Prevention of Fake Comments using web3

*T. Nithya, *A Amrita Varsheni, *S Brindha

[1]Rajalakshmi Institute of Technology, Chennai, India
[1]{nithya.t, amritavarsheni.a.2019.cse, brindha.s.2019.cse}@ritchennai.edu.in

Abstract. The volume of information on the internet is currently rising dramatically. Social media platforms/e-commerce market place is producing a lot of data, including reviews, comments, and opinions, every day. As there are a number of fake reviews should incorporate Spam detection to produce a genuine opinion. Fake reviews are growing problem in online shopping, and they have a significant impact on consumer'sdecision-making. Many people today base their decisions when choosing a product or service on social media opinions. Because so many false or phoney evaluations have been written by businesses or individuals for a variety of reasons, detecting opinion spam is a difficult and time-consuming task. They produce fictitious reviews to deceive usersor automated detection systems by elevating or degrading the reputations of their targetproducts in order to elevate or lower them. In this article, we'll regulate it by leveragingblockchain technology to make the review system more authentic by allowing only legitimate product purchasers to submit evaluations using their account credentials. Weuse the Ethereum blockchain to authenticate user credentials, and we only permit verifiedusers to purchase things. Also, we only permit customers to leave reviews or commentson products, ensuring that the reviews are accurate.

Keywords: E-Commerce site, Fake comments, bogus ratings, Blockchain Technology, Ethereum Blockchain method, product review.

1 Introduction

E-commerce is one of the sectors with the fastest development. E-commerce sites typically give customers the option to leave product reviews. These reviews' presence can be used as a source of knowledge. Companies can use it to decide how to develop their products, for instance, but regrettably, some people have sought to take advantage of the significance of evaluations by fabricating reviews to either promote the product or denigrate it. On the internet, they express their ideas. It is customary for people to research products before making a purchase. Customers can evaluate various brands and decide on a certain product based on reviews. Any information in social media platforms is trusted by the common people. Such as comments, likes, ratings of a product in online platforms which brings the product appear to be the most popular one than it is actually, also gives more visibility of the product on the platform.

There are several reasons to create a fake comment in social media or in an e-commerce website, the first reason is to increase the popularity of the product, the second reason is to

manipulate the public opinion. The fake reviews or ratings of the products to discredit the organization or a person. There are some people who create fake accounts to post fake comments in order to traffic the particular product or the website that generates revenue through the clicks of advertisements or affiliate links. In addition, fake comments can be used in the form of social engineering, an individuals or group used to comment to gain trust of others or phishing attacks. As there is a lack of physical inspection in online shopping, these comments will help the people to judge the product with the ratings, reviews by comparing the product with other websites for the quality of the product that meets their expectation. As it provides easy access to information of a product, people access the product information that make them to make decision.

According to research 2.14 billion people purchase products through online store in which approximately 90% of the people buy products on the basis of its feedback. Here the major problem is to determine whether the review is genuine or fake. In the present era, fake reviews are the most promising issues as it impacts the business world considerably. More precisely, in the e-commerce industry, customer feedback plays a role in both corporate growth and decline. The main issue is figuring out which comment is false and which is genuine. The

bogus comments or reviews are human generated or sometimes machine generated, both of these comments are difficult to identify and remove by humans. In general, people will blindly trust the reviews or ratings given for a product, these fake reviews or ratings would destroy the true quality of the product also the sales of the particular product in online store.

The difficulty in spotting phoney reviews is known as the fake review detection problem. Because fraud users frequently post phoney comments using the new capabilities to bypass detection methods, the current methods for bogus comment detection are ineffective in halting the spread of fake reviews. Consumers or consumers struggle to trust internet evaluations, thus in order to purchase a high-quality goods, they must carefully study and compare reviews. But, evaluations are advantageous since they offer crucial information that aids consumers in making decisions about whether to spend their money on high-quality goods and services. These fake reviews produce a negative view to consumers, that will damage the online platforms such as their reputation and reduces the number of consumers to purchase products on their platforms. In order to fully comprehend the fake review problem and discover a workable solution, ongoing research is necessary, as fraud users regularly alter their methods in order to evade detection. The consumers, vendors, and researchers are still impacted by this issue and are working to find solutions. To maintain their reputation, the online platforms must ensure that the customers need to experience true information that provides a guaranteed fair competition.

2 Methodologies

We formed a method for this research by consulting multiple journals, IEEE papers, Springer. This method explained the research objective, research methodology, and selection and exclusion requirements.

