Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

SPAM DETECTION BASED ON FUSION OF SPAMMER BEHAVIOR FEATURES AND LINGUISTIC FEATURES

AMNA IQBAL

Ph.D Candidate, Computer Science at Government College University Faisalabad, Pakistan.
Email: amna_iqbal133@hotmail.com.

MUHAMMAD YOUNAS

Assistant Professor, Computer Science Department, Government College University Faisalabad Pakistan. Corresponding Author E-Mail: younas.76@gmail.com

RAMZAN TALIB

Professor and Chairman of the Department of Computer Science, Government College University Faisalabad (GCUF), Pakistan. E-mail: ramzan.talib@gcuf.edu.pk

MUHAMMAD MURAD KHAN

Assistant Professor with the Department of Computer Science, Government College University Faisalabad, Pakistan. E-mail: muradtariq.tk@gmail.com

BUSHRA ZAFAR

Assistant Professor in Department of Computer Science at Government College University Faisalabad (GCUF), Pakistan. E-mail: bushrazafar@gcuf.edu.pk

Abstract

E-commerce sites, forums, and blogs have become popular platforms for people to share their views. Reviews have emerged as a crucial source of information for potential customers, influencing their purchasing decisions. Similarly for profit gain or fame, Spam reviews are deliberately written with the intention of defaming businesses or individuals. This act is known as review spamming. Spam review detection is rapidly answered by various ML techniques. Review of spamming is more challenging task in multilingual communities. Spammer behavior features and linguistic features often exhibit complex relationships that influence the nature of spam reviews. The unified representation of features is another challenging task in spam detection. Various deep learning approaches have been proposed for review spamming, including different neural networks (Convolutional Neural Network, CNN). These methods are specialized in extracting the features but lack to capture feature dependencies effectively with other features. Spam Review Detection using the Fusion Gradient Boosting Model (SRD-FGBM) is proposed with fusion of spammer behavior features and linguistic features to automatically detect and classify the spam reviews. Fusion enables the proposed model to automatically learn the interactions between the features during the training process, allowing it to capture complex relationships and make predictions based on both types of features. It apparently shows the promising result by obtaining **94.3%** accuracy.

Index Terms: Review Spamming, Linguistic features, spammer behavior features, Classification, Feature engineering, SVM, Gradient Boosting Model (GBM), Fusion.

1. INTRODUCTION

E-commerce sites, forums, and blogs are main source where users put their opinion in the form of review [1]. Online reviews hold significant importance for both customers and vendors, influencing purchasing decisions and shaping future strategies [2]. Rapid spam review attacks have become a growing concern, where anyone can write fake reviews to

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

promote products or services, leading to financial consequences for businesses and loss of trust [3], [4] ,[5]. Spammers exploit opinion sharing websites to create hype and manipulate the value of a product or service [6]. Detecting spam reviews is critical to maintain the integrity of online review sites, and major platforms like Yelp and Amazon have made progress in addressing this issue [7]. While researchers have proposed various techniques for spam review detection, there is still room for improvement, particularly with real-world datasets [8], [9]. Spam reviews differ from web or email spam as they provide misleading opinions about products or services, making manual detection challenging [10]. Existing approaches in web or email spam detection are not suitable for identifying spam reviews [11], [12]. The detection of spam reviews relies on analyzing spammer behavioral features and review text to differentiate between legitimate and spam reviews [13]. The prevalence of unsolicited and irrelevant messages across diverse digital channels underscores the need for the creation of effective and precise mechanisms for identifying and filtering out spam content. Linguistic attributes, such as textual configurations, affective assessment, and semantic data, furnish significant indicators for the detection of unsolicited and unwanted content [14]. TF-IDF captures term importance, BoW represents word presence/absence, CHI2 identifies significant linguistic features, and Word2Vec [15]captures semantic relationships. These linguistic features provide different perspectives and can contribute to detecting review spam based on linguistic characteristics [16]. Conversely, characteristics of spammer conduct such as the rate of posting, modes of interaction, and questionable behaviors provide valuable perspectives into the actions of prospective spammers. However, previous studies have often focused on linguistic methods or behavioral characteristics separately when identifying spammers and spam reviews. To overcome these limitations, this research aims to utilize fusion along with Gradient Boost method (GBM) techniques for accurate analysis of spam reviews and incorporate a comprehensive set of behavioral and linguistic features to filter spam and not-spam reviews.

This research paper answers the following research questions.

QR1. How does the fusion of linguistic characteristics and spammer behavior attributes improve the accuracy of review spam detection compared to conventional single-feature techniques?

QR2. What machine learning or statistical techniques can be employed to effectively integrate linguistic features and spammer behavior features for accurate spam detection?

QR3. How does the fusion of the GBM (Gradient Boosting Machine) model with combined spammer behavior and linguistic features improve the accuracy of spam review detection?

QR4. Can linguistic features and spammer behavior features contribute to the identification of spam reviews?

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

QR5. Can the fusion of linguistic features and spammer behavior features enhance the ability to differentiate between legitimate users and spammers in online communities?

This research contributes to contribute to answering the above research questions the research question pertains to the fusion methodology that combines linguistic characteristics and spammer behavior attributes as its main contribution. The integration of linguistic features and spammer behavior features aids in the detection of evolving spamming techniques and patterns. This study aims to evaluate the effectiveness of the fusion approach in enhancing the accuracy of review spam detection, in comparison to traditional single-feature methods. It also aims to highlight the advantages of the fusion methodology over traditional approaches. The fusion approach has been empirically proven to be effective in terms of enhanced accuracy and robustness, as demonstrated through rigorous testing and analysis. The fusion technique in spam detection, as compared to linguistic-only or behavior-only approaches, through a comparative analysis of their respective performances. It also facilitates the identification and mitigation of novel forms of spam by enabling the detection system to capture emerging patterns or behaviors, thereby enhancing its efficacy against the constantly evolving tactics employed by spammers.

