

# Sentiment Analysis Techniques: A Review

**DIVYANSHI SOOD<sup>1</sup>**

RESEARCH SCHOLAR,  
COMPUTER SCIENCE  
AND ENGINEERING,

CHANDIGARH  
UNIVERSITY, GHARUAN,  
MOHALI (PUNJAB)

[deevyanshisood@gmail.com](mailto:deevyanshisood@gmail.com)

**NITIKA KAPOOR<sup>2</sup>**

ASSISTANT PROFESSOR,  
DEPARTMENT OF CSE,

CHANDIGARH  
UNIVERSITY, GHARUAN,  
MOHALI (PUNJAB)

[nitika.cse@cumail.in](mailto:nitika.cse@cumail.in)

**DISHANT SHARMA<sup>3</sup>**

RESEARCH SCHOLAR,  
COMPUTER SCIENCE  
AND ENGINEERING,

CHANDIGARH  
UNIVERSITY, GHARUAN,  
MOHALI (PUNJAB)

[dishantsharma44.ds@gmail.com](mailto:dishantsharma44.ds@gmail.com)

## ABSTRACT

Sentiment can be described in the form of any type of approach, thought or verdict which results because of the occurrence of certain emotions. This approach is also known as opinion extraction. In this approach, emotions of different peoples with respect to meticulous rudiments are investigated. For the attainment of opinion related data, social media platforms are the best origins. Twitter may be recognized as a social media platform which is socially accessible to numerous followers. When these followers post some message on twitter, then this is recognized as tweet. The sentiment of twitter data can be analyzed with the feature extraction and classification approach. In this paper various sentiment analysis methods are reviewed and analyzed.

## KEYWORDS

Sentiment Analysis, Machine Learning, Twitter Data, Lexical Analysis

## 1. INTRODUCTION

Sentiments refer to the thoughts, beliefs or feelings of an author for something, which may include a person, thing, corporation or position. An opinion of a writer

towards some subject or on the entire relative polarity of the manuscript is summarized in the process of analyzing the sentiments. The outlook can be defined as the point of view or assessment or the emotional message of an individual. Opinions are critical influencer of a person's behavior. The opinions and insights for the reality are based on the way of others for perceiving the world. The fundamental objective of the OM (opinion mining) is to deduce the inclusive polarity of the document on some particular topic. SA is a major research sector in which various regions related to technology as well as social disciplines, such as sociology, psychology, and ethics are included. To mine the opinions is a skill in which the perception of people towards something or some particular concept is followed from an enormous judgments or reviews which are openly obtainable in web. OM plays a significant role when the decision is made after searching out other opinions. To illustrate, to buy a camera or any gadget, we can check reviews or comments or take the view of others before buying the product. Opinion Mining called as SA is a technique using which judgment of an individual is revealed for a topic or a product. This technique is utilized to classify the scrutiny of a user towards a region, event and object etc. in three classes: positive, negative or neutral. Subjective understanding related to any topic is contained in an

opinion text. The weblog, reviews, reaction etc. are involved in opinion text. These reviews are recognized as positive and negative reviews. OM and recapitulation process is executed in 3phases in which the opinion is retrieved, classified and summarized.

#### a. Opinion Retrieval

This process focuses on selecting the review text from several review sites. People are posted review on various places, news, objects and movies on the review websites. These reviews help other buyers to attain an idea about the quality and services related to that particular place or thing. The data of review text is gathered from diverse sources and stored in a database using a number of methods. This process contains a stage to retrieve the reviews, micro-blogs and comments of different users.

#### b. Opinion Classification

This process is executed for classifying the review text at first. To illustrate, in a given document  $M = \{M_1, \dots, M_i\}$  and a predefined category set  $K = \{\text{positive, negative}\}$ , the main intend is to classify each point in  $M$ . This process classifies the review in two parts such as positive and negative [9]. These kinds of tasks are accomplished using dictionary-based techniques and ML (machine learning) techniques.

