

Comprehensive Study Analysis of Fake Product Review

Mega anushya , Raamani

Assistant Professor, Kathir College of Engineering, Coimbatore, Tamilnadu, India.

Article Information

Received : 02 August 2024
Revised : 06 August 2024
Accepted : 18 August 2024
Published : 22 August 2024

Abstract— This paper In online shopping, the product sellers often take reviews from the customer for the product. As e-commerce is growing rapidly the number of reviews from the customer is also increasing day by day. Some spammers attempt to write fake(spam) reviews to affect customer opinions and increase their sales rating. It is a tedious work for the customers to identify the fake(spam) reviews. We propose spam detection using Bi-gram algorithm and IP Address to filter the genuine reviews from the fake (spam) reviews.

Corresponding Author:

Mega anushya

Keywords: *D product review, fake(spam) reviews, IP addressdetection*

Copyright © 2024:, Mega anushya, This is an open access distribution, and reproduction in any medium, provided Access article distributed under the Creative Commons Attribution License the original work is properly cited License, which permits unrestricted use.

Citation: Mega anushya, “Comprehensive Study Analysis of Fake Product Review”, Journal of Science, Computing and Engineering Research, 7(8), August 2024.

I. INTRODUCTION

In the era of internet, everything has very are becoming very faster. Social media sites are used for interacting with people all over the world. Thoughts are shared on the internet. Internet provide the facility of online shopping like Amazon, Flipcart etc. Before purchasing anything people look for reviews. Based on the reviews the people make their decision whether to purchase the product or not. Based on the evaluations customers can compare the brands and finalize their interest. People are not aware about the credibility of the reviews. It may be genuine or fake. This is not identified by the people. To address this issue, a few consumer sites have consolidated tips and clues to manually spot spam reviews. However, for online products with a huge number of reviews, it is practically impossible to manually check and distinguish the spam opinions from the real reviews. Several high-profile cases have been reported and spammers have transparently admitted to be paid to write fake reviews in media investigations. Many businesses have rewarded positive reviews with promotions and coupons. Many of the, researchers have shown great interest in identifying truthfulness of reviews. Reviewer behavior analysis is conducted in to detect suspicious activities using the temporal footprints of the reviewers, and hypothesize that there are natural distributions of review star ratings and spam reviews can lead to skewed rating distribution. use content similarity analysis which treats duplicates and near-duplicate reviews as potential spam reviews based on the evidence that spammers often sign in with different identifies to reuse similar reviews on a same product. However, research in spam detection is far from adequate to address the challenges in this area. In this study, we propose an end-to-end system that works seamlessly on large data sets and generates user consumable reports. It integrates

multiple methodologies in a practical fashion to overcome the ever-changing techniques employed by spammers.

II. RELATED WORKS

Clearly consumers value the feedback given by other users as do the companies that sell such products. Blogs, websites, discussion boards etc. are a repository of customer suggestions which are valuable and important sources of textual data. Therefore, today's individuals and older ones extensively rely on reviews available online. It means that people make their decisions of whether to purchase the products or not by analyzing and reflecting the existing opinions on those products. The fact that is if the potential customer or users gets a genuine overall impression of a product by considering the present affect for that product, it is highly probable that he will actually purchase the product. Normally if the percentage of positive and effective opinions is considerable, it is likely that the overall impression will be highly positive. Likewise, if the overall impression is not proper, it is doubtful that they don't buy the product. The main disadvantage is if the social media optimization team uses different mail accounts to review the product it is not possible to detect the fake review. To overcome this a new frame work is proposed

III. BACKGROUND OF STUDY

Many people require review about a product before spending their money on the product. People may come across various reviews in the website but these reviews are genuine or fake is not identified by the user. In some review websites some good reviews are added by the product company people itself in order to make product famous this people belong to Social Media Optimization team. Every one give good reviews for many different products manufactured by their own firm. User will don't know to find out whether the review is genuine or fake. To find out

this "Fake Product Review Analysis" system is introduced. The proposed framework consists of two main methods namely IP address detection - to find out fake reviews made by the social media optimization team by

identifying the IP address. and bi-gram algorithm – to differentiate good and bad reviews.

