Girish B. Desale, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 14, Issue 5, May, 2024, pp: 149-156

RESEARCH ARTICLE OPEN ACCESS

Comprehensive Study on Spam Review Detection Model using Linguistic and Behavioral Features

Girish B. Desale [#1], Swati K. Borse [#2]
[#1]*Assistant Professor & Head, Department of Computer Science & I.T., JET's Zulal Bhilajirao Patil College, Dhule, Maharashtra, India.*
[#2]*Associate Professor, Department of Computer Science SSVPS's Late K. Dr. P. R. Ghogrey Science College, Dhule, Maharashtra, India.*
[1]*girishdesale@gmail.com* , [2]*swatippatil28@gmail.com*

Abstract
In the field of electronic commerce, consumer reviews and ratings serve as primary sources for measuring public sentiment towards various items. Unexpectedly, the proliferation of fake reviews in digital shopping applications is steadily increasing. In general, vendors often have two distinct categories of criteria when it comes to distributing spam reviews. One strategy is to cultivate a favourable perception of their products. Furthermore, spreading unfavourable viewpoints on the merchandise of the rival. Many researchers have a keen interest in conducting surveys pertaining to the development of detection systems aimed at identifying spam reviews. The present study introduces a Comprehensive Study on Spam Review Detection (CSSRD) model, which aims to discern spam material by analysing its associated spam properties. Previous studies have indicated that spam detection strategies in prior surveys were mostly based on either linguistic or behavioural approaches. The system CSSRD that has been created incorporates both linguistic-based and behavioral-based methods. This study aimed to identify and categorize the 12 distinct characteristics of spam reviews, with each category consisting of 6 elements. Linguistic-based features rely on textual spam variations such as text similarity and emotion polarity. Behavioral-based elements are contingent upon the contextual data derived from the reviews. The determination of the outcome is contingent upon several factors such as the ratings assigned, the dates on which the reviews were posted, and the total number of reviews available. After identifying the distinguishing features of spam, we conducted a study utilising machine learning methodologies on the complete dataset, as well as separately on the linguistic features and behavioral features.

Date of Submission: 14-05-2024 Date of acceptance: 25-05-2024

Keywords: Online reviews, e-commerce spam reviews, spam review detection, behavioral-based, linguistic-based.

I. Introduction

The proliferation of spam reviews on e-commerce websites continues to rise on a daily basis. This can be caused by the growing popularity of applications for e-commerce and the services they provide. The users are better able to evaluate the products as a result of their opinions. Not only can the customers benefit from these comments, but the vendors can as well, as they can use them to improve product quality and develop more effective marketing strategies (Huang et al., 2013). Within the last couple of years, there has been an increase in the practice of spam review propagation. It can happen to anyone who hires people to submit fake reviews for what they sell or their competitors without being restricted in any way. According to Biradar et al. (2017), spammers have been responsible for publishing an immense number of product reviews in order to either alter consumers' perceptions of the goods or to promote specific brands. In previously conducted research, different kinds of spam review detection models have been investigated with the purpose of detecting spam reviews.

Research that has already been carried out has explored a variety of approaches to the detection of spam reviews. Linguistic techniques are based on text data, which means they depend on

the presence of text data in the reviews. The behavioural technique is not a text-based strategy, which implies that its application is dependent on the context data of the reviews.

The proposed system, known as CSSRD, incorporates both linguistic-based and behavioral-based spam models. Every model possesses a unique approach to discerning whether reviews should be classified as spam. Discover twelve distinctive characteristics from both models, where each category has six distinct characteristics. Features based on linguistics dependence on text-based spam variants. Various forms of text-based spam variations, including text similarity and sentiment polarity, etc. have been observed. When determining behavioral-based characteristics, the contextual data of the reviews, including ratings and dates, is considered. Once the characteristics of spam have been identified, machine learning algorithms are employed to analyse the complete dataset.

The major objective of this research is to present a taxonomy of the numerous techniques that can be utilised to identify spam comments left on online reviews. The following is a list of the significant contributions that were made to this work:

i. An analysis of the importance of spam review identification, as well as a study of the different forms of spam detection models
ii. Provide an overview of the effectiveness of the methods and technologies that are now available for the detection of spam reviews.

