

SENTIMENT ANALYSIS ON PRODUCT REVIEWS

¹Prof. Swati Powar, ²Mr. Sonu Parab, ³Mr. Sahil Panchal, ⁴Mr. Siddesh Parkar

¹Assistant Professor, ²Student, ³Student, ⁴Student ^{1,2,3,4}Department of Information Technology,
^{1,2,3,4}Finloex Academy of Management and Technology, Ratnagiri, Maharashtra, India

Abstract: In the digitalized world recommendation of any products to attract customers that meet their requirements is very crucial for the vendors to survive in the global market. Sentiment analysis is an intellectual way of identifying user's sentiments towards certain entities. Natural Language Processing (NLP) has the process of analyzing sentiments as one of its prominent fields. In today's e-commerce world Sentiment analysis is crucial part because it capture the opinion of any product. The growth in e-commerce has led to increasing customer reviews about various products which are available on the internet. Reviews are not only for product but also the service given to the customer. The objective is that identifying set potential feature in the review and extracting review. The approach proposed in this paper is novel and serves as a better alternative to rate a product based on its technical specification by analyzing large number of user reviews which are the subjective information in source materials i.e. e-commerce website by applying Natural Language processing, . Several methods have been developed in recent years in order to accomplish this task. In this paper, we discuss various levels of sentiment analysis followed by comparison among different approaches to sentiment analysis.

Index Terms – Sentiment analysis, product reviews, Naive Bayes.

I. INTRODUCTION

“We live LESS in real and MORE in virtual, this statement is turning out to be a reality with the rapid development of the Internet and Social media platforms and the ultimate increase in the number of social media users. With the expansion of e-commerce, the numbers of people that buy the products and review them increased. Different websites allows different review structure to be followed. Some websites allows user to explicitly write the advantageous and disadvantageous, in some cases along with the summary, in other cases there will not be any restrictions on user to write review, so that they can write however they want and express their feelings. Sometimes fake review is given on particular product. it has become possible to use sophisticated machine learning algorithms for this purpose. In this paper, we are providing solution to classify a particular review based on the data. Sentiment analysis is also called as opinion mining, which makes system to gather opinion about product. by obtaining important information from review it helps user while buying any product, and also helps best product between two. Many different techniques or approaches have been presented for analysis of sentiments in product reviews. Some of the methods are discussed in this paper. These approaches are basically categorized into machine learning based and lexicon based approaches. Approaches based on machine learning include some supervised and unsupervised classification algorithms. Lexicon based methodologies consist of dictionary based and corpus-based approaches. Sentiment analysis has a remarkable importance in the competitive environment of the market and the dynamic needs of people. It provides meaningful information at both customer level as well as enterprise level [3]. It has many applications in business intelligence, recommender systems, managing brand reputation, politics and government intelligence. [4]

II. LITERATURE SURVEY

The main approaches used to generate extractive summaries are (1) calculating semantics of words (2) lexical chains (3) combinations of heuristics The traditional text summarization and online product reviews summarization are completely different aspects because during summarizing user reviews, the aim is to gather and identify semantic features of products and later generate a comparative summary of products based on data obtained from feature-wise sentiment classification of the reviews that in turn will help the users and potential buyers in making a decision.[3] automated the process of gathering user review data for a particular product from any ecommerce platform and analyzed those reviews in terms of the sentiments expressed about specific features of the product. It involved: filtering of unwanted reviews; tokenizing the reviews into set of sentences, which were further broken down to a set of tags of which later dependencies were calculated like that of an adjective and noun. Sentiment Analysis is discussed in [2]. The authors implemented Dynamic language model and Naive Bayes classifier algorithms for Opinion Mining and have compared the efficiency. They also have thrown the light on fake review detection. [3] Talks about crawling the web to get all the information about a particular product and their reviews. Text Sentiments are also analyzed by using Stanford –corenlp analyzer API [1]. The review can be classify using similarity metric from balanced review sets Several investigations on the review states that it is possible to classify the reviews based on the nature. Based on the analysis, we could conclude that reviews even though were mixed in nature; it is possible to group them into clusters as needed. It is assumed that the reviews were trust worthy and aren't fake ones [7]. The Naïve Bayes classification is the most efficient algorithms for text classification of opinion mining [8].

