ECOMMERCE FAKE PRODUCT REVIEW MONITORING AND REMOVAL SYSTEM

[1]Kharat Shital Dnyaneshwar, [2]Chandgude A.S,

[1]UG Scholar, Computer Engineering Department, S.N.D collage of engineering and research center, Yeola
[2]UG Scholar, Computer Engineering Department, S.N.D collage of engineering and research center, Yeola

kharaty81@gmail.com

ABSTRACT- As the trend to shop online is increasing day by day and more people are interested in buying the products of their need from the online stores. This type of shopping does not take a lot of time of a customer. Customer 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. Customer place the order only by taking a gander at the rating and by perusing the audits connected with the specific item. Such remarks of others are the wellspring of fulfillment for the new item purchaser. Here, it could be conceivable that the single negative audit changes the point of the client not to purchase that item. In the present circumstance, it may conceivable that this one survey is phony. Thus, to eliminate this kind of phony audits and give the clients the first surveys and rating connected with the items, we proposed a Fake Product Review Monitoring and Removal System (FaRMS) which is an Intelligent Interface and takes the Uniform Resource Locator (URL) connected with results of Amazon, Flipkart and Daraz and investigates the surveys, and gives the client the first appraising. It is a one of a kind nature of the proposed framework that it works with the three online business Websites and break down the surveys in English as well as the audits written in Urdu and Roman Urdu. Past work on counterfeit surveys doesn't uphold element to break down the audits written in languages like Urdu and Roman Urdu and cannot handle the reviews of multiple e-commerce Websites. The proposed work achieved the accuracy of 87 Percent in detecting fake reviews of written in English by using intelligent learning techniques which is greater than the accuracy of the previous systems..

Index Terms- Fake Reviews Detection, Text Classification, Natural Language Processing, Machine Learning, Bigrams, Term Frequency and Inverse Document Frequency

I. INTRODUCTION

In recent years, online reviews have been playing an important role in making purchase decisions. This is because, these reviews can provide customers with large amounts of useful information about the goods or service. However, to promote factitiously or lower the quality of the products or services, spammers may forge and produce fake reviews. Due to such behavior of the spammers, customers would be mislead and make wrong decisions. Thus, detecting fake (spam) reviews is a significant problem. Opinion spamming refers to the use of excessive and illicit methods, such as creating a large volume of fake reviews, in order to generate biased positive or negative opinions for a target product or service with the intention of promoting or demoting it, respectively. The reviews created for this purpose are known as fake, spam or bogus reviews, and the authors responsible for composing such deceptive content are known as fake or spam reviewers.

Individuals compose shameful positive surveys about items to advance them. At times malevolent negative audits to other (serious) items are provided to harm their standing. A portion of these comprises of non-audits (e.g., advertisements and advancements) which contain no sentiments about the item. The main test here is, a word can be positive in one circumstance while being negative in some other circumstance. For example "long" as far as a PC's battery duration being long is a positive assessment while a similar word about the beginning time is long is a negative assessment.

This shows that the opinion mining system trained about words from opinions cannot understand this nature of the word, giving a different meaning in different situations. Another challenge is that people don't always express opinions the same way. Most of the traditional text processing techniques assume that small difference in text don't change the meaning much. However, in opinion mining, e.g. the service was great, and the service wasn't great does make a huge difference. Finally, in some cases, people give contradictory statements which were difficult to anticipate the nature of the opinion. There could be a hidden positive sense in a negative review. And sometimes there is 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 sentences with short forms and emoticons. These emoticons when used intext format are composed of punctuations and there is a good chance that they will be lost in data cleaning process while preprocessing the text in opinion mining.

After all these challenges, detecting the reviews that are not genuine or which are used to deviate the consumers opinion in a certain direction becomes even more difficult. Opinion spamming or fake review detection is thus significant problem for ecommerce sites and other service providers as the consumer these days rely highly on such opinions or reviews.
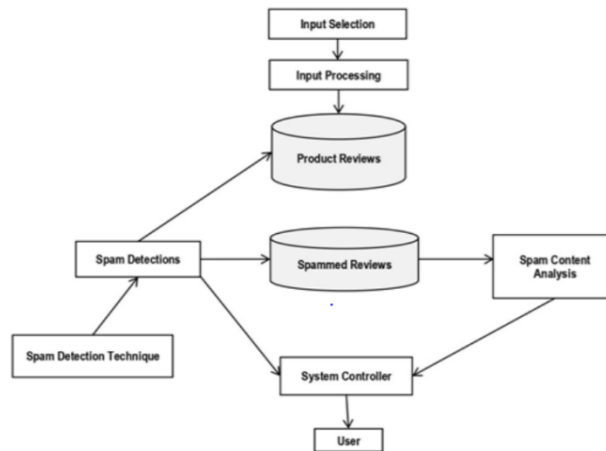
II. SYATEM ARCHITECTURE

