Spam Review Detection:A Systematic Literature Review

SUBMISSION DATE / POSTED DATE

14-09-2020 / 17-09-2020

CITATION

Farooq, Muhammad Shoaib (2020): Spam Review Detection:A Systematic Literature Review. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.12951077.v1

DOI

10.36227/techrxiv.12951077.v1

Spam Review Detection: A Systematic Literature Review

Rabia Aslam Khan, Muhammad Shoaib Farooq
University of Management and Technology, Lahore, Pakistan
Corresponding author: Muhammad Shoaib Farooq (e-mail:Shoaib.farooq@umt.edu.pj).

ABSTRACT In this era of technology, people rely on online posted reviews before buying any product. These reviews are very important for both the consumers and people. Consumers and people use this information for decision making while buying products or investing money in any product. This has inclined the spammers to generate spam or fake reviews so that they can recommend their products and beat the competitors. Spammers have developed many systems to generate the bulk of spam reviews within hours. Many techniques, strategies have been designed and recommended to resolve the issue of spam reviews**.** In this paper, a complete review of existing techniques and strategies for detecting spam review is discussed. Apart from reviewing the state-of-the-art research studies on spam review detection, a taxonomy on techniques of machine learning for spam review detection has been proposed. Moreover, its focus on research gaps and future recommendations for spam review identification.

INDEX TERMS Spam Analytics, Spam Review, Spam Detection Techniques, Fake Review.

I. INTRODUCTION

In this age of global village, online reviews are playing an important role for companies and customers, providing the base of a new type of WOM (Word-Of-Mouth) information. According to recent research, 52% of online buyers search about product-related information on the internet, whereas 24% of users browse for products while making any purchase [1]. Product reviews influence the decision of potential customers. Mostly, people buy products that have a high reputation and rating. A survey was conducted by "US Cone Communication" at Harvard university states that 64% of the buyers refer to the comments or reviews before making any purchase[2]. According to the survey, a 1% growth in star rating effects 5% to 9% increase in overall revenue of specific product. Therefore, product manufacturers give more importance to review analysis in order to carry economic activities. In comparison to honest reviews given by real buyers, skinner used to post fake reviews in order to manipulate consumers decision in purchasing of specific products.

According to recent studies about 25% to 30% online reviews are spam [2, 3]. The threat occurrence of the spam reviews, which can lose the confidence of the customer about product rating based in reviews[4]. Spam review is more difficult to identify by customer whether a specific review is fake or real [5]. Little efforts have been made on detecting and avoiding these spam reviews. The activities of spam review are performed by the users who want to manipulate the selection and purchasing decisions of customers. The term used for such manipulators is sock-puppeting. Groups of spammers post the bulk of spam reviews to manipulate the rating of products. Therefore, spam reviews can be posted casually or copied for the other user reviews. Too much inconsistency exist between review text and rating of product [6].

Jindal.L et al. [7] characterize spam reviews into three types; Brand Reviews, Non-Reviews and Untruthful Review as shown in fig 1. Brand Reviews are related to product sellers and these reviews ignores feedback about the product form customers. In a product review, confusing customers by adding reviews of other non-relevant products are called Non-reviews. Untruthful reviews are those that provide related information, but the provided information is wrong. Therefore, brand reviews and Non-Reviews can be easily be detected by customers then detecting untruthful reviews [7, 8].

A number of techniques have been found for spam review detection and most of the techniques based on machine learning [9].

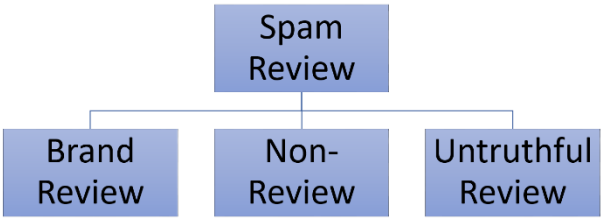

FIGURE 1. Types of Spam

Many machine learning techniques have been implemented to detect spam reviews. These techniques have been classified into three categories: supervised learning, unsupervised learning. In supervised learning, a function is used to map the input to the output depending

on examples of related input-output pair. Supervised learning techniques that have been used for spam review detection so far are; Rule based classification [5, 10], Unified model [2], Logistic Regression [4, 11, 12], K-nearest neighbor (KNN) [4], Random Forest [4, 13-15], Decision Trees [16, 17], Gradient Decent[4, 10], Genetic Algorithm [18], Conceptual Model [19], Time Series [20], Neural Network [21], Deep Neural Network [22], Multinomial Naïve Bayes [9, 11, 13], N-Gram [13], Hybrid Learning Approach (Active and supervised learning) [23], RNN, CNN [24], and Multilayer Perceptron Model (MLP)[4, 24],

Unsupervised learning is a category of machine learning that work on the unlabeled datasets. Many unsupervised learning techniques have been used in spam detection which are: Natural Language Processing [6, 9][58] Markov Network [25], Neural Auto-encoder Decision Forest [16], and PU Learning [26]. Other than these supervised and unsupervised learning techniques, there are many other techniques that have been used for spam detection such as Fuzzy Logic [27], Heterogeneous Information Network [28], Hadoop [29], Text Mining [30], Sentiment Analysis [31-35], Cuckoo Search [36][57], Adaptive Binary Flower Pollination [37], and Map Reduce[29].

Spam Review Detection has been the most active area of research in past years that covers all broad. In [2] a classier has been build based on logistic regression with content characteristics, feedback features, and rating features to identify fake reviews. Earlier studies have proposed to label datasets in two categories: duplicate reviews as spam reviews and the rest of the reviews as legitimate reviews. However, Jindal and Liu identified that many spam reviews were written in a way that it looks authentic. Hence, they determined that using duplication feature to differentiate legitimate reviews and spam reviews is not suited for creating label datasets [16].

Hernández F. et al. [26] presented PU Learning that builds a binary classifier. In PU Learning two sets were trained: set of positive instances (P) and set of both negative and positive instances but without a label (U). PU Learning technique depicts improvement in results compared to other techniques. *Heydari, A. et al*. introduces a system for detecting spam reviews using time series. They investigate fake reviews posted at doubtful time intervals. Moreover, they employ rating behaviors, context similarity, and people activeness in each time interval to differentiate between spam reviews and legitimate reviews [20].

Luyang, B, W, T., et al. [21] uses Sentence Convolutional Neural Network (SCNN) and Sentence Weighted Convolutional Neural Network (SWNN) to detect spam reviews. SCNN and SWNN were designed by modifying document-representation learning model. The time complexity of the SCNN and SWNN model is $O(n*d^2)$. The SWNN model gives an accuracy of 86.1% as compared to the basic convolution neural network. *Shreyas Aiyar, N. S.et al.* [13] proposed the spam review detection using N-gram. Their model improves the accuracy of classification, whereas we have also identified the research gaps in implementing techniques used for spam review detection.

Nidhi A. Patel at el. [33] presents the techniques and datasets used for spam review detection. Moreover, they discuss the limitations of datasets such as limited number of features and unlabeled datasets whereas along with all these we have proposed taxonomy, which classifies the existing techniques and approaches so that the most appropriate approach can be figured out.

SP. Rajamohana *at el.* [8] discusses the accuracy of adapted techniques using evaluation metrics whereas we have also discussed the open issues and challenges in the domain of detecting spam review.

In this paper, a systematic mapping process has been implemented. As a result, this mapping process has allowed summarizing the techniques used for detecting spam review. This systematic mapping study focuses on analyzing, classifying, and summarizing the context of research in view of the spam review detection. Moreover, this paper also proposed a taxonomy for spam review detection

The remainder section is organized in the following manner: Section II describes the used research methodology focusing on research questions, objectives, search strategy, screening of relevant papers, inclusion/exclusion criteria, sources, data extractions, and classification scheme; Section III define and design tables of finding and results obtained from section II; Section IV discusses the assessments of research questions; Section V and VI presents the discussion and conclusion

II. RESEARCH METHODOLOGY

The research methodology, "Systematic Literature Review" is selected for this kind of study. The objective of the systematic mapping is to provide an overview of the work that has been already done for detecting spam reviews so far. It establishes the research evidence if it exists. In order to complete this study, we used the process of systematic literature review explained by Petersen [38]. To write a systematic literature review, guidelines were implemented described by the Charters and Kitchenham [39]. The main objective of this study is to propose a taxonomy and explore existing research that has been done to detect spam review. The process followed for systematic mapping is shown in Fig. 2.

