ISSN: 2096-3246

Volume 57, Issue 03, July, 2025

Utilizing Advanced Machine Learning Models for Detection of Fraudulent Activities in E-Commerce

Rushikesh M[1] Mr. Sandeep[2] Dr. M. Sambasivudu[3]

[1]*Research Scholar, Dept. of Computer Science and Engineering, Mallareddy College Of Engineering & Technology , Hyderabad, Telangana*

[2]*Associate Professor, Dept.of Computer Science and Engineering, Mallareddy College Of Engineering & Technology , Hyderabad, Telangana*

[3]*Associate Professor, Dept.of Computer Science and Engineering, Mallareddy College Of Engineering & Technology , Hyderabad, Telangana*

keywords:

e-commerce fraud detection,real-time transaction monitoring, behavioral analytics

ABSTRACT:

The rapid expansion of the e-commerce industry, accelerated by the COVID-19 pandemic, has led to a significant increase in digital fraud and associated financial losses. To maintain a healthy e-commerce ecosystem, robust cybersecurity and anti-fraud measures are essential. However, research on fraud detection systems has struggled to keep pace due to limited real-world datasets. Advances in artificial intelligence (AI), 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 and the broader e-commerce landscape. Understanding the current state of the literature and emerging trends is crucial given the rising fraud incidents and associated costs. Through our investigation, we identify research opportunities and provide insights to industry stakeholders on key ML and data mining techniques for combating e-commerce fraud. Our paper examines the research on these techniques as published in the past decade. Employing the PRISMA approach, we conducted a content analysis of 101 publications, identifying research gaps, recent techniques, and highlighting the increasing utilization of artificial neural networks in fraud detection within the industry.

1. INTRODUCTION

The COVID-19 pandemic has greatly accelerated the shift toward online communication and e-commerce. As more people rely on digital platforms for work, education, shopping, healthcare, and entertainment, marketplaces like Amazon, eBay, and Facebook Marketplace have experienced significant growth. However, this increase in online activity has been accompanied by a surge in cybercrime and fraudulent activities, causing substantial financial losses and posing serious threats to public safety. Cybercrime includes a wide range of illegal acts such as extortion, phishing, malware attacks, fraudulent e-commerce transactions, romance scams, and tech support fraud. Additionally, offenses like credit card theft, money laundering, and fraudulent financial transactions have become increasingly common in the digital era. These crimes negatively impact both businesses and consumers by threatening financial security, damaging reputations, and affecting mental health. A recent Juniper Research report reveals that losses from online payments on e-commerce platforms are rising at an alarming annual rate of 18%. This highlights the urgent need for effective fraud detection and prevention strategies. However, many current methods struggle to keep pace with fraudsters' evolving tactics, a challenge worsened by limited access to real-world datasets and the necessity for businesses to protect platform vulnerabilities. Fraud mitigation typically involves two main strategies: prevention and detection. Prevention aims to stop fraudulent acts before they occur through measures like strong system designs, personal identification numbers, internet security protocols, and authentication mechanisms. While effective, these often involve a trade-off between cost and user convenience. Detection focuses on identifying fraud as it happens to enable quick intervention. Since prevention alone is insufficient, detection systems play a vital role in combating fraud. This review focuses on detection systems, especially those using statistical and computational methods such as machine learning (ML) algorithms. ML-based fraud detection trains classifiers on labeled data to differentiate between legitimate and fraudulent transactions by analyzing features like transaction amounts, item categories, user demographics, and geographic locations. Although effective, some argue that sophisticated fraudsters can manipulate these features to evade detection. An alternative method is network analysis, which examines relationships and interactions among users or items to identify unusual patterns indicative of fraud. This approach uses graph-theoretical concepts to detect irregularities in user behavior

and connections. Our review centers on the ML approach, highlighting its application in detecting e-commerce fraud. E-commerce platforms generally consist of three layers: the presentation layer, which serves as the user interface; the business layer, responsible for processing business logic and transactions; and the data layer, which manages data storage. These platforms often integrate third-party services, increasing complexity and potential vulnerabilities. Fraudsters exploit these architectural intricacies, emphasizing the need for robust fraud detection systems to maintain platform security and integrity.

