ISSN No:-2456-2165

Sentiment Analysis on Online Product Review

Harshit Gupta[1], Lokesh kr. Tiwari[2], Kartik[3], Lav Agarwal[4], Atul Kumar[5]
[1,2,3,4] (Bachelor of Technology, Final year student), [5]Associate Professor,
Department of Computer Science & Engineering, IMS Engineering College, Ghaziabad, U.P., India

Abstract:- In today's world opinion mining is widely used topic, internet is having variety of data which is rich in sentiments which are used by different companies for various purposes. Our aim is to develop a web application which embeds machine learning model which provides the analysis of user reviews for particular product.

It shows the positive and negative polarity of reviews for searched product which will be helpful for users. In this application when a user searches for a product, review data is collected from ecommerce websites and that data is passed to a machine learning model which is a naïve bayes classifier which classifies the reviews into positive and negative sentiments based on the features extracted by the model. We show user the overall positive and negative polarity of reviews for searched product and also we show how accurately we have obtained the results. Thus these results helps user to decide about the product.

Keywords:- *Sentiment Analysis, Negative & Positive, Reviews of the Products.*

I. INTRODUCTION

The purpose of this project is to develop website that examines the product in positive and negative polarity on the basis of comments extracted from e-commerce website given as input. Each sample will be processed for selective features and an assessment will be done based on those features in order to provide the right product.

➢ *Sentimental Analysis*

Sentiment analysis (sometimes referred as opinion mining or emotion AI) it points towards the use of natural language processing, analysis of text, computational linguistic, & biometrics to systematically identify, extract, quantify, and study important states and subjective information.

The process of computationally distinguishing and categorizing opinions indicates in a piece of text, principally in order to determine whether the author's approach towards a peculiar topic, product and many other things is positive, negative, or neutral. It is context mining of content which identifies and selects subjective knowledge in source material, and helping a business to interpret the social sentiment of their brand, product or service while observing online conversations.

II. METHODOLOGY

The current research proposes knowledge system which aims at providing better review about the any product which is present on the e-commerce website. In this we take the comments given by the customers on the product and review them and present them in such manner that the user can understand the polarity of the comments on that product. These comments/reviews may be in the form of phrases or sentences. To know the sentiment of the users, these sentence have to be separated into Bag of words. These bag of words are checked with the trained classifier bag of words and expressed these sentences into negative & positive words. In the e-commerce sites their is a high growth of data and these data is used to tell the quality of product. Now a days the popularity of these sites are becoming high as they are major source for the customers to predict the exact rating and reviews of the product.

So this the benefit of our system that it saves the time and make it easy for the user to compile the rating and reviews and obtain the opinions from all this document, providing the exact view of the product.

III. IMPLEMENTATION

In this project we managed to organize the product review into two categories - positive, negative. Below figure depicts the implementation process. The implementation steps may include,

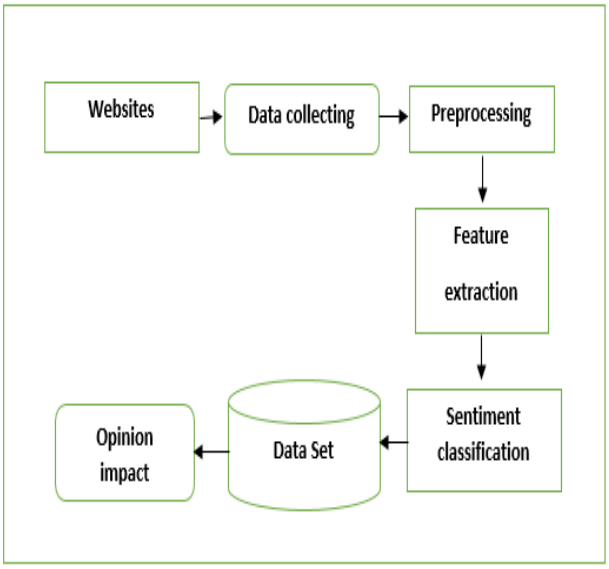

Fig 1:- Process of extracting features

A. Data Collection

Data collection is the way to collect the basic data or document on which work is done. Consumers express their sentiments about particular products on e-commerce websites like amazons. Their sentiments and opinions are expressed in different way, with different vocabulary, context of writing, usage of short forms and slang, making the data huge and disorganized. Manual analysis of sentiment data is virtually impossible. Therefore we uses the sentiment analysis to make this effort easy.

The website is made with many products inside it and the data about that products are collected in which reviews and ratings of that products are stored. The assessment is collected in the form of reviews which are given by the different consumers of that product shows their opinion about it.

B. Pre-processing

The pre-processing is nothing but filtering the extracted data before analysis. It includes identifying and eliminating non-textual content and content that is irrelevant to the area of study from the data.

C. Feature Extraction

In the feature extraction we try to find out the best suitable outcome from the reviews and ratings that are generated about the particular product on the e-commerce sites. To find out this we uses the method of bag of words which can predict the result is best way.

