International Journal of Innovative Research in Engineering
Volume 3, Issue 2 (March-April 2022), PP: 181-184.
www.theijire.com **ISSN No: 2582-8746**

Sentiment Analysis of a Product based on User Reviews

Aziz Lokhandwala[1], Usman Merchant[2],Shamim Mirajkar[3], Adil Shaikh[4]
[1,2,3,4]*Department of Information Technology, M.H. Saboo Siddik College of Engineering, Byculla, Mumbai – 400008, India.*

How to cite this paper:
Aziz Lokhandwala,Usman Merchant,Shamim Mirajkar, Adil Shaikh, "Sentiment Analysis of a Product based on User Reviews", IJIRE-V3I02-181-184.

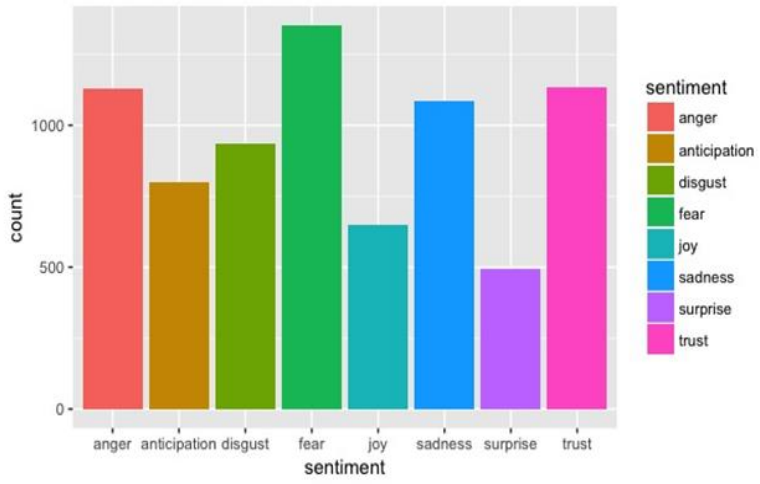
Abstract: *Sentiment analysis is one of the fastest spreading research areas in computer science, making it challenging to keep track of all the activities in the area. We present a customer feedback reviews on product, where we utilize opinion mining, text mining and sentiments, which has affected the surrounded world by changing their opinion on a specific product. We performed a comparative sentiment analysis of retrieved reviews. Sentiment Analysis is a process of analyzing and categorizing the emotion or sentiment over any given review or text piece in order to know what the reviewer wants to express in the form of positive, negative or neutral. Today, people are highly interested in buying things online from any e-commerce site or they search for a product review in order to know the quality and one's perception toward that product before buying. The product provider also gets to know about the user's opinion over a product. This can help the company to improve its marketing strategy and quality of product in their favor. Sentiment analysis uses various semantic approaches like on these online reviews to extract as much features it can and categorize the type of opinion. Some techniques also help in rating the product value based on user's opinion.*

Keywords *- Sentiment Analysis, Online Reviews, Opinion Extraction, Text Analysis, Fake Review.*

I. INTRODUCTION

Opinions are statements that reflect people's perception or sentiment. Sentiment analysis is a series of methods, techniques, and tools about detecting and extracting subjective information, such as opinion and attitudes, from language , helping in finding the mood of the customers about a purchasing of a particular product or topic . Online buying today is the most preferred and trending way for buying any product like clothes, home essentials, electronic gadgets, sport equipments, books, etc. For buying a particular thing, a user needs to be sure about the quality of the product, as only product pictures and description can be viewed online. So there is a review system for the product, reviewed by the users who bought it. They write about the product quality they received and other opinions over that product bought by the. This review helps other user buying the similar product. And these review sets can be taken from different platforms such as Amazon, Flipkart, Ebay, yelp, etc.

Suppose a review says, "This product is awesome. Good to use." It can be seen that these words indicate a positive opinion. But to mine the sentiments technically these reviews undergo assorted process and classification techniques. [1]Firstly, these reviews are gathered and proper noise removal is done in order to process the texts. Then different extraction techniques used in classification of sentiments on different types of datasets from different platforms.



The Figure 1.1 shows different sentiments like anger, fear, joy, etc based on user reviews.

