applied sciences

MDPI

Review

Spam Review Detection Techniques: A Systematic Literature Review

Naveed Hussain [1,2], Hamid Turab Mirza [1,*], Ghulam Rasool [1], Ibrar Hussain [2] and Mohammad Kaleem [3]

[1] Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Lahore 54000, Pakistan; naveed.hussain@se.uol.edu.pk; grasool@cuilahore.edu.pk
[2] Department of Software Engineering, The University of Lahore, Lahore 54000, Pakistan; ibrar.hussain@cs.uol.edu.pk
[3] Department of Electrical Engineering, COMSATS University Islamabad, Islamabad 44000, Pakistan; mkaleem@comsats.edu.pk
* Correspondence: drturab@cuilahore.edu.pk; Tel: +92-3204064547

Received: 20 February 2019; Accepted: 4 March 2019; Published: 8 March 2019

Abstract: Online reviews about the purchase of products or services provided have become the main source of users' opinions. In order to gain profit or fame, usually spam reviews are written to promote or demote a few target products or services. This practice is known as review spamming. In the past few years, a variety of methods have been suggested in order to solve the issue of spam reviews. In this study, the researchers carry out a comprehensive review of existing studies on spam review detection using the Systematic Literature Review (SLR) approach. Overall, 76 existing studies are reviewed and analyzed. The researchers evaluated the studies based on how features are extracted from review datasets and different methods and techniques that are employed to solve the review spam detection problem. Moreover, this study analyzes different metrics that are used for the evaluation of the review spam detection methods. This literature review identified two major feature extraction techniques and two different approaches to review spam detection. In addition, this study has identified different performance metrics that are commonly used to evaluate the accuracy of the review spam detection models. Lastly, this work presents an overall discussion about different feature extraction approaches from review datasets, the proposed taxonomy of spam review detection approaches, evaluation measures, and publicly available review datasets. Research gaps and future directions in the domain of spam review detection are also presented. This research identified that success factors of any review spam detection method have interdependencies. The feature's extraction depends upon the review dataset, and the accuracy of review spam detection methods is dependent upon the selection of the feature engineering approach. Therefore, for the successful implementation of the spam review detection model and to achieve better accuracy, these factors are required to be considered in accordance with each other. To the best of the researchers' knowledge, this is the first comprehensive review of existing studies in the domain of spam review detection using SLR process.

Keywords: spam review detection; opinion spam; fake review; review spammer detection

1. Introduction

Nowadays, websites have become the main source for individuals to express themselves. People can easily share their views about products and services by using e-commerce sites, forums, and blogs. Most people read reviews about product and services before buying them. Everybody on the web is now acknowledging the importance of these online reviews for other customers and for

vendors too. Vendors are also capable of designing their additional marketing strategies based on these reviews [1]. For example, if various customers buy a specific model of a laptop and write reviews regarding issues related with its screen resolution, then the manufacturer might become aware and resolve this issue to increase customer satisfaction

Recently, the trend of spam attack has increased because anybody may simply write spam reviews and post them to e-commerce websites without any constraint. Any company might hire individuals to write fake reviews for their products and services; such people are called as spammers. Spam reviews are usually written in order to earn a profit or to promote their products or services. This practice is known as review spamming.

One of the main problems about opinion sharing websites is that spammers can easily create hype about the particular product by writing spam reviews. These spam reviews may play a key role in increasing the value of the product or service. For instance, if a customer wants to purchase any product online, they usually go to the review section to know about other buyers' feedback. If the reviews are mostly positive, the user may buy, otherwise, they would not buy that specific product. This all shows that spam reviews have become the main problem in online shopping that may lead to a loss for both the customer and the manufacturer.

Review spam can financially affect businesses and might cause a sense of mistrust in the general public, therefore, due to its significance, this problem has recently attracted the consideration of the media and governments as well. Recent media news from the New York Times and BBC [2] have stated that "nowadays, spam reviews are very frequent on websites, and recently a photography company was exposed to thousands of customer spam reviews". Hence, detecting spam reviews appears to be a key area, and without solving this important issue, online review sites could become a place full of lies and completely useless [3]. To counter this issue, commercial review hosting sites, such as Yelp (www.yelp.com) and Amazon (www.amazon.com), have already made some progress in detecting spam reviews. However, there is still a lot of room for improvement in spam review detection techniques [4].

Generally, spam review detection approaches consist of the following steps. The first basic step is the gathering of the review dataset; since review datasets mostly consist of unstructured text and may contain noisy data, there is almost always a need to pre-process the datasets. The next step is to select a feature engineering approach, such as linguistic n-gram or an individual spammer-based features approach. Finally, different review spam detection techniques, such as machine learning and Lexicon-based techniques, are applied to figure out which reviews are spam.

The focus of this systematic literature review is to analyze, classify, and summarize the context of the productive research in the domain of spam review detection. Until now, there exists only a few survey papers [5–9] in the domain of opinion mining, and the focus of these studies has only been on Facebook posts, twitter data, discussion group, newsgroup posts, E-mail spam, and spam review detection models in general, whereas the current research also focuses on publicly available review datasets and feature engineering techniques to extract useful features or attributes from the datasets. Furthermore, the researchers propose a taxonomy of spam review detection methods and discusses different parameters that are used to evaluate the performance of spam review detection methods. To the best of the researchers' knowledge, this is the first attempt to compile all the work on spam review detection using the SLR process. In this study, the researchers formalize the basic search string to collect the most relevant studies in the domain of spam review detection. Studies from reputed journals and conferences are analyzed for this SLR. This study has selected 76 primary studies out of 1690 potential primary studies. The preferred studies are evaluated through multiple aspects.

This SLR proposes to answer the following research questions (RQs):

RQ1: Which feature engineering techniques are used for construction or extraction of features from review datasets?
RQ2: What methods are used to solve the problem of spam review detection?
RQ3: Which performance metrics are used to evaluate the performance of spam review detection methods?

The two main contributions of this study are as follows: (i) Provide researchers and practitioners with insight and further improvement prospects on the spam review detection problem. (ii) Present the accuracy of existing spam review detection methods to identify the most effective methods. The answer to RQ1 will help to build an understanding of suitable feature engineering approaches, elaborate commonly used features, and list openly accessible datasets in the domain of spam review detection. RQ2 will identify the existing spam review detection methods, devise taxonomy of approaches, and present a comprehensive comparison of the accuracy of those methods. The answer to RQ3 will help to identify different performance evaluation metrics for the spam review detection methods and present a comparative analysis of these matrices.

The remaining paper is divided into following main sections: Section 2 describes the details of the systematic literature review (plan, conduct, and analysis), Section 3 discusses the answers to research questions and addresses the open issues in the area of spam review detection. In the end, conclusions are discussed in Section 4 of the study.

2. Systematic Literature Review

The systematic literature review provides the answers to specific research questions, whereas the general survey paper gives a broad idea about the domain. This SLR is performed using the guidelines provided in [10] and the selection of the primary studies is performed according to [11]. The key objective of this research is to identify the best available feature extraction techniques, and present different existing models for spam review detection and available parameters to analyze these models. The process of SLR helps to determine different studies available in the domain of spam review detection and answer different research questions. The distinct phases of the systematic literature review are shown in Figure 1. Preceding the study, the researchers discuss how different steps are performed in each phase of SLR.

Figure 1. Systematic Literature Review (SLR) process.

2.1. SLR Planning

In the subsequent section, this study describes the details of how this SLR is planned.

2.1.1. Necessity of the SLR

It is necessary to collect the best evidence from the existing literature. The SLR process provides the best techniques to collect and analyze evidence from primary studies. It also addresses the importance of the different methods of each research question. This work is conducted by using the

guidelines provided by [10]. Following a search string is done to confirm that there exists no similar literature in this domain.

((‘Spam’ or ‘Fake’ or ‘Shill’ or ‘Opinion Spam’ or ‘Spammer’ or ‘Social Spam’) AND (‘Reviews’ or ‘Comments’ or ‘Online Comments’) AND (‘Detection’ or ‘Finding’) AND (‘Technique’ or ‘Method’) AND (‘Systematic Overview’ or ‘Systematic Review’ or ‘Research review’))

The studies are selected based on their title, abstract, and conclusion. The result of this search depicted that there is no SLR having the same scope.

2.1.2. Research Questions

Research questions (RQs) as shown in Table 1.

Table 1. Research Questions.

ID	Research Question	Motivation
RQ1	Which feature engineering techniques are used for construction or extraction of features from review datasets?	To understand different available review dataset and approaches of feature engineering and how these approaches help for extraction of useful features from data.
RQ2	What methods are used to solve the problem of spam review detection?	To identify existing spam review detection models and analyze these models based on their accuracy score.
RQ3	Which performance metrics are used to evaluate the performance of spam review detection methods?	Study different metrics which are used to evaluate the performance of different spam review detection methods.

2.1.3. Review Protocol

The following part describes the review protocol for performing this SLR. Moreover, it also elaborates the search process, selection of studies, and analysis of data.

Search Process: Search strings play a significant role in selecting a related set of existing studies. Therefore, the researchers first studied the main concepts and terminology in the spam review detection area and considered different keywords used in the RQs. The researchers tested the alternatives and synonyms for each of the keywords used in the specified RQs. In the end, this study used different Boolean operators (‘OR’ and ‘AND’) and wildcard (*) in order to prepare the search string. Table 2 shows the search string, which is joined by AND operator.

Table 2. Search string.

Population	Intervention
(Spam) And (Review) And (Detection)	(Spam OR Fake OR Shill OR Opinion Spam OR Spammer OR Social Spam OR Reviews OR Comments OR Online Comments OR Detection OR Finding OR Technique OR Method)

The population contains different keywords, e.g., “Spam” and “Review”, which are the major keywords that are used to filter out the search records. The search procedure produced 1690 initial studies. Out of these studies, 165 are selected as being relevant and 76 are selected as primary studies.