The present research paper expounds upon the pragmatic implications of the fusion approach in the context of spam detection systems that are operational in the real-world. The paper offers valuable insights for researchers and practitioners who aim to develop more dependable and efficient spam detection mechanisms by showcasing the advantages of integrating linguistic features and spammer behavior features.it also makes noteworthy contributions by introducing a new fusion approach, evaluating its effectiveness through empirical means, conducting a comparative analysis with individual feature-based methods, demonstrating its adaptability to evolving techniques, and highlighting its practical implications for spam detection systems. The aforementioned contributions serve to propel the domain of spam detection and furnish significant perspectives for forthcoming research and development endeavors.

The format of this research paper is as follows: Review of the literature is presented in Section 2. The study's methodology is described in Section.3. The findings and discussion are presented in Section.4. The conclusion is found in Section 5. The research's limitations and future work are finally covered in Section.6.

2. LITERATURE REVIEW

Existing research has explored various approaches for detecting spam reviews, focusing on two main methods: spammer behavioral analysis and linguistic analysis. Previous studies have investigated spam review detection by analyzing patterns and relationships among spammers. However, only a few studies have explored this method. Mukherjee et al.[17] used clustering techniques to model reviewer spamicity and identify spammer clusters. Heydari et al. [18] focused on time series features of reviewers in an Amazon dataset. Kc and Mukherjee [19] developed a text mining model using unsupervised

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

approaches and semantic language models. Li et al. [20] proposed an unsupervised model based on review posting rate and temporal patterns. Dematis et al.[21] employed a network model to capture correlations among users and products. Most of these studies only utilized time series-based behavioral features, but incorporating a richer set of features can enhance spammer identification. The problem of spam review detection was first studied by Jindal and Liu [7], who analyzed review texts from Amazon.com and identified duplicated content used by spammers. Lau et al. [22], [23] applied semantic language models and the Support Vector Machine classifier. Li et al.[24] employed supervised learning with co-training methods. Fusilier et al. [25] proposed a classification method based on N-gram characters and Naïve Bayes. Ott et al. [26] designed a dataset for spam review detection and incorporated psycholinguistic features. Hazim et al. [27] used statistically based features and different models for multilingual datasets. Kumar et al. [28] proposed a hierarchical supervised learning method, while Zhang et al. [29] recommended a supervised model based on reviewer features. Ahmed and Danti [30] utilized rule-based machine learning algorithms, and Lin et al. [31] Employed time-sensitive features and SVM. Li et al. [32]used a feature-based sparse additive generative model and the SVM classifier. The spammer behavior features used in the literature were presented in a Table.1.

Table 1: Spammer Behavior Features

Features	Description	Features	Description
F1	Content similarity	F8	The ratio of first reviews
F2	Number of reviews per day	F9	Review of a single product
F3	Review burstiness	F10	Deviation in rating
F4	Activity window	F11	Length of the review body
F5	Review count	F12	Extreme rating
F6	% Positive opinion words	F13	% Of capital words used
F7	% Negative opinion words	F14	Duplicate/near-duplicate reviews

Table 2 presents the comparative summary of state-of-the-art spammer behavior feature are discussed.

Table 2: Comparative summary of state-of-the-art spammer behavior feature

Sr	Reference	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	Accuracy
1	[4]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		92%
2	[33]				✓	✓					✓				✓	81%
3	[34]	✓	✓				✓				✓	✓				83%
4	[28]		✓		✓	✓					✓					81%
5	[35]	✓		✓		✓			✓		✓					AUC

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

It appears that the literature explores different approaches of, including TF-IDF, Bag-of-Words, CHI2, and Word2Vec, for classification or spam detection showed in Table.3.

Table 3: Linguistic Feature used for classification in state-of-the-art studies

Reference	TF-IDF	BoW	CHI2	Word2Vec
[36]		✓		
[15]				✓
[37]		✓		
[38]				✓
[39]				✓
[40]	✓			✓
[41]	✓			
[42]	✓			
[43]	✓			
[4]	✓			
[44]			✓	
[45]	✓		✓	
[46]				✓

Most of these studies did not consider important linguistic features and utilized only one classifier, so the literature shows the research gape that without fusion of linguistic features and spammer behavior features in spam detection offers improved accuracy, robust detection capabilities, synergistic information, adaptability to evolving techniques, generalization across domains, and enhanced countermeasures. These benefits contribute to the development of more effective and efficient spam detection systems, ensuring better online security and user experience

3. METHODOLOGY

The purposed frame work used state-of the-art dataset describe in Table.3. The spammer behavior features and linguistic are calculated in next step, feature scaling is applied in both types of features then fusion vectors is calculated using the both feature then the fuse vectors is used along with the gradient boosting classifier for classify the spam review or not spam review. The purposed framework **SRD-FGBM** is defined step by step in the following and the graphical representation is showed in Fig1.

3.1 Data set

The present study utilizes an authentic Amazon product review dataset, which encompasses the extensive behavioral and posting chronicles of the reviewers. The corpus comprises of 26.7 million reviews, 15.4 million reviewers, and 3.1 million products, which are predominantly classified into six distinct categories. Table 1 provides a comprehensive breakdown of the dataset's distribution across various categories, reviews, reviewers, and products. The identification of spam reviews through linguistic methodology necessitates a dataset that has been labeled for the purpose of training the classifier. However, the Amazon product review dataset utilized in this labeled investigation.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

Table 4: data set description

Category	Total Reviews	Total Reviewers	Total Products
Cell Phones and Accessories	3446396	2260636	319652
Clothing, Shoes, and Jewellery	5748260	3116944	1135948
Electronics	7820765	4200520	475910
Home and Kitchen	4252723	2511106	410221
Sports and Outdoor	3267538	1989985	478846
Toys and Games	2251775	1342419	327653
Total	26787457	15421610	3148230

The SRD-FGBM framework conducts various tasks such as data pre-processing, tokenization, review content analysis, feature extraction and selection, and classification. These tasks are accomplished through the utilization of the Natural Language Toolkit (NLTK) version 3.0. NLTK offers convenient built-in text processing libraries that are user-friendly.

