

# A STUDY ON RANKING BASED ON REVIEW MINING

Vipasha Jain<sup>1</sup>, Dr. Manvi Siwach<sup>2</sup>  
PG Student<sup>1</sup>, Assistant Professor<sup>2</sup>

Department of Information Technology  
YMCA University of Science And Technology, Faridabad, India

## Abstract

Ranking is a relationship between an arrangement of things such that, for any two things, the first is either 'positioned higher than', 'positioned lower than' or 'positioned equivalent to' the second.

Review Mining is one of the current branches of data mining. Review is a combination of words, and sentences. Review mining depends on the reviews of different clients. Review mining is utilized to classify each review as positive or negative. Review mining is known as a kind of common dialect preparing if there should arise an occurrence of recognizing the mind-set of the general population about a specific item. Review mining and sentiment analysis have been used to cover a wide extent of uses.

As there are many customer reviews available on the Internet and going through all the reviews is not possible because the number of reviews can be large or can be lengthy which can consume most of the time of the user. So subsequently, it is vital to reflect reviews in a ranking to settle on a decision effortlessly.

This survey has explored various papers and research-work which examines the suggestion strategy in light of opinion mining using to prescribe top ranking to the purchasers.

**Keywords** - naïve bayes, decision trees, recommendation, opinion mining framework, customer reviews, ranking.

## I. INTRODUCTION

Review mining is additionally called sentiment analysis since it includes building a framework to arrange reviews about an item. Essentially, review mining depends on the surveys of the clients and is utilized to arrange every review as positive or negative. To classify each review as positive or negative 'Bag of Keywords' is constructed which contains the adjectives extracted from the reviews such as good, waste etc. These keywords are partitioned into two sections on the basis of their semantic orientation as positive or negative. In view of keyword orientation, review orientation will be computed.

These days, an organization give a business benefit which needs to get reviews from the client. With the fast expansion of organization, they have more services on the web to enhance consumer satisfaction/loyalty. The supplier will read customer reviews and different clients who need to utilize services will read reviews to express opinions on the services. The quantity of customer reviews is expanding from websites, web journals, discussions and online networking. Therefore, many customers will read the reviews arbitrarily because it is hard to read all reviews and make decision on the services. If client reads a couple of reviews, client may get opinion review to be bias. Therefore, the sentiment mining is computing score automatically which can be inclined to their opinion to judge the remarks as negative and positive.

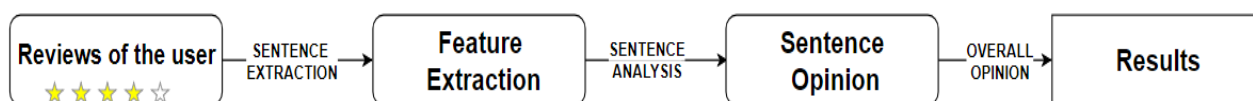


Fig 1- Process of Opinion Mining

## II. REVIEW/OPINION MINING FRAMEWORK

Review/Opinion mining cover an extensive variety of application. Argument mapping software sorts out information intelligently in a logical way around these strategy explanations by directing logical connections between them. One of the significant application of opinion mining is in Recommendation Systems. It arranges the individual's opinion into positive and negative and the framework tells which one should get suggested and the one that could be neglected. Individual's opinions and experience are extremely helpful component in basic decision making process. Opinion mining gives individual's opinion that can be adequately used as a part of decision making process.