2.1. IP ADDRESS DETECTION

2.1.1 EMAIL ID

Every E-commerce website we could the patron avail its offerings and permit transactions most effective after verifying his identification i.e. The user has to make an account on that platform. Majorly, one such piece of information that each line service company demands is the customer's Gmail ID. This work has exploited this statistics. In our prototype, a customer can put up best 1 assessment on a specific product with a given Gmail ID. One vital feature that distinguishes Gmail IDs is the usage of prevent phrases (full prevent, commas, etc.). In a situation wherein if a consumer login to a website with Gmail ID say- 'abc@gmail.Com', and the subsequent time the spammer tries to login the usage of 'ab.C@gmail.Com', then in that case majority of the present Ecommerce websites deal with these of them as separate debts or two exclusive customers. But in fact these are the same IDs with admire to Gmail. Thus, it turns into smooth for the spammers to login through distinctive bills and submit multiple evaluations with a extraordinary identification every time. To avoid this, we have proposed a new framework the elimination of prevent phrases and then compare if the email IDs given are identical or exceptional. A better idea would be permit Gmail handle the process of authentication and in preference to having separate signup, virtually permit the client surf thru the internetsite the usage of his Gmail ID. We have adopted both these processes in our prototype. But what if the spammer makes another Gmail ID? Though it will be tedious to make a new ID every time before posting a overview however yes there is a possibility. To counter assault this issue, we can hint the IP address of the device.

2.1.2 IP ADDRESS

The spammers can use any of the devices (smartphones, laptops) to publish fake critiques. However, the system getting preserve the IP cope with of the tool can help in uniquely identifying the device. we have proposed the prototype puts a prevent on the number of attempts or times a consumer allowed to submit online evaluations from his tool on one specific product. A proper user will now not try to put up multiple reviews for the one unique product greater than once. Thi technique prevents review unsolicited mail that would in any other case lie to the users, which is extremely obnoxious. Algorithm: To detect spam review Input: A review 1. Authenticate the Gmail ID of the user 2.

If the user has already posted a review for the same product, discard the review. 3. If he is posting for the first time, check the IP address of the device being used 4. If number of reviews for a specific product from that device \geq Threshold, review is manipulated. 5. If number of reviews for a specific product from that device $<$ Threshold, review is considered to be genuine. If multiple reviews from the same IP Address have been made in the same day then restrict the user from posting the review.

2.1.3 SPAM WORD DICTIONARY Another problem posing hazard is the nonevaluations. A unsolicited mail word dictionary is used to stumble on irrelevant reviews. We have taken the inspiration of that the elimination of the unsolicited mail words from the beyond studies which many of the researchers have attempted to locate junk mail emails. Phrases like 'buy direct', 'different offer', 'money back' and lots of extra are used steer the people or promote it their very own services. Thus, such opinions can not be taken into consideration genuine. We have, therefore, modelled our framework approach that does not keep in mind a evaluate containing such spam words as genuine. database is being maintained that consists of all such spam words and terms. The content analysis of the online evaluations is made in opposition to these saved spam phrases the usage of the ngram model and Jaccard's coefficient preserving the brink as 80%. The opinions containing any of these junk mail words or phrases are considered to be misleading. The textual analysis of the online opinions further restricts the spammers to a first rate extent. Algorithm: To detect non-review Input: A review If the review contains any word present in the spam dictionary, it is considered as a non-review.

IV. METHODOLOGY

The overall process is illustrated in Figure. 1 First, each review is converted into a set of bi-gram shingles that are formed by combining two consecutive words together. For example, "This product is amazing," a set of bigram shingles is generated containing "This product," "product is," and "is amazing." Usage of such bigrams is more meaningful than tokenizing each word as it helps increase the relevance between the contexts. The similarity between two opinions are regularly computed by using checking the ratio of interactions of their sets of bi-gram shingles. There are various similarity measures in the literature like Euclidean distance, Cosine similarity, Minkowski distance and Jaccard similarity. For this research, Jaccard similarity is applied due to the fact as compared to other measures that compute similarity measures for facts objects represented assets a chunk like units of bigram shingles. 2.2.1 Cyclic Redundancy Check (CRC) Hash and min-hash for simpler computation to get the Jaccard similarity cost, the shingles are first mapped to shingles ID's the use of the CRC32 hash characteristic that converts any variable duration string right