Selection of Studies: The selection of primary studies was performed by three types of searches i.e., primary search, secondary search, and snowball tracking. Major research efforts in the domain of spam review detection were started in 2007. Therefore, this study is based on a 12-year duration from 2007 to 2018. This research executed a primary search by using different online research databases (Springer, Elsevier, IEEE, Science Direct, and ACM), conference proceedings, e-journals, and review papers. Researchers of this study used Google and Google Scholar to search the research databases. In the secondary search, the primary search was evaluated by studying the titles, abstracts,

and conclusions, and lastly snowball tracking was utilized (i.e., going by the bibliography of works) to finalize the selected studies. Further, selection of studies using SLR [10,11] followed automatic and manual search mechanisms to find out the meaningful research available in the respective domain. At first, researchers performed an automatic search, which was based on search strings and is performed in search engines of relevant electronic data repositories. Then, researchers performed a manual search, which was based on studying the title, abstract, conclusion, and snowball tracking.

The experimental quality of the selected works is evaluated in the full-text-based analysis that is performed on the remaining 80 research papers. The selection criteria of different papers are dependent upon the raised RQs. The quality of any SLR depends upon the selection of relevant works. This SLR employed the subsequent criteria for the selection of existing studies:

General Criteria:

- Research is peer-reviewed.
- The research is related to the search string and area of "Spam Review Detection".
- The study contains an experimental evaluation or survey research.

Inclusion/Exclusion Criteria:

- Research publication must be in the range from January 2007 to December 2018.
- The selected study must be a full-length published paper.
- Research publications must be written in the English language.

Criteria specific to the research questions:

- The study includes a feature engineering approach and extracts features from data by using these approaches. (RQ1)
- The study contains methods which are used to resolve the problem of spam review detection. (RQ2)
- The study contains a performance evaluation of spam review detection methods. (RQ3)
- The study presents comparative analysis. (RQ1, RQ2, and RQ3)

Quality assessment criteria: While conducting any SLR, the most important part is to select high-quality literature for producing the most accurate and reliable analysis. The selection of the right keywords and well-defined exclusion and inclusion conditions are important activities in the SLR planning phase. For this work, the following criteria are employed to validate the quality of studies (Table 3).

Table 3. Quality criteria for selection of primary studies.

Type	Definition
Internal Validity	To validate that the presented study should contain context and assumptions.
External Validity	To validate that the findings of the study should be able to be applied in academia or industry.
Construct Validity	To validate the association among different research questions and their results.
Conclusion Validity	To validate the conclusion of the RQs, and that these conclusions are according to RQs.

Data Extraction: The data extraction form is designed in MS Excel to extract relevant data in a consistent manner for each RQ, as shown in Table 4. The extracted data provides information for further processing. Definitions of exact information that are captured in relation to RQs are discussed below.

1. **The contribution of extracted features or attributes:** Feature engineering is carried out to extract significant features or attributes from the dataset. Existing works have employed the following techniques for feature extraction from datasets: linguistics [9], the content of review [2], meta-data of review [10], and information about a product [11]. The extracted features are used in spam review detection models.

2. **Spam review detection model or method:** A spam review detection model supports identification of the spam review. The main purpose of this research is to categorize different review spam detection approaches and propose a taxonomy of review spam detection techniques.
3. **Performance evaluation metrics used for spam detection methods or models:** The accuracy of spam review detection methods is evaluated through different evaluation measures. This research analyzed the primary studies for the use of different performance evaluation measures.

Data Analysis and Synthesis: For analysis, this study utilized qualitative as well as quantitative methods (Section (2.2.2)).

Table 4. Data Extraction Form.

Purpose	Meta-Data
General Information	Article title, author(s) name(s), date of publication, publication venue
Specific Information	Availability of review datasets and feature engineering techniques based on linguistic and spamming behavioral approaches. Different methods or models of spam review detection based on machine learning and lexicon-based approaches. Different evaluation measures to assess the accuracy of spam review detection methods.

The following section presents the steps for conducting SLR.

2.2. SLR Conduct

In the following section, we discuss the selection process of primary studies and analysis of primary studies with respect to year of publication.

2.2.1. Search and Selection of Primary Studies

The mechanism of primary search and selection is shown in Figure 2. The initial primary search based on the search string produced 1690 studies.

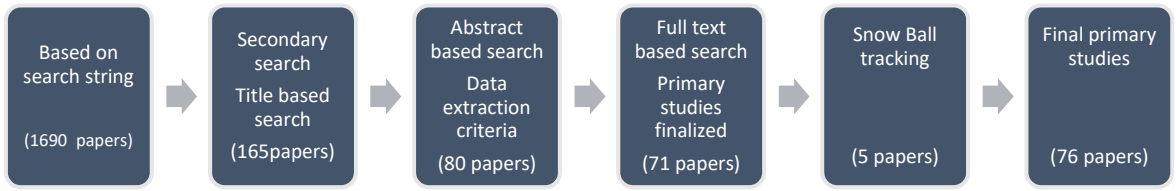

Figure 2. Primary studies selection process.

Table 5 shows the number of potential primary studies based on the search string.

Table 5. Potential primary studies after primary search.

Publisher	Number of Potential Studies
Springer	980
Elsevier	323
Science Direct	210
IEEE	98
ACM	45
IET Journal	34
Total	1690

After applying the secondary search, the result set is reduced by 165. Finally, 76 primary studies are selected based on inclusion/exclusion criteria, research questions, and quality assessment criteria (Table 3). Furthermore, snowball tracking technique is used to check the references of the studies [12]. Out of 76 studies, 41 were able to be published in the scientific journals, 29 in conferences, and 6 in workshop proceedings. The selection process is based on different stages: Pre-Stage is based on the search string from relevant journals, conferences, and workshops. Stage 1 is based on a title-based search to select potential studies. Stage 2 is based on an abstract based search to extract primary studies, and Stage 3 is based on a full text-based search.

2.2.2. Data Extraction and Analysis

The data about primary studies was extracted by using a data extraction sheet, as mentioned under the Data Extraction heading. Figure 3 shows the distribution of primary studies with respect to year of publication.

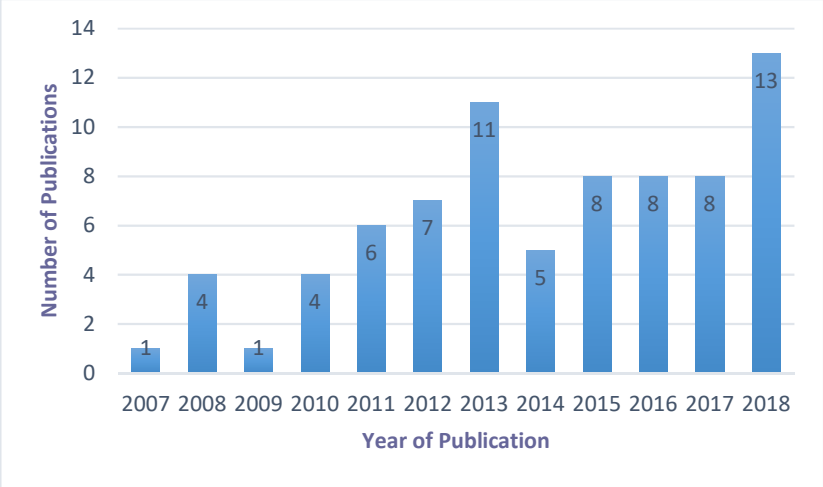

Figure 3. Distribution of selected studies with respect to time of publication.

3. Assessment and Discussion of Research Questions

This section discusses the answers of each RQ, and assessment of RQs is based on the analysis of 76 primary selected studies. Facts of each research question are extracted after the analysis of the selected studies.

3.1. Assessment of RQ1: Which Feature Engineering Techniques Are Used for Construction or Extraction of Features from Review Datasets?

This section discusses publicly available datasets and existing feature extraction techniques in the area of spam review detection. The construction or extraction of features from data is called feature engineering. Many studies have employed different types of feature engineering techniques to extract the most common features or words in reviews. The most common feature extraction

technique is the linguistic approach, and this technique is applied by the bag of words [13] approach. In a bag of words method, features for each review contain single words or small group of words found in a review text. Another feature engineering approach is based on individual spammer's behavioral characteristics. Further, spammer features can be classified into two types: review centric and reviewer centric features [14]. Features that are constructed using information contained in a single review are called review centric features. In reviewer centric features, this study not only considers all reviews written by an author on the same platform, such as Amazon.com or Tripadvisor.com, but also information about the author, such as review rating and time of posted review. Further data collection for one author from multiple platforms is at times not possible (due to change of user identity, non-availability of data, and disperse nature of data).

3.1.1. Review Datasets

Availability of a dataset is the key starting point of any spam review detection research. The key issue in the spam review detection problem is the availability of the labeled dataset. It has been observed from existing studies that only one labeled hotel review dataset is available [15], but it has only review text and no other features are available, such as review posting time and review rating. Researchers need to have access to the labeled dataset to train a classifier so that it may classify an unknown review as spam or not-spam. The alternative approach is to use an artificially created review dataset by using synthetic review spamming [16]. Ott et al. [15] have used Amazon Mechanical Turk (www.mturk.com) (AMT) for collecting labeled datasets through online workers (called Tuckers) to write fake hotel reviews (by giving one dollar for each review) in order to represent a few hotels in a positive light and some negatively. According to Mukherjee et al. [4], the process of labeling has not provided improved accuracy for spam review detection on real-life datasets. Table 6 lists review datasets used by different researchers and show total reviews, number of reviewers, and number of products for each dataset. It is observed by the review of the literature that all review datasets are not publicly available, and many times researchers use crawlers to gather required data. It has also been observed that most of the researchers used Amazon.com e-commerce website datasets in their works, as it is the biggest e-commerce platform to have product reviews, and the second largest review dataset is available from tripadvisor.com, which is an online hotel booking website. Additionally, the researchers working in the spam review detection area use these datasets provided by such websites. Through a few datasets, such as Amazon.com (http://jmcauley.ucsd.edu/data/amazon), Dianping (http://liu.cs.uic.edu/download/dianping), Resellarrating (www.trustpilot.com/review/www.resellerratings.com), and Datatang (http://www.datatang.com), products reviews and hotel review datasets are freely available, however the problem of unlabeled data exists in their datasets as well.

