IJIRCCE

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

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906120 |

Spam Reviews Detection System Using Machine Learning

Sagar Bhor[1], Vivek Vidhate [2], Dipali Pawar [3], Nikita Karpe [4], Prof. Nitin Shivale [5]

U.G. Student, Department of Computer Engineering, BSIOTR, Pune University, Pune, India[1]

U.G. Student, Department of Computer Engineering, BSIOTR, Pune University, Pune, India[2]

U.G. Student, Department of Computer Engineering, BSIOTR, Pune University, Pune, India[3]

U.G. Student, Department of Computer Engineering, BSIOTR, Pune University, Pune, India[4]

Associate Professor, Department of Computer Engineering, BSIOTR, Pune University, Pune, India[5]

ABSTRACT: People often place their trust in products based on product reviews and ratings. Reviews can have an impact on a company's or a brand's profile. The company must examine market reactions to its products. Popular reviews, on the other hand, are difficult to track and arrange. In social media, many public viewpoints are difficult to manually process. The next step is to develop a mechanism for automatically categorizing favourable and negative public feedback. Customers will be able to see how the product performs in terms of consistency, efficiency, and guidance, which will provide prospective buyers a better knowledge of the product. The applicability of web assessments from suppliers in order to fulfil client requirements by evaluating beneficial input is one such unfulfilled possibility. Good and bad reviews are important in determining customer demands and gathering product feedback from customers. Sentiment Analysis is a type of computer analysis that collects contextual information from text. A large number of internet mobile phone ratings are studied in this study. We divided the text into positive and negative categories, as well as feelings of disappointment, expectation, disgust, trepidation, delight, regret, surprise, and confidence. This clearly defined category of feedback aids in a comprehensive evaluation of the product, allowing consumers to make better decisions..

KEYWORDS: Machine Learning, Social Media, Text Mining, Text Classification, Sentiment Analysis, and Online Reviews .

I. INTRODUCTION

Many businesses and software sectors store their data in Social networking creation provides the customer with an ability to share his or her views. That means the organization can't monitor the contents of the virtual universe now. Complaints in social media are submitted by customers who are not pleased by a company's services or goods. On the other hand, consumers are still optimistic for a commodity in the social media. This view could affect other potential clients, including positive or negative ones. Potential consumers can find out about a certain product before deciding to purchase goods.

An appraisal of the sentiment is expected to immediately decide whether the feeling is negative or positive. Feeling analyses are a subset of text mining that focuses in the text of a person's feeling, mood and attitude. The fundamental theory of sentiment analysis consists of categorizing the polarity of texts and determining whether they are positive or negative.Sentiment analyses are commonly used as rapid social network growth. For different places public opinion is becoming really critical. There have been some difficulties in collecting public examination.

Many product evaluation pages have recently been published on the Internet. It invites scientists to carry out a consumer review sentiment analysis. On product evaluations, customer opinion was evaluated in this paper.

II. RELATED WORK

In this paper [1], authorprovides a method for detecting false product reviews that combines content and usage data. The suggested approach takes advantage of both product reviews and the behavioural qualities of reviewers, which are linked via specific spam indications. To properly assess reviews generated over “suspicious” time intervals, fine-grained burst pattern recognition is used in this research. The reviewer's previous reviewing history is also used to

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906120 |

determine the reviewer's overall "authorship" reputation as an indicator of the authenticity degree of their most recent reviews.

This study [2]investigates the effects of online customer evaluations on consumer agility and, as a result, product performance using a big data analytical method. Using large-scale customer review texts and product release notes, the authors construct a singular value decomposition-based semantic keyword similarity method to evaluate consumer agility. Our empirical analysis demonstrates that review volume has a nonlinear relationship with customer agility, using a mobile app data set with over 3 million online reviews. Furthermore, consumer agility and product performance have a curved relationship. This study adds to the body of knowledge in the field of innovation by proving the impact of a company's capacity to use online customer feedback on product performance. It also aids in the resolution of contradictions in the literature concerning the links between the three components.

