

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/328528135>

# Data Analysis: Opinion Mining and Sentiment Analysis of Opinionated Unstructured Data: Second International Conference, ICACDS 2018, Dehradun, India, April 20–21, 2018, Revised Sel...

Chapter · April 2018

DOI: 10.1007/978-981-13-1813-9\_25

CITATIONS

2

2 authors:



**Harshi Garg**

Mody University of Science and Technology

1 PUBLICATION 2 CITATIONS

[SEE PROFILE](#)

READS

193



**Niranjan Lal**

Mody University of Science and Technology

18 PUBLICATIONS 121 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



heterogeneous [View project](#)



Opinion Mining and Sentiment Analysis of Heterogeneous data [View project](#)



# Data Analysis: Opinion Mining and Sentiment Analysis of Opinionated Unstructured Data

Harshi Garg<sup>(✉)</sup> and Niranjan Lal

Mody University of Science and Technology, Lakshmangarh,  
Sikar, Rajasthan, India

harshil23har@gmail.com, niranjan\_verma51@yahoo.com

**Abstract.** With the evolution of technology, there is also a huge increase in unstructured data. Now a day's social media is an obvious source of current opinions and reviews and to extract the valuable suggestions on the basis of comments and opinions given on social network is very important. This paper includes data analysis and data mining with a special emphasis on Opinion mining and sentiment analysis. It is one of the most vigorous research areas in natural language processing and is also widely studied in areas like Web mining, and text mining. This paper tackles a comprehensive overview of last update in this field. For the very first time in human history, we now have a huge volume of opinionated data recorded in digital form for analysis. Users not only use the resources but also give their opinions and suggestions in the form of feedback for the improvement of the existing system. It is very important to analyze their opinions and to extract the valuable opinions and suggestions from the comments. This paper proposed algorithms of machine learning and lexicon based approaches to address the issue arised due to lack of analyzing tools.

**Keywords:** Opinion mining · Sentiment analysis · SVM · Naïve bayes  
Lexicon-based · Machine learning

## 1 Introduction

In the present digital scenario, there is a vast progress and development of the web and online world wide technologies, due to this we aspect a huge volume of data and information present from many different resources services and sites which were not available to people just a few years ago. Data Analysis is the study of analyzing data and deriving business results from it and data mining discovers patterns in large data sets.

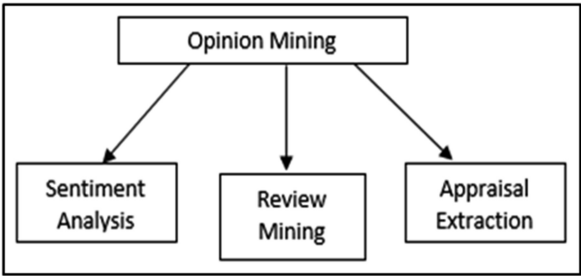
With the increase in questionnaire to get feedback about people's opinion in order to improve decision making, How to extract the valuable information from comments and suggestions, become the demand in many areas, and thus the opinion analysis for the data mining become a research focus in the field of data mining. Opinion Analysis and Sentiment Analysis is a method of automatic extraction of key knowledge from the opinion of others about some product review or problem. It is a computational mechanism that classifies the user's opinions either positive, negative or neutral comments and quotes underlying the text. This task is technically very challenging and practically very useful.

Various algorithms in recent years are proposed to address the issue such as we can develop opinion classification system using Maximum Entropy (ME) and K-Means Clustering. Various data-driven techniques like Naïve Byes, HMM Pos Tagger and SVM can also be used [6]. To classify the sentiments we can use two approaches either machine learning or lexicon based approaches. A system is also proposed [9] which extracts opinions and aspects of products and it is a two stage process, the very first is knowledge extraction and sentiment analysis. In the first stage, NLP tools are used which extracts the syntactic knowledge and then opinions aspect relations are implied. Knowledge from the first stage is used to analyze new reviews and then summary is generated.