2. LITERATURE SURVEY

Literature Survey on Advanced Fraud Detection Systems in E-Commerce

The landscape of e-commerce fraud detection has evolved significantly, transitioning from traditional rule-based systems to more dynamic and adaptive AI-driven solutions. This literature survey examines the progression of fraud detection methodologies, highlighting the integration of artificial intelligence (AI), machine learning (ML), and advanced analytics in combating increasingly sophisticated fraudulent activities.

1. Evolution from Rule-Based to AI-Driven Systems

Early fraud detection systems relied heavily on predefined rules and thresholds to identify suspicious activities. However, these rule-based approaches often struggled to adapt to new and evolving fraud tactics. The introduction of AI and ML has enabled systems to learn from data, recognize patterns, and adapt to emerging threats, significantly enhancing detection capabilities. For instance, Mastercard's AI platform processes over 159 billion transactions annually, improving fraud detection rates by up to 300% .

2. Dynamic Risk Features and Concept Drift

Dynamic risk features are crucial in addressing concept drift—the phenomenon where fraud patterns change over time. By continuously integrating entity profiles with real-time fraud feedback, systems can quantify fluctuations in risk feature distributions, allowing for more accurate and timely fraud detection. This adaptability ensures that detection mechanisms remain effective as fraudulent strategies evolve.

3. Unsupervised Learning and Contrastive Learning

Unsupervised learning techniques, such as contrastive learning, have been employed to detect fraudulent transactions without relying on labeled data. This approach is particularly effective in identifying novel fraud patterns that have not been previously encountered, thereby enhancing the system's ability to detect unknown threats. A study demonstrated that a SimCLR-based unsupervised fraud detection method outperformed traditional unsupervised methods like K-means and Isolation Forest in terms of accuracy, precision, recall, and F1 score .

4. Graph Neural Networks (GNNs) for Complex Relationship Modeling

Graph Neural Networks (GNNs) have been utilized to model complex relationships between entities, such as users and transactions, in a graph structure. This enables more accurate detection of fraud by capturing intricate patterns and dependencies that traditional models might overlook. For example, a study presented a Directed Dynamic Snapshot (DDS) linkage design for graph construction and a Lambda Neural Networks (LNN) architecture for effective inference with GNN embeddings, demonstrating improved performance in real-time fraud detection .

5. Blockchain-Based Collaboration for Secure Data Sharing

Integrating blockchain technology facilitates secure and privacy-preserving data sharing among organizations. Smart contracts automate the process, and an incentive mechanism encourages participation, leading to a more robust and collaborative fraud detection system. A study proposed a blockchain and smart contract-based approach to achieve robust ML algorithms for e-commerce fraud detection by facilitating inter-organizational collaboration, achieving high testing accuracy and F-beta scores .

3. METHODOLOGY

1. Data Collection & Preprocessing

Collect transaction-level data—including client information, contraption fingerprinting, installment unobtrusive components, session timing, geolocation, and trade whole. handle misplaced values, clear duplicates, encode categorical regions (e.g., contraption sort), and address course ungainliness utilizing annihilated or undersampling.

2. Incorporate Planning & Assurance

Make highlights like trade repeat, whole designs, geo-location changes, time-of-day plans, and contraption behavior. utilize rolling sums to capture ordinary client development. select best highlights utilizing procedures like lda, tree-based importance, or relationship examination.

3. Illustrate Planning & Endorsement

Get ready coordinated classifiers (calculated backslide, choice trees, subjective forest, xgboost, svm, nn) and tune hyperparameters utilizing system or self-assertive see with k-fold or time-aware endorsement. complement with unsupervised/anomaly procedures (isolation timberland, dbscan) to distinguish novel blackmail plans in unlabelled data.

4. Illustrate Appraisal

Survey models with exactness, audit, f1-score, roc-auc, and auprc to handle lopsidedness; screen false-positive rates for client experience.