2.1 Search Strategy

Many literature reviews in leading online journal databases such as Springer, IEEE, Science Direct are advised. The fake comment detection and removal system is used in e-commerce sites such as Amazon India, Snapdeal, India MART, Nykaa, Book My Show, Flipkart, Alibaba, Myntra are limited to research papers in India that are published after 2018.

2.2 Requirements for Inclusion and Exclusion

Based on the following requirements, the research papers were selected:

- Methods, quality of the fake comment detection and removal model in detailed papers.
- Latest Published in the journals from 2018-2023

To filter the data, the papers were excluded based on the following exclusion requirements:

- Papers that we cannot be able to the access the directly or indirectly.

- Papers that don't have detailed explanation and assessment proofs.

3 Related Works

Collecting and evaluating data on the various circumstances that can cause people to post phoney reviews or comments is the first step in creating a trustworthy fake comment monitoring system. Many sources, including Amazon, Flipkart, Meesho, and other e-commerce platforms, can be used to gather this data [6]. Then, patterns can be found and false reviews can be detected using machine learning methods as support vector machines, k closest neighbors, logistic regression, or classification algorithms [2]. The classification method utilized to identify whether or not the given product review was legitimate was Support Vector Machine (SVM) with a linearsupport vector classifier kernel. A learning method for prediction and classification is connected to this supervisedlearning model for pattern recognition. KNN, a simple algorithm, is also used for pattern prediction and categorization. The K parameter and voting scheme are just two of the many variables that affect how well the KNN model performs. Logistic regression is a machine learning method used for the task of binary classification.This logistic function is used to make predictions. Every classification algorithm includes a step for identifying bogus reviews. The bogus comment monitoring and detection system must be regularly updated and reanalyzed in light of fresh data and user feedback [4]. The effectiveness of the system can also be assessed using metrics like recall, precision, and accuracy to make sure that it offers consumers relevant and helpful real reviews.

4 Fake Comment Detection and Removal System

The fake comment removal system is based on three types of inputs. They are:

- users (the fake comment removal system is fully based on user behavior)
- Items (the fake comment removal system is fully based on products on e-commerce site)
- Payment (the fake comment removal system is fully based on the connection between the user's detailsto pay for the purchased products).

The main goal of the fake comment removal system is to remove phoney reviews on the e-commerce site. Implementing the fake comments removal system is not that simple, there are various types of fake comments detection and removal system techniques available. Thus, Fig-2 explains the overall fake comments detection andremoval system with respect to skikit-learn (SKL), support vector machine (SVM) and K nearest neighbor (KNN) where the

techniques used to filter the phoney reviews and remove them [5].

The implementation process in the fake comment detection and removal system are lengthy, choosing a best suited algorithm for this process would help in bringing the speed and accuracy of the system. That is, the system ensuresonly genuine comments that are to be available for the users on each product. Every product's review plays a major role in user's or buyer's decision making, and also the reputation of the e-commerce site and its products.[2]

The performance metrices are also determined by the speed, accuracy, precision and recall, this ensures that the performance of the algorithm is more efficient and works effectively to detect and remove the fake comments/reviews on each and every item on the e-commerce site [4]. It helps in increasing the sales growth, as it produces a desired product and it also includes the reputation of the e-commerce site or the owner.

In the fake comments detection and removal system there are many detection techniques are used. In general, to detect and remove false reviews machine learning algorithms are used widely [5]. The text is classified by data processing and extracted by training model and tested for phoney reviews then it is removed from the e-commercesite. There are several algorithms and methods under machine learning. The methods are deployed in the system to evaluate and detect the phoney comments in the real time. Here are some general techniques that are used in phoney comments detection and removal system.

4.1 General Techniques

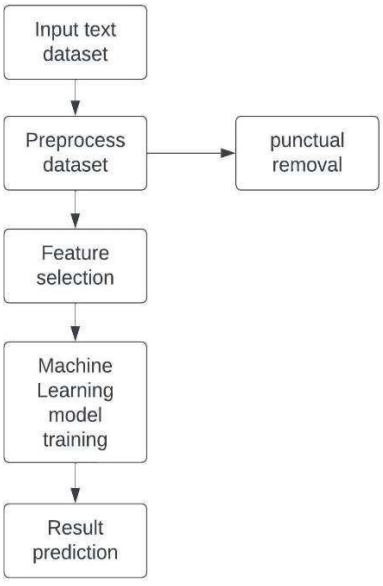

Fig. 4.1 Fake comments removal technique

4.1.1 Preprocessing: A filtering procedure is part of the data preprocessing phase. It stands for the text elementsthat we remove, such as punctuation marks, because their absence reduces the classification process's overall accuracy. The word count is determined after the

punctuation has been removed. The process of feature selectionis then completed, choosing variables or identifying qualities to build an effective model. The criteria for featureselection are as follows:

