

Research Article

Fake Review Identification Methods Based on Multidimensional Feature Engineering

Ge Wang,¹ Guanglei Shang,² Pengbo Pu ,¹ Xuxin Li,¹ and Hao Peng¹

¹College of Intelligent Equipment, Shandong University of Science and Technology, Tai'an, China

²College of Finance and Economics, Shandong University of Science and Technology, Tai'an, China

Correspondence should be addressed to Pengbo Pu; ppb@sdust.edu.cn

Received 1 July 2022; Accepted 26 July 2022; Published 25 August 2022

Academic Editor: Chia-Huei Wu

Copyright © 2022 Ge Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Product reviews in electronic platforms are very valuable to potential customers, product manufacturers, and product sellers. Their data contain huge business opportunities. Therefore, this paper analyzes the views, attitudes, and emotions expressed in these reviews. It presents three fake review identification methods based on multidimensional feature engineering. Under the premise of adding product feature extraction and opinion sentence judgment, six feature parameters are defined to identify fake reviews, and a fake review identification model based on multidimensional feature engineering is constructed. Then, the effectiveness of the selected feature engineering is verified. Based on the multidimensional feature engineering model, a fake review identification algorithm based on multidimensional feature engineering of union relationship, an identification algorithm based on weighted multidimensional feature engineering scoring, and an identification algorithm based on weighted multidimensional feature engineering classification are proposed. The execution effects of the three methods are compared. Fake review identification models based on multidimensional feature engineering can effectively filter fake reviews.

1. Introduction

More and more products have begun to sell online, and more and more people feel the convenience and speed from the e-commerce platforms. It has become a habit for potential purchasers to read product reviews before buying. Product reviews have played a crucial role in whether a purchaser is going to buy this product. If the product's review is praise, then consumers can buy this product. But in turn, if the reviews of the product are mostly bad, consumers will turn to other products. At the same time, a good review means huge business benefits for a company. Therefore, more and more businesses and companies noticed that product reviews have great influence on their operations and sales [1, 2].

How to identify fake reviews is a very important job. The identification of fake reviews is still different from the identification of spam. One of the reasons is that the publisher of fake reviews is easy to disguise themselves. Users are difficult to identify, which is not like spam where

users can easily distinguish which messages are spam. Another reason is that it is difficult to establish a recognition model for the identification of fake reviews. It is difficult to manually mark which reviews are fake and which are real. People who post these fake reviews tend to write content very much like real reviews. Overall, the identification of fake reviews can be regarded as a binary classification [3–5]; that is, they can be divided into two categories, true or false.

1.1. Related Work. In order to obtain more business benefits from the reviews, especially praise, many businesses began to exploit some loopholes or mechanisms in the platforms to write some fake reviews to cheat consumers or fight competitors [6–8]. Reviews with this deception or fraudulent behavior are spam reviews or fake reviews [9]. The problem of spam review identification is first proposed in Prof. Bing Liu's research [10] in 2007. Since then, the identification of spam reviews has become a hot problem in the field of opinion identification [11].

The identification of fake reviews is difficult. One of the reasons is the lack of reliable tagged real and fake reviews. Jindal and Liu [12] used repetitive reviews in the experiments. In their research, they studied 5.8 million reviews and 2.14 million publishers in Amazon and found that there was a large amount of repetitive and nearly repetitive data in these reviews. This showed that there were a lot of fake reviews. It was likely because publishers had never bought these products or used related services, and it was difficult for them to write innovative new reviews. So they copied others' reviews, used the same reviews for different products, or made a little modification to some reviews. They divided these copies into four categories:

- (1) Repeated reviews of the same product from the same customers.
- (2) Repeated reviews of the same product from different customers.
- (3) Repeated reviews of different products from the same customers.
- (4) Repeated reviews of different products from different customers.

They think that the first type may be due to error operation from duplicate submissions, which could be judged by viewing the date. However, the remaining three types of repeated reviews were fake reviews. They used them as fake reviews to train their algorithm model.

Lim and Nguyen [13] identified the authenticity of reviews by classifying the publisher abnormal behavior. Jindal and Liu [14] proposed a method based on unexpected rules to find fake reviews. Ott and Choi [15] proposed a supervised learning method based on standard words and POS multivariate features. Li et al. [16] also used a supervised approach to identify fake reviews by defining some additional features. Wang et al. [17] used a graph-based method to identify fake store reviewers. Wu and Greene [18] proposed distortion-based criteria to assess the degree of impact of the Tripadvisor suspect hotels. They believed that fake reviews could distort the overall popularity of the hotel collection. Mukherjee et al. [19] proposed a method of initial group spam testing. Bhuvaneshwari et al. [20] proposed a deep learning (DL) based novel framework to learn document level representation for identifying the spam reviews. Ashraf et al. [21] proposed an approach based on unsupervised learning via self-organizing maps (SOM) in conjunction with convolutional neural networks (CNN) to perform classification of the reviews. Hajek and Sahut [22] proposed a fake review identification model based on behavioral and sentiment linguistic features. They combined content analysis methods with the reader's behavior to judge the authenticity of the reviews.

The meaning of fake reviews contains more content and broad concepts. In the work of Jindal and Liu [12], they divided fake reviews into the following three types:

- (1) Untruthful opinions were some positive reviews that intentionally misled consumers or opinion identification system, or maliciously negative or unfair reviews of other products or objects.

- (2) Reviews on brands were those commenting on the brands, manufacturers, or sellers without products. Here, this type of reviews was also a fake review, because their objects described in the reviews were not the products.
- (3) Non-reviews contained two types. One was advertisement. The other was irrelevant reviews. These reviews had no opinion.

In our work, we define fake reviews, namely, “non-opinion reviews” and “untruthful reviews.”

Σ' is the collection of words for review α . Σ_f is a word set of product feature words in similar product reviews. Σ_o is a sentiment word set in similar product reviews. Then, α is a non-opinion review meeting one of the following three conditions:

- (1) $\Sigma' \cap \Sigma_f = \emptyset$, or $\Sigma' \cap \Sigma_o = \emptyset$.
- (2) $\Sigma' \cap \Sigma_f \neq \emptyset$, but $\Sigma' \cap \Sigma_o = \emptyset$.
- (3) $\Sigma' \cap \Sigma_f \neq \emptyset$, $\Sigma' \cap \Sigma_o \neq \emptyset$, but there is no $w \in \Sigma_o$ that is related to $f \in \Sigma_f$.

Untruthful reviews mentioned in this paper are untruthful opinions as defined by Jindal and Liu [12].

