Developing a Credible and Trustworthy E-Commerce Application using Blockchain and Machine Learning

Baby V[1], Avinash Reddy G[1*], Yogendra Sai K[1], Sree Harshitha K[1], Padmini Chowdary N[1]

[1] VNR Vignana Jyothi Institute of engineering and technology, Vignana Jyothi Nagar, Pragathi Nagar, Nizampet (S.O), Hyderabad, Telangana 500090
avinash150602@gmail.com

Abstract. Electronic Commerce or E-commerce is the buying or selling of goods online, over the internet. Nowadays, people are spending most of their daily life online. A business of any scale, be it a start-up or a well-grown company, can benefit from E-commerce. It helps them reach a broader consumer base and reduce operational costs. From a customer's point of view, it is convenient for the customer to shop from their home, without visiting several shops before buying a product. Also, they can compare various products across various brands easily, with the help of their computer or their Mobile Phone. E-Commerce applications have some associated problems. One major issue is that a seller might dispatch and deliver a fake product to a customer, who had already paid for the product. Though there are mechanisms such as 'Return & Replace the product', the fake products need to be identified first. This can be addressed using a Blockchain-based Supply Chain. Each product will have a QR Code, which a manufacturer provides, and at each step in the product's movement during delivery to the customer, the components of the supply chain such as the Delivery Agency can update their details associated to the QR Code and at any point of time, the QR Code can be scanned to verify the product's authenticity. One more issue is that products of the same type may have a huge variation in their prices, due to a lot of sellers selling the same type of product. So, it is helpful to predict the price a product deserves, so that a unified cost for that product can be achieved and it gives a potential customer a good idea of the estimated-deserved price for the product. The quality of a product can be estimated based on customers' reviews. This is done using various Machine Learning approaches. Also, in an e-commerce application, the customers might give spam reviews - they might glorify the product, or try to degrade the product, different from what the product actually deserves. Such spam reviews are identified using Machine Learning & Natural Language Processing approaches.

Keywords: Block chain, Price Suggestion, NLP, Fake Product, Spam Reviews, Product Quality Detection.

C. Kiran Mai et al. (eds.), *Proceedings of the Fourth International Conference on Advances in Computer Engineering and Communication Systems (ICACECS 2023)*, Atlantis Highlights in Computer Sciences 18,
https://doi.org/10.2991/978-94-6463-314-6_23

1 Introduction

E-commerce, often known as electronic commerce, is the exchange of goods and services through a network, such as the internet. The products are real and tactile if a user chooses to purchase them offline, allowing them to be touched, felt, and judged for quality prior to purchase. Everyone still wants to purchase goods online in the current digital world when the goods are practically represented by photographs and videos. Users therefore face a variety of issues, including the inability to tell if a product is real or fraudulent, the inability to assess the product's quality, and the lack of knowledge on the appropriate pricing; moreover, the product review section of the website could include spam. Before establishing an order for a product with an online retailer, the user analyzes the product characteristics, however the buyer runs the risk of receiving fake products. For instance, the customer may have ordered PUMA shoes, but when they are delivered, they are not PUMA shoes despite looking just like them. As the standard of the goods given to them by the fake branding is worse than that of the actual product, such experiences cause people to lose trust in the brand. Therefore, authentic brands lose clients without making mistakes. Therefore, the issue of selling counterfeit goods is bad for both consumers and producers. In order for other users to read the reviews and make a purchasing decision, users may also provide feedback on the products they have purchased. Therefore, product makers may create a large number of false accounts only to provide nice evaluations for the items in order to enhance sales, and ultimately, if the product itself is subpar, people might fall for the trap of purchasing a low-quality item. Comparably, rival businesses may set up a number of fictitious accounts only to post a lot of unfavorable comments about the product in an effort to reduce sales. The producers have the power to set their own prices, which provides them the advantage of being able to sell the goods for far more than their rivals are willing to. Users could fall victim to this scam by spending a lot of money on a good only to later discover it for less money. These are the issues that both the producers and the customers are having. Users lose faith in the brands, producers lose clients, and both parties suffer financial losses as a result. In the absence of procedures to address such issues, individuals may cease purchasing goods online.