1. Number of words
2. Bigram Type
3. Word for relationships
4. Word count for sentences
5. Counting Nouns and Verbs

With these parameters, the length of the review is first estimated, and then the Markov Model—also known asthe Bigram Probability Model—is used to determine the likelihood of the next word. There are some words in the describes the relationship, SKL feature selection is done by considering the relationship words. To tokenizethe sentence, Natural Language Toolkit is used and then by using Part of Speech, the words are tagged as noun,verb and adverb, etc. As part of this feature selection procedure, nouns and verbs are also counted and calculated. The dataset is then divided in an 80/20 ratio to test the samples. Cross validation is then performedon the dataset's entire classification accuracy.

4.1.2 Classification Algorithm: The act of categorizing data into two or more classes or labels is referred to as classification. Many classification methods are included with ML. When training machine learning models, feature selection is crucial to getting better classification results. Support Vector Machine (SVM) with a linear support vector classifier kernel was the classification technique used to determine whether the given product review was authentic or not. It is a supervised learning model for pattern identification that is linked to a learning algorithm used for prediction and classification. KNN is also used for classification and pattern prediction, which isa straight forward algorithm. The performance of KNN model depends on many factors, suchas K parameter and voting scheme. Binary classification is a task that makes use of the machine learning technique known as logistic regression. For making predictions, this logistic function is employed. The fake review identification process is a component of every classification algorithm.

4.1.3 Detection process: The data will evaluate the model to anticipate a proper outcome after the training phaseis over. Using SVM, KNN, and logistic regression, the model is trained.

4.1.4 Experimental design/details: Based on this investigation, it was shown that bogus reviews have a significant impact on how internet enterprises operate. The dataset for this study was self-extracted using a filtering approach from the Yelp website. Fake reviews that are taken from this website and represented as semi-real data are more realistic. Also, with realistic datasets, bogus review identification is more difficult. A dataset is developed for fake review detection bypass reviews that are having length less than 150 characters, by considering this fact, an extensive dataset has been created that includes the review length from 452 words to 15words. To this effect, a self-extracted dataset has been developed. In the process of extracting data, two aspects are evaluated: mixed reviews and class bias. To avoid imbalance issue in dataset, equal number of positive and negative reviews are taken.

A record of 20 hotel reviews have been collected from the Yelp website that includes nearly 2000 reviews in which some are harmful and some are positive reviews. This provides a balanced dataset. It is the best data splitting technique, and every classification model needs to use it. Divide the dataset into two subgroups in a ratio for training and testing for the optimum model performance evaluation. Depending on the classification process, the ratio

may change. One of the most difficult problems in fake review detection is the lack of a labelled dataset.

4.2 Evaluation of Fake Comment detection and removal System

In fake comment detection and removal system there are several criteria that involves accessing the system abilityto accurately identify and remove false comments. The speed and accuracy of the system is determined by detection and removal of phoney comments. The e-commerce site has rapidly grown, everyday billions of users shop online, the false reviews will take some time to be detected and removed by the owner.

The reviews must be analyzed and uses different algorithms to predict or find the phoney comments that are to beremoved. These are found by comparing a review with other reviews and based on the word count and similarityof the words the phoney reviews are detected and removed.

There are also many other evaluation techniques used in the fake comments removal system that include cross validations, A/B testing, and machine learning modules, A/B testing involves comparing the system to a control set for testing. While one set of users accesses the e-commerce site with the system in place for detecting and removing fake comments, the other group accesses the site without it. By contrasting the metrics of the two categories, the performance of the system is assessed.

Cross validation technique entails creating training and testing sets from the dataset. The training set is used to train the system, and the testing set is used to test it. The performance of the system is assessed using metrics suchas precision, recall, F1 score, false positive rate, and false negative rate.

In Machine learning module technique, using labelled datasets, modules can be trained to recognize false comments. Metrics like precision, recall, and F1 score can be used to assess the models' performance.

4.3 Limitations of Fake Comment Monitoring and Detection System

The following are the drawbacks associated with existing methods. This will mislead the customers to spend money with the business or on a product. It takes time to define and choose suitable algorithm [1]. This will compare the attributes with the sentimental attributes of the reviews and then take the similarity of those attributesto detect it [5]. It is intensive to large datasets. It takes more time to train large datasets [4]. The performance of SVM highly depends on choosing its kernel. The SVM model makes it difficult to understand the interpretation on how they make prediction [5]. Most common surveys have used machine learning and deep learning algorithms to detect fake reviews. The process of detecting fake comments are slow that may confuse consumersto purchase the product. This may provide wrong information that will make consumers to be unsatisfied about the product.

