Proceedings of the International Conference on Sustainable Computing and Data Communication Systems (ICSCDS-2022)
IEEE Xplore Part Number: CFP22AZ5-ART; ISBN: 978-1-6654-7884-7

Fake Reviews Detection using Support Vector Machine

R.Poonguzhali[1]
Assistant Professor (SS)
Department of Computer Science and Engineering
Periyar Maniammai Institute of Science & Technology
Thanjavur, India
rampg@pmu.edu

S.Franklin sowmiya[2]
UG Student
Department of Computer Science and Engineering
Periyar Maniammai Institute of Science & Technology
Thanjavur, India
franklinsowmiya00@gmail.com

P.Surendar[3]
UG Student
Department of Computer Science and Engineering
Periyar Maniammai Institute of Science & Technology
Thanjavur, India
surendar21101999@gmail.com

M.Vasikaran[4]
UG Student
Department of Computer Science and Engineering
Periyar Maniammai Institute of Science & Technology
Thanjavur, India
vasikaran450@gmail.com

2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) | 978-1-6654-7884-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSCDS53736.2022.9760747

***Abstract*— One of the fastest expanding business categories in the world today is internet shopping. People nowadays buy a lot of things from internet shopping sites. Customers can buy a better quality products based on the reviews given by previous buyers of the products. Reviews includes text reviews, ratings and smileys. On a product review there are hundreds of reviews in which some of the reviews would be fake reviews. Opinion mining from natural languages is a difficult method for evaluating customers' sentiments, but sentiment analysis provides the best answer. It provides crucial data for decision-making in a variety of fields. So, we propose a fake reviews detection system using support vector machine which detect the fake reviews of the products. The primary goal is to suggest higher-quality products to the user. We use the support vector machine algorithm to classify the reviews into positive and negative groups. Finally fake reviews are predicted which are posted by the users. The reviews are grouped as negative, positive and neutral. In this system, only purchased users can post the reviews and duplicates are verified based on user id and booking id. Genuine reviews are considered for product recommendation**

Keywords—Online shopping, Product Recommendation, Machine learning, Sentiment analysis, Support Vector Machine

I. INTRODUCTION

Reviews are feedbacks which shows suggestion of the customers. With the advancement of e-commerce, a rising number of consumers are enjoying the convenience of purchasing online and then sharing their thoughts on the products they purchased. This information given by the consumers in the shape of online reviews is valuable for users and therefore represents an essential element in the buying decision process and also helps in targeting more customers to select for purchasing online. Without trusted external control, every individual can give their reviews on the e-commerce site because of the various web technologies. In this paper, we have come up with a proposal to analyze the reviews posted for the products, using the concept of Sentiment Analysis with supervised machine learning. The sentiments include reviews, ratings and emotions which are gathered, analyzed and are differentiated based on their sentiment for positive, negative and neutral reviews. This helps the user to identify if the product needs to be bought or not. The reviews are then checked for their credibility and the genuine reviews are filtered. Further the opportunity to post fake reviews are prevented with the help of user ID and Booking ID verification techniques which helps the user to select a better quality product to buy and not get misleaded.

II. LITERATURE WORKS

They created a comparison between two models to describe performance for the dataset and their value in real-time software deployment. The goal is to find an algorithm that will do the work of detecting false reviews. [1]. For detecting false reviews, the author proposes a spam review detection approach. The model employs four methods. The first three phases are recommended to improve the base classifiers. [2]. Provides a method for detecting negative or fake reviews using machine learning. To extract reviewer behaviors, they used one approach that is feature extraction approach [3]. Provided a comprehensive analysis of the most well-known machine learning-based spam review detection methods. Using classical statistical machine learning to boost feature extraction and classifier development enhances text classification model performance.. [4].Using a dataset of hotel evaluations from online sources, researchers created an effective supervised machine learning strategy for classifying fraudulent online reviews [5]. To detect bogus reviews, the proposed method employs a mix of multiple feature fusion rolling collaborative training. This gives an effective results Compared with the traditional methods. Here unlabelled data is used to increase the efficiency and produced the better accuracy [6]. The proposed system is a multilayer interactive with neural network model that is based on attention. The multilevel implicit expression form of reviews is mined by this

