EasyChair Preprint

№ 7224

Review of Fake Product Review Detection Techniques

B V Santhosh Krishna, Sanjeev Sharma, Y Sahana, K Devika, K N Sharanya and K Indraja

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

December 17, 2021

Review of Fake Product Review Detection Techniques

Santhosh Krishna B V
Department of E&C Engineering
New Horizon College of Engineering
Bengaluru, India
santhoshkrishna1987@gmail.com

Dr. Sanjeev Sharma
Department of E&C Engineering
New Horizon College of Engineering
Bengaluru, India

Sahana Y
Department of E&C Engineering
New Horizon College of Engineering
Bengaluru, India
sahana.yelchuri@gmail.com

Devika K
Department of E&C Engineering
New Horizon College of Engineering
Bengaluru, India

Sharanya K N
Department of E&C Engineering
New Horizon College of Engineering
Bengaluru, India

Indraja K
Department of E&C Engineering
New Horizon College of Engineering
Bengaluru, India

***Abstract*—In the present-day world, reviews available on online websites play a vital role in the sale of products and resources. People using an online platform for purchasing different products very often read reviews of the product before making a decision on which type of product to which platform to buy and whether to buy or not. Basically, reviews are the statements that express the opinions, suggestions, or experiences of a person about any product available in the market. These reviews have their impact directly on the manufacturers and retailers as they are highly concerned about the feedback and opinion received by the consumer because customer service is the key. But nowadays spam/fake/pseudo reviews are written deliberately on the products to build/break their reputation and attract more consumers. This practice of giving falls truth untruthful deceptive reviews on the products to promote or devaluate the commodities and services is called opinion spam or review spam. Because of these reasons, it is the need of the hour to bring in new techniques for solving the problem of detecting fake reviews. With respect to this, our review paper compares the performance of different techniques Sentimental Analysis, different Boosting techniques in ML.**

***Index Terms*--: Sentiment Analysis, Opinion Mining, XGboost techniques, Machine learning techniques, Opinion Spam.**

I. INTRODUCTION

Globally, there is a tremendous increase in support to trade in electronics, sales and import of goods and services on online platforms. We all are aware that globally the web usage continues to grow in massive size and has become majorly important and hence the importance of reviews on various products on online platforms are also gaining importance equally [1].

Since online reviews play a vital role in making business decisions, positive or good reviews can lead to a major increase in financial gains and fame for organisations. but unfortunately, this gives a chance for imposters to play the system by deliberately posting fake comments and reviews to either credit or discredit business and target businesses.

Machine learning techniques can serve to be a greater contribution in obtaining a solution to the above problem of fake reviews. To retrieve and find helpful information we need to use machine learning algorithms and web mining techniques. Content mining is one such web mining technique. A very traditional example or illustration of content mining is Opinion mining. To understand the sentiments of the text, opinion mining uses machine learning. Here a classifier is built and trained to analyse and understand the features of reviews along with their respective sentiments. Most of the sentiment analysis problems are solved using SVM (support vector machine), where two classes will be analysed and evaluated. for example: A review being positive or negative. [2]

There have been multiple researchers who have come up with various techniques to find out review spam. There are few difficulties faced, one of them are unavailability of real life enriched labelled data set [3]. Mostly we can find data sets that are based on pseudo or fake reviews, but these data sets are completely different when compared to the real-world data.

There is another method introduced to detect fake reviews namely (extreme gradient boosting) XG boost, (Adaptive boosting) Ada boost and GBM (Gradient Boosting Machine). These methods accurately detect real life fake reviews when compared to that of traditional ML algorithms [3].

II. MOTIVATION

Any person can write his or her opinion review on any products on E- commerce platforms. Now these reviews can pose to be a threat, if intended to discredit any product or give a positive review on a worthless product. This can bring down the reputation of organisation drastically. Hence there is a need to come up with effective techniques and solutions to solve the above problem.

The focus of Syed Mohammad Anas in this research was to create and set up an environment of online E- commerce industry, where the users, consumers develop trust in the platform were the products the users purchase are 100% genuine and the reviews given on these products on their websites are authentic and are checked on a frequent basis by the companies. This will in return result in drastic increase of user engagement. In the present scenario, big social media organisations like Facebook and Twitter are deploying sentimental analysis in their systems to ensure the removal of derogatory posts, fake contents, news, or reviews [4].

With respect to Sifat Ahmed's perspective, there are a few people around who write honest reviews just to help other users in deciding about buying a particular product. Few imposter/ sellers are taking undue advantage by posting pseudo/ fake reviews to deceive the buyers or customers. Detecting the pseudo reviews is not a simple task, especially when the responsibility is handed over to the computer. Few fake reviews are written skilfully in such a format that even humans cannot detect that let alone computers. In addition, the freedom given to express or write their views without a proper monitoring system has made it even more difficult to detect these spam reviews. Therefore, to maintain the trust between the buyer and the seller these deceptive reviews must be detected [3].