In this research [3], authorsobserved that reviewers' posting rates (the number of reviews they write in a given period of time) follow an unusual distribution pattern that has never been documented before. That is, their rates of posting are bimodal. Multiple spammers also have a tendency to publish reviews to the same set of products in a short period of time, a practice known as cobursting. Furthermore, the author discovered some intriguing patterns in the temporal dynamics of individual reviewers as well as their co-bursting behaviors with other reviewers. The authors suggest a two-mode Labeled Hidden Markov Describe to model spamming using only individual reviewers' review posting times, based on their findings. The Coupled Hidden Markov Model is then used by the authors to capture both reviewer posting habits and co-bursting signals.

Authors [4] This study proposes using statistically based features that are modelled through the supervised boosting approach such as the Extreme Gradient Boost (XGBoost) and the Generalized Boosted Regression Model (GBM) to evaluate two multilingual datasets in order to improve the detection of opinion spams in the mobile application marketplace (i.e. English and Malay language). According to the results of the evaluation, the XGBoost is best for detecting opinion spams in the English dataset, while the GBM Gaussian is best for the Malay dataset. The application of the proposed statistical based features yielded a detection accuracy rate of 87.43 percent on the English dataset and 86.13 percent on the Malay dataset, according to the comparative study.

III. METHODOLOGY

The first step is to identify and calculate spammer behavioral features in an unlabeled Amazon review dataset. This calculation is carried out on all dataset reviews based spam review detection using behavioral features method.

IV. SYSTEM ARCHITECTURE

The Fig.1 shows the proposed system architecture.

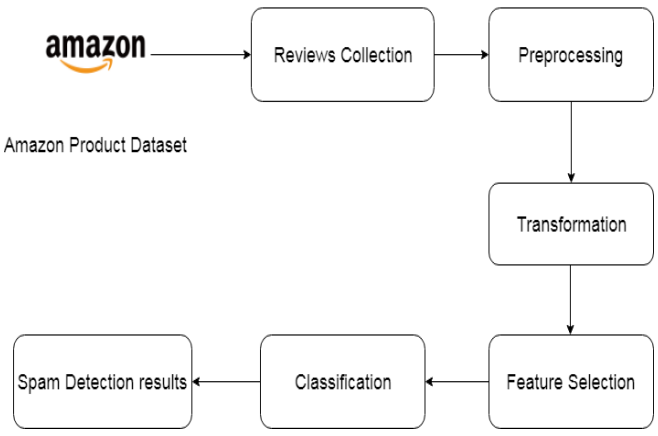

Fig 1. System Architecture

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906120 |

- SpamDup framework that is a novel network based approach which models review networks as heterogeneous information networks.
- A new weighting method for spam features is proposed to determine the relative importance of each feature and shows how effective each of features are in identifying spams from normal reviews.

• The SpamDup framework outperforms the state-of-the-art in terms of time complexity, which is heavily influenced by the number of features used to detect a spam review.

Our suggested framework's fundamental notion is to describe a given review dataset as a Heterogeneous Information Network (HIN) and transfer the challenge of spam detection into a FIN classification task. In particular, a model review dataset in which reviews are linked together using various node kinds. The flowchart of the SpamDup framework is shown in Figure 2.

Algorithm:

Feature selection:

Algorithm 1: Spam review detection using behavioral features method

```
Input: review Ri , τ = 0.5,0.55,0.6 //threshold value for labelling the review
Output: Spam or Not-Spam
1.  for each review Ri in review dataset do
2.      // behavior features (F1, F2, F3, ..., F13)
3.      for each behavior feature Fi calculate normalize value do
4.          // variable V is calculating normalize value of F
5.          Vi = calculate normalize value Fi
6.          Sum += Vi
7.      end for
8.      // calculating average score
9.          Average Score = Sum / 13
10.     for each value Vi do
11.         // calculating drop score
12.         DropScore = (Sum - Vi) / 12
13.         if | Average Score – DropScore | >= 0.05 then
14.             assign weight Wi ◄— 2
15.             Total Weight += 2
16.         else
17.             assign weight Wi ◄— 1
18.             Total Weight += 1
19.         end if
20.     end for
21.     for each value Vi do
22.         // calculating total spam score
23.         Score += Wi * Vi
24.     end for
25.     Spam Score = Score / Total Weight
26.     if Spam Score > τ then
27.         label Ri ◄— Spam
28.     else
29.         label Ri ◄— Not-Spam
30.     end if
31. end for
```

V. RESULT AND DISCUSSION

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and Jdk 1.8. The application is web application used tool for design code in Eclipse and execute on Tomcat server.