One feature of the untruthful reviews is very strong concealment. The content of such reviews is very similar to the real reviews. The other feature is being very purposeful and targeted, with a strong utilitarian purpose behind the reviews. So the identification of untruthful reviews is more difficult than that of non-opinion reviews.

1.2. Feature Engineering Setting. Feature engineering is a very critical task in the identification of fake reviews. It analyzes datasets, especially fake reviews, to obtain important parametric indicators to identify fake reviews. Here, we use the manually annotated dataset, and the rules of the Jingdong website evaluation to summarize the feature parameters of the identification model. Thus, the six feature items in the feature engineering are defined. They are the length of reviews, the correlation of the review content and the product, the similarity of the reviews, the word repetition rate, the number of reviewer comments per day, and the existence of opinion sentences. The six feature items are shown in Table 1.

1.2.1. The Length of Reviews (length). E-commerce platforms have certain restrictions on the number of review words. For example, the length limit of reviews on Taobao is 500 words, and the book reviews on Dangdang are no more than 2,000 words. But the buyers often think this is troublesome, and many are reluctant to comment, or use default comments. Therefore, many e-commerce platforms have introduced a lot of encouraging principles on the number of review words. Many sellers also encourage buyers to comment with more than 10 words with favorable cash back.

The experimental data consists of 4088 reviews obtained from Taobao. The crawled datasets are preprocessed and

TABLE 1: The relationship of feature and review category.

Classification	Feature item	Description
Non-opinion reviews	length	The length of reviews
	relativity	The relativity of review content and products
	repetition_ratio	The repetition ratio of words
	WIOS	Whether the reviews include the opinion sentences
Untruthful reviews	length	The length of reviews
	similarity	Similarity of the review content
	MNR	Customers posting the maximum number of reviews in a day

then stored in a formatted form. The data retain the basic features to be used for fake information identification.

In the dataset, we count the number of reviews and obtain the distribution of word count of reviews of some mobile power products in Taobao. The distribution results are shown in Table 2.

1.2.2. The Relativity of Review Content and Products (relativity). Jingdong's comment rules show that a condition for whether you can obtain Jingdong beans is whether the review content is related to the products. The definition of non-opinion reviews also points out that one method for judging whether a review is non-opinion is the relativity of the review content and products, that is, the usefulness of the review content.

Some non-opinion reviews have many words, but the usefulness of content is low. There is not any help to potential buyers. The review content cannot show the information of the products or the relevant services provided by the merchants.

Reviews that include more product attributes are reviews with higher relationships. Feature relativity calculation is necessary to avoid reviews copied from other products or other users or simply to meet the number of words.

The relativity of the review content and products is calculated as follows:

- (1) Extract product features in reviews.
- (2) Convert product reviews into word vectors based on product features.

TABLE 2: The distribution of review words.

Classification	Review number	Proportion
Word number ≤ 10	875	0.214
$10 < \text{word number} \leq 20$	808	0.198
$20 < \text{word number} \leq 30$	962	0.235
$30 < \text{word number} \leq 40$	673	0.165
$40 < \text{word number} \leq 50$	316	0.077
Word number > 50	454	0.111

- (3) Select the correlation product features by χ^2 statistical method.

In this work, χ^2 statistical method is used to select the correlation product features.

Here, a 2×2 matrix is used to show the relationship between feature f and category c , as is shown in Table 3.

In Table 3, $N(f, c)$ is the number of reviews where the feature f appears in category c ; $N(f, \bar{c})$ is the number of reviews where the feature f appears in category \bar{c} ; $N(\bar{f}, c)$ is the number of reviews where the feature f does not appear in category c ; $N(\bar{f}, \bar{c})$ is the number of reviews including feature f in category \bar{c} . N is the total number of reviews in the corpus. It is referred to as the following equation:

$$N = N(f, c) + N(f, \bar{c}) + N(\bar{f}, c) + N(\bar{f}, \bar{c}). \quad (1)$$

Based on the above definition, define χ^2 function as the following equation:

$$\chi^2(f, c) = \frac{N(N(f, c)N(\bar{f}, \bar{c}) - N(\bar{f}, c)N(f, \bar{c}))^2}{(N(f, c) + N(\bar{f}, c))(N(f, \bar{c}) + N(\bar{f}, \bar{c}))(N(f, c) + N(f, \bar{c}))(N(\bar{f}, c) + N(\bar{f}, \bar{c}))}. \quad (2)$$

The feature value is represented by the weighted average of feature values as follows:

$$\chi^2(f) = \sum_{i=1}^m P(c_i) \chi^2(f, c_i). \quad (3)$$

The feature values are represented by 0 and 1. If the feature appears in the review, its value is 1. In contrast, if the feature does not appear in the review, its value is 0.

1.2.3. Similarity of the Review Content (similarity). Jindal and Liu [12] pointed out in their article that fake reviews are largely repetitive reviews because the publishers of fake reviews had not purchased or used the reviewed products. It was difficult for them to describe some of the features of the products or point out the problems in the products. The easiest and fastest way was to copy the reviews of others as their own reviews. So the similarity of reviews is an important indicator to judge whether the review is fake. This indicator is used to distinguish untruthful reviews in the definition.

TABLE 3: The relationship table of feature f and class c .

Feature	c	\bar{c}
f	$N(f, c)$	$N(f, \bar{c})$
\bar{f}	$N(\bar{f}, c)$	$N(\bar{f}, \bar{c})$

The length of reviews in e-commerce websites is relatively short compared to news, blogs, and other comments. The average length of 4,088 reviews of electronic products collected from Taobao is 29 words, while the length of news and blogs is generally more than 500 words. Although the length of reviews is short, the customers' expression is relatively free, and the grammar structure is not as standardized and rigorous as in the news. In order to improve the efficiency and accuracy, some keywords are drawn from the product reviews to form our vector list. The vector space algorithm is used to calculate the similarity of product reviews. To reduce the dimension, the extracted keywords are the feature and opinion words about the products.

The similarity calculation method of the product reviews is as follows:

- (1) Extract product features in reviews.
- (2) Extract the opinion words in reviews.
- (3) Convert product reviews to word vectors based on product feature and opinion words. A review can be expressed as a vector, $R = (r_1, r_2, \dots, r_n)$, where r_i is the keyword in the review.
- (4) Calculate the similarity of the product reviews. The calculation formula is as follows:

$$\begin{aligned} \text{Similarity}(R_j, C_k) &= \frac{\vec{R}_j \cdot \vec{C}_k}{|\vec{R}_j| |\vec{C}_k|} \\ &= \frac{\sum_{i=1}^n w_{i,j} w_{i,k}}{\sqrt{\sum_{i=1}^n w_{i,j}^2} \sqrt{\sum_{i=1}^n w_{i,k}^2}}, \end{aligned} \quad (4)$$

where R_j is the sample review, C_k is the review that will be calculated, and $w_{i,j}$ is the weight of each feature item.