As for the research conducted by Rodrigo Barbado and Oscar Araque, they have found that in the E- commerce industry customer reviews on products are playing a vital role for the users, creating a new format/ type of word of mouth (WOM) information. Latest studies shows that nearly 52% of the average customers online make use of the web to search for information related to the products, on the other hand 24% of them make use of the web to surf products before making purchases [5].

Considering the importance of reviews on products for businesses and the difficulty of obtaining a very good reputation and fame on the Internet, there are many techniques that have been used to improve online presence, including the fake ones. Fake reviews are one of the most popular unethical formats which are present on sites such as TripAdvisor or Yelp. However according to research conducted by Jindal and Liu, all the fake reviews are not equally harmful. Fake negative reviews or feedbacks on good products are very much harmful for business organisations and enterprises [5].

Md Forhod Hossain of Missouri state university says that a professional or an authentic fake reviewer usually involves in writing very large number of reviews on products available on ecommerce platforms. They might be working as a freelancer, or they might be working for an organisation to write pseudo reviews on products. They get equally paid for this work of theirs. Since they are involved in writing large number of reviews there is a certain linguistic and behavioural pattern that they follow, and these patterns can easily be identified by using different data mining algorithms. But before they are caught, for such activities they might already have done a significant amount of damage to enterprises and hence this damage cannot be compensated. To make things even worse, when they're caught for spamming reviews, they leave their original accounts create a new account and start the same process all over again and this has created a lot of issue and damage to huge business enterprises and organisations costing a lot of financial loss [6].

III. OBJECTIVE

The primary objective of the paper named ''Review of Fake Product Review Detection Techniques'' is to identify all the fake or pseudo reviews posted on a product. In this paper, we will be reviewing different techniques like opinion mining/ sentiment analysis techniques, Latent semantic analysis, XGBoost, Multilingual Sentimental Analysis, Social network Analysis and fine-grained opinion analysis etc. These use traditional ML algorithms like Naïve Bayes and Support Vector Machine (SVM), Random Forest etc. In addition, we will also investigate techniques that uses boosting algorithms like XG boost, Ada boost. Therefore, drawing a comparative study between the all the techniques based on different parameters. These proposed systems will save a lot of time and efforts since it helps business organisations to identify pseudo reviews and label them authentic or spam accordingly [19-20].

