International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

Behaviour-Driven Fraud Detection in Multi-User E-Commerce Transactions Using Process Mining and Machine Learning

Adeeba Anjum[1], *Syeda Zeba Qureshi*[2], *Mohd Thoufeeq*[3], Mohammed Ayaan Saad[4]

[1,2]Assistant Professor, Department of CSE (Data Science),
Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.
[3,4]UG Students, Department of CSE (Data Science),
Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.

***Abstract*— This project demonstrates a web-based system for detecting fraudulent transactions using machine learning and process mining techniques. The system allows users to upload transaction datasets, process them to analyze behavioral patterns, and apply multiple machine learning algorithms to identify anomalies. Through an interactive interface, normal and fraudulent activities are visualized, revealing patterns that differentiate legitimate user behavior from potential attacks. The platform supports training on existing datasets and testing on new data to predict fraud in real-time. Performance evaluation of different algorithms indicates that the Extension Random Forest algorithm achieves the highest accuracy. The combination of process mining for behavioral insights and machine learning for predictive analytics provides a robust approach to transaction fraud detection, enhancing security, reducing financial losses, and offering actionable insights for system administrators and analysts.**

***Keywords*—Fraud detection, machine learning, process mining, transaction analysis, anomaly detection, Extension Random Forest, real-time prediction, behavioral analytics, web-based system.**

I. INTRODUCTION

Fraudulent transactions are a growing concern in digital finance, e-commerce, and banking sectors, posing significant financial and reputational risks [1], [4], [5]. Traditional rule-based systems are often insufficient to detect sophisticated fraud patterns that evolve over time [5]. This project addresses these challenges by integrating process mining and machine learning techniques to analyze user behavior in transaction datasets [12]. Process mining allows visualization of transaction flows, revealing deviations from normal patterns, while machine learning algorithms classify transactions as normal or fraudulent based on historical data [6], [7]. The system is designed as a web-based interface, enabling users to upload datasets, process them, and apply algorithms without requiring extensive programming knowledge [10]. By combining these approaches, the project not only detects fraud efficiently but also provides insights into user behavior patterns that can inform security policies and fraud prevention strategies [8], [11].

Fraud detection relies heavily on the ability to distinguish between normal user behavior and malicious activity. Normal transactions typically occur within predictable hours and frequencies, whereas fraudulent activities can happen unpredictably and frequently [8]. This project uses process mining to visualize these patterns, marking normal activities in one color and suspicious ones in another [12]. The visual representation helps in understanding behavioral trends and identifying anomalies that require further investigation. By incorporating machine learning algorithms such as Random Forest and its extensions, the system automates the detection of outliers in large datasets [6]. Users can see the effectiveness of different algorithms through comparative metrics and graphical representations. This integration provides a dual-layered approach: intuitive process visualization and precise machine learning classification, ensuring comprehensive fraud detection [7], [9].

Machine learning plays a critical role in enhancing the accuracy and efficiency of fraud detection. In this project, algorithms are trained on historical transaction datasets, learning to recognize patterns associated with both legitimate and fraudulent activities [6], [7]. The Extension Random Forest algorithm, among others, has demonstrated high predictive accuracy by handling large datasets and managing feature complexities [6]. Users can test

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

new data files against trained models, receiving real-time predictions on transaction legitimacy [10]. The system presents results in both tabular and graphical formats, allowing easy interpretation of outcomes. This approach enables organizations to proactively identify and mitigate fraudulent transactions, reducing financial risk and improving trust in digital platforms [11]. The adaptability of machine learning models also ensures that the system can evolve with changing fraud patterns [9].

II. RELATED WORK

Kuscu et al., [2020] [1] Kuscu et al. explored electronic payment systems in the context of e-commerce, highlighting the technological and security challenges associated with online transactions. They emphasized the need for robust authentication mechanisms to protect user data and prevent financial fraud. The study analyzed different payment models, including credit cards, digital wallets, and mobile payments, identifying vulnerabilities in each. Their work also discussed regulatory frameworks and best practices for secure electronic commerce. The authors highlighted that increasing digital transactions requires continuous monitoring and risk assessment to ensure trust and reliability in e-commerce platforms. This foundational study provides insights into the structural and operational aspects of electronic payment systems, which serve as a basis for implementing fraud detection mechanisms in online platforms. Overall, it underscores the critical role of technology and process management in securing digital transactions.