To show the differences in the customer product reviews that can be represented by the various feature items, we use TF-IDF to calculate the weight of each feature item. The differences of reviews are shown by the frequency of each feature item in the reviews. The weight of feature item is as follows:

$$w_{i,k} = \text{freq}_{i,k} \log\left(\frac{N}{n_i} + 0.01\right), \quad (5)$$

where $w_{i,k}$ is the feature item, $\text{freq}_{i,k}$ is the number of the keyword r_i occurrences in the review R_k , N is the number of the reviews in the corpus, n_i is the number of reviews containing keyword r_i .

Using the vector space model algorithm in the calculation of the similarity of the review content can simplify the problem and reduce the complexity. The vector space model fully considers the difference of the reviews. The vector space model assumes that the keywords are independent of each other.

1.2.4. The Repetition Ratio of Words (*repetition_ratio*).

Through the analysis of non-opinion reviews and untruthful reviews, we find that there are many repeated words in these reviews. Some customers want to meet the specified number of words and earn points in this way. But most real reviews are logical, with continuous expression and high consistency. Due to the short length of product reviews, most customers rarely have a word appearing repeatedly in their real reviews. Considering this, the repetition ratio of words is also used in judging whether a review is fake.

Define the equation of repetition ratio of words as follows:

$$\text{repetition_ratio} = 1 - \frac{n_{\text{unique}}}{N}, \quad (6)$$

where N is the number of words appearing in the reviews and n_{unique} is the number of words that appear only once in the reviews.

1.2.5. Customers Posting the Maximum Number of Reviews in a Day (*MNR*).

Through the research, we find that a customer does not post many reviews in one day. Many fake reviews from e-commerce websites are posted by hiring teams, commonly known as the Water Army. This is also mentioned in the research work of Mukherjee et al. [23]. They thought that a customer posting a lot of reviews in a day was an abnormal behavior. Their data analysis revealed that 25% of fake review publishers posted five reviews in a day and 75% posted more than six reviews per day. Among the real review publishers, 50% posted one review per day, and 90% had no more than 3 reviews per day.

Therefore, in our work, the number of reviews posted by a customer in a day is used to identify the fake reviews according to the analysis of Mukherjee et al. We set the threshold of this feature item to 3; that is, a customer cannot post more than 3 reviews per day.

1.2.6. Whether the Review Includes Opinion Sentences (*WIOS*).

Based on the definition of the non-opinion reviews, the number of opinion sentences in a review is also a parameter to judge whether a review is a non-opinion review. If at least one product feature word and a sentiment word are included in a sentence α . α is an opinion sentence. If a review contains at least one opinion sentence, the value of the feature item *WIOS* is set to 1; otherwise, it is set to 0.

1.3. Fake Review Identification Algorithm Based on Multidimensional Feature Engineering.

To make a comprehensive analysis of the effects of feature engineering, we used three algorithms based on feature engineering. They are the fake review identification algorithm based on multidimensional feature engineering of union relationship, the identification algorithm based on weighted multidimensional feature engineering scoring, and the identification algorithm based on weighted multidimensional feature engineering classification.

1.3.1. Fake Review Identification Algorithm Based on Multidimensional Feature Engineering of Union Relationship. In the experiment of Bhattarai et al. [24], they used union relationship in the feature engineering to identify whether the blog content is fake. We also use union relationship and our definition of six feature items in the experiment. So a review R can be written by the following equation:

$$R = F_1 \times F_2 \times \dots \times F_n, \quad (7)$$

where F_1, \dots, F_n ($n = 6$) are the six feature items.

The reviews are classified into two categories, that is, non-opinion reviews and untruthful reviews, by the experiment. Each feature item has a set threshold. Therefore, the condition where a review is defined as a fake review is to satisfy the following equation:

$$\text{fake_review} = f(F_1) \vee f(F_2) \vee \dots \vee f(F_n). \quad (8)$$

If at least one feature item is not within the threshold range we define, the review is considered as fake.

The fake review identification algorithm based on multidimensional feature engineering of union relationship is shown in Algorithm 1.

1.3.2. Fake Review Identification Algorithm Based on Weighted Multidimensional Feature Engineering. In this work, fake reviews are identified by a suitable weight value used in feature engineering. The entropy method is used to define the weight values of the feature items.

In information theory, entropy describes a kind of uncertainty. It is a measure of uncertainty. The greater the information, the less the uncertainty, that is, the less the entropy.

According to this characteristic of entropy, the degree of dispersion of each feature item can be judged by entropy values. The larger the dispersion of the feature item, the greater the impact of this feature item on comprehensive evaluation and the smaller the entropy value.

Now, a dataset $R = \{R_1, R_2, \dots, R_n\}$ includes n reviews and m ($m = 6$) feature items, $F = \{F_1, F_2, \dots, F_m\}$, where F is the feature vector of R . Define the value of the feature items in each review; that is, $X = (x_{ij})_{n \times m}$ where x_{ij} is the j th feature item of R_i , that is, the value of F_j .

To determine the value of the weight of each feature item, normalizing the original data is the first step. The original data obtained in the experiment are the data of different types of units. Therefore, before using them for calculation, all feature values x_{ij} are normalized. The method is as follows:

$$z_{ij} = \frac{x_{ij} - \min\{x_{1j}, \dots, x_{nj}\} + 0.1}{\max\{x_{1j}, \dots, x_{nj}\} - \min\{x_{1j}, \dots, x_{nj}\} + 0.1}, \quad (9)$$

where z_{ij} is the value after normalization ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$). Thereby, the original dataset $X = (x_{ij})_{n \times m}$ is converted to $Z = (z_{ij})_{n \times m}$. Then, calculate the proportion of each feature item. p_{ij} is the

proportion of the j th feature item. The calculation method is as follows:

$$p_{ij} = \frac{z_{ij}}{\sum_{i=1}^n z_{ij}} \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, m). \quad (10)$$

Thus, the entropy e_j of the j th feature term is calculated as the following equation:

$$e_j = -k \sum_{i=1}^n p_{ij} \ln(p_{ij}), \quad (11)$$

where k is as follows:

$$k = \frac{1}{\ln(n)}. \quad (12)$$

Next, the redundant degree d_j of information entropy is calculated as the following equation:

$$d_j = 1 - e_j \quad (j = 1, 2, \dots, m). \quad (13)$$

Then, get the weight w_j of each feature item indicator F_j ($j = 1, 2, \dots, m$). The weight calculation formula is as follows:

$$w_j = \frac{d_j}{\sum_{j=1}^m d_j} \quad (j = 1, 2, \dots, m). \quad (14)$$

By weighting the feature terms, the comprehensive score can be calculated for each review. The comprehensive score formula for each review is as follows:

$$\text{score}_i = \sum_{j=1}^m w_j \cdot z_{ij}. \quad (15)$$

There are two algorithms using weighted multidimensional feature engineering. One is the identification algorithm based on weighted multidimensional feature engineering scoring, which is Algorithm 2. The other is the identification algorithm based on weighted multidimensional feature engineering classification, which is Algorithm 3.