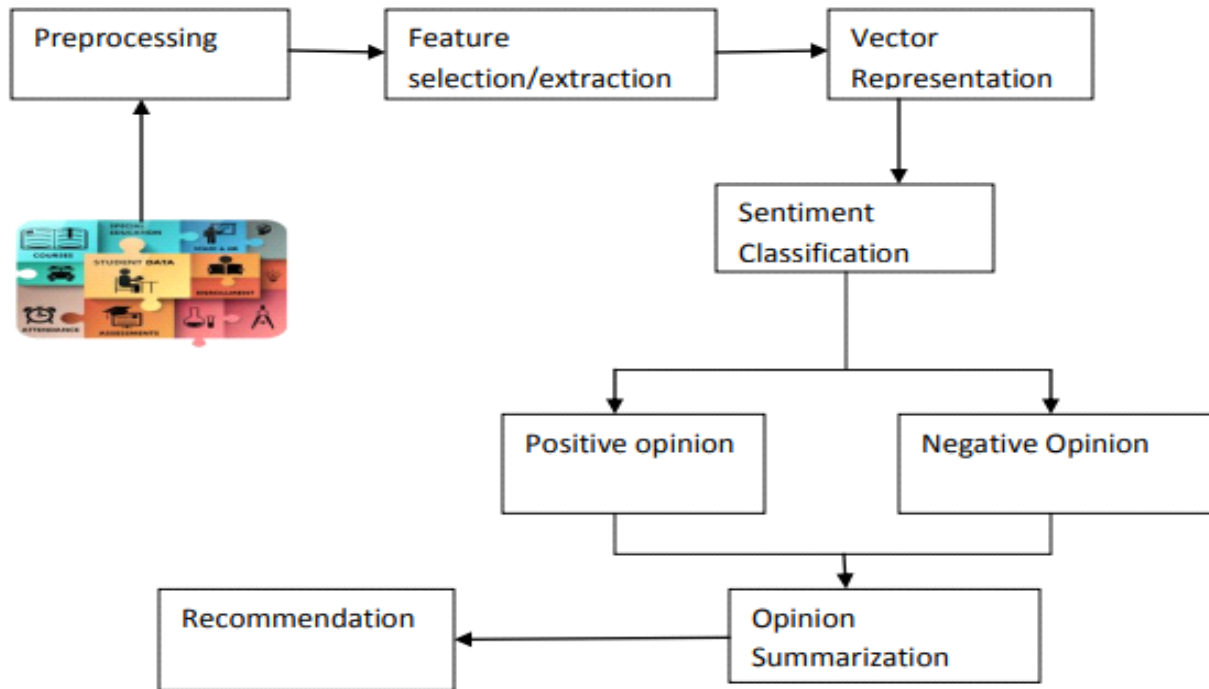


Fig 2- Opinion mining framework

The above representation gives the entire framework of opinion mining feature extraction.

### PHASES

**A. Pre-processing (Phase I):** In this phase, raw data taken and is preprocessed for feature extraction. The preprocessing phase has been further divided into number of sub phases as follows:

- 1: Tokenization- Text document has a collection of sentences which is split up into terms or tokens by removing white spaces, commas and other symbols etc.
- 2: Stop word Removal - removes articles like “a, an, the”.
- 3: Stemming - decreases relevant tokens into a single type. E.g. generalization, generally are represented as general.
- 4: Case Normalization - English texts published contains both higher and lowercase characters and this process turns the entire document or sentences into lowercase/uppercase.

**B. Feature Extraction (Phase II):** The feature extraction phase deals with feature types, feature selection, feature weighting mechanism, and reduction mechanisms.

**C. Vector Representation (Phase III):** The information collected from the reviews can be classified in one of the following ways to provide a good recommendation system.

#### Document Sentiment Classification

Supervised machine learning approaches are used for predicting the overall sentiment of the document. The review document is taken as a whole and is trained with the labeled samples. Finally the document is labeled as either positive or negative as a whole.

Sentence Level Sentiment Classification

The sentence level documents are just short documents, where the documents obtained from reviews is parsed into sentences. The sentences containing opinion words are extracted and further they are classified into subjective and objective sentences.