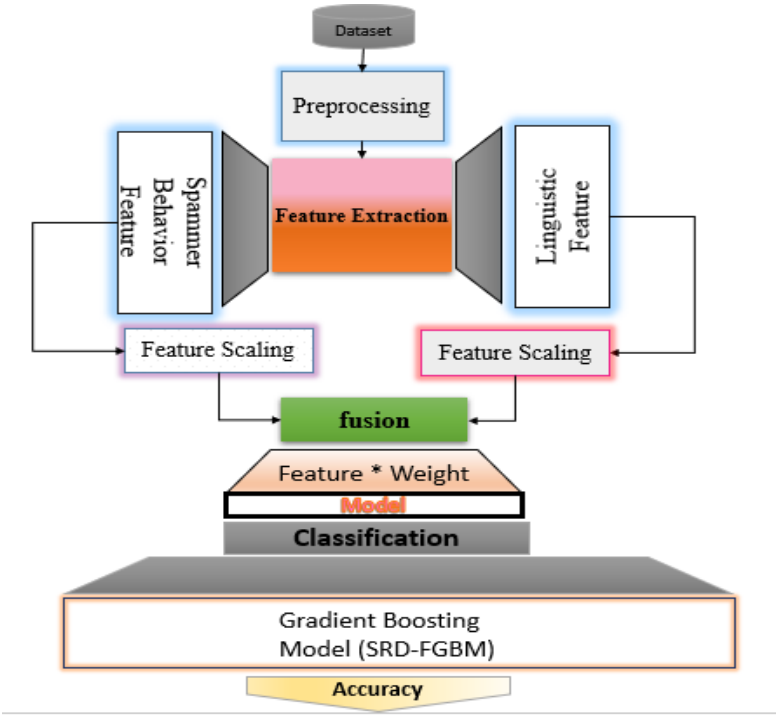


Figure 1: SRD-FGBM Framework for Review Spam Detection

3.2 Spammer Behavior Feature Calculation

The behavioral traits of a reviewer can serve as indicators of their association with spamming activities. Consequently, these traits can be utilized to differentiate between

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

spam and non-spam reviews. The aforementioned characteristics may serve as cues for detecting spammers and should not be regarded as definitive criteria for classifying a reviewer as a spammer or non-spammer. Hence, the suggested methodology employs an extensive array of behavioral characteristics and avoids dependence on a solitary behavioral trait for the identification of spammers. This section provides a discussion of the spammer behavioral features. In this research, the utilization of spammer behavior features was combined with linguistic features to separate spam reviews from non-spam reviews. The purpose was to fuse these two types of features in order to develop a method that could accurately differentiate between the two categories. The following Spammer behavior feature are utilized in **SRD-FGBM**

Calculating the spammer behavior features from a review dataset requires analyzing the relevant attributes and applying specific calculations. Here's an overview of how each feature computed:

Table 5: Notations used in this methodology

R	Reviewer
r	Review
T_r	Total number of reviews
$R_i(r_j)$	Refers to a review written by reviewer R_i
$MR(r_j)$	Maximum reviews written by a reviewer
L $R_i(r_j)$	The Last date of the report authored by reviewer R_i.
$+iv\ OW$	Number of positive words in r
$+iv\ OW$	Number of negative words in r
Tnr	Total number of words in r
$R_i(r_{first})$	Represent the first review of a reviewer
P	Products
NS	Not spam
S	Present the spam
W	All reviews of reviewer R
OW	Opinion Words

3.2.1. Content Similarity (CS):

Cosine similarity used to measure the textual or semantic similarity between pairs of reviews [47], [48]. Spammers often opt to copy reviews from similar products due to the time-consuming nature of generating new reviews. Therefore, it is advantageous to employ cosine similarity to identify the similarity in content between reviews written by the same reviewer. To identify the most undesirable behavior of spammers, in this research the maximum similarity approach employed the equation for the maximum similarity approach defined in following Eq.1.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

$$F_{CS} = \mathrm{CS}(R_i) \quad (1)$$
$$= Max\left(Cosine\left(R_i(r_j), R_i(r_k)\right)\right) \; where \; R_i(r_j), R_i(r_k) \in R_i(T_r)$$

In this equation, $R_i(r_j)$ and $R_i(r_k)$ represent two reviews written by reviewer Ri from the set of reviews $R_i(T_r)$. The cosine similarity between $R_i(r_j)$ and $R_i(r_k)$ is computed using a cosine similarity function.

3.2.2. Maximum Number of Reviews per Day (MNR)

Posting multiple reviews in a single day can be seen as a sign of deviant behavior [47], [49]. This indicator quantifies the reviewer's maximum daily review count, normalized by the overall maximum value in our dataset.

$$F_{MNR}(R_i) = \frac{MR(R_i)}{M_{R_i \in R_i(T_r)} MR(R_i)} \quad (2)$$

3.2.3. Review Burstiness (RB)

Authentic reviewers periodically publish their reviews from their personal accounts, while opinion spammers are characterized by their recent membership on the site. Currently, account's activity utilized to detect and capture instances of spamming behavior. The reviewing burstiness is defined as the difference between the first and last dates of review creation, also known as the activity window. If the time frame for a posted review is reasonable, it could include a typical activity. However, when reviews are posted in a short period of time (specifically within 28 days, as estimated in [20]), there is evidence of spam behavior occurring.

$$F_{bs}(R_i) = \begin{cases} NS & L(R\_i\,(r)) - f(R_i((r)) > \acute{r} \\ \dfrac{L(R_i(r)) - f(R_i((r))}{\acute{r}} & Otherwise \end{cases} \quad (3)$$

In the above Eq.3 $L(R_i(r))$ denotes the most recent date on which the reviewer R_i posted a review ŕ, while $F(R_i(r))$ represents the initial date of posting for the same review.