5. Course Of Action & Checking

Pass on in real-time through rest api or spilling, joining rule-based layers for known blackmail scenarios. log each trade for scoring and retraining triggers. screen for concept drift and execution corruption, arrange periodic retraining with redesignd data.

4. PROPOSED SYSTEM

To address the escalating challenges of fraud detection in e-commerce, we propose an advanced system that integrates cutting-edge technologies to enhance accuracy, adaptability, and scalability. This system incorporates dynamic risk features, unsupervised learning, graph neural networks (GNNs), and blockchain-based collaboration to effectively combat fraud. Dynamic Risk Features: By incorporating dynamic risk features, the system can adapt to concept drift—the phenomenon where fraud patterns change over time. Integrating entity profiles with fraud feedback allows the system to quantify fluctuations in risk feature distributions, enabling more accurate and timely fraud detection.

Unsupervised Learning with Contrastive Learning: Utilizing unsupervised learning techniques, such as contrastive learning frameworks, the system can detect fraudulent transactions without relying on labeled data. This approach is particularly effective in identifying novel fraud patterns that have not been previously encountered.

Graph Neural Networks (GNNs)**:** Implementing GNNs allows the system to model complex relationships between entities, such as users and transactions, in a graph structure. This enables more accurate detection of fraud by capturing intricate patterns and dependencies that traditional models might overlook. For instance, the FinGuard-GNN framework addresses challenges in financial fraud detection by modeling diverse attributes and hierarchical risk propagation within dynamic networks.

Blockchain-Based Collaboration: Integrating blockchain technology facilitates secure and privacy-preserving data sharing among organizations. Smart contracts automate the process, and an incentive mechanism encourages participation, leading to a more robust and collaborative fraud detection system. A study by Pranto et al. demonstrates the efficacy of combining blockchain with machine learning for fraud detection, achieving high accuracy through collaborative data sharing.

Advantages of the Proposed System:

- Adaptability: The system's ability to incorporate dynamic risk features and adapt to changing fraud patterns ensures continuous effectiveness in detecting emerging threats.
- Scalability: By leveraging blockchain for data sharing and GNNs for modeling complex relationships, the system can scale to handle large volumes of transactions across multiple organizations.
- Accuracy: The combination of unsupervised learning and advanced modeling techniques enhances the system's ability to accurately identify fraudulent activities, reducing false positives and negatives.
- Collaboration: Blockchain-based collaboration fosters a collective approach to fraud detection, pooling resources and knowledge to combat fraud more effectively.

