

# A SURVEY BASED ON REVIEW SPAM DETECTION AND ITS TECHNIQUES

**Irtiqia Amin**

*Research Scholar*

*School of Computer Sciences and Engineering*

*Lovely Professional University*

*Phagwara Punjab India 144411*

**Dr. Mithilesh Kumar Dubey**

*Associate Professor*

*School of Computer Sciences and Engineering*

*Lovely Professional University*

*Phagwara Punjab India 144411*

## **ABSTRACT**

*Reviews that are displayed online is a key position in customers choice to buy an item or facilities. They are vital and worthy source of information that can be used to evaluate public suggestions / points of view regarding items or services. Due to its influence, producers and distributors are deeply worried with its inputs or reviews / opinions from customers. Such opinions are based solely on either consumer or reviewer's feelings or perceptions, thereby giving rise to potential concerns that wrongdoer may generate bogus reviews to unfairly support or bring down the reputation of an item or services. As false reviews become more common on the web and can be an issue to it. These fake reviews are called as spams. A method for distinguishing between truthful and untruthful reviews is therefore necessary. This article deals with highly machine learning techniques that are put forward to address problems of spams and evaluate the performance of various approaches used for classifying reviews and analyzing their detections. The key objective of this article is to offer a detailed modified study of present research to detect review spam and also to suggest methodologies for future inspection.*

**Keywords:** Review Spam Detection, Opinion Mining, Big Data, Machine learning.

## **1: Introduction**

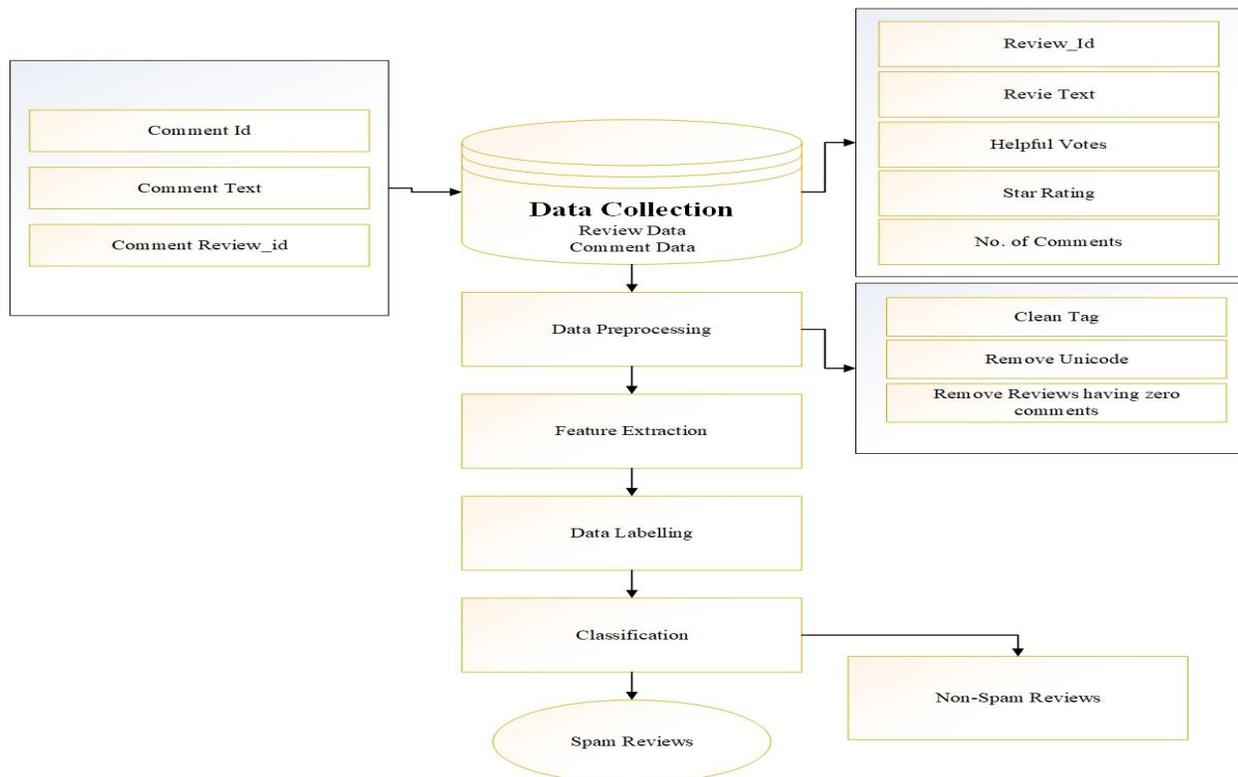
While web services continue to increase in size and value, online reviews are steadily increasing in number and effect. People are relying on customer reviews and recommendations that show how to purchase things, try new restaurants, and locate experts and other goods and services. Web reviews may be formed for a diversity of reasons. Frequently the efforts to get better and upgrade their business, web retailers and facility providers may ask their customers to provide response and share their experiences on those items and services they have already purchased and whether they were happy or not. If these customer's particularly have good or bad experience with the item they may feel free to write about it as