Fig. 1- System Architecture

Detecting fake reviews from product is very much important in this era. As there are two types of purchases in every ecommerce Website which are verified purchase and Nonverified purchase. Verified purchase means that the customer who was writing the review purchased the product from the online store and who did not receive

the product at a great discount. To detect fake reviews, data is gathered from trip advisor, MTurk and Yelp. So that model can be trained in a best possible way. In this paper we also using SVM along with the sentiment analysis. There are two approaches to for detecting fake review one for verified purchase and other for non verified purchases.

1) Verified purchase

We are going to scrap the data of product from the e-commerce store and we will apply the sentimental analysis this will take the review words, rating of product and will classify genuine and fake review based on the polarity

2) Non-verified Purchase

Here first step is going to collect the data to train the model. Here we are going to collect the data from multiple sources about the reviews like n Spam Corpus, Trip Advisor, MTurkand yelp. After collection of data we need to process the data called preprocessing of data. Data preprocessing involves following methods

- Punctuation Removal
- Stop Words Removal
- Lemmatization
- N-gram Modeling

Next step is feature extraction in this step we are going to use Term Frequency (TF) and Term Frequency Inverse Document Frequency (TF-IDF) for the selection of features related to our dataset. In the proposed approach, system uses Count Vectorizer that converts each review into bag of words and is used to tokenize the set of words described in the reviews and after it system applies TF-IDF transformer. The model converts the collection of text documents into a matrix of token counts

Next step is classifier in this step we converts each review into 2-D matrix and then apply the TF-IDF transformer that gives weight to each word. After the feature selection, the last step in the classification process is to train the classifier. The proposed architecture is tested by applying three different supervised machine learning algorithms including SVM, Naïve Bayes, Logistic Regression but SVM outclass the remaining classifier with its performance and compete the other algorithms in terms of the results.

Last step is prediction and evolution -

Following tables will help use to understand the accuracy of the different model with different data set for different languages

Table 1 :Comparison of Classifiers for Yelp Reviews

Classifier	Accuracy
Support Vector Machine (SVM)	87%
Naïve Bayes	85%
Logistic Regression	81%

Table 2 :Comparison of Classifiers for Urdu Reviews

Classifier	Accuracy
Support Vector Machine (SVM)	70%
Naïve Bayes	68%
Logistic Regression	68%

Table 3 :Comparison of Classifiers for RomanUrdu Reviews

Classifier	Accuracy
Support Vector Machine (SVM)	69%
Naïve Bayes	66%
Logistic Regression	65%

I. CONCLUSION

In the proposed work, dataset is developed that contain Urdu and Roman Urdu reviews. It is difficult to detect fake reviews by yourself. So, N-gram approach is used to detected fake reviews for multiple languages. It is observed that text categorization with SVM classifier is best approach for the detection of fake reviews

REFERENCES

[1] A. Sinha, N. Arora, S. Singh, M. Cheema, and A. Nazir, "Fake Product Review Monitoring Using Opinion Mining," vol. 119, no. 12, pp. 13203– 13209, 2018.

[2] Torbet, Georgina. "U.S. Customers Spent over $6 Billion on Black Friday Purchases." Digital Trends, Digital Trends, 25 Nov. 2018, www.digitaltrends.com/web/shopping-totals-blackfriday/.

[3] Sterling, Greg. "Study Finds 61 Percent of Electronics Reviews on Amazon Are 'Fake'." Marketing Land, 19 Dec. 2018, marketingland.com/study-finds-61-percent-of electronics-reviews-on-amazon-are-fake 254055.

[4] K. Khan, W. Khan,A. Rehman, A. Khan, Asfandyar. Khan, A. Ullah Khan, B. Saqia, "Urdu Sentiment Analysis," (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 9, 2018.

[5] A. Mukherjee, B. Liu, and N. Glance, "Spotting Fake Reviewer Groups in Consumer Reviews, " 2012.