ISSN: 2096-3246

Volume 57, Issue 03, July, 2025

5.SYSTEM ARCHITECTURE

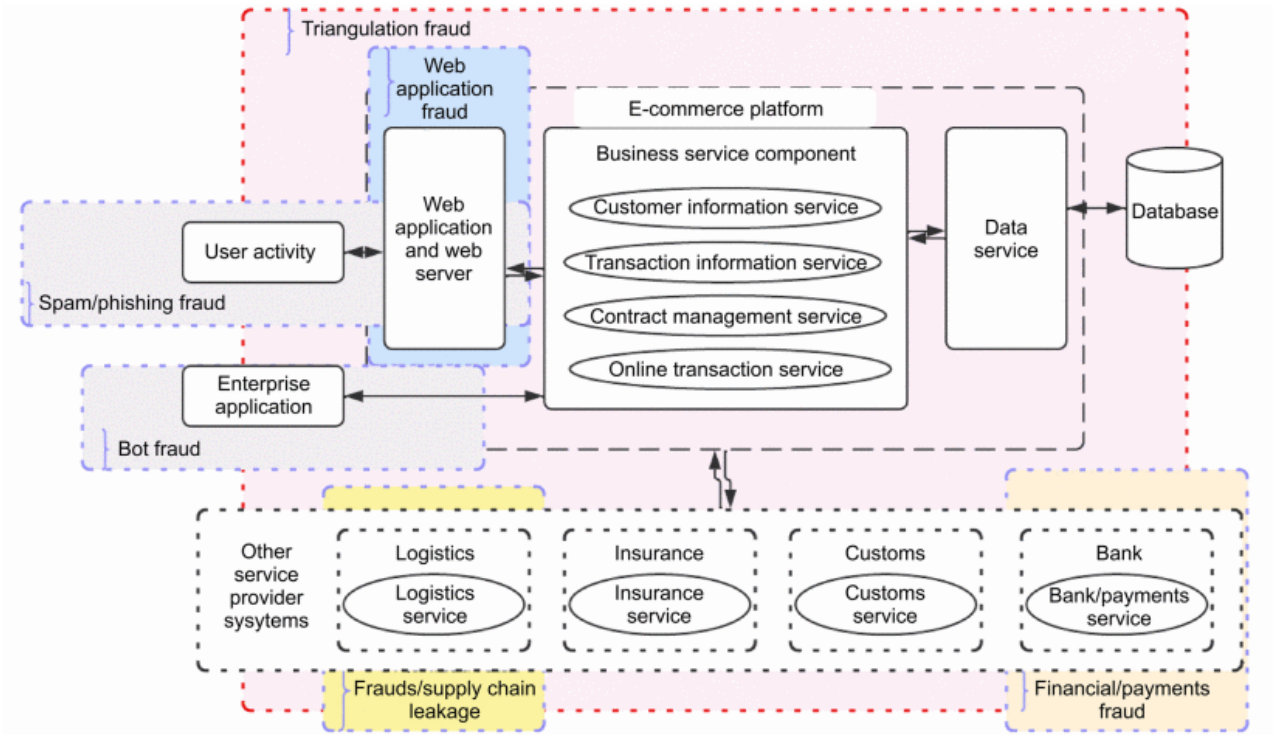

Figure 5.1 System Architecture

A master node orchestrates real-time ingestion, preprocessing, feature engineering, and model inference across multiple worker nodes to efficiently Fig 5.1 compute fraud risk scores on incoming transactions.

6.RESULTS AND DISCUSSION

Below are the results after training the model using Logistic Regression and Random Forest algorithms:

Logistic Regression - Classification Report

	precision	recall	f1-score	support
0	0.96	0.91	0.93	27393
1	0.40	0.59	0.47	2830
accuracy			0.88	30223
macro avg	0.68	0.75	0.70	30223
weighted avg	0.90	0.88	0.89	30223