3.2.4. Percentage of Positive Opinion Words (PPW)

Calculate the percentage of positive sentiment words or expressions within each review [50], [51].

$$F_{PPW} = \frac{Number\ of + iv\ OW}{tnr} \times 100 \quad (4)$$

3.2.5. Percentage Negative Opinion Words (PNOW)

Calculate the percentage of negative sentiment words or expressions within each review[50], [51] .

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

$$F_{PNW} = \frac{Nunber\ of\ -iv\ OW}{tnr} \times 100 \quad (5)$$

3.2.6. Ratio of First Reviews (RFR)

People tend to rely on the initial reviews in order to benefit from them [51]. Spammers create email accounts early on to impact initial sales. Spammers believe that controlling initial product reviews gives them the ability to manipulate public opinion. We calculate the ratio between the initial reviews and the total reviews for each author. The term "first reviews" refers to the initial evaluations of a product that are posted by the author.

$$F_{RFR} = \frac{\left|R_i(r_{first}) \in R_i(Tr)\right|}{R_i(Tr)} \quad (6)$$

3.2.7. Review of a Single Product (RSP)

If a reviewer posts multiple reviews about the same product, it can be indicative of spam behavior [48]. The mathematical formula for representing a review about a single product would be:

$$F_{RSP} = r, \text{where rP} \in \text{R(P)} \quad (7)$$

In this equation, rP represents a review specifically related to the product P. The notation "rP ∈ R (P)" indicates that the review r belongs to the set of reviews written by reviewer R for the product P.

3.2.8. Extreme Rating(ER)

Identify reviews with ratings at the extreme ends of the rating scale [47], such as extremely positive or extremely negative ratings spammer behavior feature detected using the Eq.8.

$$F_{ER} = \begin{cases} S, * (r_i) \in \{1, 5\} \\ NS * (r_i) \in \{1, 5\} \end{cases} \quad (8)$$

3.2.9. Deviation in Rating (DR):

A rational user is anticipated to provide a rating that aligns with the rating given by another reviewer for a comparable product [50] , [47], [49], [51]. Previous research has found that spammers tend to provide ratings that differ from those of genuine reviewers in order to manipulate the perception of a product, either positively or negatively. The product's mean rating value is determined using Equation (9). Next, the normalized score, also known as the rating deviation, is calculated using the mean value according to Equation (10).

$$MEAN_r = \frac{\sum_{x==1}^{|w_r|} \left| * r_p \right|}{w_r} \quad (9)$$

$$F_{DR} = \frac{\left| * r_p - MEAN_r \right|}{4} \quad (10)$$

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

3.3. Linguistics Features

The analysis of review text employs linguistic features. The aforementioned characteristics have the potential to aid in the detection of fraudulent reviews through analysis of the language utilized by the reviewers. This section employs a four-step approach to ready the dataset for the application of review-related features. (i) The process of identifying the grammatical category of each word in a given text is known as part of speech tagging. (ii) The elimination of all punctuation marks from a text is a common pre-processing step in natural language processing. (iii) Stemming refers to the process of reducing words to their root form, which can help to reduce the dimensionality of a text dataset (iv) word2vec used as a method for selecting relevant features from a text dataset. Word2Vec decreases the dimensionality of word representations in contrast to conventional one-hot encoding or sparse representations. Word2Vec employs lower-dimensional dense vectors to represent words instead of high-dimensional and sparse vectors. This decreases the computational complexity and memory demands for subsequent natural language processing (NLP) tasks.

3.4. Data Preparation

In dataset “X_SB” be the matrix representing the Spammer Behavior features, where each row corresponds to a review and each column represents a specific spammer behavior feature. Let “X_L” be the matrix representing the linguistic features, where each row corresponds to a review and each column represents a specific linguistic feature. y be the vector of labels indicating whether each review is spam “1” or not spam “0”.

The Complexity of this step is: O (1)

3.4.1. Feature Engineering: For features extracting each review in dataset DS, extract the spammer behavior features which consist of **CS, MNR, RB, PPW, PNOW, RFR, RSP, ER, RD** calculated using the above Eq.1 to Eq.10. Let's denote the set of spammer behavior features for reviews as S_i. For each review in dataset DL, extract the linguistic features. Let's denote the set of linguistic features for reviews as L_i.

3.4.2. Preprocessing: By applying feature scaling to spammer behavior features and using word embeddings for linguistic features, it leveraged the strengths of each technique to enhance the representation and capture important patterns within both types of features.

a) Feature Scaling

Applying feature scaling to the spammer behavior features (X_SB) using standardization

Mean Calculation is performed by calculating the mean (μ) for each feature in the spammer behavior features matrix X_SB.

$$\mu_i = \frac{\mathrm{sum}(\mathrm{X\_SB}[:, i])}{N} \qquad (11)$$

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

Where:

- μ_i represents the mean of the i^{th} feature.
- X_SB [:, i] represents the i^{th} column of the spammer behavior features matrix.
- N represents the total number of samples (reviews) in the dataset.

b) Standard Deviation Calculation:

Calculate the standard deviation (σ) for each feature in the spammer behavior features matrix X_SB.

$$\sigma_i = \frac{\text{sqrt}(\text{sum}((\text{X\_SB}[:, i] - \mu_i)\hat{}2)}{N} \quad (12)$$

Where:

- σ_i represents the standard deviation of the i^{th} feature.
- X_SB [:, i] represents the i^{th} column of the spammer behavior features matrix.
- μ_i represents the mean of the i^{th} feature.

N represents the total number of samples (reviews) in the dataset.

c) Transforming Features

Standardization is obtained by Transform each feature in the spammer behavior features matrix X_SB to have zero mean and unit variance using the standardization formula in Eq.13.

$$\text{X\_SB\_scaled}[:, i] = \frac{(\text{X\_s}[:, i] - \mu_i)}{\sigma_i} \quad (13)$$

- X_SB_scaled[:, i] represents the ith column of the scaled spammer behavior features matrix.
- X_SB [:, i] represents the i^{th} column of the spammer behavior features matrix.
- μ_i represents the mean of the i^{th} feature.
- σ_i represents the standard deviation of the i^{th} feature.