model. The model they suggested is demonstrated by the experimental results from available datasets. [7]. Fake reviews detection is a complex task because of the major fake reviews are connected to at least two main sources such as behavior of the reviewer and content of the review [8]. Developed a sentiment classification algorithm based on the TF-IDF that can classify sentiment value with 92% accuracy. [9]. Classified the reviews as positive or negative by using 15 linguistic features [10]. Sentiment analysis was performed on a number of academic publications in a variety of forms. Used Machine Learning techniques for classification purpose [11]. The major goal of this research is to make a contribution by offering a set of strategies that function together and complement each other. [12]. The performance of ensemble learning-based algorithms for detecting fraudulent online information is calculated in this research. Then, the datasets are collected such as restaurant reviews. The reviews are categorized using machine learning methods such as decision trees, SVM, gradient-boosting trees, random forests, and multilayer perceptron. [13]. This article is only intended for reviews in English. The planned work is based entirely on the users' star ratings. In order to detect intrusions, the average rating should be used as a guide. [14]. This paper present sentimental analysis for detecting the spam in the product reviews. Sentiment analysis play vital role to make business decision about the product then applied two machine learning algorithms for analysing the amazon product reviews [15]. Provide two sorts of features: a behavioural feature set and a semantic feature set. After that, using semantic analysis, three feature sets were presented. [16]. The study of online shopping website reviews led to the development of a novel approach for detecting false product reviews. By evaluating the temporal trends of reviews and comments, the outlier identification approach is utilized to find outlier products. [17]. To detect false online review records, many semi-supervised and supervised text mining approaches are used in this paper. [18]. An evaluation of approaches for spotting spammers on twitter was proposed. In the taxonomy of twitter spam detection techniques, false content identification, Web address-based detection and spammer identification approaches were all categorized. [19]. Importance of the qualities on the choice is validated by data collection. Experiments on a huge number of reviews collected from the internet were conducted, demonstrating the efficacy of the proposed strategy. [20]. They used several multiple language review datasets in the proposed approach, such as Urdu reviews and Roman Urdu reviews. It's tough to tell which reviews are bogus. As a result, the n-gram technique is used to detect bogus reviews in a variety of languages. The most effective method for detecting fraudulent reviews has been determined to be text classification using a support vector machine classifier. [21]. Deep learning algorithms are designed to tackle the classification problem from beginning to end. The recent comeback of deep learning architectures has benefited text review classification because of their ability to achieve high accuracy with less engineered features. [22]. The proposed method uses a set of Amazon Review Dataset to detect false reviews. Rather than using the standard strategy, they applied numerous boosting algorithms to increase the accuracy [23]. This work proposes a semi-supervised learning strategy based on self-training. This approach classifies spam based on the training data that has been labelled. [24]. The proposed CNN model outperformed existing approaches when evaluated on the Deceptive opinion spam corpus. Experiments have demonstrated that in addition to textual semantics, new factors must be examined in order to successfully identify deceptive viewpoints in reviews [25]. It is feasible to eliminate misleading news by identifying fraudulent and duplicate customer accounts. The recommended methodology is based on the customers' initial evaluations. In order to detect intrusions, the average rating should be used as a guide. [26]. To detect spam reviews, a multi-dimensional analysis was proposed. It uses three approaches to provide credibility reports for particular products: duplicate review detection, review distribution and rating count anomaly detection and incentivized review detection. All of the methods provide useful information that can be used as an overlay to detect false reviews [27]. To detect spammers from product reviews, a partially supervised learning-based method is applied. This model mines the review data for spammer group candidates using frequent item mining. [28]. In this project, we use automatic message production and receipt from the service provider. We add the app version they were using while submitting a review for their chosen app in the proposed work. [29]. Create a system for discovering and eliminating bogus reviews. Here are some product review comparisons. Using network characteristics and geolocation, the system calculates the bogus review's for computer and mobile. The device also detects the reviewer's position. [30].

III. EXISTING METHODOLOGIES

Sentences that represent observations or a positive or negative attitude presented in a positive or negative manner. Users give feedback on their e-commerce websites. Existing systems that only utilize ratings due to a lack of data. Those ratings are analyzed to determine if the review is positive or negative. The limitation of this system is need to improve some process, then only it can detect fake reviews efficiently in a short period of time. To determine if the user is a fake reviewer or not, the evaluator analyze the reviews. The evaluator then gathers evidence regarding the relationship between these reviews and compares it to other sets of reviews. Even if some of the algorithms have been utilized in the study of opinion spam, the outcomes are better. However, no algorithm can handle all of the obstacles and problems that today's youth is confronted with. Fake reviews are sometimes mistaken for genuine reviews, and the settings have been tweaked so that no one can tell what their true motivation is. As a result, some additional work will be required in the future to increase the performance.

