

Fake Product Review Monitoring And Removal

Shaikh Mohd Danish¹, Siddiqui Mohd Aasim², Shaista Hifzurrahman³

Dr. A.K Sampath⁴, Dr. Varsha Shah⁵

¹Computer Engg, Rizvi College of Engineering

²Computer Engg, Rizvi College of Engineering

³Computer Engg, Rizvi College of Engineering

⁴HOD Of Computer Engg, Rizvi College of Engineering

⁵Principal of Rizvi College of Engineering

Abstract - In the current scenario, the data on the web is growing very Fast. Social media is generating a large amount of data such as reviews, comments, and customer's opinions on a daily basis. This user generated data is worthless unless some mining operations are performed to it. As there are a number of fake reviews so opinion mining technique should incorporate Spam detection to produce a genuine opinion. Nowadays, there are a number of people using social media opinions to make their decision on shopping for product or services. Opinion Spam detection is an exhausting and hard problem as there are many faux or fake reviews that have been created by organizations or by the people for various purposes. They write fake reviews to mislead readers or automated detection system by promoting or demoting target products to promote them or to degrade their reputations. The proposed technique includes Ontology, Geo location and IP address tracking, Spam words Dictionary using Naïve Bayes, Brand only review detection and tracking account used.

Key Words: Fake reviews, Fake reviews detection, Opinion mining, Sentiment analysis, Text mining, Supervised learning, Spam, IP address

1.INTRODUCTION

As most of the people needs review about a product before spending money on the product. So people come across various reviews/comments in the websites but these comments/reviews are genuine or fake is not identifying by the consumer. In some websites some good reviews are added by the product seller itself in order to make product famous this peoples belongs to Social Media Optimization team. They write good reviews for many different products manufactured by their own industry. So user won't be able to find out whether the review on a product is genuine or fake. To find out fake and genuine review in the website this "Fake Product Review Monitoring and Removal" system is introduced. This system will find out fake reviews written by the social media optimization team by identifying the IP address. User will login to the system using his id and password and can view many products and will give review about the product. To find out the review is fake or genuine, system will find out the IP address of the user if the system observe fake review send by the same IP Address many time it

will inform the admin to remove that review from the system. This system uses data mining method. This system helps the user to find out correct reviews of the product.

2. Problem Statement:

The main agenda is to boost customer satisfaction and online shopping experience. so as to try to to so, it's become a typical practice for online merchants to enable their customers to place forward their reviews/comments on the products that they need purchased. With more computer users becoming comfortable with the net, a large number of individuals are coming forward to jot down the reviews/comments and post them on website which is becoming beneficial for other customers. It also decides profit or loss for any e-commerce websites. As an outcome of this, the amount of reviews that a product receives is grow quickly. Most of the famous products get thousands of reviews/comments at some large Websites. Now any user can write any review text or comment, this will create the individual's attention and organizations to present undeserving wrong opinions to push or to discredit some target products. the prevailing system doesn't restrict junk and invalid reviews and comments. So there's a necessity to develop a system which automatically mine opinions and classify them into spam and non-spam category. As most of the people needs review a few product before spending their money on the merchandise. So people come upon various reviews within the website but these reviews are genuine or fake isn't identified by the user. In some review websites some good reviews are added by the merchandise company people itself so as to form so as to provide false positive product reviews. they provide good reviews for several different products manufactured by their own firm. User won't be able to see whether the review is genuine or fake. to seek out fake review within the website this "Fake Product Review Monitoring and Removal for Genuine Online Product Reviews Using Opinion Mining" are used. this technique helps the user to search out correct review of the product.

3. Objective:

The main objective of our research are the following: To develop a replacement mechanism and build a replacement model for overcoming this system challenges. to search out and define the technique for detecting fake reviews/comments using Supervised learning techniques. To

detect unfair users reviews for a product, and improve the accuracy using Sentiment Analysis algorithms and supervised learning techniques. to create a Sentiment analysis Model and compute spam scores using the supervisor learning technique. To detect collusion and manipulation done by, both sides, customer and seller. to produce the developer to try to to improvement of current spam systems and take into all of the problems focused on this study.

