

Online Product Review analysis for Sentiments

Ishan Arora¹, Gagandeep Singh², Lokesh Kumar^{3*}

¹Computer Science, PSIT College Of Engineering, APJ Abdul Kalam Technical University, Kanpur, India

²Computer Science, PSIT College Of Engineering, APJ Abdul Kalam Technical University, Kanpur, India

³Computer Science, School of Engineering & Technology, Poornima University, Jaipur, India

*Corresponding Author: lokesh.skshukla@gmail.com, Tel.: +91-8840633141

Available online at: www.ijcseonline.org

Accepted: 15/May/2018, Published: 31/May/2018

Abstract— Buying and selling of things are a major part of human's since early ages , but with the development of the online market the trade got shifted from usual market to online for ease of everyone .Internet (www) has been a resource to get the user's review about the particular thing he had purchased. There are 2.4 billion active online users, who write and read online and use internet around us [1]. It will also help the companies to know what the problem the customers are facing in their use of the product. This will help the company to make better product and will surely help the customer to buy a product will large positive value [2]. With the help of the given system we classify the reviews. The paper will try to compare the various technique used to find out the opinion of the users .The proposed System will use the general algorithms of AI to find out the answer to this problem which are described in details in this paper.

Keywords – Sentiment Analysis, Naïve Bayes, Random Forest

I. INTRODUCTION

Opining mining is done to identify the customer's view toward the particular product. It helps to maintain the reputation of many company and their product. Since internet has become the major area for trade so the opinion of users are also important [3].Research have already proved that the 90 percentage of the customer decision full rely on the online review and analysis of those reviews [4]. The idea of the project was obtained from the Flipkart website as how they perform the web mining and preform the sentimental analysis of the reviews on different products. This analysis help the different companies to know their product was liked, disliked by the people buying it. Later on the other people use to see the positive negative chart at the bottom of the website to find the total no of positive ,negative review that help them to buy the products easily.

The software will work to find out the sentiments in the text file of reviews and calculate their polarity.

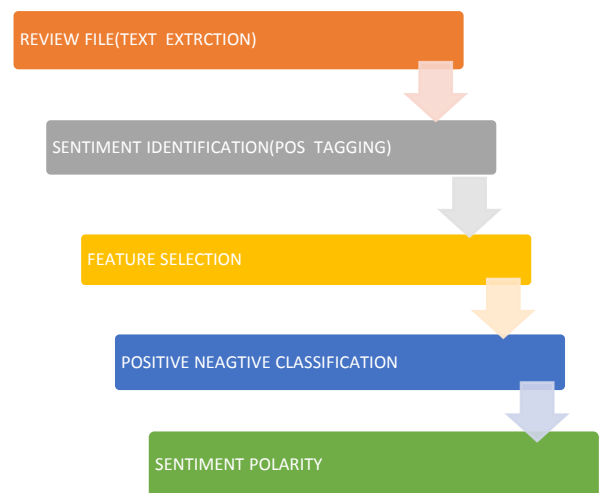


Figure 1. Process of Sentimental analysis

II. LITERATURE SURVEY

Sentimental analysis is very important in business to know customers opinion. Since much of the work is already performed on the Sentimental analysis, we can review some research work to get the essential idea about the project. Bing Liu [5] gave the list of the positive and negative words which were developed by using reviews of different customers. The

list consist of General words, some words wrongly written for the given classification problem. Xing Fang et al [6] proposed the process of this analysis. The full process was divided into 3 major parts. The techniques that were used are:

1. Naïve Bayes.
2. Random Forest.
3. SVM.

During the initial phase the objective content was filtered out and the major subjective content is retained .This context is now extracted for tagging. During the second step the identification of the NOA (Noun over adjective) and NOV (Noun over verb) is performed for sentimental extraction. At last in the Phase 3 Sentiment polarity Categorization is performed and final sentimental score is calculated.

III. RESARCH DESIGN ANDMETHODOLOGY

A. Data Collection

The review file that is used in this project was obtained from the amazon.com. Each line of the amazon review file was of the following form:

- 1) The reviewer ID
- 2) Product ID
- 3) Rating
- 4) Review time
- 5) Helpfulness
- 6) Review Text.

The cleaning of this review file was the important task because the file contains the unnecessary data which is of no use. The review and product id were kept after cleaning this large dataset. The new cleaned data set were of the following form:

- 1) Reviewer ID
- 2) Review Text

At last only the reviews of a particular id were used to find out the Sentiment of particular product at once.

B. Preprocessing

The obtained review file contains bulk of unnecessary which is of no use, hence it is pre-processed to get the valuable content from it. This valuable content is only read by the computer. Stop words are those words which are useless and hence should be removed to get the better output. The review was extracted for the review text and the text were word tokenized. The tokenized were then matched to the stop words in the list. The stop words were now removed and this list become the input of the POS tagging.