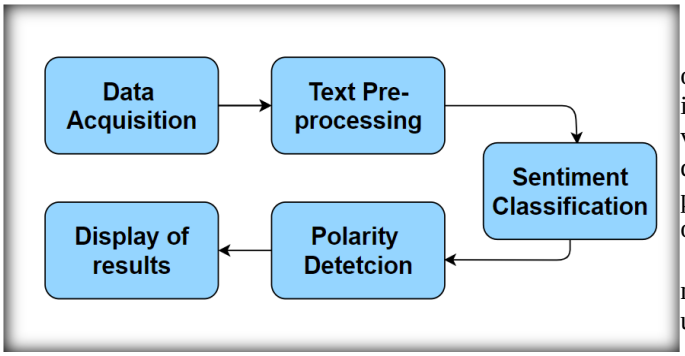
II. SENTIMENT ANALYSIS PROCESS MODEL AND TECHNIQUES

This section of the paper discusses the Sentiment Analysis Process Model and the many methodologies utilized to evaluate sentiment in it.

A. PROCESS MODEL OF SENTIMENT ANALYSIS

[13]Sentiment Analysis is a method of determining a person's feelings or opinions on a product, service, or organisation.. We will discuss each step in brief as depicted in Fig 1.

Figure 1



1. **Data Acquisition:** The acquisition or collection of data on which sentiment techniques must be done is the first step in sentiment analysis. Any online review on any product from various sites, as well as any opinion, can be included in this dataset. Without data gathering, the model will be unable to progress. Text analysis and categorization may be conducted on these datasets.

2. **Text Pre-processing:** Following data collection, data must be pre-processed, which includes the elimination of undesirable noise, emoticons, duplicate words, stemming, and various forms of URLs, among other things.

3. **Sentiment Classification:** To classify the collected feature into various sentiments, various sentiment analysis approaches are now applied. Support Vector Machine (SVM), Nave Bayes, Neural Network, and other approaches are widely utilized.

4. **Polarity Detection:** Following the sentiment classification phase, polarity detection has been used to identify whether a statement is positive, negative, or neutral.

5. 5.**Validation and Evaluation**: Lastly, the obtained result is validated and evaluated in order to verify the accuracy of the result.

B. TECHNIQUES OF SENTIMENT ANALYSIS

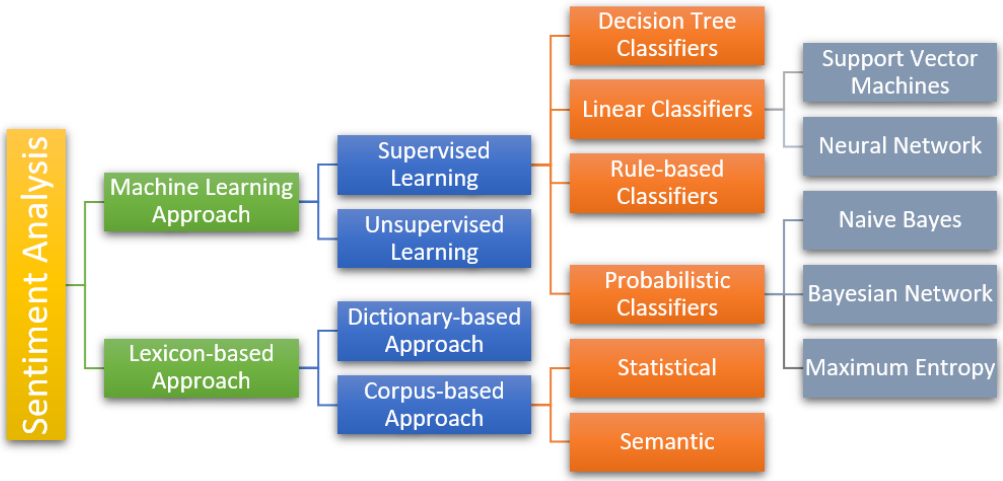
[14]A. Supervised learning

The computer is given labelled training data consisting of a set of example input and target output pairs. The goal is to teach the computer a generic rule that connects inputs and outputs. A computer's job is to learn a function that maps an input to an output from examples of input-output pairs. In supervised learning, the algorithms evaluate the training data in order to generalise a function that can be used to map fresh input samples.

B. Semi-supervised learning

In this scenario, the computer is given a training data set that is missing part of the target outputs. The computer is programmed to learn by combining a small amount of labelled data with a large amount of unlabeled data. The model's performance is improved by using all labelled data and making intelligent use of unlabeled data.

Figure 2



C. Reinforcement learning
In a dynamic environment, the computer makes decisions and takes actions, and the reward (or penalty) is supplied as feedback for its actions.The ideal policy will be discovered after multiple trials.

D. Unsupervised learning
Unlabeled data is fed into the learning algorithm, which infers a function that describes the structure of the data.
The term "unlabeled data" refers to information that has not been categorised or categorised.

E. Naive Bayes Algorithm
It is a classification method based on the Bayes theorem. It presupposes that the presence or absence of a class feature is unrelated to the presence or absence of other class features. It is the most straightforward approach for classifying enormous data sets.