III. DATA

For training the classification models, we are using a labelled dataset of 31,962 amazon reviews. The dataset is in the form of csv file. Each record in the dataset contains following attributes

- 1] Review id
- 2] Review
- 3] Review label

The used Dataset has 29720 tweets labelled as ‘ 0’ and 2242 tweets labelled as ‘ 1’ .The memory usage of the Dataset is 749.2 KB.

IV. METHODOLOGY

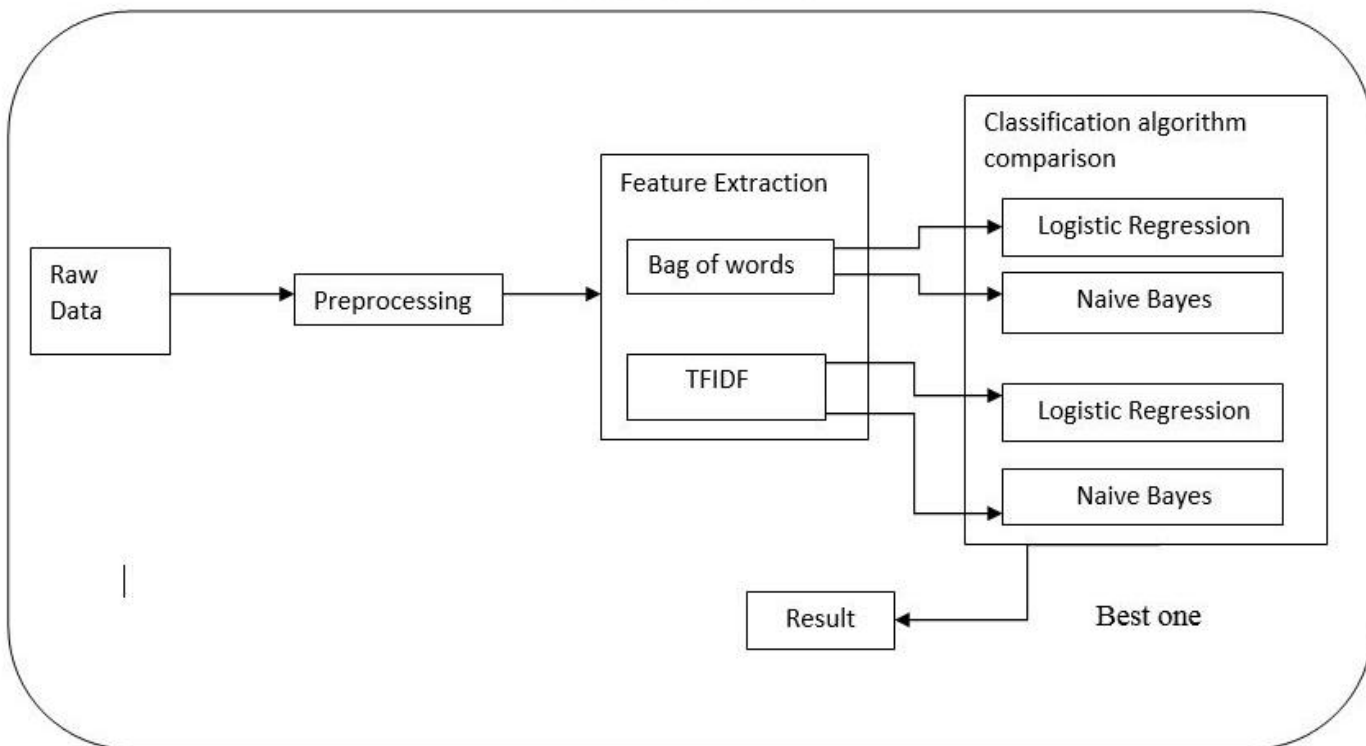


fig1. Architecture diagram 3.1 Population and Sample

We propose the above methodology for solving the problem. First we perform data pre-processing for noise removal which includes discarding language stop words URLs or links, mentions, hash tags, special characters and word stemming. Once we are done with pre-processing, then we perform feature extraction using bag of words (BOW) and Term frequency- inverse definition frequency (TFIDF). Then we use logistic regression classifier and naive bayes classification algorithms on both of the extracted features i.e. BOW and TFIDF to classify the tweets.