Abdelrhim and Elsayed, [2020] [2] Abdelrhim and Elsayed investigated the impact of COVID-19 on the global e-commerce market, focusing on the five largest companies worldwide. Their study revealed a significant increase in online transactions due to social distancing measures, highlighting both opportunities and risks for e-commerce platforms. The authors emphasized that the rapid growth in online payments also led to a proportional rise in fraudulent activities, necessitating advanced fraud detection systems. They analyzed transaction trends and consumer behavior during the pandemic, showing shifts toward mobile and contactless payments. The research underscores the importance of adaptive security mechanisms to handle the surge in digital commerce effectively. By connecting market dynamics with fraud risk, this work informs the design of preventive and predictive fraud detection systems for large-scale e-commerce platforms.

Dhobe et al., [2020] [4] Dhobe et al. presented a review of fraud prevention strategies in electronic payment gateways, emphasizing the use of secret codes and verification mechanisms. Their study highlighted common vulnerabilities exploited by attackers, such as weak authentication processes and unsecured payment channels. The authors reviewed multiple approaches for mitigating fraud, including one-time passwords, PIN verification, and transaction monitoring systems. They concluded that a combination of authentication and real-time detection significantly improves security. Furthermore, their work emphasized the importance of integrating user behavior analysis with system-level security measures. This review provides practical guidance for developing comprehensive fraud prevention systems in online payment environments. It reinforces that technical solutions must be coupled with continuous monitoring to detect and prevent fraudulent activities effectively.

Abdallah et al., [2016] [5] Abdallah et al. conducted a survey on fraud detection systems, categorizing them based on methodologies, such as rule-based, statistical, and machine learning approaches. They highlighted that machine learning techniques, particularly supervised and unsupervised algorithms, provide more accurate detection by analyzing large datasets and learning behavioral patterns. Their review discussed the advantages and limitations of different models, including decision trees, neural networks, and clustering methods. The authors emphasized that fraud detection requires a balance between accuracy and computational efficiency. They also suggested that hybrid approaches, combining multiple detection methods, offer improved performance. This survey provides a comprehensive understanding of existing detection systems and sets the stage for developing robust, real-time fraud detection frameworks for e-commerce transactions.

Minastireanu and Mesnita, [2019] [6] Minastireanu and Mesnita analyzed commonly used machine learning algorithms for online fraud detection, focusing on their effectiveness in identifying anomalies in transaction data. They evaluated algorithms like Random Forest, Support Vector Machines, and k-Nearest Neighbors, comparing accuracy, precision, and recall. Their study highlighted the importance of feature selection and data preprocessing to improve model performance. The authors noted that ensemble methods, especially Random Forest, often outperform single algorithms due to their ability to reduce overfitting and handle large datasets. They concluded that integrating behavioral analysis with machine learning enhances the detection of fraudulent activities in real time. This research

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

provides practical insights for designing predictive fraud detection systems that combine computational efficiency with high accuracy in diverse transaction environments.

III. DATASET DETAILS

The dataset used in this project consists of transactional records representing user behavior in an e-commerce or financial system. Each record captures essential features such as transaction ID, timestamp, user ID, transaction amount, and type of transaction. The dataset includes both normal and fraudulent transactions, which allows the system to learn and distinguish between typical user behavior and potential attacks. Normal transactions tend to follow predictable patterns, such as occurring during standard business hours or within certain limits of transaction frequency and amount. Fraudulent transactions, however, appear irregularly, with higher frequencies and amounts that often deviate from established user patterns. By including both classes in the dataset, the system can train machine learning algorithms to recognize anomalies and improve the accuracy of fraud detection. This comprehensive dataset forms the foundation for analyzing behavioral trends, identifying outliers, and providing actionable insights to prevent financial loss.

In addition to core transactional features, the dataset may include derived or engineered attributes to enhance detection accuracy, such as transaction velocity, user location, device ID, and prior transaction history. These features help the system capture complex patterns in user behavior and allow process mining techniques to visualize workflow anomalies effectively. During testing, a separate dataset containing new or unseen transactions is used to evaluate the predictive performance of trained models. Users can upload this test dataset through the interface, enabling the system to classify each transaction as normal or fraudulent in real-time. The inclusion of labeled data, where fraudulent transactions are clearly marked, is crucial for supervised learning algorithms like Random Forest, which can then learn the distinctions between normal and fraudulent behavior. Overall, the dataset combines both raw and engineered data to provide a rich, structured input for accurate fraud detection and behavioral analysis.