2 Literature Survey

Wasnik, K. [1] in which manufacturer adds the product details to the database and generates a QR code, and sends it with the product to the supplier. Next, the supplier scans the QR and adds his details, and sends this product to the customer. Finally, the customer scans the QR to check the authenticity of the product. They have used ReactJS for the frontend, the blockchain that is used is Ethereum, web3.js is used for connecting the JavaScript code with the Ethereum blockchain, MetaMask valet is used to manage the accounts, and MySQL database to store the data.

Shreekumar, T. [2] developed a mobile application, used dart for the frontend, Node.js for writing the server-side code, and the database used is Firebase and they have used Ethereum-based blockchain and solidity for writing the smart contracts,

web3.js is used to connect the JavaScript code with the blockchain. They have used the QR-Code-based system to determine the authenticity of the products. A notification verifying the legitimacy of the goods will be given to the consumer if the QR Code matches; else, a notification stating that the product is false or counterfeit will be issued.

In Jambhulkar, S. [3] various users involved are Manufacturers, Distributors, Retailers, and Customers. The manufacturer will add the data to the blockchain, distributor and retailer will update the data in the blockchain by adding their information. Whenever the user scans the QR Code, they will be able to access all the information, from the manufacturer to the retailer. The front end is developed using HTML, and CSS and they have used PHP for server-side code, and the database used is MySQL.

Ma, J., Lin [4] developed smart contracts using solidity. The data is stored in Geth, which is a private Ethereum-based blockchain and can be accessed using Metamask. Manufacturers will include information about themselves and the seller, such as the seller's address and the quantity of products that can be sold from a certain seller. Then the product is sent to the supplier by attaching a QR code to it. The supplier will send this product to a specific user who has bought it. This application also provides the feature of exchanging the product where the manufacturer initially verifies the identity of the user and then the exchange process will be initiated by changing the status in the blockchain.

Tejaswini Tambe [5] developed a mobile application. They used firebase as their database to store the data, Used SHA 256 algorithm to generate the unique QR code for each product, and these QR codes are scanned using the built-in feature present in the mobile application which they have developed without using a separate application only for the purpose of scanning the QR codes. When the QR codes are scanned it will provide a label indicating “fake product” if the product is fake, or it will be labelled as “received” if the product is not fake.

Zehtab-Salmasi, A. [6] proposed several forecasting models to predict the price range of cell phones, based on their specifications. Based on the price range threshold, classes are assigned. Out of several models developed, CNN clubbed with Inception-V3 with Convolutional Image Feature Extraction and Dense Concatenation gave the best results, with an F1-Score of 88% on the CD18 dataset.

Mahoto, N. A. [7] proposed a method, to decide the price of the product which the seller should sell for, based on the customer’s buying behaviour, which is a Business Intelligence-Machine Learning fusion model. They built three models: Random Forest, Logistic Regression, and One-vs-All model. The One-vs-All model had a mean Recall and Precision of around 90%.

Han, L. [8] proposed a model that provided price suggestions for non-identical items across different categories. First, based on product images, it is classified as qualified or not. If found not qualified, the seller will have to give a text description. Finally, based on the photographs and text description, a pricing recommendation is made for the product. RMSLE (Root Mean Squared Logarithmic Error) for the project was a respectable 0.69.

In Fathalla, A. [9] for estimating the price of a used item based on the picture and written description of the item for various item kinds, a deep model architecture composed of LSTM & CNN was proposed. The model obtained a respectable R^2 score of 0.77.

