Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

E-Commerce Fraud Detection Based on Machine Learning Techniques

Kommu Manikanta, G.Rajamani, R.Vinay, P.Sowmya, C.Sandeep, N.Pavan Kumar

Department of Computer Science and Engineering Gates Institute of Technology , Andhra Pradesh, India.

Correspondence

Kommu Manikanta
Department of Computer Science and Engineering, Gates Institute of Technology, Andhra Pradesh, India.

- Received Date: 11 Feb 2026
- Accepted Date: 02 Apr 2026
- Publication Date: 25 Apr 2026

Keywords

E-commerce fraud detection, Machine learning, Cybersecurity, Data mining, Anomaly detection, Artificial neural networks, Digital marketplaces, PRISMA methodology, Predictive modeling, Fraud prevention

Abstract

Sarcasm The e-commerce industry's rapid growth, accelerated by the COVID-19 pandemic, has led to an alarming increase in digital fraud and associated losses. To establish a healthy e-commerce ecosystem, robust cyber security and anti-fraud measures are crucial. However, research on fraud detection systems has struggled to keep pace due to limited real-world datasets. Advances in artificial intelligence, Machine Learning (ML), and cloud computing have revitalized research and applications in this domain. While ML and data mining techniques are popular in fraud detection, specific reviews focusing on their application in e-commerce platforms like eBay and Facebook are lacking depth. Existing reviews provide broad overviews but fail to grasp the intricacies of ML algorithms in the e-commerce context. To bridge this gap, our study conducts a systematic literature review using the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) methodology. We aim to explore the effectiveness of these techniques in fraud detection within digital marketplaces.

Introduction

The rapid expansion of the e-commerce industry, significantly accelerated by the COVID-19 pandemic, has transformed the way consumers and businesses interact in the digital economy. Alongside this growth, however, there has been a substantial rise in fraudulent activities, including payment fraud, identity theft, and fake transactions. These challenges pose serious threats to both consumers and online platforms, leading to financial losses and reduced trust in digital marketplaces.

Traditional rule-based fraud detection systems are increasingly inadequate in handling the dynamic and evolving nature of cyber fraud. As a result, there is a growing need for intelligent and adaptive solutions. In this context, machine learning (ML) and data mining techniques have emerged as powerful tools for detecting and preventing fraudulent activities by analyzing large volumes of transactional data and identifying hidden patterns.

Recent advancements in artificial intelligence, neural networks, and cloud computing have further enhanced the effectiveness and scalability of fraud detection systems. However, despite the increasing number of studies in this domain, existing research often provides only general overviews and lacks a focused analysis of ML techniques specifically applied to e-commerce platforms. To address this gap, this study presents a systematic literature review using the PRISMA methodology, analyzing recent research on machine learning-based fraud detection in e-commerce. The objective is to evaluate current approaches, identify research gaps, and highlight emerging trends, thereby providing valuable insights for researchers and industry practitioners aiming to develop more robust and efficient fraud detection systems. In order to derive graphtheoretical variables or scores that specifically characterize nodes of fraud, the concept makes use of the connectedness between the nodes, which are often users or items in a dataset.

Literature Review

Fraud detection in e-commerce has gained significant attention over the past decade due to the rapid increase in online transactions and associated cyber threats. Early approaches primarily relied on rule-based and statistical methods, which were limited in detecting complex and evolving fraud patterns. With the advancement of machine learning, researchers have increasingly adopted data-driven techniques to improve detection accuracy and efficiency.

Supervised learning algorithms such as Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines have been widely used for classifying fraudulent and legitimate transactions. These models require labeled datasets and have shown

Citation: Kommu M, Rajamani G, Vinay R, Sowmya P, Sandeep C, Pavan Kumar N. E-Commerce Fraud Detection Based on Machine Learning Techniques. GJEIIR. 2026;6(4):245.

strong performance in structured environments. However, their effectiveness is often constrained by data imbalance and the scarcity of real-world fraud datasets.

To overcome these limitations, unsupervised and semi-supervised learning techniques, including clustering and anomaly detection, have been explored to identify unusual transaction behaviors without requiring labeled data. In recent years, deep learning models, particularly Artificial Neural Networks, have demonstrated promising results in capturing complex, non-linear patterns in large-scale e-commerce data.

