Journal of Advances in Developmental Research (IJAIDR)
E-ISSN: 0976-4844 ● Website: www.ijaidr.com ● Email: editor@ijaidr.com

E-Commerce Fraud Detection Based on Machine Learning

N. Shravani[1], N. Renuka[2], M. Pavan Naik[3], B Sree Theja[4], S.Siva Sankar[5]

[1,2,3,4,5]Department Of CSE, Tadipatri Engineering College , Tadipatri

ABSTRACT:
Nowadays, there are so many applications available on internet because of that user cannot always get correct or true reviews about the product on internet. In this project, we propose the system by developing web application which help to detect fraud apps using sentiment comments and data mining. We can check for user's sentimental comments on multiple application. The reviews may be fake or genuine. But after comparing reviews of admin as well as user's, we can get more clear idea. Hence, we can get higher probability of getting real reviews. So we are proposing a system to develop a web application that will take reviews from registered users for single product, and analyse them for positive negative rating. For every users reviews and comments will be fetched separately and analysed for positive negative rating. Then their rating/comments will be judged by the admin and it would be easy for admin to predict the application as Genuine or Fraud. In Review Based Evidences, besides ratings, most of the App stores also allow users to write some textual comments as App reviews. Such reviews can reflect the personal perceptions and usage experiences of existing users for particular mobile Apps. Indeed, review manipulation is one of the most important perspective of App ranking fraud.

Keywords: E-commerce fraud, machine learning, fraud detection, data mining, Random Forest, Decision Tree.

INTRODUCTION

The E-Commerce has become an essential part of modern shopping, enabling consumers to purchase products and share their experiences through online reviews and ratings. These reviews significantly influence customer purchasing decisions and help build trust between buyers and sellers. However, the rapid expansion of e-commerce platforms has also created opportunities for fraudulent activities, including fake reviews, manipulated ratings, spam comments, and coordinated review attacks. Such deceptive practices mislead customers, provide unfair advantages to certain sellers, undermine platform credibility, and ultimately weaken consumer confidence in online shopping systems. Traditionally, review fraud detection has relied on rule-based approaches such as keyword filtering, predefined user activity thresholds, and manual moderation. While these methods can identify simple or previously known fraud patterns, they lack flexibility and scalability. As fraudsters continuously adapt their strategies to evade detection, static rule-based systems struggle to identify new and complex manipulation techniques. This limitation results in lower detection accuracy, higher false positives, and an inability to respond effectively to evolving fraud behaviours.

Machine learning offers a more intelligent and adaptive solution to review fraud detection by leveraging large volumes of data to uncover hidden and complex patterns. By analyzing review text, reviewer

behavior, rating distributions, and user interaction networks, machine learning models can automatically identify suspicious activities that may not be evident through manual inspection or fixed rules. These models are capable of learning from historical data and continuously improving their performance as new fraud patterns emerge, leading to faster detection and more reliable results. This project proposes a machine learning–based framework for detecting fraudulent reviews in e-commerce platforms. A comprehensive feature extraction and analysis process is conducted to capture key indicators such as linguistic characteristics of review text, reviewer credibility and activity patterns, rating deviations, and temporal behavior. To address the common issue of class imbalance in review fraud datasets, appropriate data balancing techniques are applied to enhance the detection of minority fraudulent instances and improve overall model robustness.

Multiple machine learning algorithms are implemented and systematically evaluated using performance metrics such as accuracy, precision, recall, and F1-score. The experimental results demonstrate that the proposed approach achieves higher accuracy and significantly reduces false positives compared to traditional rule-based methods. By improving the reliability of review systems, this framework contributes to increased consumer trust, fair competition among sellers, and a more transparent and dependable e-commerce environment.