F. Multinomial Naive Bayes Algorithm
Each feature has a multinomial distribution, resulting in a feature vector that represents the frequency of occurrence of that characteristic in a specific instance. It's a variant of the Naive Bayes classifier.

G. Bernoulli Naive Bayes Algorithm
Rather than probabilities, like in Multinomial Naive Bayes, the features are independent binary variables that describe the existence or absence of something.

H. Logistic Regression Algorithm
It is a binary data-based mathematical model that estimates the likelihood of occurrence of a feature. If food quality is a feature, for example, the likelihood of being good quality is 1 and the probability of being terrible quality is 0.

I. Linear SVC (Support Vector Clustering) Algorithm
It returns a hyperplane that divides the data into good and poor categories.

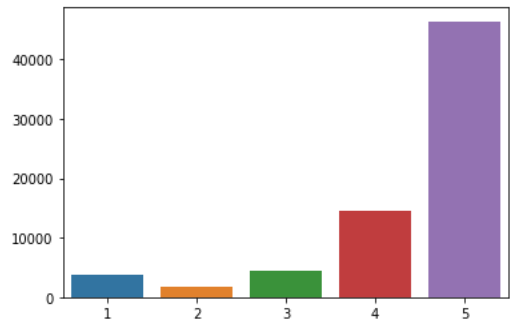
III. PROPOSED METHOD

This paper describes how to use Machine Learning algorithms to handle textual and statistical data from a dataset of different brands. It's a supervised learning application. The goal is to organise the customer reviews into graded categories that may be broadly sorted from terrible to good or negative to positive.[2]
From gathering raw data to determining whether a review is favourable or negative, and to what extent, there is a series of stages required. Tokenization, part of speech tagging, and assigning probabilities of good or bad based on recurrence to every adjective that appears in the reviews are just a few of the important steps to be taken [3].
The steps taken are

1. Data cleaning
The practise of correcting or deleting incorrect, corrupted, improperly formatted, duplicate, or incomplete data from a dataset is known as data cleaning. There are numerous ways for data to be duplicated or mislabeled when merging multiple data sources.

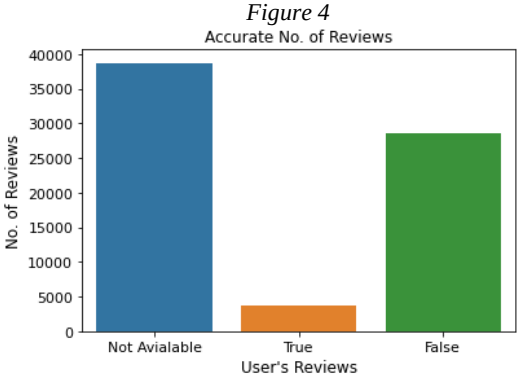
2 .Visualizing
The graphical depiction of information and data is known as data visualisation. Data visualisation tools make it easy to examine and comprehend trends, outliers, and patterns in data by employing visual elements like charts, graphs, and maps.

Figure 3



Need to map 1,2 - Negative, 3-Neutral and 4,5 - Positive

Figure 4



3. Fake review analysis
True reviews are less in number and as you can see poeple who didn't even purchase the product has provided reviews these are all fake reviews which are much higher than those who have purchased and provided the review.

4.Svc Algorithm
The objective of a Linear SVC (Support Vector Classifier) is to fit to the data provided , returning a "best fit" hyperplane that divides, or categorizes, data

IV. EXPARIMENTATION AND RESULTS

The experimentation is carried out in Python, in a machine with intel core i5 processor and 4GB RAM. The accuracy of each algorithm depends on the quality of the training and testing sets. The accuracy obtained from the algorithms and the accuracy of the proposed model is **90.007%**