into a hexadecimal value of 32-bit binary sequence. Now, the set for every review is represented as a set of integers instead of substring shingles. Still the scale of the set is large to compute the similarity. Our primary goal is to have a smaller illustration of these large sets known as “signatures”. The key belonging to such signatures is that they should be an excellent representation of the large set with mile smaller size. Signatures for each set are derived the use of min hash scheme with k hash functions, where k is a fast and hard integer. The valve of k is ready to one zero five for this research. According to chernoffBound the predicted mistakes charge by using min-hash is $0(1/\sqrt{k})$ and in general having a okay around 100 ends in a small error probability. Using min hash values prevents from having to explicitly compute random diversifications of all of the shingles IDs 2.2.2 Inverted index After reap min-hash values to represent each review, the Jaccard similarity is converted into the range of equal signatures in opinion A and B divided by the full wide variety of specific signatures in them. However, to examine min-hash values among critiques still requires the complexity of $O(n^2)$ where n is the wide variety of evaluations indexes are constructed the usage of the min-hash values. Inverted index is the broadly adopted technique in the statistics retrieval. In our study, a dictionary is constructed by the use of all min-hash value of a product as index keys, where every index keys factors to opinions containing the corresponding min-hash cost. This inverted index returned all the goods with a given min-hash fee in a $O(\log n)$ time.

V. RESULTS AND DISCUSSIONS IP ADDRESS DETECTION

The maximum widevariety of instances a reviewer can post evaluations from a given e-mail ID was set to 1, with aselected IP dealwith to2, and from aspecific location to 3. The proposed prototype changed into testedand opinions from special users.It confirmed that the accuracy percentage of 65%.The proposed model couldn't determine out the spammers in a few cases. For example, in a rooted mobile device, the person can alternate the ways of configurations of his tool inclusiveof its IP address anytime. Secondly, there are few area converting applications available thatpermit the users togovern the place of the device easily. Another element that are at the times interfered with the right working of these prototype turnedinto the precision and accuracy of the geo-coordinates retrieved. In certain cases, the range and longitude of an area have been retrieved of then up to number that is four decimal places, whereas in some other cases up to2or three places of the decimals. In thissortof scenario, the server facet database shops exclusive values of geocoordinates for the identical place. Hence, this inconsistency of these recent facts needs to be resolved when similarly paintings is performed inthis domain.

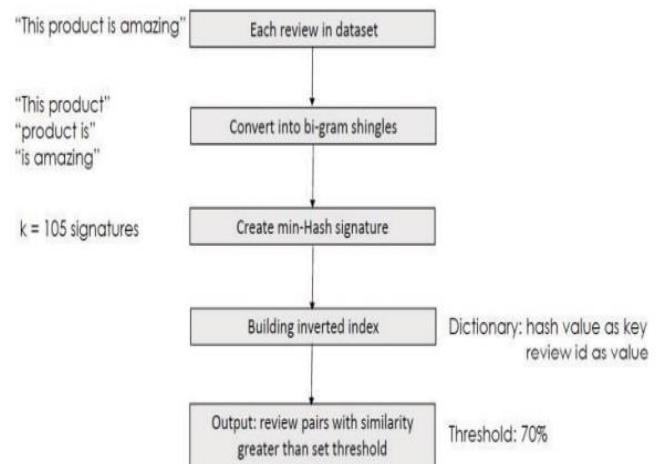


Figure 1. Overall process of bi-gram algorithm

VI. DETECTION OF DUPLICATEREVIEWES USING BIGRAM ALGORITHM

The output from duplicate detection approach is kept in a CSV file one by one for every product category. Product X’s asin, review rating, unixtime, product Y’s asin, review rating, unixtime, and similarity score between X and Y are the column headings for this CSV file. Figure 3. Screenshot of the example duplicate detection output In order to examine the results of the output, example duplicate detection results are presented. The highlighted rows 366 and 367 show high similarity scores that indicate potential duplicate reviews between products B00155237W and B00063X7KG; and between B00155237Wand B002XOXS12. To validate the results, the product asins are used to find the product description fromAmazon.com: • B00155237W: Cruiser Accessories 76200 Tuf Flat Shield Novelty / License Plate Shield • B00063X7KG: Meguiar’s G1016 Smooth Surface ClayKit • B002XOXS12: Meguiar’s G110V2 Professional Dual ActionPolisher Based on the unixtime, the reviewer “Mack Wu” haswritten similar reviews to all the three different types of products and at the same time which indicates spams. IV. EFFICIENCY The system can perform fake review analysis efficiently. By running the test on a MacBook Pro, the system is in a position to generate the credibility report for the product categories within an hour. This makes the system usable for online website deployment. It’s well adapted to large datasets and also might be used for other datasets with little changes to the feature names.