4. Scope:

Finding the review spam from huge amount of unstructured data has become a very important problem. Now businesses industries, specialists and academics are insisting their efforts, researches and concepts to seek out the most effective system for opinion review analysis. Although, a number of the algorithms are employed in opinion spam analysis gives fare results, but still no algorithm can resolve all the challenges and difficulties faced by today's consumers. More future work and knowledge is required on further improving the performance of the opinion review analysis. within the future we'll do further investigate different types of features to create more accurate predictions.

5. Methodology to accomplish objectives of the research:

We build a supervised binary classifier to classify reviews as fake or not. Figure13 shows thethree phases conducted after feature extraction. We begin by preprocessing the data. Then,we compare the results of different classification algorithms. We optimize the algorithmsby feature selection and hyperparameter tuning. Last, we evaluate the importance of the classification features. Afterwards, we conduct an in-the-wild experiment to evaluate howour classifier performs in practice, i.e., on imbalanced data. Fig – 1.

6. Data Preprocessing

We pre-processed the info in three steps. We began by performing data cleaning, i.e., filling null values rather than removing affected columns. Of the chosen features only one column includes null values, the review frequency is in 1,734 cases undefined because only one review was provided by the reviewer. during this case, we set the frequency to lifetime of the app store, which is 9 years. Then, we normalized the dataset so individual samples to possess unit norm. We used the normalize() method with standard parameters of the preprocessing module provided by scikit-learn .Last, we standardized the dataset so the individual features are standard normally distributed, i.e., Gaussian with zero mean and unit variance. this is often a standard requirement for several classification algorithms, like the radial basis function (RBF) kernel of support vector machines. If not standardizing the info, features with a far higher variance compared to others might dominate the target function. As a result, the classification algorithm is unable to be told from other features. We used the scale()method with standard parameters of the preprocessing module.

7. Feature extraction:

The preprocessed data is converted into a group of features by applying certain parameters. the subsequent features are extracted: Normalized length of the comment/review-Fake reviews tend to be of smaller length. Reviewer ID-A reviewer posting multiple reviews with the identical Reviewer ID. Rating-Fake reviews in most scenarios have 5 out of 5 stars to entice the customer or have least rating for the competitive products thus it plays a vital role in fake detection. Verified Purchase-Purchase reviews that are fake have lesser chance of it being verified purchase than genuine reviews. Thus these combination of features are selected for identifying the fake reviews. This successively improves the performance of the prediction models.

