© 2025 IJRTI | Volume 10, Issue 8 August 2025 | ISSN: 2456-3315

Identification of Malicious Injection Attacks in Dense Rating and Co-Visitation Behaviors

[1]Kuthuru Kavyasri, [2]A. Nageswari

[1]PG Scholar, [2]Assistant Professor

[1]Information Technology,

[1]G. Narayanamma Institute of Technology and Science, Hyderabad, India

[1]kuthuruakshara@gmail.com, [2]a.nageswari@gnits.ac.in

Abstract: In on-line e-commerce Web sites, product reviews formed by the user input have been found to impact on purchasing behavior. But, the rising number of malicious injection-based attacks-light-weight malicious injection attacks-via-high-density rating games and through co-visitation spam actions-endangers the reputation of said platforms. This paper has therefore come up with a new model to combat this challenge, the “Comprehensive study of the spam review detection (CSSRD)” which amalgamates linguistics-based features and behavioural-based features in detecting spam reviews. The number of spam characteristics obtained by the proposed hybrid model is 12, with six belonging to the textual analysis (e.g., sentiment polarity, content repetition) and six to the contextual behavior (e.g., abnormal rating patterns, review frequency). All the reviews are spammed-scored with the mean-based scoring and then they are classified by ML algorithms, such as “Naive Bayes, Decision Tree, Support Vector Machine, and Neural Network”. Empirical findings based on Amazon Datafiniti review dataset indicate that the neural network model is the most accurate with 97 percent accuracy in recognition of both forms of spam pattern as compared to other classifiers. CSSRD model provides an extendable real-time e-commerce detection framework and gives importance to the type of multi-feature fusion in spam detection.

“Index Terms *- spam review detection, machine learning, neural network, linguistic analysis, behavioral analysis, e-commerce security, hybrid model, fake review identification”.*

1. INTRODUCTION

In the digital commerce landscape, user-generated reviews significantly influence purchasing decisions and serve as a foundation for consumer trust. However, the credibility of these reviews is increasingly threatened by the proliferation of spam and fake content. Malicious users and automated bots inject deceptive reviews to either falsely elevate product ratings or damage competitors' reputations. These manipulations distort consumer perception and compromise the fairness of recommendation systems. As a result, e-commerce platforms face mounting challenges in maintaining the integrity of user feedback. To counteract this, “machine learning (ML) and natural language processing (NLP)” techniques have emerged as effective tools for detecting such fraudulent activity [1].

Traditional spam detection systems often rely solely on either textual or behavioral data. Text-based models identify suspicious content through language patterns, sentiment, and repetition [3], while behavioral models assess rating trends, review bursts, and reviewer credibility [4][5]. Hybrid models that integrate both linguistic and behavioral features have demonstrated superior accuracy and adaptability in detecting diverse spam tactics [6]. Public datasets like the Amazon reviews dataset [2] provide a robust foundation for training such models.

Given the increasing sophistication of spam strategies, including “coordinated and adversarial attacks [7]”, this study proposes an intelligent hybrid spam detection system. By combining linguistic analysis with behavioral tracking, the system aims to accurately identify manipulated reviews, enhance data reliability, and support trustworthy online marketplaces.

2. RELATED WORK

Many studies have discussed in detail how easy it is for shilling attacks, false reviews and harmful injection behavior to get into the recommended systems and electronic trading platforms. Since online control systems have increasing power over how people buy and how businesses are visible, attackers are increasingly using these sites to change product evaluation, cheat consumers and damage the integrity of platforms. These new concerns have led to many scientific work to find and stop various types of information and fraud revisions.

These changes were mostly made possible by “neural networks and sentiment analysis”. Paliwal et al. [8] they have shown that neural networks can accurately capture the emotional polarity of written assessments. This is a step to find patterns that show a review spam. Shenepoor et al. [9] they created this by creating a Netspam frame that uses graph structures to simulate the behavior of spam by combinations of information and text content. It shows how users, reviews and items are connected to each other.

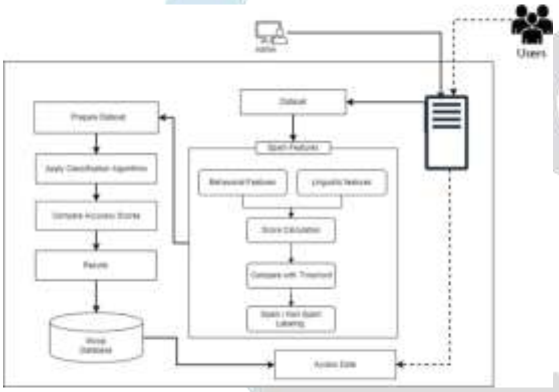
Recently, methods of contradictory and amplification have been used to imitate more complex attacks. Fan et al. [10] He came up with a crusader, a frame based on strengthening learning that can find out the best time and way to provide poisoned data for collaborative filter systems. Yang et al. [11] they proposed a multimodal method that analyzes user goals and social interactions to find hidden shillistic behavior, even if attackers act as real users.