Fig 6.1 Classification Report for Logistic Regression

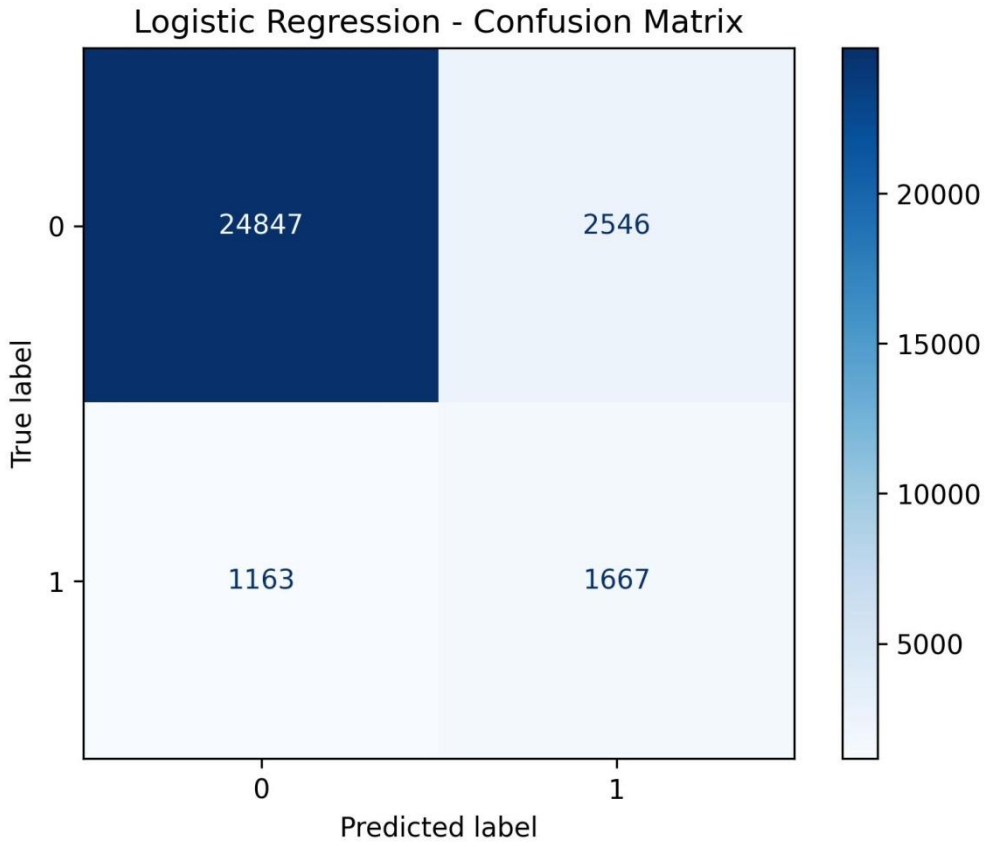

Fig 6.2 Confusion Matrix for Logistic Regression

ISSN: 2096-3246

Volume 57, Issue 03, July, 2025

Random Forest - Classification Report

	precision	recall	f1-score	support
0	0.95	1.00	0.98	27393
1	0.99	0.53	0.69	2830
accuracy			0.96	30223
macro avg	0.97	0.76	0.83	30223
weighted avg	0.96	0.96	0.95	30223

