



A Products Fake Review Detection and Deletion System

Ankita J. Patil¹, Tabbu S. Mulla², Gauri P. Dongare³, Sufiya B. Sheakh⁴, Sagar P. Mali⁵

Student, Computer Science and Engineering, AITRC, Vita, India¹

Student, Computer Science and Engineering, AITRC, Vita, India²

Student, Computer Science and Engineering, AITRC, Vita, India³

Student, Computer Science and Engineering, AITRC, Vita, India⁴

Assistant Professor, Computer Science and Engineering, AITRC, Vita, India⁵

Abstract: Consumer's reviews on ecommerce websites, online services, conditions and experience stories are useful for the stoner as well as the seller. The critic can increase their brand's fidelity and help other guests understand their experience with the product. Also reviews help the merchandisers gain further biographies by adding their trade of products, if consumers leave positive feedback on their product review. But unfortunately, these review mechanisms can be misused by merchandisers. Unlike the being work, rather of using a constrained dataset design chose to have a wide variety of vocabulary to work on similar as different subjects of datasets combined as one big data set. Sentiment analysis has been incorporated grounded on emoji's and textbook content in the reviews. Fake reviews are detected and distributed. The testing results are attained through the operation of Naïve Bayes, Linear SVC, Support Vector Machine and Random Forest algorithms.

Keywords: Fake Review, Naïve Bayes, Linear SVC, Support Vector Machine and Random forest algorithms.

I. INTRODUCTION

The data on the web is growing exponentially. Social media is generating a large quantum of data similar as reviews, commentary, and client's opinions on a diurnal base. This huge quantum of stoner generated data is empty unless some mining operations are applied to it. As there are a number of fake reviews, so opinion- mining fashion should incorporate Spam discovery to produce a genuine opinion. currently, there are a number of people using social media opinions to produce their call on hopping for product or service. Opinion Spam discovery is an exhausting and hard problem as there are numerous dummy or fake reviews that have been created by associations or by the people for colorful purposes. They write fake reviews to mislead compendiums or automated discovery system by promoting or distracting target products to promote them or to degrade their reports. The proposed fashion includes Ontology, Geo position and IP address shadowing, Spam words Dictionary using Naïve Bayes, Brand only review discovery and tracking account used. One of the veritably rapid-fire growth area is ecommerce. Generally-commerce give installation for guests to write reviews related with its service. The actuality of these reviews can be used as a source of information. For exemplifications, companies can use it to make design opinions of their products or services but unfortunately, certain parties who tried to produce fake reviews, both aimed at raising the fashion ability or to discredit the product abuse the significance of the review. They partake their studies on internet. Before coping anything, it's a normal mortal gets to do a check on that product. Grounded on reviews, guests can compare different brands and can finalize a product of their interest. These online reviews can change the opinion of a client about the product. However, also this can help the druggies to elect proper product that satisfy their conditions, If these reviews are true. On the other hand, if the reviews are manipulated or not true also this can mislead stoner. This boosts us to develop a system, which detects fake reviews for a product by using the textbook and standing property from a review. The honesty value and measure of a fake review will be measured by exercising the data mining ways. An algorithm could be used to track client reviews, through mining motifs and sentiment exposure from online client reviews and will also blocked the fake reviews.

II. LITERATURE REVIEW

Existing System

A E.Madhorubagan 1, Tamilselvi .V et. al presented as the trend to shop online is adding day by day and further people are interested in buying the products of their need from the online stores. This type of shopping doesn't take a lot of time of a client. Consumer goes to online store, search the item of his/ her need and place the order. But, the thing by which people face difficulty in buying the products from online store is the bad quality of the product. Client place the order only by looking at the standing and by reading the reviews related to the particular product. Similar commentary of other people is the source of satisfaction for the new product buyer. Then, it may be possible that the single negative review changes the angle of the client not to buy that



product. In this situation, it might possible that this one review is fake. So, in order to remove this type of fake reviews and with the original reviews and standing related to the products, project proposed a Fake Product Review Monitoring and junking System which is an Intelligent Interface and takes the Uniform Resource Locator related to products of Amazon, Flip kart and analyzes the reviews, and provides the client with the original standing. It's a unique quality of the proposed system that it works with the three-commerce Websites and not only analyzes the reviews in English but also the reviews written in Urdu and Roman Urdu. former work on fake reviews doesn't support point to dissect the reviews written in languages like Urdu and Roman Urdu and can't handle the reviews of multiple-commerce Websites. The proposed work achieved the delicacy of 87in detecting fake reviews of written in English by using intelligent literacy ways which is lesser than the delicacy of the former systems