Chen et al. [12] He came up with a differing model, a diffusion model that follows how false reviews spread through user interactions. This allows you to find coordinated inspection attacks. Yu et al. [13] They proceeded to contradictory approaches to the next level by introducing Rup-Gan, a generative contradictory network that generates fake user profiles that seem authentic to do the SPAM detection in the Black Box settings.

Nguyen et al. [14] they conducted a thorough study of attacks and talked about their methods and ways to stop them. They also pointed out some restrictions in the current model, as it does not increase or do not conceal. NAWARA et al. [15] they emphasized how important socio -cultural and behavioral features are, and call for real data files that show how complicated attacks can be. All these research shows how important hybrid models and adaptive detection systems to combat new hazards to set up, where people give recommendations based on reviews.

3. METHODOLOGY

This proposed system is a hybrid framework for SPAM detection that uses both language and behavioral aspects to find false reviews on e -trading websites. It pulls out 12 important functions: six from the review text, including the polarity of the sentiment and the similarity of the topic, and six of behavior such as the assessment of extremes and inspection explosions. The method uses the Amazon Datafiniti data file [2] to obtain an average spam score for each category. We use ML techniques such as “naive Bayes, decision tree, Support Vector Machine, and Neural Network”, do the best work [8]. This method of two features causes things to be more accurate, limits false positives [9] and amplifies spam protection [15].



“Fig. 1. System Architecture”

Figure 1 shows a SPAM detection system with two parts: an algorithm rating and real -time detection. We analyze user data by analyzing language and behavioral functions, providing scores and comparing it with the threshold. The administrator starts the system while the MySQL database stores and gets results such as accuracy measurement.

i) Dataset Collection:

To find spam reviews in electronic trading, you need a data file with lots of structured information. This research uses Amazon Customer Reviews from Datafiniti (2018) [2], which has user, product ID, text check, evaluation and time stamp. These attributes make it possible to perform a language analysis, such as the view of the sentiment and the similarity of content and analysis of behavior, such as that we look at the frequency of reviews and anomalies of evaluation. The width of the real world allows 12 important functions of linguistic and six behavioral functions-which are needed to train and test classification models of machine learning that are accurate and strong.

ii) Pre-Processing:

To turn Raw Review data into a structured input for spam detection, preliminary processing is required. This step involves getting rid of punctuation, stopping words and elements of HTML, changing text to lowercase and its tokenization. We throw up duplicate or incomplete items to make sure the data is good. For language analysis we obtain the polarity ratio, content similarity and the ratio of capitalization. For behavior analysis we use time stamps and activities of activities to find things such as inspection rupture and abnormalities. Elakkiya et al. [3] they reported that good pre -processing is more accurate and faster models of deep learning.

iii) Training & Testing:

Training and testing are important components of the detection of SPAM control detection that works well. After extracting functions, the data file is divided, usually with 80% for training and 20% for testing. During training, the model learns patterns from language and behavior. It is then tested with data she has never seen before. This procedure ensures that the model can be used in other situations and prevents excessive quantities that are generally backed up by cross validation. Hussain et al. [5] they emphasized that good training and testing causes systems to be more scalable and better to find different types of spam injection behavior.

iv) Algorithms:

“Naïve Bayes” is a probability classifier that uses

Bayes' sentence and assumes that functions are independent. It is often used to find spam because it is simple and works well with text data. Although it does certain prerequisites, it works efficiently on noisy data files. Hassan and Islam [4] have proven that it works well when it finds false reviews looking at things like the mood and frequency of words.

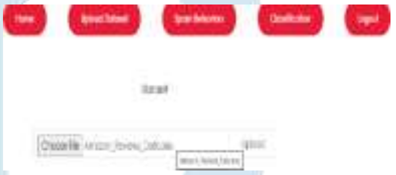
“Decision Tree” is an algorithm under supervision that divides data into groups based on elements and creates the rules of decision -making in the tree structure. This can be understood, working with numbers and categories and finding patterns that are not lines. Hussain et al. [5] they have shown that they are trying to find spam by looking for an unusual activity, such many reviews or very high ratings, which are signs of false or manipulative behavior.

