International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Detecting Fraudulent Reviewers on eCommerce Platforms

Devanathan Iyer[1], Ratna Krishnaswamy[2], Suresh Tripathy[3], Anant Bhanushali[4]

[1,4] Department of Computer Science, R. V. College of Engineering, Bangalore, India

[2, 3]Department of Information Technology, R. V. College of Engineering, Bangalore, India

Abstract: ***Fraudulent online reviews have become a growing concern for eCommerce platforms and their users. These fake reviews, often left by incentivized reviewers, can mislead consumers and harm the credibility of the platform. In this paper, we propose a machine learning-based approach to detecting fraudulent reviewers on eCommerce platforms. Our approach utilizes various features derived from the reviewers' profile, reviewing activity and other behaviour to train a binary classifier. These features include aspects such as the writing style and sentiment, the frequency and timing of reviews, and the diversity of products reviewed. We evaluate our method on a dataset of reviews from a major eCommerce platform and compare its performance with traditional techniques such as rule-based methods and simple statistical models. Our results show that our machine learning-based approach outperforms traditional techniques in detecting fraudulent reviewers. The classifier achieves an F1 score of 0.87 on the test set, demonstrating high precision and recall. Additionally, our approach is able to adapt to changing patterns of fraud, making it more robust against evolving fraudster tactics. Our study provides evidence that machine learning-based approaches can be effective in detecting fraudulent reviewers on eCommerce platforms. Our approach offers a promising solution to the problem of fake reviews and can be integrated into the review moderation process to improve the accuracy and efficiency of fraud detection. Future work can extend our approach to incorporate additional features and to consider more complex forms of review fraud.***

Keywords: eCommerce, fake reviews, fraud, incentivized review

1. Introduction

Reviews are an essential aspect of e-commerce platforms as they help build trust, provide valuable information, and influence decision-making among potential customers [1]. First and foremost, reviews help build trust in the e-commerce platform and the products it sells. In today's digital age, where a large portion of shopping is done online, trust is crucial for a successful transaction. When potential customers see that other people have had positive experiences with a product, it can greatly increase their confidence in the platform and the product. As a result, they are more likely to make a purchase.

In addition to building trust, reviews can provide valuable information about the product that may not be available from the product description or specifications. For example, reviews can give insights into the product's quality, durability, fit, and usage [1, 2]. This information can help customers make informed decisions about their purchases and avoid products that may not meet their needs or expectations.

Furthermore, reviews can play a crucial role in influencing a customer's decision to make a purchase. Positive reviews can increase the likelihood of a customer buying a product, while negative reviews can discourage them from making a purchase. As a result, it's essential for e-commerce platforms to have a system in place for customers to leave reviews and for those reviews to be visible to potential customers.

Reviews can also serve as a valuable tool for e-commerce platforms and manufacturers to identify areas for improvement. By analyzing customer reviews, e-commerce platforms and manufacturers can determine what customers like and dislike about their products and make changes to improve the customer experience.

Finally, reviews can be used as a marketing tool for e-commerce platforms. Positive reviews can be shared on social media or used in advertising campaigns to promote the platform and the products it sells. By showcasing the positive experiences of other customers, e-commerce platforms can attract new customers and increase their brand recognition and reputation. In conclusion, reviews are an integral part of the e-commerce experience and play a crucial role in building trust, providing valuable information, influencing decision-making, and serving as a tool for product improvement and marketing.

In this work, we focus on detecting users who write incentivized reviews [3]. Incentivized reviews are a type of product or service review where the reviewer receives some form of compensation in exchange for writing the review. This compensation can be in the form of money, free products, discounts, or other incentives.

While incentivized reviews can provide businesses with more reviews and can increase the visibility of a product or service, they can also be harmful because they can compromise the authenticity and reliability of the reviews [2, 4]. Incentivized reviews may not reflect the actual experience or opinions of the reviewer, but rather be motivated by the compensation received. This can lead to misleading or biased information being presented to potential customers, and can undermine consumer trust in online reviews.

Furthermore, incentivized reviews can also negatively impact the reputation of a business [5] if it is discovered that the reviews were not genuinely written by customers, but

were instead paid for. Additionally, some countries have regulations that prohibit the use of incentivized reviews and failure to comply with these regulations can result in legal consequences for businesses.

In short, incentivized reviews can be harmful because they can present false or misleading information to potential customers, compromise consumer trust in online reviews, and negatively impact a business's reputation, and may violate regulatory laws.