5 Proposed System

In our proposed system, we use a technology called as blockchain to fulfill our objective. Our main aim is to prevent the spammers from giving a fake review on the products, we allow only genuine comments to be added for the products in the e-commerce industry. We

use Ethereum blockchain to authenticate the user. In this technology we use public key and private key to authenticate a user. As this technology isa two-factor authenticator, it is providing more security.

Blockchain authentication refers to the system developed for increasing the security of the user and verifies user identity and allows users to connect to online portal. It uses authentication methods to enhance the privacy and security of authentication systems. Also contains data integrity. Since identities are decentralized on blockchain,it gives end-user more control over digital identity, due to this nature of blockchain technology, data cannot be tampered with. Therefore, personal identities are secure on blockchain. There are some procedure and process that are in our system to comment on the products in E-Commerce website or industry. The frame for for the process is specified in the Figure 5.1 below.

After the verification of user with authentication process, the user will be directed to the online portal and willbe able to access the products that are available on the shopping portal. It displays a list of products or items with few information such as its image, reviews/ratings, and product description. The user will not have an option to comment on the products until they purchase.

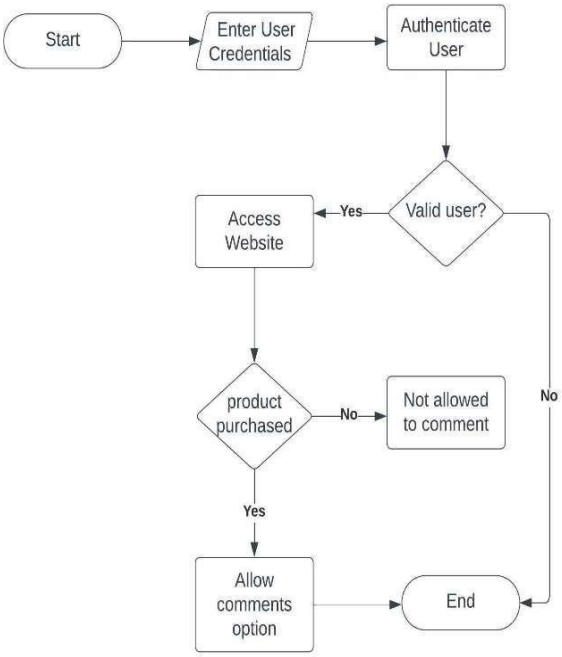

Fig. 5.1 Workflow of proposed system

Once the user buys a product, only then a comment option will be visible for the products purchased, where thebuyer can comment on the product. The buyer can then edit their comment on the shopping portal, if necessary,with same id in which they have purchased the product, these comments are then replaced using Ethereum Blockchain technology. In this way we can avoid fake reviews on online shopping portal.