II. Literature Review

The process of segregating spam detection techniques for distinguishing between 'positive' and 'negative' content identification is accomplished through the analysis of diverse entities such as e-mails internet URLs, computer IP addresses, and other relevant factors. These models have the capability to operate on the client side of the systems or apps. The efficacy of these filtering algorithms is contingent upon the implementation of two primary strategies: whitelists and blacklists. In their study, Dada et al. (2019) discussed the practise of spam filter models incorporating a "whitelist" feature to mitigate the risk of legitimate content being mistakenly identified as spam or rejected. This whitelist functionality often allows for the inclusion of specified objects such as computer IP addresses, website domains, and email IDs, among others. The client-side of the programme automatically generates a list of identified "bad" objects as part of the blacklisting procedure, which is periodically updated. In several instances, search engines and users commonly categorise undesirable entities as being included in "blacklists," which encompass domain names, email addresses, and similar elements.

In their study, Elakkiya et al. (2020) conducted an analysis of previous research that utilised text classification algorithms for the purpose of identifying spam reviews. Subsequently, they put forth the Text Spam Detector approach. The Convolutional Neural Network (CNN) system use the Deep Learning technology to effectively identify spam and fake reviews. The proposed approach involves the classification of spam and ham reviews through the utilisation of tagged datasets for training purposes.

R. Hassan and M. R. Islam's (2019) method for identifying spam reviews employs linguistic-based criteria. Term frequency, sentiments, and word count were used as linguistic variables in this essay. Additionally, they applied machine learning techniques to the features in order to predict spam reviews. The Nave Bayes method generated the most favourable results, with an accuracy score of 83%, when they compared Nave Bayes and SVM models.

The present study investigated the efficacy of machine learning and deep learning algorithms in classifying reviews through the utilisation of a feature-based model. The categorization of text-based spam material is a prevalent issue, often addressed by employing deep learning algorithms that provide comprehensive solutions. The utilisation of deep learning algorithms in text data categorization has proven to be advantageous due to its superior accuracy compared to traditional machine learning techniques. In their study, Naveed et al. (2020) put up a Spam Review Detection

model that incorporates linguistic and behavioural aspects. The calculation is performed for each review in the dataset, based on its behavioural characteristics. By assigning normalised values to each behavioural parameter, the average score for the respective reviews in the complete dataset is calculated. Subsequently, the classification of a review as spam or non-spam is established by comparing this spam score with a variable threshold.

III. Proposed Methodology

Spam characteristics refer to the methods or behaviours that are derived from the patterns of spammers and subsequently compared with the patterns exhibited by regular users. Prior research has predominantly relied on utilising data mining or machine learning methods to compare review content with spam text. In these instances, our reliance is solely on text-based methodologies. In recent times, several studies have begun to make predictions regarding the filtering of spam in emails, the identification of phishing URLs, and the evaluation of e-commerce reviews. These studies employ a combination of text-based and context-based methodologies. This study employs a combination of linguistic, which depends on text-based variables, and behavioural, which are extracted based on context data, to forecast the occurrence of spam reviews. This study presents the architectural design of the CSSRD system, illustrating the methodology overview as depicted in Figure 1.