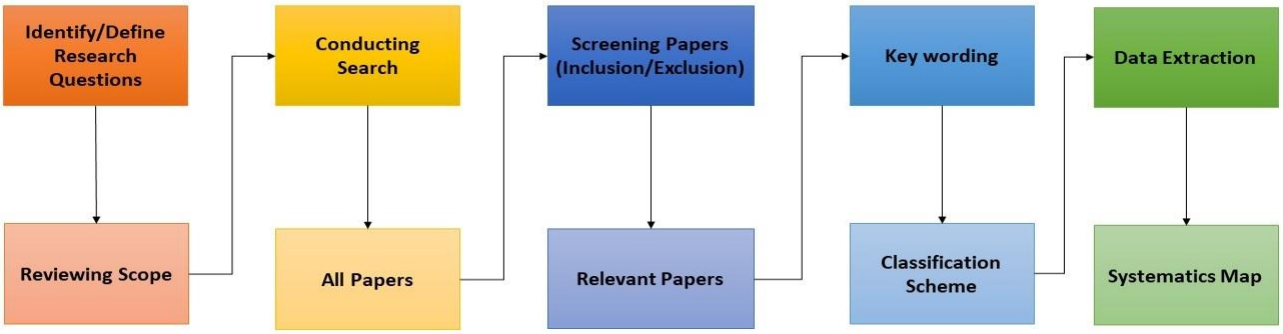

FIGURE 2. **Systematic Mapping Process**

A. RESEARCH OBJECTIVES

This research consists of the following objectives.

RO1: Characterize and categorize existing techniques in the domain of spam review detection.

RO2: A taxonomy is proposed that shows the adopted techniques and approaches used for detecting spam review.

RO3: Identify the challenges and research gaps.

RO4: More focused research has been done in the domain of spam review detection.

B. IDENTIFY/ DEFINING RESEARCH QUESTIONS

Basic and important step of systematic literature review is identifying and defining the research questions.

- The study focuses on gaps in approaches that were utilized to determine the issue of "spam review detection".
- The research contains detailed information of "spam review detection" techniques.
- The study presents relative information and analysis.

After having detailed literature three most important research questions related to spam review detection are shown in Table I.

TABLE I
RESEARCH QUESTIONS

RQ-Id	Research Question	Motivation
RQ-1	Which approaches have been used to detect spam reviews?	To identify that which techniques and approaches have been used for detecting spam review.
RQ-2	What are the gaps between previously defined techniques to detect spam reviews?	To discover the gaps between previously defined technologies that is there any limitation for using these techniques.
RQ-3	What information and features have been found in the datasets of reviews?	To identify the feature of datasets that have been used for spam review detection.

R-Q1: Which approaches have been used to detect spam reviews?

A systematic literature review enables to understand what technology already has been used to detect spam reviews. How it was implemented and what are the pros and cons of the previously used technologies?

R-Q2: What are the gaps between previously defined techniques to detect spam reviews?

Systematic Mapping Process of the research helps in understanding the gaps and current research of the spam reviews detection.

R-Q3: What information and features have been found in datasets of reviews?

This research question has helped us to identify the datasets and their features. These features help to select the method for detecting spam reviews.

C. CONDUCTING SEARCH

Conducting search is the second stage of the SLR. In the stage, all the relevant papers are searched related to the research topic. Methods defined by the research protocol has been used to undertake a specific search of systematic literature. A search string is defined to collect all papers which are related to the research topic using scientific databases. The terms or phrases used in search string were selected after the initial searches, where all possible keywords were tested. Therefore, the goal of this systematic mapping process is to search and map papers that relate to the technical aspects of the spam reviews.

The search string that was used to collect all the related papers is described in Table II. After the design and test of search string, we selected all the authentic scientific databases for searches. To conduct a systematic literature review, peer-reviewed, and high-quality papers which are published in workshops, books, conference, journals, and symposium that are related to the research topic. There are five scientific databases that are used for paper retrieval. Selected databases are Science Direct, IEEE Xplore, Spring Link, ACM Digital Library and Elsevier.

1) SEARCH STRATEGY

In this step, relevant studies were identified for review. The articles used for conducting searches were consulted from five scientific databases: ACM Digital Library, Springer Link, Science Direct, IEEE Xplore and Elsevier. Another source Google Scholar was also used in order to access the gray literature in this field like white papers or technical reports. The search string was generally defined using the following equation 1 whereas K_P represents primary keywords, K_S represents secondary keywords and K_A represents additional keywords.

$$\forall K_P \wedge \forall K_S \wedge \forall K_A \quad (1)$$

Following is the search string that was used to perform an automatic search in scientific or databases. Fig. 2 is the representation of the Search String that how it works.

"Spam" AND ("Fake" OR "Junk") AND ("Review" OR Opinion") AND ("Identification" OR "Detection" OR "Analytics") AND ("Predictive" OR "Descriptive")

The search string for all scientific databases were checked and modified. Table II depicts the search strings that were used to search the five scientific databases.

TABLE II
SEARCH STINGS OF SCIENTIFIC DATABASES

Scientific Database	Search String
IEEE Xplore	"Spam" AND ("Fake" OR "Junk") AND ("Review" OR Opinion") AND ("Identification" OR "Detection" OR "Analytics") AND ("Predictive" OR "Descriptive")
Science Direct	Title, abstract, keywords: Spam ("Fake" OR "Junk") Review (Comments OR Opinion") Identification ("Detection" OR "Analytics")
ACM digital Library	((Spam) AND ("Fake" OR "Junk") AND ("Review" OR Opinion") AND ("Identification" OR "Detection" OR "Analytics") AND ("Predictive" OR "Descriptive"))
Elsevier	"Spam" AND ("Fake" OR "Junk") AND ("Review" OR Opinion") AND ("Identification" OR "Detection" OR "Analytics") AND ("Predictive" OR "Descriptive")
Springer Link	"Spam" AND ("Fake" OR "Junk") AND ("Review" OR Opinion") AND ("Identification" OR "Detection" OR "Analytics") AND ("Predictive" OR "Descriptive")

D. INCLUSION/EXCLUSION CRITERIA

Inclusion criteria refers to the characteristic that should be considered and exclusion criteria refers to the characteristic that prohibit viewpoint subjects from including them in the study. Following Fig 3. shows the inclusion exclusion criteria for this systematic mapping study.

	Inclusion Criteria
i.	Research articles should be peer-reviewed.
ii.	Research articles should be related to the search string.
iii.	Studies that present spam detection frameworks.
iv.	Studies that present techniques used for spam detection.
v.	Studies that present case studies related to spam detection.
vi.	All the research articles published from 2007 to 2019.
vii.	Published literature as chapters of books, books, and technical/non-technical reports.
	Exclusion Criteria
i.	Title of research articles that were irrelevant or not related to spam detection.
ii.	Research articles that were not full research papers for example tutorials, conference abstracts, essays, thesis or presentation.
iii.	Research articles that were not in the English language.
iv.	Research articles that were without abstract.
v.	Research articles that did not present solutions explicitly.

Figure 3. Inclusion/Exclusion Criteria

D. SEARCH AND SELECTION RESULTS

The results of the search and selection of research papers are shown in Fig. 5. Initially, 6099 papers were retrieved when the designed search string was implemented to selected five scientific databases. The search was performed in three phases i.e., primary search, secondary search, and snowballing. In first-round papers were included or excluded on the basis of titles of retrieved research papers. Titles of all the research papers were examined which result in selection of 154 papers. The high number of papers was excluded because they were not relevant to the topic. For example, most of the research papers discuss the business perspective of spam review detection. Hence, those papers were not part of our study.