2. Experiment

A very important issue in the identification of fake reviews is the acquisition of reliable data. Since fake reviews do not have obvious distinguishing characteristics like spam, the concealment of fake reviews is very high. Many of the fake reviews are published by the customers who have bought the products before. The sellers pay for the customers and ask them to help publish the fake reviews. The purchase process is also strictly followed. So the identification of these fake reviews is very difficult.

2.1. Data Preprocessing. Due to the lack of features for the experiment in the open dataset collected online, we did not use the existing dataset. Instead, we collected 4,088 electronic product reviews from e-commerce platform page through crawler tools and marked the fake reviews by manual annotation.

```

Input: Review corpus  $R\_list$ , feature sequences in feature engineering  $FE\_list$ , feature item threshold  $thre\_list$ 
Output: Review realistic label sequence  $Fake\_list$ 
Procedure FakeReviews_union( $R\_list, FE\_list, thre\_list$ )
begin
  for each  $R_i$  ( $R_i \in R\_list$ )
  begin
    for each  $FE_j$  ( $FE_j \in FE\_list$ )
    begin
      calculate  $f(FE_j)$ ;
      if  $f(FE_j)$  is beyond the defined threshold value  $t_j, t_j \in thre\_list$ ;
       $f(FE_j) = true$ ;
    elseif;
       $f(FE_j) = false$ ;
    endif;
  endfor;
   $fg_i = f(F_1) \vee f(F_2) \vee f(F_3) \vee f(F_4) \vee f(F_5) \vee f(F_6)$ ;
  add  $fg_i$  in  $Fake\_list$ ;
endfor;
return  $Fake\_list$ ;
end;

```

ALGORITHM 1: FakeReviews_union.

```

Input: Dataset  $R\_list$ , feature sequence in feature engineering  $F\_list$ 
Output: Score sequence for reviews  $score\_list$ 
Procedure multiFeatureScore( $R\_list, F\_list$ )
begin
   $R\_F \leftarrow F\_list$ ;
  for each  $R_i$  in  $R\_list$ 
  begin
    calculate the value of the feature vector in the feature engineering,  $R_i-F_j$ ;
    add  $R_i-F_j$  to  $R\_F$ ;
  endfor;
  for each  $R_i$  in  $R\_list$ 
  begin
    for each  $F_j$  in  $R\_F$ 
    begin
      calculate the weight value  $w_j$  of  $F_j$ ;
    endfor;
    calculate the score value  $score_i$  of  $R_i$ ;
    add  $score_i$  to  $score\_list$ ;
  endfor;
  return  $score\_list$ ;
end;

```

ALGORITHM 2: The identification algorithm based on weighted multidimensional feature engineering scoring (multiFeatureScore).

The principle of labeling is based on the assumption that the higher the usefulness of the review, the higher the authenticity of this review. The assumption of “usefulness” is reflected in the following four aspects: (1) whether the product name appears in the review; (2) whether there is a relevant description of the attributes of the product in the review; (3) whether the author’s view on the purchased product is expressed in the review; (4) whether a picture is attached. If three or all are met, the review is judged as of high usefulness. If only two of them are met, the usefulness is determined to be moderate. If only one or none of the four

conditions is met in the review, the usefulness of the review is low. To improve the accuracy of the experiment, all empty reviews are deleted.

2.2. The Annotation of the Dataset. The dataset used in the experiment was manually annotated. Most voting strategies were used to reduce individual bias. Labels numbered over half were selected as the final labels for reviews.

In order to improve the rigor of manual annotation, the dataset was classified at the usefulness level during the annotation process. This work is to facilitate better

```

Input: Dataset  $R\_list$ , feature sequence in feature engineering  $F\_list$ 
Output: Classification results  $classify\_list$ 
Procedure multiFeatureScore( $R\_list, F\_list$ )
begin
   $R\_F \leftarrow F\_list$ ;
  for each  $R_i$  in  $R\_list$ 
  begin
    calculate the value of the feature vector in the feature engineering,  $R_i F_j$ ;
    add  $R_i F_j$  to  $R\_F$ ;
  endfor;
  for each  $R_i$  in  $R\_list$ 
  begin
    for each  $F_j$  in  $R\_F$ 
    begin
      calculate the weight value  $w_j$  of  $F_j$ ;
      add  $w_j$  to  $w\_list$ ;
    endfor;
  endfor;
  classify  $R\_list$  by using SVM algorithm based on  $F\_list$  and  $w\_list$  and get the classification result  $classify\_list$ ;
  return  $classify\_list$ ;
end;

```

ALGORITHM 3: The identification algorithm based on weighted multidimensional feature engineering classification (multiFeatureClassify).

discrimination of fake reviews. The dataset is divided into high, middle, and low categories according to review usefulness. Then, the dataset is labeled using the three categories. The distribution of the dataset is shown in Table 4.

In terms of the distribution of the dataset, the 4,088 reviews collected from the web page contained 949 fake reviews, accounting for 23% of all reviews. As also seen from Table 4, the previous hypothesis is also reasonable. About 89% of the fake reviews come from the low usefulness, 9% from the middle usefulness, and 2% from the high usefulness dataset. This distribution also verifies that reviews with low usefulness are more likely to be fake.

In the experiment, six feature items are set for identifying the fake reviews. They are the length of reviews, the relativity of review content and products, the similarity of the review content, the repetition ratio of words in reviews, the maximum number of reviews customers posting in a day, and whether there is an opinion sentence in a review. The six indicators are used to distinguish “non-opinion reviews” and “untruthful reviews.” The thresholds are set for five of the feature items. The settings are shown in Table 5.