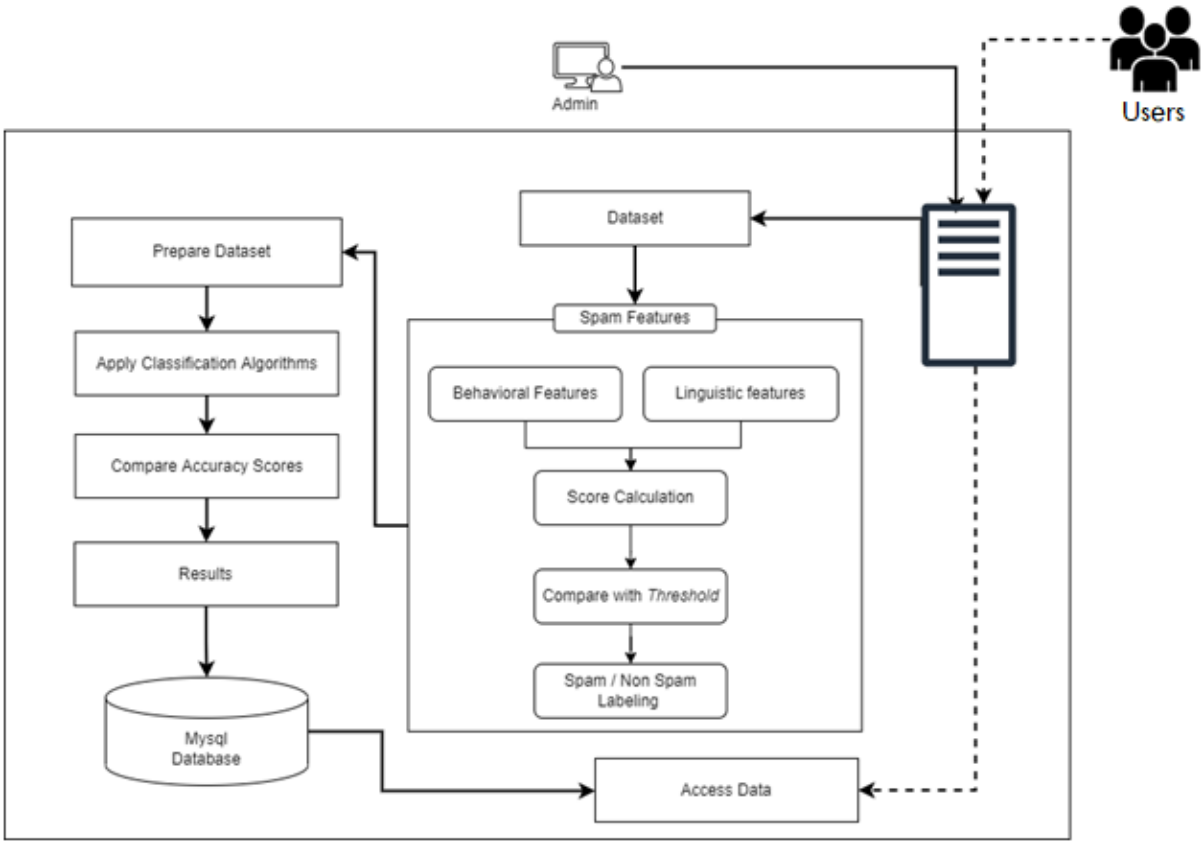

Figure 1: CSSRD Architecture. Within the field of architecture, there exist two distinct entities that play vital roles in the system: administrators and users. The administrator employs linguistic and behavioural characteristics to detect spam data. By utilising a dataset containing spam features, the administrator will employ classification algorithms to determine the most accurate algorithm for predicting spam in different scenarios. Users are those who use data for the purpose of accessing and reviewing product information. They are granted the ability to examine data pertaining to items and reviews that is submitted by the administrator.

3.1 Spam characteristics

3.1.1 Behavioral characteristics

The behavioral characteristics depends on the contextual data of the reviews, including ratings, dates, etc. Within this particular category, there exist six distinct features that are explicitly outlined in the Table 1.

Table 1: Behavioral characteristics

1	The maximum number of reviews [Naveed et al. (2019)]	This functionality computes the quantity of product reviews throughout a 24-hour period. According to this feature, the act of a person posting multiple reviews or ratings for a single product is classified as spam. $$if\ r_p \in \mathrm{d} > 1: res = 1$$ $$else: res = 0$$
2	Analysis of Rush [*Shehnepoor et al. (2017)*]	This feature facilitates the computation of the temporal duration between successive product reviews. If a user submits repeated reviews within a timeframe that falls below the specified threshold, it is classified as spam. $$if\ date\_diff(u_r) \in \mathrm{p} > threshold\ t: res = 1$$ $$else: res = 0$$
3	Examination of Daily Review Count [Naveed et al. (2019)]	This function performs a computation to determine the aggregate number of reviews and ratings attributed to various products. If a user submits several reviews that exceed the predetermined threshold value, such reviews are classified as spam. $$if\ count(u_r) > threshold\ t: res = 1$$ $$else: res = 0$$
4	Evaluation of Individual Product Review [*Hussain et al. (2020)*]	If an account has submitted multiple reviews or ratings for the same item, it might be classified as spam. $$if\ u_r == p: res = 1$$ $$else: res = 0$$
5	Assessment of Rating Variability [*Shehnepoor et al. (2017)*]	Any ratings that differ from the mean rating of an item are classified as spam. $$if\ (u_r - mean(p_r)) \geq \mathrm{threshold}\ t: res = 1$$ $$else: res = 0$$
6	The highest or lowest rating . [*Hussain et al. (2020)*]	This feature will take into account instances of spam when users submit ratings that are either 1 or 5. $$if\ u_r\ has - only\ [1,5]: res = 1$$ $$else: res = 0$$