Moreover, we also studied papers that were relevant to spam detection other than spam reviews e.g. Spam emails e.tc. This review has been done from 2012 to 2019. This research has been done by primary search using scientific databases, conference procedures, journal papers, books and review papers. In the second round of inclusion or exclusion, primary search results were assessed by considering the titles, abstract, methodology and in the final round of inclusion or exclusion snowballing was used. Therefore, 63 research papers were selected for studies. The criteria for the selection of various research papers were reliant upon the define research questions. The nature of SLR relies on the determination of significant works. This SLR utilized the criteria defined in Fig. 3 for the selection of existing research studies. The search

criteria and the result of selections are presented in given below Fig. 4.

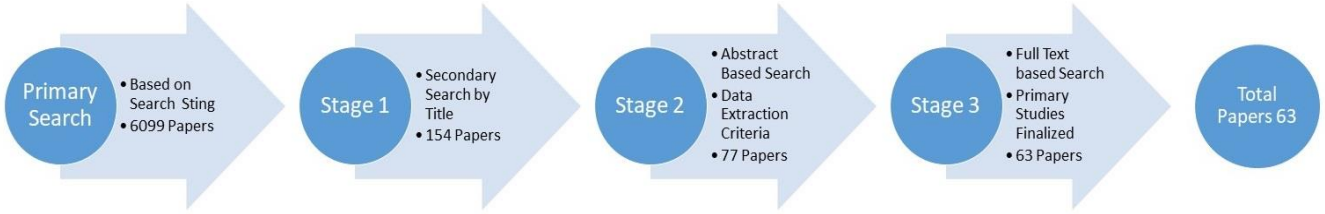

FIGURE 4. Selection Process

E. KEYWORDING USING ABSTRACT

In this stage of systematic literature review, after selecting relevant papers keywording was done on the basis of the abstract. We have used the process for keywording defined by Petersen [38]. This process has been completed in two phases. In the first phase, abstracts have been examined and identify concepts that reflect the contribution of studies. In the second phase, higher-level understanding has been developed on the basis of keywords. These keywords have been used to cluster categories.

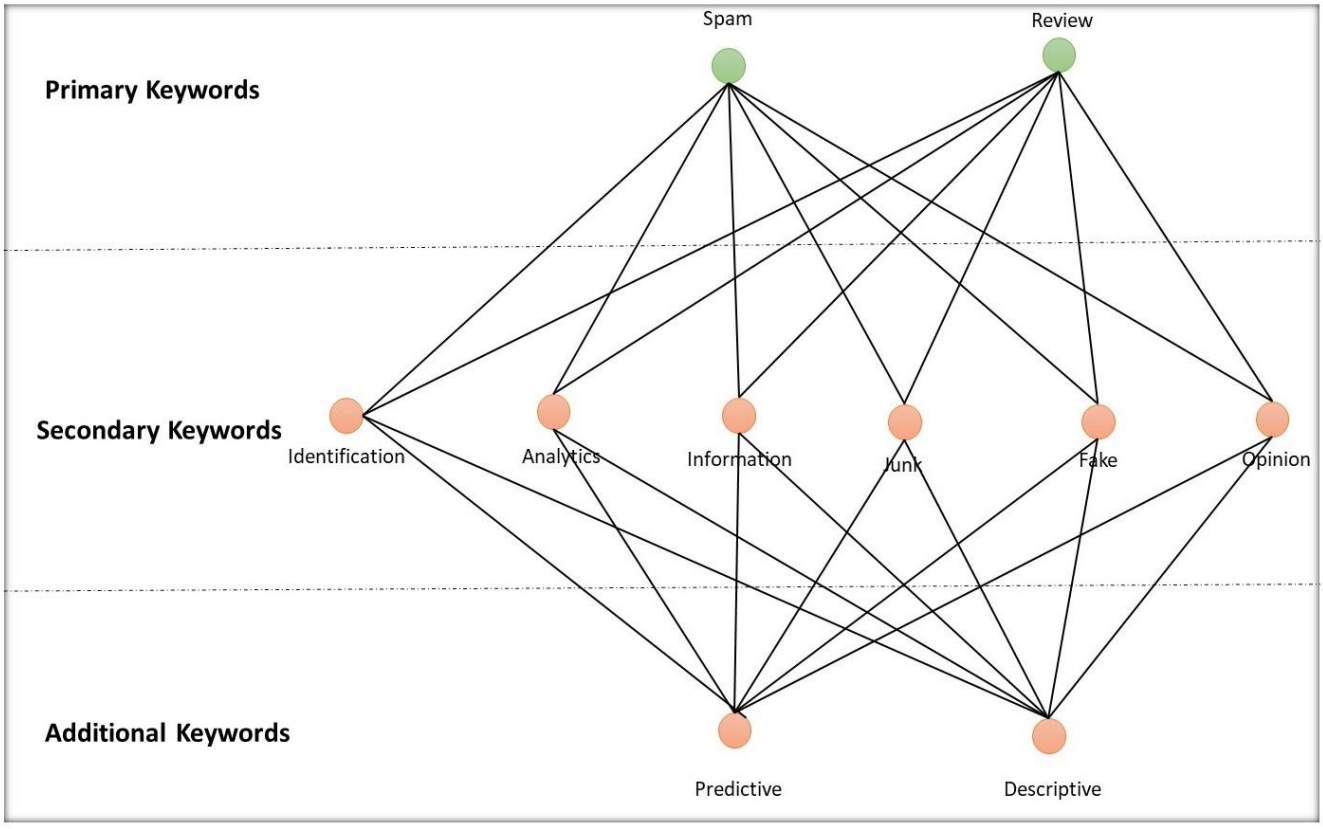

FIGURE 5. Search String Formulation

F. QUALITY ASSESSMENT CRITERIA

In SLR, quality assessment criteria are defined to assess the quality of selected papers after screening. A set of questionnaires has been designed for scoring and measuring the quality of papers selected for studies. Table VIII depicts possible answers to the criteria for ranking all the research papers selected for a systematic literature review.

TABLE III
CRITERIA FOR RANKING STUDIES

Criteria	Rank	Score
a. The study provides a clear contribution to spam detection	Yes	1
	No	0
b. The study documented the clear limitation of work while detecting	Yes	1
	No	0
c. The study was methodologically explained so that it can be trusted	Yes	1
	Partially	0

Criteria	Rank	Score
	No	0.5
d. The size of the selected dataset and the collection methods clearly mentioned	Yes	1
	No	0
e. Source ranking Journal/Conference or Symposium	Q1	2.5
	Q2	2
	Q3	1.5
	Others	1
	Core A	1.5
	Core B	1
	Core C	0.75
	IEEE and ACM identifies	0.25
	Others	0

1) PUBLICATION SOURCES

Table V depicts the details of sources, type of sources and their URL address that were used to access the selected studies for SLR.

TABLE V
INFORMATION SOURCES

Source	Type	URL
IEEE Xplore	Digital Library	https://ieeexplore.ieee.org/Xplore/home.jsp
Elsevier	Digital Library	https://www.elsevier.com
ACM Digital Library	Digital Library	https://dl.acm.org
Science Direct	Digital Library	https://www.sciencedirect.com
Springer	Digital Library	https://www.springer.com/gp

TABLE IV
SELECTION PROCESS OF RETRIEVED ARTICLES

Phases	Process Name	Criteria	ACM Digital Library	Springer	Science Direct	Elsevier	IEEE Xplore	Total
1	Search	Keywords	02	2354	1699	1929	115	6099
2	Screening (Phase 1)	Title and Duplication Removal	02	33	40	35	44	154
3	Screening (Phase 2)	Abstract	02	14	23	14	24	77
4	Inspection	Full Article	02	12	20	10	19	63

G. DATA EXTRACTION AND CLASSIFICATION SCHEME

The data extraction method has been implemented to get the possible answers to the research questions defined in Table I. Table VI was designed to extract data in order to collect the relevant information required to address the research questions defined in this systematic literature review. Data extraction id from DE1 to DE7 all collect the basic information related to the papers. These data extraction features include the title of research papers, name of the author's country, author name, publication type, source of the publications and name of the place where the publication was held. All other data from DE8 to DE11 were extracted after studying the paper.

1) PRIMARY SEARCH RESULTS

Primary search results were obtained from the five scientific databases ACM digital library, Elsevier, IEEE Xplore, Science Direct and Springer. Table IZV presents primary search results obtained after using the search string on scientific databases.