IV. ML ARCHITECTURE

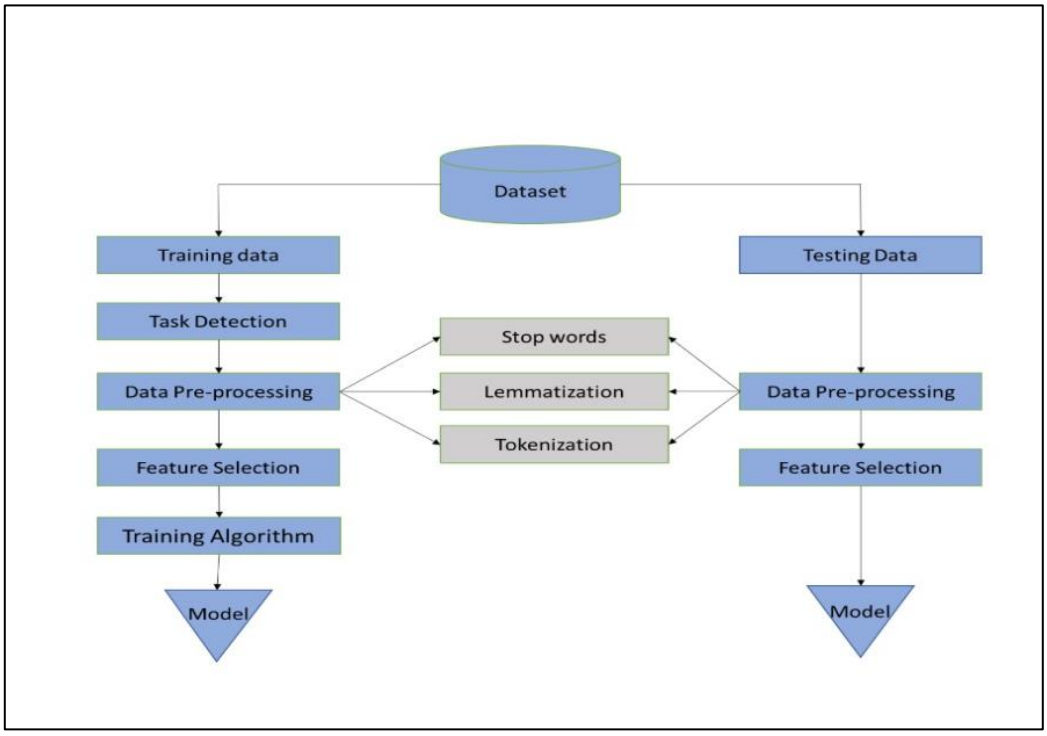

Fig 1. Machine Learning Architecture

- **Acquisition of required Data for further analysis:** In this step, we initially collect the data set which is obtained from the reviews given by people in different websites or applications such as amazon, flip cart, Myntra. These records are chosen in random from the reviews given by people on any of the products available on any of the mentioned Ecommerce platforms.
- **Splitting the dataset:** The obtained data set is then divided into two other sets such as training data set (Train data) and testing data set (Test data), where 80% of the data will be used in training data set and 20% of the data will be used in the testing data set.
- **Task Detection:** Analyzing and interpreting the task that needs to be performed using ML algorithms.
- **Pre- processing of data:** It is the step in which inconsistent, missing, and noisy data are analyzed and unnecessary punctuations, attributes and redundant words are removed. This step ensures that the data set is efficient for the training process of the model.
- **Feature Selection:** This process mainly focuses on reducing the count of input variables while developing a predictive model.
- **Classifier training:** Based on the extraction of various features we train the classifier using feature extraction. Both the testing and training data sets are compared with their features by which classifier's prediction is made on the data set.

V. EXISTING TECHNIQUES THAT ARE USED TO DETECT FAKE PRODUCT REVIEWS:

SL. NO	AUTHOR	TITLE	TECHNIQUES	ALGORITHM USED	OBSERVATION
1	Veselin Stoyanac.et al. (2008)	Topic identification for Fine-Grained Opinion analysis	Opinion mining	Lexicon semantic analysis	Introduced an approach to opinion topic identification that relies on the identification of topic coreferent opinions
2	Kaki leelaprasad.et al. (2009)	Combining social network analysis and sentiment analysis	Social network analysis, sentimental analysis, lexical analysis	Term frequency, User frequency and Document Frequency	Capturing of unique language is not possible, hence supervised learning will have to be deployed.
3	Wei Wang.et al. (2010)	Sentiment analysis of online product reviews with semi-supervised topic sentiment mixture model	Sentiment Analysis	Semi-supervised Co-LDA model	The results show that the semi-supervised Co-LDA is effective to topic-sentiment analysis.
4	Erik Tromp.et al. (2011)	SentiCorr: Multilingual sentiment analysis of personal correspondence	Multilingual sentiment analysis	AdaBoost Boosting Algorithm	Our system is aimed to help individual users to become more aware of the sentiment in their correspondence.
5	S.Chandrakala. et al. (2012)	Opinion mining and sentiment analysis classification: A survey	Opinion mining and sentiment analysis	Naïve Bayes, Support Vector Machines and Maximum Entropy	Many of the applications of Opinion Mining are based on bag-of-words, which do not capture context which is essential for Sentiment Analysis
6	C.K Sindhu.et al. (2013)	A survey on opinion mining and sentiment polarity classification	opinion mining and sentiment polarity classification	Decision Trees, Naïve Bayes	Many of the applications of Opinion Mining are based on bag-of-words, which do not capture context which is essential for Sentiment Analysis

7	Hoda Korashy.et al. (2014)	Sentiment analysis algorithms and applications: A survey	Sentiment analysis	SVM, Neural Network, Naïve Bayes, Bayesian, and Maximum entropy	This gives the overview of updates
8	Tamzid nahian.et al. (2016)	Review spam detection using active learning	Active learning	Perceptron Algorithm, stochastic Gradient Classifier and linear SVM	Among the 3 classifiers, Linear SVM works the best, with an accuracy of 88.2% and precision of 87.0.
9	M.Dharani.et al. (2012)	A survey on online review SPAM detection techniques	Modified method of iterative computation framework	ICF++ method	This paper reveals several approaches used for review spam detection and performance measures were identified. This paper identified Type I, Type II and Type III reviews
10	Jitendra Kumar Rout.et al. (2018)	A framework for fake review detection	Feature Engineering techniques	Supervised ML Algorithms.	The experiments from this framework conclude an accuracy of 90.19% for supervised learning and 83.70% for semi supervised learning.
11	Dr. Prakash S.Prasad.et al. (2019)	Enhancing NLP techniques for fake review detection	Active learning, Feature weighing like TF- IDF.	Rough set classifier, Decision tree and random forest	This paper suggests different factors for identifying fake review & all those factors are weighted using active learning.
12	Sifat Ahmed, Faisal Muhammad Shah (2019)	Using Boosting Approaches to Detect Spam Reviews	Boosting Algorithms	XGBoost, AdaBoost and GBM	Boosting algorithms result in high accuracy when compared to traditional ML algorithms. And also, they are effectively predicting in real life datasets