We have organized this paper as follows. Section 2 defines Opinion Mining and its synonyms used, Sect. 3 explains the previous work done in searching and ranking, Sect. 4 illustrates the classification approaches present, Sect. 5 explains the architecture of system, Sect. 6 explains the application areas, Sect. 7 explains the tools used in this field, Sect. 8 explains the research challenges arised, Sect. 9 conclude the paper with future directions.

## 2 Opinion Mining and Sentiment Analysis

Opinion Mining is a kind of natural language processing which helps in evaluating “What other people think”? Opinion mining deals with the opinion of the text while sentiment analysis score the opinions either positive, negative or neutral. It is well suited to various intelligence applications to improve their decision making process (Fig. 1).



**Fig. 1.** Synonyms for opinion mining

## 3 Related Work

Pang and Lee [1] illustrated that availability of huge resources such as blogs, articles and information on web makes it necessary for us to deal with the computational usage of opinion, sentiments and subjectivity of words in text. They include wider issues related to manipulation, privacy and economic effect that the development of opinion-oriented information gives rise to access services based on review of people.

Haseena Rahmath [2] illustrated that people's opinion are very valuable for decision making. Due to the increased popularity of sites like FaceBook, twitter resulted huge collection of data in an unstructured manner. This give rise to emerging field opinion mining so that valuable opinions can be extracted from feedbacks. Various challenges and research applications like detection of fake reviews and arrogant words are discussed by author.

Kalarani et al. [3] illustrated that opinions reflect person's perspective, understanding, desires and attitudes. Several websites on web encourages users to express their feedbacks related to product, suggestions, comments related to product, policies and services. Extraction of useful opinions from these sources is a challenging task. Author discussed various challenges related to opinion mining.

Mago [4] illustrated that researchers aims to distinguish the opinions as positive or negative and summarize them in such a way which is easily understood by decision makers. Opinion Mining can be used for information retrieval and web data analysis. Author discussed various approaches which can be used to identify the sentiments and can be used for classify the polarity of opinions.

Songpan [7] illustrated that opinions can be open and close opinion. The open opinion refers the comment which shows emotions directly from consumer. But issue arises when customer give rating contrast along with comments. Author proposes the study and prediction rating related to customer reviews using probability's classifier model.

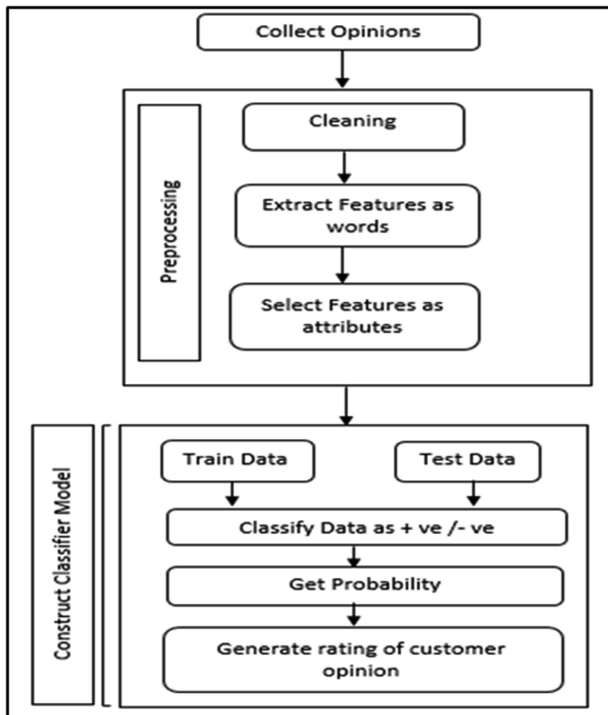


Fig. 2. Common methodology used for opinion mining

Guellil et al. [9] illustrated a survey in the form of table which includes all the research work done in the field of opinion mining and sentiment analysis. They considered 60 research papers and compare the work among various criterias.

Osimo et al. [10] presents the research challenges arised in this field. GC1: Model based collaborative Governance and GC2: Data Powered Collective Intelligence and Action (Fig. 2).