The papers selected for the primary study were extracted using the data extraction sheet. First, an Excel sheet was maintained, which consist of all the Metadata of research papers, books or articles obtained from the primary search results. Then, using the data extraction and classification scheme defined under the heading "Data Extraction and Classification Scheme" was used for a secondary search. In secondary search, we select all the papers by the title of the research paper that were best suited according to our selected topic. In the next round, we select the papers by reading the abstract and conclusion of the research papers that are obtained after the secondary search. We select 90 papers after implementing the abstract and conclusion-based search. Then we implemented a full text-based search on those 90 papers and selected 63 papers as final papers that were used to write this systematic literature review. Fig. 6 describes the number of publications and their year of publication of final results. The classification criteria were divided into categories, established with the help selected primary studies. Information for RQ1 is further categorized into the year in which the studies are produced, publication channels of the primary selected studies and the citation ratio. The Google scholar engine was used to obtain the citation count because it is an important feature that reflects the quality of the select research study. All of this information is collected directly from the respective study sources. RQ2 includes categories and techniques used for detecting spam review. RQ3 categories include information and features of publicly available datasets. From Table VIII it can be seen that most of the selected

research papers are published between 2015 to 2019. Fig. 6 shows the distribution of these article publications by year.

TABLE VI
DATA EXTRACTION FEATURES

Data Extract ID	Data Extract	Description
DE1	Study Identifier	DOI no
DE2	Title	Paper Title
DE3	Authors	Name of Authors
DE4	Publication Information	Name of the place where publication is held.
DE5	Country	Country of the Authors
DE6	Publication Type	Nature of the Publication (workshops, journals, conferences, etc.)
DE7	Source of the Publication	Industry/Academia
DE8	Abstract	Abstract of Papers
DE9	Objectives	Aim of Papers
DE10	Research Goals	The research goals and achievement that is defined in papers
E11	Discussion and Conclusion	Major findings of the research or study.

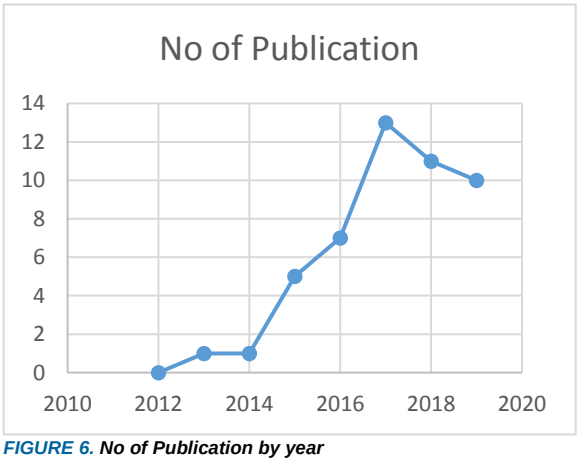

FIGURE 6. No of Publication by year

Fig. 7 describes the type of publication of the final selected papers. Publication type tells about the channel where the papers are published. In this systematic literature review, we have included different publication types which are conferences, book chapter, and journals

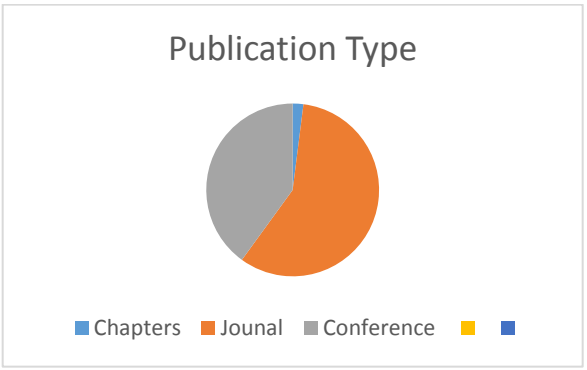

FIGURE 7. Publication Type

III. QUALITY ASSURANCE

When performing Systematic Literature Review quality assurance is the most influential part in order to select the literature of high quality, so that accurate and reliable analysis can be produced. Selection of inclusion/exclusion and well-defined keywords are the most important tasks for the planning phase of Systematic Literature Review. To accomplish this task following criteria is defined in order to validate the quality of research studies

In the final set, each publication was assessed for its quality. The quality assessment was performed during the data extraction phase and ensures that remaining included studies contribute a valuable study to the SLR. Hence, to fulfill this task questionnaires were designed. The questions were written in a way inspired by [4]. These criteria are labeled as a, b, c, d, and e. The scores of the quality criterion shows that journals are advantageous than conferences, workshops, and symposia because it is more difficult to publish papers in the journals of rank Q1 and Q2 than in conference, symposium or workshop. The final score of each paper is calculated by taking the sum of all five related questions. Table VIII depicts the evaluation results based on the criteria defined in Table III.

TABLE VIII
MAPPING AND CLASSIFICATION

Sr. No	Ref.no.	Year	Publication channel	Techniques used for spam detection	Scoring					
					(a)	(b)	(c)	(d)	(e)	Total
1	[1]	2019	Journal	Feature Oriented Framework	**1**	**1**	**1**	**1**	**2.5**	**6.5**
2	[2]	2018	Journal	Unified Model (Review Deviation	**1**	**1**	**1**	**1**	**2.5**	**6.5**
3	[4]	2019	Journal	KNN, Logistic, SVM, Random Forest , Gradient Boosting, MLP	**1**	**1**	**1**	**1**	**2.5**	**6.5**
4	[5]	2015	Journal	Rule based Classification	**1**	**0**	**1**	**1**	**2.5**	**5.5**
5	[7]	2018	Conference	Node2vec, Doc2vec	**1**	**0**	**1**	**0.5**	**2.5**	**5**
6	[6]	2013	Journal	Natural Language Processing	**1**	**1**	**1**	**1**	**2.5**	**6.5**
7	[10]	2019	Journal	Rule Based Classifiers, Classical Machine Leaning Classifier, Majority Voting Ensemble Classifier, Stacking Ensemble Classifiers	**1**	**0**	**0.5**	**1**	**2.5**	**5**
8	[18]	2019	Journal	Genetic Algorithm, Random Weight Network	**1**	**0**	**0.5**	**1**	**2.5**	**5**
9	[19]	2016	Journal	Conceptual Model	**1**	**0**	**1**	**1**	**2.5**	**5.5**
10	[26]	2015	Journal	PU Learning	**1**	**0**	**0.5**	**1**	**2.5**	**5**
11	[20]	2016	Journal	Time Series	**1**	**0**	**0.5**	**1**	**2**	**4.5**
12	[40]	2015	Journal	Binomial Regression, Classification	**1**	**1**	**1**	**1**	**2**	**6**
13	[21]	2017	Journal	Neural Network based Model	**1**	**1**	**0.5**	**1**	**2**	**4.5**
14	[22]	2018	Journal	Deep Neural Network	**1**	**0**	**0**	**0**	**1**	**2**
15	[13]	2018	Conference	N-Gram, Random Forest, SVM, Multinomial Naïve Bayes	**1**	**0**	**0**	**1**	**1.5**	**3.5**
16	[16]	2018	Journal	Neural Auto encoder Decision Forest	**1**	**0**	**0.5**	**0**	**0.25**	**1.75**
17	[41]	2019	Journal	Descriptive approach, Hybrid Content Mining, Data Mining and Learning Methods, Topological Measures	**1**	**0**	**0.5**	**1**	**0.25**	**2.75**
18	[25]	2017	Book	Markov Network	**1**	**0**	**1**	**0**	**1.5**	**3.5**
19	[17]	2017	Conference	Naïve Bayes, SVM, Decision Tree, Random Forest, Gradient Boosted Trees	**1**	**1**	**1**	**0**	**1.5**	**4.5**

