IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 7, July 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107040 |

Machine Learning for Identifying Inauthentic Reviews

Sonu S[1], Dr Raghavendra S P[2]

PG Student, Dept. of Master of Computer Applications, Jawaharlal Nehru New College of Engineering, Shivamogga, India

Assistant Professor, Dept. of Master of Computer Applications, Jawaharlal Nehru New College of Engineering, Shivamogga, India

ABSTRACT: A customer's decision to purchase a product or service are primarily influenced by online reviews. Customers use online reviews, which are valuable sources of information to understand the public opinion on products and/or services. Dependability on online reviews can give rise to the potential concern that violator could give deceitful reviews in order to synthetically promote or decry products and services. This practice is known as Opinion Spam, where spammers manipulate reviews by making fake, untruthful, or deceptive reviews to get profit and boost their products, and devalue a competitor's products. In order to tackle this issue, we propose to build a fraud risk management system and removal model. This captures fraudulent transactions based on user behaviour and network, analyses them in real-time using Data Mining, and accurately predicts the suspicious users and transactions. In this system, we use two algorithms Logistic Regression and Naïve Bayes to differentiate between fake and genuine reviews or feedback received by the customer.

KEYWORDS: Machine Learning, Linear Regression ,Naïve Bayes, Random Forest algorithms

I.INTRODUCTION

The internet is continuing to grow in size and importance, and hence the quantity and impact of online reviews is increasing continuously as well. Reviews can influence people of all areas but notably e-commerce is one of the most important aspect, where comments and reviews regarding products and services are often the most convenient and easy way for buyers to make a decision on whether or not to buy a product. Whenever we buy any product online on any online sites like amazon.com, flipkart.com, and many others, we first tend to check for the products reviews. If we think the product has nice and glorious reviews, we tend to immediately press the buy button, trusting these reviews. Even though most of these reviews are from genuine customers posting and sharing their feedback, there is a huge possibility that some of these reviews might be fake that are given to synthetically gain profit by attracting customers to these fake reviews.

II.RELATED WORK

Here we have selected few key literatures after exhaustive literature survey and listed as below:

Abhishek Punde, Sanchit Ramteke al [1], any E-commerce website gets bad reputation if they sell a product which has a bad review so in this paper it seen that sentimental analysis play a vital role to make business decision about product services.

Chengai Sun, Gang Tian et al [2], due to the purpose of profit, reviewers game the system by posting fake reviews for promoting or demoting the target products. This paper exploits the product related review features for fake review detection.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107040 |

Geng Cui, Hon-Kwong et al [3], this study examines the effect of online reviews on new product sales for consumer electronics and video games . the non cumulative data of sales rank of new product is present to significant challenge from this website.

Gyandeep Dowari, Divya Jyothi et al [4], made a literature survey on online product review on shopping experience in social media as promoted user to provide a feedback. In this proposed system the detailed discussion about the existing techniques to find out the reviews spam or not.

Huayi Li et al [5], discussed how online audits have turned into an undeniably significant asset for direction and item planning. However, surveys frameworks are frequently battered by assessment spamming.

Manoj Hudnurkar, Akash et al [6] , the purpose of this paper is to gain insights from the online product review of e-commerce sights such as online shoppings. the sentimental analysis was used in this study it is not limited to it.

Piyush Jain , Mihir Jain et al [7], the data on the web is growing exponentially. Social media is generating a large amount of data such as reviews ,comments and customer opinion on a daily basis , so in this article the detailed discussion about the existing techniques to find out the reviews spam or not.

R Divya ,S Hema et al [8], evaluation in the websites this fake product review monitoring and removal for genuine reviews through ip address tracking application is delivered. The use of this proposed system the people will receive valuable goods from a trust worthy websites.

Shashank Kumar ,Anupam Goel et al [9],most of the consumers prefer e-commerce ,because of the inacurative offers. In this paper, they incorporate sentiment analysis of reviews techniques into the spam review detection.

Yuming Lin et al [10], depicts that recognizing spam surveys is significant for current online business applications. In his paper, he and his group investigate the issue on counterfeit audit diminish online assessment spam. The attributes of phony audits, right off the bat, are analyzed.

