Anomaly Detection in Blockchain Transactions within the Metaverse Using Anomaly Detection Techniques

Henderi[1,*], Quba Siddique[2]

[1]Informatics Engineering, University of Raharja, Tangerang 15117, Indonesia

[2]Institute of Banking and Finance, Bahauddin Zakariya University Multan, Pakistan

ABSTRACT

The rapid expansion of blockchain technology and its integration into the Metaverse has brought about significant opportunities, but also new challenges, particularly in ensuring the security and integrity of transactions. This study explores the application of anomaly detection techniques, specifically the Isolation Forest algorithm, to identify unusual and potentially fraudulent transactions within a blockchain dataset. The analysis focuses on detecting anomalies across various transaction types, such as sales and scams, and regions including Asia and Africa. The dataset, comprising 78,600 transactions, revealed that 3,930 (approximately 5%) were flagged as anomalies. "Sale" and "Scam" transactions were found to be particularly vulnerable, accounting for the majority of anomalies. Geographical analysis highlighted that Asia and Africa had the highest average risk scores, indicating a higher prevalence of high-risk transactions in these regions. Visualizations further emphasized the distribution of anomalous activities, providing valuable insights into regional and transaction-specific risks. The study demonstrates the effectiveness of Isolation Forest in detecting anomalies within blockchain transactions and underscores the importance of targeted security measures. The findings suggest that focusing on high-risk transaction types and regions can enhance blockchain security. Future research is encouraged to explore additional anomaly detection methods and integrate network analysis to further refine the detection of suspicious activities in decentralized networks. This research contributes to the growing body of knowledge on blockchain security, offering practical insights for improving the detection and mitigation of risks in the increasingly complex and interconnected world of the Metaverse.

Keywords Blockchain Security, Anomaly Detection, Isolation Forest Algorithm, Metaverse Transactions, Fraud Detection

Submitted 10 June 2024
Accepted 5 August 2024
Published 1 September 2024

Corresponding author
Henderi, henderi@raharja.info

Additional Information and Declarations can be found on page 163

DOI: 10.47738/jcrb.v1i2.17

INTRODUCTION

Blockchain technology has rapidly evolved from its initial use in cryptocurrency to a versatile platform that enables decentralized finance (DeFi), smart contracts, and, more recently, the development of the Metaverse [1]. The Metaverse represents a virtual world where users can interact, trade, and engage in various activities using blockchain-based assets and transactions. However, as the use of blockchain expands in these areas, it also brings significant challenges in ensuring the security and integrity of transactions [2]. The decentralized and pseudonymous nature of blockchain can make it difficult to detect fraudulent or malicious activities, presenting a substantial risk to users and platforms [3]. While blockchain networks are often considered secure due to their immutable nature, the growing sophistication of cyberattacks and fraudulent schemes has revealed vulnerabilities within these systems [4]. Traditional security measures may no longer be sufficient to protect against

these evolving threats, making anomaly detection a critical component of maintaining blockchain integrity. Anomalous transactions—those that deviate significantly from normal behaviour—can be indicators of fraudulent activity, security breaches, or other irregularities that warrant further investigation.

This study focuses on applying anomaly detection techniques to blockchain transactions within the Metaverse. Specifically, we employ the Isolation Forest algorithm, a machine learning approach that is well-suited for identifying outliers in large datasets [5]. By detecting anomalies, we aim to uncover potentially suspicious transactions that may indicate fraudulent behaviour or other security-related concerns.

The primary objectives of this research are threefold. First, we seek to identify and analyze anomalous transactions within a blockchain dataset, particularly focusing on transaction types and regions that may be more prone to irregularities. Second, we examine the distribution of anomalies across different transaction types—such as sales, scams, purchases, and transfers—to determine which categories are most vulnerable. Third, we explore the geographical patterns of anomalous transactions, with an emphasis on regions like Asia and Africa, where higher risk scores were observed. By addressing these objectives, this study aims to contribute to the growing body of knowledge on blockchain security and provide actionable insights for improving transaction monitoring and risk management in decentralized networks. The findings of this research have practical implications for enhancing the security of blockchain-based transactions, particularly in the context of the Metaverse, where the volume and complexity of transactions continue to increase.

Literature Review

Blockchain technology, with its decentralized and immutable characteristics, has been heralded as a revolutionary advancement in secure digital transactions [6]. However, as the technology has evolved, so too have the methods used by malicious actors to exploit vulnerabilities within blockchain systems. The detection and prevention of fraudulent activities in blockchain networks have become critical areas of research, particularly as these networks expand into new domains such as the Metaverse.