#### c. Opinion Summarization

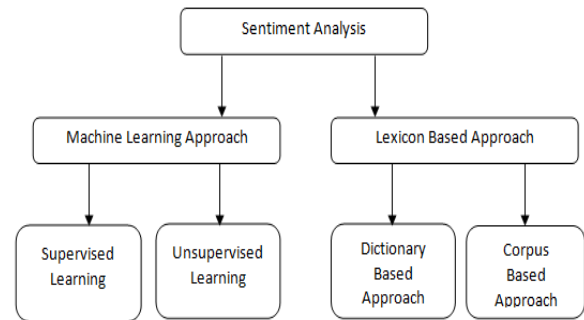
The process to summarize the opinions is the major stage of OM. The reviews are summarized on the basis of sub concepts or attributes present in reviews.

### 1.2. TECHNIQUES OF SENTIMENT CLASSIFICATION

Various prime DM (data mining) techniques are adopted for extracting the facts and information. The methods of OM are represented in Figure 1.4.

The entire procedure is consisted of various stages, in which text is cleaned online, white spaces are eliminated, acronym is amplified, stemming is done, stop words are removed elimination, refusal is managed, and the attributes are selected.

Thereafter, the classifiers are exploited to classify the opinions as positive, negative, and neutral.



**Figure 1.4 Techniques of Sentiment Analysis**

#### 1.2.1 Machine Learning Approaches

ML is planned on the basis of its diverse techniques. The issue related to classify the sentence level is resolved using these techniques. Moreover, a judgment is made on the syntactic attribute. ML is categorized in two kinds: supervised learning and unsupervised learning. ML assists the machines in adjusting their interior configuration in such a way for predicting the upcoming performance boost.

##### a. Supervised Learning

Supervised learning is effective for dealing with the classification issues. This category is emphasized on acquiring the workstation so that a classifier can be investigated. The major instance of classification learning is the digit recognition. Any issue, in which classification learning is significant and the classification can be detected easily, is tackled using the classification learning. In some scenarios, the execution of programmed classifications is not required for each occurrence of a problem in case the technique is capable of performing the classification itself.

##### b. Unsupervised Learning

Such algorithms are utilized to assume the patterns of any dataset for which no labeled outcomes are considered. Unlike the initial category, these methods are ineffective for dealing with the problem of regression or a classification. Due to this, it becomes difficult to train the algorithm normally. The unsupervised learning is useful to discover the

underlying data structure. The prior unknown data patterns are exhibited using this approach.

### 1.2.2. Lexicon Based Approach

The method of classifying the sentiment is utilized to estimate the opinion lexicons from text. Opinion lexicons are classified as: positive and negative. The most expected phase is confirmed using positive outlook terms and the negative outlooks are assisted in representing the most redundant phase. Opinion lexicons are called Opinion clauses or idioms.

#### a. Dictionary Based Approach

This technique is a suitable technique to collect the sentiment words due to the involvement of synonyms and opposites in the majority dictionary list for each and every word. Therefore, some seed sentiment words are generated with the help of simple and easy technique for reset in accordance with antonym and synonym organization of a lexicon.

#### b. Corpus-Based Approach

The above-mentioned technique has limitation that it is incapable of discovering the view or sentiment words according to the application domains. Thus, this technique is suggested for tackling the issue of dictionary-based system. This technique is able to acquire the opinion with the region-based orientations. This technique becomes popular as it provides opinion words with domain-based systems.

## 1.3 APPLICATIONS

Various applications utilized to analyze the sentiments are discussed as:

1. **Aid in decision making**—This application is become significant in daily life. The SA (sentiment analysis) is useful to select a particular product and decide from the accessible options in accordance with the views of other people.
2. **Designing and Building Innovative Products**—The SA is implemented to analyze the products based on the public reviews and opinions. The utilization and adaptive nature of a product are considered while analyzing the products.
3. **Recommendation System**—Various applications make the implementation of the recommendation system. This system assists the users related to books, online-media, entertainment, music, film industry and other art types. This system deploys the information regarding an individual, history, likes and dislikes to make proposals.
4. **Products Analysis**—SA is capable of examining several goods and to make selections. This analysis is efficient to select a product with regard to its specified attributes.
5. **Business Strategies**—The reactions of the public are considered to plan any business. The major goal of industries is to meet the demands of consumers. Hence, the strategies of companies are decided on the basis of public opinions and remarks
6. **User Modeling**—This application provides a mechanism of interfaces and utilizes to produce more interactive design. This mechanism emphasizes on establishing a communication amid human and computer.
7. **Information Diffusion**—This procedure is deployed to transmit and spread the information by the means of interactions. This theory is utilized to make same decisions in a sequential manner. The behavior of individuals acts significantly to analyze the sentiments.

