International Journal of Innovative Research in Engineering
Volume 3, Issue 2 (March-April 2022), PP: 33-36.
www.theijire.com **ISSN No: 2582-8746**

Product Review Monitoring System by Machine Learning

Prof. Sachin Darekar[1], Vidya Dada Atpadkar[2], Swati Dnyaneshwar Banchhere[3], Saloni Hridayanand Verma[4]
[1]*Assistant Professor, Information Technology, BVCOE/Mumbai University, India.*
[2,3,4] *B.E. Student, Information Technology, BVCOE/Mumbai University, India.*

How to cite this paper: Prof. Sachin Darekar[1], Vidya Dada Atpadkar[2], Swati Dnyaneshwar Banchhere[3], Saloni Hridayanand Verma[4], "Product Review Monitoring System by Machine Learning", IJIRE-V3I02-33-36.

Abstract: *As now more people are love to do the online shopping. People don't want to waste a time in shopping that's why they choose the online shopping. People goes to online store, search the product of his need and place the order. But sometimes there is a bad quality of the product, and also there is different different reviews are present, so while looking towards the reviews people or customers make his mind to buy that product or not to buy that particular product.Some of the negative reviews are also present their, through watching thatnegative reviews people make his mind not to buy that product. So for removing such type of negative reviews we proposed this Fake Product Review Monitoring System.*

Key Word*: Sentiment Analysis, Fake review detection, Machine learning algorithm, Fake Review, Spam Detection.*

I. INTRODUCTION

In our surrounded areas there are different ways to shop like you can buy a any product of your need by going to a shop or near by store. In this way of shopping the seller of the product gives you the feedback about that of product. Likethis you will get the review about the product. But now the situation is different, people are doing online shopping and for examine the product they just go to the comments or reviews of that product. Sometimes the reviews are real but sometimesthey are fake. There are many types of fake reviews like sometimes it posted for misleading the customer and sometimes for promoting the product some social media optimization posted such fake reviews. But because of this User or customer cant make his mind like to buy that product or not to buy that product. So currently this is very critical situation.

II. MATERIAL AND METHODS

The flowchart given below gives us an overview of how our review monitoring system will work and how the each and every module of the process will function like Product Review Dataset, Data Preprocessing, Spam Review Removal, Sentiment Score Count, Spam Review Detection, Feature Tagging, Analysis, Result.
The flow chart is given as the overall Workflow of Good or Bad Reviews monitoring, In this all the respected data will Preprocess and then it remove the spam reviews in it. Then the Sentiment Score Count will work, in it they count overall score count then the Detection process is done. Then in the Feature Tagging all the feature will tag in it. At the end the analysis and it shows the Result.

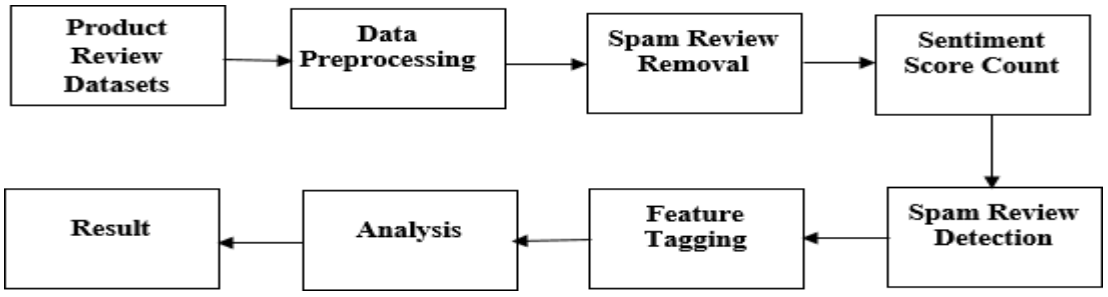


Fig1: Work Flow

A. Data Collection

Product reviews provide valuable information to online retailers about their effective gain and loss, profit and their worth in the market. For our paper, we take a dataset of Amazon product reviews.
To retrieve that dataset, we have used three techniques.

1. Using the tool WebHarvy web scraper,
2. Crawl data by developing a code,
3. Collecting data sets available open-source online.

B. Pre-Processing and Reviews Filtering

Online informal text of product reviews may involve bad and good words. It requires some methods to clean noise in raw text. We have created a 'bunch of words', containing around 1lakh words of expletives, and some other non-relevant words. This is also known as product feature extraction.

It involves-

• Removal of special characters and punctuations, like (#, ^, *,@,$,%,(,), etc.)

• Removal of malicious words

• Removal of repeating letters

• Remove fake comments from the dataset like "This is a faulty product" in this the word "faulty"Indicates fake review so it can directly remove from dataset.

C. Sentiment Analysis

The process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral.