Table 6. Summary of review datasets used by researchers.

Paper ID	Dataset Category	Data Collection Source	Number of Reviews	Number of Reviewers	Number of Products
[17]	Different hotel reviews	Tripadvisor.com	27,952	1000	20,622
[18]	Different reviews about stores	Resellarratings.com	628,707	561,703	8737
[19]	Different reviews about stores	Resellarratings.com	408,470	343,603	1456
[20]	Different reviews about product	Datatang.com	10,020	291	9384
[21]	Reviews about different restaurant	Dianping.com	493,982	206,586	278
[22]	Different Reviews about product	Amazon.com	65,098	1703	53,353
[23]	Different Reviews about product	Amazon.com	109,518	53,469	39,392
[24]	Different Reviews about product	Amazon.com	195,174	141,501	300,864
[25]	Different Reviews about product	Amazon.com	3,794,69	1,037,621	962,234

[26]	Reviews about different electronics items	Amazon.com	542,085	424,519	35,992

Based on an analysis of existing studies, it is observed that very limited real-world labeled datasets are available. Hence, there is a need to have publicly available labeled standard review datasets that may be used by researchers for analyzing and comparing the results of different spam review detections techniques.

3.1.2. Feature Engineering Using the Linguistic Approach

The linguistic approach is the dominant approach for feature extraction or construction from review datasets. Moreover, this approach utilizes only review text and performs different steps, such as data preprocessing, tokenization, transformation, and feature selection [27]. In this study, researchers analyze and discuss commonly used feature engineering techniques.

1. Pre-Processing

- **Removing Stop Words or Punctuation**

Generally, the review text contains unnecessary words like "is", "the"," and", "a". These words are not helpful in detecting spam reviews, therefore, it is better to remove them before tokenizing to avoid noise and unnecessary tokens. For instance, take a review "This is a very good car". After removing stop words and punctuation, the review appears as "good car".

- **Part of speech tagging**

This basically involves tagging word features with parts of speech based on the context within the review text in which it is found [15,28]. Moreover, the relationship with the adjacent and related words in a review text is also tagged. A simplified form of POS tagging is the identification of words as nouns, verbs, adjectives, adverbs, etc.

- **Stemming Word**

A stemming algorithm converts different forms of the word into a single recognized form. For instance, considering the words "works", "working", and "worked" as instances of the word work. Stemming must be applied to the review text before tokenizing it.

2. Tokenization

In this method, single words or group of words are used as features. This linguistic technique is called uni-gram when one word is selected, bi-gram when two words are selected, tri-gram when three words are selected, and so on. This technique is called n-gram in general [28–30]. For example, consider a review "good car" and application of different n-gram techniques on it. Unigram: ["good", "car"], Bi-gram: ["good car"], Uni + Bi-gram: ["car", "good", "good car"]. This work employed different n-gram combinations on review data.

3. Transformation

Document term matrix is used to represent tokens generated by the n-gram model in the form of a sparse matrix. A sparse matrix defines the frequency of terms or tokens in the collection of the reviews. It was observed by the literature review that most of the researchers use the following two techniques for transformation.

- **Simple Count**

In simple count technique, the value of the term in the matrix is determined by the number of times it occurs in a document, e.g., if a token occurs 2 times its value will be 2, or 0 if it does not exist in the document. Moreover, the occurrence frequency of each term in a review is considered [31,32]. Simple count approach uses review text of the review to determine the number of times a term appears in the review text of a single review. Term frequency according to their structure is shown in Table 7.

Review 1: The car is very comfortable to drive and fuel consumption is good
Review 2: The interior is very comfortable
Review 3: The fuel consumption is good

Table 7. Example of text features frequencies dataset structure, for Review 1, 2, and 3.

Review	the	Car	Is	Very	Comfortable	to	Drive	Its	Performance	Good	Interior	Fuel	Consumption	and
Review 1	1	1	2	1	1	1	1	0	0	1	0	1	1	1
Review 2	1	0	1	1	1	0	0	0	0	0	1	0	0	0
Review 3	1	0	1	0	0	0	0	0	0	1	0	1	1	0

- **Term frequency and Inverse document frequency (TF-IDF)**

TF-IDF technique is intended to consider how significant a word or token is to a document in a collection of the corpus. The TF-IDF value increases proportionally to the number of times a token occurs in the document but is often decreases by the occurrence of the word in the corpus, which helps to adjust for the fact that few words occur more often generally. Nowadays, TF-IDF is one of the most popular term-weighting schemes and provides better results than a simple count technique. TF-IDF can be mathematically represented as,

$$\mathrm{TF}(t,d) = \frac{f_{t,d}}{\sum_{t'} f_{t',d}}$$

In the above equation, $f_{t,d}$ is frequency of term t in document d and $\sum_{t'} f_{t',d}$ is number of terms in document d.

$$IDF(t,D) = log\frac{N}{|d\epsilon D: t\epsilon d|}$$

where N is total number of documents and $|d\epsilon D: t\epsilon d|$ is number of documents with Term t in them. TF-IDF according to their structure is shown in Table 8.

Review 4: This is a very good car. A good car indeed.
Review 5: This book is very awesome, I loved it very much.

After removing Stop words review 4 and review 5 looks like "awesome", "book", "car", "good", "loved".

Table 8. Example of term frequency and inverse document frequency dataset structure, for Review 4 and 5.

Review	Awesome	Book	Car	Good	Loved
Review 4	0.0	0.0	0.707	0.707	0.0
Review 5	0.577	0.577	0.0	0.0	0.577

Feature selection

Feature selection is the most important technique after removal of unnecessary words, tokenization, and transformation techniques from the review dataset. Moreover, this technique is useful because of the greater dimensionality of text features and existence of noisy features and selects an optimal set of features and removes the irrelevant feature in order to improve the classification performance [33]. In this section, the researchers describe a few of the processes which are helpful for feature selection in the text classification.

- **Gini Index**

The most common method for calculating the discernment level of features is the use of a measure called the Gini-index. In the following equation, $p_i\,(w)$ is the conditional probability that the review belongs to class *i*. Moreover, the Gini index for the word w denoted by G(w) is described as follows.

$$G(\mathrm{w}) = \sum_{i=1}^{k} p_i(w)^2$$

The higher value of $G(\mathrm{w})$ indicates the higher discriminative power of word *w*.

- **Information Gain**

Another common method utilized for text feature selection is information gain or entropy. Let P_i be the global probability of class *i* and $p_i(w)$ is the probability of class *i*, given that the document contains of the word w. Moreover, $F(w)$ is the fraction of the document which contains the word w. The information gain measure $I(w)$ for a given word w is described as follows:

$$I(w) = -\sum_{i=1}^{k} P_i \,.\, log(p_i) + F(w).\sum_{i=1}^{k} p_i(w).log(p_i(w)) + (1 - F(w)).\sum_{i=1}^{k}(1 - p_i(w)).log(1 - p_i(w))$$

The greater the value of information gain $\boldsymbol{I(w)}$, the greater the discriminative power of word *w*.

- **χ2-Statistic**

The χ2-Statistic used a different method to calculate the lack of independence between the word w and a class *i*. In the following equation, *n* is the total number of reviews in the dataset and $\boldsymbol{p_i(w)}$ is the probability of class i for reviews which contain *w*. P_i is the global fraction of reviews consisting in class *i* and *F(w)* is the global fraction of reviews containing the word *w*. The χ2-Statistic between word *w* and class *i* is computed as follows:

$$X_i{}^2 = \frac{n.F(w)^2.(p_i(w) - P_i)^2}{F(w).(1 - F(w)).P_i\,.(1 - P_i))}$$

The above equation calculates the normalized value of $X_i{}^2$ and identifies the relevant words for different classes.

3.1.3. Feature Engineering Using Spammer Behavioral Features Analysis

The spammer behavioral features deal with metadata of the review rather than the text content of the review [34]. Table 9 shows different characteristics of reviews: review ID, reviewer ID, product ID, rating, useful, review word count, review date, and time. These review characteristic features may help in spam review detection.

Table 9. Review characteristics dataset structure.

Review	Review ID	Reviewer ID	Product ID	Rating	Useful	Number of Words	Date	Time
Review 1	192	332	4455	3	1	7	3-5-2013	07:22
Review 2	193	456	4455	5	1	8	3-5-2013	09:33
Review 3	194	287	4455	5	0	5	4-5-2013	22:11

It was observed from the literature review that the researchers used reviewer's features [28,35,36] to identify the spammer. Moreover, the spammer may share common features, such as activity pattern, IP address, geographical location, and profile characteristics. With the help of these features or a combination of these features, spammer, spam, and non-spam reviews can be identified [27,37,38]. Table 10 shows an example of reviewer-centric features.

Table 10. Reviewers characteristics dataset structure.

Reviewer Number	Product ID	Reviewer ID	E-mail	Number of Reviews	Date 1st Review Written	Date the Last Review Was Written	Max Number of Reviews per Day
Reviewer 1	4456	3312	jsmith @gmail.com	44	4-6-2013	8-9-2014	15
Reviewer 2	4456	4411	Johan27@yahoo.com	70	7-9-2014	6-6-2015	23
Reviewer 3	4456	2211	Stefan002@gmail.com	5	5-8-2015	8-9-2015	3

Some common individual spammer features are discussed below:

A maximum number of reviews: Existing works show that mostly spammers write more than one review on a particular day. Hence, this spammer behavioral feature may help to detect spammers [39–41].

Percentage of positive reviews: It is observed that most spammers write positive and favorable reviews, hence a high percentage of positive reviews about any product or service might be an indication of spam reviews [28].

Review length: Existing literature shows that most spammers do not write detailed reviews about a product or service, so this useful reviewer-centric feature may help to identify spammers [35,42].

Reviewer deviation: Mostly spammers' ratings deviate from the average review rating. Generally, spammers give a high rating for product or services [40,43].

Maximum content similarity: Similar text content of reviews about different products by a similar reviewer is found to be a strong indication of a spammer [2,4].