LITERATURE REVIEW

In[1] This paper by Mutemi and Bacao is a systematic literature review analyzing how machine learning (ML) techniques are applied to detect fraud in e-commerce. It reviews 101 publications to identify research gaps, trends, and the increasing use of artificial neural networks in combating digital, transaction-based fraud. [2] Zhouhang shao, xuran wang used heterogeneous graphs methods and amazon dataset in the GNN-EADD for dual stage learning and Perform extensive experiments on large dataset. To experiments on amazon demonstrate superior accuracy, precision and recall compared to existing methods. [3] Emmanuel lleberi, yanxia sun are used the Random forest, SMOTE, AdaBoost for to identify spike, credit card fraud detection performance on imbalance datasets and the drawback is overfitting and increase the computational cost and it simplifies them with 99% accuracy. [4] Ebenezer Esenogho, The G swart are used the k- reverse nearest neighbour (KRNN) method to eliminate the extreme outliers from the fraud class in A neural network ensemble with feature engineering for improved credit card fraud detection, it detects the fraud with accuracy and precision of 98.40% and 97.34%.

[5] Elias Dritsas, Maria Trigka are used the machine learning techniques across key e commerce domains in machine learning e commerce : trends and future challenges. It study about the scalability, interpretability and cold start issues and detect fraud. The main drawback is performance comparisions. [6] Bertrand Lebichot, Liyun He-Guelton are used the Fraud Detection system (FDS) method in Transfer learning for credit card fraud detection for the most commonly used the input features are not cause of output binary class. The main drawback is negative tranfer transaction in online, it improves the detection accuracy. [7] Suyuan Luo, Shaohua wan, are uses the Bigdata in Leveraging product characteristics for online collusive detection it plays a crucial role to identify the fraud in user characteristics. It can solve the continued and growing problems by adopting data mining techniques. It depends on mostly rich and accurate produce metadata.[8] Xiaonan Sun, Yuan Cao are uses the SWIFT format in standardized interface frameworks for intelligent financial platforms: A pre standardized study for established messaging standards such as ISO 2002. The standardized interface framework for intelligent financial platforms(SIFFP). The drawback is real world implementations.

[9] Seyedeh khadijeh hashemi, Sergio Greco used the lightgbm, XGBoost and catBoost with Bayesian methods in fraud detection banking data by machine learning techniques. This approach involves high computational cost and complex hyper parameters.it detects the fraud banking reviews.[10] Yadong Zhou, Dae wook kim, lili liu, Hongbo jin, are found the new platforms to host the variety of business activities such as online promotion videos in Pro guard: detecting accounts in social based networks . Experimental results have demonstrated the detection rate of 96.67% at a very low false positive rate of 0.3%.[11] Josus Genaro, Jose Antonio used the random forest and logistic regression models in Hyphatia: A card-not-present fraud detection system based on self-supervised tabular learning for to worked on existing studies on credit cards. [12] Arbena musa, kamer vishi used the deep learning methods, worked on Our digitals traces in cybersecurity for to improve the security challenges.it provide more security in digital systems. [13] Samin M. ABD-ALHALEM, NAGLAA F. SOLIMAN used CNN methods in advancing E-Commerce Authenticity: A novel fusion approach based on deep learning and aspect features for detecting false reviews worked for to detect the fake review with 97.73% accuracy rate.[14] JAQUELINE D. DUARTE, PEDRO CHAGAS JUNIOR, JOÃO PAULO JAVIDI DA COSTA are used k-fold cross validation in Machine Learning for Early Detection of Phishing URLs in Parked Domains. An Approach Applied to a Financial Institution. They worked on LightGBM-based framework effectively detects phishing URLs early especially new and parked domains achieving ~97% accuracy using SSL and domain-based features. The drawback is performance may drop on highly imbalanced patterns. [15] M. Salimi and P. Fränti are used Machine learning techniques in Joint Use of Time Series and Graph Data for Fake Comment Detection in CafeBazaar Dataset. They worked on an LSTM autoencoder–based model detects fake comments by identifying anomalous user behavior over time, achieving 99% precision and removing 1.9M fake reviews on Cafebazaar. The drawback is making it less effective for user with sparse or short activity sequences.