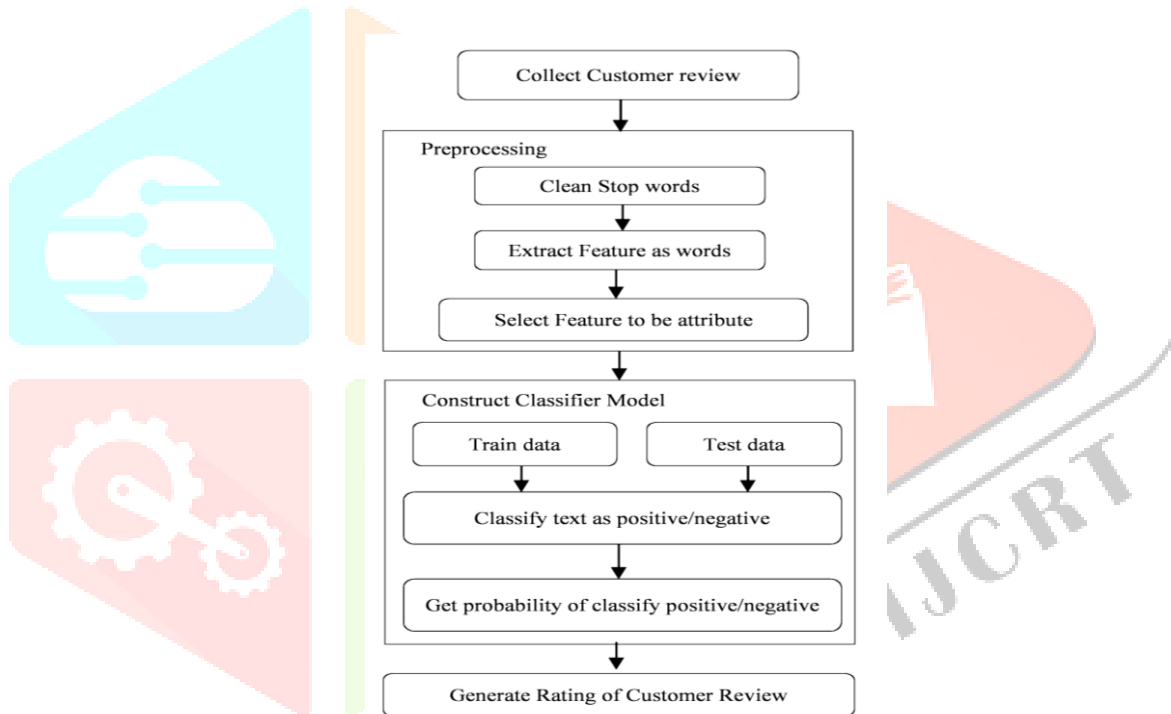
Word or Phrase Sentiment Classification

Here, the word level consolidation of sentiments has been done. The words used are mostly adjectives or adverbs that have semantic orientation which classifies the given word into positive, negative and neutral classes.

**D. Sentiment Classification (Phase IV):** As a part of sentiment analysis, sentiment classification tries to classify the nature of document/sentence using machine learning and lexicon based approaches.

**E. Opinion Summarization (Phase V):** Opinion Summarization is proposed to present extracted information components in a graphical form that facilitates users to have a quick view of a product features and users' sentiments expressed over them, without reading the pile of review documents.

**III. SCORE GENERATION OF CUSTOMER REVIEWS USING OPINION MINING**



**Fig 3- Score generation using opinion mining**

**1. Pre-processing:** The process is started from collected data and pre-processing is cleaned data by removal stop words and using the high frequency of word which will be selected into attribute for using classifier model. The classifier model will be solve the text of customer review that is positive of negative from training data and test data. Some of these keywords are listed in Table 1

**Table 1. Bag of Keywords**

Positive Keywords	Negative Keywords
Good, excellent, delight, very good, awesome, cheap, reliable, high speed, faster, superior.	Bad, worst, awful, extremely bad, costly, fake, expensive, not, slow, slower, inferior.