## 2. LITERATURE REVIEW

**Pankaj Gupta et al.** (2016) [11] performed a case study of many unexplored fields of emotion analysis. The deployment of right knowledge is required to enhance the earlier methods. To summarize the text is a suitable method that was used to extract only valuable information for users from massive volume of gathered textual data. An intelligent system was planned using ML methods for extracting the data and analyzing the sentiments. With regard to summarize the text and analyze the review, a survey was conducted in this work. The merits and demerits of existing technologies were ascertained in this work.

**S. Jirpe and B. Joglekar et. Al** (2017) [12] discussed that many schemes to analyze the

sentiments were reviewed for polarity shift discovery. Reviews have shown that all types of polarity shifts can be noticed and removed through polarity shift discovery, removal, and hybrid models. Therefore, the polarity shift was capable of detecting and removing a variety of issues related to detect the polarity shift. The performance of ML algorithms was outstanding in this work.

**M. Bouazzi and T. Ohtsuki, et.al** (2019) [13] discussed that the multi-class classification techniques were utilized to classify the Twitter users' online posts. This work analyzed the advantages and disadvantages of this approach. This research suggested a novel framework for representing diverse thoughts and presenting the potential of this model in understanding the relationship between emotions. The accuracy of multi-class classification gets improved. Also, this model resolved the existing issues.

**M. Bouazzi and T. Ohtsuki** (2016) [14] presented a novel approach to sarcasm on Twitter. The predictive technique employs different components of the Tweet. His plan uses part-of-speech-tags for exposing the blueprints showing the level of disrespect for tweets. Although the outcomes obtained are found good, a huge training set is deployed to enhance these outcomes. This implies that the dataset assisted in extracting the tweets probably unable of covering all the possible samples of sarcasm. The authors also envisioned a more efficient way to grow their set having an initial training set of 6000 tweets and a more robust prototype with the hash tag "#sarcasm".

**Aisa M. Alshari et al.** (2018) [15] suggested a novel technique so that the distinction between SentiWordNet and corpus terminology was extended to improve the feature set for analyzing the sentiments. Learning from SentiWordNet was done by assigning polarity scores to available non-opinion words in the corpus vocabulary. A labeled dataset from Film Review was used to gauge this approach. The outcomes depicted the efficiency of the suggested technique over the standard SentiWordNet.

**Annet John et al.** (2019) [16] reviewed that semantic gap was assessed to measure the polarity difference through making up hybrid lexicons. The acronyms were employed to remove the sentiment scores

present in the text to assign the scores in a precise way. The hybrid lexicon was implemented to deal with this issue. This problem could be avoided by the hybrid lexicon, however, could not remove all the issues. With the aim to enhance the output, a hybrid technique was devised that integrated two techniques with other contextual sentiment alteration schemes that provided very accurate outcomes.

**Soonh Taj, et al.** (2019) [17] projected a lexicon dependent scheme in order to analyze the sentiment on news articles. BBC News dataset was executed to implement the projected scheme. The experiments were conducted to validate the projected scheme. The results indicated that the business and sports categories had a higher number of positive articles and the entertainment and technology-based categories had negative articles.

**S. Ahmed and A. Danti** (2016) [18] presented a novel method that was efficient for punters to decide on online reviews available on the web. The focus was on generating an effective technique to web reviews through mixed rule-based ML algorithms. The exploratory results demonstrated the efficiency of the new approach with maximal accuracy. The comprehensive experiments have been conducted on a variety of rule-based ML algorithms for classifying emotions.