4.1 Logistic Regression

Logistic regression measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function. The recommended sample size for each category of independent variable is at least 10 observations per estimated parameter. Logistic regression uses large sample size which decreases the chances of errors

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k - E \tag{1}$$

Y= Dependent variables b= coefficient of variable X X= independent variable E= error term

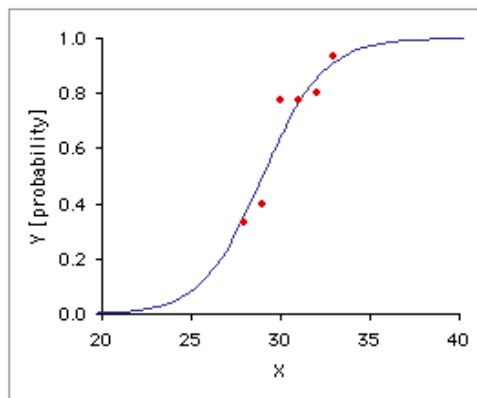


Fig. 2 Logistic Regressions.

4.2 Naïve Bayes

It is a classification technique based on Bayes Theorem. Naïve Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. It is easy to predict class of test data set.

$$P(c | x) = P(x | c) P(c) / P(x) \tag{2}$$

$P(c | x)$ = posterior probability of the target class

$P(c)$ = prior probability of class

$P(x | c)$ = likelihood which is the probability of predictor class

$P(x)$ = prior probability of predictor

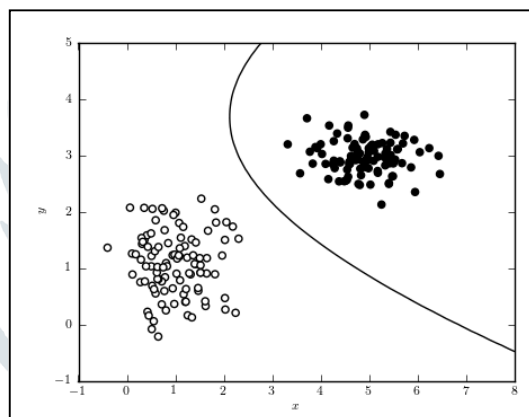


Fig. 3 Naive Bayes.

V. CONCLUSION

Sentiment analysis or opinion mining is the study that is used to analyze people emotions, sentiments towards the product. This research is used to perform evaluation measure on comments obtained from the customer. It is seen that sentiment analysis also plays vital role to make decision about product for improvement in the product and marketing strategies. Sentiment analysis is crucial part of business for product manufacturing companies as it helps them to identify satisfaction level of users, locate the drawbacks in product and find the area of improvement. The work can be further extended for improving sales of products on online platform. Creating multiple applications for E-commerce websites is possible with help of this study.

REFERENCES

[1] Ms. Swati Powar, Dr. Subhash Shinde “Named Entity Recognition and Tweet Sentiment Derived From Tweet Segmentation using Hadoop”, IEEE & CSI sponsored 1st International Conference on Intelligent Systems & Information Management Oct.2017

[2] Sayali Pednekar, Komal Patil, Rutweek Sawant “SENTIMENT ANALYSIS ON ONLINE PRODUCT REVIEWS” International Journal of Research Education

- [3] Prashant Kumar Singh, Arjit Sachdeva, Dhruv Mahajan, Nishtha Pande, Amit Sharma “An approach towards feature specific opinion mining and sentimental analysis across ecommerce websites”, 5th International Conference- Confluence the Next Generation Information Technology Summit, (Confluence) 2014
- [4] Anil Singh Parihar, Bhagyanidhi. “A Study on Sentiment Analysis of Product Reviews” 2018 International Conference on Soft-computing and Network Security (ICSNS)
- [5] A.Nisha Jebaseeli., E.Kirubakarm.” A Survey on Sentiment Analysis of (Product) Reviews” International Journal of computer Applications (0975 – 888) Volume 47– No.11, June 2012
- [6] Raheesa Safrin, K.R.Sharmila, T.S.Shri Subangi, E.A.Vimal “SENTIMENT ANALYSIS ON ONLINE PRODUCT REVIEW” in International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 04 | Apr -2017
- [7] Periakaruppan Sudhakaran, S hanmugasundaram Hariharan “Classifying product reviews from balanced datasets for Sentiment Analysis and Opinion Mining”, Joan Lu3 in 2014 6th International Conference on Multimedia, Computer Graphics and Broadcasting.
- [8] Santhosh Kumar K L, Jayanti Desai, Jharna Majumdar, “Opinion Mining and Sentiment Analysis on Online Customer Review” 2016 IEEE International Conference on Computational Intelligence and Computing Research

