# Sentiment Analysis of Online Product Based Review System

**Ankit[1], Deepak Sharma[2], Ankit Kumar[3], Deepak Panchal[4] ,*Fatima Ziya[5]***

[1]Computer Science And Engineering, KEC, Gzb

[2]Computer Science And Engineering, KEC, Gzb

[3]Computer Science And Engineering, KEC, Gzb

[4]Computer Science And Engineering, KEC, Gzb

[5] Assistant Professor, KEC, Gzb

---------------------------------------------------------***---------------------------------------------------------

**Abstract -** In the  present scenario much more population are dependent on different social platform. And due to the presence of population a huge amount of sentiment loaded data are also presented here and by using these data we have to predict their likes dislikes but this is a tricky task for the simple human. Here likes and dislikes depends on their sentiment your words that describe it in their review and by e data may be positive or negative or neutral. And these pre-processing techniques are done by using Machine Learning Algorithms. The main motive of this research paper is to explain some methods and tools which is useful in the sentiment analysis.

**Keywords:** Sentiment Analysis, Machine Learning, Tools.

## I.    INTRODUCTION

Sentiment analysis is the process of using natural language processing (NLP), Text Analysis, and Statistics for analyzing the person's sentiment [1]. The best way to understand the sentiment of a person is to get the necessary requirements and preconditioning of a particular person. Customer sentiment can be found in tweets, comments, reviews, or other places where people mention your brand. Sentiment Analysis is the technique to understand the emotion of the customer and it's a must-understand for developers and business leaders in a modern workplace.  Because of Huge hike in the use of social sites brought people in the era of E-Commerce where people can sell and buy different products[5]. And due to this huge amount of unstructured sentimental data is available on the sites. And these unstructured data may be in the form of using these review evaluation of a good or bad product depends.  Sentiment Analysis is done to show a product that how many people like a product and how valuable it is. Here when data is gathered then before performing sentiment analysis the data will be pre-processed by which the confidence and quality of data is checked i.e. Her

emails blogs which have quantity around 80% of world's data[7]. With the help of sentiment analysis techniques we try to collect information like opinion in text, emotional words by using Natural Language Processing. when any authentic user provides their opinion it also provides a new way by which selection can be done. When a purchaser got posting information then it becomes possible that the product should be purchased or not[3]. and also the opinions of old user of any product is helpful in the identification of user interest prediction and product research are also in different election polls. So there may be some positive opinion while some negative opinion and it is difficult to classify views and sentiment of people[2]. Hence due to high availability of an unstructured format of reviews and by using these reviews we have to analyse the reviewers sentiments needs special machine learning techniques. and also we can use techniques like clustering and nearest neighbour methods. here in this paper we are using data from Kaggle.com and stored in a csv file[8].

In Section II & III explanation of the related work and, the explanation of methodology and its classification is discussed. The conclusion is shown in Section IV.

## II.    RELATED WORK:

Sentiment Analysis is used to distinguish the viewpoint of reviewer which is used in the consideration and also helps in the prediction of positive and negative word[6]. Here for analysis of these types of words we perform different methods which is shown in Fig. 1.1.
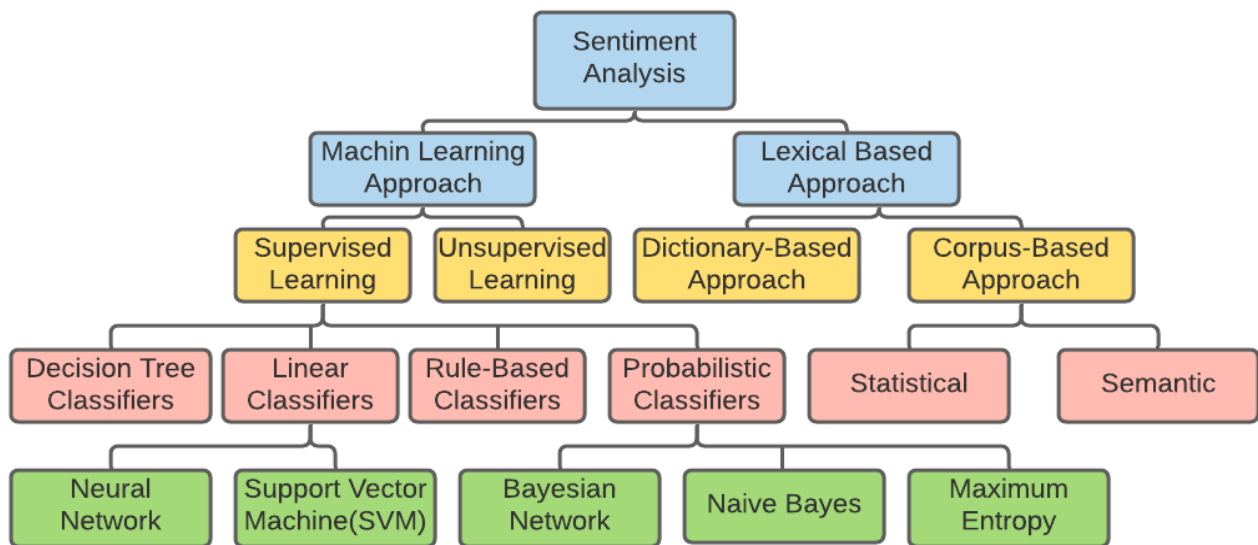


Fig. 1.1: Sentiment Analysis Techniques

### A. Machine Learning Approach

Machine learning approach is also divided as shown in fig 1.1.

a. **Supervised Learning** is machine learning approaches by which we can done analysis over a large data set[9].

b. **Unsupervised Learning** is also known as a sentiment lexicon in this method the network learn without specifying desired output[9].