2.3. Experiment Procedure. To validate the effectiveness of each step during the experiment, we used the standard evaluation parameters, that is, precision (p), recall (r), and F-value (F), to test the performance of the experimental method.

2.3.1. Number of Words Selected. We collected 4,088 reviews for the dataset. It has about 89% of reviews within 50 words. The word number distribution of the dataset is shown in Figure 1.

Figure 2 shows the test of different thresholds in the experiment. It indicates the precision of fake reviews identified in the dataset when the threshold is set to different

TABLE 4: Dataset distribution.

Usefulness category	Number of reviews	Number of fake reviews	Number of real reviews
High	1521	19	1502
Middle	1345	86	1259
Low	1222	844	378

TABLE 5: Threshold values set for different features.

Feature item	Threshold
Length of reviews	50
Relativity of review content and products	0.4
Similarity of review content	0.5
Repetition ratio of words in reviews	0.5
Customers posting the maximum number of reviews in a day	3

parameter values. As is shown from the figure, when word count is set to 50, the precision is the highest, and the fake reviews in the dataset can be effectively identified.

2.3.2. The Relativity Experiment. The correlation test of review content and products is also an important indicator to measure whether the product review content is untruthful. The review content of many products has nothing to do with the product itself. These reviews are worthless for potential buyers; that is, the usefulness of the reviews is very low. According to the assumptions, these reviews need to be filtered and identified through the correlation evaluation.

In the correlation calculation, the statistical method χ^2 is used to select the relevant degree features. Then, the classification algorithms are used to verify the validity. In the procedure, the naive Bayes, support vector machine, and maximum entropy method are used to verify the

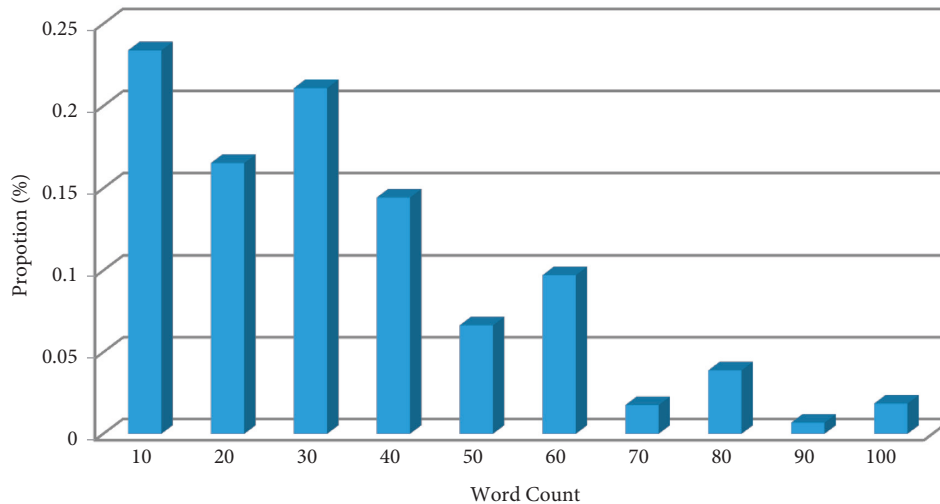


FIGURE 1: Word count distribution ratio.

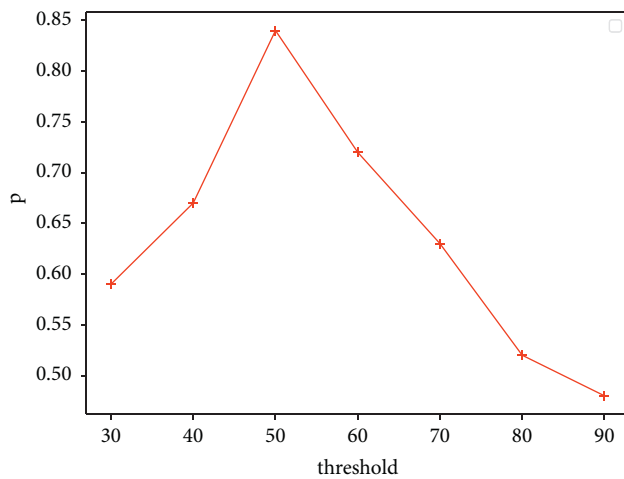


FIGURE 2: Precision for different thresholds.

effectiveness of selected features. The experimental results are given in Table 6.

From the results of Table 6, the method of testing product correlation by extracting the keywords of product attributes is effective. The correlation algorithm based on product attributes for naive Bayes algorithm, support vector machine algorithm, and maximum entropy algorithm is not very different. Through these three methods of test, the effectiveness of the correlation test of the review content and the product can be explained. Figure 3 and Figure 4 show the results of correlation test based on different methods at different thresholds.

2.3.3. Similarity Experiment. The similarity of review content is also an important indicator defined in the feature engineering, which is a reference basis for distinguishing “untruthful reviews.” We select keywords to reduce the dimension, reduce the computation amount of the algorithm, and improve the execution efficiency of the algorithm.

TABLE 6: Validity results of correlation feature selection based on different methods.

Method	Precision	Recall	F-score
NB	0.7316	0.7090	0.7201
SVM	0.8017	0.7844	0.7930
ME	0.7643	0.7115	0.7370

The results of review content similarity for different data sizes are shown in Table 7.

Much of product review content is similar, especially the features of the product. Different people use different ways to describe the features of the same product. For example, some people use some acronyms to describe product features. Although the name is different, the features described with others are the same. If the similar features of the product are not merged and the dimension of the review matrix is not reduced, the important product features cannot be well extracted, and the principal component of the product cannot be effectively analyzed. Table 7 shows the results after similarity incorporation, that is, the analysis results after dimension reduction. Table 8 shows the results of no dimensionality reduction of the review matrix.

As can be seen from the data shown in Table 7 and 8, the execution effect of the algorithm does not decrease after dimension reduction. It is better than the no reduction algorithm. Due to the loose grammar in the reviews, the language structure is not rigorous. If the components of the sentences are not reduced, the interference and noise in the algorithm will be large. It will have a great impact on the execution effect of the algorithm. So the key factors and the principal components that can represent the meaning of the sentences are extracted in the process of computing similarity. The effect is obvious by the calculation of these main parts. Figures 5 and 6 show the precision and recall of similarity calculations at different thresholds.