D. Sentiment Score

1) Sentiment Score: Here the sentiment score shows a review's of sentiment polarity. That is, the degree of how genuine a review is. We calculated the sentiment score of each review. The score will be in the scale of -1 to 1.

Equation:-i=N

Score(r) = ™{(F)* (W)} / L

i=1 Where,

N= number of different Sentiment wordsF= frequency of the sentiment word

W= weightage of the sentiment wordr = r^{th} of review

L= square root of length of unique tokens

E. Spam Review Detection

We normalize the rating of product (1 to 5) to a scale of -1 to 1. All the reviews have their own ratings from -1 to 1. We calculate the sentiment score. The reviews where in the difference of the sentiment score and the rating is greater than 0.5, is considered as a spam. This is considered a word as a spam because a greater rating difference between the two.

For example, the review "The working of product is not so good in my opinion." has original rating 4 and normalized rating 0.5. Its sentiment score is -0.658. The difference between the two is 0.832, which, being greater than 0.5, constitutes spam.

F. Product Feature Analysis

In our paper, we consider the features of mobile phone as a product, we consider display, processor, version, camera, battery life, speakers, bluethooth,etc. Then, we divide the features in a series of sub-features. Eg- Speaker:- Sound, Volume, High and Low Volume. For review, we check all the features. And hence, the reviews remaining after spam filtering are categories good reviews.

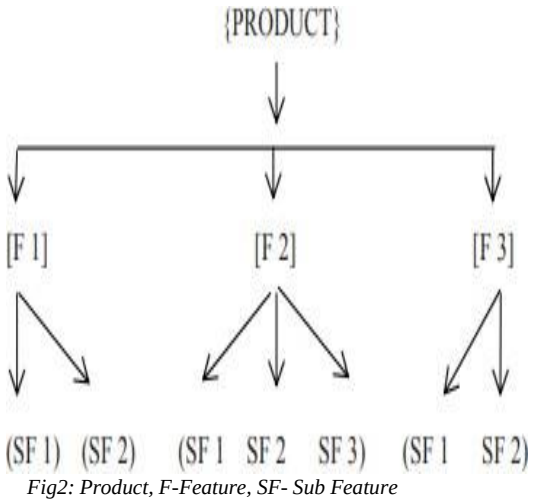


Fig2: Product, F-Feature, SF- Sub Feature

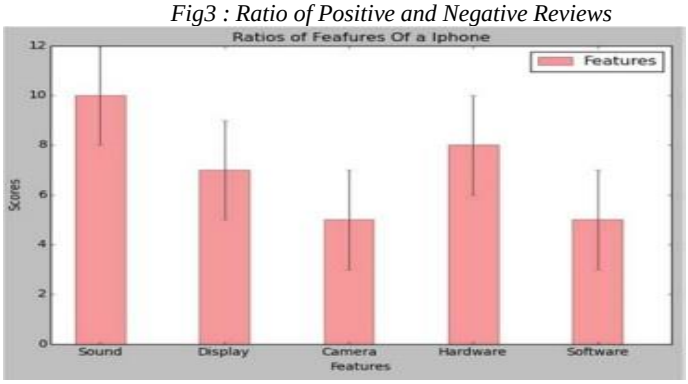
III. RESULT

Our study specifically takes into consideration 1Lakh random reviews of the product from the Amazon dataset.

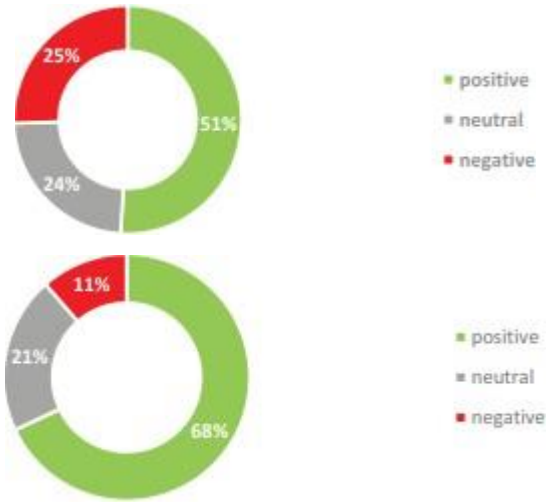
A. Dictionary Accuracy – We check the accuracy by comparing the sentiment score of the review calculated by it and the score calculated by NLTK text processing of online portal . Accuracy is the ratio of the no. of reviews with correct sentiment score and the total no. of reviews. The accuracy deduced by these methods comes out to be 68.2%.

B. Filtering and Spam Detection –By applying the algorithm We find the number of fake review. Out of 1Lakh product reviews, 50 reviews are found to be fake, and hence, removed, and 48 are found to be spam.
We check the product popularity as per the remaining reviews.