“Support Vector Machine (SVM)” is a subordinate learning model that creates the best hyperplane to divide classes in space with lots of dimensions. It uses core functions to work with thin non -linear data. SVM works well to find spam when looking at language and behavior. Paliwal et al. [8] They have shown that it is relatively accurate and can be used for a wide range of sentiment and spam tasks.

“Neural Networks” are brain -based models that include layers of neurons that are associated with others and learn from data through weighted connections. They are quite good in finding complicated, non -linear patterns in spam by learning from language and behavior. Elakkiya et al. [3] they have shown that deep learning models are better than standard methods in search of spam using semantic and syntactic analysis by a user - generated inspection material to find small behavior of spam.

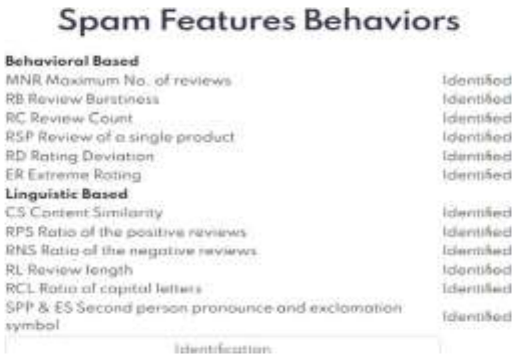
4. RESULTS AND DISCUSSIONS

We tested the SPAM detection system with four algorithms that used a combination of characteristics. “The neuron network had the highest accuracy of 97%, followed by SVM to 94%, the decision -making tree at 90%and naive Bayes at 86%”. These findings show that the integration of linguistic and behavioral factors will improve detection. Neuron networks are particularly good in finding complicated spam patterns and fine handling actions.



“Fig. 2. Upload Dataset”

“Fig. 3. Identification of BB and LB”



Spam Features Behaviors

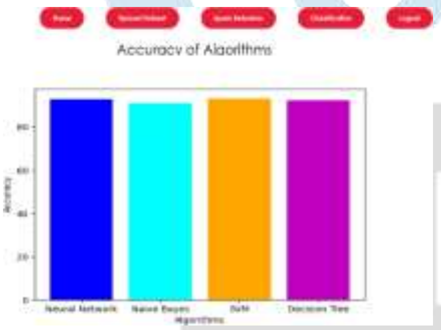
Behavioral Based	
MNR Maximum No. of reviews	Identified
RB Review Burstiness	Identified
RC Review Count	Identified
RSP Review of a single product	Identified
RD Rating Deviation	Identified
ER Extreme Rating	Identified
Linguistic Based	
CS Content Similarity	Identified
RPS Ratio of the positive reviews	Identified
RNS Ratio of the negative reviews	Identified
RL Review length	Identified
RCL Ratio of capital letters	Identified
SPP & ES Second person pronounce and exclamation symbol	Identified

“Fig. 4. Identification of Spam Behaviour BB and LB”

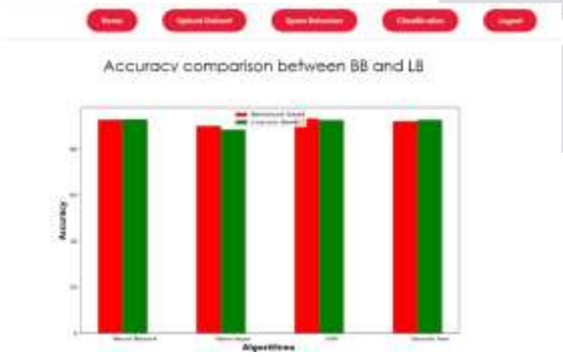
“Fig. 5. Upload Features Data for Classification”

Algorithm Name	Overall Accuracy	Behavioral Based Accuracy	Linguistic Based Accuracy
Neural Network	[illegible]	[illegible]	[illegible]
Naive Bayes	[illegible]	[illegible]	[illegible]
SVM	[illegible]	[illegible]	[illegible]
Decision Tree	[illegible]	[illegible]	[illegible]

“Fig. 6. Classification of Four Algorithm with their Accuracy”



“Fig. 7. Accuracy of Four Algorithm”



“Fig. 8. Accuracy Comparison between BB and LB”

