

Detection of Faux Review using Collaborative Filtering and Stemming Algorithm

M.Rajeswari¹, R.Alamelu², J.Jayanisha³, V.Kalyani⁴

Associate Professor, Department of Information Technology, Panimalar Institute Of Technology, Chennai, Tamil Nadu, India¹

U.G. Student, Department of Information Technology, Panimalar Institute Of Technology, Chennai, Tamil Nadu, India^{2,3,4}

ABSTRACT: The new technologies are enabling people to have their work done efficiently. Such an evolution of technology is online marketplace. Now-a-days everyone buys a product based on the reviews given. People can make fake reviews for promotion of their own products; also competitive companies can try to damage each other reputation by negative reviews. Hence to overcome this problem we have proposed a system in which it allows only genuine users to give reviews. The reviews are accepted only after the successful authentication, provided by sending a transaction ID and OTP to the user along with the product details. It also finds whether the reviews are positive or negative reviews. Our system also removes rude words in the reviews.

KEYWORDS: Collaborative Filtering, Stemming Algorithm, Random Number Generation, Faux Reviews, Removal of Rude Words, OTP Generation.

I. INTRODUCTION

Retrograding the days where the customer buys product based on the recommendation of the friends, relatives and neighbours rating the product. The customer buys the product based on the rating given. However due to the rapid use of the Internet and the fussy lives of the humans made the evolution of ecommerce widely spread. Nowadays customer tends to buy a product in the online based on the reviews given. As many people rely only on the reviews. These reviews can be faux or genuine. People can make faux reviews for promotion of their own products, also competitive companies can try to damage each other reputation by giving fake reviews. To overcome this problem our project allows only genuine users to give reviews. The user has to create an account by filling the basic details such as name, emailid, and telephone no, and address. Secondly the user has to log in to the website and search for the product they need. After the successful payment for the product the admin sends the transaction ID and OTP to the user along with the product details through E-mail ID. By entering the correct OTP the user is allowed to give the reviews for the product that they purchased. The admin groups the users based on the similar interest of the users by the collaborative filtering algorithm. It also removes the rude words by the stemming algorithm.

II. LITERATURE SURVEY

Chengai Sun, Qiaolin Du, Gang Tian [1] The product reviews are being used by all the people for making the decisions in day today activities. But, for the sake of lucrative purposes, the reviewers are posting the faux reviews for either the promotion of their own products or for deprecating products. This paper proposes the convolutional neural network model to integrate the review using the product word composition model. To reduce the overfitting and high variance, a bagging model is introduced to bag the neural network model with two efficient classifiers.

Atefeh Heydari, Mohammadali Tavakoli, Naomie Salim, Zahra Heydari [2]: The contemporary world, online reviews play the significant role for the customer reviews on the product. The fraudsters activities mislead potential customers and organizations reshaping their business and opinion mining techniques form research accurate

conclusions. This paper is categorized into three groups to detect spam reviews, individual spammers and group spam.

Myleott, Yenjin Choi, Claire Cardie [3]: The rise in consumer reviews in online products are becoming the targets of opinion spam. This paper focuses on the deceptive opinion spam fictitious opinions that are sound authentic. Integrating work from psychology and computational linguistics. Comparing the three approaches to detecting deceptive opinion spam dataset. It also reveals the relationship between the deceptive opinions and imaginative writing.

Jiwei Lil, Myleott, Claire Cardie, Eduard Hovy [4]: In this paper, generalized method for checking the web deceptive opinion spam supported the quality dataset which is comprised of knowledge from three different domains, each of which contains three kinds of reviews i.e. customer generated truthful reviews, turker generated deceptive reviews and employee generated deceptive reviews. This approach tries to make a difference between the varied kinds of reviews made which is beneficial for the purchasers to make decisions.

Jitendra Kumar Rout, Anmol Dalmia, Kim-Kwang Raymond Choo [5]: As many consumers using online opinion reviews to make their decisions this has an economical impact on the line of business. This paper uses the semi-supervised learning methods are used to detect spam reviews.

Nitin Jindal, Bing Liu, Hady W. Lauw [6]: This paper is used to detect spam reviews or review spammers. It identifies the behaviours of review spammers and design these behaviors to detect these spammers. First, spammers may point the specific products to increase their effect. Second they need a bent to deviate from the other reviewers in their ratings of products. This method measures the degree of spam for each reviewer. With the help of a web based spammer evaluation.

Ahlan Alrehili, Kholood Albalawi [7]: This paper plays a major role of finding whether the review is positive or negative. This is done by Ensemble method. This also provides a better view of products to the business proprietors. This paper uses Random Forest Technique for achieving high accuracy.