Sr. No	Ref.no.	Year	Publication channel	Techniques used for spam detection	Scoring					
					(a)	(b)	(c)	(d)	(e)	Total
20	[9]	2018	Journal	Natural Language Processing, Multinomial Naïve Bayes, Stochastic Gradient Decent	**1**	**1**	**1**	**1**	**1.5**	**5.5**
21	[31]	2019	Conference	Sentiment Analysis	**1**	**1**	**1**	**1**	**2.5**	**6.5**
22	[30]	2017	Conference	Text mining, supervised technique, support vector machine, Naïve Bayes	**1**	**1**	**1**	**1**	**2**	**6**
23	[32]	2019	Conference	Sentiment Analysis, Classification	**1**	**0**	**1**	**1**	**1.5**	**4.5**
24	[33]	2018	Conference	Sentiment Analysis, Machine Learning Techniques	**1**	**1**	**1**	**1**	**2.5**	**6.5**
25	[8]	2017	Conference	Classification	**1**	**1**	**0.5**	**1**	**1.5**	**5**
26	[37]	2017	Conference	Adaptive Binary Flower Pollination 27Algorithm, Classification	**1**	**0**	**0.5**	**0**	**0.25**	**1.75**
27	[36]	2017	Conference	Cuckoo search, Harmony search	**1**	**0**	**0.5**	**1**	**0.25**	**1.75**
29	[23]	2016	Conference	Hybrid Learning Approach (Active Learning and Supervised Learning)	**1**	**0**	**1**	**0**	**1.5**	**4.5**
30	[42]	2016	Conference	Ensemble Techniques (Boosting, Bagging and Random Forest)	**1**	**1**	**1**	**0**	**1.5**	**4.5**
31	[43]	2016	Conference	Semantic Modeling and Emotion Modeling	**1**	**1**	**1**	**1**	**1.5**	**5.5**
32	[44]	2015	Conference	Three-View Semi Supervised Method	**1**	**1**	**1**	**1**	**2.5**	**6.5**
33	[11]	2017	Conference	Multinomial Naïve Bayes classifier, Bernoulli Naïve Bayes classifier, logistic regression classifier	**1**	**1**	**1**	**1**	**2**	**6**
34	[27]	2017	Conference	Fuzzy Logic	**1**	**0**	**1**	**1**	**1.5**	**4.5**
35	[45]	2017	Conference	Ensemble Classification	**1**	**1**	**1**	**1**	**2.5**	**6.5**
36	[12]	2018	Conference	Logistic Regression	**1**	**1**	**0.5**	**1**	**1.5**	**5**
37	[46]	2019	Conference	GSCPM	**1**	**1**	**1**	**1**	**2.5**	**6.5**
38	[47]	2017	Conference	Conversion Matching Algorithm, LCS Algorithm	**1**	**0**	**1**	**1**	**2.5**	**5.5**
39	[34]	2014	Conference	Sentiment Analysis, Classification	**1**	**0**	**1**	**0.5**	**2.5**	**5**

Sr. No	Ref.no.	Year	Publication channel	Techniques used for spam detection	Scoring					
					(a)	(b)	(c)	(d)	(e)	Total
40	[28]	2017	Journal	Heterogeneous Information Networks.	1	1	1	1	2.5	6.5
41	[48]	2018	Conference	Naïve Bayes	1	0	0.5	1	2.5	5
42	[49]	2019	Conference	SVM, Naïve Bayes	1	0	0.5	1	2.5	5
43	[35]	2015	Conference	Sentiment Analysis	1	1	1	0	1.5	4.5
44	[50]	2017	Conference	Naïve Bayes, Random Forest, AdaBoost	1	1	1	1	1.5	5.5
45	[51]	2016	Journal	Active Learning	1	1	1	1	2.5	6.5
46	[52]	2018	Conference	SVM	1	1	1	1	2	6
47	[53]	2016	Conference	Semi-supervised recursive auto encoders	1	0	1	1	1.5	4.5
48	[29]	2018	Conference	Map Reduce, Hadoop	1	1	1	0	1.5	4.5
49	[24]	2019	Conference	CNN, RNN, MLP, LSTM.	1	1	1	1	1.5	5.5

IV. ASSESSMENTS OF RESEARCH QUESTION

In this section, detailed answers to each research question are given which depends on analysis of the 63 selected primary research papers. Detail discussion of all three research question are extracted after analyzing and studying the selected research paper.

A. RQ1 Assessment: Which approaches have been utilized to distinguish spam contents or reviews?

Many techniques have been used to identify spam reviews and most of them are based on machine learning and rest uses semantic analysis. Mainly these techniques were classified in Unsupervised Learning (30%). Supervised Learning (20%), Active Learning (17%), Hybrid Learning (7%), Semantic Analysis (15%) and Data Mining (11%). The classification of these techniques is shown below in Fig. 7. Each of these techniques has throe own domain and limitations. Table IX depicts details of all the limitations, techniques and datasets used for detecting spam reviews.

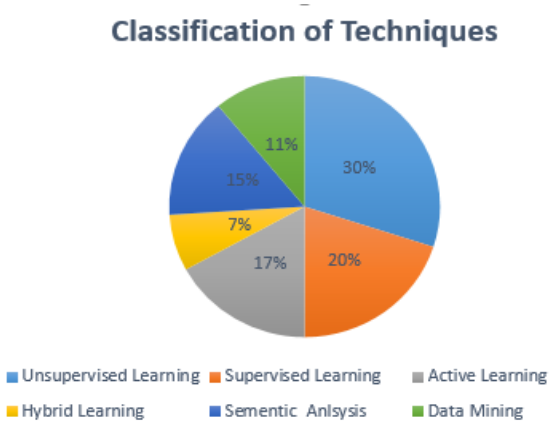

FIGURE 8. Classification of Techniques

TABLE IX
TECHNIQUES USED FOR SPAM REVIEW DETECTION.

Publication	Approach	Technique	Classifiers	Data Sets	Limitations
[1]	Semantic	Semantic Analysis	NLP	Online Electronic Retailer Review	
[40]	Supervised	Linguistic clues Cognition Indicators	Supervised machine learning algorithms; Random forest, SVM, Naive Bayes	Hotel Reviews	Suitable Only for labeled data sets - Not for Unlabeled dataset.

Publication	Approach	Technique	Classifiers	Data Sets	Limitations
[4]	Semi-supervised	Linguistic clues Cognition Indicators	k-NN, logistic, Support Vector Machines (SVM), Random Forest, Gradient Boosting, and Multilayer Perceptron (MLP).	Hotel Reviews	Suitable for labeled datasets.
[22]	Semi-supervised	Hybrid	Hybrid machine learning, Random weight network, Genetic Algorithms	Spam Assassin dataset	
[19]	Supervised	Classification	Conceptual Model		
[26]	Semi-supervised	Data Mining	PU-Learning	Amazon Mechanical Turk, Trip Advisor, Expedia, Hotels.com Reviews	Require Labeled datasets in the large amount
[20]	Data Frame	Regression	Time Series	Products Datasets	Require a limited number of datasets.
[40]	Supervised	Regression	Binomial regression, Classification	Amazon products reviews	
[54]	Multi-View Learning	Classification	KNN, Random Forest	Twitter Social Honeypot Dataset	
[17]	Supervised	Linguistic clues N-gram	Decision tree, Naïve Bayes, Random forest, SVM	Hotel Review	Limited Number of Features
[7]	Semi-Supervised		Decision tree, Random forest, SVM, K-nearest neighbor, Logistic regression, Naive Bayes	Hotel Reviews	require Labelled Data
[22]	Deep Learning	Neural Networks	Recurrent convolutional vector	TripAdvisor	
[55]	Supervised and Unsupervised Learning	Classification	Clustering	Iris Dataset	Performance relies on number of the initial neurons
[45]	Representation learning	Neural Networks	Convolutional neural network (CNN), Recurrent neural network (RNN)	Three Domain Datasets (Hotel, Restaurant, Doctor)	Unlabeled datasets
[41]	Supervised	Classification, Data Mining	Clustering, Pattern Mining	RepLab	
[56]	Supervised Learning	Classification	Feature reweighting		
[42]	Supervised Un-Supervised Learning	Sentiment Score, Linguistic feature and Unigram Reviewer data, Review Data, and Product information	SVM, Decision tree, Naive Bayes	Hotel Reviews, Amazon electronic product Reviews	Limited features
[26]	Semi-Supervised Learning	Unigram and Bigram	SVM, Naive Bayes	Hotel Reviews	Require Labelled Data