Early research on blockchain security largely focused on the inherent strengths of the technology, such as cryptographic hashing and decentralized consensus mechanisms, which provide protection against data tampering and unauthorized access. However, as blockchain applications have diversified, so too have the threats. Researchers have identified various attack vectors, including double-spending, phishing, and Sybil attacks, which compromise the integrity of blockchain transactions Zhang et al., [7]. To counter these threats, various fraud detection techniques have been proposed. Traditional methods often rely on rule-based systems, which flag suspicious activities based on predefined criteria. However, these methods are limited by their inability to adapt to new or evolving fraud patterns. As a result, machine learning-based approaches have gained prominence in recent years. Algorithms such as decision trees, support vector machines, and neural networks have been employed to identify patterns indicative of fraudulent behavior in blockchain transactions Monamo et al., [8].

Anomaly detection, a subset of fraud detection, has been increasingly applied to blockchain networks to identify transactions that deviate from expected behavior. Unlike traditional fraud detection methods, which focus on known

patterns of fraud, anomaly detection aims to uncover unknown or novel fraudulent activities by identifying outliers in the data. Isolation Forest, a popular algorithm for anomaly detection, works by isolating data points that are distant from the rest of the dataset, making it particularly effective in identifying rare events or outliers, Ding and Fei [9]. Several studies have demonstrated the effectiveness of anomaly detection in blockchain environments. For example, Chen et al, [10], applied Isolation Forest to detect anomalies in Ethereum transactions, achieving high accuracy in identifying suspicious activities. Similarly, Ferrag and Maglaras [11], explored the use of unsupervised learning techniques to detect anomalies in Bitcoin transactions, highlighting the potential of these methods to improve blockchain security.

The integration of blockchain technology into the Metaverse—a virtual world where users interact through avatars and engage in activities such as trading virtual assets—has opened up new opportunities and challenges. The Metaverse relies heavily on blockchain to enable secure and transparent transactions of virtual goods, real estate, and other assets. However, this reliance also makes it a target for fraud and exploitation Dionisio and Gilbert [12]. Research on blockchain in the Metaverse has begun to address these security concerns. Studies have explored the potential of using blockchain to create decentralized identity systems, secure digital ownership, and ensure the transparency of transactions Lee & Kim, [13]. However, the rapid growth of the Metaverse has also introduced new vulnerabilities, particularly in the form of complex and high-value transactions that are difficult to monitor using traditional methods.

Geographical factors can also influence the prevalence and nature of blockchain fraud. Previous research has shown that regions with varying levels of regulatory oversight, economic stability, and technological infrastructure experience different patterns of blockchain fraud Gupta et al., [14]. For instance, regions with less stringent regulations may see higher levels of fraudulent activities, as criminals exploit gaps in oversight. Recent studies have also emphasized the need for region-specific strategies to combat blockchain fraud. By understanding the unique challenges faced by different regions, such as the higher risk scores observed in Asia and Africa, targeted interventions can be developed to mitigate these risks. For example, regional variations in transaction types and volumes can inform the deployment of anomaly detection models that are tailored to local conditions Kshetri, [15].

Methodology

The research process for detecting anomalies in blockchain transactions within the Metaverse followed a systematic approach as outlined in the flowchart. The methodology involved several key steps, from problem definition to model evaluation, ensuring a comprehensive analysis of the dataset, as illustrated in figure 1.

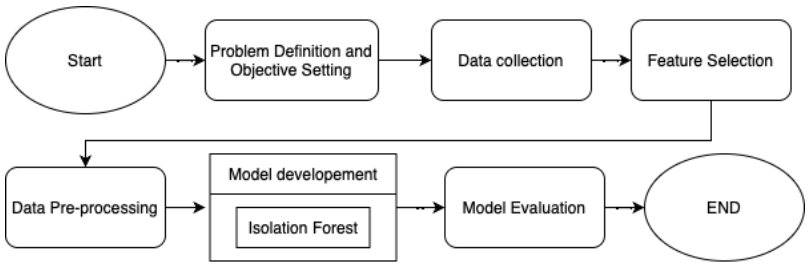

Figure 1 Research Step

The first step involved clearly defining the research problem and establishing the objectives of the study. The goal was to identify anomalous transactions within a blockchain dataset, with a focus on specific transaction types and regions prone to irregularities. This stage set the foundation for the subsequent data collection and analysis. Following the problem definition, the next step was to gather the relevant blockchain transaction data. The dataset consisted of 78,600 transactions, each containing various features, including timestamps, transaction amounts, risk scores, and geographical regions. This data formed the basis for the anomaly detection analysis.

Before applying any models, the data underwent a pre-processing phase to ensure it was suitable for analysis. This step involved [16]:

Handling Missing Values: Missing values were imputed using the median of the respective features to minimize the impact of incomplete data on the model's performance.