Complexity: O (n_samples * n_Features_s)

3.4.3. Feature Fusion:

Let X_SB and X_L be spammer and linguistic features respectively for each review in dataset. Concatenation is performed. Eq. 14 concatenate the spammer behavior features and linguistic features into a single feature vector for each review:

$$\text{X\_Fusion} = [\text{X\_SB}, \text{X\_L}] \quad (14)$$

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

The resulting unified representation

$$\mathrm{X\_Fusion} = \mathrm{f}(\mathrm{n\_samples}, \mathrm{n\_Features\_SB} + \mathrm{n\_Features\_L}) \quad (15)$$

In Eq. (15) presents the n_samples is the number of reviews and n_Features_SB and n_Features_L are the number of spammer behavior and linguistic features, respectively.

Weighted Combinations are designed by Assigning weights w_s to the spammer behavior features and weights w_l to the linguistic features based on their relative importance. In Eq. (16) showed the multiply each feature in X_SB by its corresponding weight. Using the Eq. (17) multiply each feature in X_L by its corresponding weight. The Sum up the weighted features to create a fusions unified representation as showed in Eq. (18). The complexity is measured by using the Eq.19.

$$\mathrm{X\_SB\_weighted} = \mathrm{X\_SB} * \mathrm{w\_s} \quad (16)$$

$$\mathrm{X\_L\_weighted} = \mathrm{X\_L} * \mathrm{w\_l} \quad (17)$$

$$\mathrm{X\_Fusion\_weighted} = \mathrm{X\_SB\_weighted} + \mathrm{X\_L\_weighted} \quad (18)$$

$$Complexity: O(n\_samples * (n\_Features\_SB + n\_Features\_L)) \quad (19)$$

Feature Interaction is performed by feeding both X_SB and X_L separately as inputs to the GBM model. The complexity is measured by using the Eq.19.

3.5. Spam Review Detection using the Fusion Gradient Boosting Model (SRD-FGBM)

Using the fusion step combines the spammer behavior features X_SB and linguistic features X_L into a single vector, allowing the model to capture information from both types of features simultaneously. Gradient Boosting is then applied to this fused vector to learn the relationships between the features and the labels algorithm SRD-FGBM showed in Fig.2. The spammer behavior matrix and linguistic features were used as separate inputs for the GBM (Gradient Boosting Machine) model. Each input was fed into a separate GBM model. Afterward, the outputs from both models were combined or merged, and this combined output was then used as the final input for the model's further processing or analysis. GBM model learn the interactions between the spammer behavior and linguistic features during the training process. The data is partitioned into training and testing sets, and the model is trained on the training data to minimize error. The trained model predicts labels for the testing data. The prediction step's complexity is determined by the size of the testing data in terms of samples and features.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

Algorithm: *SRD-FGBM Model for Spam Review detection*

Input: *Fused vector* — *- X_SB: Spammer Behavior Features*; *- X_L: Linguistics Features*

Output: *Spam Review Classification*

1. Split the data into training and testing sets:

- *Split X_SB and X_L into X_train_SB, X_test_SB, X_train_L, X_test_L*

2. Train the SRD-FGBM model:

2.1. Initialize the GBM model

2.2. Fit the SRD-FGBM model on the training data:

- *X_train_SB: Spammer Behavior Features for training*
- *X_train_L: Linguistics Features for training*

3. Predict using the trained model

3.1. Predict the labels for the testing data:

- *X_test_SB: Spammer Behavior Features for testing*
- *X_test_L: Linguistic Features for testing*

3.2. Store the predicted labels in y_pred

4. Accuracy calculation

6. Over all complexity $O(num\_iterations * n\_samples * (n\_Features\_SB + n\_Features\_L))$.

Figure 2: Algorithm for proposed SRD-FGBM

4. RESULTS AND DISCUSSION

The finding of this research provides an inclusive analysis of spam review detection methods, encompassing framework Fusion based GBA (SRD-FGBM) consist of spammer behavioral and linguistic approaches. The research findings indicate that fusion of these two methods enhances accuracy in spam review identification. The fusion technique demonstrated promising results, achieving an accuracy in differentiating between genuine and spam reviews by combining the strengths of both approaches. This approach utilized behavioral features to capture anomalies and relationships among spammers during fusion. It also employed linguistic features, including word2vec and stemming techniques, to analyze review content using the Gradient Boost method. Furthermore, the fusion technique evaluated the performance of several classifiers, such as Naïve Bayes, Logistic Regression, sport Vector machine, Logistic Regression, and Random Forest, in order to improve the accuracy of spam review prediction. Fig. 3 shows the fusion-Based results accuracy comparison with deferent classifiers and which study utilized which feature, and which classifier was employed. What were the outcomes or results. In the previous studies, linguistic features and spammer behavior features were used separately with various classifier models such as Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and the Mean

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

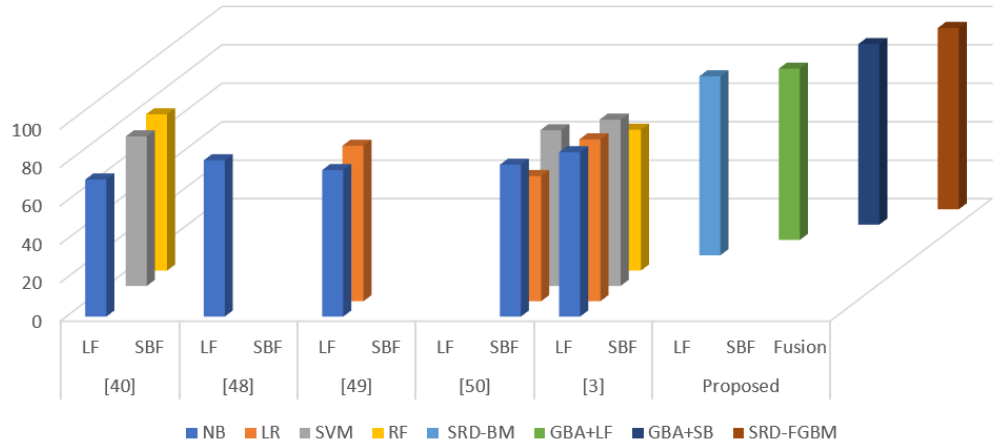
Value Method. In this research, combined linguistic features and spammer behavior features with the GBA algorithm. The accuracy achieved when employing the GBA algorithm in conjunction with spammer behavior features was 87.2%. When the GBA algorithm was integrated with linguistic features, the resulting accuracy was determined to be 82.1%. Additionally, by combining the features related to spammer behavior and linguistic characteristics, and employing the Fusion based GBA (SRD-FGBM) algorithm, were able to attain a peak accuracy of 94.3%. This outcome highlights the effectiveness of integrating both spammer behavioral and linguistic models. The SRD-FGBM framework has exhibited significant promise in effectively addressing the issue of spam reviews and offering more dependable information for both users and businesses.