5. CONCLUSION

A growing number of fake reviews on e -shop websites caused users generated information less trustworthy and changed the way people make purchases. This article solves an important problem finding spam reviews on e -trading websites by suggesting a hybrid spam detection access that uses both language and behavior data. 12 unique functions were obtained and studied to improve identification accuracy. Six of them came from the text itself and six came from the behavior of users. Using the Amazon Datafinitie Review data file, we have changed pre -processed input data with many features, allowing ML models to accurately classify data. We tested the system very carefully using a number of different algorithms to find the best way to find spam. The neural network method had the best performance of all tested models with a rate of accuracy of 97%. This result shows that DL models are quite good to find complicated patterns and interactions between several spam indicators. The hybrid method reduced false positives by a lot and generally made spam detection more reliable. The results of this study show that the integration of linguistic and behavioral characteristics provides a more complete and reliable way of identifying spam reviews. This will help platforms of electronic trading to maintain feedback generated by a user - friendly and open user.

In the future, the system could be enhanced to handle reviews in many languages, which is more useful on electronic trading sites around the world. Adding deep context language models, such as Bert, can even improve linguistic extraction of functions. Adding real -time detection capabilities and increasing the data file to include rating from other platforms can also be more robust.

REFERENCES

[1] Dada, E.G., Bassi, J.S., Chiroma, H., Abdulhamid, S.M., Adetunmbi, A.O., & Ajibuwa, O.E. (2019). Machine learning for email spam filtering: review, approaches and open research problems. Heliyon, 5.

[2] Datafiniti, (2018), *"Consumer reviews of Amazon products"*, Available at: https://data.world/datafiniti/consumer-reviews-of-amazon-products. [Accessed: 1st August, 2023].

[3] Elakkiya, E. & Selvakumar, Santhanalakshmi & Velusamy, R.. (2021). TextSpamDetector: textual content based deep learning framework for social spam detection using conjoint attention mechanism. Journal of Ambient Intelligence and Humanized Computing.

[4] Hassan, R., & Islam, M.R. (2019). Detection of fake online reviews using semi-supervised and supervised learning. 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 1-5.

[5] Hussain Naveed & Mirza, Hamid & Rasool, Ghulam & Hussain, Ibrar & Kaleem, Mohammad. (2019). "Spam Review Detection Techniques: A Systematic Literature Review". Applied Sciences.

[6] J. Huang, T. Qian, G. He, M. Zhong, and Q. Peng, (2013) "Detecting professional spam reviewers", Springer.

[7] Rathore, Shailendra & Loia, Vincenzo & Park, Jong. (2017). "SpamSpotter: An Efficient Spammer Detection Framework based on Intelligent Decision Support System on Facebook." Applied Soft Computing.

[8] S. Paliwal, S. Kumar Khatri and M. Sharma, (2018) "Sentiment Analysis and Prediction Using Neural Networks," ICIRCA.

[9] S. Shehnepoor, M. Salehi, R. Farahbakhsh and N. Crespi, (2017) "NetSpam: A Network-Based Spam Detection Framework for Reviews in Online Social Media," IEEE Transactions on Information Forensics and Security.

[10] Fan, J., Li, M., Sun, Y., & Chen, P. (2025). DRLAttack: A Deep Reinforcement Learning-Based Framework for Data Poisoning Attack on Collaborative Filtering Algorithms. *Applied Sciences*, *15*(10), 5461.

[11] Zhihai Yang, Qindong Sun, Zhaoli Liu, "Three Birds With One Stone: User Intention Understanding and Influential Neighbor Disclosure for Injection Attack Detection", *IEEE Transactions on Information Forensics and Security*, vol.17, pp.531-546, 2022.

[12] Chen, W., Ma, X., Li, S., & Liu, B. (2025). Difshilling: A Diffusion Model for Shilling Attack. *Available at SSRN 4928287*.

[13] Yu, S., Duan, M., Wang, K., & Yang, S. (2025). RUP-GAN: A Black-Box Attack Method for Social Intelligence Recommendation Systems Based on Adversarial Learning. *Big Data Mining and Analytics*, *8*(4), 820-836.

[14] Thanh Toan Nguyen, Nguyen Quoc Viet Hung, Thanh Tam Nguyen, Thanh Trung Huynh, Thanh Thi Nguyen, Matthias Weidlich, Hongzhi Yin, "Manipulating Recommender Systems: A Survey of Poisoning Attacks and Countermeasures", *ACM Computing Surveys*, 2024.

[15] Dina Nawara, Ahmed Aly, Rasha Kashef, "Shilling Attacks and Fake Reviews Injection: Principles, Models, and Datasets", *IEEE Transactions on Computational Social Systems*, vol.12, no.1, pp.362-375, 2025.