Katarya, R [10] proposed CapsMF, which is a combination of Capsule Network and Matrix Factorization. Capsule network uses embedding layer and Bi-Directional Gated recurrent Unit for representation of textual descriptions of users and items and the Matrix Factorization is used to generate the improved recommendations. The cold start problem with recommender systems was addressed by this approach.

Hwangbo, H [11] developed a model called K-recSys, which is an extension to the existing collaborative filtering recommendation system. It considered both online and offline preferences of the users to generate the recommendations. When this model is tested in the actual operating environment then it is found that the model adopted substitute recommendations more frequently than the complementary recommendations.

Hendrawan, R [12] proposed a method to assess the quality of reviews as they are helpful to the user in making a buying decision. The quality of the review is determined based on 3 characteristics: structural, readability and metadata. The weighted sum is calculated for the reviews and they are sorted by considering the final score in the order of highest to lowest.

In Singla, Z., [13] the sentiment of reviews are analyzed as they provide insights about the performance of the products and customer satisfaction. The dataset is collected from Amazon and an undersampling method is used to deal with the data imbalance problem.They have trained various classification models like Naive Bayes, Decision trees, and SVM.The accuracy of the models is identified using 10-fold cross-validation. The highest accuracy is achieved for SVM which is 81.75% and Naive Bayes has the worst accuracy.

Maranzato, R., [14] aimed at addressing the issue where few sellers try to deceive the reputation systems in e-commerce for their benefit. This approach describes certain characteristics in transactions that indicate fraud. Then they incorporated some other possible fraud characteristics and logistic regression is applied to both the datasets. The improved set containing other potential fraud characteristics performed better than the unimproved set with logistic regression, it increased the number of fraudsters identified by 110%.

Hooi, B., [15] primarily focuses on identifying users' or products' suspicious behaviour when it deviates from other accepted practices by building a Bayesian model for rating behaviour and developing a measure for spam identification. This method uses two alternative approaches: one where a single product receives multiple reviews from various users of the same text, and the other where a single person provides multiple reviews on other goods with identical reviews. When the model is run on the Flipkart dataset, it successfully detects fraud reviews in significant real-world applications with a precision of 84%. When this was applied to 250 of the application's most suspect users of Flipkart, 211 of them actually submitted reviews of fraud.

Liu, Y., [16] developed a hierarchical attention network in order to extract and analyse sentence representation through word embeddings, They used n-gram CNN to extract the multi-granularity and informative sentence representation, and they used BI-

LSTM and convolution structures to extract complete data as well as the history of sentences. The model starts by comprehending the review's texture. On some datasets, this method's detection accuracy (F1) increased by 5% when it was evaluated on mixed and cross-domain datasets.

Shahariar, G., [17] approaches by performing both traditional machine learning algorithms as well as deep learning techniques and compared both the performances and concluded that deep learning techniques were better. Here they also used three different deep learning classifiers like Multilayer perceptron, RNN(LSTM) and CNN. They used both labelled and unlabelled datasets preprocessed them using NLP methods and by active learning algorithm converted the unlabelled to labelled data and for feature selection TF -IDF for MLP and Word Embeddings for both CNN and LSTM. For the classification all the three classifiers performed well with high accuracies as compared to other existing techniques and among the three LSTM has higher accuracy.

3 Proposed System

The flow of the proposed system is depicted in the below figure 1:

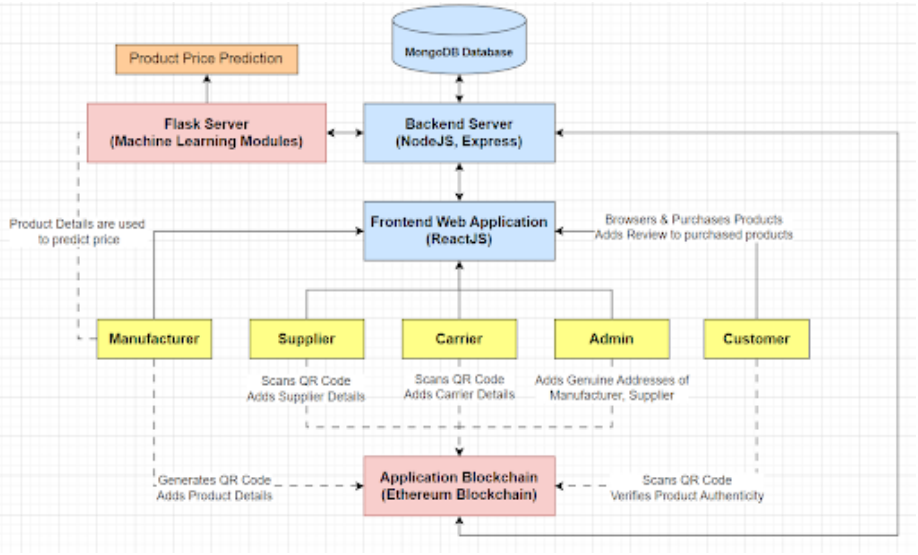

Fig. 1. System Architecture

This proposed system considers 5 types of users: manufacturer, supplier, carrier, customer, and admin.

Manufacturers will enter the details of themselves and the product which they have manufactured but their address will be fetched automatically by using the GeoLocation Web API. This Web API will fetch the latitude and longitude of the user's current location. When the manufacturer enters the details, a QR Code will be generated for that

particular product. This QR Code will be attached to the product and it will be shipped to the supplier. Next, Supplier will scan the QR Code present on the product, then they will be able to see the manufacturer details and they will enter their own information but again the address will be fetched automatically using the same mechanism. Next, when a customer orders the product, the carrier will pick up the product and they will also scan the QR Code and add their details and the pick-up date and when the product is being delivered, they will also add the delivered date. Finally, when the customer receives the product, they will scan the QR code present on the product that is delivered then the details of the whole supply chain along with the authenticity of the product is displayed. Admin will be adding the addresses of manufacturers and suppliers, these addresses are the actual addresses which the admin verifies with the manufacturer and supplier before adding them to the blockchain and the addresses which are fetched automatically will be matched with these addresses to determine whether the manufacturer and supplier are fake or not.

For Product Price Prediction, the dataset consists of Fashion products having attributes such as Description, Brand, Color, Gender and the target variable as Price. With such a dataset, the following methodology is used to predict the price given the other attributes. First, data is preprocessed and features are extracted from the Description and Brand, Gender and Color fields using techniques such as TF-IDF Vectorization and Count Vectorization. Then, a Blended/Stacked Machine Learning Model is trained, by using Light Gradient Boosting Machine (LGBM) and Ridge Regressor. 57% of the predictions are allotted to the LGBM Model and 43% to the Ridge Regression model. Finally, the model is evaluated using metrics such as R^2 Score, Mean Absolute Error, etc. For the fusion model the R^2 score for training data is 84.5% and for test data it is 83%. Mean Absolute Error (MAE) for training data is 441 and for test data it is 468.

For Product Quality Prediction from reviews and Spam review classification, the following method is proposed. For this approach we have used a dataset with 300000 records with equal split for both spam and non-spam reviews i.e., 150000 records are not spam and 150000 records are spam. The reviews are first preprocessed using TF-IDF vectorization then a Logistic Regression Classifier is trained. The accuracy score for training data is 88% and for test data it is 86.7%. Then, Sentiment Analysis is performed on the reviews, with the target labels representing a rating range. Finally, all the sub-modules are integrated into a modern MERN Stack web application.

4 Experiment and Result

The below pictures are the screenshots of our E-Commerce web application built using React for frontend and Node.JS, Express.JS, Flask (Python), MongoDB for backend. We used Solidity for writing the smart contracts and the following features have been implemented:

4.1 Fake Product Detection using Block chain

By reading the QR code, this feature enables the customer to determine the product's authenticity. The buyer is shown the authenticity of the goods as well as the whole supply chain facts. Beginning with the manufacturer's information, supplier's information, and carrier's information, the entire product's information is kept in the blockchain. Authenticity of the product is given in figure 2 and figure 3.