Several studies have also emphasized the role of data preprocessing, feature engineering, and real-time detection systems in enhancing model performance. Additionally, the integration of cloud computing has enabled scalable and efficient processing of large datasets. Despite these advancements, existing literature often lacks a focused analysis of machine learning techniques specifically tailored for e-commerce platforms. Moreover, issues such as data privacy, high false positive rates, and adaptability to evolving fraud strategies remain key challenges. This highlights the need for more comprehensive and domain-specific research, motivating the present study.

Proposed System

The system is designed to identify fraudulent transactions in real time by analyzing user behavior and transaction patterns. Initially, transaction data is collected from e-commerce platforms, including features such as user details, transaction amount, time, location, and payment method. The collected data undergoes preprocessing steps such as data cleaning, normalization, and handling of missing values to ensure data quality. Feature engineering techniques are then applied to extract relevant attributes that contribute to fraud detection.

The system employs a hybrid machine learning approach by integrating supervised and unsupervised learning methods. Supervised algorithms such as Random Forest and Support Vector Machine are used for classification, while anomaly detection techniques are incorporated to identify unusual patterns in unlabeled data. Additionally, Artificial Neural Networks are utilized to capture complex and non-linear relationships within the dataset.

To address the issue of imbalanced datasets, techniques such as oversampling and undersampling are applied to improve model performance. The trained model is then deployed in a real-time environment, where incoming transactions are continuously monitored and classified as either legitimate or fraudulent. The system also includes a feedback mechanism that updates the model periodically based on new data, ensuring adaptability to evolving fraud patterns. Performance metrics such as accuracy, precision, recall, and F1-score are used to evaluate the effectiveness of the proposed model.

Research Method

The PRISMA approach generates high-quality results and supports reproducibility. It is structured in a manner that allows the identification and summarization of problems (domains), techniques, and methods used to solve the problem. The implementation of this approach follows a checklist of title, abstract, introduction, methods, results, discussion, and funding. In this structure, the title and abstract are constructed to achieve comparable objectives to any other approach, but the introduction must provide the rationale for the review and the questions to be addressed.

Data and search strategy:

1. By extracting potential search terms from the titles, abstracts, and subject indexing of three pertinent publications[17, 23, 24], we develop an initial search strategy. We use its results to expand the list of key words and restrict it to only English-language articles in order to further hone this strategy. We then test the validity of our search strategy by checking whether it could retrieve the three known relevant studies and two more studies referenced in Ref. [17]. All the five studies are successfully identified by the strategy. A group of peer reviewers approves the final search strategy.
2. Using an iterative search approach, we look for publications within our search period (2010−2023) that have the following keywords in their title or abstract: e-commerce, fraud detection, machine learning, systematic review, organized retail fraud, data mining, and digital marketplace. We display the iterative approach in the workflow diagram shown in Fig.
3. To reduce the amount of noise in the results, our search strategy employs the search logics “AND”, “OR”, “LIMIT TO”, and “EXCLUDE”.

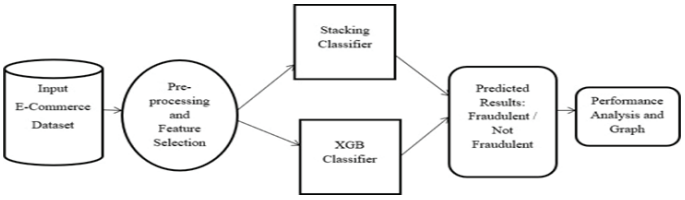
Feature Extraction

The relevant features are derived from multiple aspects of e-commerce transactions, including user behavior, transaction details, and device information. Transaction-based features include attributes such as transaction amount, frequency, time of transaction, and payment method. Behavioral features capture user activity patterns, such as purchase history, browsing behavior, and login frequency. Additionally, location-based features (e.g., IP address, geographic location) and device-related features (e.g., device ID, browser type) are considered to identify suspicious access patterns. Device id , browser type. are the features.

Model Architecture

The architecture consists of data input, preprocessing, feature extraction, model training, and prediction layers.