the internet is available to anyone. Whereas web reviews might be useful, sightless belief in them is hazardous equally to advertisers and purchaser. Most people go through these web reviews before purchasing a particular product, therefore the reviews may be tainted for benefit, and the decisions must be taken with caution prior to the purchasing of such items. The reviews are considered as phony or spam as it can have major effect on the web shopping place due to the significance of these review.

“Spam comments can be grouped into three categories: a: Untruthful or Fake reviews, b: Reviews on brands, c: Non reviews [1].” False reviews are those words which are formed with mean intention of a person. Brand reviews aren't specific to products; instead they are related to professionally naming the product or services. Other reviews contain all words and saying except describing a product. Detection of untruthful reviews is a challenging task [2][3]. These reviews are created by doubtful person called as “spammers”, just to harm or boost the status of any associations. As these reviews are given by the unknown person it becomes very difficult to determine whether the product belongs to a reputable source or not. “Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed and focuses on the development of computer programs that can access data and use it learn for them”. “Machine learning is categorized into three methods: a: Supervised learning- used for labeled datasets, b: Unsupervised learning- used for unlabelled datasets, c: Semi supervised learning- used for both.” Supervised machine learning algorithms are trained in labeled data sets and are used to identify reviews as true and false based on defining the features of reviews. Such labeled datasets include the collection of instances (reviews), the characteristics of a specific review and their class labels. Describing features are categorized as: “a: Review centric, b: Reviewer centric, c: Product centric”. The features explaining the review usually include text review and details contained in it e.g. rating, date, time, ID. Reviewer oriented features include his or her description produced in the review e.g. user-ID, date, time, IP address, name, email-ID.

It is known as high dimensionality to have a broad feature area, creating challenges such as high computing costs and worthless features [4]. To counter high dimensionality, machine learning technique known as feature selection may be implemented [5][6]. In addition to the set of features, one of the techniques used is called ensemble techniques were used to detect the performance of the classifier and minimize over fitting [7]. Data mining and machine learning strategies for web and text mining lead excitingly to identifying fake reviews [8], defined “web mining as the process for finding useful information and relations from the content available on the web by largely relying on the available machine learning techniques and methods”. “Web mining can be divided into three groups: a: Structure b: Content c: Usage mining. Content mining deals with the knowledge and information extraction and categorization them using machine learning and data mining methods. The true example of content mining is opinion mining”. Opinion mining is used for providing the sentimental analysis of a text message by analyzing the feature of that review like content feature and natural language processing (NLP), text features

(bag of words, POS tagging, reviewer centric features) .“A classifier can be skilled to categorize latest occurrence by inspecting the text features related with different suggestions next to with their sentiments”. The primary issue found by [10] is the absence of any distinguished terms (features) capable of providing a definite clue for classifying reviews as fake or real. They used “NLP” and “bag of words” for individual word and small groups. The extraction of features was considered to be the future work [11][12][13 ], using individual words from the analysis text as features as language features (NLP and word bag)[14 ] used in syntactic and lexical terms. An additional study by [15] on datasets for the detection of false and real examples created by him. His work was related on the review characteristics features, unigram features frequencies and reviewer classification and the reviewer behavior features. The reviewer behavioral studies can never be carried out from the content of a solo review. Reviewer (spammer) behavioral analysis include finding several client IDs for the identical author[16] and recognizing them by observing their signatures and irregular behavioral characteristics of the review performed better than LIWC (Linguistic Inquiry Word Count) and POS (Part of Speech) [17][18][19]. An alternative graph theory based on relationship between review and reviewer and results showed are quite promising [20][21]. To address the questions related with upgrading the analysis of spam detection, data collection must be taken into account. Data is a most important element of a machine-based learning model, although a large quantity of data is available on the internet, it’s a difficult task to collect and classify an adequate amount of data to teach a spam classifier algorithm. The different approach given by [26] is to unnaturally construct review spam datasets by using “synthetic review spamming method”, which takes previous genuine reviews and generate bogus reviews from them. The work done by them was based on a: Feature selection or Feature engineering (performance measurements) and comparative analysis (machine learning methods and results evaluated by them). The experimentation is done using a: Real world unsupervised and semi supervised methods, b: Data sample addressing, c: Data quality to check the presence of noise due to mislabeled class instances, d: Ensemble techniques (performance of algorithm) for detection of review spam. Future work will be based on live streaming review spam detection the enhanced version of previous method. The model for review spam detection is shown in Fig 1:



## 2. Methodology

Review spam detection is divided into three major classification based on feature selection: a: Review centric, b: Reviewer centric, c: Product centric. Centric features of the reviewers are those apps that construct using the information contained in the analysis and take comprehensive look at the writers along with the information about him. Review centric apps are based on software engineering or extraction of features from documents. This survey based article deals with analyzing and addressing some of the widely used features in the spam detection area of study and the datasets used by the authors in their analysis.