**2. Classifier Model:** The classifier models are used 2 models which are decision Tree (C4.5) and Naïve Bayes to classify texts as class labels: positive or negative. Each data set is trained to model and test model that given predicted class labels follows probability trending of classifier model. The classifier models are described as below:

• Decision Tree

The decision tree learning was proposed as a model of data classification for a class label. Decision tree is clearly represented through a tree diagram. It starts from the first node as a root node. The root node selects an attribute as words in opinion from the best value of measurement. Each attribute has its own values i.e. true/false, which are separated by branch links composed of original attributes. At the end, the data reveals a class which represents a leaf node (i.e. positive/negative). The advantage of the decision tree is for ordering attributes that are the best measurement as Eq. (1).

$$I(s_1, s_2, \dots, s_n) = -\sum_{i=1}^n \frac{s_i}{s} \log_2 \frac{s_i}{s} \quad (1)$$

where,

$n$  is the number of class label.

$S_i/s$  the number of data  $S_i$  of class  $i$

After the distinguished information of attribute is calculated, the entropy value is also calculated to define the summary of each branch needed be clearly separated from attribute A as Eq. (2)

$$E(A) = \sum_{j=1}^m \frac{s_{1j} + \dots + s_{nj}}{s} I(S_{1j}, \dots, S_{2j}) \quad (2)$$

where,

$m$  is the number of branch of attribute A.

The highest gained value of the attribute A results in the best attribute to classify data set which is calculated and range between 0 and 1 by Eq. (3)

$$Gain(A) = I(s_1, s_2, \dots, s_n) - E(A) \quad (3)$$

• Naïve Bayes

Naïve Bayes is an algorithm of probability based on Bayes theorem of learning. It aims to create a model in the form of probability. The advantage of Naïve Bayes is an effective method which is easy processing. A few examples where this algorithm used are spam detection, sentiment analysis, classifying news articles. The algorithm makes assumption that occurrence of a certain feature is independent of the occurrence of the other features present in the list. For example, to identify the fruit based on the color, shape, taste then an orange color, spherical shape and tangy flavor fruit is most likely to be an orange. With Naive Bayes we can build the model fast and make quick prediction. The range of probability is between 0 and 1 as Eq. (4).

$$V_{NB} = \arg \max P(v_j) * \prod_{i=1}^n P(a_i | v_j) \quad (4)$$

• Evaluation Model

The evaluation model is used k-fold cross validation with test data which are generated all training data. The k defines the number of grouping data. For example, k is 10-fold cross validation of 400 training data, means each group as 40 records and 10 groups, whereas the testing data will be groups 1 of 40 records and evaluation this groups to calculate average of the accuracy collected until N as 10 groups,

$$Accuracy = \sum_{i=1}^{fold} \sum_{j=1}^{fold} \frac{\delta_{ij}}{N} \quad (5)$$

where,

$$\delta_{ij} \begin{cases} 1 = \text{predicted class label is correct} \\ 0 = \text{predicted class label is incorrect} \end{cases}$$

In addition, the results is evaluation by rating, the root mean square error is used in this case. The comparison results are generated rating with classifier model and rating from actual customer review as Eq. (6).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (P_i - O_i)^2} \quad (6)$$

where,

$P_i$  is prediction from probability value of classifier model.

$O_i$  is actual score from customer review.

#### IV. LITERATURE REVIEW

**1. Gnana Kirubaharan** introduced a ranking system which helps the consumers to save time and get a precise ranking scale from 1 to 5 based on the extraction of reviews and processing it by sentiment classification where each and every word from the numerous number of review is analysed and given positive and negative values to each words and then the overall ranking is given to the consumer.

**2. T.Sangeetha** used the aspect ranking method which identifies the key aspects of product from customer reviews.

**3. Blety Babu Alengadan** developed an aspect based opinion mining model that can identify opinionated sentences from huge data set of reviews with a high average precision and can classify the polarity of the reviews with a good average accuracy in comparison to the existing models and algorithms.

**4. Jordan Rodak** used the data set for electronic products sold through Amazon, for which there were over million electronics reviews.

**5. P. Venkata Rajeev and V. Smrithi Rekha** have shown work on opinion mining of online customer reviews of mobiles and tablets. The aim of this proposed system is to help the user to select the best product he needs.