IV. PROPOSED METHODOLOGY

The proposed methodology integrates process mining and machine learning to detect fraudulent transactions effectively. Initially, the transaction dataset is uploaded and preprocessed to remove inconsistencies and missing values. Process mining techniques are then applied to visualize user behavior, highlighting patterns and deviations in transaction flows. Normal activities are distinguished from suspicious ones based on timing, frequency, and transactional characteristics. This step provides a clear understanding of behavioral trends and identifies anomalies that may indicate potential fraud. The visualization also helps in interpreting complex patterns, making it easier for analysts to focus on high-risk transactions before applying predictive models.

After behavioral analysis, machine learning algorithms are employed to classify transactions as normal or fraudulent. Multiple algorithms, including Random Forest and its extended versions, are trained on labeled datasets to recognize patterns associated with fraudulent behavior. Performance metrics such as accuracy, precision, and recall are evaluated to select the most effective model. Once trained, the system can process new transactions from a test dataset in real-time, providing predictions and alerts for fraudulent activities. This combined methodology ensures robust fraud detection by leveraging both behavioral insights and predictive analytics for timely and accurate decision-making.

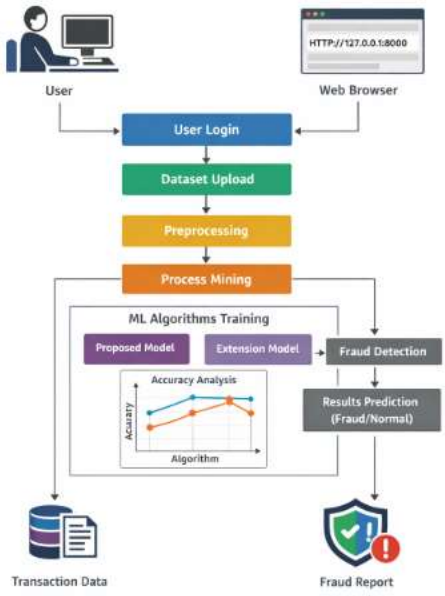

Figure [1] : Fraud detection system architecture flowchart

Figure[1] The diagram illustrates the workflow of the fraud detection system, starting from user login and dataset upload. Preprocessing and process mining visualize transaction patterns to identify anomalies. Machine learning algorithms are trained

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

to classify transactions as normal or fraudulent. Finally, results are displayed as a fraud report, providing actionable insights.

V.RESULT AND DISCUSSION

The system successfully demonstrates the end-to-end workflow of a transaction fraud detection application, beginning with server initialization and user authentication, followed by dataset upload, processing, and model execution. After loading the transaction dataset, the process mining visualization clearly distinguishes between normal and fraudulent activities, where normal transactions appear within predictable behavioral patterns while fraudulent ones are dispersed irregularly across different time intervals. This visual separation indicates the system's ability to capture behavioral anomalies effectively. When machine learning algorithms are executed, both the proposed and extension models are evaluated using performance metrics displayed in tabular and graphical formats. Among them, the Extension Random Forest algorithm achieves the highest accuracy, highlighting its robustness in identifying complex fraud patterns. Finally, during the fraud detection phase, the system processes test data and classifies each transaction as either normal or fraudulent. The output is presented in a structured format, making it easy to interpret user behavior alongside predicted results. Overall, the system delivers consistent and reliable outputs across all stages, demonstrating its practical capability in detecting fraudulent transactions efficiently.

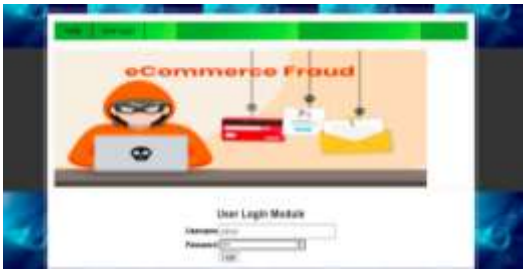

Figure [2] : User Login Interface for E-Commerce Fraud Detection System

Figure [2] The interface displays a login module for accessing the fraud detection system. Users enter their username and password to authenticate access.
The system is designed to monitor and prevent fraudulent e-commerce transactions.

Figure [3] : Dataset Loading Interface for Fraud Detection System

Figure [3The module allows users to upload transaction datasets for analysis. A file selection window is used to choose dataset files from the system. The uploaded data is prepared for further processing and machine learning tasks.