2.3.4. Repetition Ratio Experiment. One of the problems that often occur in the “non-opinion reviews” is to repeat words

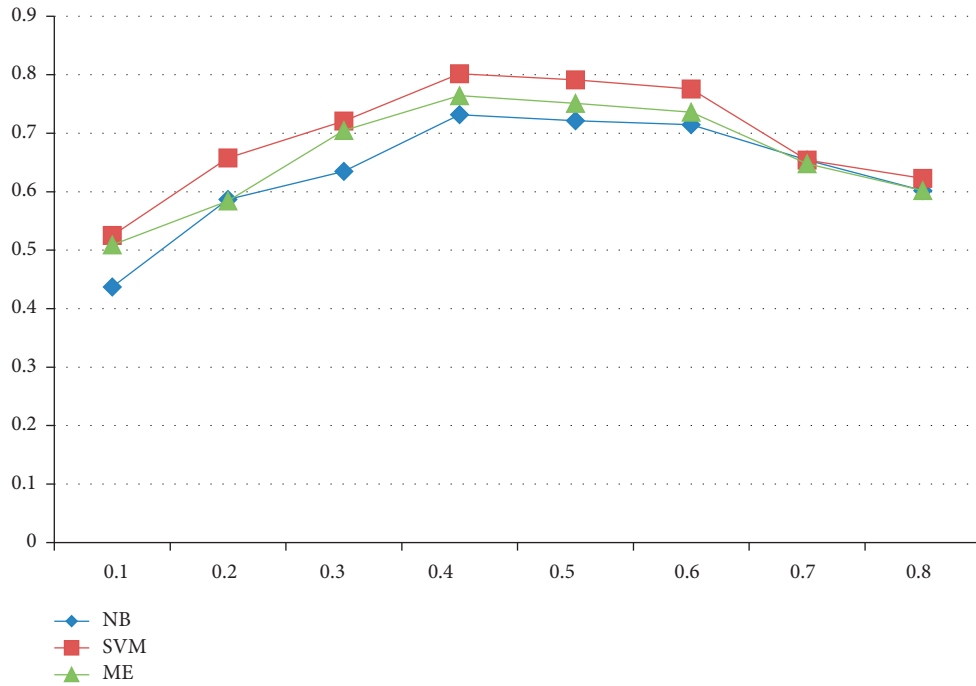


FIGURE 3: Precision of correlation at different thresholds.

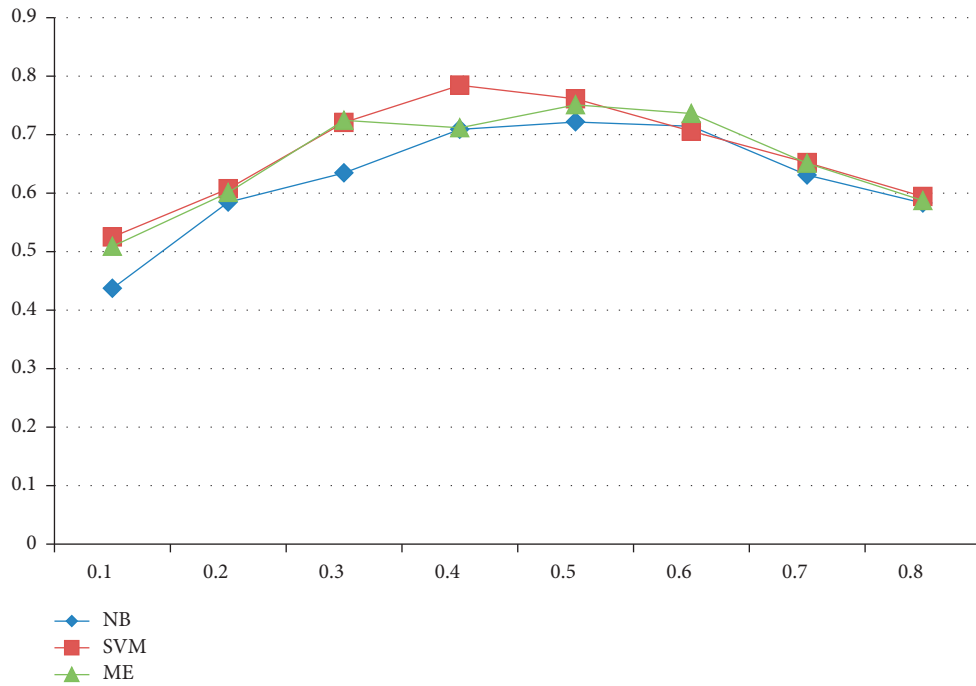


FIGURE 4: Recall of correlation at different thresholds.

TABLE 7: Similarity calculation results with dimensionality reduction based on different data sizes.

Data size	Precision	Recall	F-score
1000	0.8053	0.7664	0.7854
2000	0.8374	0.7841	0.8099
3000	0.8407	0.7923	0.8158

TABLE 8: Similarity calculation results without dimensionality reduction based on different data sizes.

Data size	Precision	Recall	F-score
1000	0.7247	0.6988	0.7115
2000	0.7563	0.7149	0.7350
3000	0.7609	0.7365	0.7485

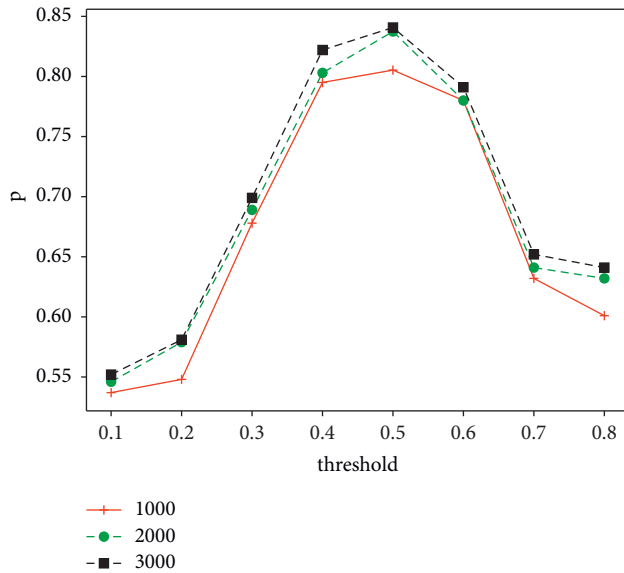


FIGURE 5: Precision of similarity calculation for different thresholds.