Fig. 2. Authentic Product

Fig. 3. Not an authentic Product

4.2 Product Price Prediction

In order to compare prices and make an informed choice, the buyer can use this tool to forecast the product's normalized or deserved price. Utilizing the product data that the manufacturer enters, the price is predicted as shown in figure 4. Customers can view the estimated price when they view the product details because it is stored in the block chain.

Fig. 4. Predicted price

4.3 Spam Review Classification

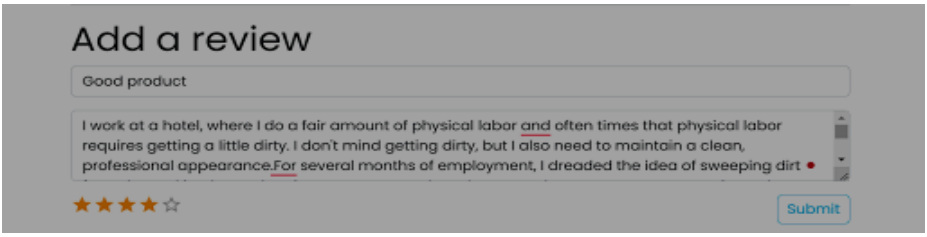

Fig. 5. Spam review

When a customer enters a review and clicks the submit button, the review content as given in Figure 5 is provided to our machine learning model, which predicts whether the review is spam or not. If the review is determined to be spam, the user will see the popup as seen in Figure 6, and the review will be removed.

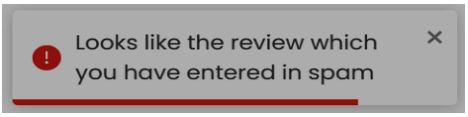

Fig. 6. Popup displaying that the review is spam

4.4 Quality Prediction from Reviews

Fig. 7. Quality Prediction

The percentage of consumers who considered our product useful is displayed as seen in figure 7 by analyzing the sentiment of the reviews that customers have left. For the analysis, only reviews that aren't labelled spam are taken into account.

5 Conclusion And Future Scope

The usage of online shopping has increased drastically so it is required to ensure that the products that are being sold online are not fake and the suppliers are trustworthy. So, we have developed a credible and trustworthy e-commerce application and reached the following objectives: Using Block chain to store the products and trace the

supply chain to overcome counterfeit goods and fraud sellers' problem, using QR Codes & Geolocation Web API. Predicting the price of products using a Blended Machine Learning Model of Light Gradient Boosting Machine (LGBM) and Ridge Regression. Giving an estimate of the quality of a product based on customer reviews, by making use of a Deep Learning LSTM Model. Classify reviews as Spam or Not, by using a Logistic Regression Classifier. The future scope of this product is to include Linguistic & Behavioral features of a custom while predicting the quality from reviews and to detect spam reviews, have a more robust Machine Learning model on a feature rich dataset to deal with price variations in a better way.