Discussion

Feature engineering is the base for any spam review detection work. The whole process of spam review detection relies upon the selection of a suitable feature engineering approach because the extracted features from the dataset become the input for the spam review detection method. The accuracy of the spam review detection method is highly dependent upon the extracted features [44]. Table 11 shows that linguistic-based approaches provide the best accuracy, as compared to review content-centric and reviewer behavioral feature approaches [4].

Table 11. Comparison of previous works and results for spam review detection using different feature engineering approaches.

Paper ID	Dataset Used	Feature Engineering Approaches	The Accuracy of Spam Detection
[35]	Crawled from Amazon website	Review and reviewer features	78%
[28]	Crawled from Amazon website	Product characteristics, review and reviewer features	78%
[15]	Hotel review dataset through Amazon Mechanical Turk (AMT)	Bi-gram	89.6%
[27]	Hotel review dataset through Amazon Mechanical Turk (AMT) + doctor and hotel reviews from domain experts	Part-of-speech tagging + Unigram + Linguistic Inquiry and Word Count	65%
[4]	Yelp's real dataset	Behavioral features + Bi-gram	86.1%

It has been observed from the existing studies that there is no real world labeled dataset available in the domain of spam review detection [4]. Labeled datasets help the researcher to implement the supervised spam review detection methods, therefore, to overcome this problem, usually people are hired to write reviews about products and services. However, this study has observed that online workers (called Tuckers) usually provide a large collection of word distributions. As a result, linguistic n-gram approach could falsely detect spam reviews with high accuracy. There is a need for publicly available real-world labeled datasets so that researchers become able to train their proposed models using standard real-world datasets.

It is also observed that most of the review datasets have a limited number of attributes and present attributes are not able to detect spam reviews accurately. For example, Amazon.com review dataset contains only the following attributes: product ID, product name, product title, product price, user ID, profile name, helpfulness, review score, review time, review summary, and review text. To achieve better accuracy in spam review detection, there is a need for multi-dimensional review datasets which contain more attributes, such as IP address and posting location of the reviewer [45].

3.2. Assessment of RQ2: What Methods Are Used to Solve the Problem of Spam Review Detection?

This section presents an in-depth analysis of each primary study to answer the above research question. Different spam review detection methods used by previous works are listed and the pros and cons of each are discussed.

Proposed Taxonomy

Based on the analysis of primary studies, this work proposes a new taxonomy that may be used by other researchers to classify existing approaches and to figure out the most appropriate technique to solve a spam review detection problem. It is observed that there are 2 major approaches to handle the spam review detection problem: Machine learning and Lexicon-based approaches. Further, these main approaches have been classified into different methods with associated classes to resolve the problem of spam review detection. The proposed taxonomy is presented in Figure 4. To the best of the researchers' knowledge, this is the first attempt to compile all the different approaches pertaining to spam review detection.

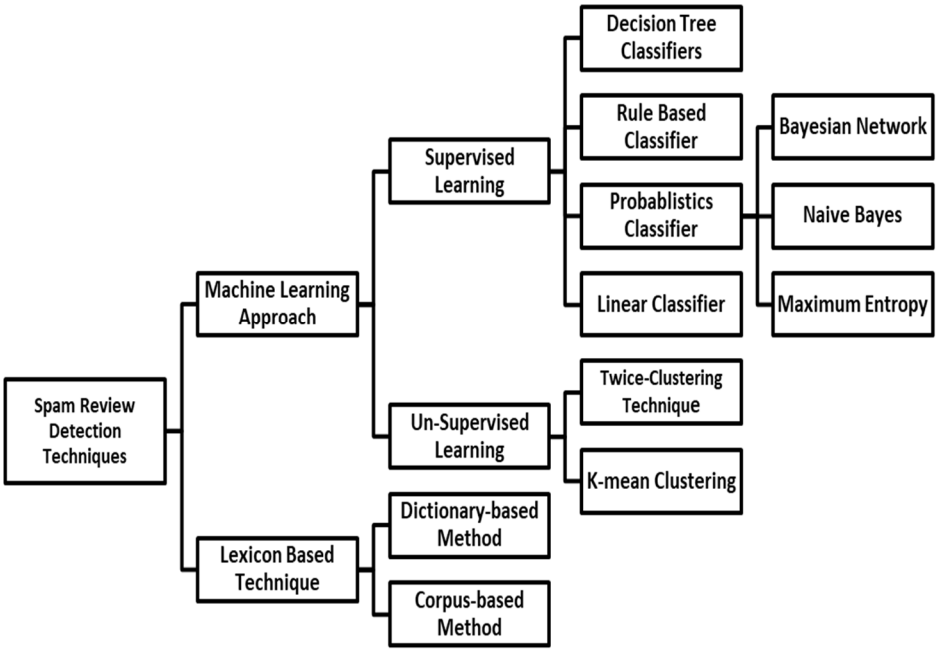

Figure 4. Taxonomy of Spam Review Detection Techniques.

The classification of spam review detection under this taxonomy is as follows below.

3.2.1. Machine Learning Approaches

Machine learning is one of the most important and prominent approaches for spam review detection and is generally categorized into supervised and unsupervised learning [40]. Below, researchers discuss different machine learning methods that have been proposed for spam review detection.

1. Supervised Learning

Supervised learning approaches used for spam review detection are commonly based on the classification methods [15]. In this learning technique, two datasets are required: training data and test data. Training data is utilized to train the classifier and afterwards test data is utilized to evaluate the performance of a classifier [28,46]. Methods such as Support Vector Machine (SVM) and Naïve Bayes (NB) have already shown great success in opinion mining.

Researchers usually start by gathering and crawling the dataset. The next step is to prepare and pre-process the dataset according to the domain. Once the dataset is prepared then features are extracted from the dataset by using the feature engineering approach. The next step is to train the classifier by using training data. Finally, the performance of a classifier can be validated by using test data [13]. Table 12 shows the comparison of different supervised learning techniques used in the spam review detection works.

i. Decision Tree Classifier

Decision tree (DT) classifiers give a hierarchical decomposition of the training data space and are used to learn the rules to identify the authenticity of the review [47,48]. A tree is formed by using different features and their values. Information gain is calculated by using a list of features. The feature that has maximum information gain is used as the root node of the decision tree. The interior nodes of the decision tree are labeled with unique features and these features have low information gain as compared to the root node. This procedure is repeated until all reviews are classified as spam or not-spam reviews. In the study by Jotheeswaran et al. [49], the IMDb movie reviews dataset is used, and inverse document frequency method is used to extract unique features and decision tree induction selects relevant features. They claimed to have correctly classified 75% of the reviews as spam or not-spam.

ii. Rule-Based Classifier

Rule-based (RB) classifiers use different rules to classify spam or not-spam reviews [31,50]. Rules may be applied to reviewer attributes, the content of the review, or the product. Ka et al. [51] used a rule-based approach to emotion cause component detection for a Chinese micro-blog dataset and they claimed 65% accuracy. A rule might be based on font size, time to write reviews, how often reviewer writes the reviews, length of the review, and how frequently sentimental words like “bad” and “good” are written [52,53]. The following four sample rules elaborate on the process of identifying spam or not-spam review class.

Rule_1: If a reviewer writes review 1 for product X and he again writes review 2 for product X within one minute, then the second review belongs to spam class.
Rule_2: If a reviewer writes review 1 for product X and he again writes review 2 for product X with the same font size and style, then the second review belongs to spam class.
Rule_3: If there are two reviews for the same product and the length of the reviews is also same, then the second review will be considered as spam.
Rule_4: If a reviewer writes the review for a product with too many sentimental words, such as “bad” and “good”, the review belongs to spam class.

iii. Probabilistic Classifier

The probabilistic approach is different from other approaches in a way that certain changes between different reviews are expressed statistically rather than some rules that are written by a human or machine [54].

Bayesian Network:

A Bayesian network shows the probability of the relationship among different nodes (features) [55], and the feature is an element of a review that is being used to classify the review. Moreover, each node of the graphical model represents a random variable and the edge represents the probability dependence between random variables. The relationship between different edges is represented by Directed Acyclic Graphs. The probability of a node occurring is the product of the probability that the random variable in the node occurs given that the parents have occurred. In the following equation, $P(x_1, \dots \dots x_n)$ is the probability of any node x_i and $P_a(x_i)$ is the probability of the parent. Moreover, $P(x_i \,|P_a(x_i))$ is also called conditional probability.

$$P(x_1, \dots \dots x_n) = \prod P(x_i \,|P_a(x_i))$$

This network model has been used in previous works to find spam reviews about any product or group of spam reviewers. Li et al. [27] crawled product reviews from Epinions.com and applied the Naïve Bayesian algorithm. They claimed 63% accuracy in the detection of spam reviews. Halees et al. [34] used Arabic opinion reviews from TripAdvisor and applied the Naïve Bayesian classifier for the spam detection. They claimed 99% accuracy. Similarly, the system proposed by reference [56] reported 94% accuracy on a customer review dataset by employing the Naïve Bayesian classifier.

Naïve Bayes:

Naïve Bayes (NB) classifier is also called a linear classifier and is used for both classification as well as training purposes. This is a probabilistic classifier method based on Bayes' theorem. Moreover, Naïve Bayes classifiers are based upon the naïve assumptions that the features in a dataset are mutually independent. The following equation is the mathematical representation of Naïve Bayes classifier.

$$\mathrm{P(C|X)} = \frac{\mathrm{P(X|C)P(C)}}{P(x)}$$

$\mathrm{P(C|X)}$ is the posterior probability of the target class with the given predicate attribute. $\mathrm{P(C)}$ is the prior probability of class. $\mathrm{P(X|C)}$ is the probability of the predictor class. $\mathrm{P}(x)$ is the prior probability of the predictor. There are different types of Naïve Bayes with different uses. It was observed from the literature review that the Naïve Bayes method is further divided into two text classification methods: (1) Multi-variate Bernoulli Naïve Bayes is used when feature vector is represented by 0 s and 1 s, where 0 s indicate a feature that does not occur in the review and 1 s represents a feature that occurs in the review; (2) Multinomial Naïve Bayes is typically used for discrete counts to determine how often a word occurs in the document.