C. Sentiment Extraction

The P.O.S tagger is extremely essential [7] as the rest model depends on the tags produced by the tagger. P.O.S Tagger become important because of

- 1) Elimination of the noun pronouns from the text as they have no sentiments attached to them.
- 2) Tagging of words can also be used to find the difference between two words used in two ways in the same text.

Every line is tagged using this P.O.S tagging. By using python program we are able to increase the efficiency and accuracy of P.O.S Tagging .The Decision tree algorithm that was used in making of P.O.S which has the training score of 75.63 percent.

The Tagging Table is defined as:

Table 1. P.O.S Tags Description

POS TAGS	DESCRIPTION
CC	Coordinating conjunction
CD	Cardinal digit
DT	Determiner
RP	Particle
SYM	Symbol
PDT	Pre-determiner
RB	Adverb

D. Sentiment Processing

Every word used with in review text preforms an import task in sentimental analysis and can be identified using the P.O.S. Each review is tagged by using an algorithm in P.O.S tagging and with the help of python program we are able to it with more accuracy and efficiently [8]. The Output from tagger is then passed to the corresponding classifiers trained whose accuracy and working is explained in in the topic. The classifiers are used for predicting the sentences and total sentiments are plotted.

The classifiers and the methods used in the system are:

1. Naïve Bayes
2. Support vector machine
3. Logistic Regression

E. Plotting the Output

The use of chart-js library was done to visualize the total no of review. The output is given in the figure 7.1

- (a) And figure 7.1 (b). The visualization is very important as millions of reviews cannot be read with their sentiments on the screen, hence plotting the total positive and negative reviews would be helpful in all cases.

IV.METHODS USED ANDDISCUSSION

A. Evaluation and Methods

Software for building the proposed system were *Scikit-learn* NLTK and Chart-js.

1) Naïve Bayes Classifier

The principal of naïve Bayes algorithm is that no feature is dependent on each other. If every feature are dependent of each other than naive Bayes algorithm will automatically identify that the given features are not independent of each other. For the above reason it is termed as naïve.

This algorithms works on the Bayes theorem [9]. By using

this method we find out whether the given word of the document appears in negative word list or positive. The formulae for naïve Bayes Algorithm is:

$$P(A | B) = (P(B | A) * P(A)) / P(B)$$

Naïve Bayes algorithm works more efficiently than many other common algorithm. There are variety of applications build on it such as spam detection in email, Classification of the given document which are build using this algorithm. The training score of Naïve Bayes on two datasets were 72.43 and 87.0 accordingly.

2) Support vector Machine

SVM is also called the Kernel machine [10]. We are allowed to modify the working of SVM by using these kernel machines. Linear kernel is the most used kernel machine in text classification.

The Training Score obtained with on two datasets were 70.180 And 68.67 percent.

B. Analysis of some important model

The accuracy obtained were plotted using a graph and it can be confirmed that the Accuracy sore of the Naïve

Bayes is higher than both the Support vector machine and Logistic Regression.

Table 2. Accuracy of different method

METHOD NAME	ACCURACY
NAÏVE BAYSE ALGORITHM	87.0 %
SUPPORT VECTOR MACHINE	70.18 %
LOGISTIC REGRESSION	70.63 %
METHOD NAME	ACCURACY

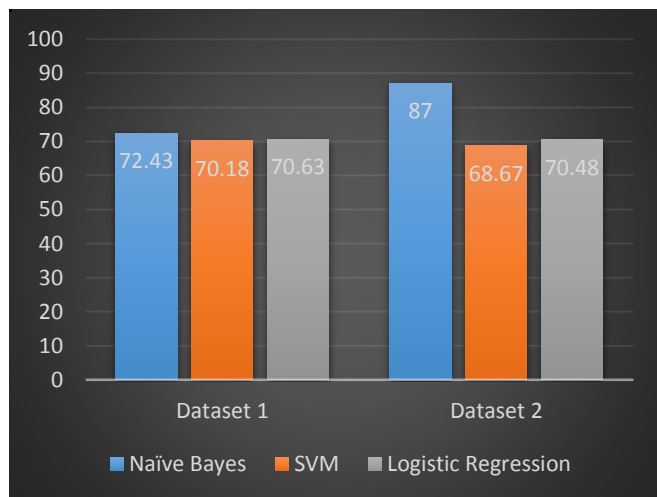


Figure 2. Comparisons of different methods on two dataset

V. WORKING OF THE PROPOSED SYSTEM

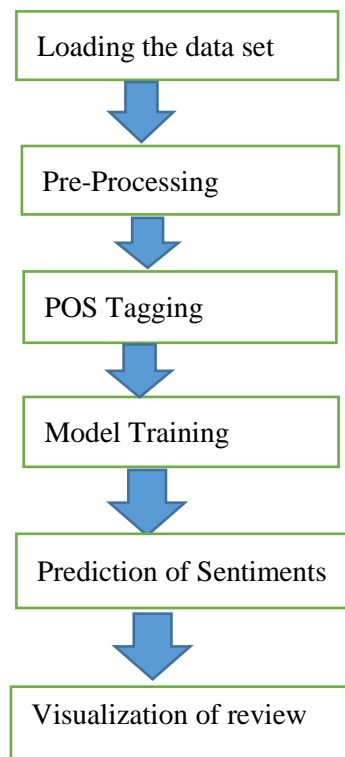


Figure 3. Working of Proposed System

VI. RESULTS

The result obtained from proposed system were in the form of graphs for better visualization. The Three figure denote sentimental analysis performed on the test file with 30 reviews and the output were the following. We have plotted the total no of positive and negative reviews on different charts. This will help the users to visualizes the result very easily when the reviews are more than thousands.