III.PROBLEM STATEMENT

Spammers might furnace and make fake reviews in order to artificially promote or devalue the product's quality and services. Now, customers could be deceived to make wrong decisions because of such behaviour of spammers. Therefore, detecting fake reviews is significantly important.

IV.DESIGN AND IMPLEMENTATION

To solve the major problem faced by online websites due to opinion spamming,this project proposes to identify any such spammed fake reviews by classifying them intofake and genuine. The method attempts to classify the reviews obtained from freelyavailable datasets from various sources and categories including service based, productbased, customer feedback, experience based and the crawled Amazon dataset with agreater accuracy using Naïve Bayes, Linear SVC, SVM, Random forest , DecisionTrees algorithm.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107040 |

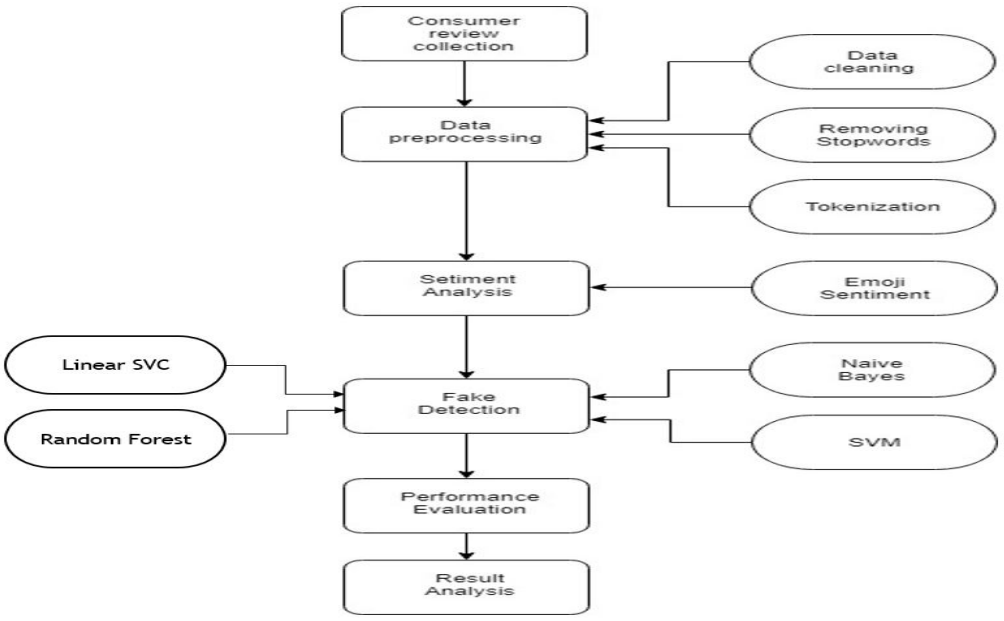

Figure1:Flow Chart of the system

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

Data Preprocess :
Processing and refining the data by removal of irrelevant and redundant information as well as noisy and unreliable data from the review dataset.

Feature extraction:
The preprocessed data is converted into a set of features by applying certain parameters. The following features are extracted.

Sentiment Analysis:
Classifying the reviews according to their emotion factor or sentiments being positive, negative or neutral. It includes predicting the reviews being positive or negative according to the words used in the text, emojis used, ratings given to the review and so on.

Fake Review Detection:
Classification assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. Each data in the review file is assigned a weight and depending upon which it is classified into respective classes - Fake and Genuine.

Performance Evaluation and Results:
Comparison of the accuracies of various models and classifiers with enhancements for better results, as discussed in Accuracy Enhancements.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107040 |

	A	B	C	D	E	F	G
1		review_date	review_rati	review_title	review_text	verified_purchase	
2	0	01-01-2019	5	Dove Men	As you get c	TRUE	
3	1	01-02-2019	5	Great for a	Three gigan	TRUE	
4	2	01-02-2019	4	Vitamin B12	Excellent	TRUE	
5	3	01-02-2019	5	A Very Hanc	A great flav	TRUE	
6	4	01-02-2019	5	Very handy	Does what i	FALSE	
7	5	01-02-2019	4	Vitamin B12	Excellent	TRUE	
8	6	01-02-2019	5	Great for a	Three gigan	TRUE	
9	7	01-02-2019	5	A Very Hanc	A great flav	TRUE	
10	8	01-05-2019	4	nearly 5 sta	Excellent ite	FALSE	