IV. PROPOSED METHODOLOGY

Using the Support Vector Machine (SVM) technique, we offer a method for detecting online fraudulent reviews based on Sentiment Analysis. In addition, recommend the products based on genuine positive reviews. Recommendation systems being a part of information extracting system. In this work, we predict the reviews that a customer is likely to leave for a product. We can construct an E-Commerce Interface first in our model. It is used to make internet purchases. Admin can add the products with specification. User can search the products and analyse the sentiment. If the product is positive means, automatically recommend the products otherwise recommend other positive products. Here, only purchased user can posts their reviews, it is based on User ID and Booking ID. Reviews includes text reviews, ratings and

smileys. Then, Sentiment Analysis will perform. It analyse the sentiments of the people. In text reviews it extract some keywords. In ratings it calculate the star count values. In smiley reviews it analyse the emotions. Now by using Support Vector Machine algorithm, those reviews are classified as positive or negative. After this process the identified fake reviews will be eliminated and positive reviews will be updated in the database. Then, genuine products will be recommended. The customer can give the reviews only once by using the User ID. So, this will prevent the Fake Reviews in our proposed work.

V. ARCHITECTURE

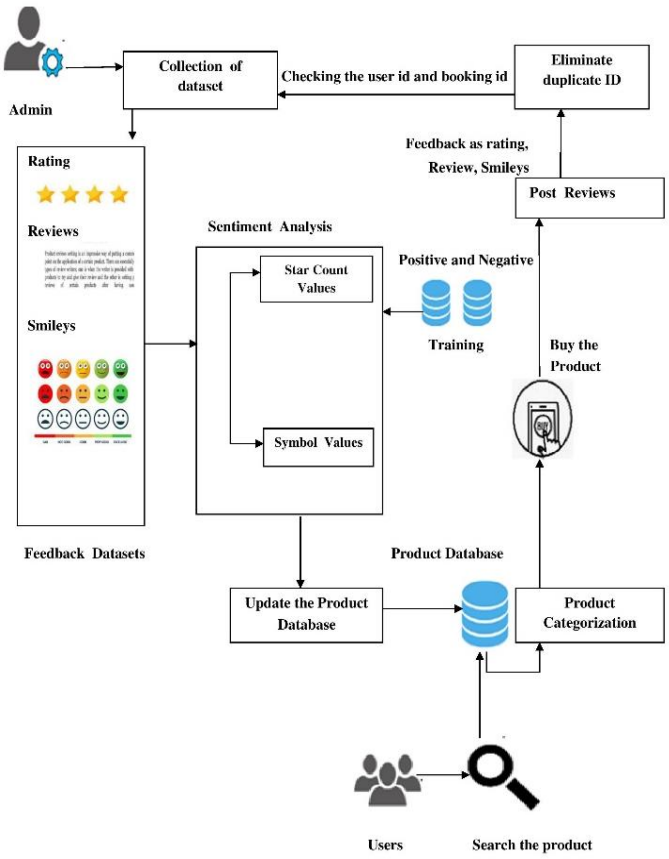

Fig 1: Architecture

In this fig 1, the datasets are collected, which is called as feedback datasets. The feedback datasets includes ratings, reviews and smileys. Then, Sentiment Analysis will be performed. It analyze the sentiments of the people. In text reviews it extracts some keywords. In ratings it calculates the star count values. That is, suppose the given rating is 4 stars then it will be categorized as good product and if the given rating is 1 star the product will be categorized as not a better quality product. In smiley reviews it analyses the emotions. After this process the identified fake reviews will be eliminated and the genuine reviews will be updated in the database. The customer checks the reviews and purchases the product, after that they post their experienced reviews. The customer will update the reviews with user ID and Booking ID. The customer can give the reviews only once. So, this will prevent the Fake Reviews in our proposed method.

VI. MODULES

There are five modules implemented in our model which is given below

A. Online E-Shopping Interface

This module is used to create an online shopping website for recommending best products. Admins are in charge of keeping track of information on the server. Admin and user accounts are the different kinds of accounts available. Admins can access the system and add a new product. Then a product details will be stored in a database. User can create an account and login to the website. After that they can view the products.