Fig3 : Ratio of Positive and Negative Reviews



Visual Representation of Different Products Review (Smart Phones, Washing Machine, AC)



We used Textblob, which is a Naïve Bayes Analyzer based sentiment text senti- ment analysis library and classified the polarity and subjectivity. We obtain results after classification as follows:

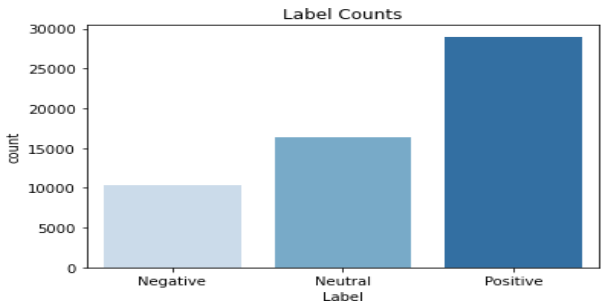


Fig4: Sentiment counts from dataset

IV. DISCUSSION

Number of methodology have been proposed in the field of fake product review monitoring. In past work different attributes have been utilized like the review words, sentences. One of the person Liu et al who proposed solutions on the deceptive or fake reviews. Liu provided a detail information about fake reviews detection. Liu et al. has proposed the algorithm is a supervised review learning method in which fake reviews or genuine reviews are regarded as a set of duplicate or near repeat reviews and other reviews are non spam reviews. In the detection procedure a classifier can be built to separate spamand non-spam reviews by employing natural language processing. In a supervised manner the part-of-speech n-gram featureis used to learn the fake and non-fake reviews.

Ohana B, Tierney B they take the e commerce review dataset then by applying some algorithm they count score of positive and negative reviews to find out sentiment analysis. From SentiWordNet as a source authors presented an improvement by building a dataset of that relevant features. They also applied machine learning classifier.

Rohit et al. use Positive-Unlabeled learning algorithm using various classifiers. In this, they uses six different classifiers to detect fake and truthful reviews. Here the logistic regression classifier gives awesome performance compared to all six different algorithms. Algorithms are decision tree, naive bayes, random forest, support vector machine, logistic regression and k-nearest neighbor classifiers are used.

Jitendra et al. applied different algorithm based on content similarity and sentiment polarity for identifying fake and good reviews. They used the linguistic and unigram as feature in that Fake Review Detection Techniques. They use three different algorithms like 1) support vector machine, 2) naive bayes and 3) decision tree.

V. CONCLUSION

Now day's people buying products online from E-commerce website improvement then also fake and real reviews are increases. Because of this huge reviews there is no good review has been identified. Some of the false reviews causes' bad selection of products happen. Therefore in this research sentiment analysis of reviews techniques into the spam review detection .In this work explains about various fake review detection techniques based on supervised and unsupervised methodologies. First we have made our own dictionary having sentiment words along with the weight given to the word according to its polarity. Then a method is proposed to calculate the sentiment score of the reviews from the natural language

Future work includes collecting abundant review data from other review web sites, computer assisted labeling of reviews to reduce the workload of human experts, more efficient model of detecting the relationship of reviews, reviewers and stores. The user to get the products from Daraz, Flipkart and Amazon with the satisfaction of their mind and pay for the good quality product. As, there are a lot of e-commerce stores like AliExpress and Alibaba which have reviews of multiplelanguages. It would be great if the proposed system finds the way, to process and filter the reviews for other multiple languages. future we would try to improve the method of calculating the sentiment score of the reviews. We would also tryto update our dictionary containing sentiment word.

References

[1] A. Rastogi, M. Mehrotra, "Opinion spam Detection in Online Reviews", Journal of information and Knowledge Management, vol. 16, no. 04, pp. 1-38, 2017.

[2] J. Rout,S. Singh, S. Jena, and S. Bakshi, "Deceptive review detection using labeled and unlabelled data", Multimedia Tools and Applications,vol.76, no. 3, pp. 3187-3211, 2016.

[3] S. Banerjee, A. Chua, J. Kim, "Using Supervised Learning to Classify Authentic and Fake Online Reviews ", Proceeding of the 9th International Conference on Ubiquitous Information Management and Communication", ACM, 2015.

[4] P.Rosso, D.Cabrera, M. Gomez, "Using PU-Learning to Detect Deceptive Opinion Spam", pp.38-45, 2013.

[5] R.Narayan,J. Rout and S. Jena, "Review Spam Detection Using Semisupervised Technique", Progress in Intelligent Computing Techniques: Theory, Practice, and Applications, pp. 281-286, 2018.

[6] W. Etaiwi,G. Naymat, "The impact of applying preprocessing steps on review spam detection", The 8th international conference on emerging ubiquitous system and pervasion networks, Elsevier, pp. 273-279, 2017.

[7] W. Zhang,R. Y. K. Lau and Li. Chunping, "Adaptive Big Data Analytics for Deceptive Review Detection in Online Social Media", Thirty Fifth International Conference on Information Systems, Auckland 2014,pp.1-19,2014.