**6. Wararat Songpan** analysed and predicted the customer review rating using opinion mining. They used customer reviews hotels from a website of hotel agent service, which service in hotel reservation directly.

**7. Mohamad Syahrul Mubarak** focused on aspect-based sentiment analysis which tries to find an aspect that is being discussed in an opinion and its sentiment polarity.

**8. Yoichi Saito and Vitaly Klyuev** proposed ranking products according to opinions of users. This ranking is based on the rates of positive sentences and negative ones. The number of positive words and negative ones are investigated to decide the polarity of a sentence.

**Table 2 . Comparison Table**

Technique	Proposed Work	Methodology	Limitations
1. Gnana Kirubaharan	Introduced a ranking system which helps the consumers to save time and get a precise ranking scale from 1 to 5 based on the extraction of reviews and processing it by sentiment classification where each and every word from the numerous number of review is analysed and given positive and negative values to each words and then the overall ranking is given to the consumer.	To acquire the ranking of a product, the reviews and specification are extracted using stream clustering method.	Going through all the review of a given product can be exhausting and time consuming at times.
2. T.Sangeetha	Used the aspect ranking method which identifies the key aspects of product from customer reviews.	This Framework encloses four main constituent, i.e., Aspects extractor, Aspect grouping, Sentiment score prediction and Aspects ranking based on aspect dictionary and opinion. It helps us to extract the important aspects of the products to be considered before buying that	The performance can be improved by using machine learning algorithms.

		particular product.	
3. Blety Babu Alengadan	Developed an aspect based opinion mining model that can identify opinionated sentences from huge data set of reviews with a high average precision and can classify the polarity of the reviews with a good average accuracy in comparison to the existing models and algorithms.	The ranking of products is done based on important aspects. If the product has equal number of pros and cons, it is identified together as a final result. So it can make the consumer to choose the product based on other aspects.	Modifications can be done on interpretation of aspects.
4. Jordan Rodak	Used the data set for electronic products sold through Amazon, for which there were over million electronics reviews.	The classifiers used are Support Vector Machine, Naïve Bayes classifier with linear, sigmoid, radial basis function, and polynomial kernels. These classifiers are very powerful and helped model the data well.	Linear SVM has achieved the highest accuracy at certain points but it was unreliable and as such it would not be recommended for the task at hand. Naive Bayes is definitely helpful for very small sample sizes, and while it does not attain the accuracy of a polynomial SVM, it gets pretty close.
5. P. Venkata Rajeev and V. Smrithi Rekha	Shown work on opinion mining of online customer reviews of mobiles and tablets. The aim of this proposed system is to help the user to select the best product he needs.	Ranking of products, ranking of products based on features, comparison of accuracy of algorithms like Naive Bayes classifier, maximum entropy classifier, and decision tree classifier has been shown. All the results are shown on a python GUI.	The number of product categories, the number of products in each product category, and the number of websites to collect the customer reviews can be increased.
6. Wararat Songpan	Analysed and predicted the customer review rating using opinion mining. They used customer reviews hotels from a website of hotel agent service, which service in hotel reservation directly.	A model is compared between decision Tree and naive Bayes. The advantage of the decision tree shown the factors ordered by level of tree to help analysing service improvement factors. Naive Bayes model is able to use probability which is similar value rating, which the system is computing automatically.	Data pre-processing can be done to extract words from a sentence using machine learning.
7. Mohamad Syahrul Mubarak	Focused on aspect-based sentiment analysis which tries to find an aspect that is being discussed in an opinion and its sentiment polarity.	As the first step, data preprocessing is aimed to clean and prepare data for next step. The second step is feature selection in which they employed Chi Square to select a subset of relevant terms to be used in the construction of Naïve Bayes model. The last step is classification of aspect and its sentiment using Naïve Bayes.	

