



IJIRCCCE

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



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 5, May 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Fake Product Review Monitoring System

Prof. Anu C S¹, Rohit N Hiremath², Sahana A E³, Shashank S⁴, Sneha G R⁵

Assistant Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering & Technology,
Davangere, Karnataka, India¹

B.E Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering & Technology,
Davangere, Karnataka, India²

B.E Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering & Technology,
Davangere, Karnataka, India³

B.E Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering & Technology,
Davangere, Karnataka, India⁴

B.E Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering & Technology,
Davangere, Karnataka, India⁵

ABSTRACT: People read reviews on websites because they want to know all of the pros and cons of a product before they buy it. The fact that these reviews are out there can be used as a way to learn more. Online sellers often post fake reviews that are either positive for their own products or negative for their competitors' products in order to make money. The goal of this project is to use opinion mining and latent semantic analysis to find fake product reviews. By keeping track of the user's IP address, reviews from the same IP address that promote or criticise the same product are deleted. So are reviews that are posted in a flood by the same person.

KEYWORDS: Opining Mining; Latent Semantic Analysis Demoting; Fake Reviews

I. INTRODUCTION

E-commerce is an area that is growing very quickly. Most e-commerce sites have a way for customers to write reviews about the services they offer. The fact that these reviews are out there can be used as a way to learn more. Before buying something, it is common for people to do research on that product. Customers can compare different brands based on reviews and choose the one that best fits their needs. Customers can change their minds about a product based on what they read about it online. But the customer doesn't know if these reviews are real or not. Some review sites have good reviews added by the company that makes the product. This gives a false impression that the product is good. They write good reviews for many of the products that their company makes. The customer won't be able to tell if the review is real or made up. This "Fake Product Review Monitoring System" is put in place to find fake reviews on the site. This system will find fake reviews made by spammers who post fake comments about a product by finding the spammer's IP address.

II. RELATED WORK

Opinion Mining has been the subject of a lot of research in the past. But there hasn't been a lot of work done in this area. Review: Spam is hard to find unless you read it by hand. Here are some of the things that were planned and done before.

In their paper "A framework for fake review detection in online consumer electronics retailers," Barbado R, Araque O, and Iglesias CA wrote, "Scraping processing is used to build the data set from Yelp, and Fake Feature Framework is used to organise the extraction and characterization of features in fake detection." Their framework is made up of two main types of features: those that focus on reviews and those that focus on users. Review-centric features only have to do with the review text, while user-centric features focus on how the user acts on the site. [1]

The paper "A survey on online review SPAM detection techniques" by Rajamohana SP, Umamaheswari K, Dharani M, and Vedackshya R used Naive Bayesian classifier, which led to a very low error rate. [3]

Elmurngi E. and Gherbi A. wrote a paper called "Detecting fake reviews through sentiment analysis using machine learning techniques." In it, they say, "Spam detection technique using J48 Algorithm is used to check for spams in the reviews, and we find that J48 algorithm has more accuracy than Naive Classifier." [4]

The paper, "Detection of Spam Reviews and Spammers in E-Commerce Sites that Used Naive Bayesian Classifier," was written by Kotian H. and Meshram BB. It had a very low error rate. [5]

The paper "Detecting Fake Reviews Using Semantic and Emotion Model" was written by Li Y, Feng X, and Zhang S. It proposed three new types of features, such as review density, semantics, and emotions, and gave the model and algorithm for building each of these features. Even so, it is not a good measure, and the change is not very big. [6]

The paper "Online review spam detection by new linguistic features" by Karami A. and Zhou B. proposes to use categories of lexical, semantic, and linguistic features to find spam reviews on the web.[7]

"Fraud detection in online reviews using machine learning techniques" is the title of a paper that Shivagangadhar K, Sagar H, Sathyan S, and Vanipriya Ch wrote. In it, they suggested that linguistic features like the presence of unigrams, the frequency of unigrams, the presence of bigrams, the frequency of bigrams, and the length of a review could be used to build a model and find fake reviews. But the main problem is that there isn't enough data, and it needs both linguistic and behavioural features. [8]

"An opinion spam analyzer for product reviews using supervised machine learning method" is the title of a paper that Mevada D. and Daxini V. wrote. It said, "Used SVM with a small number of support vectors that can have good generalisation, even when the dimensionality of the data is high." [9]

In their paper "Sentiment analysis using product review data," FangFang X and Zhan J proposed using POS tagging to read text and figure out the part of speech for each token. The sentiment score is then used to show the polarity of a review's sentiment with 57.2% accuracy. [10]

The paper "Fake review and brand spam detection using J48 classifier" by Kokate S. and Tidke B. proposed a behavioural approach to find review spammers who try to change the ratings on some target products.