### 2.1. Research Gap:

As outlined in the introduction, several different types of features were used in the previous study, most often in the text of the analysis, and is normally done via “bag of words” approach wherever all review features consists of either single words or group of words contained in it. Using the combination of features to train a classifier had shown the high performance than any feature used solely as demonstrated by [20][21][31][33][24][9]. [25] Stated that using the combination of features like (LIWC and POS) in addition with bag of words are more strongly accepted as that of the features used solely in “bag of words”. A study by [23] showed that the feature used as abnormal behavior of the reviewer performed better as that of the linguistic feature of review themselves.

## Features categorization

The review based features are divided into various categories: a: Bag of words b: LIWC c: POS d: Semantic and Syntactical features. “Text categorization uses bag of words to extract text features or attributes from the individual or small groups of reviews”. These attributes are known as “n-grams”, picks one or two continuous terms from the text and are denoted respectively as unigrams (n=1), bigrams (n=2), trigrams (n=3), used by [21][9][22]. Term frequencies are related to the “bag of words” as they are concerned about the occurrences of the reviews. “LIWC (Linguistic Inquiry word Count) is the software tool in which users can build their own languages and analyze the dimensionality of the their language and POS (Part of Speech) is used for tagging word features with part of speech based on definition and its context within the review in which its used”. [27] Used lexical or syntactical styles as features for review. [28] worked on labeled datasets as small no of datasets was used as a training datasets, review centric features or content features, sentimental features, product features, reviewer centric features.[29] proposed a graph based model to detect review spam detection with three different nodes. Iterative algorithm was used to detect individual spammers as detected by positive or negative and loopy belief propagation (LBP) was used to detect signed networks. KDE (kernel density estimation) technique was used by [30] to sense the burst pattern in review for a particular product. Similarly Markov Random fields (MRF) model was used to detect the reviewer burst with same LBP framework to detect review spam with 83.7% precision. The major datasets used by most of the authors are AMT (Amazon Mechanical Turk), Trip Advisor. [31] Showed the concept of the psycholinguistic deception of reviews.

## 3. Conclusion

This study reviewed the basic techniques of the machine learning algorithm used for reviewing spam detection, as well as the review development and reviewer features that have been released to date. A serious issue in the detection of review spam is the pool of standard datasets, as most scientists have used spam reviews created by them with its characteristics. Hence the precision of the given procedures cannot be accurately measured. Thus the selection of these data sets i.e. to be used in the spam analysis is still an unsettled issue in this field. We found that the most effective strategy for the identification of spam is to catch burst pattern for feedback. Future work can be done by detecting live streaming review spams for different social websites, and the psycholinguistic specifications of review related with reviewer.

## References

[1] Dixit S, Agrawal AJ. Survey on review spam discovery. Int J Comput Commun Technol ISSN (PRINT) 4:0975–7449,2013.

