Comment Sentiment Analysis and Fake Product Review Detection

Jyoti Bist[1], Neha Hulsurkar[2], Shraddha Bhalerao[3], Deepali Narkhede[4]

[1,2,3]Student, Information Technology, Shivajirao S. Jondhale College of Engineering, Dombivli, Mumbai, India
[4]Assistant Professor, Information Technology, Shivajirao S. Jondhale college of Engineering, Dombivli, Mumbai, India

---

Abstract—*Opinion mining has become very important with the increase of E-commerce websites and gives clear bifurcation to customers about the product reviews and service reviews. Sentiment analysis helps computer to extract emotions from customer's reviews further helping users in decision making process while shopping. Naive Bayes, a machine learning algorithm will be used for sentiment analysis and fake review detection along with other methods. In this paper we propose a system which improves users shopping experience by recognizing emotions behind the reviews and detecting fake or false reviews posted by opponent with wrong intentions*

Keywords-naïve bayes, opinion mining, sentiment analysis

1. INTRODUCTION

Now-a-days, due to the advent of technology and internet, shopping is mostly based on reviews or feedbacks. Traditionally when e-commerce websites were not considered. Buying product then using it will reveal its quality. E-commerce have played a vital role in changing this shopping culture. Today everything can be purchased with just a phone and network. As the use of e-commerce websites is increasing, it is more susceptible to fault intensions. Fake product reviews can lead to massive growth or great financial losses. We propose a project which focuses on removing fake reviews and analyzing users review for great shopping experience.

2. EXISTING SYSTEM

Several methods have been suggested to understand and implement opinion exploration and sentiment analysis. Scientists have developed models to identify the polarity of words, sentences, and the entire document. There are now several tools available to explore opinions, analyze feelings, and synthesize opinions. Previously, the basic concept among all the algorithms and models has been identifying emotional words first. These words are used to find where opinion is present inside the document. Then the opinions extracted are analyzed to find out polarity of the opinion.

2.1 Sentiment analysis of product reviews for ecommerce recommendation

This article describes the methods for performing magazine analysis. Here semi-supervised method suggested in which approach is on important opinion words finds using Word Net. This method is beneficial comparing with properly supervised or unsupervised algorithms with the help of advanced learning method as ANN to increase accuracy. Sentiment analysis is used at the sentence level with NLTK with the probability model Naïve Bayes. The representation of the results is done graphically and statistically [1].

2.2 Fake product review monitoring using opinion mining

Product reviews play an important role in deciding to sell a particular product on "e-commerce websites" or apps like "Flipkart", "Amazon", "Snapdeal", etc. In sentiment analysis, the goal is to figure out the opinion of a customer through a piece of text. First it is checked if the review is related to the specific product with the help of Decision tree. Spam dictionary is used to identify the spam words in the reviews. In Text Mining several algorithms are applied and on the basis of these algorithms specific results are obtained[2].

2.3 Fake review detection using opinion mining

As e-commerce grows and becomes more and more popular day by day, the number of comments received from customers about any product increases rapidly. Nowadays, people rely heavily on reviews before buying anything. This leads many people to write unnecessary scams and reviews about other related products or services. Some organizations in the marketplace even hire professionals to write false reviews and promote their products or defame the products of their competitors. Therefore, this article aims to develop a method that detects and records false reviews. The proposed method automatically classifies users' opinions into "suspicious", "clear" and "fuzzy" categories by phase processing. The fuzzy category recursively reveals suspicious or clear elements. This results in richer detection and can be useful to both the business organization and customers. The sales

organization can monitor the sale of its products by analyzing and understanding what customers say about the products. This can help customers buy valuable products and spend their money on quality products. Finally, end users view each individual exam with scored polarity scores and a credibility score [3].

2.4 Amazon review classification and sentiment analysis

Biggest online shopping mall in the world is "Amazon". Before buying the item people always belief items by looking at product reviews. But reviews on "Amazon" are not necessarily items, but a mix of item review and item service review ("Amazon"-linked or item-company-linked). Customer is mistaken as a general feeling (grading ranking) that "Amazon" provides feedback as a collective opinion and that there is no difference between a service review and a item review. The proposed model makes a satisfactory distinction between service review and product review, in addition to that, if the user talks about particular product feature it also classifies the review as a feature review. A featured review is nothing more than a product review, this template also gives an impression of the text about the functionality of the product. For example, if the user writes in their review, "The storage for this phone is excellent," then we also rate the storage function as positive. This was to construct a system that envision the feel of the magazine in the form of graphics [4].