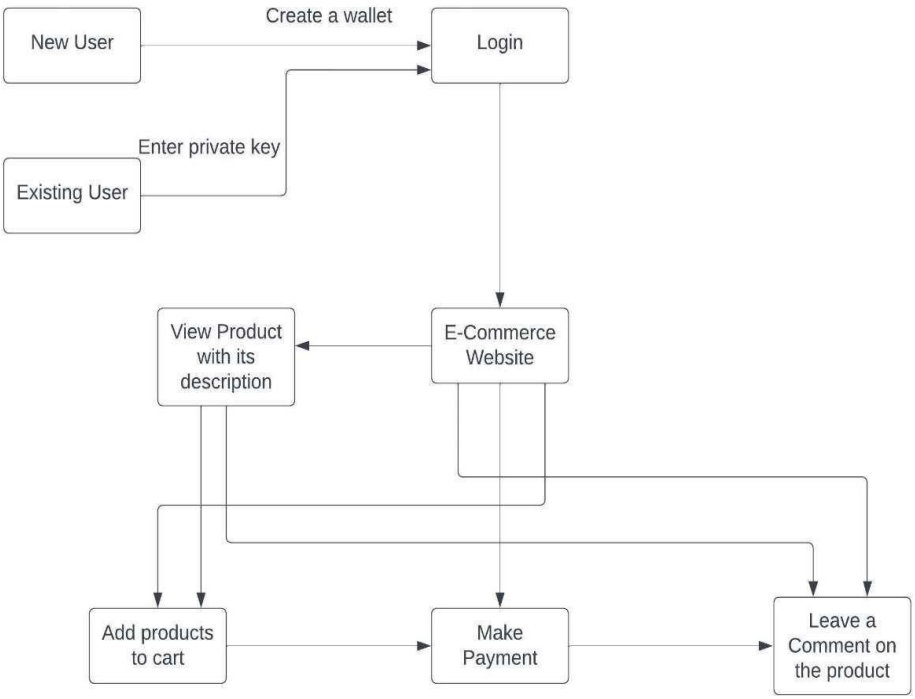

Fig. 5.2. Architecture diagram

In this system, we incorporate Ethereum blockchain, a decentralized application widely used in blockchain technology for secured platform. This Ethereum blockchain technology is used for authentication in our system,that authenticates the user login and prevents from hacking. When it is applied in our system, Individuals or organizations can build and manage their own digital identities using Ethereum. This is accomplished by buildinga smart contract that contains the identity data and can be used to confirm the legitimacy of the person or group. On the system, the user starts a transaction that needs authentication. When a transaction is detected, Metamask asks the user to validate the transaction's specifics, such as the petrol fee and the destination address. Once the user has confirmed the transaction, Metamask asks them to sign it with their private key. With their private key, which is safely kept in the Metamask wallet, the user signs the transaction. The signed transaction is broadcast tothe Ethereum network using Metamask. The transaction is received by the Ethereum network, which then processes it and updates the blockchain's state appropriately. The system adjusts its state and completes the user'sdesired action as soon as the transaction is confirmed. Users can connect with system on the Ethereum blockchain in a secure fashion thanks to the authentication procedure offered by the Metamask wallet. The user's private keysare managed by Metamask, which also offers a simple user interface for transaction confirmation and signing. The Ethereum blockchain guarantees decentralized and trustless transaction processing, giving consumers a safeand dependable authentication method.

5.1 Module Description

This web application has four working modules: the user login module, the add to cart module, the payment module, and the user review module.

User Login module:

- To access their account on an online store, a user must log in using their own credentials, such as a password.
- The user can access their account, which may contain features like customized shopping recommendations, order tracking, stored goods, and access to special specials, once the login information has been validated by the website's authentication system.
  The login procedure is a crucial security feature because it guards against unwanted access to a user's personal data and online buying activities.

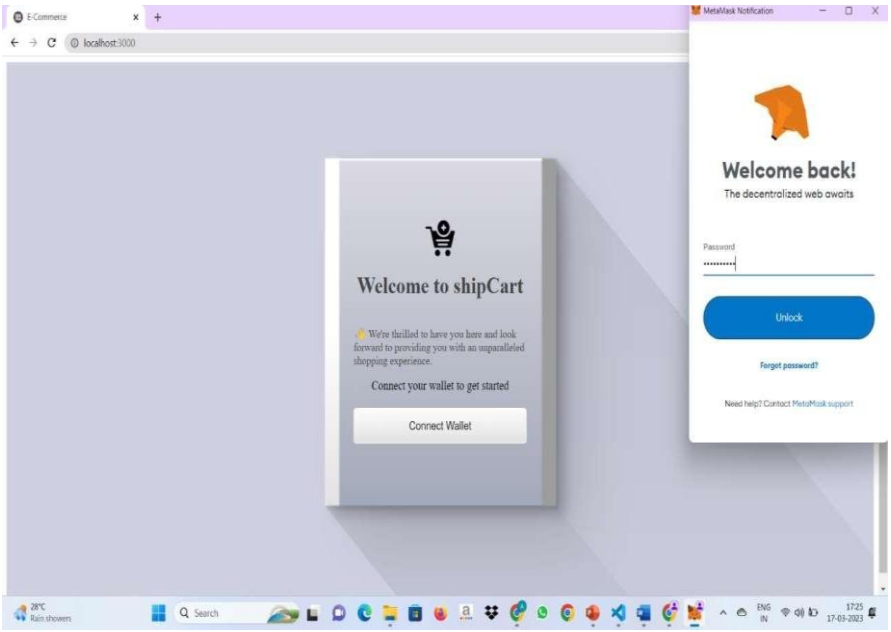

Fig. 5.1.1 User login Module

- The Ship Cart website offers a page called "see products" that provides details on a selection of the itemsthat may be purchased there.
- The product name, description, price, photos, and customer reviews are often included on this page.
- The see products page could also include other, suggested products that the client might find interesting.
- Customers are given all the information they require on the see products page so they can decide forthemselves whether or not to buy the product.