Maximum Entropy:

Maximum entropy (ME) is used when there are only two outcomes of the classification. Maximum entropy model assigns a class by computing a probability from an exponential function of different features and assigns a different weight to each class [57]. Logistic Regression (LR) extracts a set of weighted features from the author's reviews, takes a log, and then each different feature is multiplied by the weight and calculated. Nitin et al. [35] used a crawled dataset from the Amazon website and applied Logistic Regression learner. They extracted review content and reviewer specific features and reported 78% accuracy.

iv. Linear Classifier

Linear classifiers utilize a linear combination of feature values of reviews and work well for the review classification problem, as it takes less time to train as compared to a non-linear classifier [58,59]. In linear classifiers, Support Vector Machine (SVM) classification is best suited for the text data. This is because of the sparse nature of the text where features are not related to each other, but they tend to correlate to one another, and generally, these features are organized into separate categories [15]. Support Vector Machine method analyzes data and defines decision boundaries by having hyper-planes. In binary classification problem, the hyper-plane separates the document vector in one class from other class, where the separation between hyper-planes is desired to be kept as large as possible. Support Vector Machine optimization procedure maximizes the predictive accuracy while automatically avoiding over-fitting of the training data. Moreover, SVM projects the input data into the kernel space, and then it builds a linear model in this kernel space. For dataset $(\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n)$, where y represents the class and x is the attribute which belongs to class y. Therefore, any hyper plane can be written as $\vec{w}.\vec{x} - b = 0$, where w is the normal vector to the hyperplane. SVM works very well for the small amount of training data and provides better results for good tokenizers. Several studies [4,15,33,36] used a SVM learner, where Ott et al. [15] and Shojaee et al. [33] used hotels review datasets through Amazon Mechanical Turk (AMT) and reported 89.9% and 84% accuracy, respectively. The variation in accuracy is because both works used different feature engineering techniques. On the other hand, Mukherjee et al. [4] used Yelp's real dataset and claimed 86.1% accuracy in detecting spam reviews and Fei et al. [36] used an Amazon dataset and reported 71% accuracy.

Table 12. Comparison of different Supervised Learning Techniques.

Paper ID	Dataset	Learner	Accuracy
[35]	Amazon.com	Logistic Regression (LR)	78%
[27]	Epinions.com	Naive Bayes (NB)	63%
[15]	Hotel reviews through Amazon Mechanical Turk (AMT)	Support Vector Machine (SVM)	89.9%
[36]	Amazon.com	Support Vector Machine (SVM)	71%
[4]	Yelp's real-life data	Support Vector Machine (SVM)	86.1%
[33]	Hotel reviews through Amazon Mechanical Turk (AMT)	Support Vector Machine (SVM)	84%
[34]	Arabic reviews from Tripadvisor.com and Booking.com	Naive Bayes (NB)	99%
[49]	IMDb movie	Decision Tree (DT)	75%
[51]	Chinese Language micro-blog	Random Forest (RF)	65%
[52]	Yelp restaurant reviews	Logistic Regression (LR)	79%
		Naive Bayes (NB)	72%
		Random Forest (RF)	76%
		Support Vector Machine (SVM)	78%
[53]	Amazon product reviews	Random Forest (RF)	91%

2. Unsupervised Learning

Publicly available review datasets with labeled classes are very scarce [4]. Hence, unsupervised learning methods that do not require a dataset with the class label are usually employed on such data [5]. Unsupervised learning methods drive the structure by considering the relationship between data;

this structure is known as clustering. Data in one cluster is dissimilar to the data in another cluster and a domain expert may suggest a label to any cluster by observing the characteristics of the data within that cluster. Table 13 shows the comparison of different unsupervised learning techniques.

Table 13. Comparison of different Unsupervised Learning Techniques.

Paper ID	Dataset	Method	Accuracy
[60]	Vietnamese mobile phone, product review	K-nearest neighbors (KNN) classification	72%
[61]	Product reviews from 360buy.com	Twice-clustering	66%
[62]	Chinese reviews from the web	K-means Constrained k-means clustering (COP K-mean)	Precision 71% Recall 61%
[63]	Real-life Amazon review dataset	Aspect-based review deviation, latent content deviation	78.15%

i. Twice-Clustering Technique

Twice-clustering is used to improve the precision and diversity of an unsupervised learning method [64–66]. Twice-clustering works in a series of steps. First, the original dataset is divided by using k-fold cross-validation. Second, all the training data to cluster is chosen for the first time to form a cluster subclass and then clustering is applied to each subclass to form a sample subset of each subclass. The sample subset of each subclass may be introducing some biasness. Therefore, to overcome this problem, it was observed by literature review that non-uniform random sampling is a good approach to form a sample subset of each subclass [67]. Finally, a subset of each subclass is selected to construct a training set to train an unsupervised learner. Jia et al. [61] used the twice-clustering method on a product review dataset from 360buy.com and reported 66% accuracy in the detection of spam reviews.

ii. K-means Clustering

K-means clustering has been shown to work well for large-scale data and its accuracy level is also high compared to other clustering algorithms [68]. The K-means clustering algorithm collects the extracted terms according to their feature values into K number clusters, and K is any positive number that is used to determine the number of clusters. The K-means clustering algorithm performs the following steps.

1. Pick a number (K) of cluster centers (at random)
2. Assign every item to its nearest cluster center (e.g., using Euclidean distance)
3. Move each cluster center to the mean of its assigned items
4. Repeat steps 2 and 3 until convergence is achieved (change in cluster assignments less than a threshold)

Specifically, previous works reported that K-means clustering yields promising results in the domain of opinion mining and spam detection [69]. Jia et al. [62] reported 71% precision by employing K-means clustering on Chinese language reviews about products. Ha et al. [60] used a K-means approach on mobile phone reviews and reported 72% accuracy.

3.2.2. Lexicon Based Technique

In this technique, different features of a given text are compared against sentimental lexicons and sentimental values are determined before their usage. People usually use different sets of words and expressions to express their feelings and opinions about a product or services. This list of words and expressions is stored in sentimental lexicons. A document is positive if it has more positive word lexicons, otherwise it is considered negative. Specifically, the following steps are carried out: (i) Each text is pre-processed by removal of HTML tags and noisy characters. (ii) Text sentimental score is initialized by 0. (iii) Tokens are assigned to each text and each token is checked, whether it is present in the sentimental directory or not. (iv) If the total sentimental score is greater than the threshold then

the review is classified as negative, otherwise it is positive. This technique falls under unsupervised learning, as it does not have labeled data for the training [70]. Table 14 presents the accuracy of different lexicon-based approaches.

There are two different methods for the construction of sentimental lexicon: Dictionary-based method and Corpus-based method.

1. Dictionary-based Method

In the dictionary-based method, targeted opinion words with an identified orientation are collected and are then searched from the WordNet dictionary for their antonyms and synonyms. The newly found words are added to the seed list. This iterative process is continued until no new words are found. The limitation of this method is that it is usually difficult to find different opinion words for a specific domain. Ben et al. [71] used the dictionary-based method on a review dataset from Blog06 and reported 78% accuracy in the detection of spam reviews. Taboada et al. [72] employed a dictionary-based method on an Amazon Mechanical Turk (AMT) dataset and reported 89% accuracy.

2. Corpus-based Technique

This technique is based on syntactic patterns in large corpora [73]. It produces a large collection of opinion words with high accuracy and needs large training data. Moreover, this approach can find opinion words with domain-specific orientation. The main benefit of this approach as compared to the dictionary-based approach is that Corpus-based technique produces specific opinion words in the respective domain and their orientations is better to understand. It may also help find domain and context specific opinion words and their orientations utilizing a domain corpus. The corpus-based technique, which is based on the domain-specific orientation, is best suited, as a word or phrase listed in an opinion lexicon does not mean that it is expressing an opinion in a sentence. For instance, in the sentence, "I am looking for good health insurance", "good" does not express either a positive or negative opinion on any insurance. Aurangzeb et al. [74] used corpus-based approach on customer reviews data and claimed 86.6% accuracy. Zhang et al. [75] employed a corpus-based technique on a Chinese language review dataset and proposed an aspect-based sentimental analysis system. They claim to achieve 82% accuracy through their system. Medinas et al. [76] used a combination of machine learning and lexicon approach. They claimed 82% accuracy by using CNET and IMDb datasets.

Table 14. Comparison of different Lexicon-based techniques.

Paper ID	Dataset	Technique	Accuracy
[74]	IMDB, Skytrax, Tripadvisor	Croup based technique	86.6%
[75]	Luce, Yoka	Croup based technique	82.6%
[72]	Amazon Mechanical Turk (AMT)	Dictionary-based method	89%
[71]	Blog06	Dictionary-based method	78%
[76]	CNET, IMDB	Machine learning (ML)+ Lexicon	82%

Discussion

This section reviewed existing literature on spam review detection methods published between 2007 and 2018. An attempt has been made to provide researchers with a comparative analysis of different spam review detection methods and their reported accuracy. Generally, spam review detection techniques are classified into two categories. The first one is machine-learning-based methods, which are further classified into two categories, supervised and unsupervised learning. The accuracy of different supervised-learning-based works is presented in Table 12. It shows that Support Vector Machine and Naïve Bayes perform better as compared to other supervised learning methods. Table 13 shows that Aspect based and K-nearest neighbors approaches perform better in unsupervised learning approaches. The second approach is Lexicon-based, which is further divided into two categories, Dictionary-based and Corpus-based methods. The Dictionary-based approach is more efficient in terms of processing time as compared to supervised learning but yields less

accuracy. The Corpus-based technique depends upon the dictionary related to the specific seed words of the domain. Table 14 shows that Corpus-based and Dictionary-based approach produce better accuracy as compared to other Lexicon-based techniques. It was observed in existing literature that all spam review detection methods are effective in identifying spam reviews; however, machine-learning-based supervised approaches generally yield better results. In recent years, new network-based filtering algorithms have been proposed, which filter out good or bad opinions from review datasets to aid the potential user, and these proposed algorithms produce better accuracy as compared to existing network-based spam review detection methods [77,78].