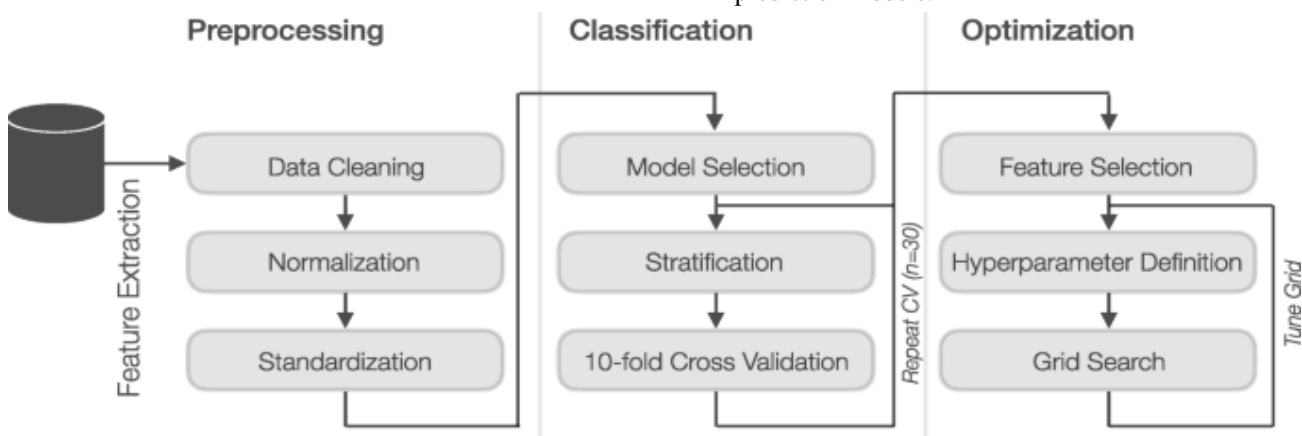


Fig -1: Overview of review classification

8. Sentiment Analysis:

Classifying the reviews per their emotion factor or sentiments being positive, negative or neutral. It includes predicting the reviews being positive or negative consistent with the words employed in the text, emoji is use, ratings given to the review and then on. Related research shows that fake reviews has stronger positive or negative emotions than true reviews. the explanations are that, fake reviews are wont to affect people opinion, and it's more significant to convey opinions than to plainly describe the facts. The Subjective vs Objective ratio matters: Advertisers post fake reviews with more objective information, giving more emotions such show happy it made them than conveying how the merchandise is or what it does. Positive sentiment vs negative sentiment: The sentiment of the review is analyzed which successively help in making the choice of it being a fake or genuine review.

9. ALGORITHM:

1. Take input five documents of each category.
2. Extraction of 100 topics (cluster of words) using LDA implement in MALLET
3. Topic models are used as training set to calculate global topic distribution of two sub categories.
4. Probability of every topic is calculated with the review to be tested.
5. Weighted topic distribution is use as a differentiator by calculating the score of the review
6. Review identified as fake or true.
 - Based on style of word used: -
 - a. Input the five documents of every category
 - b. Extract hundred topics from the input documents.
 - c. Counting number of nouns, adjective, adverbs and comparative words in each topic using POS tagging.
 - d. Identify the category of topic supported word count of every type (categories mentioned above)
 - e. Divide the topics into two classes truthful and false.
 - For each topic if $CE+DI \text{ words} > GC+CA+RR \text{ words}$
 - then class 1
 - else class2
 - f. Train the SVM classifier using this data.
 - g. Convert each review to be tested into vector
 - h. Test the review to be fake or true.