References

1. Author, Wasnik, K., Author, Sondawle, I., Author, Wani, R., Author, Pulgam, N.: Detection of Counterfeit Products using Blockchain. In: ITM Web of Conferences (Vol. 44, p. 03015). EDP Sciences. (2022)
2. Author, Shreekumar, T., Author, Mittal, P., Author, Sharma, S., Author, Kamath, R. N., Author, Rajesh, S., Author, Ganapathy, B. N.: Fake Product Detection Using Blockchain Technology. In: Journal of Algebraic Statistics, 13(3), 2815-2821. (2022)
3. Author, Jambhulkar, S., Author, Bhoyar, H., Author, Dhore, S., Author, Bidkar, A., Author, Desai, P.: Blockchain based Fake product Identification System. In: International Research Journal of Modernization in Engineering Technology and Science, 2582-5208. (2021)
4. Author, Ma, J., Author, Lin, S. Y., Author, Chen, X., Author, Sun, H. M., Author, Chen, Y. C., Author, Wang, H.: A blockchain-based application system for product anti-counterfeiting. In: IEEE Access, 8, 77642-77652. (2020)
5. Author, Tambe, T., Author, Chitalkar, S., Author, Khurud, M., Author, Varpe, M., Author, Raut, S. Y.: Fake product detection using blockchain technology. In: International Journal of Advance Research, Ideas and INNOVATIONS in Technology, 7, 314-319. (2021)
6. Author, Zehtab-Salmasi, Author, A., Feizi-Derakhshi, Author, A. R., Nikzad-Khasmakhi, Author, N., Asgari-Chenaghlu, Author, M., & Nabipour, S.: Multimodal price prediction. In: Annals of Data Science, 1-17. (2021)
7. Author, Mahoto, N. A., Author, Iftikhar, R., Author, Shaikh, A., Author, Asiri, Y., Author, Alghamdi, A., Author, Rajab, K.: An intelligent business model for product price prediction using machine learning approach. In: Intelligent Automation & Soft Computing, 29(3), 147-159. (2021)
8. Author, Han, L., Author, Yin, Z., Author, Xia, Z., Author, Guo, L., Author, Tang, M., Author, Jin, R.: Price Suggestion for Online Second-hand Items. In: 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 5920-5927). IEEE. (2021)
9. Author, Fathalla, A., Author, Salah, A., Author, Li, K., Li, K., Author, Francesco, P. (2020).: Deep end-to-end learning for price prediction of second-hand items. In: Knowledge and Information Systems, 62(12), 4541-4568. (2020)
10. Author, Katarya, R., Author, Arora, Y.: Capsmf: a novel product recommender system using deep learning based text analysis model. In: Multimedia Tools and Applications, 79(47), 35927-35948. (2020)
11. Author, Hwangbo, H., Author, Kim, Y. S., Author, Cha, K. J.: Recommendation system development for fashion retail e-commerce. In: Electronic Commerce Research and Applications, 28, 94-101. (2018)

12. Author, Hendrawan, R. A., Author, Suryani, E., Author, Oktavia, R.: Evaluation of e-commerce product reviews based on structural, metadata, and readability characteristics. In: Procedia Computer Science, 124, 280-286. (2017)
13. Author, Singla, Z., Author, Randhawa, S., Author, Jain, S.: Sentiment analysis of customer product reviews using machine learning. In: 2017 international conference on intelligent computing and control (I2C2) (pp. 1-5). IEEE. (2017)
14. Author, Maranzato, R., Author, Pereira, A., Author, do Lago, Author, A. P., Author, Neubert, M.: Fraud detection in reputation systems in e-markets using logistic regression. In: Proceedings of the 2010 ACM symposium on applied computing (pp. 1454-1455). (2010)
15. Author, Hooi, B., Author, Shah, N., Author, Beutel, A., Author, Günnemann, S., Author, Akoglu, L., Author, Kumar, M., Author, Faloutsos, C.: Birdnest: Bayesian inference for ratings-fraud detection. In: Proceedings of the 2016 SIAM International Conference on Data Mining (pp. 495-503). Society for Industrial and Applied Mathematics. (2016)
16. Author, Liu, Y., Author, Wang, L., Author, Shi, T., Author, Li, J.: Detection of spam reviews through a hierarchical attention architecture with N-gram CNN and Bi-LSTM. In: Information Systems, 103, 101865. (2022)
17. Author, Shahariar, G. M., Author, Biswas, S., Author, Omar, F., Author, Shah, F. M., Author, Hassan, S. B.: Spam review detection using deep learning. In: 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 0027-0033). IEEE. (2019)