3. PROBLEM DEFINITION

The main objective of this software is to help users gather the best details about a particular item they want, based on the opinions of other consumer, and to help make a choice about any product. This system detects the hidden sentiments like good or bad reviews in comments and rate them accordingly and also it detects the fake reviews of product. The system uses opinion mining methodology in order to achieve desire functionality. This system is beneficial to the user which provides the analysis sentiment and fake product review.

4. ARCHITECTURE

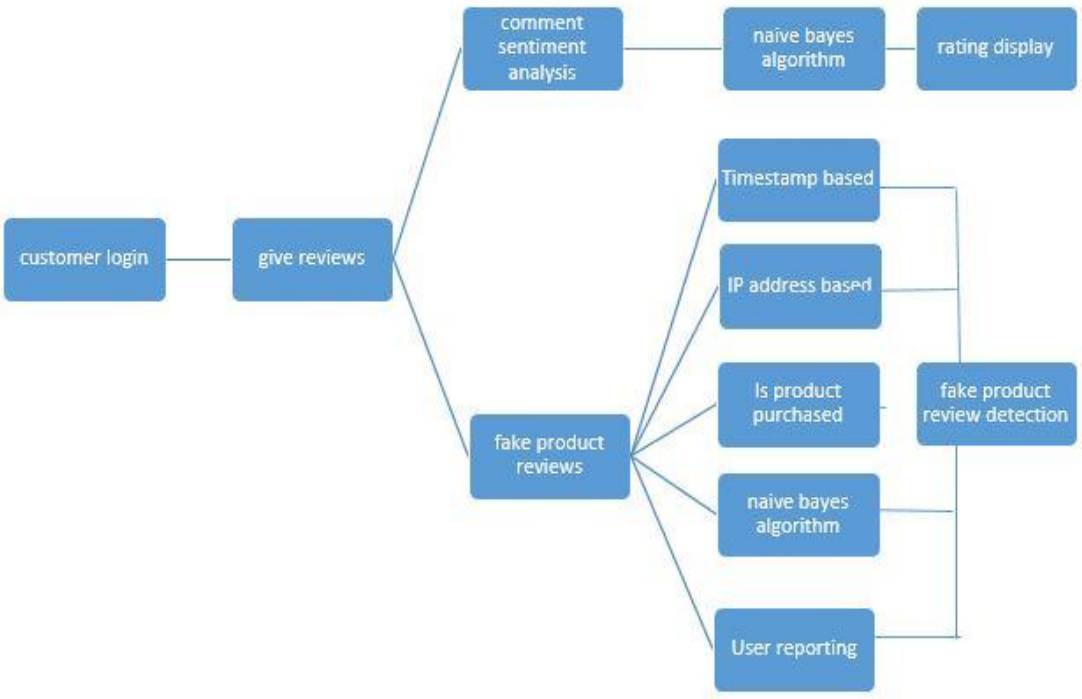


Fig-1 System Architecture of Comment Sentiment Analysis and Fake Product Review Using Opinion Mining

System works as follows:-

1. User registers using email id.

2. User logins to the system

3. User once access the system, can see product and can post review about theproduct.

4. Overall rating will be calculated by the system for a particular product.

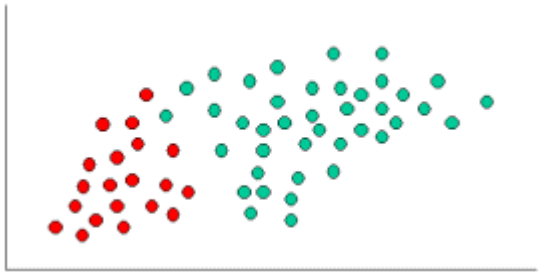
5. Fake reviews are detected by the system and will be removed by the admin

5. METHODOLOGY

5.1 Comment Sentiment Analysis

- **Naïve Bayes Algorithm**

This algorithm is also known as Bayesian theorem. It is mostly used when the dimensionality of inputs are high.



Consider example displayed in the illustration above. Let the red color indicate apples and green color indicates grapes. Our aim is to classify the new cases on arrival, that is, to decide which class label they belong to, according to the objects that currently come out.