- [2] Ott M, Choi Y, Cardie C, Hancock JT. Finding deceptive opinion spam by any stretch of the imagination. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies Association for Computational Linguistics.-V1:309–319,2011.
- [3] Jindal N, Lui B. Opinion spam and analysis. In: Proceedings of the 2008 international conference on web search and data mining,2008.
- [5] Crawford M, Khoshgoftaar TM, Prusa JD. Reducing feature set explosion to facilitate real-world review spam detection. In: Proceedings of the 29th international FLAIRS conference, 304–309,2011.
- [4] Haykin S. Neural networks: a comprehensive foundation, 2nd edn. Prentice Hall, Upper Saddle River,1998.
- [6] Mukherjee A, Venkataraman V, Liu B, Glance N. What yelp fake review filter might be doing? In: Seventh international AAAI conference on weblogs and social media, 2013.
- [7] Dietterich TG. Ensemble methods in machine learning. In: International workshop on Multiple classifier systems, 1–15, 2000.
- [8] Jindal N, Lui B. Opinion spam and analysis. In: Proceedings of the 2008 international conference on web search and data mining, 2008.
- [9] Bandakkanavar RV, Ramesh M, Geeta H. A survey on detection of reviews using sentiment classification of methods. IJRITCC 2(2):310–314,2014.
- [10] Abbasi A, Zhang Z, Zimbra D, Chen H, Nunamaker JF Jr. Detecting fake websites: the contribution of statistical learning theory. MIS Q 34(3):435–461,2010.
- [11] Jindal N, Liu B, Lim EP. Finding unusual review patterns using unexpected rules. In: Proceedings of the 19th ACM international conference on Information and knowledge management, ACM, Toronto, ON, Canada. 1549–1552, 2010..
- [12] Li F, Huang M, Yang Y, Zhu X. Learning to identify review spam. In: IJCAI Proceedings-International Joint Conference on Artificial Intelligence, V(22), 3:2488, 2010.
- [13] Mukherjee A, Liu B, Glance N. Spotting fake reviewer groups in consumer reviews. In: Proceedings of the 21st international conference on World Wide Web, ACM, Lyon, France. 191–200, 2012.
- [14] Shojaee S, Murad MAA, Bin Azman A, Sharef NM, Nadali S. Detecting deceptive reviews using lexical and syntactic features. In: Intelligent Systems Design and Applications (ISDA), 2013 13th International Conference IEEE, Serdang, Malaysia 53–58, 2013.

- [15] Ott M, Cardie C, Hancock JT .Negative Deceptive Opinion Spam. In: HLT-NAACL., 497–501, 2013  
jjjjjjjjjj
- [16] Qian T, Liu B. Identifying Multiple Userids of the Same Author. In: EMNLP., 1124-1135, 2013.
- [17] 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, ACM 632–640, 2013.
- [18] Feng S, Xing L, Gogar A, Choi Y. Distributional footprints of deceptive product reviews. ICWSM 12:98–105, 2012.
- [19] Xie S, Wang G, Lin S, Yu PS. Review spam detection via temporal pattern discovery. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, Beijing, China 823–831, 2012.
- [20] Jindal N, Liu B. Review spam detection. In: Proceedings of the 16th international conference on World Wide Web, ACM, Lyon, France pp. 1189–1190, 2007.
- [21] Jindal N, Liu B. Opinion spam and analysis. In: Proceedings of the 2008 International Conference on Web Search and Data Mining, ACM, Stanford, CA 219–230, 2008.
- [22] Fei G, Mukherjee A, Liu B, Hsu M, Castellanos M, Ghosh R. Exploiting Burstiness in reviews for review spammer detection. ICWSM 13:175–184, 2013.
- [23] Mukherjee A, Venkataraman V, Liu B, Glance NS. What yelp fake review filter might be doing? Boston, In ICWSM, 2013.
- [24] Hammad ASA. An Approach for Detecting Spam in Arabic Opinion Reviews. Doctoral dissertation, Islamic University of Gaza, 2013.
- [25] Li J, Ott M, Cardie C, Hovy E. Towards a general rule for identifying deceptive opinion spam. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, Maryland, USA, June 23-25 2014. ACL 1566–1576, 2014.
- [26] Morales A, Sun H, Yan X. Synthetic review spamming and defense. In: Proceedings of the 22nd international conference on World Wide Web companion. International World Wide Web Conferences Steering Committee, Rio de Janeiro, Brazil, 155–156, 2013 .

- [27] Shojaee S, Murad MAA, Bin Azman A, Sharef NM, Nadali S. Detecting deceptive reviews using lexical and syntactic features. In: Intelligent Systems Design and Applications (ISDA), 2013 13th International Conference on IEEE, Serdang, Malaysia 53–58, 2013.
- [28] Ahmed, I., Ali, R., Guan, D., Lee, Y. K., Lee, S., & Chung, T. Semi-supervised 838 learning using frequent itemset and ensemble learning for SMS classification. 839 Expert Systems with Applications, 42(3), 1065–1073, 2015.
- [29] Wang, Guan, et al. "Review graph based online store review spammer detection." 2011 IEEE 11th International Conference on Data Mining. IEEE, 2011.
- [30] Fei et al. Exploiting burstiness in reviews for review spammer detection. In 849 Seventh international AAI conference on weblogs and social media, 2013.
- [31] Pennebaker et al. The development and psychometric properties of 908 LIWC2007. Austin, TX, LIWC.Net,2007.
- [32] Wang G, Xie S, Liu B, Yu PS. Identify online store review spammers via social review graph. ACM Transactions on Intelligent Systems and Technology (TIST) 3(4):61, 2012.
- [33] Wang G, Xie S, Liu B, Yu PS. Review graph based online store review spammer detection. In: Data mining (icdm), 2011 iee 11th international conference on Vancouver, Canada, . IEEE,.1242–1247, 2011.