**M. Tsytsarau and T. Palpanas** (2016) [19] discussed the issue of dissenting perceptions of feelings and opinions and their recognition in terms of a single or time factor for each case. He proposed procedures for the information-preservation emporium of several emotions and for understanding negativity for vast and coarse data reasons, which was the first comprehensive and systematic explanation for FIX.A probe estimate together with the model and actual data demonstrates the appropriateness and effectiveness of the proposed interpretation.

**Ankita and Nabizath Saleena** (2018) [20] presented a new ensemble classification system to improve efficacy to classify the sentiments in tweet. A single superior classification algorithm was developed by integrating the base learning classification algorithms. The experiments were carried out to assess the working of the presented approach. A

comparative analysis was conducted on the results obtained from the presented system against existing algorithms and the results exhibited the effectiveness of the presented classifier over the traditional classifier. Customers were able to choose the best products according to public opinion by the means of this innovative technique. The work ahead can expand this investigation to study neutral tweets present in the dataset.

**Benwang Sun, et al.** (2018) [21] conducted a study which focused on computing several techniques adopted for classifying the Tibetan micro-blog sentiments with regard to accuracy. For this, DL algorithms were implemented. The hybrid DL model was analyzed on the basis of several evaluation parameters and the results revealed an improvement of 1.22% in accuracy by using the presented technique.

**Chae Won Park and Dae RyongSeo**(2018) [22] presented a novel framework in order to analyze tweets and select an artificial intelligence approach to make better decisions. Users' opinion was taken to understand the tweets better. The purpose of sentiment analysis was to improve the sentiment of

products or services by actively researching them and developing NLP.

**Mukesh Yadav and VarunakshiBhojane**(2019) [23] presented three approaches based on which Hindi multi-domain review were utilized to analyze the sentiments. The Devanagari script available in the UTF-8 encoding system was utilized for the input in this research. In the first approach, Neural Network for pre-classified words was used for classifying the data. The second technique made the implementation of IIT-Bombay Hindi Center to classify the data. The third technique employed NN prediction in pre-classified sentences as labeled data with the objective of classifying the data. Finally, the accuracy of each approach was obtained. In the results, the first technique provided the accuracy up to 52%, second around 71% and third approach offered accuracy up to 70%.

**Table 1: Table of Comparison**

Author Name	Year	Description	Outcomes
<b>Mukesh Yadav</b>	2019	Hindi multi-domain review were utilized to analyze the sentiments. The Devanagari script available in the UTF-8 encoding system was utilized for the input in this research. In the first approach, Neural Network for pre-classified words was used for classifying the data. The second technique made the implementation of IIT-Bombay Hindi Center to classify the data. The third technique employed NN prediction	Finally, the accuracy of each approach was obtained. In the results, the first technique provided the accuracy up to 52%, second around 71% and third approach offered accuracy up to 70%.

		in pre-classified sentences as labeled data with the objective of classifying the data.	
<b>M. Bouazzi and T. Ohtsuki</b>	2019	It discussed that the multi-class classification techniques were utilized to classify the Twitter users' online posts. This work analyzed the advantages and disadvantages of this approach. This research suggested a novel framework for representing diverse thoughts and presenting the potential of this model in understanding the relationship between emotions.	The accuracy of multi-class classification gets improved. Also, this model resolved the existing issues.
<b>Annet John</b>	2019	Semantic gap was assessed to measure the polarity difference through making up hybrid lexicons. The acronyms were employed to remove the sentiment scores present in the text to assign the scores in a precise way.	A hybrid technique was devised that integrated two techniques with other contextual sentiment alteration schemes that provided very accurate outcomes.
<b>Soonh Taj</b>	2019	A lexicon dependent scheme in order to analyze the sentiment on news articles. BBC News dataset was executed to implement the projected scheme	The experiments were conducted to validate the projected scheme. The results indicated that the business and sports categories had a higher number of positive articles and the entertainment and technology-based categories had negative articles.
<b>Aisa M. Alshari</b>	2018	A novel technique so that the distinction between SentiWordNet and corpus terminology was extended to improve the feature set for analyzing the sentiments.	A labeled dataset from Film Review was used to gauge this approach. The outcomes depicted the efficiency of the suggested technique over the standard SentiWordNet.