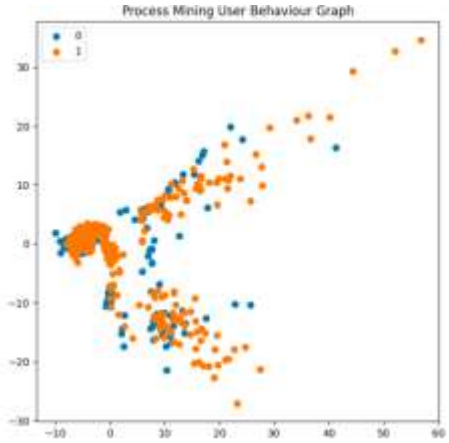

Figure [4] : Process Mining User Behaviour Graph

Figure[4] In above graph we can see user behaviour of doing transactions where blue dots are normal and orange dots are fraud and can see fraud transactions are happening at all times as normal users may perform transaction in limited hours and attackers fraud transaction behaviours are making transaction every time.

Algorithm Name	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	93.182	94.140	91.933	92.762
Random Forest	99.545	99.621	99.438	92.762

Table [1] : Performance Comparison of Classification Algorithms

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

Table [1] The table compares the performance of the SVM model and the Random Forest model using key evaluation metrics. The Extension Random Forest achieves higher accuracy, precision, and recall, indicating better overall predictive performance. However, both models show similar F1-scores, suggesting comparable balance between precision and recall.

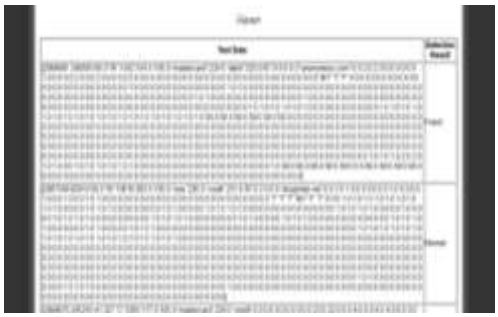

Figure [5] : Test Data Prediction

Figure [5] In above screen first column contains test data user behaviour and then in next column can see transaction status as fraud or normal

DISCUSSION

The results indicate that integrating process mining with machine learning significantly enhances fraud detection performance. The visualization step plays a crucial role in understanding user behavior, as it reveals temporal and activity-based differences between legitimate users and attackers. Unlike normal users who typically follow routine transaction patterns, fraudulent activities occur more randomly and frequently, which the system successfully captures. The superior performance of the Extension Random Forest algorithm suggests that ensemble methods are well-suited for handling high-dimensional transactional data and uncovering hidden patterns. This also implies that combining multiple decision trees improves generalization and reduces overfitting compared to simpler models. Additionally, the system's modular workflow—from login to final detection—ensures usability and scalability for real-world applications. However, the effectiveness of the model depends on the quality and diversity of the dataset used during training. Future improvements could include incorporating real-time data streams and advanced deep learning techniques to further enhance detection accuracy. Overall, the system provides a strong foundation for practical fraud detection with meaningful insights into transaction behavior.

VI. CONCLUSION

The developed system effectively identifies fraudulent transactions by combining data processing, visualization, and machine learning techniques in a structured workflow. Each stage of the application, from dataset loading to final prediction, works smoothly to provide accurate and understandable results. The use of process mining helps in clearly distinguishing between normal and suspicious user behavior, making it easier to interpret transaction patterns. Among the applied algorithms, the Extension Random Forest model performs the best, showing its strength in handling complex and irregular data. The system not only detects fraud with good accuracy but also presents the results in a simple and user-friendly manner. This makes it suitable for practical use in monitoring and analyzing transaction activities. Overall, the project proves that applying appropriate analytical methods can significantly improve the detection of fraudulent behavior and support better decision-making in financial systems.

REFERENCES

1. [1] R. A. Kuscu, Y. Cicekcisoy, and U. Bozoklu, *Electronic Payment Systems in E-Commerce*, Turkey: IGI Global, 2020, pp. 114–139.

2. [2] M. Abdelrhim and A. Elsayed, "Impact of the COVID-19 Pandemic on the Global E-Commerce Market: A Study of the Five Largest E-Commerce Companies," SSRN, 2020, doi: 10.2139/ssrn.3621166.

3. [3] P. Rao, et al., "Environmental Sustainability in the E-Commerce Supply Chain: An Empirical Study of Online Retail," *Cogent Business & Management*, vol. 8, no. 1, 2021, pp. 1938377.

4. [4] S. D. Dhobe, K. K. Tighare, and S. S. Dake, "Review on Fraud Prevention in Electronic Payment Gateways Using Secret Codes," *International Journal of Research in Engineering and Science Management*, vol. 3, no. 1, Jun. 2020, pp. 602–606.