Bag of Word is the technique in computer science field known as natural language processing to extract features from text. These features can be used for training machine learning algorithms. It creates a vocabulary of all the unique words occurring in all the documents in the training set. The output of the bag of words model is a frequency vector.

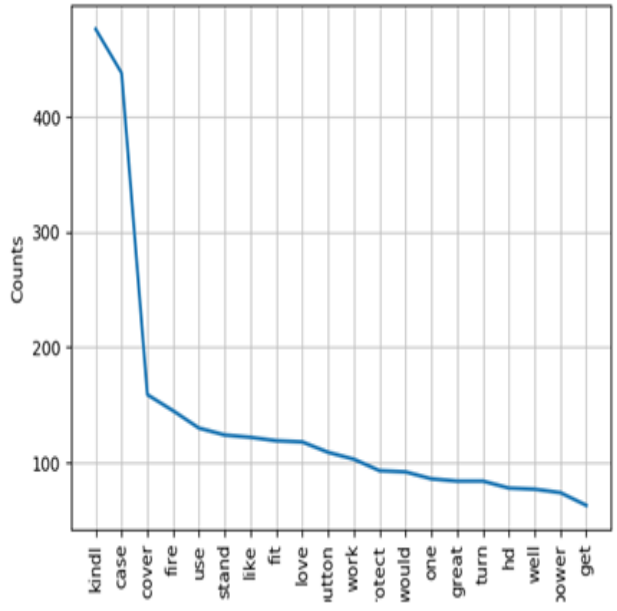

Fig 2:- Different word count

D. Sentiment Classification

➢ *Naive Bayesian Classifier*

Consider the case with one normally distributed predictor and two classes. We classify to the highest input density, taking the prior into account. A generative classifier is a model that specifies how to generate the data given the class conditional densities p(x|y = c) and the (prior) class probabilities p(y = c). This is a model for the joint distribution p(y, x). We compute the conditional probabilities for classification using Bayes' theorem,

p(y = c|x) = p(x|y = c)p(y = c) P c 0∈Y p(x|y = c 0)p(y = c 0) .

The Naive Bayes classifier (NBC) is a simple generative model based on the assumption that the predictors are conditionally independent given the class label. The class conditional density the becomes

p(x|y = c) = Y p j=1 p(xj |y = c).

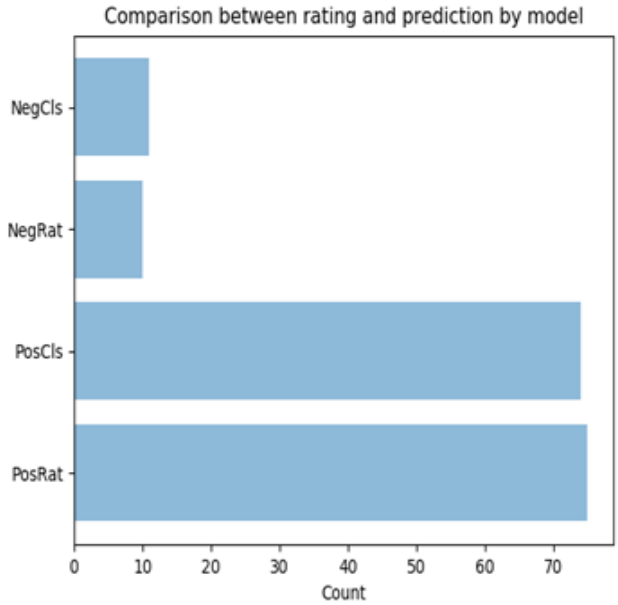

Fig 3:- This shows a comparison between the results on model from sentiment calculated by taking rating we have taken 3 star and above to be positive and 1 and 2 star are taken as negative.

E. Data Set

In the data set it is distributed in 2 parts:

- Training data set : Training data set is taken from UCI machine learning repository which contains 1000 reviews which are labelled positive and negative i.e. 1 and 0 . We have used 75% of data to train the classifier and 25% to evaluate the performance of classifier.
- Second type of data is unlabelled data which are reviews of various product which are to be classified as positive and negative polarity by using our trained classifier to review the product. We have collected these product reviews from Amazon.

IV. RESULT

The result displays the report about the product based on the features(user review) and this report is generated on the basis of general classification of positive or negative polarity of the features given by the users on the website.