Reviewers are ranked based on how their behaviour is scored as a whole.[11]

Kolhe NM, Joshi MM, Jadhav AB, and Abhang PD wrote a paper called "Fake reviewer groups' detection system." In it, they said that it was hard to spot individual fake reviews, but it was easier to spot groups. To do this, they used a method called "frequent item set mining" (FIM) to look at the dataset. [12]

The paper "Fake review detection" by Mukherjee A, Venkataraman V, Liu B, and Glance N said that they compared real-life filtered (fake) reviews and unfiltered (not fake) reviews on Yelp. The results showed that it is much harder to classify real-world data, with only 67.8 percent accuracy. [13]

The paper "Spotting and Removing Fake Product Review in Consumer Rating" by Reddineelima C, Haritha V, Dinesh U, Kalpana B, and Kumar PN suggested that a fake review could be found by finding the same IP address of the user ID more than once.[14]

Ioannis D. wrote in the paper "Fake Review Detection via Exploitation of Spam Indicators and Behaviour Characteristics" that Evaluation of the used Multiplayer Perceptron (MLP) classifier showed high accuracy in detecting review spam based on content. MLP can be used to review text and decide whether it is spam or real. [15]

III. LITERATURE SUMMARY

Based on the review of the literature, it could be said that the process of figuring out if a review is real or fake in the current system is not very good because there isn't much information that can be used. Sometimes people think that fake reviews are just as good as real ones. There is a lot of worry about the effects of fake reviews, and many studies have been done to show what those effects are. But, as far as we know, no study has yet looked into why and how people, review platforms, and customers post fake reviews in detail. Review spam is hard to spot unless it is read by hand. Features like the number of reviews, their meaning, and how they make people feel have also been used to make models, but this is not a good metric. To build a model and find fake reviews, you need linguistic features like unigram presence, unigram frequency, bigram presence, bigram frequency, and review length. The main problem is that there isn't enough data, and you need both linguistic and behavioural features. A lot of research has been done on opinion mining, where the opinion of a customer is linked to a piece of text.

Existing System

Online reviews have become a big part of how people decide what to buy. The current system lets customers write reviews about the products they've bought, but it doesn't have a way to spot and get rid of fake reviews. Some spammers will make up and post fake reviews to boost or lower the quality of a product.

Problem Statement

Customers would be led astray by spammers' fake reviews. These fake reviews trick customers into buying and have a direct effect on sales. In the proposed system, the IP address is used to tell if a review about a certain product is real or not.

Proposed Solution

In the proposed system, it is clear whether a review about a certain product is real or not. By tracking the IP address of the user, reviews that are posted in a flood by the same person are deleted. This also applies to reviews that are posted by the same person more than once. This would help other people who use the e-commerce platform buy something worthwhile.

Proposed Objectives

The objectives of the proposed approach are:

- To acquire various datasets for differentiating positive and negative reviews.
- To clean the dataset by handling missing values and outliers.
- To understand and analyse the correlation among the dataset features and feature selection for training and testing models.
- To detect spam reviews by Sentimental analysis and Latent semantic analysis.
- To present opining mining on spam filtered data.