Given that there are twice as many grapes as apples, it is obvious to trust that a new case (which has not yet been observed) is twice as likely to have homemade grapes instead of apples. In Bayesian analysis, this belief is called an earlier probability. The above probabilities are based on previous experience.

$$\text{Prior probability for Grapes} \propto \frac{\text{Number of Grapes}}{\text{Total number of fruits}}$$

$$\text{Prior probability for Apples} \propto \frac{\text{Number of Apples}}{\text{Total number of fruits}}$$

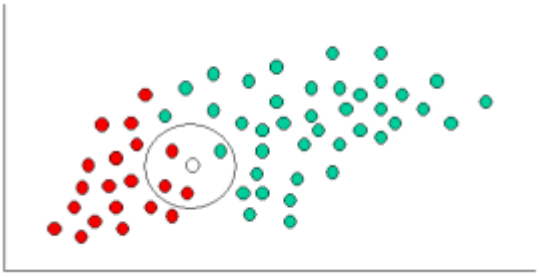
Since there is a total of 60 objects, 40 of which Are Grapes and 20 Apples, our prior probabilities for class membership are:

After formulation of prior probability, we are now ready to classify a new object (WHITE circle).

Since more likely that the new cases belong to that particular fruit.

$$\text{Prior probability for Grapes} \propto \frac{40}{60}$$

$$\text{Prior probability for Apples} \propto \frac{20}{60}$$



$$\text{Likelihood of X given grapes} \propto \frac{\text{Number of Grapes in the vicinity of X}}{\text{Total no of grapes}}$$

$$\text{Likelihood of X given apples} \propto \frac{\text{Number of apples in the vicinity of X}}{\text{Total no of apples}}$$

From the illustration, it is clear that the likelihood of X given grapes is smaller than apples as the circle encompasses 3 apples and 1 grapes.

$$\text{Probability of X given grapes} \propto \frac{1}{40}$$

$$\text{Probability of X given apples} \propto \frac{3}{20}$$

In "Bayesian analysis" the concluding classification is made by combining the two sources of data, i.e. the previous one and the probability, to form a posterior probability using the so-called "Bayes"rule.

$$\text{Posterior probability of X being grapes} \propto \text{Prior probability of grapes x likelihood of X given grapes}$$

$$= \frac{4}{6} \times \frac{1}{40} = \frac{1}{60}$$

$$\text{Posterior probability of X being apples} \propto \text{Prior probability of apples x likelihood of X given apples}$$

$$= \frac{2}{6} \times \frac{3}{20} = \frac{1}{20}$$

Lastly, we classify X as apple since its class membership reaches the largest posterior probability.

Note. The above probabilities are not standardized. However, this does not change the result of the classification since their normalization constants are the same.

5.2. Fake product review detection

- IP address method

In this method, IP address of the user is tracked. If same IP address is used for posting multiple reviews then the Reviews are considered Spam.

- Timestamp Method

Timestamp is a digital proof of the time of occurrence of a particular event. The system detects if the timestamp of the reviews are same.

- Is product purchased

System check if the user posting the review about the particular product has purchased the product or not.

- User reporting

System keeps record of reviews which are reported by other users. Admin can delete the reported comments.