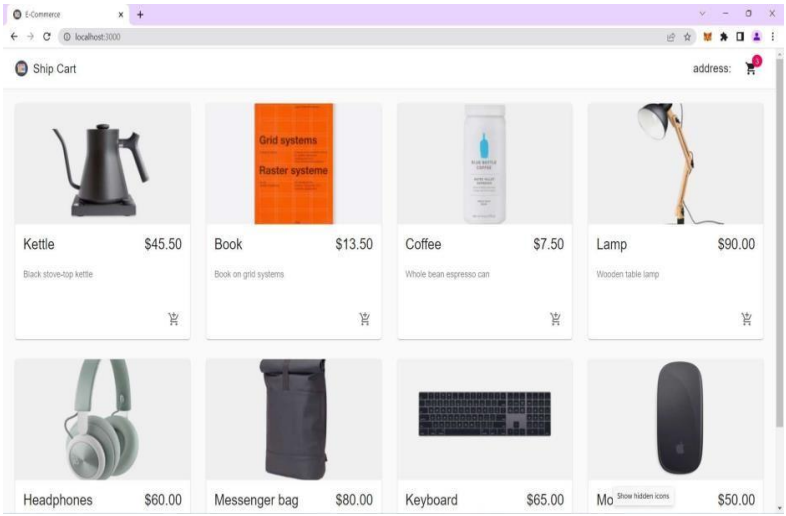

Fig. 5.1.2. E-commerce website

Add to cart Module:

- When shopping online, adding items to your cart entails choosing the items you want to buy and addingthem to a virtual basket or cart that symbolizes your upcoming transaction.
- The items you have added are normally listed in alphabetical order on the trolley page, along with theirnames, quantities, prices, and total costs.
- Before finishing your purchase, you can also remove products from your trolley or change the quantityas necessary.
- Putting items in your trolley is a crucial stage in the online purchasing process because it enables you toreserve items while you surf the website and make selections.

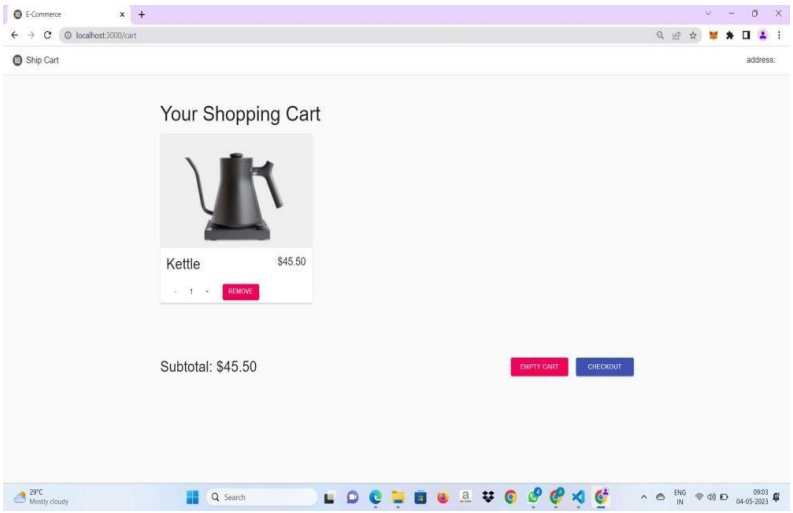

Fig. 5.1.3 Add to Cart Module

Payment Module:

- It establishes a link between the e-commerce website and the payment gateway, which serves as amiddleman between the website and the financial institution handling the payment processing.
- Payment modules often include a variety of payment methods, such as bank transfers, e-wallets, creditcards, and debit cards.
- It enables customers to make purchases and businesses to receive payments quickly and securely, bothof which are necessary for fostering client confidence and boosting sales.

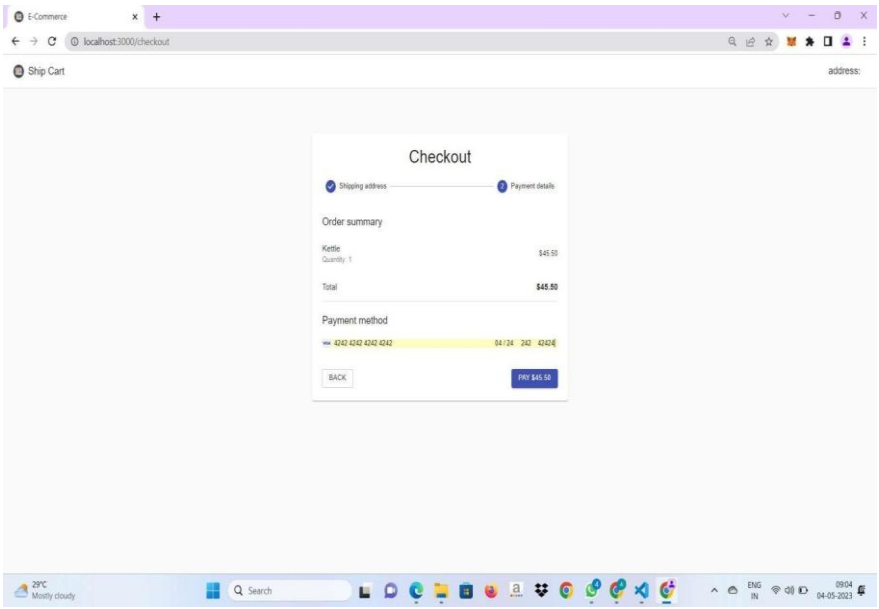

Fig. 5.1.4. Payment Module

User review module:

- In this module, the system allows only buyers to leave feedback on the product they have purchased. It is a tool that assists in giving potential customers who are thinking about buying the goods useful input.
- The goal of the user review module is to give prospective consumers who are thinking about buying theproduct insightful comments. They can gain a sense of the product's performance and quality, as well asany potential flaws or problems they should be aware of, by reading reviews from other consumers.
- Customers can normally evaluate products on a scale of one to five stars in the review module, and thereis also usually a place where they can write a lengthy evaluation of their experience using the product. Also, customers can add pictures or videos to support their reviews.
- An open environment that will help them draw in and keep customers over time by giving users a placeto express their ideas and opinions.

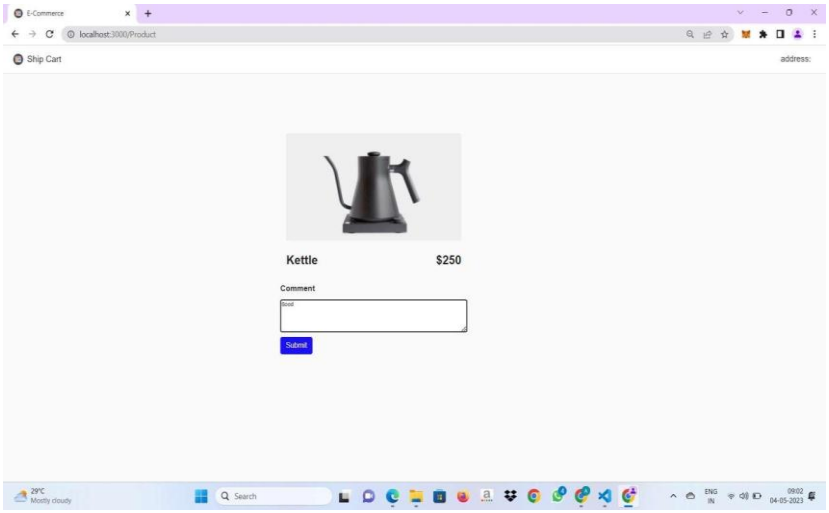

Fig. 5.1.5. User review module

6 Result

- This system is an online shopping portal, that connects users and the E-Commerce website owner.
- As this system uses Ethereum blockchain algorithm, it provides a secure authentication.
- This system enables easy access to every product on the site, which contains product description foreach product, that makes it easier for the user to choose the desire product on the site.
- Only a buyer would be able to comment on a product, that makes a genuine comment on the product.
- A buyer can also return the product within the warranty time of a product.
- As the comments are genuine in our system, that boost up the sales for the site owner.

7 Conclusions

Throughout the research, it is observed that fake reviews have been hard to handle. After a study, most of the researchers have been using many algorithms and its method to detect and remove the fake comments from the review session. we have proposed a methodology using the Ethereum blockchain technique to avoid fake comments from the spammers. The result proves 90% of the fake reviews can be prevented form the business industry.

Our technology provides more accuracy and trustful information on the product to the users, which increases the popularity of the quality product. In future work, we can test the performance of this proposed system by installing it across various shopping portals. And also, we can get some inputs from the customers side through a survey or from a sort of data collection techniques to make it user-friendly.

References

[1] Zhang Shunxiang; Zhu Aoqiang; Zhu Guangli; Wei Zhongliang; Li KuanChing(2023) "Building Fake Review Detection Model Based on Sentiment Intensity and PU Learning" doi: 10.1109/TNNLS.2023.3234427

[2] Thi-Kim-Hien Le & Yi-Zhen Li1 & ShengTun Li (2022), "Do Reviewers' Words and Behaviours Help Detect Fake Online Reviews and Spammers? Evidence From a Hierarchical Model", doi: 10.1109/ACCESS.2022.3167511, South Vietnam.

[3] Mujahed Abdulqader & Abdallah Namoun & Azed Alsaawy (2022), "Fake online reviews: A unified detection model using deception theories", doi: 10.1109/ACCESS.2022.3227631, Saudi Arabia.