Publication	Approach	Technique	Classifiers	Data Sets	Limitations
[5]	Supervised Learning	Classification	Rule-Based Classification		Require labeled datasets
[7]	Supervised Learning	Regression	Logistic Regression, NodeVec, DocVec	Yelp	
[31]	Semantic	Big Data	Sentiment Analysis, NLP	"The Shape of Water" extracted from IMDB	
[32]	Semantic		Sentiment Analysis	Digikala.com	Require labeled datasets
[23]	Supervised Learning, Active Learning	Classification	Linear SVM classifier, Stochastic Gradient Descent classifier (SGD), Perceptron algorithm.	Crawled data from the websites (did not use publically available datasets)	
[43]	Semantic		Sentiment Analysis	www.dianping.com	Best Suitable labeled data
[11]	Semi-Supervised	Linguistic clues Cognition Indicators	Multinomial Naïve Bayes classifier, Bernoulli Naïve Bayes classifier, logistic regression classifier	Yelp.com	Unlabeled dataset
[50]		Artificial Intelligence	Fuzzy Logic	Amazon	A limited number of features

B. RQ2 Assessment: What are gaps between previously defined technology and application to detect spam reviews?

This research distinguished that there have been several open issues and gaps in defined techniques defined for detecting spam review. Significant gaps in research are explained below:

1) INACCESSIBILITY OF THE LABELED DATASETS

Availability of the labeled datasets is the main issue in the area of "spam review detection". Only one dataset is found about the hotel reviews was available publicly. However, this dataset has a limited number of features. Researchers need typically labeled datasets in order to design the classifiers for distinguishing proof of spam or legitimate reviews.

2) GROWTH IN DATASETS:

A large number of reviews exist on websites that are review based for example amazon.com, alibab.com, etc. The frequency of reviews and the reviewers are growing continuously and rapidly. Datasets with such frequency require high computation power.

3) LIMITATIONS OF DATA ATTRIBUTES

Review datasets that publicly available have limited features. Lack of dataset attributes limits the researcher to detect spam reviews more precisely and accurately.

4) DETECTING MULTILINGUAL SPAM REVIEW

Review is a user-created content and user can compose a review in any language of their decision. Up until this point, not many researchers have worked on the dataset other than English, for example, Arabic, Chinese, or Malay. There is a need to have top to bottom research on the detection of spam in multilingual reviews.

C. RQ3 Assessment: What information and features have been found in datasets of reviews?

This section discusses about publicly available datasets and their features used for spam detection. The extraction of the features from data is called feature extraction. Numerous considerations have utilized various methods for feature extraction to remove the most widely identified features or words in reviews.

1) REVIEW DATASETS:

Accessibility of a dataset is the beginning stage of any "spam review detection" research. The key issue in the "spam review detection" is the accessibility of the labeled dataset. It has been seen from existing research that just one labeled review dataset is available, however, it has just reviewed the content and availability of different features. "Amazon Mechanical Turk (AMT)" is likewise uses labeled datasets through online laborers (called Exhausts As indicated by Mukherjee et al. the way toward marking has not given improved precision to "spam review detection" on genuine datasets. Table X records review datasets utilized by various researchers and show absolute reviews, frequency of reviewers and frequency of items for each dataset. It is seen by the researcher that all review datasets are not openly available, and common researchers use crawlers to accumulate required data. It has additionally been seen that the greater part of the researchers utilized Amazon.com, internet business site datasets in their works, as it is the greatest web-based business stage to have item reviews, and the second biggest review dataset is available from booking.com and yelp.com, which is an online lodging booking site. Furthermore, the researchers working in the "spam review detection" utilize these datasets gave by such sites. In view of the existing researcher, it is seen that constrained true labeled datasets are available. Thus, there is a need to have freely available labeled standard review

datasets that might be utilized by the researchers for dissecting and learning the consequences of various "spam review detection systems. Table X presents the details of the datasets used for the spam review detection.

TABLE X
DATASETS INFORMATION

Data Sets	Domain	Instances	Description
SMS Spam Collection Dataset	Linguistics email and messaging languages	5573 Instances	This dataset contains a collection of SMS. The dataset is unlabeled.
Spam Text Message	linguistics email and messaging languages	5467 Instances	This dataset contains a collection of spam or legitimate SMS. The dataset is labeled.
Restaurant Customer Review	Restaurant reviews, food, and drink	1001 Instances	This dataset comprises information about who likes or dislike restaurant. Moreover, the dataset is labeled.
Amazon Unlocked Mobile	Telecommunication, Consumer Resources	413841 Instances	This dataset comprises mobile rating, prices, brand, reviews, and product name.
Spam mail	Email and messages	703 Instances	This dataset contains spam emails that were generated for spam research.
YouTube spam classified comment	Online media	1956 Instances	This is the public dataset of YouTube comments generated for spam research. It contains five datasets comprises of 1956 messages
Amazon Reviews	Business	3 Million Instances	This dataset consists of a few millions of reviews obtained from amazon customer.
Hotel Reviews	Business, Hotel	515739 Instance	This data set is extracted from the Booking.com which is publicly available.
Yelp	Business	Millions of instances	This Dataset was consisted of 7 CSV files.

V. DISCUSSION

In this section detail, the discussion has been presented about the techniques used for spam review detection. A taxonomy has been proposed to sum up the findings and results of the research.

A. PROPOSED TAXONOMY:

As best of our knowledge Fig. 9 presents the proposed taxonomy based on analysis of the selected studies. This work results in new taxonomy that may help researchers to categorize the techniques used for problems related to detection of spam review.

Spam Review Detection

Spam Review can be detected using following approaches:

Machine learning

Various studies have been done on techniques of machine learning. Most of these techniques of machine learning have been implemented in the area of the spam detection especially for detecting spam emails. There are two types of machine learning approaches that have been used span detection and these approaches have been classified in to different techniques. Following are the techniques that have been used for spam detection.

Supervised Learning

In supervised learning, machine is trained for prediction by using labeled data. In other words, we can say that some of the data is tagged with accurate answers. Following are the supervised learning techniques that have been used for spam detection.

- *Decision Tree Classifiers:* It is a systematic classification approach, which is used to build model of classification with the help of input dataset. It is specially used for problems which are complex to classify.
- *Regression:* It is a statistical approach used to find relationship between the variables or features of datasets.
- *Rule Based Classifiers:* It is used to refer classification scheme which uses IF-THEN rule for prediction. It consists of three components; Rule Induction Algorithm, Rule Ranking Measure and Class Prediction Algorithm.
- *Probabilistic Classifier:* These classifiers are designed by undertaking generative models. Following are the probabilistic classifiers that have been used for spam detection.

- *Bayesian Network:* It is the "probabilistic graphical model". Bayesian Network uses "Bayesian Interference" for computation done with the help of probability.
- *Naïve Bayes:* It is another type of the probabilistic graphical model. It is the most powerful algorithm used for predictive modeling. Naïve Bayes produces methods to estimate or predict the probability from the dataset.
- Maximum Entropy: This principle is used to select most unpredictable assumption where we have single parameter of probability distribution.

- *Linear Classifiers:* Algorithm of linear classifiers make classification on the basis of linear function for prediction.

Unsupervised Leaning

In unsupervised learning, machine is trained for prediction by using unlabeled data. Following unsupervised learning techniques have been used to detect spam detection.

- *K-means Clustering:* It is quite simple algorithm of supervised learning. It helps in solving problems related to clustering. Data set is classified in to clusters which is defined using letter "k".
- *Twice Clustering Technique:* It is another unsupervised learning technique that have been used for spam detection.
- Neural Networks: We can define the neural network as computing system. These computing systems consist of interconnected nodes and nodes are known as neurons. Information is processed using these nodes that are organized in form of layers. Following are the techniques of neural networks that have been used for spam detection.
  - *Auto Encoders:* It is another technique of unsupervised learning used to encode or compress the data and reconstruct data using the "reduced encoded representation.
  - *Multilayer Perceptron Model:* It is a computational graph. It has some layers. First layer is called input layer that usually contain features that we have as input. Second layer is called hidden layer. There may be one or more than one hidden layers. Third layer is output layers that contain prediction.