Review	ID	IP	Spam Words	Impact
Great product	abc@email.com	192.168.6 5.1	-	No spam
Nice features	abc@email.com	192.168.6 5.1	-	Spam
Amazing quality	id@email.com	223.190.3 3.131	-	No spam
Not up to the mark	mvd@email.com	223.180.2 3.1	-	No spam
Money guarantee	sim@email.com	190.120.1 0.1	money back	Spam

Figure 2. Results of sample test cases

All the circuits are designed by using Cadence Virtuoso tool and simulated in 90 nm CMOS technology with the supply voltage of 0.6V. The output waveform of comparator shown in Fig 8. Power waveform of the single tail comparator is shown in Fig 9.

VII. CONCLUSION

In this paper, we present a multi-dimensional analysis to detect the fake reviews. It generates credibility reports for given products detection based on two methods: IP Address detection to and Bi-gram algorithm. All the methods provide useful information, serving as an overlay to enable the discovery of fake reviews. To our best knowledge, no other work in the literature has investigated the integration of algorithms into a practical and deployable framework to determine the credibility. The results show the effectiveness and efficiency of our proposed approach

REFERENCES

- [1]. P. Nirmala, T. Manimegalai, J. R. Arunkumar, S. Vimala, G. Vinoth Rajkumar, Raja Raju, "A Mechanism for Detecting the Intruder in the Network through a Stacking Dilated CNN Model", *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 1955009, 13 pages, 2022. <https://doi.org/10.1155/2022/1955009>.
- [2]. D. Sathyanarayanan, T. S. Reddy, A. Sathish, P. Geetha, J. R. Arunkumar and S. P. K. Deepak, "American Sign Language Recognition System for Numerical and Alphabets," 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 2023, pp. 1-6, doi: 10.1109/RMKMATE59243.2023.10369455.
- [3]. J. R. Arunkumar, Tägele berihun Mengist, 2020" Developing Ethiopian Yirgacheffe Coffee Grading Model using a Deep Learning Classifier" *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN: 2278-3075, Volume-9 Issue-4, February 2020. DOI: 10.35940/ijitee.D1823.029420.
- [4]. Ashwini, S., Arunkumar, J.R., Prabu, R.T. et al. Diagnosis and multi-classification of lung diseases in CXR images using optimized deep convolutional neural network. *Soft Comput* (2023). <https://doi.org/10.1007/s00500-023-09480-3>
- [5]. J.R.Arunkumar, Dr.E.Muthukumar," A Novel Method to Improve AODV Protocol for WSN" in *Journal of Engineering Sciences*" ISSN NO: 0377-9254 Volume 3, Issue 1, Jul 2012.
- [6]. R. K. A. Shameem, P. Biswas, B. T. Geetha, J. R. Arunkumar and P. K. Lakineni, "Supply Chain Management Using Blockchain: Opportunities, Challenges, and Future Directions," 2023 Second International Conference on Informatics (ICI), Noida, India, 2023, pp. 1-6, doi: 10.1109/ICI60088.2023.10421633.
- [7]. Arunkumar, J. R. "Study Analysis of Cloud Security Challenges and Issues in Cloud Computing Technologies." *Journal of Science, Computing and Engineering Research* 6.8 (2023): 06-10.
- [8]. J. R. Arunkumar, R. Raman, S. Sivakumar and R. Pavithra, "Wearable Devices for Patient Monitoring System using IoT," 2023 8th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2023, pp. 381-385, doi: 10.1109/ICCES57224.2023.10192741.
- [9]. S. Sugumaran, C. Geetha, S. S, P. C. Bharath Kumar, T. D. Subha and J. R. Arunkumar, "Energy Efficient Routing Algorithm with Mobile Sink Assistance in Wireless Sensor Networks," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ACCAI58221.2023.10201142.
- [10]. R. S. Vignesh, V. Chinnammal, Gururaj.D, A. K. Kumar, K. V. Karthikeyan and J. R. Arunkumar, "Secured Data Access and Control Abilities Management over Cloud Environment using Novel Cryptographic Principles," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ACCAI58221.2023.10199616.
- [11]. Syamala, M., Anusuya, R., Sonkar, S.K. et al. Big data analytics for dynamic network slicing in 5G and beyond with dynamic user preferences. *Opt Quant Electron* 56, 61 (2024). <https://doi.org/10.1007/s11082-023-05663-2>
- [12]. Krishna Veni, S. R., and R. Anusuya. "Design and Study Analysis Automated Recognition system of Fake Currency Notes." *Journal of Science, Computing and Engineering Research* 6.6 (2023): 16-20.
- [13]. V. RamKumar, S. Shanthi, K. S. Kumar, S. Kanageswari, S. Mahalakshmi and R. Anusuya, "Internet of Things Assisted Remote Health and Safety Monitoring Scheme Using Intelligent Sensors," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ACCAI58221.2023.10199766.
- [14]. R. S. Vignesh, R. Sankar, A. Balaji, K. S. Kumar, V. Sharmila Bhargavi and R. Anusuya, "IoT Assisted Drunk and Drive People Identification to Avoid Accidents and Ensure Road Safety Measures," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ACCAI58221.2023.10200809.