<b>Ankita and Nabizath Sale</b>	2018	A new ensemble classification system to improve efficacy to classify the sentiments in tweet. A single superior classification algorithm was developed by integrating the base learning classification algorithms.	The experiments were carried out to assess the working of the presented approach. A comparative analysis was conducted on the results obtained from the presented system against existing algorithms and the results exhibited the effectiveness of the presented classifier over the traditional classifier.
<b>Benwang Sun</b>	2018	A study which focused on computing several techniques adopted for classifying the Tibetan micro-blog sentiments with regard to accuracy.	The hybrid DL model was analyzed on the basis of several evaluation parameters and the results revealed an improvement of 1.22% in accuracy by using the presented technique.
<b>Chae Won Park</b>	2018	A novel framework in order to analyze tweets and select an artificial intelligence approach to make better decisions. Users' opinion was taken to understand the tweets better.	The purpose of sentiment analysis was to improve the sentiment of products or services by actively researching them and developing NLP.
<b>S. Jirpe and B. Joglekar</b>	2017	It discussed that many schemes to analyze the sentiments were reviewed for polarity shift discovery. Reviews have shown that all types of polarity shifts can be noticed and removed through polarity shift discovery, removal, and hybrid models.	The polarity shift was capable of detecting and removing a variety of issues related to detect the polarity shift. The performance of ML algorithms was outstanding in this work.
<b>Pankaj Gupta</b>	2016	The deployment of right knowledge is required to enhance the earlier methods. To summarize the text is a suitable method that was used to extract only valuable information for users from massive volume of gathered textual data. An intelligent system was planned using ML methods for extracting the data and analyzing the sentiments.	With regard to summarize the text and analyze the review, a survey was conducted in this work. The merits and demerits of existing technologies were ascertained in this work.

<b>M. Bouazzi and T. Ohtsuki</b>	2016	It presented a novel approach to sarcasm on Twitter. The predictive technique employs different components of the Tweet. His plan uses part-of-speech-tags for exposing the blueprints showing the level of disrespect for tweets.	The authors also envisioned a more efficient way to grow their set having an initial training set of 6000 tweets and a more robust prototype with the hash tag "#sarcasm".
<b>S. Ahmed and A. Danti</b>	2016	A novel method that was efficient for punters to decide on online reviews available on the web. The focus was on generating an effective technique to web reviews through mixed rule-based ML algorithms	The exploratory results demonstrated the efficiency of the new approach with maximal accuracy. The comprehensive experiments have been conducted on a variety of rule-based ML algorithms for classifying emotions.
<b>M. Tsytarau and T. Palpanas</b>	2016	He proposed procedures for the information-preservation emporium of several emotions and for understanding negativity for vast and coarse data reasons, which was the first comprehensive and systematic explanation	The model and actual data demonstrates the appropriateness and effectiveness of the proposed interpretation.

### 3. Conclusion

Sentiment analysis is an umbrella term that covers many diverse fields, related to computer science as well as social disciplines such as sociology, psychology, and ethics. There are various stages of sentiment analysis methods proposed so far. The pre-processing step involves removing missing and unnecessary values from the dataset. The technique of extracting attributes establishes the relationship amid the attribute and the target set. The final stage applies a classifier to classify the data into certain classes such as positive, negative and neural. The various techniques for the sentiment analysis are reviewed in terms of certain parameters.

It is analyzed that machine learning are best performing technique for the sentiment analysis

### References

- [1] J. M. Wiebe, R. F. Bruce, and T. P. O'Hara, "Development and use of a Gold-standard Data Set for Subjectivity Classification." Proceeding of the 37<sup>th</sup> Annual Meeting of the Association for Computational Linguistics on Computational Linguistics, USA, pp. 246-253, 1999.
- [2] D. Pelleg and A. Moore, "X-means: Extending K-means with Efficient Estimation of the Number of Clusters," in Proc. of the 17<sup>th</sup> Int. Conference on Machine Learning, San Francisco, USA, pp. 727-734, 2000.