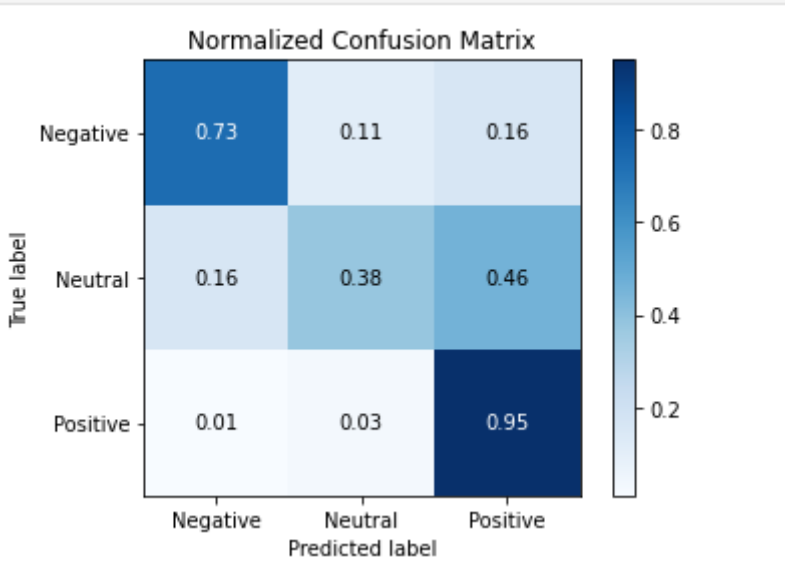
Figure 5

```
Accuracy Test : 0.9000987724001693
              precision    recall  f1-score   support

    Negative       0.75      0.73      0.74      1126
     Neutral       0.38      0.38      0.38       898
    Positive       0.95      0.95      0.95     12150

    accuracy                           0.90     14174
   macro avg       0.69      0.69      0.69     14174
weighted avg       0.90      0.90      0.90     14174
```

Figure 6



V. LIMITATIONS AND FUTURE SCOPE

Accuracy and classification diversity are two of the project's drawbacks. The training set will never be ideal because separating 20,000 reviews into excellent and bad simply by reading them is a difficult undertaking. The quality of the training and testing data sets is a critical factor in evaluating the efficacy of any algorithm. Furthermore, the classification is binary, yielding an overall conclusion of the review's goodness or badness. It is incapable of detecting sarcasm and does not take into account particular characteristics like as food, ambiance, prices, or service, to name a few. The future scope is to score businesses based on several characteristics, to train and identify sarcasm using better and larger datasets.

IV. CONCLUSION

This paper contain strategies for detecting polarity in online reviews or classifying them into various sentiments. On their various datasets, Support Vector Machine technique has delivered good and adept outcomes. SVM also produces the best results when determining the polarity of ambiguous sentiment or reviews. One of the most important and fascinating topics to research is machine learning.

Most things can be reduced to a collection of numbers and probabilistic events. The power that mathematics has given us, amplified by a computer's computation abilities, is immense, and we can only benefit from it if we learn to master it.

References

[1] A. Ethem, Introduction to Machine Learning, MIT Press. , 2010..
[2] D. Jurafsky and H. M. James, Speech and Language Processing, 2000.
[3] Q. Jiangtao, L. Chuanhui, L. Yinghong and L. Zhangxi, "Leveraging sentiment analysis at the aspects level to predict ratings of reviews," Information Sciences, vol. 451–452, p. 295–309, 2018.
[4] A. Bagheri, M. Saraee and F. de Jong, "An Unsupervised Aspect Detection Model for Sentiment Analysis of Reviews," Lecture Notes in Computer Science., vol. 7934, pp. 140-151, 2013.
[5] O. Alvaro, M. M. José and M. C. Rosa, "Sentiment analysis in Facebook and its application to e-learning," Computers in Human Behavior, vol. 31, pp. 527-541, 2014.
[6] A. Orestes, C. Francisco, C. Jenny and Hamido, "A hybrid approach to the sentiment analysis problem at the sentence level," Knowledge-Based Systems, vol. 108, pp. 110-124, 2016.
[7] M. Aminu, W. Nirmalie and L. Robert, "Contextual sentiment analysis for social media genres," Knowledge-Based Systems, vol. 108, pp. 92- 101, 2016.
[8] C. Tao, X. Ruifeng, H. Yulan, X. Yunqing and W. Xuan, "Learning User and Product Distributed Representations Using a Sequence Model for Sentiment Analysis," IEEE Computational intelligence magazine, pp. 34-44, 2016.
[9] R. Marco, S. Fabio and Z. Markus, "Analyzing user reviews in tourism with topic models," Information Technology Tourism , vol. 16, p. 5–21, 2016.
[10] P. J. Anuja and D. Padma, "Application of Machine Learning Techniques to Sentiment Analysis," in IEEE conference, 2016.
[11] Z. Xueying and Z. Xianghan, "Comparison of Text Sentiment Analysis based on Machine Learning," in 15th International Symposium on Parallel and Distributed Computing, 2016.
[12] Z. Yan, M. Melody and M. Teng-Sheng, "Multi-Layer Text Classification with Voting for Consumer Reviews," in IEEE International Conference on Big Data (Big Data), 2016.
[13] Rahul , Vasundhara Raj ,Monika Sentiment Analysis on Product Reviews 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)
[14] Hemalatha S , Ramathmika Sentiment Analysis of Yelp Reviews by Machine Learning Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2019)