- [15].I. Chandra, G. Sowmiya, G. Charulatha, S. D, S. Gomathi and R. Anusuya, "An efficient Intelligent Systems for Low-Power Consumption Zigbee-Based Wearable Device for Voice Data Transmission," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10083856.
- [16].G. Karthikeyan, D. T. G, R. Anusuya, K. K. G, J. T and R. T. Prabu, "Real-Time Sidewalk Crack Identification and Classification based on Convolutional Neural Network using Thermal Images," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 1266-1274, doi: 10.1109/ICACRS55517.2022.10029202.
- [17].R. Meena, T. Kavitha, A. K. S, D. M. Mathew, R. Anusuya and G. Karthik, "Extracting Behavioral Characteristics of College Students Using Data Mining on Big Data," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10084276.
- [18].S. Bharathi, A. Balaji, D. Irene, J, C. Kalaivanan and R. Anusuya, "An Efficient Liver Disease Prediction based on Deep Convolutional Neural Network using Biopsy Images," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1141-1147, doi: 10.1109/ICOSEC54921.2022.9951870.
- [19].I. Chandra, G. Sowmiya, G. Charulatha, S. D, S. Gomathi and R. Anusuya, "An efficient Intelligent Systems for Low-Power Consumption Zigbee-Based Wearable Device for Voice Data Transmission," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10083856.
- [20].Revathi, S., et al. "Developing an Infant Monitoring System using IoT (INMOS)." International Scientific Journal of Contemporary Research in Engineering Science and Management 6.1 (2021): 111-115.
- [21].J.R.Arunkumar, Dr.E.Muthukumar,| A Novel Method to Improve AODV Protocol for WSN| in Journal of Engineering Sciences| ISSN NO: 0377-9254Volume 3, Issue 1, Jul 2012.
- [22].R. S. Vignesh, A. Kumar S, T. M. Amirthalakshmi, P. Delphy, J. R. Arunkumar and S. Kamatchi, "An Efficient and Intelligent Systems for Internet of Things Based Health Observance System for Covid 19 Patients," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ICECONF57129.2023.10084066.
- [23].I. Chandra, K. V. Karthikeyan, R. V, S. K, M. Tamilselvi and J. R. Arunkumar, "A Robust and Efficient Computational Offloading and Task Scheduling Model in Mobile Cloud Computing," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ICECONF57129.2023.10084293.
- [24].R. K, A. Shameem, P. Biswas, B. T. Geetha, J. R. Arunkumar and P. K. Lakineni, "Supply Chain Management Using Blockchain: Opportunities, Challenges, and Future Directions," 2023 Second International Conference on Informatics (ICI), Noida, India, 2023, pp. 1-6, doi: 10.1109/ICI60088.2023.10421633.
- [25].J. R. Arunkumar, and R. Anusuya, "OCHRE: A Methodology for the Deployment of Sensor Networks." American Journal of Computing Research Repository, vol. 3, no. 1 (2015): 5-8.