Rakibul Hassan, Md. Rabiul Islam [8]: Generally, in online shopping the products are purchased by viewing the reviews posted for the particular product. So here they are finding whether given reviews are original or not. So that users can buy the products without any hesitation and they will know the true nature of product.

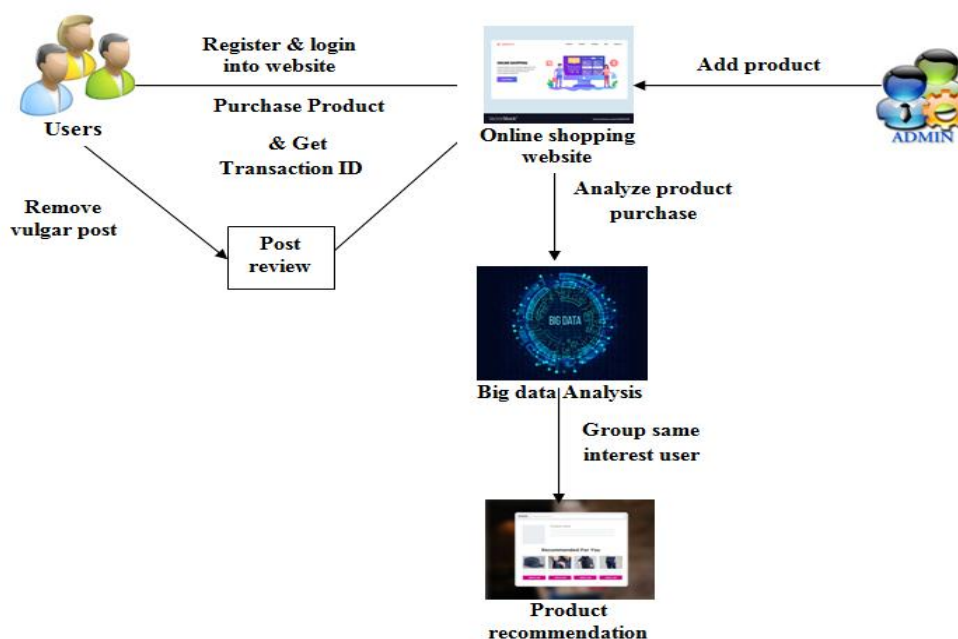


Fig.1 BLOCK DIAGRAM



A. COLLABORATIVE FILTERING

Collaborative filtering is used for making automatic prediction on the interest of the client.

It is used by the recommended system. It can be approached by two ways:

The narrow one and the general one. The main objective of this filtering is to make the best recommendation from someone with similar taste to themselves.

It requires :

- Users active participation.
- Easily represent users interest .
- Algorithms that match people with similar interest.

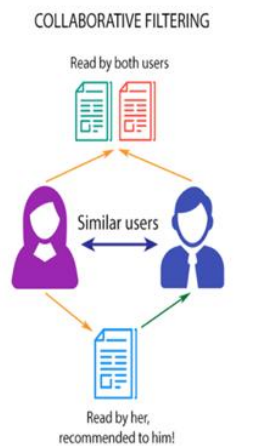


Figure 2. (Collaborative Filtering)

B. STEMMING ALGORITHM

Stemming is the process of removing or truncating the words to a root word. Stemming program are commonly referred to as Stemming algorithm or stemmers .

The two mainly occurred error in stemming algorithm is :

- Over stemming
- Under stemming

OVER STEMMING:

It occurs when two words are stemmed to same root that are different stems.

UNDER STEMMING:

It occurs when two words are stemmed to same root that are not different stems.

	original_word	stemmed_words
0	connect	connect
1	connected	connect
2	connection	connect
3	connections	connect
4	connects	connect

Figure 3: (Stemming Algorithm)

The submitted subject matter separated into part

1. User Registration
2. Rude Word Removal
3. OTP Generation and Verification
4. Best Product Recommendation
5. Product Recommendation and Grouping

IV. METHODOLOGY

A. USER REGISTRATION

In client side the user has to enter their user details Then user can login using the username and password which is given while registering. All the searched and needed items are added into the product list. User selects wanted items and it is added to cart products with count of the each item. A warning message will display in dialogue box when the customer type the number above the constraint value mentioned within the database. All selected items are displayed within the cart product list. Then purchase the required items.

B. RUDE WORD REMOVAL

In this module we remove the rude words in the review. Usually the reviews are either good or bad but some reviews contain the abusive words. Before posting the review the content of the review is checked if contains any rude words the review will get discarded .