Table 6: Comparison results on evaluation metrics of proposed approach and state-of-the-art study [40], [48], [49], [50], [3]. Key: Linguistic Feature-LF, Spammer Behavior Feature-SBF, Mean Value- MV, Spam Review Detection - Fusion Based GBA (SRD-FGBM)

	[44] Accuracy	**[52] Accuracy**	**[53] Accuracy**	**[54] Accuracy**	**[4] Accuracy**		**Proposed Methodology Accuracy**		
	Features	Features	Features	Features	Features		Features		
Model	LF	LF	LF	SBF	LF	SBF	LF	SBF	Fusion
NB	71.27	81.3	76.2	79	85.5	×	×	×	×
LR	×	×	80.9	65.1	84.2	×	×	×	×
SVM	77.81	×	×	81	86.5	×	×	×	×
RF	81.3	×	×	×	73.3	×	×	×	×
SRD-BM	×	×	×	×	×	93.1	×	×	×
GBA+LF	×	×	×	×	×	×	89.1	×	×
GBA+SB	×	×	×	×	×	×	×	93.9	×
SRD-FGBM	×	×	×	×	×	×	×	×	**94.3**

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

Figure 3: Fusion-Based results comparison with deferent classifiers in term of Accuracy



5. CONCLUSION

The Fusion of spammer behavior features and linguistic features in spam review detection using a GBM framework has shown promising results which increased accuracy. By applying feature scaling techniques such as standardization to the spammer behavior features, the features are normalized and brought to a similar scale, which enhanced the modeling process. The GBM model, with its ability to capture complex interactions and learn relationships, effectively leverages the combined features for accurate spam detection. The model automatically learns feature interactions during training, allowing it to capture intricate relationships and make predictions based on both types of features. Feeding the separately processed X_SB and X_L as inputs to the GBM model does not involve any additional computations that scale with the size of the dataset. The purposed Model **SRD-FGBM**, in fusion with spammer behavior and linguistic features, offers a promising approach for effective spam review detection leading to improved performance or a more comprehensive understanding of the data. The integration of the GBM model with the combined spammer behavior and linguistic features presents a promising solution for effective spam review detection, resulting in enhanced performance and a deeper insight into the underlying data.

6. FUTURE WORK AND LIMITATIONS

Future work includes exploring ensemble techniques and integrating deep learning approaches (such as RNNs or CNNs) to enhance spam review detection. Limitations include the need to validate model generalizability, address data imbalance, optimize feature selection, and ensure scalability for large-scale datasets.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

References

1) R. Amos, R. Maio, and P. Mittal, “Reviews in motion: a large scale, longitudinal study of review recommendations on Yelp,” Feb. 2022, [Online]. Available: http://arxiv.org/abs/2202.09005

2) H. Paul and A. Nikolaev, “Fake review detection on online E-commerce platforms: a systematic literature review,” *Data Min Knowl Discov*, vol. 35, no. 5, pp. 1830–1881, Sep. 2021, doi: 10.1007/s10618-021-00772-6.

3) G. Wang, S. Xie, B. Liu, and P. S. Yu, “Review graph based online store review spammer detection,” *Proceedings - IEEE International Conference on Data Mining, ICDM*, no. December, pp. 1242–1247, 2011, doi: 10.1109/ICDM.2011.124.

4) N. Hussain, H. Turab Mirza, I. Hussain, F. Iqbal, and I. Memon, “Spam Review Detection Using the Linguistic and Spammer Behavioral Methods,” *IEEE Access*, vol. 8, pp. 53801–53816, 2020, doi: 10.1109/ACCESS.2020.2979226.

5) C. Xu, J. Zhang, K. Chang, and C. Long, “Uncovering collusive spammers in Chinese review websites,” pp. 979–988, 2013, doi: 10.1145/2505515.2505700.

6) Y. Ren and D. Ji, “Learning to Detect Deceptive Opinion Spam: A Survey,” *IEEE Access*, vol. 7, no. c, pp. 42934–42945, 2019, doi: 10.1109/ACCESS.2019.2908495.

7) N. Jindal and B. Liu, “Analyzing and detecting review spam,” *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 547–552, 2007, doi: 10.1109/ICDM.2007.68.

8) K. Ravi and V. Ravi, *A survey on opinion mining and sentiment analysis: Tasks, approaches and applications*, vol. 89, no. June 2015. 2015. doi: 10.1016/j.knosys.2015.06.015.

9) F. Hemmatian and M. K. Sohrabi, “A survey on classification techniques for opinion mining and sentiment analysis,” *Artif Intell Rev*, pp. 1–51, 2017, doi: 10.1007/s10462-017-9599-6.

10) M. Ben Khalifa, Z. Elouedi, and E. Lefevre, “An evidential spammer detection based on the suspicious behaviors’ indicators,” *Proceedings of 2020 International Multi-Conference on: Organization of Knowledge and Advanced Technologies, OCTA 2020*, 2020, doi: 10.1109/OCTA49274.2020.9151805.