### B. Lexicon Based Approach

We can further divide this approach in dictionary based approach and Corpus based approaches.

### C. Dictionary Based Approach

In this approach we called collect a few set of opinion as base then we use a we will dictionary to match these opinions word by adding its synonyms and antonyms.

### D. Corpus Based Approach

There was some limitations in dictionary based approach and those limitations can be removed by Corpus based approach this approach recognise some unique words.

### E. Support Vector Machine (SVM)

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems.

### F. Bayesian Network

A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph.

### G. Naïve Bayes

Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems. It is mainly used in *text classification* that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast

machine learning models that can make quick predictions.

The formula for Bayes' theorem is given as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

**Where**

**P(A|B) is Posterior probability**: Probability of hypothesis A on the observed event B.

**P(B|A) is Likelihood probability**: Probability of the evidence given that the probability of a hypothesis is true.

**P(A) is Prior Probability**: Probability of hypothesis before observing the evidence.

**P(B) is Marginal Probability**: Probability of Evidence.

**III.    Methodology**

**a.    Data Collection**

The data set is used from kaggle.com to train the algorithm. it contains 71045 reviews which is done for 1000 products[7].

**b.    Data Pre-processing**

The main motive of this step is to convert unstructured data in a specific format as per requirement of the system. all the text containing null value will be neglected[6].

**c.    Feature Selection**

In this section such a text will be selected by which analysis of product will be done in a flexible manner[10].

**d.    Analysis Process**

Here analysis of sentiment is done with the help of a long attributes-

- True Positive

It describes exact positive review.

- False Positive

Incorrect and fake Positive reviews.

- True Negative

Correct and Negative reviews.

- False Negative

Incorrect and Fake Negative Reviews.

**e.    Sentiment Classification**

The last process is sentiment classification and done with the help of random forest algorithm for the testing of system the source of data is kaggle.com. With the help of Kaggle firstly random data is cleaned and stored as csv. and then csv file will be used on Jupyter basically reviews is presented by five different classes. Where 0,1 used for negative reviews and 2,3 denotes neutral reviews and at last 4,5 used for positive reviews. the data set consists of Id, brand, categories, name, data updated and price[4].

Sentiment Analysis is the automated process of understanding the sentiment are opinion of a given text. you can use it to automatically analyse

| Feature | Description |
|---|---|
| Id | Product id |
| Brand | Brand of the product |
| Categories | Category based on the review |
| Name | Product name |

product reviews and sort them by positive neutral negative. when we got a negative rating then it describes the defect of a product and it is also seen that when we got positive then it will be used in reflecting the goodness of that product.

**IV.    Conclusion**

According to this paper we have performed text mining over reviews by using machine learning algorithm and here   Natural Language Processing is used to assess different sentimental text. because a reviewer reviews a product in a text format over any social media sites. and when review is seen by different provider and user then provider can improve their brand and new user feel helped after seeing these reviews[5].And for analysing things we have used Naïve Bayes algorithm which helps in the extraction of different positive and negative text. When we got these types of text we can rate them according to the text.

## REFERENCES

[1] Alexandra Cernian, Valentin Sgarciu, Bogdan Martin: Sentiment analysis from product reviews using SentiWordNet as lexical resource, 978-1-4673-6647-2/15/$31.00 ©2015 IEEE

[2] Pang, B., Lee and vaithyanatham, S., Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing- volume 10 (EMNP 02), vol. 10. Association for computational Linguistics, Pa, USA, 79-86, 2002.

[3] Zhai, Zhonghu, Hua, Bada kang, and Peifa Jia. "Exploiting effective features for Chinese sentiment classification." Expert systems with Applications, vol. 38, PP.9139-9146, 2011.

[4] Y.ko and J. seo, "Automatic text categorization by unsupervised learning" in proceedings of the 18Th conference on computational linguistics-volume 1. Associations for computational Linguistics, 2000, pp.453-459.

[5] A. otiose, J.M martin, and R.M Carro, "Sentiment analysis in Facebook and its applications to e-learning." Computers in Human Behaviour, vol.31, pp, 527-541, 2014

[6] P.Ajitha and R.Jenitha Sri "Survey of Product Reviews Using Sentiment Analyis, Indian Journal of Science and Technology, vol 9(21), pg no 1-4, 2016.

[7] Sukhchandan Randhawa, Sushma Jain and Zeenia Singla, "Sentiment Analysis of Customer Product Reviews Using Machine Learning", International Conference on Intelligent Computing and Control (12C2),volume 3, pg no 1-5, 2017.

[8] Fang Luo et al., "Affective-feature-based Sentiment Analysis using SVM Classifier", IEEE 20th International Conference on Computer Supported Cooperative Work in design 2016. International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017)

[9] Ebru Aydoan et al., "A Comprehensive Survey for Sentiment Analysis Tasks Using Machine Learning Techniques", IEEE, 2016.

[10] J. A. Konstan and J. Riedl, ''Recommender systems: From algorithms to user experience,'' User Model. User-Adapted Interact., vol. 22, nos. 1–2, pp. 101–123, Apr. 2012, doi: 10.1007/s11257-011-9112-x.

[11] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, Collaborative filtering recommender systems, in The Adaptive Web (Lecture Notes in Computer Science), vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Cham, Switzerland: Springer, 2007.