8. Yoichi Saito and Vitaly Klyuev	Proposed ranking products according to opinions of users. This ranking is based on the rates of positive sentences and negative ones. The number of positive words and negative ones are investigated to decide the polarity of a sentence.	In this research, some methods for sentiment classification are compared for accuracy. All methods are divided into two groups. One group is for the methods that only use review text. Another is for the methods that use user and product information. The result of this research shows that it is efficient to apply the neutral network to sentiment classification	
-----------------------------------	---	---	--

## V. PROBLEM STATEMENT

Most of the reviews are lengthy but less number of sentences will have the actual aspect/focus upon the product. This makes it difficult for customer to look through them and to create knowledgeable decision upon purchasing the products. If the customer reads only some reviews, he/she will get influenced by them. Also, the large number of reviews on various online shopping websites makes it difficult for the product maker to keep a track of customer opinion for their products.

However, it is inappropriate for customer to manually recognize the best aspect of the particular product from large reviews. For these reasons, there is a need for an approach which could automatically recognize the best aspects that are highly wanted.

## VI. CONCLUSION AND FUTURE WORK

The opinion mining of client reviews is essential to enhance benefits. The decision tree arranges the factors by level of tree in order to analyse service improvement and priority factors. However, Naive Bayes model can use probability with comparative rating value, which the framework can compute automatically. Many client will read reviews, however the framework can be condensed entire rating consistency with the reviews. In this way, the clients can make a choice quickly.

In the future work, other machine learning algorithms can be applied and opinion mining model could prove effective and one can spare a lot of time.

## REFERENCES

- [1] Shahab Saquib Sohail, Jamshed Siddiqui, Rashid Ali, "Book recommendation system using opinion mining technique," International Conference on Advances in Computing, Communications and Informatics ICACCI ),2013.
- [2] Blessy Selvam, S.Abirami, "A Survey on opinion mining framework," International Journal of Advanced Research in Computer and Communication Engineering, 2013.
- [3] P. Venkata Rajeev and V. Smrithi Rekha, "Opinion Mining of User Reviews Using Machine Learning Techniques and Ranking of Products Based on Features,"
- [4] Wararat Songpan,"The Analysis and Prediction of Customer Review Rating Using Opinion Mining," (IEEE), 2017.
- [5] Yoichi Saito and Vitaly Klyuev," Opinion Mining: Book-Ranking Application," IEEE, 2017.
- [6] Anand Shanker Tewari, Tasif Sultan Ansari, Asim Gopal Barman Opinion," Based Book Recommendation Using Naïve Bayes Classifier," IEEE, 2014.
- [7] Lina L. Dhande and Dr. Prof. Girish K. Patnaik," Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier," International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 2014
- [8] Aashutosh Bhatt, Ankit Patel, Harsh Chheda, Kiran Gawande," Amazon Review Classification and Sentiment Analysis," (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6) , 2015.
- [9] Jordan Rodak, Minna Xiao, Steven Longoria," Predicting Helpfulness Ratings of Amazon Product Reviews."
- [10] Muhammad Rafi, M.Rafay Farooq, Usama Noman, Abdul Rehman Farooq, Umair Ali Khatri, "Study of Automatic Extraction, Classification, and Ranking of Product Aspects Based on Sentiment Analysis of Reviews",

Proceedings of the World Congress on Engineering and Computer Science 2015 Vol II WCECS 2015, October 21-23, 2015, San Francisco, USA.

[11] Gnana Kirubaharan, Elumalai P. S. and Yovan Felix , “Online Product Ranking Based On Reviews,” ARPN Journal of Engineering and Applied Sciences ©2006-2017.

[12] Blety Babu Alengadan, Shamsuddin S.Khan, “A Proposed System for Modifying Aspect Based Opinion Mining for Ranking of Products,”2017 IEEE 3rd International Conference on Sensing, Signal Processing and Security (ICSSS) 978-1-5090-4929-5©2017 IEEE.

[13] T.Sangeetha, Mr.N.Balaganesh, Dr.K.Muneeswaran, “Aspects based Opinion Mining from Online Reviews for Product Recommendation”2017 International Conference on Computational Intelligence in Data Science(ICCIDS).