IV. PROPOSED METHODOLOGY

A. System Design:

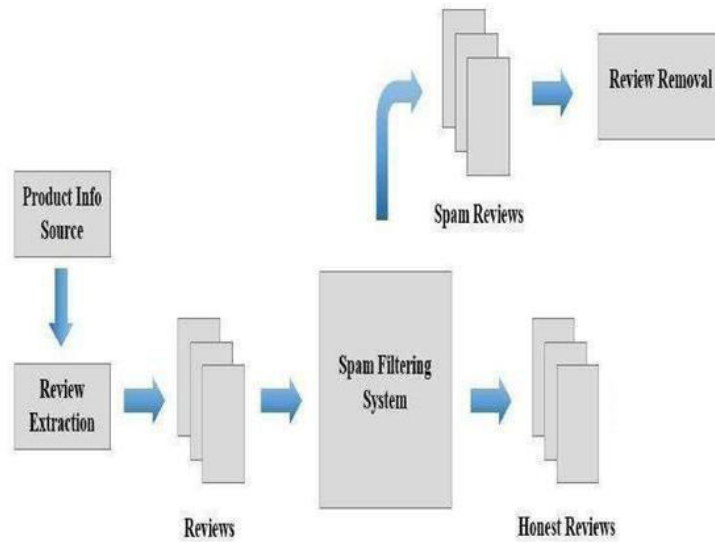


Fig: System Architecture overview

B. Proposed Methodology:

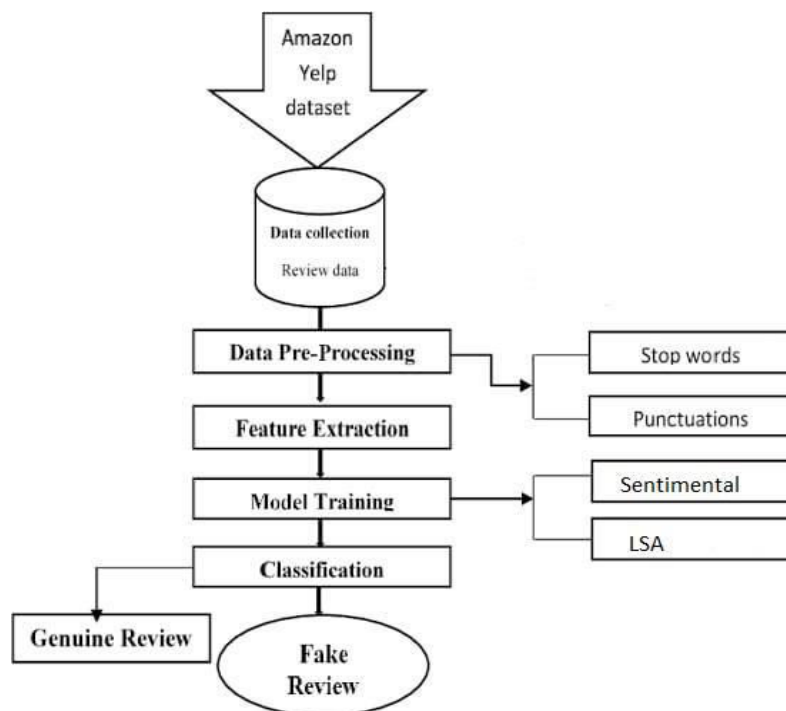


Fig: Methodology Diagram

Detailed Description of The Methodology

- **Data Acquisition:** In this step, we make an in-house data set of spam reviews and reviewers using human data from online e-commerce websites or apps like Amazon and Flipkart that have different features and sizes. The records are picked at random from all of the records on the website.
- **Data Integration:** In this step, we combine the data from multiple review sourcedatasets into a coherent form.
- **SpamIdentificationLabelling:**In this step, we look for various types of thespam in the data integrated set, and labelled each record as spam and non-spammanually.
- **Pre-processing:** In this step, we use different types of pre-processing techniques to deal with data that is missing, noisy, or doesn't match up. There are a number of pre-processing methods, such as case folding, deleting unwanted characters, tokenization, handling slang words, getting rid of stop words, stemming, and handling numbers.
- **ProcessingStage:**In this step, we will first have to implement the followingsteps: 1. Data mining classification. 2. Text mining classification. 3. Data-Textminingclassification.Nowweapplyeachprecedingstepbymorethanoneclassificationmethod.
- **Evaluate the Approach:** In this step, analyzation of the outcome and rationalizethe feasibility of the approach we followed by comparing it with other previousapproaches.

V. SIMULATION RESULTS

Dataset with 92933 reviews have been traversed using Latent Semantic Analysis and Sentimental Analysis. The traversal resulted in a CSV file with genuine.