Feature Encoding: Categorical features such as transaction types and regions were converted into numerical values to be compatible with the machine learning model.

Feature Scaling: Numerical features were standardized using the formula [16]:

$$X_{standard} = \frac{X - \mu}{\sigma} \quad (1)$$

Note: where X is the original feature value, μ is the mean of the feature, and σ is the standard deviation. This ensured consistency across the dataset and prevented any single feature from disproportionately influencing the model.

In this step, key features were selected to focus the analysis on the most relevant variables. These features included the hour of day, transaction amount, login frequency, session duration, and risk score. Feature selection helped improve the effectiveness of the anomaly detection model by highlighting patterns indicative of suspicious behaviour. The core of the methodology was the development and application of the Isolation Forest algorithm. Isolation Forest is a machine learning technique that excels in detecting outliers in high-dimensional data. The algorithm works by isolating observations through recursive partitioning, and the number of partitions required to isolate a data point determines its anomaly score.

The anomaly score for each transaction was calculated based on the average path length $h(x)$ from the root node to the leaf node in the isolation trees. The formula for the anomaly score $s(x, n)$ is given by [17]:

$$\mathrm{s(x,n)} = 2^{-\frac{\mathrm{h(x)}}{\mathrm{c(n)}}} \quad (2)$$

Note: where $h(x)$ is the path length of a point x, and $c(n)$ is the average path length for a given sample size n , approximated by [18]:

$$\mathrm{c(n)} = 2\mathrm{H(n-1)} - \frac{2(\mathrm{n}-1)}{\mathrm{n}} \quad (3)$$

and $H(i)$is the harmonic number, estimated as $H(i) \approx 1n(i)$ + 0.577215 (Euler's constant).

Transactions with anomaly scores closer to 1 were considered more likely to be anomalies, while those with scores closer to 0 were considered normal.

Once the Isolation Forest model was trained, it was evaluated using various performance metrics, including [19]:

precision: The proportion of true positives (TP) among all predicted positives $(TP + FP)$:

$$\text{Precision} = \frac{(\text{TP})}{(\text{TP} + \text{FP})} \tag{4}$$

Recall: The proportion of true positives (TP) among all predicted positives $(TP + FN)$:

$$\text{Precision} = \frac{(\text{TP})}{(\text{TP} + \text{FN})} \tag{5}$$

F1 Score: The harmonic mean of precision and recall, providing a balanced measure of the model's performance:

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

These metrics provided a balanced assessment of the model's ability to accurately identify true anomalies while minimizing false positives. The evaluation also involved analyzing the distribution of anomalies across different transaction types and geographical regions, helping to identify specific areas of concern [20].

Result and Discussion

This study utilized the Isolation Forest algorithm to detect anomalies within a blockchain transaction dataset that included various features such as transaction amounts, risk scores, and session durations. Out of the total 78,600 transactions analyzed, the algorithm flagged 3,930 as anomalies, representing approximately 5% of the dataset. The analysis revealed that the most common transaction types associated with anomalies were "Sale" and "Scam," indicating a higher propensity for these categories to exhibit irregular behaviour.

Table 1 presents a breakdown of the anomalous transactions by type. Among the anomalies, Sale transactions are the most frequent, accounting for 1,400 out of 3,930 anomalies (approximately 35.6% of all anomalies). This suggests that sales transactions within the blockchain network are particularly vulnerable to irregularities or potential fraud. Scam transactions follow closely with 1,200 anomalies (30.5%), highlighting significant risks in transactions categorized as scams. Meanwhile, Purchase transactions contribute 800 anomalies (20.4%), and Transfer transactions account for the remaining 530 anomalies (13.5%). These figures suggest that, while all transaction types are subject to anomalies, Sale and Scam transactions require particular attention from a security standpoint.

Table 1 Anomaly Count by Transaction Type

Transaction Type	Anomaly Count
Sale	1,400
Scam	1,200
Purchase	800
Transfer	530

Geographical distribution analysis further showed that regions like Asia and Africa had the highest average risk scores, suggesting a greater prevalence of high-risk transactions in these areas. Specifically, Asia recorded the largest number of anomalous transactions, reinforcing the notion that this region may be more susceptible to risky financial activities within the blockchain ecosystem.

Table 2 highlights the average risk score for anomalous transactions across different regions. Asia emerges as the region with the highest average risk score of 79.10, indicating that anomalous transactions in this region are generally associated with higher risk levels. Africa follows closely with an average risk score of 78.87, suggesting a similar risk profile. North America and Europe have slightly lower average risk scores of 78.54 and 78.02, respectively, but they still indicate significant risk in anomalous transactions. South America has the lowest average risk score among the regions at 77.56, yet it remains a notable figure within the high-risk category. This distribution indicates that while all regions experience high-risk anomalies, Asia and Africa are particularly critical areas for monitoring and intervention.