6. RESULT

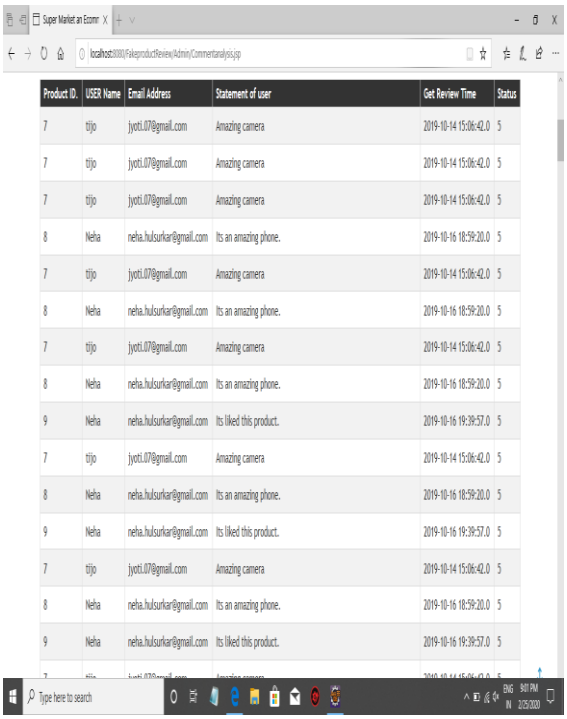


Fig-2 Comment Sentiment Analysis Result

Fig.-3 User Report Method

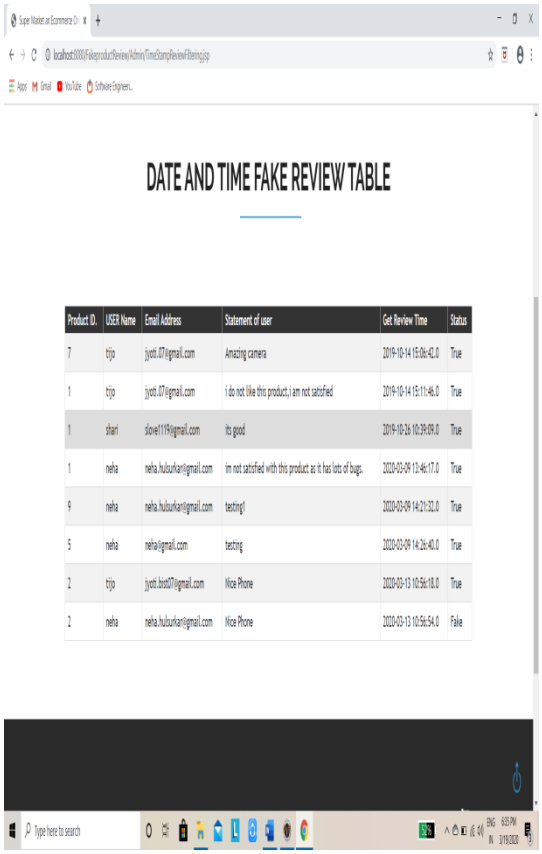


Fig-4 Fake Product Review Detection Using Timestamp Method

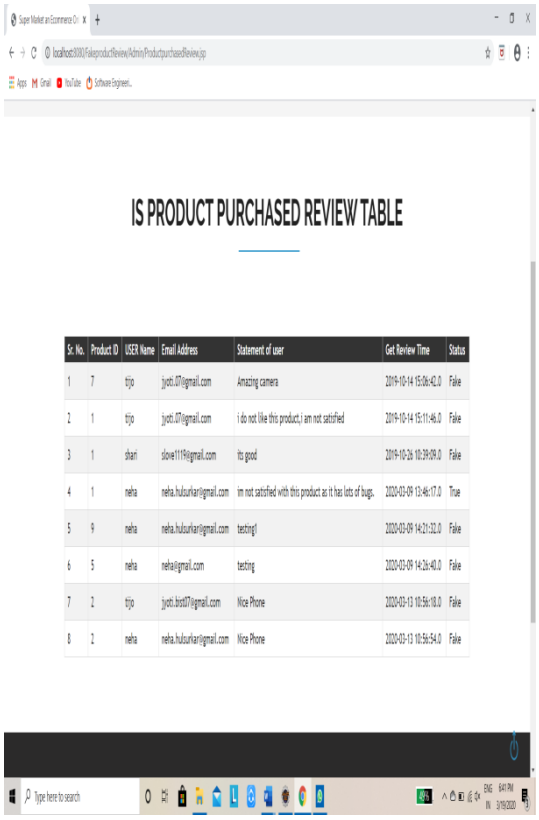


Fig.-5.Fake Product Review Detection Using Product Purchased Method

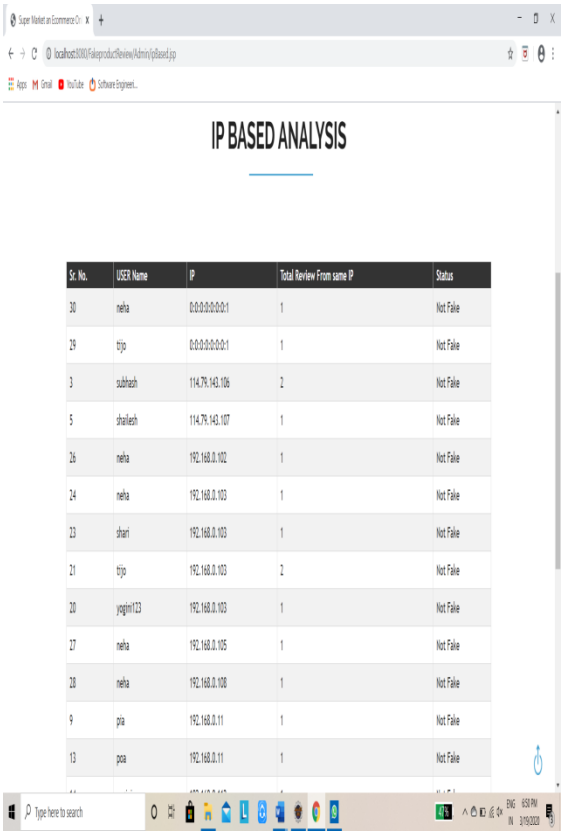


Fig-7. Fake Product Review detection using IP based analysis

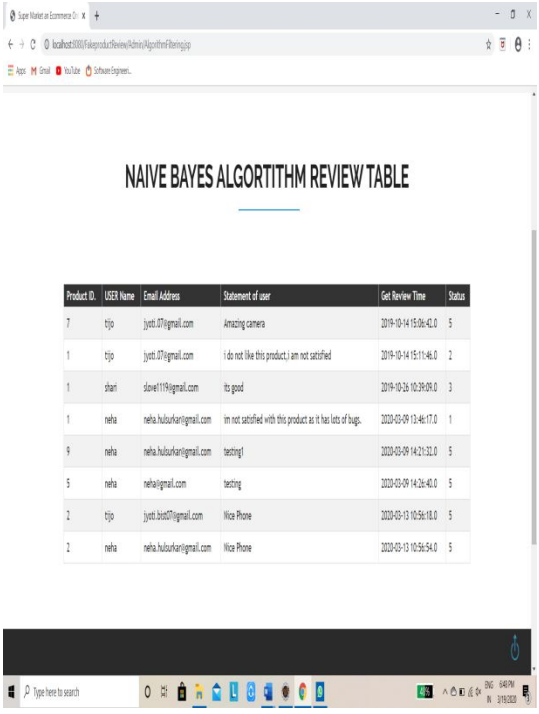


Fig-6. Fake Product review detection using naïve Bayes algorithm

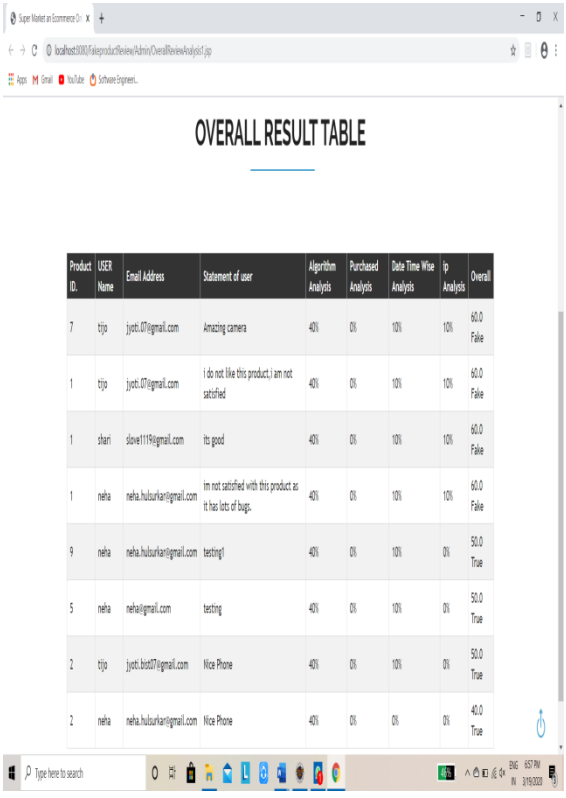


Fig-8. Overall Result Table

7. CONCLUSIONS

Now a day's technology is growing day by day and there are so many website and application available in the online market by which they sell their product. Every product contains millions of reviews and on basis of these reviews user buy the product most of the time. There are some organizations which post fake reviews on genuine product for some reasons. Reasons are defamation of the organization by the opponent company also for promoting its own products organization often pay its employees to comment on its products. So our software will do analysis of these comments which display the overall rating of the product which will help customer for choosing the product and provide only genuine reviews by detecting the fake review.

8. REFERENCES

[1] Abhyankar, Bhavarthi, Gaidhar, Bangare ,"SentimentAnalysis Of Product Reviews For Ecommerce Recommendation"Mali,Conference Paper · January 2016

[2] Anusha Sinha, Nishant Arora, Shipra Singh, MohitaCheema,AktharNazir, "Fake Product Review Monitoring Using Opinion Mining", International Journal of Pure and Applied Mathematics, Volume 119 No. 122018

[3] Dhairya Patel, Aishwerya Kapoor, SameetSonawane, "Fake Review Detection Using Opinion Mining "International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue:12|Dec2018

[4] Aashutosh Bhatt, Ankit Patel, Harsh Chheda, KiranGawande," Amazon Review Classification and Sentiment Analysis",International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 6 (6) 2015

[5] Madhura N Hegde ,Sanjeetha K Shetty , Sheikh Mohammed Anas , Varun K, "Fake Product Review Monitoring" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, Volume: 05 Issue: 06 | June2018

[6] SanthoshKumar K L Jayanti Desai Jharna Majumdar "OpinionMiningand SentimentAnalysisOnOnlineCustomerReview"Conference Paper · December [2016]