B. Reviews Collection

Admin can collect different kinds of reviews, such as users' text reviews, star ratings then smileys. All of the reviews are saved in a database for future analysis. Product database contains Booking Id, Product Id, Product Name, User Name and Reviews. Here Ratings, Reviews and Smileys are classify in terms of Support Vector Machine.

C. Sentiment Analysis

Sentiment analysis has great importance in fake reviews detection system. It analyse the sentiments of the people. In this module, by using sentiment analysis, admin can analyse the review whether it is positive or negative. Sentiment analysis can be used to calculate the star count values in star rating. It extract keywords from text reviews, match them to the database and classify the data using the Support Vector Machine technique. Then smiley reviews are calculated based on Support Vector Machine algorithm.

D. Fake Reviews Monitoring

In this module duplicate reviews are monitored then removed. Admin has access to user account information, mobile address information, and order id information. As a result, users can post reviews that are genuine. If the user can post the reviews from same user id, booking id means, automatically consider as duplicate reviews and it will be eliminated. Genuine reviews are forward to next module for recommendation.

E. Recommendation System

Recommender systems are a type of information extraction system that attempts to predict a user's feedback of an item. The user can search for a product and get a list of products based on pricing and also get the product reviews information. Support vector machine will classify the reviews as positive or negative, then genuine reviews are displayed in the recommendation panel. If a product has a negative review means, the other positive (real) products will appear in the recommendation panel automatically.

VII. CONCLUSION

Finding fake reviews is a difficult issue for researchers. In existing system they have used several types of method to detect fake reviews. In this system, we implemented fake reviews detection using Support Vector Machine. Online e-commerce interface is developed, in that admin and user login are created. Then, the user can easily find the products. Positive and negative attitudes are determined by analyzing ratings, reviews, and smileys. Negative reviews will be eliminated and genuine reviews are provided for product recommendation. The objective of this system is to recommend higher-quality products based on genuine feedback. Using our present model, we were able to achieve a better result.

ACKNOWLEDGMENT

We express our heartfelt gratitude to our Department of Computer Science Engineering faculties for their valuable guidance, spontaneous assistance, and total support in ensuring the successful completion of our project. We will never forget our parents' support and encouragement in the preparation of our project work, as well as the help and encouragement we received from our friends. We are thankful to everyone who has assisted us.

REFERENCES

[1] Santoshi Kumari, Syed Mohammed Anas "Opinion Mining based Fake Product review Monitoring and Removal System" 6th International Conference on Inventive Computation Technologies (ICICT). IEEE,2021.

[2] Hui Jiang, Jianrong Yao, Yuan Zheng "An Ensemble Model for Fake Online Review Detection Based on Data Resampling, Feature Pruning, and Parameter Optimization " IEEE Access 9 (2021): 16914-16927.

[3] Ahmed M. Elmogy, Ammar Mohammed, Atef Ibrahim, Usman Tariq "Fake Reviews Detection using Supervised Machine Learning." International Journal of Advanced Computer Science Applications.Vol. 12 (2021).

[4] Matthew Springer, Rami Mohawesh, Robert Ollington, Shuxiang Xu, Son N. Tran, Sumbal Maqsood, Yaser Jararweh "Fake Reviews Detection: A Survey." IEEE Access,Vol 9 pp. 65771-65802, 2021.

[5] Md. Rabiul Islam, Rakibul Hassan "A Supervised Machine Learning Approach to Detect Fake Online Reviews." 23rd International Conference on Computer and Information Technology (ICCIT). IEEE 2020.

[6] Fanqi Meng, Genhua Shi, Haitao Kan, Jingdong Wang, Qizi Mu, Xixi Xiao "Fake review detection based on multiple feature fusion and rolling collaborative training" IEEE Access 8 2020. 182625-182639.

[7] Jiyun Zhou, Meiling Liu, Qi Yue, Yue Shang "Detecting Fake Reviews Using Multidimensional Representations With Fine-Grained Aspects Plan." IEEE Access 9 2020. 3765-3773.

[8] Nour Jnoub, Wolfgang Klas "Declarative Programming Approach for Fake Review Detection." 15th International Workshop on Semantic and Social Media Adaptation and Personalization SMA. IEEE, 2020.