B Somayeh Shojae, Azreen Azman et. al. said that the effectiveness of opinion mining relies on the vacuity of believable opinion for sentiment analysis. Frequently, there's a need to filter out deceptive opinion from the spammer; thus, several studies are done to descry spam reviews. It's also problematic to test the validity of spam discovery ways due to lack of available annotated dataset. Grounded on the being studies, experimenters perform two different approaches to overcome the mentioned problem, which are to hire evaluators to manually label reviews or to use crowd sourcing websites similar as Amazon Mechanical Turk to make artificial dataset. The data collected using the ultimate system couldn't be generalized for real world problems. Likewise, the former system of detecting fake reviews manually is a delicate task and there's a high chance of misclassification. In this paper, project propose a new fashion to annotate review dataset for spam discovery by furnishing further information and meta data about both reviews and pundits to the evaluators for effective spam reflection. Project proposed a frame and developed an on- line reflection system to ameliorate the review reflection process. The system is tested for several reviews from theamazon.com and the result is promising with0.10 error on labeling.

C Shilpa yadav, Dr.Gulbakshee Dharmela , Khushali Mistry explained Online reviews play a veritably important part in moment's e-commerce for decision- forest. Large part of the population i.e. guests read reviews of products or stores before making the decision of what or from where to buy and whether to buy or not. As writing fake/ fraudulent reviews comes with financial gain, there has been a huge increase in deceptive opinion spam on online review websites. Fake review or fraudulent review or opinion spam is an untruthful review. Positive reviews of a target object may attract further guests and increase deals; negative review of a target object may lead to lower demand and drop in deals. These fake/ fraudulent reviews are designedly written to trick implicit guests in order to promote/ hype them or defame their reports. Project work is end at relating whether a review is fake.

III. PROBLEM STATEMENT

People write unworthy positive reviews about products to promote them. In some cases vicious negative reviews to other (competitive) products are given in order to damage their character. Some of these consist of non-reviews (e.g., advertisements and elevations) which contain no opinions about the product. The first challenge then is, a word can be positive in one situation while being negative in any other situation. For example the word" long" in terms of a laptop's battery life being long is a positive opinion while the same word about the launch time is long is a negative opinion. This shows that the opinion mining system trained about words from opinions can't understand this nature of the word, giving a different meaning in different situations. Another challenge is that people do not always express opinions the same way. Utmost of the traditional textbook processing ways assume that small difference in textbook do not change the meaning much. still, in opinion mining ,e.g. the service was great, and the service wasn't great does make a huge difference. Eventually, in some cases, people give antithetical statements which were delicate to anticipate the nature of the opinion. There could be a retired positive sense in a negative review. And occasionally there's both positive and negative opinion about the product. An emotion factor can add a lot to what a person says or expresses. Adding a negative emoji to a positive comment or vice versa. In the millennial world of texting people have replaced long rulings with short forms and emoticons. These emoticons when used in textbook format are composed of punctuations and there's a good chance that they will be lost in data drawing process while preprocessing the textbook in opinion mining. After all these challenges, detecting the reviews that aren't genuine or which are used to diverge the consumers opinion in a certain direction becomes indeed more delicate. Opinion spamming or fake review discovery is therefore significant problem for ecommerce spots and other service providers as the consumer these days calculate largely on similar opinions or reviews.

IV. WORKING PROCESS

Overall Working

The high-level architecture of the implementation can be seen system architecture and the problem is solved in the following seven steps

- i.Data Collection
- ii.Data Preprocess
- iii.Feature extraction
- iv.Sentiment Analysis



- v. Fake Review Detection
- vi. Performance Evaluation and Deletion
- vii. Results

i. Data Collection

Consumer review data collection-Raw review data was collected from different sources, similar as Amazon, websites for reserving Airlines, Hotel and Restaurant, CarGurus, etc. reviews. Doing so was to increase the diversity of the review data. A dataset of 21000 was created

ii. Data Preprocess

Processing and enriching the data by junking of inapplicable and spare information as well as noisy and unreliable data from the review dataset.