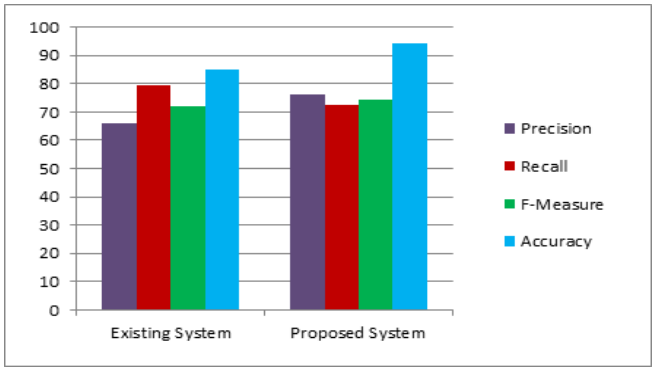

Fig 2. Performance Analysis between existing and proposed system

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906120 |

The proposed SpamDup framework time complexity is $O(e^2 n)$. The SpamDup framework accuracy is 94.06\% which is better than SPaglePlus Algorithm accuracy is 85.14\% on using product dataset.

VI. CONCLUSION

Sentiment Analysis is a case study that looks at the feeling, mood, entropy or feelings of people. This paper addresses a basic issue of the study of feelings and the classification of feelings of polarity. Data was compiled from online product reviews of Amazon.com. A method known as the categorization of emotion polarity and POS along with thorough explanations of each phase was proposed. These measures include pre-processing, pre-filtering, partitioning, data consistency. Functionality that include machine learning expertise. Much work has been done in opinion mining and consumer evaluation in the form of a study of documents, sentences, and features. Opinion Mining can become a most interesting field of study for potential preferences by using a number of found function expressions derived from the reviews. More novel and successful approaches need to be invented to address the existing difficulties of mining opinion and sentiment analysis.

REFERENCES

[1] Dematis, E. Karapistoli, and A. Vakali, "Fake review detection via exploitation of spam indicators and reviewer behavior characteristics,'' in Proc. Int. Conf. Current Trends Theory Pract. Inform. Cham, Switzerland: Edizioni Della Normale, 2018, pp. 581–595.

[2] S. Zhou, Z. Qiao, Q. Du, G. A. Wang, W. Fan, and X. Yan, "Measuring customer agility from online reviews using big data text analytics," J. Manage. Inf. Syst., vol. 35, no. 2, pp. 510–539, Apr. 2018.

[3] H. Li, G. Fei, S. Wang, B. Liu, W. Shao, A. Mukherjee, and J. Shao, "Bimodal distribution and co-bursting in review spam detection," in Proc. 26th Int. Conf. World Wide Web (WWW), 2017, pp. 1063–1072.

[4] M. Hazim, N. B. Anuar, M. F. A. Razak, and N. A. Abdullah, "Detecting opinion spams through supervised boosting approach," PLoS ONE, vol. 13, no. 6, 2018, Art. no. e0198884.

[5] N. Kumar, D. Venugopal, L. Qiu, and S. Kumar, "Detecting review manipulation on online platforms with hierarchical supervised learning," J. Manage. Inf. Syst., vol. 35, no. 1, pp. 350–380, Jan. 2018.

[6] N. Hussain, H. Turab Mirza, G. Rasool, I. Hussain, and M. Kaleem, "Spam review detection techniques: A systematic literature review," Appl. Sci., vol. 9, no. 5, p. 987, 2019.

[7] C. Pandey and D. S. Rajpoot, ''Spam review detection using spiral cuckoo search clustering method,'' Evol. Intell., vol. 12, no. 2, pp. 147–164, Jun. 2019.

[8] R. Narayan, J. K. Rout, and S. K. Jena, "Review spam detection using opinion mining,'' in Progress in Intelligent Computing Techniques: Theory, Practice, and Applications. Singapore: Springer, 2018, pp. 273–279.

[9] R. Barbado, O. Araque, and C. A. Iglesias, "A framework for fake review detection in online consumer electronics retailers," Inf. Process. Manage., vol. 56, no. 4, pp. 1234–1244, Jul. 2019.

[10] R. Ghai, S. Kumar, and A. C. Pandey, "Spam detection using rating and review processing method," in Smart Innovations in Communication and Computational Sciences. Singapore: Springer, 2019, pp. 189–198.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

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