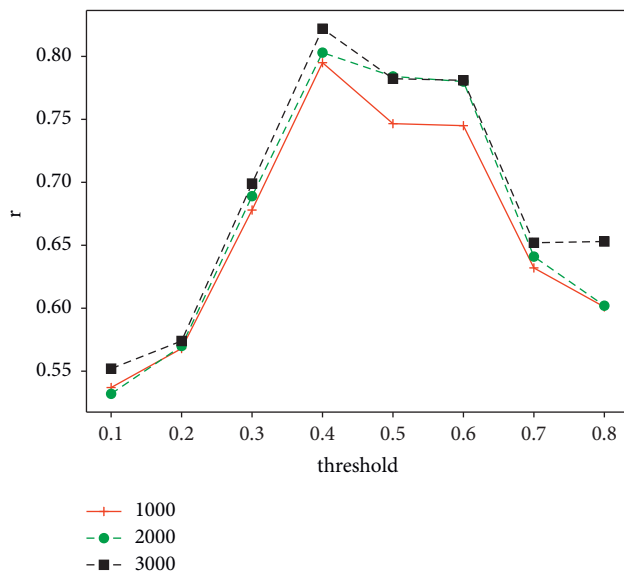


FIGURE 6: Recall of similarity calculation for different thresholds.

or phrases in order to meet the required number of words. These reviews are not useful. So the repetition ratio of words is also an important indicator in identifying fake reviews. Some reviews have a repetition rate of words up to 100%. Such reviews are worthless for potential customers. Figure 7 and Figure 8 demonstrate the calculation results of the repetition ratio of words at different thresholds.

2.3.5. Fake Review Identification Experiment Based on Multidimensional Feature Engineering. In this work, the fake review identification algorithm based on multidimensional feature engineering of union relationship and the fake review identification algorithm based on weighted multidimensional feature engineering are compared. The setting

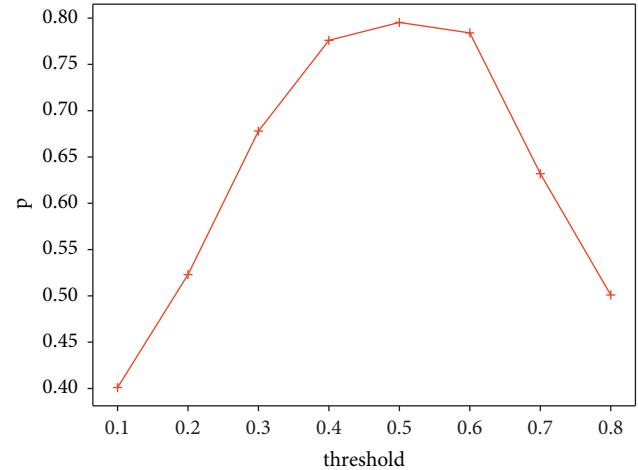


FIGURE 7: Precision of repetition rate for different thresholds.

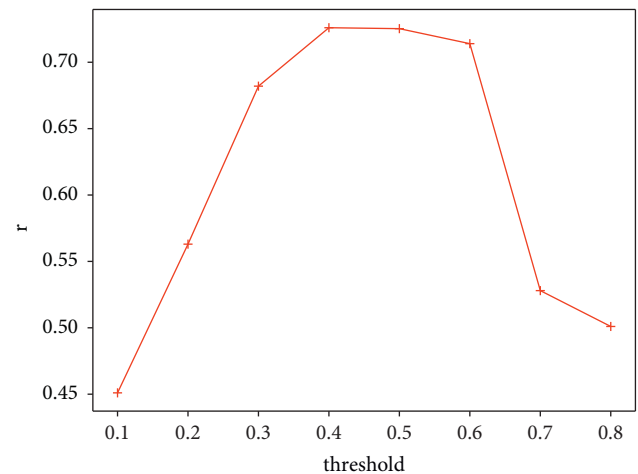


FIGURE 8: Recall of repetition rate for different thresholds.

of various parameters in the fake review identification algorithm based on multidimensional feature engineering of union relationship is discussed earlier.

In the identification algorithm based on weighted multidimensional feature engineering, the weight value of the feature is calculated according to the current data value. When the data size and content are different, the weight values are also different. According to the experiment, the weight values of different data sizes are tested, and the test results are shown in Table 9.

Experiments are performed separately using the method of scoring each review and the weight-based classification method. In the identification algorithm based on weighted multidimensional feature engineering scoring, the task of scoring each review is achieved by calculating the weight value. The threshold of scoring is 0.1. Reviews that scored above the threshold are identified as fake reviews; otherwise, they are true. The identification algorithm based on weighted multidimensional feature engineering classification uses support vector machine to identify fake reviews.

In the identification algorithm based on weighted multidimensional feature engineering scoring, the

TABLE 9: Weight calculation results.

Data size	F_1	F_2	F_3	F_4	F_5	F_6
100	0.0618	0.3688	0.0298	0.4000	0.1182	0.0213
500	0.0487	0.2929	0.2694	0.2864	0.0847	0.0179
1000	0.0493	0.2879	0.2710	0.2899	0.0843	0.0175
1500	0.0487	0.2875	0.2723	0.2916	0.0841	0.0158

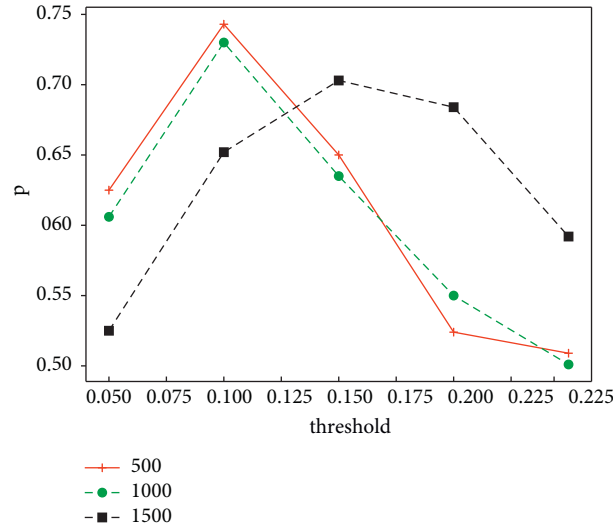


FIGURE 9: Relationship between threshold value and data size.

TABLE 10: Comparison results of three methods.

Methods	Precision	Recall	F-score
The fake review identification algorithm based on multidimensional feature engineering of union relationship	70.56	67.94	69.23
The identification algorithm based on weighted multidimensional feature engineering scoring	74.13	70.78	72.42
The identification algorithm based on weighted multidimensional feature engineering classification	76.69	72.47	74.52

relationship between the threshold and the data size is shown in Figure 9. The results of precision, recall, and F-score of the three fake review identification algorithms are shown in Table 10.