[9] Md. Rabiul Islam, Rakibul Hassan "Impact of Sentiment Analysis in Fake Online Review Detection." International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD). IEEE, 2021.

[10] Akbar Siami Namin, David R. W. Sears, Faranak Abri, Keith S. Jones, Luis Felipe Gutierrez"Linguistic features for detecting fake reviews." 19th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2020.

[11] Aditya Uday Naik, Jane Crystal Rodrigues, Jane Trinity Rodrigues, Pratiksha Shetgaonkar, Shailendra Aswale, Vialli Luis Keji Gonsalves "Machine & deep learning techniques for detection of fake reviews: A survey." International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE). IEEE, 2020.

[12] Anand Godbole, Kaustubh Gaikwad, Kunan Mandalya, Ruchit Furia "Tool For Review Analysis Of Product." International Conference on Emerging Trends in Information WTechnology and Engineering (ic-ETITE), IEEE 2020.

[13] Akbar Siami Namin, David R. W. Sears, Faranak Abri, Luis Gutierrez-Espinoza, "Ensemble learning for detecting fake reviews." 44th Annual Computers,Software, and Applications Conference (COMPSAC).IEEE, 2020.

[14] Mallidi. A.E.Reddy, Ganta Rojamani, Praveena Pillala, Priyanka Kumari Bhansali "Random Forest Model for Intrusion Detection in Crowd-Sourced Reviews."International Conference on Smart Electronics and Communication (ICOSEC). IEEE, 2020.

[15] Dibya Jyoti Bora, Gyandeep Dowari, Temjensangba Aier "Fake Product Review Monitoring and Removal using Opinion Mining.“ 2020.

[16] Wang, Xinyue, et al. "Identification of fake reviews using semantic and behavioral features." 4th International Conference on Information Management (ICIM). IEEE, 2018.

[17] Liu, Wenqian, et al. "A method for the detection of fake reviews based on temporal features of reviews and comments." IEEE Engineering Management Review (2019): 67-79.

[18] Hassan, Rakibul, and Md Rabiul Islam. "Detection of fake online reviews using semi-supervised and supervised learning." International conference on electrical, computer and communication engineering (ECCE). IEEE, 2019.

[19] Masood, Faiza, et al. "Spammer detection and fake user identification on social networks." IEEE Access 7 (2019): 68140-68152.

[20] Sanjay, K. S., and Ajit Danti. "Detection of fake opinions on online products using Decision Tree and Information Gain." 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2019.

[21] Danish, Nazir M., et al. "Intelligent interface for fake product review monitoring and removal." 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE). IEEE, 2019.

[22] Shahariar, G. M., et al. "Spam review detection using deep learning." IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). IEEE, 2019.

[23] Ahmed, Sifat, and Faisal Muhammad. "Using Boosting Approaches to Detect Spam Reviews." 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT). IEEE, 2019.

[24] Navastara, Dini Adni, Ana Alimatus Zaqiyah, and Chastine Fatichah. "Opinion Spam Detection in Product Reviews Using Self-Training Semi-Supervised Learning Approach." International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA). IEEE, 2019.

[25] Archchitha, K., and E. Y. A. Charles. "Opinion Spam Detection in Online Reviews Using Neural Networks." 19th International Conference on Advances in ICT for Emerging Regions (ICTer). Vol. 250. IEEE, 2019.

[26] Prabakaran, Anusha, and Min Chen. "Product review credibility analysis." International Conference on Computing, Networking and Communications (ICNC). IEEE, 2019.

[27] Kudo, Wataru, Mao Nishiguchi, and Fujio Toriumi. "Fraudulent user detection on rating networks based on expanded balance theory and GCNs." Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

[28] Zhang, Lu, Zhiang Wu, and Jie Cao. "Detecting spammer groups from product reviews: a partially supervised learning model." IEEE Access 6 (2017): 2559-2568.

[29] Manoj, K., et al. "Genuine ratings for mobile apps with the support of authenticated users' reviews." Second International Conference on Green Computing and Internet of Things (ICGCIoT). IEEE, 2018.

[30] Raja, B., et al. "Eliminating products' fake reviews using network parameters and geolocation." IOP Conference Series: Material Science and Engineering. Vol. 981. No. 3. IOP Publishing, 2020.