11) A. A, “Review on Effective Email Classification for Spam and Non Spam Detection on Various Machine Learning Techniques,” *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 3, no. 3, pp. 1621–1624, 2015, doi: 10.17762/ijritcc2321-8169.1503158.

12) Z. Gyongyi and H. Garcia-Molina, “Web spam taxonomy,” *Proceedings of the 1st International Workshop on Adversarial Information Retrieval on the Web, AIRWeb 2005 - Held in Conjunction with the 14th International World Wide Web Conference*, pp. 39–47, 2005.

13) D. Zhang, W. Li, B. Niu, C. W.-D. S. Systems, and undefined 2023, “A deep learning approach for detecting fake reviewers: Exploiting reviewing behavior and textual information,” *Elsevier*, Accessed: Mar. 15, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167923622001828

14) J. Pouramini, B. Minaei-Bidgoli, and M. Esmaeili, “A novel feature selection method in the categorization of imbalanced textual data,” *KSII Transactions on Internet and Information Systems*, vol. 12, no. 8, pp. 3725–3748, 2018, doi: 10.3837/tiis.2018.08.010.

15) A. S. Kale, V. Pandya, F. Di Troia, and M. Stamp, “Malware classification with Word2Vec, HMM2Vec, BERT, and ELMo,” *Journal of Computer Virology and Hacking Techniques*, vol. 19, no. 1, pp. 1–16, Mar. 2023, doi: 10.1007/S11416-022-00424-3.

16) V. Geetha, C. G.-J. Of P. Negative, and undefined 2023, “A Study on Spam Review Detection Using Linguistics,” *pnrjournal.com*, Accessed: Mar. 15, 2023. [Online]. Available:

ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

https://pnrjournal.com/index.php/home/article/view/6130

17) A. Mukherjee *et al.*, “Spotting opinion spammers using behavioral footprints,” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. Part F1288, pp. 632–640, 2013, doi: 10.1145/2487575.2487580.

18) A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, “Detection of review spam: A survey,” *Expert Syst Appl*, vol. 42, no. 7, pp. 3634–3642, 2015, doi: 10.1016/j.eswa.2014.12.029.

19) K. C. Santosh and A. Mukherjee, “On the temporal dynamics of opinion spamming: Case studies on yelp,” in *25th International World Wide Web Conference, WWW 2016*, International World Wide Web Conferences Steering Committee, 2016, pp. 369–379. doi: 10.1145/2872427.2883087.

20) H. Li, G. Fei, S. Wang, B. Liu … W. S.-P. Of the 26th, and undefined 2017, “Bimodal distribution and co-bursting in review spam detection,” *dl.acm.org*, Accessed: Mar. 28, 2019. [Online]. Available: https://dl.acm.org/citation.cfm?id=3052582

21) “Fake Review Detection via exploitation of Spam Indicators and Author Behavior Characteristics,” no. February, pp. 1–92, 2017.

22) R. Y. K. Lau, S. Y. Liao, R. Chi-Wai Kwok, K. Xu, Y. Xia, and Y. Li, “Text mining and probabilistic language modeling for online review spam detection,” *ACM Trans Manag Inf Syst*, vol. 2, no. 4, pp. 1–30, 2011, doi: 10.1145/2070710.2070716.

23) R. Y. K. Lau, S. Y. Liao, R. Chi-Wai Kwok, K. Xu, Y. Xia, and Y. Li, “Text mining and probabilistic language modeling for online review spam detection,” *ACM Transactions on Management Information Systems*, vol. 2, no. 4. Dec. 2011. doi: 10.1145/2070710.2070716.

24) F. Li, M. Huang, Y. Yang, and X. Zhu, “Learning to identify review spam,” *IJCAI International Joint Conference on Artificial Intelligence*, pp. 2488–2493, 2011, doi: 10.5591/978-1-57735-516-8/IJCAI11-414.

25) D. H. Fusilier, M. Montes-Y-Gómez, P. Rosso, and R. G. Cabrera, “Detection of Opinion Spam with Character n-grams,” pp. 285–294, 2015, doi: 10.1007/978-3-319-18117.

26) M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, “Finding Deceptive Opinion Spam by Any Stretch of the Imagination,” pp. 309–319, 2011, doi: 10.1145/2567948.2577293.

27) M. Hazim, N. B. Anuar, M. F. Ab Razak, and N. A. Abdullah, “Detecting opinion spams through supervised boosting approach,” *PLoS One*, vol. 13, no. 6, Jun. 2018, doi: 10.1371/journal.pone.0198884.

28) N. Kumar, D. Venugopal, L. Qiu, and S. Kumar, “Detecting Review Manipulation on Online Platforms with Hierarchical Supervised Learning,” *Journal of Management Information Systems*, vol. 35, no. 1, pp. 350–380, 2018, doi: 10.1080/07421222.2018.1440758.

29) D. Zhang, L. Zhou, J. L. Kehoe, and I. Y. Kilic, “What Online Reviewer Behaviors Really Matter? Effects of Verbal and Nonverbal Behaviors on Detection of Fake Online Reviews,” *Journal of Management Information Systems*, vol. 33, no. 2, pp. 456–481, 2016, doi: 10.1080/07421222.2016.1205907.

30) S. Sakib, N. Ahmed, A. J. Kabir, and H. Ahmed, “An Overview of Convolutional Neural Network : Its Architecture and Applications,” no. February, 2019, doi: 10.20944/preprints201811.0546.v4.

31) Y. Lin, T. Zhu, H. Wu, J. Zhang, X. Wang, and A. Zhou, “Towards online anti-opinion spam: Spotting fake reviews from the review sequence,” *ASONAM 2014 - Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 261–264, 2014, doi: 10.1109/ASONAM.2014.6921594.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

32) W. Zhang, Q. Wang, X. Li, T. Yoshida, and J. Li, “DCWord: A Novel Deep Learning Approach to Deceptive Review Identification by Word Vectors,” *J Syst Sci Syst Eng*, vol. 28, no. 6, pp. 731–746, 2019, doi: 10.1007/s11518-019-5438-4.