PROPOSED METHODOLOGY

The proposed system aims to detect fraudulent reviews in e-commerce or mobile application platforms using sentiment analysis and machine learning techniques. User reviews and ratings for a specific product or application are collected through a web application, where only registered users are allowed to submit feedback to ensure data reliability. The collected reviews are then preprocessed by removing duplicates, eliminating stop words, tokenizing the text, converting it to lowercase, and handling missing values to improve data quality. After preprocessing, sentiment analysis is performed on each review to classify it as positive or negative and to identify abnormal patterns such as overly positive or repetitive reviews. Important features such as sentiment score, rating value, review length, and frequency of user reviews are extracted and used as input for machine learning models. Classification algorithms like Decision Tree and Random Forest are applied to categorize reviews as genuine or fraudulent. The system also provides an admin interface where analyzed reviews, sentiment results, and fraud predictions are displayed, enabling the admin to make a final decision on whether the application or product is genuine or fraudulent. The performance of the proposed system is evaluated using metrics such as accuracy, precision, and recall, and the results show improved effectiveness compared to traditional rule-based fraud detection methods.

Decision Tree: Decision Trees classify data by learning hierarchical decision rules from input features. The model works by recursively splitting the dataset into smaller subsets based on feature values that best separate different classes, using measures such as Gini Index or Information Gain. The final classification is obtained by following a path from the root node to a leaf node. Decision Trees are widely used due to

their simplicity and interpretability, as the tree structure clearly shows how decisions are made at each step. They can model non linear relationships and interactions among features without requiring complex data preprocessing. However, Decision Trees may be sensitive to noise and prone to overfitting, particularly when trained on large or imbalanced datasets. These limitations can be reduced using pruning methods or ensemble techniques such as Random Forests.

Random Forest: Random Forest is an ensemble learning method that builds multiple decision trees and combines their predictions, typically through majority voting for classification tasks. By aggregating the outputs of many trees trained on different subsets of data and features, it improves overall prediction accuracy and reduces the risk of overfitting compared to a single decision tree.The algorithm introduces randomness through bootstrapped sampling of the training data and random selection of features at each split, which increases diversity among the trees and enhances robustness. Random Forest can handle high-dimensional data, capture complex and non-linear relationships, and is relatively resistant to noise. Additionally, it provides feature importance metrics, which help identify which variables contribute most to predictions. The main limitations are higher computational requirements and reduced interpretability compared to a single decision tree.

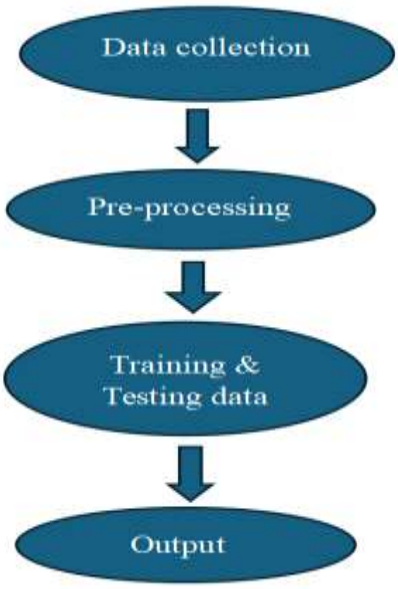

Fig:1. E-Commerce Fraud Detection Modules

Fig:1.1 DATA COLLECTION & PROCESSING

Fig:2. Data Collection: The data collection in machine learning is the crucial process for to gathering diverse customer and products data like browsing history, purchase records, demographics, product

attributes from sources like website interactions, CRMs, and analytics to train the models for personalization. It can collect the data from the user for to process the information