3.3. Assessment of RQ3: Which Performance Metrics Are Used to Evaluate the Performance of Spam Review Detection Methods?

This section presents different evaluation measures that are used in the domain of spam review detection. Review of primary studies shows that evaluation measures to assess the accuracy of the spam review detection methods are precision, recall, and f-measure [75]. Precision is used to measure the percentage of correct instances among the known positive instances, whereas recall is used to calculate the percentage of correct instances that can be known among the entire positive instance. F-measure is a geometric mean of precision and recall.

Accuracy is calculated by the following formula:

$$\text{accuracy} = \frac{\sharp\ \text{hits}}{\sharp\ \text{total\_review}}$$

where $\sharp\ \text{total\_review}$ represents the number of reviews used for the experiment and # hits represents the total positive or negative reviews that were correctly classified by the spam review detection method. Precision value can be calculated for positive or negative reviews that have been correctly classified. Therefore, the following equation is used to calculate positive precision:

$$\text{Positive\_precision} = \frac{\sharp\ \text{positive\_well\_classified}}{\sharp\ \text{positive\_well\_classified} + \sharp\ \text{positive\_bad\_classified}}$$

where $\sharp\ \text{positive\_bad\_classified}$ shows a number of negative reviews that were incorrectly classified as positive, and $\sharp\ \text{positive\_well\_classified}$ represents original positive reviews submitted to the proposed spam review detection method. The same procedure is used for negative or neutral reviews. Recall parameter can be computed only on positive or negative reviews that have been properly classified. Therefore, poitive_recall is defined by the following equation:

$$\text{poitive\_recall} = \frac{\sharp\ \text{positive\_well\_classified}}{\sharp\ \text{total\_positive\_reviews}}$$

where $\sharp\ \text{total\_positive\_reviews}$ represent those reviews that were originally considered as positive within the dataset, and $\sharp\ \text{positive\_well\_classified}$ is the number of positive reviews submitted to the spam review detection model and labelled as positive. Moreover, negative or neutral reviews are computed in the same manner. A false positive error occurs when the proposed model identifies reviews as spam that are not-spam, and a false negative error occurs when there is a failure to identify reviews as spam that actually are spam. Moreover, it was observed in the literature review that Precision is affected by the number of false positive errors and Recall is affected by the number of false negative errors.

For evaluation purposes, AUC (Area under the receiver operating characteristic curve) is also employed to evaluate the accuracy of classification models [35]. Moreover, to draw a receiver operating characteristic curve (ROC), only the true positive rate (TPR) and false positive rate (FPR) are needed. The TPR defines how many correct positive results occur among all positive samples available during the test, while FPR on the other hand defines how many incorrect positive results occur among all negative samples available during the test. Therefore, ROC space is defined by FPR and TPR as x and y axes. AUC and ROC are standard measures for evaluating the quality of machine learning algorithms. Lift curve is used to visualize the performance of spam detection models and to predict outlier reviews [28]. Lift curve is also used for class distribution. It is observed from existing

studies that researchers also used Rank Normalized Discounted Cumulative Gain to evaluate a spammer's individual features, e.g., rating, review text, etc.

Table 15. Performance metrics and performance validation methods.

Paper ID	Performance Metric	Performance Validation Method
[79]	F-measure	Graph-based & Clustering
[80]	F-measure and Precision	n-gram
[81]	Precision	Human Evaluation
[36]	Accuracy	K-means
[40]	Accuracy	Support Vector Machine
[24]	Accuracy	Support Vector Machine
[25]	Accuracy	Rank Boost
[26]	Receiver operating characteristic (ROC) Curve	Spammer behavioral features
[82]	Normalized discounted cumulative gain (NDCG) and Area under the receiver operating characteristic curve (AUC)	Classification algorithms
[43]	Rank Normalized discounted cumulative gain (NDCG)	Rating
[22]	Rank Normalized discounted cumulative gain (NDCG)	Rating
[18]	Rank Normalized discounted cumulative gain (NDCG)	Rating

Discussion

Accuracy is a basic measure to evaluate spam review detection methods. Previous works show that it is not practical to implement spam review detection methods without training [83]. The learner's accuracy is much higher when trained on real-world datasets [84]. Table 15 shows that most of the existing studies used precision, f-measure, accuracy and ranked normalized discounted cumulative gain parameters to evaluate the performance of spam review detection methods. Review of the state-of-the-art research shows that supervised learning methods mostly use the area under the ROC curve to evaluate the accuracy, whereas the lift curve is used to visualize the performance and outlier review of the methods.

3.4. Open Issues and Future Directions

This study identified that there are still several research gaps and open issues in spam review detection research. Major research gaps are elaborated below:

1. Unavailability of labeled datasets

The scarcity of labeled datasets is an open issue and challenge in the domain of spam review detection. One labeled dataset about hotel reviews [15] is available but it has a limited number of attributes. Researchers need to have access to standard labeled datasets to train the classifiers for the identification of spam or not-spam reviews.

2. The growing rate of review datasets

Millions of reviews already exist on review-based websites, such as Amazon.com, and the number of reviews and reviewers are growing rapidly. Such large datasets involve high computing power for experiments [6] and the implementation of semantic algorithms is one of the principal challenges in this domain. Semantic analysis of words depends upon SentiWordNet (http://sentiwordnet.isti.cnr.it/) and WordNet, both of which have an enormous dictionary of words that is utilized for sentiment analysis of reviews. Until now, no semantic-based model has been proposed for spam review detection.

3. Limited data attributes

Current publicly available review datasets have limited attributes. This limitation makes it challenging for researchers to detect spam reviews accurately. The main challenge here is the unavailability of multi-dimensional datasets. Many researchers depend upon the datasets gathered by crawling; however, such datasets also suffer from limited attribute problems. To improve the accuracy of the algorithms there is a need for more attributes, such as the IP address of the spammer, registered email address for the review website, and location where the reviewer signed in to write the review.

4. Multilingual review spam detection

A review is user-generated content and users can write a review in any language of their choice. So far, few researchers have worked datasets in languages other than English, such as Arabic, Chinese, or Malay [4]. There is a need to have in-depth research on the detection of spam in multilingual reviews.

5. Recognizing the spammer by analysis of feedback of other users on their written reviews

To detect spam reviews, researchers have made some progress by analyzing the content of the review and the reviewer's behavior. However, so far the reviewer's profile information has not been exploited by any work. Usually, there are follow-up comments or reviews by other users on the given reviews. For example, many websites ask such questions as "Did you find this review useful?" So far, such feedback or comments on given reviews have not been exploited as features for detection of spam reviews.

4. Conclusions

This study presented a systematic literature review of the spam review detection domain and highlighted recent research contributions in the form of different feature engineering approaches, spam review detection methods, and different measures used for performance evaluation. To extract precise pragmatic evidence, this work planned review methodology, focused on the search string, raised research questions, selected papers from renowned publishers, and used formal inclusion and exclusion study assessment criteria. A total of 1690 papers published from January 2007 to December 2018 were selected based on the search string, and after applying a title-based search 165 papers were shortlisted. Finally, by applying an abstract-based search, full-length analysis, and snow ball tracking, 76 publications were finalized for further study. Moreover, quality assessment criteria to determine the relevance and validity of the research domain appropriate to the selected publications were utilized. Table 5 presented a summary of the selected publications in the form of publisher and proceedings. A principle advantage of this study is that to the best of our knowledge, this is the first attempt to compile all the existing studies of spam review detection techniques using the SLR approach. Furthermore, the output of this study can be useful for further research in the area of spam review detection.

Main findings of this systematic literature review are summarized below:

Feature engineering techniques: Feature engineering can have a considerable influence on the performance of spam review detection methods [85–87]. Recent works on spam review detection used the same dataset, method, and evaluation measures, but reported varying results after applying different feature engineering approaches [4,15,32,33]. Table 11 presents the accuracy of works that used different feature approaches. It is also observed that the individual linguistic features and combination of linguistic and behavioral features yield better accuracy.

Spam review detection methods: Review of existing works showed that there are mainly two spam review detection techniques, namely machine learning and Lexicon-based. However, this domain is relatively new, therefore only a few studies on spam review detection have been conducted to date [88]. It is found that most of the existing studies focused on supervised machine learning approaches. Supervised learning must have a labeled dataset. Real-world datasets are difficult to acquire in this domain and most of the available datasets are synthetically created. However, building

models based on synthetically created datasets are not very reliable. For example, when an artificial AMT dataset [15,31,32] and Yelp's filtered real review dataset were used by the same framework, it was observed that Yelp's real-world dataset produced low accuracy, especially with n-gram linguistic features [4]. Lexicon approach is based on a set of precompiled sentimental terms, idioms, and phrases. Previous works have primarily used two subclasses of the Lexicon approach: Dictionary and Corpus-based. The Dictionary-based approach is used to collect a preliminary set of terms that are typically gathered in a manual way, whereas the Corpus-based approach is used to provide dictionaries linked to a particular area, and therefore produce better accuracy as compared to the Dictionary-based approaches.

Evaluation measures: It is observed that earlier works in the area of spam review detection have been evaluated using precision, recall, f-measure, and area under ROC curve [59] evaluation measures. Supervised machine learning methods mostly used AUC evaluation measures. Since most works are based on synthetically created review datasets, there is a need to evaluate their performance more rigorously with respect to how these methods will perform on real-world data. Future research should also focus on availability of standard labeled datasets for the researcher to train the classifier, and more attributes should be added to the dataset so as to improve the accuracy and reliability of the spam review detection models, such as IP address of the spammer and location where the reviewer signed in to write the review. Furthermore, there is a need for in-depth research on the detection of spam in multilingual reviews.

Author's Contributions: Every author contributed equally.

Funding: This research was not funded.

Acknowledgements: The authors acknowledge COMSATS University Islamabad, Lahore Campus, Pakistan, for providing facilities.

Conflicts of Interest: The authors state that they have no conflicts of interest.