3.1.2 Linguistic characteristics

The linguistic characteristics depends on the text data of the reviews, including word count, sentiment, etc. Within this particular category, there exist six distinct features that are explicitly outlined in the Table 2.

Table 2: Linguistic characteristics

1	Examination of Content Similarity [*Naveed et al. (2019)*]	This function identifies statements that have been said previously or statements with substance that is similar and matches them with the threshold. $$if\ match(u_r) < threshold\ t : res = 1$$ $$else: res = 0$$
2	Proportion of Positive Reviews [*Hussain et al. (2020)*]	It is considered spam when a user publishes many reviews all containing positive feedback from the same individual. $$if\ senti(u_r) == 'pos' : res = 1$$ $$else: res = 0$$
3	Proportion of Negative Reviews [*Hussain et al. (2020)*]	It is considered spam when a user publishes many reviews all containing negative feedback from the same individual.. $$if\ senti(u_r) == 'neg' : res = 1$$ $$else: res = 0$$
4	The length of the review	A review is classified as spam if its entire letter count is below the threshold value. $$if\ length(u_r) < threshold\ t : res = 1$$ $$else: res = 0$$
5	The portion of capital letters	It is assumed that any review that is written entirely in capital letters has been tampered with and should be ignored. $$if\ u_r.isUpperCase() : res = 1$$ $$else: res = 0$$
6	The use of the second person pronoun and exclamation mark	If a review is written in the second person and contains punctuation like question marks or exclamation marks (!), then it is deemed spam.

3.2 Detection of Spam reviews

With the existing review dataset, the system performs its own computations on the dataset in order to determine all 12 features, after which it will record the results of each feature. When all 12 components of the total data have been successfully captured, the system will proceed to do the mean computation. This work calculated the mean score for the collective characteristics. Upon calculating the mean of all the reviews, the subsequent procedure involves evaluating each individual mean score in relation to the predetermined threshold value. The steps of detection of spam reviews are mentioned in the following.

Step 1: Calculate characteristics results for reviews

Step 2: Mean calculation

$$\text{Mean of the review characteristics (M)} = \frac{\sum_n (c1,c2,\ldots c12)}{n}$$

Step 3: Labelling

It involves evaluating individual mean scores for all reviews and in relation to the predetermined threshold value.

$$if\ M \geq threshold\ t: Labeling\ = 1$$

$$else: Labeling\ = 0$$

3.2 Classification Analysis

At this point in the research process, four distinct classification algorithms are to be employed on the feature dataset with the aim of acquiring knowledge about the two primary research questions of the study. Which classification method has the highest level of reliability in predicting spam reviews? When considering the prediction of spam reviews, which category of features is the most suitable? In this particular case, this work employed four distinct classification algorithms, each of which is described in more detail below.

- Naïve Bayes algorithm
- Decision Tree algorithm
- Support Vector Machine algorithm
- Neural Network algorithm

3.3 Dataset

In order to identify spam reviews pertaining to e-commerce products, it is necessary to obtain a dataset including specific information. This dataset should include the UserID of the individual responsible for publishing the review, the date on which the review was posted, the unique identifier of the product being reviewed, the content of the review itself, and the corresponding rating assigned to the product. The Amazon reviews public dataset from Datafiniti (2018) has been utilised in accordance with the project criteria.

IV. Results

The spam characteristics dataset was divided into a 70:30 ratio for the purposes of classification analysis. The classification analysis was then conducted in three different modes.