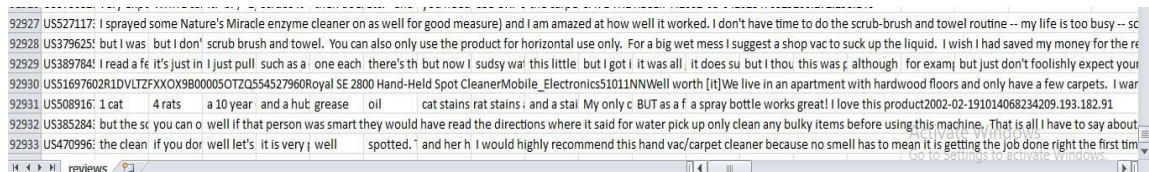


Fig: Initial Dataset with both Genuine and Fake Reviews

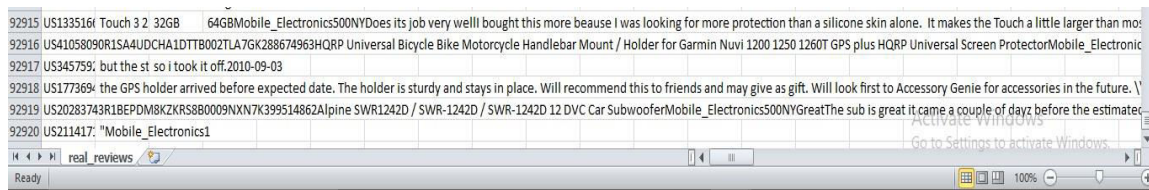


Fig: Dataset with both Genuine Reviews

VI. CONCLUSION AND FUTURE WORK

The fake review detection system is made to weed out the fake reviews. In this research, Latent Semantic Analysis and Sentimental Analysis were better at classifying than the Naive Bayes classifier for the testing dataset. To use a real-time/time-based dataset that will let us compare the user's timestamps of the reviews to see if a certain user is posting too many reviews in a short amount of time. To expand the research to deep learning techniques, use and

compare other machine learning algorithms like logistic regression.

REFERENCES

- [1] Barbado R, Araque O, Iglesias CA. A framework for fake review detection in online consumer electronics retailers. *Information Processing & Management*. 2019 Jul 1;56(4):1234-44.
- [2] Sinha A, Arora N, Singh S, Cheema M, Nazir A. Fake Product Review Monitoring Using Opinion Mining. *International Journal of Pure and Applied Mathematics*. 2018;119(12):13203-9.
- [3] Rajamohana SP, Umamaheswari K, Dharani M, Vedackshya R. A survey on online review SPAM detection techniques. In 2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT) 2017 Mar 16 (pp. 1-5). IEEE.
- [4] Elmurngi E, Gherbi A. Detecting fake reviews through sentiment analysis using machine learning techniques. *IARIA/data analytics*. 2017 Nov;65-72.
- [5] Kotian H, Meshram BB. Detection of Spam Reviews and Spammers in E-Commerce Sites. In 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC) 2017 Sep 8 (pp. 299-304). IEEE.
- [6] Li Y, Feng X, Zhang S. Detecting fake reviews utilizing semantic and emotion model. In 2016 3rd International Conference on Information Science and Control Engineering (ICISCE) 2016 Jul 8 (pp. 317-320). IEEE.
- [7] Karami A, Zhou B. Online review spam detection by new linguistic features. *Conference 2015 Proceedings*. 2015 Mar 15.
- [8] Shivagangadhar K, Sagar H, Sathyan S, Vanipriya CH. Fraud detection in online reviews using machine learning techniques. *International Journal of Computational Engineering Research (IJCER)*. 2015 May;5(5):52-6.
- [9] Mevada D, Daxini V. An opinion spam analyzer for product reviews using supervised machine learning method. *Journal of Information, Knowledge and Research in Computer Engineering*. 2015.
- [10] Fang Fang X, Zhan J. Sentiment analysis using product review data. *Journal of Big Data*. 2015 Dec 1;2(1):5.
- [11] Kokate S, Tidke B. Fake review and brand spam detection using J48 classifier. *IJCSIT Int J Comput Sci Inf Technol*. 2015;6(4):3523-6.
- [12] Kolhe NM, Joshi MM, Jadhav AB, Abhang PD. Fake reviewer groups' detection system. *Journal of Computer Engineering (IOSR-JCE)*. 2014;16(1):6-9.
- [13] Mukherjee A, Venkataraman V, Liu B, Glance N. Fake review detection: Classification and analysis of real and pseudo reviews. Technical Report UIC-CS-2013-03, University of Illinois at Chicago, Tech. Rep.. 2013.
- [14] Reddineelima C, Hariitha V, Dinesh U, Kalpana B, Kumar PN. Spotting and Removing Fake Product Review in Consumer Rating.
- [15] Ioannis D. *Fake Review Detection via exploitation of Spam Indicators and Author Behavior Characteristics* (Doctoral dissertation, Aristotle University of Thessaloniki)



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165

doi[®]
cross **ref**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



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