## 4 Sentiment Analysis Classification Approaches

The goal of Sentiment Classification (SC) is to discover opinions, identify the sentiments and classify the polarity of opinions accordingly. Sentiment analysis classification is illustrated in the below Fig. 3.

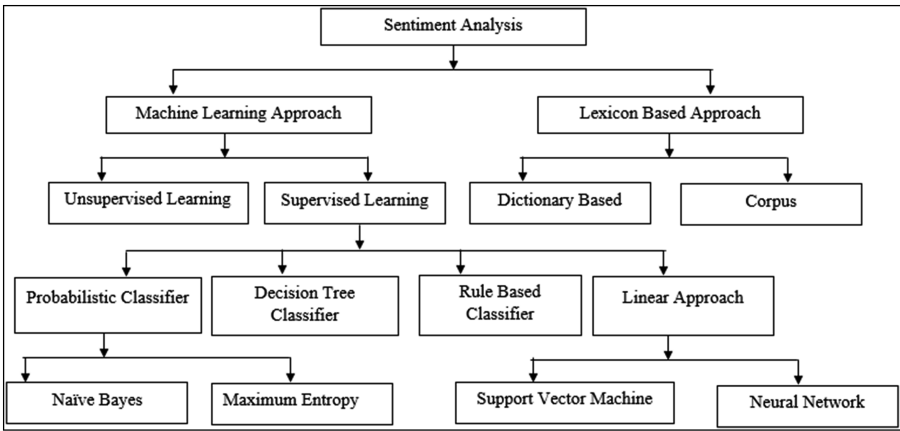
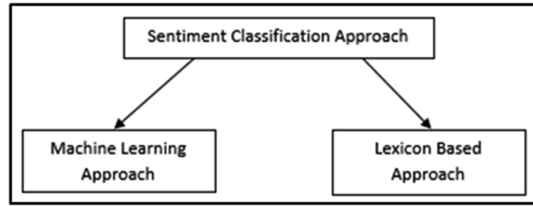


Fig. 3. Sentiment classification categories

There are three main classification levels exists:

- **Document-Level:** It classifies an opinion into positive and negative opinion and considers the whole text as an elementary information unit.
- **Sentence-Level:** It classifies sentiment in each sentence however there exist no fundamental difference between document and sentence level because sentences are just short text documents available.
- **Aspect Level:** It classifies sentiments with respect to specific aspects of entities [7].

Sentiment Classification (SC) techniques is based on two approaches (Fig. 4):



**Fig. 4.** Sentiment classification approach

#### 4.1 Machine Learning Approach

In this mainly SVM which is used to classify text as either positive or negative and Naïve Bayes classifiers which is used to classify sentiment and this sentiment orientation performs well with more accuracy are used to classify sentiments and survey stated that SVM performs better than Naïve Bayes [6].

##### 4.1.1 Naïve Bayes Classifier(NB)

This is the most frequently used classifier which computes the posterior probability of a class, this is based on the distribution of the words in the document. This classifier uses Bayes theorem to calculate the probability that a given feature set belongs to a particular label.

$$P(\text{label}|\text{features}) = \frac{P(\text{label}) * P(\text{features}|\text{label})}{P(\text{features})} \quad (1)$$

where  $P(\text{label})$  is the prior probability of a label.  $P(\text{features}|\text{label})$  is the prior probability that a given feature set is being classified as a label.  $P(\text{features})$  is the prior probability that a given feature set is occurred.

##### 4.1.2 Support Vector Machine Classifier (SVM)

It's principle is to determine linear separators in the search space which can best separate different classes. SVM can construct a nonlinear decision surface in the original feature space by mapping the data objects non-linearly to an inner product space where the classes can be separated linearly with a hyperplane. A separating hyper plane is written as:

$$W * X + b = 0 \quad (2)$$

where  $W = \{w_1, w_2, w_3, \dots, w_n\}$

$w_n$  is defined as weight vector of  $n$  attributes and  $b$  is defined as bias.