Figure2:Data Preprocessing Model

V.RESULT AND DISCUSSION

As we discussed above the method is implemented to remove the fake product review from the dataset. By using the Naïve Bayes algorithm we can get the most accuracy , F-measure, Precision and recall values . The accuracy of the system is around 80% percent. As we discussed if we give an example to find positive or negative. “Shirt is of very good quality” for this sentence we have to do pre-processing first. There the punctuations are removed from the sentence. Then the feature extraction process. There the sentence is compared with the stopwords then “is”, “of”, words were removed from the sentence.

Figure 3: To Detect Review Counts

In the following diagram predicts the accuracy of fake and real reviews from datasets. The Processing and refining the data by removal of irrelevant and redundant information as well as noisy and unreliable data from the review dataset.Which are devided by number of rows and columns, which is 2000 above columns and 15 rows present in the dataset. Classifying the reviews according to their emotion factor or sentiments being positive, negative or neutral. It includes predicting the reviews being positive or negative according to the words used in the text, emojis used, ratings given to the review and so on.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107040 |

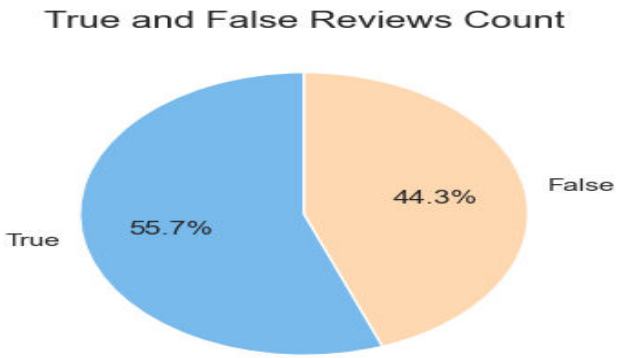

Figure 4: Predicting True and False Reviews

The following diagram is to compare the accuracies of various models and classifiers with enhancements for better results, as discussed in accuracy enhancements. Comparison of the accuracies of various models and classifiers with enhancements for better results, as discussed in Accuracy Enhancements.By predicting the reviews from the dataset which are of 55.7% true revies and 44.3% false reviews.

VI.SNAPSHOTS

Figure 5:Fake product Review Monitoring

REFERENCES

[1] Abhishek Punde, Sanchit Ramteke, “Fake product review Monitoring& Removal and Sentimental Analysis of Guenuine Reviews”. International Journal of Engineering and Management Research vol.9,2019.

[2] Chengai Sun, Gang Tian, “Exploiting Product related Review features for Fake review Detection”. Hindawi Publishing Corporation vol.2,2016.

[3] Geng Cui, Hon-Kwong , ”The effect of Online Consumer Reviews on new Product Sales”. International Journal of Electronic Survey vol.17,2012.

[4] Gyandeep Dowari, Divya Jyothi, ”Fake Product Review Monitoring and Removal using Opinion Mining”.UGC Care Group I Listed Journal vol.9 2020.

[5] Huayi Li , Aishwarya M,”Fake Review Detection System Using Machine Learning”. International Journal of Research in Engineering, Science and Management vol.19 2018.

[6] Manoj Hudnurkar, Akash , “Online Product Reviews and Their Impact on Thid Party Sellers Using Natural Language Processing”. International Journal of Business Inteligence Research vol.12 2021.

[7] Piyush Jain , Mihir Jain, “Fake Product Review Monitoring System”. International journal of Trend in Scientific Research and Development vol.3 2019.

[8] R Divya ,S Hema, “Fake Product Review Monitoring and Removal for Genuine Online Product Review Using IP Address Tracking”. International Journal od Advanced Research in Electrical, Electronics and Instrumentation Engineering vol.9 2020.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107040 |

[9] Shashank Kumar ,Anupam Goel, "Research on Product Review Analysis and Spam Review Detection". International Journal of Advances in Engineering and Management vol.5 2017.

[10] Yuming Lin, Goel, "Fake product Reviews Monitoring and Detecting Spam".Internation Research journal management vol.9 2014.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details