DATA VIZUALIZATION FOR REVIEWS



Figure 4. Line graph for positive negative Sentiments

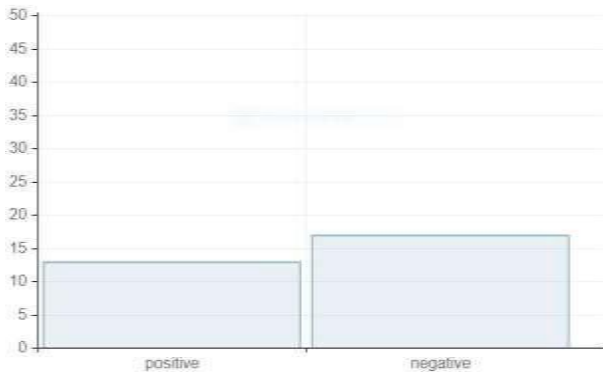
DATA VIZUALIZATION FOR REVIEWS

Figure 5. Bar graph for Positive negative sentiments

VII. CONCLUSION

The Paper helps to find the solution of the Sentimental analysis problem of the review. The Comparison of the accuracy obtained in the figures and the flowchart explains in depth the process of sentimental analysis. We have made this proposed system to work on many kind of reviews. The Sentiments are then plotted using the bar chart, pie chart and line chart for the better visibility and clear output. The output from the proposed system can be easily used by the companies and individual to find the sentiments of the user to their new product or defects in the product. The proposed system proves how Naïve Bayes is better for text classification than any other algorithm. The obtained results tells that how the company and people will be benefited as they can look the detailed graphs of positive and negative reviews. There will be no need to read the full bulk of reviews as our proposed system is fully automated [11].

REFERENCES

- [1] Peng, L., Cui, G., Zhuang, M., & Li, C. (2014). Hong Kong: Hong Kong Institute of Business Studies, Lingnan University.
- [2] AR. PonPeriasamy, G. Vijayasree, "Data Mining Techniques for Customer Relationship Management", International Journal of Computer Sciences and Engineering, Vol.5, Issue.4, pp. 120- 126, 2017.
- [3] U. Aggarwal, G. Aggarwal, "Sentiment Analysis: A Survey", International Journal of Computer Sciences and Engineering, Vol.5, Issue.5, pp. 222-225, 2017.
- [4] Pang B, Lee L (2004): A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics, ACL '04 .Association for Computational Linguistics, Stroudsburg, PA, USA
- [5] Bing Liu," Exploring User Opinions in Recommender Systems", Proceeding of the second KDD workshop on Large Scale Recommender System and the Netflix Prize Competition", April 2012, Las Vegas, USA.
- [6] Xing Fang, Justin Zhan, "Sentiment Analysis using product review data", Springer: Journal of Big data", 2015, North Carolina, A&T State university, Greensboro, NC, USA.
- [7] Opinion Mining and Sentiment Analysis Bo Pang and Lillian Lee Yahoo! Research, 701 First Avenue, Sunnyvale, CA 94089, USA, Computer Science Department, Cornell University, Ithaca, NY 14853, USA, llee@cs.cornell.edu
- [8] Roth D, Zelenko D (1998) Part of speech tagging using a network of linear separators. In: Coling-Acl, the 17th International Conference on Computational Linguistics. pp. 1136–1142
- [9] Wilson T, Wiebe J, Hoffmann P (2005) Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, Stroudsburg, PA, USA. pp. 347–354.
- [10] Peter D. Turney "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July2002, pp. 417-424.
- [11] Maks Isa, Vossen Piek. A lexicon model for deep sentiment analysis and opinion mining applications. Decision Support System 2012.

AUTHORS PROFILE

Ishan Arora is pursuing Bachelor of Technology in Computer Science and Engineering from PSIT College of Engineering, Kanpur. His area of interest include neural networks, opinion mining, sentiment analysis, etc.



Gagandeep Singh is pursuing Bachelor of Technology in Computer Science and Engineering from PSIT College of Engineering, Kanpur. His area of interest include Web Mining, Opinion Mining, etc.



Lokesh Kumar pursued Master of Technology from Amity University, Noida, 2014. He is currently working as an assistant professor in Computer Science Department, Poornima University. His area of interest include data mining, web mining, artificial intelligence, opinion mining, etc. He has 3 years of teaching experience.