[4] Noureddine Seddari & Abdelouahid Derhab & Mohamed Belaoued & Waleed Halboo & Jalal Al-Muhtadi & Abdelghani Bouras (2022), "A Hybrid Linguistic and Knowledge-Based Analysis Approach for Fake News Detection on Social Media", doi: 10.1109/ACCESS.2022.3181184.

[5] Hina Tufail & M. Usman Ashraf & Khalid Alsubhi & Hani Moaiteq Aljahdali (2022), "The Effect of Fake Reviews on e-Commerce During and After Covid-19 Pandemic: SKL-Based Fake Reviews Detection", doi: 10.1109/ACCESS.2022.3152806, Arab.

[6] L. Chen and Q. Li, "How Does fake review Influence E-Commerce Platform Revenue?," 2022 23rd Asia- Pacific Network Operations and Management Symposium (APNOMS), Takamatsu, Japan, 2022, pp. 1- 4, doi: 10.23919/APNOMS56106.2022.9919902.

[7] R. Agarwal and D. K. Sharma, "Detecting Fake Reviews using Machine learning techniques: a survey," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022, pp. 1750-1756, doi: 10.1109/ICACITE53722.2022.9823633.

[8] V. P. Sumathi, S. M. Pudhiyavan, M. Saran and V. N. Kumar, "Fake Review Detection Of E-Commerce Electronic Products Using Machine Learning Techniques," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 2021, pp. 1-5, doi: 10.1109/ICAECA52838.2021.9675684.

[9] P. Devika, A. Veena, E. Srilakshmi, A. R. Reddy and E. Praveen, "Detection of fake reviews using NLP &Sentiment Analysis," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 2021, pp. 1534-1537, doi: 10.1109/ICCES51350.2021.9489210.

[10] D. Agnihotri, S. Ahmed, D. Darekar and C. Gadkari, "Examination of Fake Reviews of Product and Approach to Filter them from Legitimate Reviews," 2021 IEEE Pune Section International Conference (PuneCon), Pune, India, 2021, pp. 1-4, doi: 10.1109/PuneCon52575.2021.9686477.

[11] Meiling Liu & Yue Shang & Qi Yue & Jiyun Zhou (2021), "Detecting Fake Reviews Using Multidimensional Representations With FineGrained Aspects Plan", doi: 10.1109/ACCESS.2020.3047947, China.

[12] Jianrong Yao & Yuan Zheng & Hui Jiang (2021), "An Ensemble Model for Fake Online Review Detection Based on Data Resampling, Feature Pruning, and Parameter Optimization", doi: 10.1109/ACCESS.2021.3051174, China.

[13] Songkai Tang & Luhua Jin & Fan Cheng (2021), "Fraud Detection in Online Product Review Systems via Heterogeneous Graph Transforme", doi: 10.1109/ACCESS.2017.DOI, China.

[14] Rami Mohawesh & Shuxiang Xu & Son N. Tran & Robert Ollington & Matthew Springer & Yaser Jararweh & Sumbal Maqsood (2021), "Fake Reviews Detection: A Survey", doi: 10.1109/ACCESS.2021.3075573, Australia.

[15] Matthew Carter & Michail Tsikerdekis & Sherali Zeadally (2021), "Approaches for Fake Content Detection: Strengths and Weaknesses to Adversarial Attacks", doi: 10.1109/MIC.2020.3032323, USA.
[16] Punit Rathore & Jayesh Soni & Nagarajan Prabakar & Marimuthu Palaniswami & Paolo Santi (2021), "Identifying Groups of Fake Reviewers Using a Semisupervised Approach", doi: 10.1109/TCSS.2021.3085406.
[17] Jingdong Wang & Haitao Kan & Fanqi Meng & Qizi Mu & Genhua Shi & Xixi Xiao (2020), "Fake Review Detection Based on Multiple Feature Fusion and Rolling Collaborative Training", doi: 10.1109/ACCESS.2020.3028588, China.
[18] Wenqian Liu & Jingsha He & Song Han & Fangbo Cai & Zhenning Yang & Nafei Zhu (2019), "A Method for the Detection of Fake Reviews Based on Temporal Features of Reviews and Comments", doi: 10.1109/EMR.2019.2928964, China.
[19] Estée Van Der Walt & An Eloff (2018), "Using Machine Learning to Detect Fake Identities: Bots vs Humans", doi: 10.1109/ACCESS.2018.2796018, South Africa.
[20] Lu Zhang & Zhiang Wu & Jie Cao (2018), "Detecting Spammer Groups From Product Reviews: A Partially Supervised Learning Model".
doi: 10.1109/ACCESS.2017.2784370, China