Fig: 1.2 Data Pre- Processing

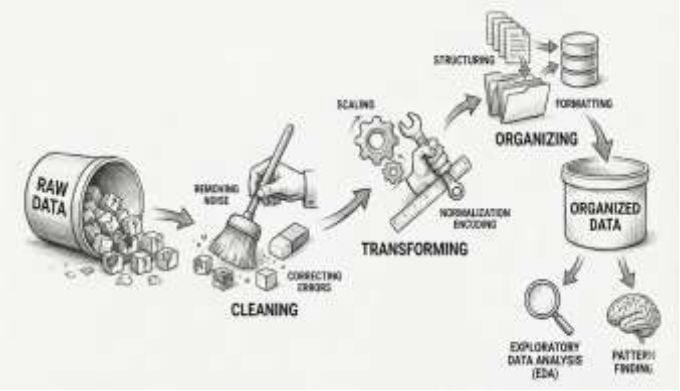

Fig:1.2. Data Pre-Processing: The data pre-processing is analyzes the data or information of users data. It is the first step in any data analysis. It involves cleaning, transforming and organizing raw data. It organized data supports better Exploratory Data Analysis(EDA), maing patterns.

Fig:1. Training and Testing data: The data we used is divided into two important parts : Training data and Testing data. Training data is the dataset used to teach a machine learning model. During training, the model looks at input and outputs pairs.

Fig:1. Output: The final result is generated, such as classifying reviews as genuine or fraudulent and providing insights for fraud prevention.

SYSTEM ARCHITECTURE

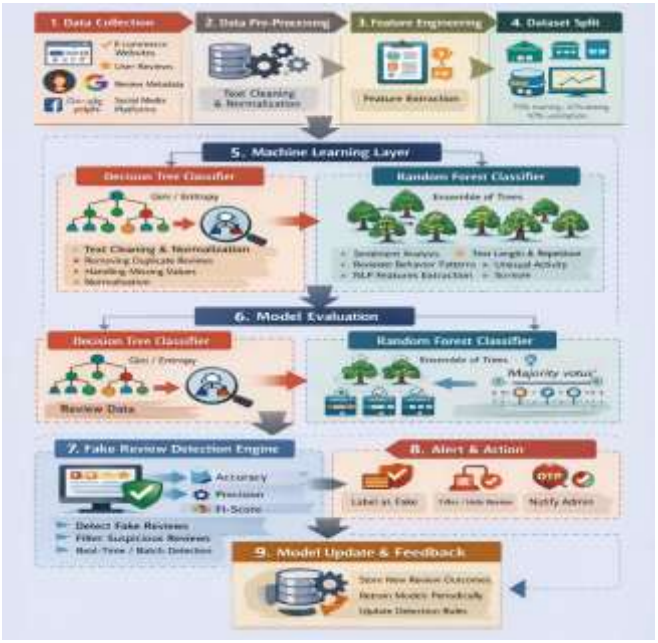

FIG 3.SYSTEM ARCHITECTURE

This system provides an intelligent approach to detect fake reviews in e-commerce platforms using machine learning techniques. It collects user reviews from online sources, preprocesses and extracts

meaningful features from the data, and applies Decision Tree and Random Forest classifiers to identify fraudulent reviews. The system evaluates model performance using standard metrics and automatically suspicious reviews for further action. Through continuous feedback and model updates, the solution improves accuracy over time, helping maintain trust and reliability in online review systems.