In the first stage, raw transaction data is collected from various sources and passed through a preprocessing layer, where noise removal, missing value handling, and normalization are performed. The processed data is then fed into the feature extraction module, which generates meaningful features representing user behavior, transaction patterns, and contextual information.



Model Training and Evaluation:

The training phase involves applying multiple machine learning algorithms, including Random Forest, Support Vector Machine, and Artificial Neural Networks. Each model is trained using the extracted features, and hyperparameter tuning is performed to optimize performance. Cross-validation techniques are also employed to ensure model reliability and prevent overfitting.

Once trained, the models are evaluated using the testing dataset. Performance is measured using standard evaluation metrics such as accuracy, precision, recall, and F1-score. Special

emphasis is placed on recall and precision, as they are critical in minimizing false negatives (missed frauds) and false positives (incorrect fraud detection).

Deployment and Real-World Testing

The deployment is carried out using a scalable cloud-based infrastructure, enabling efficient handling of large volumes of transaction data with low latency. The trained model is integrated into the transaction processing pipeline, where each transaction is evaluated instantly and classified as legitimate or potentially fraudulent. An API-based architecture is used to facilitate seamless communication between the e-commerce platform and the fraud detection system. When a transaction is initiated, relevant features are extracted and passed to the model, which generates a prediction score. Based on predefined thresholds, suspicious transactions are either flagged for manual review or automatically blocked. For real-world testing, the system is evaluated using live or near real-time transaction data to assess its practical effectiveness. Key performance indicators such as detection rate, false positive rate, response time, and system scalability are monitored. This phase helps in identifying potential issues such as model drift, latency challenges, and adaptability to new fraud patterns.

Requirements Gathering Stage

Output screens:

Screen-1:

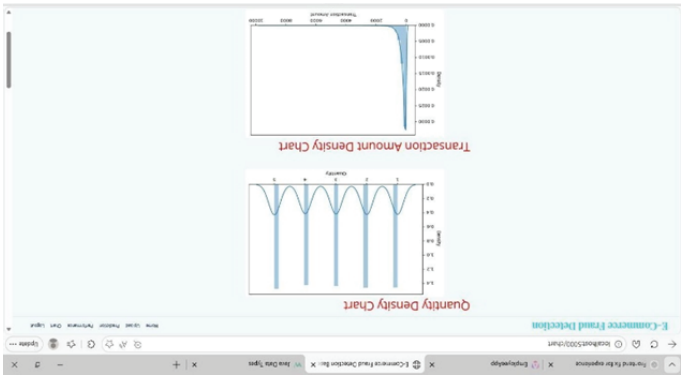
Screen-1: Login page

The developed system includes a user-friendly web-based interface for accessing the e-commerce fraud detection platform. As shown in the figure, the homepage provides a clean and simple login module where authorized users can enter their credentials (username and password) to access the system.

The interface is designed with usability and accessibility in mind, ensuring smooth interaction for administrators or analysts. Upon successful login, users can access the fraud detection dashboard, where transaction data can be monitored and analyzed in real time.

The front-end is integrated with the backend machine learning model, enabling seamless communication between the user interface and the fraud detection system. This ensures that user inputs and transaction data are processed efficiently, and results are delivered instantly.

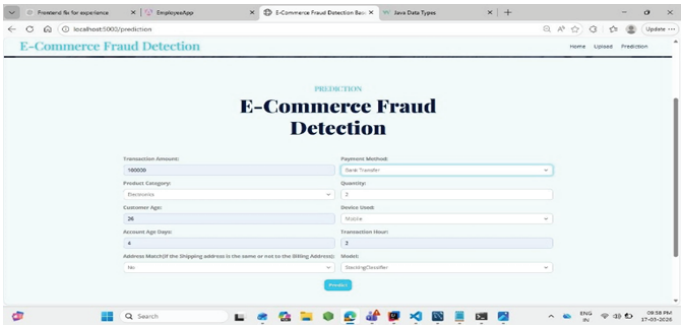
Screen-2:



Screen -2 : Upload Transaction Dataset page

The screenshot enables administrators or analysts to input bulk transaction data into the system for fraud analysis. Once the file is uploaded, the backend processes the data by performing preprocessing, feature extraction, and applying the trained machine learning model to detect fraudulent transactions.The design ensures ease of use with minimal steps, allowing users to quickly upload and analyze data. The integration between the front-end interface and the backend model ensures efficient processing and timely generation of result.