3. Conclusions

As is shown in Table 10, the identification algorithm based on weighted multidimensional feature engineering classification is better than the other two algorithms. Since the weight is calculated from the current dataset, the weight values of different datasets are different. And the weight values for different data sizes in the same dataset are also different. Therefore, the experimental results of the identification algorithm based on weighted multidimensional feature engineering scoring and the identification algorithm based on weighted multidimensional feature engineering classification will be different. However, from the overall effect of the experiment, the F-score of the identification algorithm based on weighted multidimensional feature engineering classification is the highest. The stability of the algorithms and the difference in different datasets are the problems that we will focus on solving in the future. In the

product reviews, many customers have uploaded the product photos. These pictures can also be used to judge whether the reviews are fake. In the future, image recognition and other related methods can be used to identify the fake reviews.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to thank all the students and teachers who participated in the following project research. This research was supported by the Natural Science Foundation of China (No. 91746104), Qingdao Philosophy and Social

Science Planning Project (No. QDSKL2101139), and Tai'an Social Science Project (No. 22-YB-033).

References

- [1] M. Syamala and N. J. Nalini, "A speech-based sentiment analysis using combined Deep learning and language model on real-time product review," *International Journal of Engineering Trends and Technology*, vol. 69, no. 1, pp. 172–178, 2021.
- [2] K. K. Gupta, S. Chennabasavaraj, N. Garera, and A. Ekbal, "Product Review Translation: Parallel Corpus Creation and Robustness towards User-Generated Noisy Text," in *Proceedings of the 4th Workshop on E-Commerce and NLP*, Bangkok, Thailand, 2021.
- [3] S. P. Sagar, K. Oliullah, K. Sohan, and M. F. K. Patwary, "PRCMLA: Product Review Classification Using Machine Learning Algorithms," in *Proceedings of the International Conference on Trends in Computational and Cognitive Engineering*, Singapore, 2021.
- [4] L. Qu, Y. Zhang, and F. Sun, "Effects of Online Product Review Characteristics on Information Adoption," *International Conference on Big Data Analytics for Cyber-Physical-Systems*, Springer, Singapore, 2021.
- [5] R. Mousavi, B. Hazarika, K. Chen, and M. Razi, "The Effect of Online Q&As and Product Reviews on Product Performance Metrics: Amazon.Com as a Case Study," *Journal of Information & Knowledge Management*, vol. 20, 2021.
- [6] A. H. Oluwadamilare, "Emoticon aware aspect based sentiment analysis of online product review," *Research In Computer Science and Library and Information Science*, vol. 7, no. 2a, pp. 166–179, 2021.
- [7] T. Sureshkumar, "Sentimental Analysis of Product Review Data Using Deep Learning," in *Proceedings of the 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Coimbatore, India, September 2021.
- [8] J. K. Rout, A. Dalmia, S. K. Rath, B. K. Mohanta, S. Ramasubbareddy, and A. H. Gandomi, "Detecting product review spammers using principles of big data," *IEEE Transactions on Engineering Management*, vol. 8, no. 99, pp. 1–12, 2021.
- [9] B. Liu, *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, San Rafael, CA, USA, 2012.
- [10] N. Jindal and B. Liu, "Review Spam detection," in *Proceedings of the International Conference on World Wide Web*, pp. 1189–1190, ACM, Banff, Canada, May 2007.
- [11] S. S. Shekhawat, H. Sharma, and S. Kumar, "Memetic Spider Monkey Optimization for Spam Review Detection Problem," *Big Data*, vol. 6, no. 99, 2021.
- [12] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proceedings of the international conference on web search and data mining*, pp. 219–230, ACM, Palo Alto, California, USA, Feb 2008.
- [13] E. P. Lim, V. A. Nguyen, N. Jindal, and W. Hady, "Detecting product review spammers using rating behaviors," in *Proceedings of the ACM International Conference on Information and Knowledge Management*, pp. 939–948, ACM, Toronto, ON, Canada, 2010.
- [14] N. Jindal, B. Liu, and E. P. Lim, "Finding unusual review patterns using unexpected rules," in *Proceedings of the ACM International Conference on Information and Knowledge Management*, pp. 1549–1552, ACM, Toronto, ON, Canada, 2010.
- [15] M. Ott, Y. Choi, C. Cardie, and T. H. Jeffrey, "Finding deceptive opinion spam by any stretch of the imagination," vol. 1, pp. 309–319, 2011, <https://arxiv.org/abs/1107.4557>.
- [16] F. Li, M. Huang, Y. Yang, and Z. Yi, "Learning to Identify Review spam," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 2488–2493, AAAI Press, Barcelona, Catalonia, Spain, 2011.
- [17] G. Wang, S. Xie, B. Liu, and P. S. Yu, "Review graph based online store review spammer detection," in *Proceedings of the IEEE, International Conference on Data Mining*, pp. 1242–1247, IEEE, Vancouver, BC, Canada, 2011.
- [18] G. Wu, D. Greene, and B. Smyth, "Distortion as a Validation Criterion in the Identification of Suspicious reviews," in *Proceedings of the Workshop on Social Media Analytics*, pp. 10–13, Washington D.C., Columbia, June 2010.
- [19] A. Mukherjee, B. Liu, J. Wang, and N. Jindal, "Detecting group review spam international conference companion on world wide web," pp. 93–94, ACM, Hyderabad, India, 2011.
- [20] P. Bhuvaneshwari, A. N. Rao, and Y. H. Robinson, "Spam Review Detection Using Self Attention Based CNN and Bidirectional LSTM," *Multimedia Tools and Applications*, vol. 80, pp. 1–18, 2021.
- [21] A. Neisari, L. Rueda, and S. Saad, ". Spam Review Detection Using Self-Organizing Maps and Convolutional Neural Networks," *Computers & Security*, vol. 106, no. nmb, Article ID 102274, 2021.
- [22] P. Hajek and J. M. Sahun, "Mining behavioural and sentiment-dependent linguistic patterns from restaurant reviews for fake review detection," *Technological Forecasting and Social Change*, vol. 177, 2022.
- [23] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What yelp fake review filter might be doing?" in *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM*, vol. 1, no. 1, pp. 409–418, Cambridge, MA, USA, July 2013.
- [24] A. Bhattarai, V. Rus, and D. Dasgupta, "Characterizing Comment Spam in the Blogosphere through Content analysis," in *Proceedings of the Computational Intelligence in Cyber Security*, pp. 37–44, IEEE, Nashville, TN, USA, 2009.