RESULT AND DISCUSSION

The proposed machine learning based fraud detection framework was evaluated on an e-commerce transaction dataset containing both legitimate and fraudulent records. Due to the severe class imbalance typically present in fraud data, imbalance handling techniques were applied to improve minority class detection. Experimental results show that ensemble based models outperformed single classifiers. Random Forest and achieved higher recall, indicating improved detection of fraudulent transactions. This improvement is critical, as undetected fraud leads to direct financial losses. Feature analysis identified transaction amount, transaction frequency, and time based behave oral attributes as the most influential features in fraud prediction. Multiple machine learning models, Decision Tree and Random Forest, were implemented and evaluated using precision, recall and accuracy. These metrics provide a reliable assessment of performance in imbalanced classification problems. The results demonstrate that addressing class imbalance significantly reduces false negatives and improves overall model robustness. Compared to traditional rule based systems, the proposed machine learning approach provides better adaptability to evolving fraud patterns and improved detection accuracy. Although some false positives remain, this trade off is acceptable in fraud detection systems, where preventing fraud is prioritized over minimal user inconvenience. Overall, the proposed framework enhances the effectiveness and reliability of e-commerce fraud detection and supports safer online transaction environments.

SCREEN SHOTS

FIG 4.INDEX PAGE

FIG 6.LOGIN PAGE

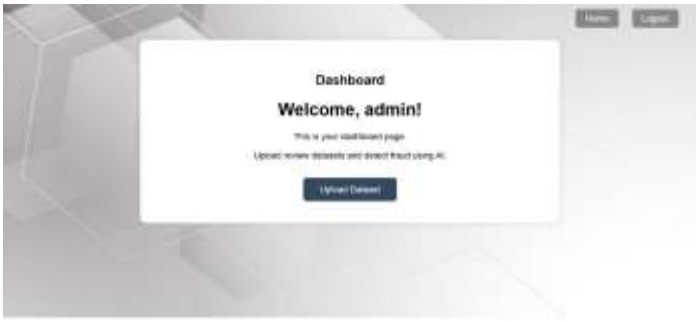

FIG 7.DASHBOARD

FIG 8.PREVIEW PAGE

FIG 9.PREDICTION PAGE

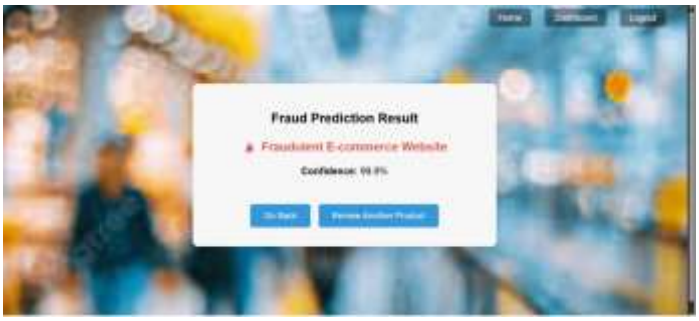

FIG 10.RESULT PAGE

CONCLUSION

In this project, a general machine learning based approach for e-commerce fraud detection has been presented to address the increasing problem of fraudulent activities in online platforms. With the rapid growth of e-commerce, detecting fraud accurately and efficiently has become essential to protect both consumers and businesses. The proposed system uses machine learning techniques to analyze user data

and user reviews based information to identify fraudulent activities. Compared to traditional rule based methods, machine learning models are more flexible and can adapt to changing fraud patterns. Algorithms such as Decision Tree and Random Forest were applied to classify reviews as genuine or fraudulent.The results show that machine learning models can effectively improve fraud detection accuracy while reducing the risk of undetected fraud. Handling imbalanced data and selecting important features play a key role in improving system performance. Although some false alerts may occur, this is acceptable in fraud detection systems where preventing financial loss is the primary goal. Overall, the proposed approach provides a reliable and efficient solution for e-commerce fraud detection. It helps build trust in online systems and ensures safest application reviews. In the future, the system can be further enhanced by using advanced algorithms and real time fraud detection techniques.