The link of the whole project where you can check the whole working of the project with the code:

https://github.com/lokesh0412/product-review-using-sentiment-analysis.git

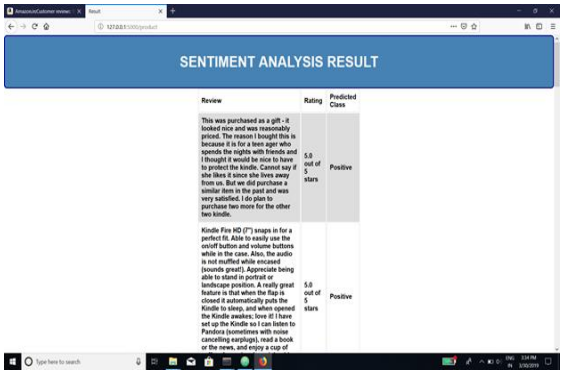

Fig 4:-: Showing the polarity of different reviews

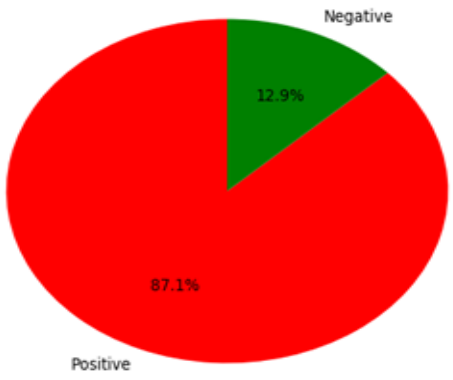

Fig 5:- Shows the percentage of positive and negative reviews for product named cello ball pen

V. PERFORMANCE EVALUATION

There are many ways in which you can obtain performance metrics for evaluating a classifier and to understand how accurate a sentiment analysis model is. Common methods used for evaluation include recall, precision, accuracy and f1 score.

Abbreviations as follows: true positive (tp), false Positive (fp), true negative (tn) and false negative (fn).

A. Recall

Recall is the ratio of correct positive classifications to the total number of actual positive instances.

Recall =tp/(tp+fn)

The percentage obtained during recall is 74.4%.

B. Precision

Precision is the ratio of correct positive classifications to the total number of predicted positive instances.

Precision = tp / (tp+fp)

The percentage obtained during precision is 77.5%.

C. Accuracy

Accuracy is the ratio of total number of correct predictions (true positive and negative) to the total number of instances.

Accuracy = (tp+tn)/ (tp+tn+fp+fn)

The percentage obtained during accuracy is 76.4%.

D. F1 Score

F1 which is a function of Precision and Recall.F1 Score is needed when you want to seek a balance between precision and recall.

F1= 2*[(precision*recall)/(precision+recall)]

The percentage obtained during f1 score is 75.91%.

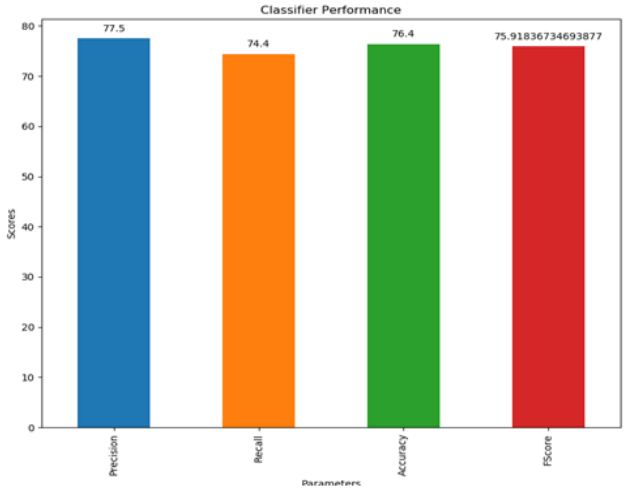

Fig 6:-: performance measure of a classifier

VI. CONCLUSION

Sentiment analysis is a area in which we finds about the consumer's sentiments, their way of thinking or emotions about the certain products. The problem on which this project works is the categorization of the sentiments in the different polarity. In this project various reviews and ratings of the different online product are taken from e-commerce sites are stored and used as a basic data on which classifier is imposed on reviews and ratings from which different bag of words are generated and using that bag of words polarity is decided. The sentiment generated from the website are shown in the best way so that customer can easily understand the polarity of the reviews generated on that e-commerce website. This is how we get the best suitable product review.

REFERENCES

[1]. Muhammad Taimoor Khan, Mehr Durrani 2, Armughan Ali, Irum Inayat, Shehzad Khalid and Kamran Habib Khan "Sentiment analysis and the complex natural language" Khan et al. Complex Adapt Syst Model (2016) 4:2
[2]. Kim S-M, Hovy E (2004) Determining the sentiment of opinions In: Proceedings of the 20th international conference on Computational Linguistics, page 1367.
[3]. Liu B (2010) Sentiment analysis and subjectivity In: Handbook of Natural Language Processing, Second Edition.. Taylor and Francis Group, Boca.Google Scholar
[4]. Stanford (2014) Sentiment 140. http://www.sentiment140.com/
[5]. M.Z.Asghar, A Review of Feature Extraction in Sentiment Analysis, Journal of Basic and Applied Scientific Research, 4(3)(2014), 181-186.
[6]. M.Eirinaki, S.Pisal and J.Singh, Feature-based opinion mining and ranking, Journal of Computer and System Sciences, 78(4)(2012), 1175-1184.
[7]. J.Isabella, Analysis and evaluation of Feature selectors in opinion mining, Indian Journal of Computer Science and Engineering, 3(6) (2012).