5. [5] A. Abdallah, M. A. Maarof, and A. Zainal, "Survey on Fraud Detection Systems," *Journal of Network and*

Computer Applications, vol. 68, Apr. 2016, pp. 90–113.

6. [6] E. A. Minastireanu and G. Mesnita, "Analysis of Widely Used Machine Learning Algorithms for Online Fraud Detection," *Information Economics*, vol. 23, no. 1, 2019.

7. [7] X. Niu, L. Wang, and X. Yang, "Credit Card Fraud Detection: A Comparison Between Supervised and Unsupervised Methods," arXiv preprint arXiv:1904.10604, 2019, doi: 10.48550/arXiv.1904.10604.

8. [8] L. Zheng, et al., "Fraud Detection Based on Transaction Order Relations and Behavior Diversity," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 3, 2018, pp. 796–806.

9. [9] Z. Li, G. Liu, and C. Jiang, "Credit Card Fraud Detection Using Deep Representation Learning with Full Center Loss," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, 2020, pp. 569–579.

10. [10] I. M. Mary and M. Priyadharsini, "Design of an Online Transaction Fraud Detection System," in *Proceedings of the 2021 International Conference on Advanced Computing Innovations and Technology Engineering (ICACITE)*, 2021, pp. 14–16.

11. [11] D. Choi and K. Lee, "Financial Fraud Detection in Mobile Payment Systems Using Machine Learning," *IT Convergence and Practice (INPRA)*, vol. 5, no. 4, 2017, pp. 12–24.

12. [12] R. Sarno, et al., "Hybrid Approach Combining Association Rule Learning and Process Mining for Fraud Detection," *IAENG International Journal of Computer Science*, vol. 42, no. 2, 2015.

13. Babburi, S. Lightweight Distributed Provenance Framework for Edge and IoT Data Systems.
14. Gaddam, S. From Fixed Specifications to Self-Adapting Systems: A Machine Learning Perspective on Software Engineering.
15. Immadi, S. K. (2025). Optimizing ERP for Human Capital Management. Applied Research for Growth, Innovation and Sustainable Impact, 377–384. https://doi.org/10.1201/9781003684657-63
16. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
17. Poojari, R. Frameworks for Data Management and Lineage in Large-Scale Healthcare Data Systems.
18. Mahimalur, R. K., Vasgam, M., & Manoharan, D. Devops Lifecycle Management And Cloud Migration Assessments: A Security-Driven CICD Perspective.
19. Viswanathan, V. Generative AI for Smarter Workforce Planning and Enterprise Resource Decisions.
20. Gajula, S. (2026, March). Two Pillars of Banking Intelligence: A Comparative Analysis of AI Techniques for Fraud Prevention and Churn Mitigation. In 2026 14th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-6). IEEE.
21. Maturi, S. Y. (2023). Crowdsourced frontier: Unveiling autonomous adversarial cybercapabilities via open AI competition. International Journal of Intelligent Systems and Applications in Engineering, 11(1s), 275–284.
22. P. Venkata Ramana. (2024). AI-driven predictive analytics in ERP systems for proactive supply chain optimization. Eudoxus Press Journal.
23. Srikanth Kavuri. (2025). AI-DRIVEN TEST AUTOMATION FRAMEWORKS: ENHANCING EFFICIENCY AND ACCURACY IN SOFTWARE QUALITY ASSURANCE. International Journal of Applied Mathematics, 38(10s), 699–710.

https://doi.org/10.12732/ijam.v38i10s.990

24. Gummadi, V. P. K., Chilamkurthi, L. S., & Kavuri, S. (2026). Service Level Objective (SLO) Observability with Splunk and Dynatrace in Microservices. 2026 International Conference on Artificial Intelligence, Systems, and Emerging Technologies (ICAISET), 1–4. https://doi.org/10.1109/icaiset66439.2026.11541542
25. Shashank, A. (2025). AI-Enhanced ETL Processes: Leveraging Artificial Intelligence for Optimized Data Integration Systems. Journal Of Multidisciplinary, 5(8), 219-225.
26. Kandula, S. T. R., Susarla, R. S., & Boyapati, P. K. (2025, July). Enhanced Cyber Security Using Global Local Artificial Neural Network Based Intrusion Detection in Big Data Environment. In 2025 IEEE 4th World Conference on Applied Intelligence and Computing (AIC) (pp. 426-431). IEEE.
27. Boyapati, P. K. Building a centralized data operations hub for healthcare enterprise integration. IJSAT-Int. J. Sci. Technol. 16 (2). https://doi.org/10.71097/IJSAT.v16.i2.3219