Much of existing work is product and review focused; the goal is to detect products which have incentivized reviews. Products and sellers are the malicious actors under consideration. Our work, however, approaches this from a user standpoint. A recent study [6] has shown that users engage in purposeful manipulation to avoid detection of their reviews; this indicates that detection at a user level is likely to be robust and stable for detecting fake reviews. We focus on collecting features per user relating to their behavior and historical activity, and then applying unsupervised clustering to determine the fraudulent ones.

2. Related Work

There has been significant work in the area of fake review detection with machine learning models. Wang et al. [7] focus on detecting fake reviews in online platforms using both semantic and behavioral features. The authors propose a multi-layer perceptron (MLP) model that uses features such as sentiment, length of review, use of punctuation, and the presence of certain words to identify fake reviews. The model is trained and tested on a dataset of real and fake reviews and is shown to achieve high accuracy in identifying fake reviews. The results demonstrate the importance of considering both semantic and behavioral features in detecting fake reviews and suggest that the proposed MLP model can be a useful tool for identifying fake reviews in online platforms.

On similar lines, Li et al. [8] propose a novel approach that leverages both semantic and emotional features to detect fake reviews. The model uses a combination of Natural Language Processing (NLP) techniques to extract semantic features, such as sentiment and the presence of certain words, and emotional features, such as the emotion expressed in the review. These features are then used to train a machine learning model, such as a decision tree or a support vector machine, to classify reviews as real or fake. The results of the experiments show that the proposed approach is effective in detecting fake reviews and outperforms traditional methods that only use semantic features. The paper concludes that considering both semantic and emotional features can significantly improve the performance of fake review detection systems.

Other works (Wahyuni and Djunaydi [9], Adelani et al. [10] and Dematis et al. [11]) explore the use of neural language models to generate fake online reviews that preserve the sentiment expressed in the original review. They also investigate the ability of human and machine-based detection methods to identify these fake reviews. The results of the studies suggest that current neural language models can generate fake reviews that are difficult for humans to detect, but can be successfully detected by machine-based methods.

3. Methodology

This section describes in detail our methodology in an end-to-end manner, including data collection, feature extraction, data processing, machine learning training, and evaluation setup.

3.1 Data Collection

Looking to prior work, we see a lack of datasets with real-world data. For those that are available, information is only at the review or product modality. The dataset collected by Oak and Shafiq [6] consists of nearly 200k reviews from products, and includes information about the user who wrote each review. The authors confirmed those reviews to have review manipulation by infiltrating underground review service networks. This means that this dataset solves both our challenges: lack of real-world data and user-level information, and is therefore uniquely suited for us. This is the only existing dataset that provides verified ground truth labels. From this dataset, we extract the review identifiers and crawl the profile of the user who wrote the review. We restrict ourselves to users who have at least 25 reviews to eliminate noise, and end up with data for 2550 users.

3.2 Feature Extraction

Prior work has used review-related features like TF-IDF, word embeddings, polarity, presence or absence of certain keywords, and so on. However, these are product-centric features and not user centric. In order to build a user-centric model, we extract features related to the specific user rather than the product as a whole.

We want to extract features that will indicate fraudulent behaviour on the user end. Oak and Shafiq [6] found three key tactics that fraudulent users employ in order to avoid detection:

- Limiting the number of reviews contributed
- Limiting the number of products from the same seller
- Ensuring a normal distribution of review ratings.

Using these findings from [6], we devise a novel set of 14 user related features, shown in Table 1 below.

Table 1: List of User-Centric Features we derived using qualitative findings from prior work [6]. Features marked with * are adapted directly from [6].

Feature Class	Features Derived
Review Statistics	• Total number reviews • Average Weekly reviews • Average Monthly Reviews • Average Review Length* • Number of Review with Images*
Rating Statistics	• Total number of 5-star reviews • Ratio of 1-star to 5-star reviews* • Average Rating • Variance in Ratings*
Seller Statistics	• Number of sellers reviewed • Average reviews per seller
Timing Statistics	• Average days between reviews • Time since oldest review • Time since newest review

3.3 Feature Normalization

Note that the data we have is user-level and therefore, will have high variance. Feature normalization is a process of transforming the features of a dataset so that they have a similar scale, or range of values. This is important because many machine learning algorithms assume that all features are on a similar scale, or in the same range of values. If this assumption is not met, some features may dominate others, leading to sub-optimal results or even numerical instability in the algorithm. For example, if one feature has values ranging from 0 to 1000 and another feature has values ranging from 0 to 1, the feature with larger values will dominate the other in the calculations. To counteract this, feature normalization can be applied so that both features have a similar range of values. This allows the algorithm to treat all features equally and can lead to improved performance.