Screen-3:

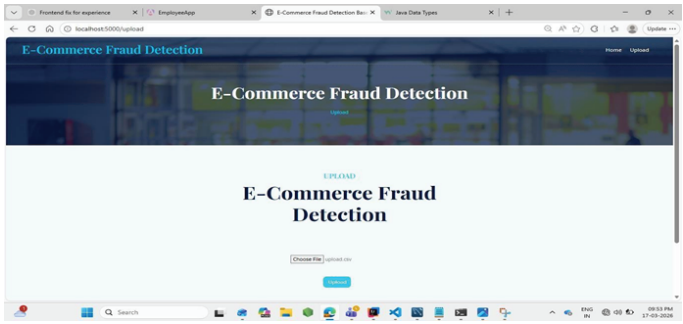


Screen-3: Transaction Prediction Input Page

The screen includes a prediction interface that allows users to input transaction details and obtain real-time fraud detection results. As shown in the figure, the interface consists of multiple input fields such as transaction amount, payment method, product category, customer age, device type, transaction time, and shipping details.Users can manually enter transaction information through this form, which is then processed by the backend machine learning model. Upon clicking the “Predict” button, the system analyzes the input data and classifies the transaction as either legitimate or fraudulent.The interface is designed to be interactive and user-friendly, enabling quick evaluation of individual transactions. It serves as a practical tool for testing the model and demonstrating its real-time prediction capability.

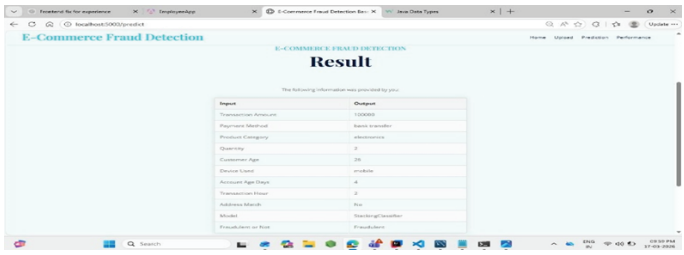
Screen-4:

The screen provides a data visualization module to analyze transaction patterns and support fraud detection. As illustrated in the figure, the interface displays graphical representations such as the Quantity Density Chart and Transaction Amount Density Chart.These visualizations help in understanding the



Screen-4: Quantity Density & Transaction Amount Density Charts

distribution and behavior of key features within the dataset. The Quantity Density Chart highlights the frequency distribution of purchased items, while the Transaction Amount Density Chart shows the variation and concentration of transaction values.Such graphical analysis aids in identifying anomalies and unusual patterns that may indicate fraudulent activities. By providing clear and interactive visual insights, this module enhances the interpretability of the data and supports better decision-making.

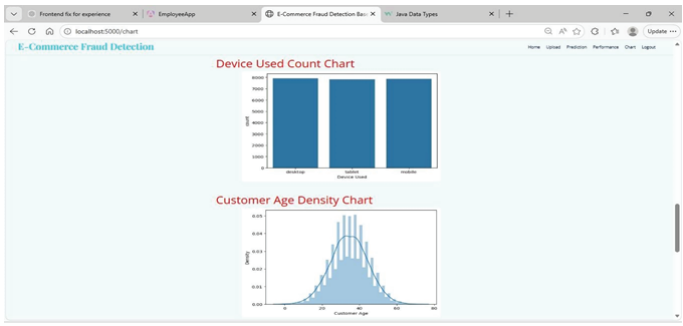


Screen 5: *Final Prediction Result Page*

The screen includes a result analysis interface that displays the outcome of the fraud prediction process. As shown in the figure, the interface presents both the input transaction details and the corresponding prediction output in a structured tabular format.

The input section includes parameters such as transaction amount, payment method, product category, customer details, and transaction characteristics. Based on these inputs, the machine learning model generates a prediction indicating whether the transaction is fraudulent or legitimate.