Lexicon based learning

It is also used to extract sentiments from the text or document and predicting using sentiment analysis. Following are the lexicon learning approaches that have been used to detect spam reviews.

- Dictionary based methods
- Corpus based methods

B. OPEN ISSUES AND CHALLENGES

Some open issues and challenges have been identified from the literature which is given below:

- Mainly the open issue and challenges are related to the unavailability of labeled datasets. There is only one labeled dataset of hotel reviews, but it has limited features or attributes.
- All the datasets which are publicly available have a limited number of attributes. This results in a lack of accuracy as more attributes are required to improve the accuracy of the implemented models or algorithm.
- With the passage of time review datasets are growing rapidly, which in return requires higher computing power. Sentiment analysis will become more challenging in the domain of spam review detection with growth in datasets.
- The feedback of the reviewer's review is not evaluated for spam detection. For example, some websites ask "Did you find this review helpful?" Such comments of other reviews are not considered for spam review detection.

VI. CONCLUSION

This study exhibited a systematic literature review of the spam review detection area and featured late research commitments as various component designing methodologies, spam review detection techniques, and various measures utilized for quality assessment. To separate valid shreds of evidence, this work uses review technique, concentrated on the search string, brought up investigate issues, chose papers from scientific databased and uses the inclusion/exclusion criteria. A sum of 6099 papers distributed from 2012 to December was chosen dependent on the search string, and in the wake of applying search based on title, 154 papers were selected. At last, by reading full length and snowballing, 63 research papers were concluded for additional study. Moreover, the criteria of quality assessment were defined to decide the importance and legitimacy of the research area fitting to selected publications.

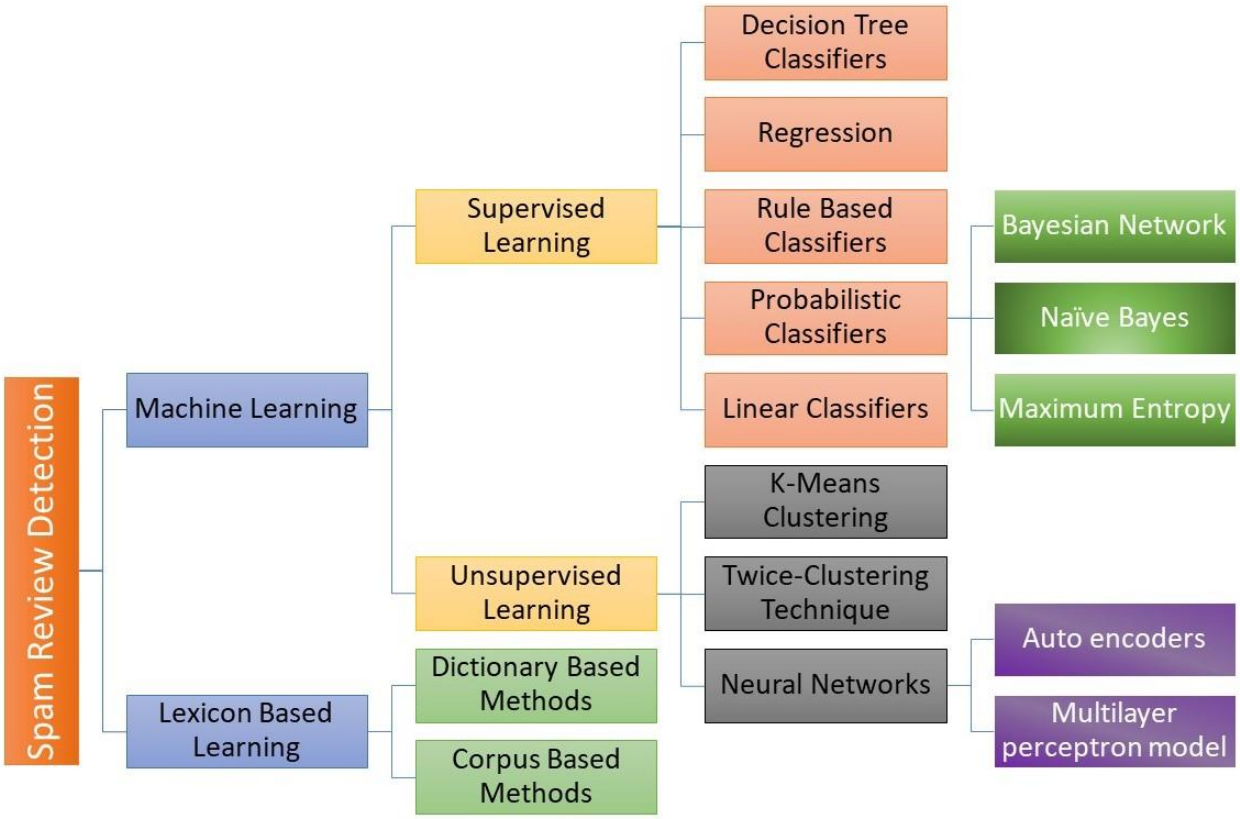

FIGURE 9. **Spam Review Detection Taxonomy**

REFERENCES:

[1] R. Barbado, O. Araque, and C. A. Iglesias, "A framework for fake review detection in online consumer electronics retailers," *Information Processing & Management*, vol. 56, no. 4, pp. 1234-1244, 2019.

[2] Y. Liu and B. Pang, "A unified framework for detecting author spamicity by modeling review deviation," *Expert Systems with Applications*, vol. 112, pp. 148-155, 2018.

[3] M. Luca, "Reviews, reputation, and revenue: The case of Yelp. com," *Harvard Business School NOM Unit Working Paper*, pp. 12-16, 15 March 2016 2016

[4] M. R. Martinez-Torres and S. L. Toral, "A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation," *Tourism Management*, vol. 75, pp. 393-403, 2019.

[5] D. T. Tanya Gera , Jaiteg Singh "Identifying Deceptive Reviews Using Networking.pdf," *InternationalConference on Computing and Communications Technologies* 2015

[6] K.-I. L. Kuldeep Sharma, "Review Spam Detector with Rating Consistency Check," *ACM*, 2013.

[7] A. O. D. Cennet Merve Yılmaz " SPR2EP A Semi-Supervised Spam Review Detection.pdf," *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* August 2018 2018.

[8] D. K. U. SP.Rajamohana, M.Dharani, R.Vedackshya "A SURVEY ON ONLINE REVIEW SPAM DETECTION TECHNIQUES " *IEEE International Conference on Innovations in Green Energy and Healthcare Technologies(ICIGEHT'17)* 2017.

[9] E. F. Cardoso, R. M. Silva, and T. A. Almeida, "Towards automatic filtering of fake reviews," *Neurocomputing*, vol. 309, pp. 106-116, 2018.

[10] R. M. K. Saeed, S. Rady, and T. F. Gharib, "An ensemble approach for spam detection in Arabic opinion texts," *Journal of King Saud University - Computer and Information Sciences*, 2019.

[11] Anna V. Sandifer , Casey Wilson, and A. Olmsted, "Detection of fake online hotel reviews," *The 12th International Conference for Internet Technology and Secured Transactions (ICITST-2017)*, 2017.

[12] Shaohua Jia and X. W. Xianguo Zhang, Yang Liu "Fake Reviews Detection Based on LDA " 2018.

[13] N. S. Shreyas Aiyar, "N-Gram Assisted Youtube Spam Comment Detection," *International*

Conference on Computational Intelligence and Data Science pp. 174–182, 2018.

[14] G. N. Wael Etaiwi, "The Impact of applying Different Preprocessing Steps on Review Spam Detection," *The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017)*, pp. 273–279, 2017.

[15] Y. Z. Faisal Khurshid, Chubato Wondaferaw Yohannese, Muhammad Iqbal, "Recital of Supervised Learning on Review Spam Detection An Empirical Analysis," 2017.

[16] M. Dong, L. Yao, X. Wang, B. Benatallah, C. Huang, and X. Ning, "Opinion fraud detection via neural autoencoder decision forest," *Pattern Recognition Letters*, 2018.

[17] G. N. Wael Etaiwi, "The Impact of applying Different Preprocessing Steps on Review Spam Detection," pp. 273–279, 2017.