References

1. Xue, H.; Li, F.; Seo, H.; Pluretti, R. Trust-aware review spam detection. In Proceedings of the 2015 IEEE Trustcom/BigDataSE/ISPA, Helsinki, Finland, 20–22 August 2015; Volume 1, pp. 726–733.
2. Lau, R.Y.; Liao, S.Y.; Kwok, R.C.; Xu, K.; Xia, Y.; Li, Y. Text mining and Probabilistic language modeling for online review spam detection. *ACM Trans. Manag.* **2011**, *2*, 25.
3. Serrano-Guerrero, J.; Olivas, J.A.; Romero, F.P.; Herrera-Viedma, E. Sentiment analysis: A review and comparative analysis of web services. *Inf. Sci.* **2015**, *311*, 18–38.
4. Mukherjee, A.; Venkataraman, V.; Liu, B.; Glance, N.S. What Yelp fake review filter might be doing? In Proceedings of the International Conference on Web and Social Media, Cambridge, MA, USA, 8–11 July 2013; pp. 409–418.
5. Rashid, A.; Anwer, N.; Iqbal, M.; Sher, M. Areas, Techniques, Challenges of Opinion Mining. *Int. J. Comput. Sci.* **2013**, *10*, 18–31.
6. Mukherjee, A.; Venkataraman, V.; Liu, B.; Glance, N.S. A Survey on Detection of Shill Reviews by Measuring its Linguistic Features. *Int. J. Emerg. Trends Technol. Comput. Sci. (Ijettcs)* **2014**, *3*, 269–272.
7. Spirin, N.; Han, J. Survey on web spam detection: Principles and algorithms. *ACM Sigkdd Explor. Newsl.* **2011**, *13*, 50–64.
8. Chakraborty, M.; Pal, S.; Pramanik, R.; Chowdary, C.R. Recent developments in social spam detection and combating techniques: A survey. *Inf. Process. Manag.* **2016**, *52*, 1053–1073.
9. Peng, J.; Choo, K.K.; Ashman, H. User profiling in intrusion detection: A review. *J. Netw. Comput. Appl.* **2016**, *72*, 14–27.
10. Keele, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*; Ver. 2.3 EBSE Technical Report: Software Engineering Group, School of Computer Science and Mathematics Keele University, UK *and Department of Computer Science University of Durham, UK*: 2007; pp. 1–57.
11. Kitchenham, B. *Procedures for Undertaking Systematic Reviews*; Joint technical report; Computer Science Department, Keele University (TR/SE-0401) and National ICT,*Sydney*Australia Ltd.: 2004; Volume 51, pp. 7–15.

12. Wohlin, C. Guidelines for snowballing in systematic literature studies and replication in software engineering. In Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, London, UK, 13–14 May 2014; p. 38.
13. Tripathy, A.; Agrawal, A.; Rath, S.K. Classification of sentiment reviews using n-gram machine learning approach. *Expert Syst. Appl.* **2016**, *57*, 117–126.
14. Xue, H.; Li, F. A Content-Aware Trust Index for Online Review Spam Detection. In Proceedings of the IFIP Annual Conference on Data and Applications Security and Privacy, Data and Applications Security and Privacy, Philadelphia, PA, USA, 19–21 July 2017; pp. 489–508.
15. Ott, M.; Choi, Y.; Cardie, C.; Hancock, J.T. Finding deceptive opinion spam by any stretch of the imagination. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Portland, OR, USA, 19–24 June 2011; Volume 1, pp. 309–319.
16. Cardoso, E.F.; Silva, R.M.; Almeida, T.A. Towards automatic filtering of fake reviews. *Neurocomputing* **2018**, *309*, 106–116.
17. Fayazbakhsh, S.K.; Sinha, J. *Review Spam Detection: A Network-Based Approach*; Final Project Report CSE 590(Data Mining and Network); 2012.
18. Peng, Q. Store review spammer detection based on review relationship. In *Advances in Conceptual Modeling*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 287–298.
19. Wang, G.; Xie, S.; Liu, B.; Philip, S.Y. Review Graph-based Online Store Review Spammer Detection. In Proceedings of the 11th IEEE International Conference on Data Mining, ICDM, Vancouver, BC, Canada, 11–14 December 2011; pp. 1242–1247.
20. Jiang, B.; Chen, B. Detecting product review spammers using activity model. In Proceedings of the International Conference on Advanced Computer Science and Electronics Information ICACSEI, Beijing, China, 25–26 July 2013; Atlantis Press, Paris, France: 2013; pp. 650–653.
21. Huang, J.; Qian, T.; He, G.; Zhong, M.; Peng, Q. Detecting professional spam reviewers. In *Advanced Data Mining and Applications*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 288–299.
22. Wang, J.; Liang, X. Discovering the rating pattern of online reviewers through data coclustering. In Proceedings of the IEEE International Conference on Intelligence and Security Informatics (ISI), Seattle, WA, USA, 4–7 June 2013; pp. 374–376.
23. Mukherjee, A.; Liu, B.; Glance, N. Spotting fake reviewer groups in consumer reviews. In Proceedings of the 21st International Conference on World Wide Web, Lyon, France, 16–20 April 2012; ACM: New York, NY, USA, 2012; pp. 191–200.
24. Lu, Y.; Zhang, L.; Xiao, Y.; Li, Y. Simultaneously detecting fake reviews and review spammers using factor graph model. In Proceedings of the 5th annual ACM Web Science Conference, Paris, France, 2–4 May 2013; ACM: New York, NY, USA, 2013; pp. 225–233.
25. Aye, C.M.; Oo, K.M. Review spammer detection by using behaviors-based scoring methods. In Proceedings of the International Conference on Advances in Engineering and Technology, Singapore, 29–30 March 2014; pp. 350–355.
26. Choo, E.; Yu, T.; Chi, M. Detecting opinion spammer groups through community discovery and sentiment analysis. In *Data and Applications Security and Privacy*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 170–187.
27. Li, J.; Ott, M.; Cardie, C.; Hovy, E. Towards a general rule for identifying deceptive opinion spam. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MD, USA, 22–27 June 2014; Volume 1, pp. 1566–157.
28. Li, F.; Huang, M.; Yang, Y.; Zhu, X. Learning to identify review spam. In Proceedings of the Twenty-Second International Joint Conference on Artifical Intelligence, Barcelona, Spain, 16–22 July 2011; Volume 22, p. 2488.
29. Jindal, N.; Liu, B. Opinion spam and analysis. In Proceedings of the International Conference on Web Search and Data Mining, Palo Alto, CA, USA, 11–12 February 2008; ACM: Stanford, CA, USA, 2008; pp. 219–230.
30. Fusilier, D.H.; Montes-y-Gómez, M.; Rosso, P.; Cabrera, R.G. Detection of opinion spam with character n-grams. In Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics, Cairo, Egypt, 14–20 April 2015; Springer: Cham, Switzerland, 2015; pp. 285–294.

31. Jindal, N.; Liu, B.; Lim, E.P. Finding Unusual Review Patterns Using Unexpected Rules. In Proceedings of the 19th ACM Conference on Information and Knowledge Management, Toronto, ON, Canada, 26–30 October 2010; ACM: New York, NY, USA, 2010; pp. 1549–1552.
32. Bajaj, S.; Garg, N.; Singh, S.K. A Novel User-based Spam Review Detection. *Procedia Comput. Sci.* **2017**, *122*, 1009–1015.
33. Shojaee, S.; Murad, M.A.; Azman, A.B.; Sharef, N.M.; Nadal, S. Detecting deceptive reviews using lexical and syntactic features. In Proceedings of the 13th International Conference on Intelligent Systems Design and Applications (ISDA), Delhi, India, 8–10 December 2013; IEEE: Serdang, Malaysia, 2013; pp. 53–58.
34. Hammad, A.S.; El-Halees, A. An Approach for Detecting Spam in Arabic Opinion Reviews. Ph.D. Dissertation, Islamic University of Gaza, Gaza Strip, Palestine, 2013.
35. Jindal, N.; Liu, B. Review spam detection. In Proceedings of the 16th International Conference on World Wide Web, Banff, AB, Canada, 8–12 May 2007; ACM: Lyon, France, 2007; pp. 1189–1190.
36. Fei, G.; Mukherjee, A.; Liu, B.; Hsu, M.; Castellanos, M.; Ghosh, R. Exploiting Burstiness in reviews for review spammer. In Proceedings of the International Conference on Web and Social Media, Cambridge, MA, USA, 8–11 July 2013; pp. 175–184.
37. Jiang, M.; Cui, P.; Faloutsos, C. Suspicious behavior detection: Current trends and future directions. *Ieee Intell. Syst.* **2016**, *31*, 31–39.
38. Algur, S.P.; Ayachit, N.H.; Biradar, J.G. Exponential Distribution model for Review Spam Detection. *Int. J. Adv. Res. Comput. Sci.* **2017**, *8*, 938–947.
39. Li, Y.; Lin, Y.; Zhang, J.; Li, J.; Zhao, L. Highlighting the Fake Reviews in Review Sequence with the Suspicious Contents and Behaviors. *J. Inf. Comput. Sci.* **2015**, *4*, 1615–1627.
40. Mukherjee, A.; Kumar, A.; Liu, B.; Wang, J.; Hsu, M.; Castellanos, M., Ghosh, R. Spotting opinion spammers using behavioral footprints. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; ACM: New York, NY, USA, 2013; pp. 632–640.
41. Li, H.; Fei, G.; Wang, S.; Liu, B.; Shao, W.; Mukherjee, A.; Shao, J. Bimodal distribution and co-bursting in review spam detection. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; pp. 1063–1072.
42. Fayazi, A.; Lee, K.; Caverlee, J.; Squicciarini, A. Uncovering crowdsourced manipulation of online reviews. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, 9–13 August 2015; ACM: New York, NY, USA, 2015; pp. 233–242.
43. Lim, E.P.; Nguyen, V.A.; Jindal, N.; Liu, B.; Lauw, H.W. Detecting product review spammers using rating behaviors. In Proceedings of the 19th ACM International Conference on Information and Knowledge Management, Toronto, ON, Canada, 26–30 October 2010; ACM: New York, NY, USA, 2010; pp. 939–948.
44. Yuan, L.; Li, D.; Wei, S.; Wang, M. Research of Deceptive Review Detection Based on Target Product Identification and Metapath Feature Weight Calculation. *Complexity* **2018**, *2018*, doi:10.1155/2018/5321280.
45. Cao, J.; Xia, R.; Guo, Y.; Ma, Z. Collusion-aware detection of review spammers in location-based social networks. In *World Wide Web*; Springer, Berlin Heidelberg,2018; pp. 1–31.
46. Hooi, B.; Shah, N.; Beutel, A.; Günnemann, S.; Akoglu, L.; Kumar, M.; Makhija, D.; Faloutsos, C. Birdnest: Bayesian inference for ratings-fraud detection. In Proceedings of the SIAM International Conference on Data Mining, Miami, FL, USA, 5–7 May 2016; Society for Industrial and Applied Mathematics **,**University of Pennsylvania, United States: 2016; pp. 495–503.
47. Zhang, Y.; Wang, S.; Phillips, P.; Ji, G. Binary PSO with mutation operator for feature selection using decision tree applied to spam detection. *Knowl. -Based Syst.* **2014**, *64*, 22–31.
48. Crawford, M.; Khoshgoftaar, T.M.; Prusa, J.D. Reducing Feature Set Explosion to Facilitate Real-World Review Spam Detection. In Proceedings of the FLAIRS Conference, Key Largo, FL, USA, 16–18 May 2016; pp. 304–309.
49. Jotheeswaran, J.; Kumaraswamy, Y.S. Opinion Mining Using Decision Tree Based Feature Selection through Manhattan Hierarchical Cluster Measure. *J. Theor. Appl. Inf. Technol.* **2013**, *58*, 72–80.
50. Erik, C. Affective computing and sentiment analysis. *Ieee Intell. Syst.* **2016**, *31*, 102–107.
51. Gao, K.; Xu, H.; Wang, J. A rule-based approach to emotion cause detection for Chinese micro-blogs. *Expert Syst. Appl.* **2015**, *42*, 4517–4528.
52. Kumar, N.; Venugopal, D.; Qiu, L.; Kumar, S. Detecting review manipulation on online platforms with hierarchical supervised learning. *J. Manag. Inf. Syst.* **2018**, *35*, 350–380.