C. OTP GENERATION AND VERIFICATION

A one-time password (OTP) may be a password that's valid for less than one login session or transaction. OTPs avoid variety of shortcomings that are related to traditional (static) passwords. The most important shortcoming that's addressed by OTPs is that, in contrast to static passwords, they are not vulnerable to replay attacks. This means that a possible intruder who manages to record an OTP that was already used to log into a service or to conduct a transaction won't be able to abuse it, since it'll be not valid. On the downside, OTPs are difficult for citizenry to memorize. Therefore they require additional technology to work. In this module we implement a system to send the transaction ID and OTP to the user through mail for the confirmation of purchase before they give the review about the product.

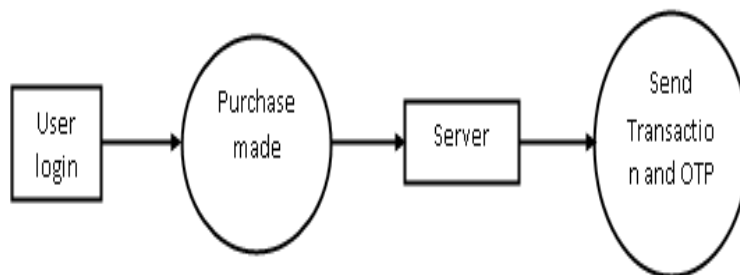


Figure 4: OTP Generation and Transaction Id

D. BEST PRODUCT RECOMMENDATION

Sentiment analysis is the process of gathering user review and stem the sensitive words from given review. Filter the keywords and fetch positive reviews from the user . When user search keywords on the websites, application will filter all the key and suggest the best product to customer.

E.PRODUCT RECOMMENDATION AND GROUPING

Based on the feedback value we suggest the product to the customer. Candidate can give keywords on the searching page based on that they will get product on the screen item sets can be generated efficiently with only two scans of database. So the user can give the feedback base product to purchase .this will be useful for the new user to by the product.User can give the review about the product only after sharing the transaction ID and OTP of that particular purchased product. And also our system will group the user who are all purchasing the same product and recommend that product to other members

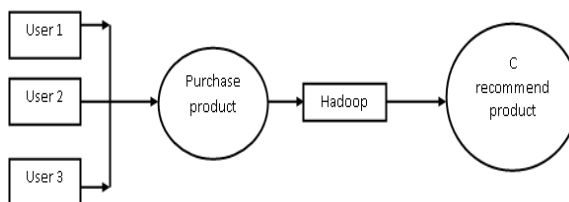


Figure 5: Grouping and Recommendation

V. CONCLUSION

Generally, it is difficult to find whether the given reviews in the particular website are original or fake reviews. But with the help of our project we can find whether the given reviews are fake or original . Also we can find the number of positive and the negative reviews . Then the rude Words are removed automatically. This is done by giving some sets of Rude words as input ,if it matches the reviews then the reviews will not be submitted. Also if anyone tries to post review for the product not purchased by the person , then that review will also not posted. We are using Collaborative filtering and Stemming algorithm .

REFERENCES

- [1] Rakibul Hassan , Rabiul Islam , “Detection of fake online reviews using semi –supervised and supervised learning”,2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019 .
- [2]Ahlam Alrehili ,Kholood Albalawi “Sentiment Analysis of Customer Reviews Using Ensemble Method” , 2019 International Conference on Computer and Information Sciences (ICCIS),3-4 february,2019 .
- [3]Jitendra Kumar Rout , Anmol Dalmia , Kim-Kwang Raymond Choo, “Revisiting Semi-Supervised Learning for Online Deceptive Review Detection” IEEE Access ,Volume: 5 ,18 January 2017.
- [4]Chengai Sun,1,2 Qiaolin Du,1 and Gang Tian “Exploiting Product Related Review Features for Fake Review Detection”, Mathematical Problems in Engineering 2016(1) ,Volume 2016 , 1-7 • January 2016.
- [5] AtefehHeydaria,MohammadaliTavakolia,NaomieSalima,ZahraHeydarib “Detection of review spam: A survey”, 2015 International Conference ,volume 42, May 1, 2015.
- [6] Jiwei Li1 , Myle Ott2 , Claire Cardie2 , Eduard Hovy1, “Towards a General Rule for Identifying Deceptive Opinion Spam” ,in proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics,(ACL),2014.
- [7] Myle Ott, Yejin Choi, Claire Cardie ,Jeffrey T. Hancock “Finding Deceptive Opinion Spam by Any Stretch of the Imagination” Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 309–319, Portland, Oregon, June 19-24, 2011. c 2011 Association for Computational Linguistics,volume-1,2012.
- [8] Nitin Jindal, Bing Liu, Hady W. Lauw, “Detecting Product Review Spammers using Rating Behaviors” In proceedings of the 19th ACM International Conference on Information and Knowledge Management(CIKM),2010.