33) N. Kumar, D. Venugopal, L. Qiu, and S. Kumar, “Detecting Review Manipulation on Online Platforms with Hierarchical Supervised Learning,” *Journal of Management Information Systems*, vol. 35, no. 1, pp. 350–380, 2018, doi: 10.1080/07421222.2018.1440758.

34) D. Radovanović and B. Krstajić, “Review spam detection using machine learning,” in *2018 23rd International Scientific-Professional Conference on Information Technology, IT 2018*, 2018, pp. 1–4. doi: 10.1109/SPIT.2018.8350457.

35) E. Choo, T. Yu, and M. Chi, “Detecting opinion spammer groups through community discovery and sentiment analysis,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9149, pp. 170–187, 2015, doi: 10.1007/978-3-319-20810-7_11.

36) S. C. Eshan and M. S. Hasan, “An application of machine learning to detect abusive Bengali text,” *20th International Conference of Computer and Information Technology, ICCIT 2017*, vol. 2018-Janua, pp. 1–6, 2018, doi: 10.1109/ICCITECHN.2017.8281787.

37) P. V Arivoli and T. Chakravarthy, “Document Classification Using Machine Learning Algorithms - A Review,” vol. 5, no. 2, pp. 48–54, 2017.

38) Q. Chen and M. Sokolova, “Word2Vec and Doc2Vec in Unsupervised Sentiment Analysis of Clinical Discharge Summaries”.

39) G. Song, X. Huang, G. Cao, Z. Tao, W. Liu, and L. Yang, “Better Word Representations with Word Weight,” *IEEE 21st International Workshop on Multimedia Signal Processing, MMSP 2019*, 2019, doi: 10.1109/MMSP.2019.8901756.

40) S. Qaiser and R. Ali, “Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents,” *Int J Comput Appl*, vol. 181, no. 1, pp. 25–29, 2018, doi: 10.5120/ijca2018917395.

41) G. Pant and P. Srinivasan, “Learning to crawl: Comparing classification schemes,” *ACM Trans Inf Syst*, vol. 23, no. 4, pp. 430–462, 2005, doi: 10.1145/1095872.1095875.

42) M. Yang, W. Zhao, L. Chen, Q. Qu, Z. Zhao, and Y. Shen, “Investigating the transferring capability of capsule networks for text classification,” *Neural Networks*, vol. 118, pp. 247–261, 2019, doi: 10.1016/j.neunet.2019.06.014.

43) D. Radovanović and B. Krstajić, “Review spam detection using machine learning,” *2018 23rd International Scientific-Professional Conference on Information Technology, IT 2018*, vol. 2018-Janua, pp. 1–4, 2018, doi: 10.1109/SPIT.2018.8350457.

44) S. Krishnamoorthy, “Linguistic features for review helpfulness prediction,” *Expert Syst Appl*, vol. 42, no. 7, pp. 3751–3759, 2015, doi: 10.1016/j.eswa.2014.12.044.

45) F. Rustam, A. Mehmood, M. Ahmad, S. Ullah, D. M. Khan, and G. S. Choi, “Classification of Shopify App User Reviews Using Novel Multi Text Features,” *IEEE Access*, vol. 8, pp. 30234–30244, 2020, doi: 10.1109/ACCESS.2020.2972632.

46) Y. Ren, Y. Zhang, M. Zhang, and D. Ji, “Improving twitter sentiment classification using topic-enriched multi-prototype word embeddings,” *30th AAAI Conference on Artificial Intelligence, AAAI 2016*, pp. 3038–3044, 2016.

47) M. Ben Khalifa, Z. Elouedi, and E. Lefevre, “An evidential spammer detection based on the suspicious behaviors’ indicators,” *Proceedings of 2020 International Multi-Conference on: Organization of Knowledge and Advanced Technologies, OCTA 2020*, 2020,

ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 07:2023
DOI10.17605/OSF.IO/CVM6N

doi: 10.1109/OCTA49274.2020.9151805.

48) M. Zhong, Z. Li, S. Liu, B. Yang, R. Tan, and X. Qu, “Fast Detection of Deceptive Reviews by Combining the Time Series and Machine Learning,” *Complexity*, vol. 2021, pp. 2–6, 2021, doi: 10.1155/2021/9923374.

49) X. Tang, T. Qian, and Z. You, “Generating behavior features for cold-start spam review detection with adversarial learning,” *Inf Sci (N Y)*, vol. 526, pp. 274–288, 2020, doi: 10.1016/j.ins.2020.03.063.

50) M. Z. Asghar, A. Ullah, S. Ahmad, and A. Khan, “Opinion spam detection framework using hybrid classification scheme,” *Soft comput*, vol. 24, no. 5, pp. 3475–3498, 2020, doi: 10.1007/s00500-019-04107-y.

51) N. Hussain, H. Turab Mirza, I. Hussain, F. Iqbal, and I. Memon, “Spam Review Detection Using the Linguistic and Spammer Behavioral Methods,” *IEEE Access*, vol. 8, pp. 53801–53816, 2020, doi: 10.1109/ACCESS.2020.2979226.

52) Y. Dang, Y. Zhang, and H. Chen, “A lexicon-enhanced method for sentiment classification: An experiment on online product reviews,” *IEEE Intell Syst*, vol. 25, no. 4, pp. 46–53, 2010, doi: 10.1109/MIS.2009.105.

53) R. Moraes, J. F. Valiati, and W. P. Gavião Neto, “Document-level sentiment classification: An empirical comparison between SVM and ANN,” *Expert Syst Appl*, vol. 40, no. 2, pp. 621–633, 2013, doi: 10.1016/j.eswa.2012.07.059.

54) G. Vinodhini and R. M. Chandrasekaran, “Opinion mining using principal component analysis based ensemble model for e-commerce application,” *CSI Transactions on ICT*, vol. 2, no. 3, pp. 169–179, 2014, doi: 10.1007/s40012-014-0055-3.