53. Saumya, S.; Singh, J.P. Detection of spam reviews: A sentiment analysis approach. *Csi Trans. Ict* **2018**, *6*, 137–148.
54. Crawford, M.; Khoshgoftaar, T.M.; Prusa, J.D.; Richter, A.N.; Al Najada, H. Survey of review spam detection using machine learning techniques. *J. Big Data* **2015**, *2*, 2–23.
55. Vidisha, M.; Vala, J.M.; Balani, P. A survey on Sentiment Analysis Algorithms for opinion mining. *Int. J. Comput. Appl.* **2016**, *133*, 7–11.
56. Jeyapriya, A.; Selvi, C.K. Extracting aspects and mining opinions in product reviews using a supervised learning algorithm. In Proceedings of the 2nd International Conference on Electronics and Communication Systems (ICECS), Karpagam College of Engineering, Tamilnadu, India, IEEE26–27 February 2015; pp. 548–552.
57. Xuan, H.N.; Le, A.C.; Nguyen, L.M. Linguistic Features for Subjectivity Classification. In Proceedings of the International Conference on Asian Language Processing, Hanoi, Vietnam, 13–15 November 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 17–20.
58. Khan, M.T.; Durrani, M.; Ali, A.; Inayat, I.; Khalid, S.; Khan, K.H. Sentiment analysis and the complex natural language. *Complex Adapt. Syst. Modeling* **2016**, *4*, 2.
59. Moraes, R.; Valiati, J.F.; Neto, W.P. Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Syst. Appl.* **2013**, *4*, 621–633.
60. Ha, Q.T.; Vu, T.T.; Pham, H.T.; Luu, C.T. An upgrading feature-based opinion mining model on Vietnamese product reviews. In *Active Media Technology*; Springer, Berlin Heidelberg, 2011; pp. 173–185.
61. Jia, W.J.; Zhang, S.; Xia, Y.J.; Zhang, J.; Yu, H. A novel product features categorize method based on twice-clustering. In Proceedings of the International Conference on Web Information Systems and Mining (WISM), Sanya, China, 23–24 October 2010; IEEE: Piscataway, NJ, USA, 2010; Volume 1, pp. 281–284.
62. Jia, W.J.; Zhang, S.; Xia, Y.J.; Zhang, J.; Yu, H. Opinion mining based on feature-level. In Proceedings of the 5th International Conference on Image and Signal Processing (CISP), Agadir, Morocco, 28–30 June 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1596–1600.
63. Liu, Y.; Pang, B. A Unified Framework for Detecting Author Spamicity by Modeling Review Deviation. *Expert Syst. Appl.* **2018**, *112*, 148–155.
64. Wong, T.L.; Lam, W. An unsupervised method for joint information extraction and feature mining across different web sites. *Data Knowl. Eng.* **2009**, *68*, 107–125.
65. Zhan, T.J.; Li, C.H. Product Feature Mining with Nominal Semantic Structure. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Toronto, AB, Canada, 31 August–3 September 2010; IEEE: Piscataway, NJ, USA, 2010; Volume 1, pp. 464–467.
66. Ghode, M.; Bere, S.; Kamal, M.; Moholkar, M. Sentiment Analysis over Online Product Reviews: A Survey. *Int. J. Recent Innov. Trends Comput. Commun.* **2014**, *2*, 3766–3774.
67. Shang, Y. Subgraph robustness of complex networks under attacks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017, 1-12.
68. Deng, X.; Li, Y.; Weng, J.; Zhang, J. Feature selection for text classification: A review. *Multimed. Tools Appl.* **2018**, 1–20, doi:10.1007/s11042-018-6083-5.
69. Ding, X.; Liu, B.; Yu, P.S. A Holistic Lexicon-Based Approach. In Proceedings of the International Conference on Web Search and Data Mining, Palo Alto, CA, USA, 11–12 February 2008; pp. 231–240.
70. Annett, M.; Kondrak, G. A comparison of sentiment analysis techniques: Polarizing movie Blogs. In Proceedings of the Conference of the Canadian Society for Computational Studies of Intelligence, Windsor, ON, Canada, 28–30 May 2008; pp. 25–35.
71. He, B.; Macdonald, C.; He, J.; Ounis. An Effective Statistical Approach to Blog Post Opinion Retrieval. In Proceeding of the 17th ACM Conference on Information and Knowledge Management, Napa Valley, CA, USA, 26–30 October 2008; pp. 1063–1072.
72. Taboada, M.; Brooke, J.; Tofiloski, M.; Voll, K.; Steve, M. Lexicon-based methods for sentiment analysis. *Comput. Linguist.* **2011**, *37*, 267–307.
73. Kharde, V.; Sonawane, P. Sentiment analysis of Twitter data: A survey of techniques. *Int. J. Comput. Appl.* **2016;Volume 139, pp.5-15**.
74. Khan, A.; Baharudin, B.; Khan, K. Sentiment classification from online customer reviews using lexical contextual sentence structure. In Proceedings of the International Conference on Software Engineering and

Computer Systems, Pahang, Malaysia, 27–29 June 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 317–331.
75. Zhang, W.; Xu, H.; Wan, W. Weakness Finder: Find product weakness from Chinese reviews by using aspects-based sentiment analysis. *Decis. Support Syst.* **2012**, *39*, 10283–10291.
76. Medinas, A.; Zhang, D.; Levene, M. Combining lexicon and learning based approaches for concept-level sentiment analysis. In Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, Beijing, China, 12 August 2012; ACM: New York, NY, USA, 2012; Article 5, pp. 1–8.
77. Shang, Y. Hybrid consensus for averager–copier–voter networks with non-rational agents. *Chaossolitons Fractals* **2018**, *110*, 244–251.
78. Shang, Y. Resilient multiscale coordination control against adversarial nodes. *Energies* **2018**, *11*, 1844.
79. Xu, C.; Zhang, J.; Chang, K.; Long, C. Uncovering collusive spammers in Chinese review websites. In Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, San Francisco, CA, USA, 27 October–1 November 2013; ACM: New York, NY, USA, 2013; pp. 979–988.
80. Sandulescu, V.; Ester, M. Detecting singleton review spammers using semantic similarity. In Proceedings of the 24th international conference on World Wide Web, Florence, Italy, 18–22 May 2015; ACM: New York, NY, USA, 2015; pp. 971–976.
81. Wang, G.; Xie, S.; Liu, B.; Yu, P.S. Identify online store review spammers via social review graph. *Acm Trans. Intell. Syst. Technol.* **2012**, *3*, 1–21.
82. Wang, X.; Liu, K.; Zhao, J. Handling Cold-Start Problem in Review Spam Detection by Jointly Embedding Texts and Behaviors. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, BC, Canada, 30 July–4 August 2017; pp. 366–376.
83. Singh, A.; Batra, S. Ensemble-based spam detection in social IoT using probabilistic data structures. *Future Gener. Comput. Syst.* **2018**, *81*, 359–371.
84. Hazim, M.; Anuar, N.B.; Ab Razak, M.F.; Abdullah, N.A. Detecting opinion spams through supervised boosting approach. *PLoS ONE* 2018, *13*, e0198884.
85. Li, L.; Qin, B.; Ren, W.; Liu, T. Document representation and feature combination for deceptive spam review detection. *Neurocomputing* **2017**, *254*, 33–41.
86. Banerjee, S.; Chua, A.Y.K. Authentic versus fictitious online reviews: A textual analysis across luxury, budget, and mid-range hotels. *J. Inf. Sci.* **2017**, *43*, 122–134.
87. Wang, B.; Xiong, S.; Huang, Y.; Li, X. Review Rating Prediction Based on User Context and Product Context. *Appl. Sci.* **2018**, *8*, 1849.
88. Brar, G.S.; Sharma, A.P.A. Sentiment Analysis of Movie Review Using Supervised Machine Learning Techniques. *Int. J. Appl. Eng. Res.* 2018, *13*, 12788–12791.

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).