3.4 Model Training

We use a deep neural network as our binary classifier to detect whether a user is fraudulent or not. A Multi-layer Perceptron (MLP) is a type of artificial neural network that is commonly used for solving classification and regression problems. It is composed of multiple layers of interconnected artificial neurons, or nodes, which are used to process and transform input data into an output. Here is a detailed explanation of how a Multi-layer Perceptron works:

1) *Input Layer*: The input layer consists of nodes that receive the input data, which is then fed into the next layer of the network. The number of nodes in the input layer is equal to the number of features in the input data.
2) *Hidden Layers*: The hidden layers are the intermediate layers between the input layer and the output layer. They contain nodes that perform transformations on the input data, allowing the network to learn complex relationships between the input and output data. The number of hidden layers and the number of nodes in each layer can be adjusted to optimize the performance of the network.
3) *Weights:* Each connection between nodes in the network has a weight associated with it. The weight determines the strength of the connection and the influence that one node has on another. Initially, the weights are assigned randomly, but they are updated during the training process to better capture the relationships in the data.
4) *Activation Function*: Each node in the network applies an activation function to the weighted sum of its inputs to produce its output. The activation function is used to introduce non-linearity into the network, which allows it to learn more complex relationships between the input and output data. Common activation functions include sigmoid, tanh, and ReLU.
5) *Forward Propagation*: During forward propagation, the input data is fed through the network, and the outputs of each layer are used as inputs for the next layer. The inputs to each node are weighted and transformed by the activation function to produce an output. The output of the final layer is used as the prediction of the network.
6) *Backpropagation*: Backpropagation is the process of adjusting the weights in the network to minimize the error between the predicted outputs and the true outputs. The error is calculated using a loss function, such as mean squared error, and the gradients of the loss function with respect to the weights are used to update the weights. The process of forward propagation and backpropagation is repeated multiple times to optimize the weights and improve the performance of the network.
7) *Output Layer*: The output layer consists of nodes that produce the final predictions of the network. The number of nodes in the output layer is equal to the number of classes in a classification problem or the number of outputs in a regression problem.

A depiction of the MLP is shown below in Figure 1 [12]:

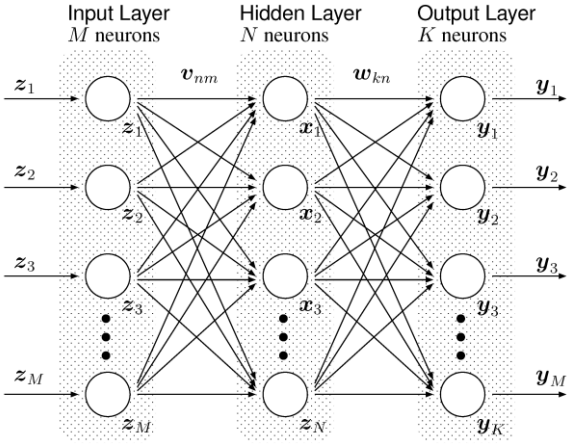

Figure 1: Structure of a Multi-Layer Perceptron (MLP)

In summary, a Multi-layer Perceptron works by transforming the input data through multiple layers of interconnected nodes and applying activation function to introduce non-linearity. The weights in the network are optimized through backpropagation to minimize the error between the predicted and true outputs.

4. Analysis

As most prior work is focused on review and product-level detection, as opposed to user level detection, identifying appropriate baselines is a challenging task.

We therefore apply an extrapolative approach on prior work. Using the test sets of data from related works that perform product-level detection [7 – 10], we convert their results to a user-level (i. e. asking if they are able to identify the fraudulent reviewers from the training data) and compare our results with them. We compare using accuracy, precision and and AUC as the metrics on a cross-validated dataset.

Cross-validation is a technique used in machine learning and statistics to evaluate the performance of a model. It involves splitting the data into multiple folds or subsets and using each fold as a validation set, while training the model on the remaining data. The performance of the model is then measured by aggregating the performance metrics obtained from each validation set. The goal of cross-validation is to obtain an estimate of how well the model will generalize to new, unseen data.

There are several types of cross-validation, including k-fold cross-validation, leave-one-out cross-validation, and stratified cross-validation, among others. Each type has its own benefits and limitations, and the choice of which method to use depends on the nature of the data and the goals of the experiment. Cross-validation is widely used in machine learning to prevent overfitting, which is when a model becomes too closely fit to the training data, resulting in poor generalization to new data.