- [3] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," Conference on Empirical Methods in Natural Language Processing, USA, pp. 79-86, 2002.
- [4] E. Riloff and J. Wiebs, "Learning Extraction Patterns for Subjective Expressions," Conference on Empirical Methods in Natural Language Processing, Japan, pp. 105-112, 2003.
- [5] T. Wilson and J. Wiebe, "Annotating opinions in the world Press," 4<sup>th</sup> SIGdial Workshop on Discourse and Dialogue, Sapporo, Japan, pp. 13-22, 2003.
- [6] Hu and Liu, "Mining and Summarizing Customer Reviews," in International Conference on Knowledge Discovery and Data Mining, Seattle, USA, pp. 168-177, 2004.
- [7] B. Pang and L. Lee, "A sentimental education: Sentimental analysis using Subjectivity Summarization based on Minimum cuts," Proceeding of the 42<sup>nd</sup> Annual Meeting on Association for Computational Linguistics, USA, pp. 271-278, 2004.
- [8] S. M. Kim and E. Hovy, "Determining the Sentiment of Opinions." Proceedings of the 20<sup>th</sup> International Conference on Computational Linguistics, USA, pp. 1367-1373, 2004.
- [9] A. M. Popescu and O. Etzioni, "Extracting Product Features and Opinions from Reviews," Conference on Human Language Technology and Empirical Methods in Natural Language Processing, British Columbia, pp. 339-346, 2005.
- [10] T. Wilson, J. Wiebe, and Paul Hoffmann, "Recognizing Contextual Polarity in Phrase-level sentiment analysis," Proceedings of the conference on human language technology and empirical methods in natural language processing, USA, pp. 347-354, 2005.
- [11] Pankaj Gupta, Ritu Tiwari, Nirmal Robert, "Sentiment analysis and text summarization of online reviews: A survey", 2016 International Conference on Communication and Signal Processing (ICCSPP)
- [12] Sayali Zirpe, Bela Joglekar, "Polarity shift detection approaches in sentiment analysis: A survey", 2017 International Conference on Inventive Systems and Control (ICISC)
- [13] Mondher Bouazizi, Tomoaki Ohtsuki, "Multi-class sentiment analysis on twitter: Classification performance and challenges", Big Data Mining and Analytics, 2019, vol. 2, issue. 3
- [14] M. Bouazizi and T. Ohtsuki, "A Pattern-Based Approach for Sarcasm Detection on Twitter," IEEE Access, vol. 4, pp. 5477 – 5488, 2016.
- [15] Eissa M. Alshari, A. Azman, S. Doraisamy, N. Mustapha, M. Alkeshr, "Effective Method for Sentiment Lexical Dictionary Enrichment Based on Word2Vec for Sentiment Analysis", 2018, Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)
- [16] Annet John, Anice John, Reshma Sheik, "Context Deployed Sentiment Analysis Using Hybrid Lexicon", 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)
- [17] Soonh Taj, Baby Bakhtawer Shaikh, Areej Fatemah Meghji, "Sentiment Analysis of News Articles: A Lexicon based Approach", 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)
- [18] S. Ahmed and A. Danti, "Effective Sentimental Analysis and Opinion Mining of Web Reviews Using Rule Based Classifiers," Advances in Intelligent Systems and Computing, vol. 410, pp. 171-179, 2016.
- [19] M. Tsytsarau and T. Palpanas, "Managing Diverse Sentiments at Large Scale," IEEE Transactions on Knowledge and Data Engineering, vol. 28, issue 11, pp. 3028 – 3040, 2016.
- [20] Ankita, Nabizath Saleena, "An Ensemble Classification System for Twitter Sentiment Analysis", 2018, International Conference on Computational Intelligence and Data Science (ICCIDS 2018) Procedia Computer Science 132, pp-937-946

[21] Benwang Sun, Fang Tian, Li Liang, “Tibetan Micro-Blog Sentiment Analysis Based on Mixed Deep Learning”, 2018 International Conference on Audio, Language and Image Processing (ICALIP)

[22] Chae Won Park,DaeRyongSeo, “Sentiment analysis of Twitter corpus related to artificial intelligence assistants”, 2018 5th International Conference on Industrial Engineering and Applications (ICIEA)

[23] MukeshYadav, VarunakshiBhojane, “Semi-Supervised Mix-Hindi Sentiment Analysis using Neural Network”, 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)