective sentence, and (b) Sentence-level sentiment classification: If s is subjective, determine whether it expresses a positive or negative opinion.

## III.APPLICATIONS

1) It is mostly used in E-commerce activities. When any customer buys any item or service from the e-commerce websites, then it permits them to submit their opinions about qualities of shopping services and products. A summary for the product and various features of the product is provided by assigning ratings.
2) It is used in Entertainment by helping people to choose which movie or series to watch.
3) It is also used in Marketing. Nowadays, each company makes available the facility to its users to provide opinions about its products and services. Hence, it is helpful for businesses to save money as well as time because there is no need any more to conduct surveys as the feedbacks related to all the products are available on their sites.
4) It is also used in education domain, to help students to determine which university is good for studies

## IV.RESEARCH CHALLENGES

There are various challenges in Sentiment analysis. A few of them are discussed in this paper.
1) The very first challenge is "opinion word" which can be considered to be positive in one way but may be considered negative in another way.
2) Second challenge is that sometimes user may convey their sentiments in an unusual way. The text in a sentence can be difficult to identify as ironic or sarcastic and this can lead to faulty polarization and misleading sentiment analysis. Reference [8] discusses this problem.

3) The third challenge is the language i.e, the majority of the work done in opinion mining is focused on two languages: English and Chinese and other languages needs to be explored.

4) Now, the fourth challenge is the sentiment given on twitter is difficult to comprehend as it consists of poor abbreviations, lack of capital letters, spelling mistakes, no proper punctuations, and grammatical errors and so on.

5) Sixth challenge is in "detection of spam and fake comments, mainly through the recognition of duplicates, the association of qualitative with summary feedbacks, the recognition of outliers, and also the reputation of the reviewers".

## V.  CONCLUSION

Sentiment Analysis is one of the important research areas as it summarizes opinions and reviews of public. This survey highlights the main idea behind Sentiment Analysis and explains literature review, Sentiment Classification, challenges in detail. Sentiment analysis is one of the active research areas and several interesting works have been done in this field. Still, a fully useful and highly efficient system has not been introduced till now. But business organizations and academics are working hard to find the best system for sentiment analysis. Sentiment analysis helps in decision making and knowing people review by analyzing or giving rating to their views such as product reviews. By making use of this system, user can get suggestion for product to buy.Naive Bayes and Support Vector Machines are the most frequently used ML algorithms for solving SC problem.

## VI. REFERENCES

[1]. Sentiment Analysis: A Survey of Current Research and Techniques

[2]. Jeevanandam Jotheeswaran, Dr. S. Koteeswaran International Journal of Innovative Research in Computer and Communication Engineering

[3]. Xing Fang and Justin Zhan "sentiment analysis using product Review data" Department of computer science, North Carolina a&T State University Greensboro, NC, USA, 2015 Springer journal.

[4]. WalaaMeddhat , Ahmed Hassan ,Hoda Korashy "Sentiment analysis algorithms and applications: A survey, Ain Sham University, Faculty of Engineering, Computer & Systems Department, Egypt 19 April 2014.

[5]. Ayesha Rashid et al, "A Survey Paper: Areas, Techniques and Challenges of Opinion Mining", International Journal of Computer Science (IJCSI), Vol 10 Issue 6 No 2, Nov 2013.

[6]. 5A Survey On Challenges In Sentiment Analysis Lincy W and Naveen kumar M International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume 21 Issue 3 – APRIL 2016.

[7]. F. Luo, C. Li, and Z. Cao, Affective- feature-based sentiment analysisusing SVM classifier, 2016 IEEE 20th Int. Conf. Comput. Support. Coop.Work Des., pp. 276281, 2016.

[8]. Eirinaki, M., Pisal, S., Singh, J.: Feature-based opinion mining and ranking. J. Comput. Syst. Sci. 1175–1184 (2012)

[9]. Aggarwal Charu C, Zhai Cheng Xiang. Mining Text Data. Springer New York Dordrecht Heidelberg London: _ Springer Science+Business Media, LLC'12; 2012.