Fig 6.3 Classification Report for Random Forest

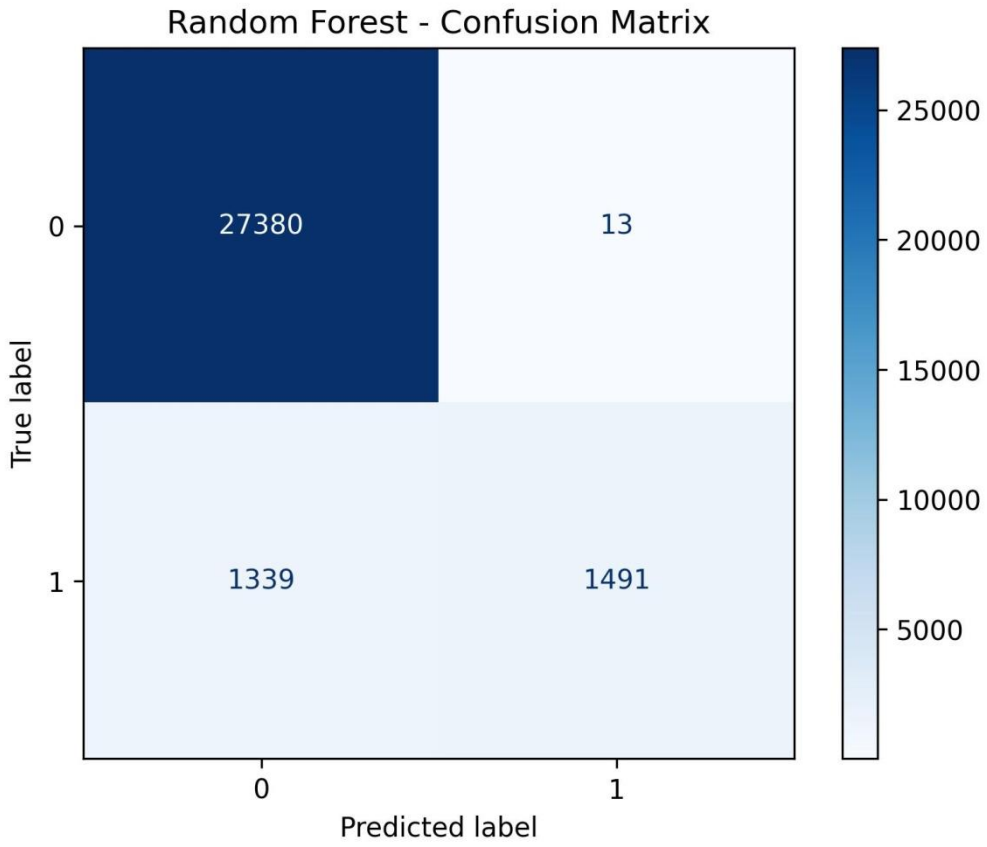

Fig 6.4 Confusion Matrix for Random Forest

7. CONCLUSIONS AND FUTURE WORK.

Conclusion

We employed a combined PRISMA methodology and content synthesis approach to systematically identify and analyze studies focused on fraud detection in e-commerce using machine learning and data mining techniques. Our review encompassed 101 articles, with 16 classified as “other” due to their use of less common data mining methods, while the remaining studies fell under mainstream machine learning categories.To guide our analysis, we formulated four research questions, where the first two provided context for the primary inquiry. Among the machine learning algorithms reviewed, artificial neural networks (ANNs) were the most frequently applied, followed closely by random forests. Most studies concentrated on detecting credit card fraud, highlighting its dominant presence in the field. However, we identified a notable gap in detailed research on reseller fraud—also known as product flipping or scalping—an area with significant economic and household impacts that deserves more focused investigation using machine learning and related techniques. Our review also uncovered emerging fraud types such as triangulation and bot fraud, which have received limited attention in current machine learning and data mining research. This highlights the need for expanded studies to address these evolving challenges effectively.Additionally, we observed a growing interest in imbalanced learning techniques aimed at improving fraud detection systems, reflecting the importance of handling the common issue of imbalanced datasets in fraud scenarios. These findings offer practical insights for e-commerce professionals, enabling them to adopt effective strategies to detect and prevent fraud, reduce losses, and protect brand reputation. Moreover, our survey contributes valuable knowledge to the academic community, providing a foundation for future research directions in e-commerce fraud detection.

References:

1.S. Monteith, M. Bauer, M. AIda, 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.

Show in ContextCrossRef Google Scholar

2.S. Kodate, R. Chiba, S. Kimura, and N. Masuda, Detecting problematic transactions in a consumer-to-consumer e-commerce network, *Appl. Netw. Sci.*, vol. 5, no. 1, p. 90, 2020.

Show in ContextCrossRef Google Scholar

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.

Show in ContextGoogle Scholar

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.

Show in ContextCrossRef Google Scholar

5.*Sam Smith and Juniper Research*, Online payment fraud: Market forecasts, emerging threats & segment analysis 2022–2027, https://www.juniperresearch.com/press/losses-online-payment-fraud-exceed-362-billion/, 2024.

Show in ContextGoogle Scholar

6.Abdallah, M. A. Maarof, and A. Zainal, Fraud detection system: A survey, *J. Netw. Comput. Appl.*, vol. 68, pp. 90–113, 2016.

Show in ContextCrossRef Google Scholar

7.R. J. Bolton and D. J. Hand, Statistical fraud detection: A review, *Statistical Science*, vol. 17, no. 3, pp. 235–255, 2002.

Show in ContextCrossRef Google Scholar

8.Phua, V. Lee, K. Smith, and R. Gayler, A comprehensive survey of data mining-based fraud detection research, *arXiv preprint* arXiv: 1009.6119, 2010.

Show in ContextGoogle Scholar

9.L. Akoglu, H. Tong, and D. Koutra, Graph based anomaly detection and description: A survey, *Data Min. Knowl. Discov.*, vol. 29, no. 3, pp. 626–688, 2015.

Show in ContextCrossRef Google Scholar

10. Irani, S. Webb, and C. Pu, Study of static classification of social spam profiles in MySpace, *Proc. Int. AAAI Conf. Web Soc. Med.*, vol. 4, no. 1, pp. 82–89, 2010.

Show in ContextCrossRef Google Scholar

11. Bhowmick and S. M. Hazarika, Machine learning for E-mail spam filtering: Review, techniques and trends, *arXiv preprint* arXiv: 1606.01042, 2016.

Show in ContextGoogle Scholar