This clear presentation of results improves transparency and allows users to easily verify the relationship between input features and the model's decision. It also assists analysts in understanding the factors influencing fraud detection.



Conclusion

The aim of this work was to propose ways to extend the lexicon algorithm in order to build systems that would be more efficient for sarcasm detection. we employed a combined PRISMA and content synthesis approach to identify and analyze relevant articles focusing on fraud detection in the ecommerce domain using machine learning and data mining techniques. Our survey encompassed a total of 101 articles, with 16 of them classified as "other" due to being unclustered data mining techniques, while the remaining articles fell under the mainstream machine learning cluster.

To structure our analysis, we formulated four research questions, with the first two providing context for our main question. Among the machine learning algorithms utilized, ANNs emerged as the most frequently employed, closely followed by random forest. Notably, the majority of articles centered around the detection of credit card fraud, showcasing its prevalence in the field. However, we found a dearth of detailed research addressing reseller fraud, also known as product flipping or scalping, within our corpus, highlighting a potential avenue for future investigation given its significance in the e-commerce domain and potential impact on the economy and households. Further exploration of various techniques, including machine learning, to combat reseller fraud could be a fruitful area for future work.

Our review also shed light on emerging fraud types in e-commerce, namely triangulation and bot fraud, which have received limited attention in the realm of machine learning and data mining techniques. This observation underscores the need for further research to address these novel fraud types effectively.

Furthermore, our analysis revealed a growing demand for the application of imbalanced learning techniques to enhance future fraud detection systems. This indicates an opportunity for the concerted use of such techniques to tackle the challenge posed by imbalanced datasets in fraud detection.

References

1. F S. Monteith, M. Bauer, M. Alda, J. Geddes, P. C. Whybrow, and T. Glenn, Increasing cybercrime since the pandemic: Concerns for psychiatry, Curr. Psychiatry Rep., vol. 23, no. 4, p. 18, 2021.
2. S. Kodate, R. Chiba, S. Kimura, and N. Masuda, Detecting problematic transactions in a consumer-toconsumer e-commerce network, Appl. Netw. Sci., vol. 5, no. 1, p. 90, 2020.
3. R. Samani and G. Davis, McAfee mobile threat report, https://www.mcafee.com/enterprise/en-us/assets/reports/ rp-mobile-threat-report-2019.pdf, 2019.
4. E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature, Decis. Support Syst., vol. 50, no. 3, pp. 559–569, 2011.
5. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, et al., The PRISMA 2020 statement: An updated guideline for reporting systematic reviews, BMJ, vol. 372, p. n71, 2021.
6. . West and M. Bhattacharya, Intelligent financial fraud detection: A comprehensive review, Comput. Secur., vol. 57, pp. 47–66, 2016.
7. S J. Omar, K. Fred, and K. K. Swaib, A state-of-the-art review of machine learning techniques for fraud detection research, in Proc. 2018 Int. Conf. Software Engineering in Africa, Gothenburg, Sweden, 2018.
8. A. O. Adewumi and A. A. Akinyelu, A survey of machine-learning and nature-inspired based credit card fraud detection techniques, International Journal of System Assurance Engineering and Management, vol. 8, no. 2, pp. 937–953, 2017.