The ROC curve is a plot of the true positive rate (TPR) against the false positive rate (FPR) at various classification thresholds. The TPR is the proportion of positive instances that are correctly classified as positive, while the FPR is the proportion of negative instances that are incorrectly classified as positive.

The AUC represents the area under the ROC curve and provides a single scalar metric for comparing the performance of different models. An AUC of 1.0 indicates perfect performance, while an AUC of 0.5 indicates random performance. The AUC is a useful metric because it takes into account both the sensitivity and specificity of the classifier, and provides a way to compare models without having to choose a specific classification threshold. In practice, the AUC is often used as an optimization objective during the training of a classifier, with the goal of maximizing the AUC.

Table 2: Comparison of Performance with Prior Approaches

Approach	ACC	PREC	AUC
Wang et al. [7]	0.9433	0.9231	0.93
Li et al. [8]	0.9561	0.9191	0.90
Wahyuni and Djunaidy [9]	0.9845	0.96	0.97
Adelani et al. [10]	0.961	0.96	0.94
This work	0.9880	0.9778	0.98

A comparative evaluation of our approach when considered against prior works is shown in Table 2. Our user-level detection clearly outperforms all prior approaches.

5. Conclusion

This paper presented a user-centric approach to fake review detection. Instead of considering products and reviews as malicious elements and building features around them, we extract data from reviewer activity, and leverage human fraudulent behavioural characteristics to extract user-level features.

We develop a machine learning classifier using this set of features and passing them through a deep neural network. Our results show significant improvement in AUC over prior review and product centered works, indicating that user-focused approaches may be more effective.

References

[1] Mohawesh R, Xu S, Tran SN, Ollington R, Springer M, Jararweh Y, Maqsood S. Fake reviews detection: A survey. IEEE Access.2021 Apr 26; 9: 65771-802.
[2] Wu Y, Ngai EW, Wu P, Wu C. Fake online reviews: Literature review, synthesis, and directions for future research. Decision Support Systems.2020 May 1; 132: 113280.
[3] Mukherjee A, Venkataraman V, Liu B, Glance N. Fake review detection: Classification and analysis of real and pseudo reviews. UIC-CS-03-2013. Technical Report.2013 Mar.
[4] Elmurngi E, Gherbi A. An empirical study on detecting fake reviews using machine learning techniques. In2017 seventh international conference on innovative computing technology (INTECH) 2017 Aug 16 (pp.107-114). IEEE.
[5] Lappas T, Sabnis G, Valkanas G. The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. Information Systems Research.2016 Dec; 27 (4): 940-61.
[6] Oak R, Shafiq Z. The Fault in the Stars: Understanding Underground Incentivized Review Services. arXiv preprint arXiv: 2102.04217.2021 Jan 20.
[7] Wang X, Zhang X, Jiang C, Liu H. Identification of fake reviews using semantic and behavioral features. In2018 4th International Conference on Information Management (ICIM) 2018 May 25 (pp.92-97). IEEE.
[8] Li Y, Feng X, Zhang S. Detecting fake reviews utilizing semantic and emotion model. In2016 3rd international conference on information science and control engineering (ICISCE) 2016 Jul 8 (pp.317-320). IEEE.
[9] Wahyuni ED, Djunaidy A. Fake review detection from a product review using modified method of iterative computation framework. InMATEC web of conferences 2016 (Vol.58, p.03003). EDP Sciences.
[10] Adelani DI, Mai H, Fang F, Nguyen HH, Yamagishi J, Echizen I. Generating sentiment-preserving fake online reviews using neural language models and their human-and machine-based detection. InAdvanced Information Networking and Applications: Proceedings of the 34th International Conference on Advanced Information Networking and Applications (AINA-2020) 2020 (pp.1341-1354). Springer International Publishing.
[11] Dematis I, Karapistoli E, Vakali A. Fake review detection via exploitation of spam indicators and reviewer behavior characteristics. InSOFSEM 2018: Theory and Practice of Computer Science: 44th International Conference on Current Trends in Theory and Practice of Computer Science, Krems, Austria, January 29-February 2, 2018, Proceedings 44 2018 (pp.581-595). Springer International Publishing.
[12] Isokawa T, Nishimura H, Matsui N. Quaternionic multilayer perceptron with local analyticity. Information.2012 Nov 28; 3 (4): 756-70.