REFERENCES

1. Mutemi and F. Bacao, "E-commerce fraud detection based on machine learning techniques: systematic Literature review," in Big data mining and Analytics , vol. 7, no. 2, pp. 419-444, June 2024,doi:10.26599/BDMA. 2023.9020023.
2. Z. Shao, X. Wang, E.Ji, S.Chen and J. Wang, "GNN-EADD: Graph Neural Networks-Based E-Commerce Anomaly Detection via Dual-Stage Learning," in IEEE Access, vol. 13,pp.8963-8976, 2025,doi: 10.1109/ACCESS.2025.3526239.
3. E. lleberi, Y. Sun and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBOOST," in /EEE Access, vol. 9, pp. 165286- 165294, 2021 , doi: 10.1109/ACCESS. 2021.3134330.
4. E.Esenogho, i.d.mienye, T.G. Swart, K. Aruleba and G. Obaido, " A neural network ensemble with feature engineering for improved credit card fraud detection," In IEEE Access, vol.10,pp. 16400-16407, 2022,doi:10.1109/ACCESS. 2022.3148298.
5. E.Dritsas and M. Trigka ,"machine learning in e-commerce: Trends, applications and future challenges", In IEEE access, vol. 13,pp.99048-99067, 2025,doi: 10.1109/ACCESS. 2025.3572865.
6. B. Lebichot, T. Verhelst, Y. -A. Le Borgne, L. He-Guelton, F. Oble and G. Bontempi,"Trasfer learning for credit card fraud detection", In IEEE acess, vol. 9, pp. 114754-114766,2021, doi: 10.1109/ACCESS. 2021. 3104472.
7. S. Luo and S. Wan, "Leveraging product characteristics for online collusive detection in big data transactions.", in IEEE access, vol. 7,pp. 40154-40164, 2019, doi: 10.1109/ACCESS. 2019.2891907.
8. X. Sun, S. Yang, Y. Cao, Y, Zhao and Z. Wang, "Standardized interface Framework for intelligent financial platforms: A Pre-standardization study," in Journal of ICT Standardization, vol. 13., no. 2, pp. 181-210, June 2025, doi. 10.13052/jicts2245-800X.1325.
9. S. K. Hashemi, S. L. Mirtaheri and S. Greco,"Fraud detection in Banking Data by machine learning techniques," in IEEE Access, vol. 11, pp. 3034-3043, 2023, doi: 10.1109/ACCESS.2022.3232287.
10. Y.Zhou et al.,"ProGuard: Detecting accounts in social-network-based online promotions," in IEEE Access, vol. 5,pp. 1990-1999,2017, doi: 10.1109/ACCESS.2017.2654272.
11. J. G. Almaraz-Rivera, J. A. Cantoral-Ceballos, J. F. Botero, F. J. MuñOz and B. D. Martinez, "Hyphatia: A Card-Not-Present Fraud Detection System Based on Self-Supervised Tabular Learning," in IEEE Open Journal of the Computer Society, vol. 6, pp. 812-821, 2025, doi: 10.1109/OJCS.2025.3570600.

12. A. Musa, K. Vishi, E. Martiri and B. Rexha, "Our Digital Traces in Cybersecurity: Bridging the Gap Between Anonymity and Identification," in IEEE Access, vol. 13, pp. 46909-46924, 2025, doi: 10.1109/ACCESS.2025.3551095.
13. E. Ileberi and Y. Sun, "Advancing Model Performance With ADASYN and Recurrent Feature Elimination and Cross-Validation in Machine Learning-Assisted Credit Card Fraud Detection: A Comparative Analysis," in IEEE Access, vol. 12, pp. 133315-133327, 2024, doi: 10.1109/ACCESS.2024.3457922.
14. J. D. Duarte et al., "Machine Learning for Early Detection of Phishing URLs in Parked Domains: An Approach Applied to a Financial Institution," in IEEE Access, vol. 13, pp. 145736-145753, 2025, doi: 10.1109/ACCESS.2025.3599454.
15. M. Salimi and P. Fränti, "Joint Use of Time Series and Graph Data for Fake Comment Detection in CafeBazaar Dataset," in IEEE Access, vol. 13, pp. 159217-159230, 2025, doi: 10.1109/ACCESS.2025.3607849.