[10]. Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proc. of EMNLP, pages 79–86.

[11]. E. Marrese-Taylor, J. D. Velasquez, F. Bravo-Marquez, "Opinion Zoom: A Modular Tool to Explore Tourism Opinions on the Web", In the Proceedings of the 2013 IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT), CA, pp. 261–264, 2013.

[12]. S ChandrakalaAnd C Sindhu: Opinion Mining And Sentiment Classification: A SURVEY DOI: 10.21917/ijsc.2012.0065 12.G.Vinodhini, RM.Chandrasekaran "Sentiment Analysis and Opinion Mining: A Survey", Volume 2 Issue 6, June 2012.

[13]. S. ChandraKala, C. Sindhu2 "OPINION MINING AND SENTIMENT CLASSIFICATION: A SURVEY", ICTACT Journal on soft computing, Volume: 03, Issue: 01, October 2012

[14]. BakhtawarSeerat, FarouqueAzam "Opinion Mining: Issues and Challenges (A survey)", International Journal of Computer Applications (0975 – 8887)Volume 49– No.9, July 2012

[15]. Liu, B.: Sentiment analysis: a multi- faceted problem. In: IEEE Intelligent Systems, pp. 1–5 (2010)

[16]. Michael Wiegand and Alexandra Balahur, "A Survey on the Role of Negation in Sentiment Analysis", Proceedings of the Workshop on Negation and Speculation in Natural Language Processing, 2010.

[17]. Adam L. Berger, Stephen A. Della Pietra and Vincent J. Della Pietra, "A maximum entropy approach to natural language processing", Computational Linguistics, Vol. 22, No. 1, pp. 39– 71, 1996.

[18]. Thorsten Joachims. "Text categorization with support vector machines: Learning with many relevant features", Proceedings of the European Conference on Machine Learning, pp. 137–142, 1998.

[19]. Subhabrata Mukherjee, ―Sentiment Analysis : A Literature Survey ―,Indian Institute of Technology, Bombay. Department of Computer Science and Engineering, June 29, 2012.

[20]. Ms. KrantiVithalGhag, Dr.Ketan Shah, ―Comparative Analysis of Effect of StopwordsRemoval on Sentiment Classification‖, IEEE International Conference on Computer, Communication and Control (IC4-2015). 21.ZohrehMadhoushi, AR Hamdon, S Zainudin, ―Sentiment Analysis Techniques in Recent Works‖, Science and Information Conference 2015 July 28-30, 2015.

[21]. M. Biltawi, W. Etaiwi, S. Tedmori, Hudaib, and A. Awajan, "Sentiment Classification Techniques for Arabic Language: A survey," In Information and Communication Systems (ICICS), 7th International Conference, pp. 339- 346, 2016.

[22]. O. Kolchyna, T.TP. Souza, P.Treleaven, and T. Aste, "Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination," arXiv preprint arXiv: 1507.00955, 2015

[23]. M.Graña, C. Toro, "Advances in Knowledge-based and Intelligent Information and Engineering Systems," Volume 1", IOS Press, pp. 2273, 2012

[24]. "Sentiment analysis", https://en.wikipedia.org/wiki/Sentiment_ analysis, Retrieved, 16-Feb-2017

[25]. W. Medhat, A. Hassan, and H. Korashy, "Sentiment Analysis Algorithms and Applications: A Survey," Ain Shams Engineering Journal, Vol.5, No. 4, pp.1093-1113, 2014

[26]. A. Das, S. Banyopadhyay and B. Gambäck, "The 5W Structure for Sentiment Summarization Visualization-Tracking," ERCIM, Retrieved, 16- Feb- 2017

[27]. A. Selamat, H. Fujita, H. Haron, "New Trends in Software Methodologies, Tools and Techniques: Proceedings of the Thirteenth SoMeT_14," IOS Press, pp. 1128, 2014

## Cite this article as :