#### 4.2 Lexicon Based Approach

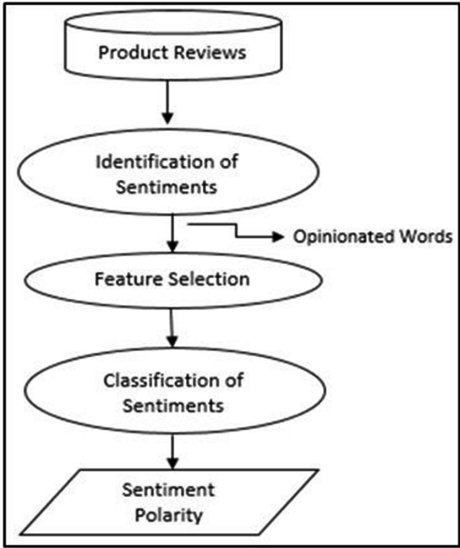
In this Semantic orientation (SO) of expressions is determined as positive if it is more related to “best” and is considered to negative if it is more related to “poor”. So SO values are dependent on calculation result which is calculated by taking average of SO value of all expressions which has been extracted (Table 1).

**Table 1.** Comparison between lexicon-based and machine learning

Parameters	Machine Learning	Lexicon-Based										
Classification Approach	Supervised	Unsupervised										
Domain	Dependent	Independent										
Statistical Significance	More	Less (Small Dataset)										
Require prior training of Dataset	Yes	No										
Adaptive Learning	Yes	No										
Accuracy	High	Low, Depends on Lexical Resources.										
		<table><tr><th>Resource</th><th>Coverage</th></tr><tr><td>SentiWordNet</td><td>117,659</td></tr><tr><td>WordNet Affect</td><td>200</td></tr><tr><td>SenticNet</td><td>14,000</td></tr><tr><td>MPQA</td><td>8,222</td></tr></table>	Resource	Coverage	SentiWordNet	117,659	WordNet Affect	200	SenticNet	14,000	MPQA	8,222
		Resource	Coverage									
		SentiWordNet	117,659									
		WordNet Affect	200									
		SenticNet	14,000									
MPQA	8,222											
Sensitive to quality and quantity of data	More	Little										
Time of Result Generation	Slow	Fast										
Maintenance	Not Required	Need maintenance of corpus										
Training Require	Yes	No										

## 5 Proposed Architecture of Opinion Mining

Proposed architecture is shown in Fig. 5 involves three key steps: the very first step is Opinion Retrieval then Opinion Classification and last Opinion Summarization. Opinions are retrieved from several review websites and many sources which can be



**Fig. 5.** Sentiment analysis process

either positive or negative review. At the end a summary is generated by considering it's frequent appearing features (Fig. 6).

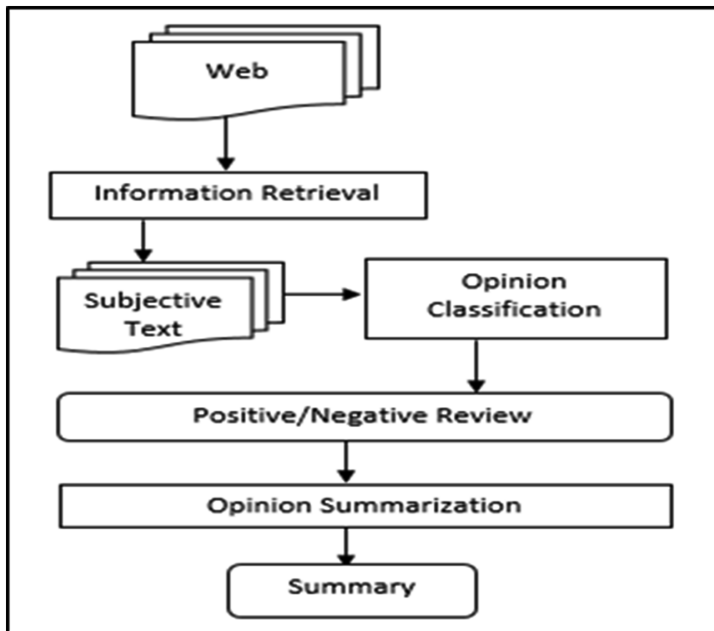


Fig. 6. Architecture of opinion mining

## 6 Applications of Opinion Mining

Different application are present in opinion mining:

- **Decision Making:** Opinion extraction plays a very important role in decision making. It gives an analyzed user's opinion which helps in improving the decision system.
- **Business Intelligence:** Manufacturers need to focus more on consumer reviews so that they can improve the product based on consumer feedback.
- **Junk-email Identification:** With the increase in amount of content, users intentionally upload junk contents or we can say spam contents so it is necessary to distinguish junk and authentic content.
- **Buying a Service or Commodity:** Valuable reviews extracted from opinions helps in taking right decision about the product that whether to buy it or not from a variety of options (Figs. 7, 8).



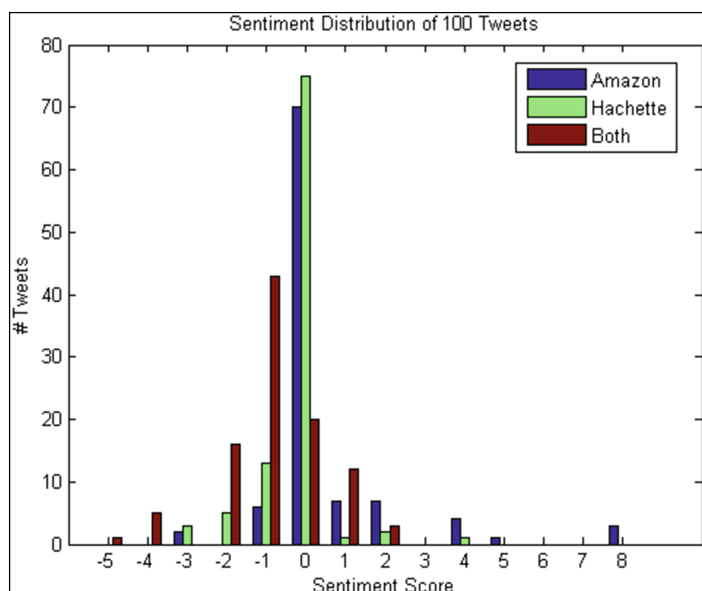


Fig. 7. Sentiment distribution graph of 100 tweets

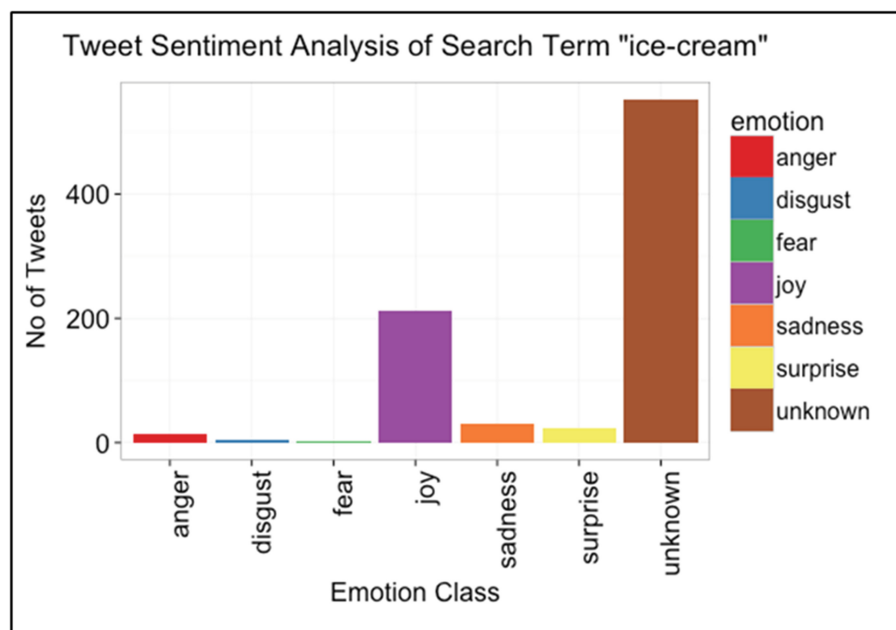


Fig. 8. Customer sentiments of search IceCream

## 7 Tools Used in Opinion Mining

Different tools can be used for opinion extraction and these are:

- **Opinion Observer:** This tool analyze and compare opinions on the web using user generated contents and suggestions.
- **Social Mention:** This tool identify keywords in events, comments, news, blogs, videos and even audio media.
- **OpenNLP:** This tool perform NLP tasks such as POS tagging and entity extraction.
- **NTLK:** This tool is used for text processing, parsing, classification and tokenization.
- **Ling Pipe:** This tool is used for clustering classification and linguistic processing.

## 8 Research Challenges

There are many research challenges arises in this field, on the basis of our research study we analyze following issues:

- Major challenge arise in handling ambiguity in NLP as user may use correct semantics or may not use correct syntax [4].
- Some linguistic issue arise in opinion mining as language is not necessarily always English.
- Another challenge is the cost of tools which can only be afforded by large organizations and government funded groups.
- Another challenge is the domain dependency of words. One feature set may give good performance in one domain and poor in another.
- There is an asymmetry in the availability of opinion mining tools.

## 9 Conclusion

Opinion Mining is a vast research area and an emerging field of data mining which extract key knowledge from a huge volume of data available that may be about consumer reviews, feedbacks, suggestions, product review. It contains wide variety of tools and techniques to extract the data. Opinion mining is important for decision making, companies want to know what people think about their product so that they can improve the product based on their feedbacks and individuals also want to know about the summarized opinions of product when they want to buy it. This paper illustrates the algorithms used in the past to address the problem arises due to lack of analyzing opinionated data and also discuss two approaches used to extract valuable suggestions. Machine Learning approach is better as this approach have higher precision but slightly lacks in result generation as comparative to lexicon based approach.

This will establish a better customer relationship by providing them what they actually need. Several challenges are also exist in this field as the vocabulary of natural

language is very large and these need to be tackled by using some solutions to improve the methods of sentiment analysis and classification.

## References

1. Pang, B., Lee, L.: Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.* **2**, 1–135 (2008)
2. Haseena Rahmath, P.: Opinion mining and sentiment analysis—challenges and applications. *Int. J. Appl. Innov. Eng. Manag.* **3**(5), 1–3 (2014)
3. Kalarani, P., Selva Brunda, S.: An overview on research challenges in opinion mining and sentiment analysis. *Int. J. Innov. Res. Comput. Commun. Eng.* **3**(10), 1–6 (2015)
4. Mago, N.: Opinion mining: applications, techniques, tools, challenges and future trends of sentiment analysis. *Int. J. Comput. Eng. Appl.* **10**(4), 1–10 (2016)
5. Vaghela, V.B., Jadav, B.M.: Analysis of various sentiment classification techniques. *Int. J. Comput. Appl.* **140**, 1–6 (2016)
6. Hamzah, A., Widyastuti, N.: Opinion classification using maximum entropy and K-means clustering. In: *International Conference on Information, Communication Technology and System*, IEEE International Conference, pp. 162–166 (2016)
7. Songpan, W.: The analysis and prediction of customer review rating using opinion mining. In: *IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, pp. 71–77 (2017)
8. Hailong, Z., Wenyan, G., Bo, J.: Machine learning and lexicon based methods for sentiment classification: a survey. In: *IEEE International Conference on Web Information System and Application (WISA)*, pp. 262–265 (2014)
9. Guellil, I., Boukhalfa, K.: Social big data mining: a survey focused on opinion mining and sentiments analysis. In: *12th International Symposium on Programming and Systems (ISPS)*. IEEE (2015)
10. Osimo, D., Mureddu, F.: Research challenge on opinion mining and sentiment analysis. In: *The CROSSROAD Roadmap on ICT for Governance and Policy Modeling* (2010)
11. Vo, A.-D., Nguyen, Q.-P., Ock, C.-Y.: Opinion—aspect relations in cognizing customer feelings via reviews. In: *IEEE*, pp. 5415–5426 (2018)