- Step 1 judgment tokenization The entire review is given as input and it's tokenized into rulings using NLTK package.
- Step 2 junking of punctuation marks Punctuation marks used at the starting and ending of the reviews are removed along with fresh white spaces.
- Step 3 Word Tokenization Each individual review is tokenized into words and stored in a list for easier reclamation. Step 4 junking of stop words

iii. Feature extraction

The preprocessed data is converted into a set of features by applying certain parameters. The following features are uprooted

- Normalized length of the review-Fake reviews tend to be of lower length.
- Critic ID- A critic posting multiple reviews with the same Critic ID.

Standing-Fake reviews in utmost scripts have 5 out of 5 stars to allure the client or have the smallest standing for the competitive products therefore it plays an important part in fake discovery.

iv. Sentiment Analysis

Classifying the reviews according to their emotion factor or sentiments being positive, negative or neutral. It includes prognosticating the reviews being positive or negative according to the words used in the textbook, emoji's used, conditions given to the review and so on. Affiliated exploration shows that fake reviews have stronger positive or negative feelings than true reviews. The reasons are that, fake reviews are used to affect people opinion, and it's more significant to convey opinions than to plainly describe the data. The private vs ideal rate matters Advertisers post fake reviews with further objective information, giving further feelings similar as how happy it made them than conveying how the product is or what it does. Positive sentiment vs negative sentiment The sentiment of the review is anatomized which in turn help in making the decision of it being a fake or genuine review.

v. Fake Review Detection

Bracket assigns particulars in a collection to target orders or classes. The thing of bracket is to directly prognosticate the target class for each case in the data. Each data in the review train is assigned a weight and depending upon which it's classified into separate classes-Fake and Genuine.

vi. Performance Evaluation and Results

Comparison of the accuracies of various models and classifiers with enhancements for better results, as discussed in Accuracy Enhancements

V. SYSTEM ARCHITECTURE

5.1 System Architecture

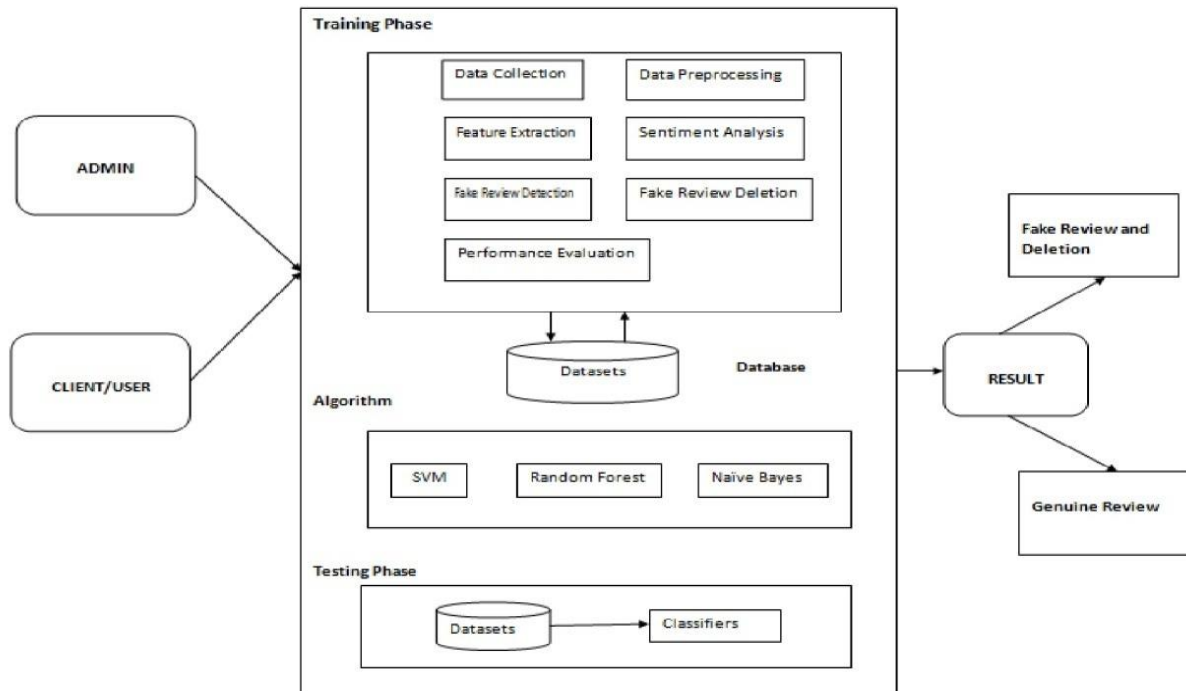


Fig 5.1 System Architecture

VI. ALGORITHMIC DETAILS

Technology Details

- **NLP based Text blob Classifier**

The two classifiers used in this configuration are:

a. Naive Bayes classifier

b. Decision Tree classifier

The experimental configuration for both classifiers was kept the same, and this section consists of the configurations used to set up the models for training the Python customer. Naïve Bayes and Decision Tree Classifier are used for detecting the genuine(T) and fake(F) reviews across a wide range of data set. The probability for each word is calculated is given by the rate of(sum of frequency of each word of a class to the total words for that class). The dataset is resolve into 80 training 20 testing, 16800 for training and 4200 for testing. Eventually, for testing the data using a test set where the probability of each review is calculated for each class. The class with the loftiest probability value using which the review is assigned the marker i.e. true/ genuine (T) or fake (F) Review. The datasets used for training areF-train.txt andT-train.txt. They include Review ID(for e.g. ID- 1100) as well as the Review textbook (Great product).

- **SKlearn Based Classifiers**

The Sklearn based classifiers were also used for classification and compared which algorithm to get better and accurate results.

a. Multinomial Naïve Bayes: Naive Bayes classifier is used in natural language processing (NLP) problems by predicting the tag of text, calculate probability of each tag of a text and then output choose the highest one.

b. LinearSVC: This classifier classifies data by providing the best fit hyper plane that can be used to divide the data into categories.

c. SVC: Different studies have shown If project use the default kernel in SVC (), the Radial Basis Function (rbf) kernel, then project probably used a more non-linear decision boundary on the case of the dataset; this would be vastly outperform a linear decision boundary

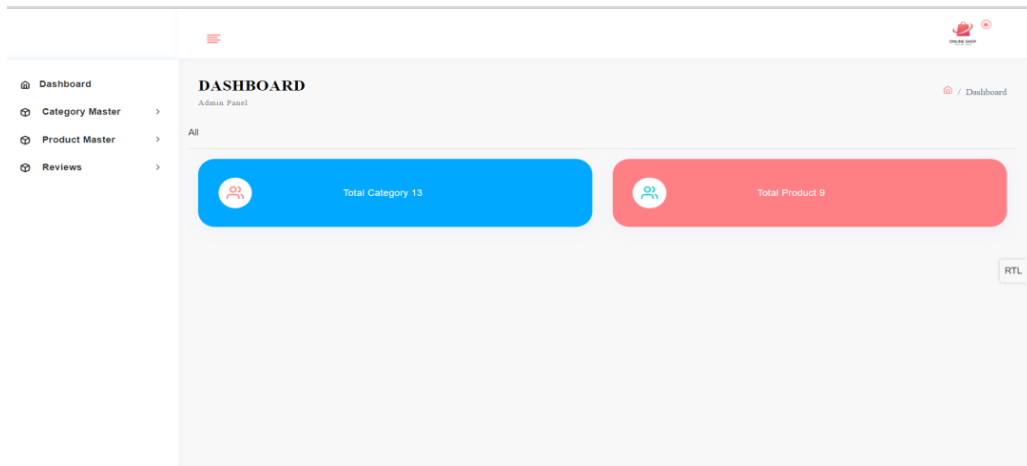
d. Random Forest: This algorithm has also been used for classifying which is provided by sklearn library by creating multiple decision trees set randomly on subset of training data.