1. Total characteristics with label (c1, … c12, Label)
2. Behavioral characteristics with label (c1, … c6, Label)
3. Linguistic characteristics with label (c7, … c12, Label)

The results of accuracy of classification algorithms are mentioned in the Table 3.

Table 3: Accuracy Results

Algorithm	Overall Accuracy	Behavioral characteristics Accuracy	Linguistic characteristics Accuracy
Neural Network	97.86772487	97.20634921	96.10405644
Naive Bayes	92.10758377	92.72486772	91.75485009
SVM	92.6366843	90.82892416	92.37213404
Decision Tree	91.97530864	92.54850088	91.84303351

In this experimental study, it was shown that the Neural Network method exhibited superior accuracy in comparison to other machine learning models for overall data. The present investigation additionally conducted a calculation and comparison of performance results between data pertaining to linguistic features and data pertaining to behavioural features. In the conducted

experiment, it was observed that the Neural Network algorithm exhibited enhanced accuracy in processing individual data. Furthermore, the disparity in accuracy between the linguistic and behavioural data was found to be minimal. Through conducting this experiment, it was determined that both spam characteristics have a significant role in the identification of spam reviews.

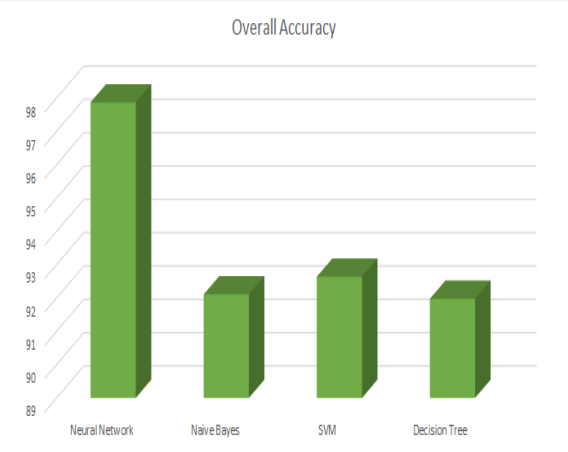

Figure 2: Overall Accuracy Graph

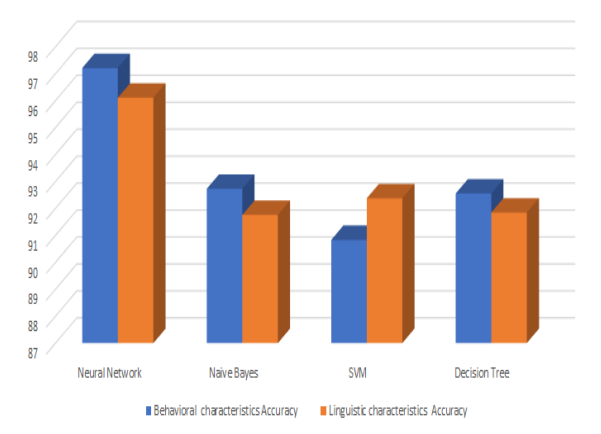

Figure 3: Comparison of linguistic features and behavioural features results

V. Conclusions

The success of purchasing goods online is contingent upon the evaluations and ratings provided by users on the products. The impact of these feedback methods on product sales is evident, as evidenced by the growing prevalence of spam reviews. Prior research has mostly concentrated on the development of network-based and text-mining-based models to identify spam reviews. This study primarily examines linguistic- and behavioural-based variables and assesses the significance of these features in spam detection through the application of classification models. Drawing from many prior studies, this paper aims to delineate the twelve distinct spam traits derived from linguistic and behavioural data. This study delineated the approaches employed in identifying spam features and subsequently applied them to a real-time dataset of e-commerce reviews. The spam reviews were found by comparing their features with the mean computation. In the context of classification analysis, the objective was to determine the most accurate machine learning model for a dataset including attributes related to spam. The results indicate that the neural network algorithm achieved a high level of accuracy, specifically 97%, for both linguistic and behavioural features. Based on the results of the classification study, it has been determined that both spam characteristics play a significant role in the identification of spam reviews. In subsequent research endeavours, it is recommended to incorporate the integration of multilingual reviews as a means to identify spam reviews through the utilisation of linguistic-based characteristics.

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.