VI. CONCLUSION

This paper studies about the various techniques that are involved to detect fake product reviews effectively, so that business enterprises do not face huge loss in terms of finance and brand reputation. Here traditional supervised, boosting ML algorithms, Latent Semantic analysis, ICF++ and many more techniques are used to detect fake product reviews on the E commerce platforms. We majorly Opinion mining technique that uses Naïve Bayes and Random Forest as one of the methods and on the other hand we use XGBoost, AdaBoost and GBM boosting algorithms as well. From the results, we have observed that boosting algorithms are performing better in terms of accuracy when compared to the traditional ML algorithms. In addition, boosting algorithms have the capability to detect accurately in real life datasets whereas traditional ML algorithms fail to do so. Since Active learning algorithm is involved in creation of real-life datasets, the performance and efficiency of the boosting techniques are high. In this way the fake product reviews ca be detected effectively.

REFERENCES

[1] Pilaka Anusha and Kaki Leela Prasad, "Survey on fake online reviews using machine learning algorithms", Journal of Critical Reviews, Vol 7, Issue 18, pp. 2752, 2020.

[2] Ahmed M. Elmogy, Usman Tariq, Atef Ibrahim, Ammae Mohammed, "Fake Reviews Detection using Supervised Machine Learning", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 12, No 1, pp. 601, 2021.

[3] Sifat Ahmed, Faisal Muhammad Shah, "Using Boosting Approaches to Detect Spam Reviews",1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), .

[4] Syed Mohammed Anas and Santoshi Kumari," Opinion Mining based Fake Product review Monitoring and Removal System", Proceedings of the Sixth International Conference on Inventive Computation Technologies (ICICT), pp. 985, February 2021.

[5] Rodrigo Barbado, Oscar Araque, Carlos A. Iglesias, "A Framework for Fake Review Detection in Online Consumer Electronics Retailers", Information Processing and Management, Volume 56, Issue 4, pp 1234, July 2019.

[6] Md Forhad Hossain, "Fake Review Detection using Data Mining", Missouri State University, August 2019.

[7] M.N. Istiaq Ahsan, Abdullah All Kafi, Tamzid Nahian, Faisal Muhammad Shah, "An Ensemble approach to detect Review Spam using hybrid Machine Learning Technique",19th International Conference on Computer and Information Technology (ICCIT), December 2016.

[8] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "Fake Review Detection: Classification and Analysis of Real and Pseudo Reviews," Technical Report, Department of Computer Science, University of Illinois at Chicago, 2013.

[9] Veselin Stoyanov and Claire Cardie, "Topic Identification for Fine-Grained Opinion Analysis", Proceedings of the 22nd International Conference on Computational Linguistics, Coling, 2008.

[10] Wei Wang,"Sentiment analysis of online product reviews with Semi-supervised topic sentiment mixture model", IEEE, 2010.

[11] Erik Tromp and Mykola Pechenizkiy "SentiCorr: Multilingual Sentiment Analysis of Personal Correspondence", IEEE, 2011.

[12] S. Chandrakala and C Sindhu, "Opinion Mining and Sentiment Classification: A Survey", ICTACT Journal on Soft Computing, Volume: 03, Issue: 01, October 2012.

[13] C.K.Sindhu and S Chandrakala, "A Survey On Opinion Mining And Sentiment Polarity Classification" ICTACT Journal on Soft Computing, Volume: 03, Issue: 01, October 2012.

[14] Hoda Korashy, Walaa Medhat and Ahmed Hassan, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, Volume 5, Issue 4, pp 1093, December 2014.

[15] Tamzid Nahian, M.N. Istiaq Ahsan, Abdullah All Kafi, Faisal Muhammad Shah, "Review spam detection using active learning", IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2016.

[16] S P. Rajamohana and M.Dharani, "A survey on online review SPAM detection techniques", International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT), 2012.

[17] Jitendra Kumar Rout, Amiya Kumar Dash and Niranjan Kumar Ray, "A Framework for Fake Review Detection: Issues and Challenges", International Conference on Information Technology (ICIT), 2018.

[18] Ms. Rajshri P. Kashti and Dr. Prakash S. Prasad, "Enhancing NLP Techniques for Fake Review Detection ", International Research Journal of Engineering and Technology (IRJET), 2012.

[19] S Mohan Kumar, Darpan Majumder, A Shajin Naragunam and D V Ashoka, "A Symmetrically Diminished Interconnected Database Segmentation Framework Using Data Mining", Journal of Physics: Conference Series, Volume 1964, Advances in Computer Science Engineering, 2021.

[20] KR Shylaja, MV Vijayakumar, Darryl N Davis, EV Prasad , "Cognitive Architecture to Evolve Conscious Cognitive Tasks into Common Sense Actions on Agents", Proceedings of the World Congress on Engineering and Computer Science, 2013.