9. S. Sorournejad, Z. Zojaji, R. E. Atani, and A. H. Monadjemi, A survey of credit card fraud detection techniques: Data and technique oriented perspective, arXiv preprint arXiv: 1611.06439, 2016.
10. B. B. Sagar, P. Singh, and S. Mallika, Online transaction fraud detection techniques: A review of data mining approaches, in Proc. 3rd Int. Conf. Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2016, pp. 3756–3761.
11. A. Amir and G. Hamid, Fraudulent transactions detection in credit card by using data mining methods: review, https:// www.researchgate.net/publication/348732395_Fraudulent_ Transactions_Detection_in_Credit_Card_by_usi ng_Data_ Mining_Methods_A_Review, 2022.
12. Nagesh, C., Chaganti, K.R. , Chaganti, S. , Khaleelullah, S., Naresh, P. and Hussan, M. 2023. Leveraging Machine Learning based Ensemble Time Series Prediction Model for Rainfall Using SVM, KNN and Advanced ARIMA+ E-GARCH. International Journal on Recent and Innovation Trends in Computing and Communication. 11, 7s (Jul. 2023), 353–358.
13. DOI:https://doi.org/10.17762/ijritcc.v11i7s.7010.
14. S. Khaleelullah, P. Marry, P. Naresh, P. Srilatha, G. Sirisha and C. Nagesh, "A Framework for Design and Development of Message sharing using Open-Source Software," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 639-646, doi:10.1109/ ICSCDS56580.2023.10104679.
15. Ramesh Kumar Ramaswamy, Pannangi Naresh, Chilamakuru Nagesh, Santhosh Kumar Balan, Multilevel thresholding technique with Archery Gold Rush Optimization and PCNN-based childhood medulloblastoma classification using microscopic images, Biomedical Signal Processing and Control, Volume 107, 2025,107801, ISSN 1746-8094, https://doi.org/10.1016/j. bspc.2025.107801.
16. K. R. Chaganti, B. N. Kumar, P. K. Gutta, S. L. Reddy Elicherla, C. Nagesh and K. Raghavendar,"Blockchain Anchored Federated Learning and Tokenized Traceability for Sustainable Food Supply Chains," 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS), Gobichettipalayam, India, 2024, pp. 1532-1538, doi: 10.1109/ ICUIS64676.2024.10866271.
17. Mohammed Inayathulla and Karthikeyan C, "Image Caption Generation using Deep Learning For Video Summarization Applications" International Journal of Advanced Computer Science and Applications(IJACSA), 15(1), 2024. http:// dx.doi. org/10.14569/IJACSA.2024.0150155.
18. Mohammed, I., Chalichalamala, S. (2015). TERA: A Test Effort Reduction Approach by Using Fault Prediction Models. In: Satapathy, S., Govardhan, A., Raju, K., Mandal, J. (eds) Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1. Advances in Intelligent Systems and Computing, vol 337. Springer, Cham. https://doi. org/10.1007/978-3-319-13728-5_25
19. Inayathulla, M., Karthikeyan, C. (2022). Supervised Deep Learning Approach for Generating Dynamic Summary of the Video. In: Suma, V., Baig, Z., Kolandapalayam Shanmugam, S., Lorenz, P. (eds) Inventive Systems and Control. Lecture Notes in Networks and Systems, vol 436. Springer, Singapore. https://doi. org/10.1007/978-981-19- 1012-8_18.
20. Mohammed and C. Karthikeyan, "Object detection in video summarization for video surveillance applications," Yugoslav Journal of Operations Research, vol. 35, no. 3, pp. 523–546, 2025. doi:10.2298/YJOR240615010I.
21. Inayathulla Mohammed and K. R. Rao, "Enhancing real-time violence detection in video surveillance using hybrid deep learning model," Journal of Web and User Applications, vol. 16, no. 1, 2025. doi:10.58346/JOWUA.2025.I1.021.
22. G. Raju, K. Sushmitha, M. Priyanka, E. Sai Leela, M. Vani Chowdary, and A. Yashwantha, "Real Time Hand Gesture Recognition Using CNN," Global Journal of Engineering Innovations & Interdisciplinary Research (GJEIIR), vol. 5, no. 2, 2025, doi: 10.33425/3066-1226.1101.
23. G. Raju, K. Sattar Hadiya, K. Penna Obulesu, M. P. Pavan Raj, and M. Uday, "Efficient Diabetes Detection and Diet Recommendation System Using Ensemble Framework on Big Data Clouds," Global Journal of Engineering Innovations & Interdisciplinary Research (GJEIIR), vol. 5, no. 2, 2025, doi: 10.33425/3066-1226.1108.
24. G. Raju, L. R. Buchupalli, A. Thudimela, K. Kodikondla, K. Palaku, and D. Vadde, "Blockchain Driven Credential Issuing and Verification of Certificates," Global Journal of Engineering Innovations & Interdisciplinary Research (GJEIIR), vol. 5, no. 2, 2025, doi: 10.33425/3066-1226.1083.