[18] H. Faris *et al.*, "An intelligent system for spam detection and identification of the most relevant features based on evolutionary Random Weight Networks," *Information Fusion*, vol. 48, pp. 67-83, 2019.

[19] A. Munzel, "Assisting consumers in detecting fake reviews: The role of identity information disclosure and consensus," *Journal of Retailing and Consumer Services*, vol. 32, pp. 96-108, 2016.

[20] A. Heydari, M. Tavakoli, and N. Salim, "Detection of fake opinions using time series," *Expert Systems with Applications*, vol. 58, pp. 83-92, 2016.

[21] L. Li, B. Qin, W. Ren, and T. Liu, "Document representation and feature combination for deceptive spam review detection," *Neurocomputing*, vol. 254, pp. 33-41, 2017.

[22] W. Zhang, Y. Du, T. Yoshida, and Q. Wang, "DRI-RCNN: An approach to deceptive review identification using recurrent convolutional neural network," *Information Processing & Management*, vol. 54, no. 4, pp. 576-592, 2018.

[23] T. N. M.N. Istiaq Ahsan*, Abdullah All Kafi£, Md. Ismail Hossain‡, Faisal Muhammad Shahμ "An Ensemble approach to detect Review Spam using hybrid Machine Learning Technique " *19th International Conference on Computer and Information Technology, December 18-20, 2016*.

[24] G. M. Shahariar, Swapnil Biswas, Faiza Omar, Faisal Muhammad Shah, and S. B. Hassan, "Spam Review Detection Using Deep Learning," 2019.

[25] G. Fei, H. Li, and B. Liu, "Opinion Spam Detection in Social Networks," pp. 141-156, 2017.

[26] D. Hernández Fusilier, M. Montes-y-Gómez, P. Rosso, and R. Guzmán Cabrera, "Detecting positive and negative deceptive opinions using PU-learning," *Information Processing & Management*, vol. 51, no. 4, pp. 433-443, 2015.

[27] Harshita Kotian and D. B. B. Meshram, "Detection of Spam Reviews and Spammers in E- Commerce Sites " *International Conference on Current Trends in Computer, Electrical, Electronics and Communication (ICCTCEEC-2017)*, 2017.

[28] S. Shehnepoor, M. Salehi, R. Farahbakhsh, and N. Crespi, "NetSpam: A Network-Based Spam Detection Framework for Reviews in Online Social Media," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 7, pp. 1585-1595, 2017.

[29] D. V. R. Lekshmi M B, "Spam Detection Framework for Online Reviews Using Hadoop's Computational Capability," 2018.

[30] N. J. Chirag Visani , Manali Modi, "A Study on Different Machine Learning Techniques for Spam Review Detection " *International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017)*, pp. 676-679, 2017.

[31] C. Yao, J. Wang, and E. Kodama, "A Spam Review Detection Method by Verifying Consistency among Multiple Review Sites," pp. 2825-2830, 2019.

[32] N. S. Mohammad Ehsan Basiri , Hadi Khosravi Farsani "A Supervised Framework for Review Spam Detection in the Persian Language " *2019 5th International Conference on Web Research (ICWR)* pp. 203-207, 2019.

[33] P. R. P. Nidhi A. Patel "A Survey on Fake Review Detection using Machine Learning Techniques " *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, 2018.

[34] H. A. N. a. X. Zhu, "iSRD: Spam Review Detection with Imbalanced Data Distributions " vol. IEEE IRI 2014, 2014.

[35] J. G. B. Siddu P. Algur "Rating Consistency and Review Content based Multiple Stores Review Spam Detection," 2015.

[36] K. U. S.P.Rajamohana, S.Vasantha Keerthana "An Effective Hybrid Cuckoo Search with Harmony Search for Review Spam Detection," *3rd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics* 2017.

[37] D. K. U. SP.Rajamohana, B.Abirami "ADAPTIVE BINARY FLOWER POLLINATION ALGORITHM FOR FEATURE SELECTION IN REVIEW SPAM DETECTION " *IEEE International Conference on Innovations in Green Energy and Healthcare Technologies(ICIGEHT'17),* 2017.

[38] R. F. Kai Petersen, Shahid Mujtaba, Michael Mattsson, "Systematic Mapping Studies in Software Engineering," 2008.

[39] K. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," July 2007 2007.

[40] D. Savage, X. Zhang, X. Yu, P. Chou, and Q. Wang, "Detection of opinion spam based on anomalous rating deviation," *Expert Systems with Applications*, vol. 42, no. 22, pp. 8650-8657, 2015.

[41] S. M. H. Bamakan, I. Nurgaliev, and Q. Qu, "Opinion leader detection: A methodological review," *Expert Systems with Applications*, vol. 115, pp. 200-222, 2019.

[42] T. M. K. Brian Heredia, Joseph Prusa and Michael Crawford, "An Investigation of Ensemble Techniques for Detection of Spam Reviews," *2016 15th IEEE International Conference on Machine Learning and Applications*, pp. 127-133, 2016.

[43] Y. Li, X. Feng, and S. Zhang, "Detecting Fake Reviews Utilizing Semantic and Emotion Model," *2016 3rd International Conference on Information Science and Control Engineering*, pp. 317-320, 2016.

[44] D.-K. K. Ji Chengzhang, "Detecting the Spam Review Using Tri-training " *ICACT*, 2015.

[45] M. M. S. Alhassan J. Ibrahim, Mazura Mat Din "Ensemble Classifiers for Spam Review Detection " *2017 IEEE Conference on Application, Information and Network Security (AINS)*, 2017.

[46] M. H. Guangxia Xu, Chuang Ma, and M. Daneshmand, "GSCPM: CPM-based group spamming detection in online product reviews," 2019.

[47] P. Liu, Z. Xu, J. Ai, and F. Wang, "Identifying Indicators of Fake Reviews Based on Spammer's Behavior Features," *2017 IEEE International Conference on Software Quality, Reliability and Security* pp. 396-403, 2017.

[48] Alok Katiyar , Rishi Kumar , D. S. K. ., and D. Y. D. S. Arya, "Pervasive Online Spam Review Detection based on Ontology using Naive Bayesian " *International Conference on Advances in Computing, Communication Control and Networking (ICACCCN2018)* 2018.

[49] A.-j. Li and L. Shi, "Product Spam Reviews Detection Based on Index Optimization," *2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, pp. 266-269, 2019.

[50] Y. Z. Faisal Khurshid, Chubato Wondaferaw Yohannese, Muhammad Iqbal "Recital of Supervised Learning on Review Spam Detection: An Empirical Analysis " 2017.

[51] T. N. M.N. Istiaq Ahsan, Abdullah All Kafi, Md. Ismail Hossain‡, Faisal Muhammad Shah, "Review Spam Detection using Active Learning " 2016.

[52] B. K. Draško Radovanović, "Review Spam Detection using Machine Learning," 2018.

[53] J. H. Baohua Wang, HaihongZheng, Hui Wu, "Semi-Supervised Recursive Autoencoders for Social Review Spam Detection," 2016.

[54] H. Shen, F. Ma, X. Zhang, L. Zong, X. Liu, and W. Liang, "Discovering social spammers from multiple views," *Neurocomputing*, vol. 225, pp. 49-57, 2017.

[55] J. Z. Lei and A. A. Ghorbani, "Improved competitive learning neural networks for network intrusion and fraud detection," *Neurocomputing*, vol. 75, no. 1, pp. 135-145, 2012.

[56] T. C. Alberto, J. V. Lochter, and T. A. Almeida, "Post or Block? Advances in Automatically Filtering Undesired Comments," *Journal of Intelligent & Robotic Systems*, vol. 80, no. S1, pp. 245-259, 2014.

[57] Abid, A., Hussain, N., Abid, K., Ahmad, F., Farooq, M.S., Farooq, U., Khan, S.A., Khan, Y.D., Naeem, M.A. and Sabir, N., 2016. A survey on search results diversification techniques. *Neural Computing and Applications*, *27*(5), pp.1207-1229.

[58] Abid, A., Ali, W., Farooq, M. S., Farooq, U., Sabir, N., & Abid, K. (2020). Semi-Automatic Classification and Duplicate Detection from Human Loss News Corpus. *IEEE Access*.