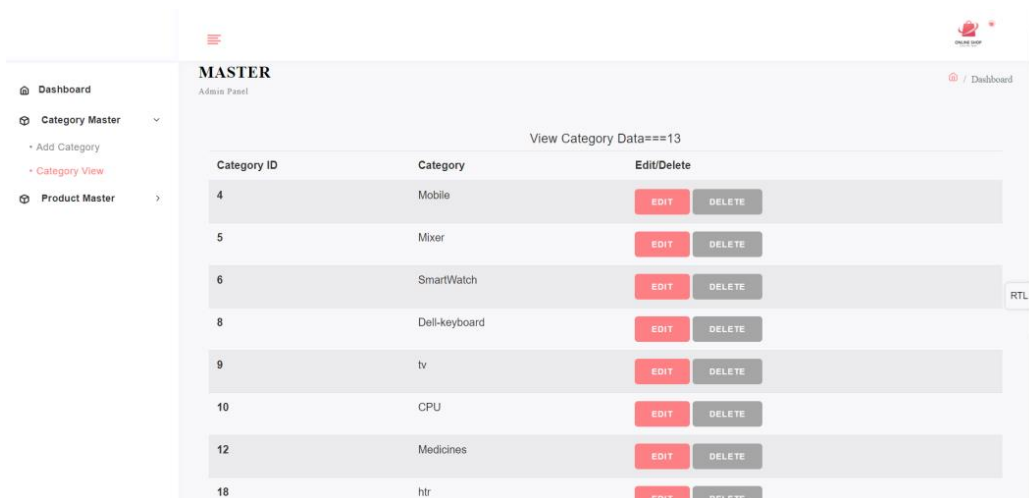
VII. RESULT



Home Page




Admin Dashboard



Category View





4
Redmi
₹10000
4GB Ram ,32Megapixel camera

Buy

Billing Address

First Name Last Name

E-mail Mobile No

Address Line 1 City

State Pin Code

Payment

Direct Check

Place Order

Buy Product

- Dashboard
- Category Master
- Product Master
- Reviews
 - Reviews
 - Delete Fake Review

DASHBOARD

Admin Panel

All

User	IP	Time	Review	Rating
9359856729	192.168.0.199	2023-03-02 11:20:35.208181	good	0.5
9078675645	192.168.0.199	2023-03-19 11:02:09.843287	good	0.5
9307238454	192.168.0.112	2023-04-21 17:26:26.269561	sigr w	4.5
9307238454	192.168.0.112	2023-04-21 17:36:20.885032	Easy to use	3.5
9307238454	192.168.0.112	2023-04-21 17:41:31.434340	nigi	4.5
9307238454	192.168.0.112	2023-04-21 17:49:13.133771	msejgr	1.5
9307238454	192.168.0.112	2023-04-21 17:59:39.553078	dfhfigh	3.5
9762379223	192.168.0.112	2023-04-22 12:20:29.976020	good one	2.5

DELETE FAKE REVIEW

Reviews

VII. CONCLUSION

In moment's world of e-commerce there's a strong need of relating fake reviews. Numerous of the papers are using machine literacy algorithm for chancing manipulated reviews. Further focus is given on the gets of the critic and different textbook parcels of commentary. So, n-gram approach is used to descry fake reviews for multiple languages. It's observed that the textbook categorization with SVM classifier is stylish approach for the discovery of fake reviews. Now days, as the technology is growing day by day and there are so numerous Websites and operations available in the online request by which dealer can vend their products and on that products there are millions of reviews available. There are some associations posting fake reviews for the products of the dealer in order to increase or drop the standing of the products. In this phase of design different algorithms can use for discovery of fake review. In coming phase discovery and complete omission of fake review is going to apply.

REFERENCES

- [1] E.Madhurabagan 1, Tamilselvi .V2, Sowmiya .M3, Sneha .S4 ,” Intelligent Interface for Fake Product Review Monitoring and Removal “ April 2021| IJIRT | Volume 7 Issue 11 | ISSN: 2349-6002
- [2] Shilpa yadav1, Dr.Gulbakshee Dharmela2 , Khushali Mistry3 “Fake Review Detection Using Machine Learning Techniques” 2021 JETIR April 2021, Volume 8, Issue 4 (ISSN-2349-5162)
- [3] Rajashree S. Jadhav , Prof. Deipali V. Gore “A New Approach for Identifying Manipulated Online Reviews using Decision Tree” Rajashree S. Jadhav et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1447-1450 ISSN: 0975-9646
- [4] MupparamSowjanya,K.Shnatilatha,Ch.hyma,K.Naresh “Implementation of fake product review monitoring system and real review generation by using data mining mechanism” Journal of Xi'an University of Architecture & Technology Volume XII, Issue II, 2020 ISS No : 1006-7930
- [5] Aishwarya M. Kashid1, Ankita K. Lalwani2, Samiksha S. Gaikwad3, Rajal A. Patil4, R. G. Sonkamble5, S. S. More “Fake Review Detection System Using Machine Learning” International Journal of Research in Engineering, Science and Management Volume-1, Issue-12, December-2018 www.ijresm.com | ISSN (Online): 2581-5792