10. Flow Chart:

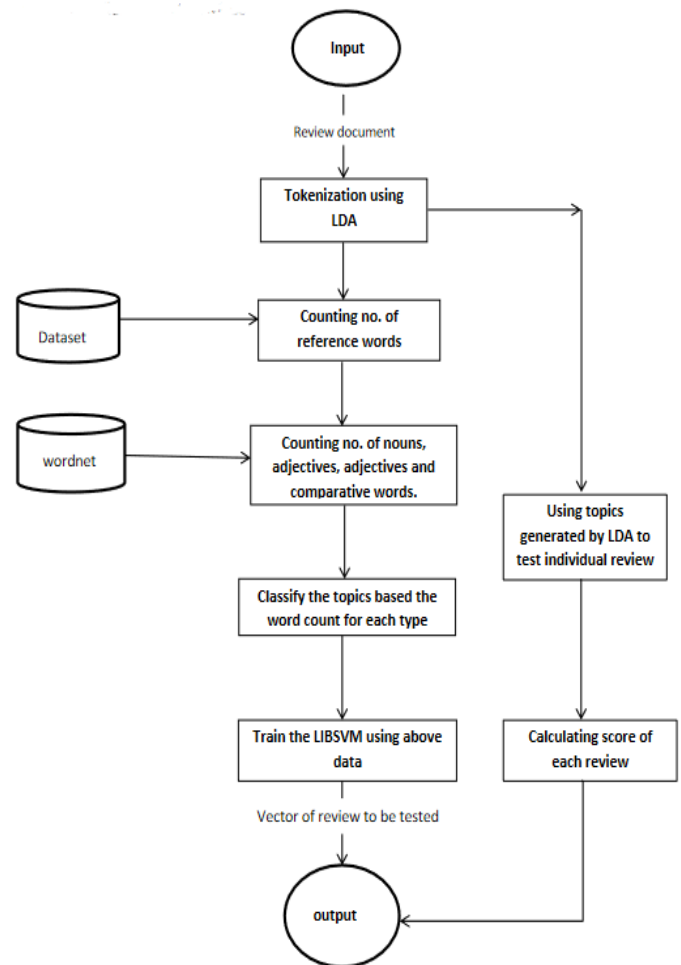


Fig -2: Flow Chart of a Model

11. Performance Evaluation and Result:

Last, we analyzed the relative importance of the extracted features with relevance to the pre-dictability of whether a given review is fake or not, called feature importance. The feature importance is calculated on how often a features is used in the split points of a tree. More frequently used features are more important. The three most vital feature are the overall number of reviews an app received (30%), the whole number of reviews a user provided (24%), and therefore the frequency thereupon a user provides reviews (13%). We assume the entire number of reviews received by an app is that the most vital feature as apps with a particular amount of reviews, e.g., 2–9 reviews (cf. Section4.1), are most frequently targeted by fake reviews. The full number of reviews a user provided in addition review frequency have a high importance as fake reviewers provide way more reviews than regular viewers, with a better frequency. The percentage of 1- and 2-star ratings an app receive dare important, because the difference between those star ratings provided are the very best when comparing site with fake reviews and regular site.

12. CONCLUSION

From our work we've got come to a conclusion that finding the opinion junk from huge amount of unstructured data has become a vital research problem. Although, a number of the algorithms are utilized in opinion spam analysis gives good results, but still no algorithm can resolve all the challenges and difficulties faced by today's generation. it's important to contemplate certain quality measures like helpfulness, usefulness and utility while analyzing each review. Our application which is able to help the user to procure the correct product with none stepping into any scams. Our application will do analysis and so post the real reviews on genuine product. And user are often sure about the products availability thereon application and reviews too. In future we might attempt to improve the strategy of calculating the sentiment score of the reviews. we'd also attempt to update our dictionary containing sentiment word. we might attempt to add more words in our dictionary and update the weights given to those words to induce more accurate calculated score of the reviews. Sentiment analysis or opinion mining is applied for any new applications which follow data processing rules. A direction for future research is to implement the system and check performance by applying proposed approach to numerous benchmark data sets. the most objective of our work is to make a system which is able to detect spam and redundant reviews and to filter them so user correct knowledge about the merchandise. Aim of our project is to boost customer satisfaction furthermore on make online shopping reliable. The project will detect the fake reviews by deploying opinion mining algorithms and creating a word dictionary.

REFERENCES

1. Eka Dyar Wahyuni and Arif Djunaidy, "Fake Review Detection from a Product Review using Modified Method of Iterative Computation Framework" MATEC Web of Conferences BISSTECH 2015
2. Kaushik Varadha Rajan, Nivasse Ajagane, Shubham Srivastav, "Evaluating Performance of Semi-Supervised Self Training in Identifying Fake Reviews," North Carolina State University , Raleigh NC 27606, USA
3. Shashank Kumar Chauhan, Anupam Goel, Prafull Goel, Avishkar Chauhan and Mahendra K Gurve "Research on Product Review Analysis and Spam Review Detection" 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)
4. Shivaprasad TK Jyothi Shetty "Sentiment Analysis of Product Reviews: A Review" International Conference on Inventive Communication
5. E. I. Elmurngi and A.Gherbi, "Unfair Reviews Detection on Amazon Reviews using Sentiment Analysis with Supervised Learning Techniques," Journal of Computer Science, vol. 14, no. 5, pp. 714–726, June 2018.
6. Benjamin Snyder and Regina Brazil, "Multiple Aspect ranking using the Good Grief Algorithm "Computer Science and Artificial Intelligence Laboratory Massachusetts Institute of Technology 2007.
7. Kolhe N.M., Joshi M.M., Jadhav A.B., Abhang P.D., "Fake Reviewer Groups" Detection System." IOSR-JCE 2014.
8. SciKit (2018) SciKit Learn: 4.3. Preprocessing data. <http://scikit-learn.org/stable/modules/preprocessing.html>