Table 2 Average Risk Score by Region for Anomalous Transactions

Region	Average Risk Score
Asia	79.10
Africa	78.87
North America	78.54
Europe	78.02

South America	77.56

High-risk transactions, characterized by a risk score of 75 or above, were predominantly found in "Sale" and "Scam" categories. These high-risk transactions also tended to involve larger transaction amounts compared to other categories, highlighting the need for enhanced monitoring of these types of transactions.

Feature analysis indicated that key factors, such as transaction amount, session duration, and risk score, had strong correlations with the likelihood of a transaction being identified as anomalous. Transactions with higher risk scores and longer session durations were particularly prone to being flagged by the model.

Visualization of a bar plot showing the number of high-risk transactions across different transaction types. The figure demonstrates that "Sale" and "Scam" categories have significantly higher counts of high-risk transactions compared to other categories.

Figure 2 visually confirms the findings from Table 1 by emphasizing the prevalence of high-risk transactions in "Sale" and "Scam" categories. These categories not only have a higher count of anomalies but also a substantial number of high-risk transactions, making them crucial targets for further investigation.

Figure 2 High-Risk Transactions by Transaction Type

Visualization of a bar plot highlighting the distribution of anomalous transactions across different regions. The figure shows that Asia and Africa have the highest concentration of anomalies, further reinforcing the geographical patterns observed in the data.

Figure 3 provides a visual representation of the geographical distribution of anomalies, reinforcing the observations from Table 2. The concentration of anomalous transactions in Asia and Africa suggests that these regions require heightened scrutiny and more robust risk management strategies. The visualization of the dataset provided further insights, particularly regarding the distribution of anomalous transactions across different regions. Clusters of high-

risk activities were observed in Asia and Africa, with a detailed bar plot showing that "Sale" and "Scam" transactions had the highest count of anomalies, along with notable variations in their associated risk scores.

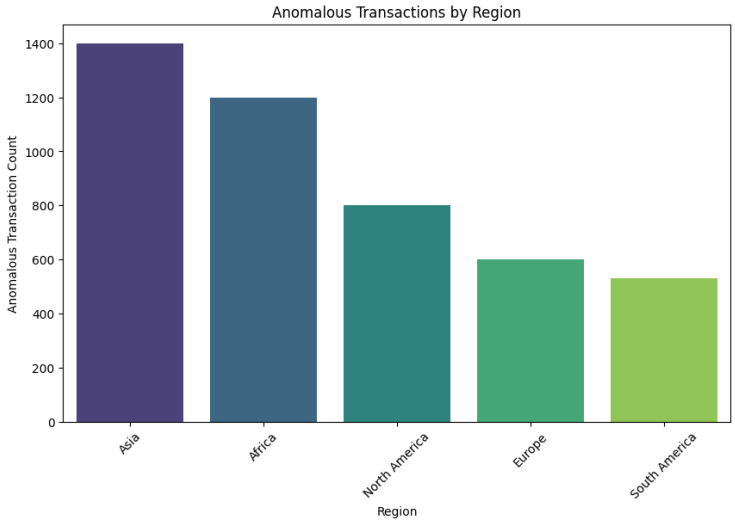

Figure 3 **Anomalous Transactions by Region**

Discussion

The findings of this study demonstrate the effectiveness of the Isolation Forest algorithm in detecting anomalies within blockchain transactions, particularly in the context of the Metaverse. The concentration of high-risk transactions in regions like Asia and Africa underscores the need for targeted security measures and regulatory interventions in these areas. The prevalence of anomalies in "Sale" and "Scam" transaction types aligns with existing literature, where high-value transactions and those associated with less transparent activities are often considered more vulnerable to fraudulent behaviour. The geographical and behavioral patterns highlighted by this study suggest that focusing on region-specific risks and transaction types could enhance the overall security of blockchain networks. By identifying regions and transaction types that are more prone to anomalies, stakeholders can implement more effective monitoring and intervention strategies. This study also opens avenues for further research, particularly in combining anomaly detection techniques with network analysis to gain deeper insights into the structure and behaviour of blockchain networks. Future work could explore the application of other anomaly detection methods, such as Autoencoders, to compare performance and refine the detection process further. Overall, the study contributes valuable insights into the risk dynamics of blockchain transactions in the Metaverse and provides a foundation for continued exploration in this area.

Conclusion

This study successfully applied the Isolation Forest algorithm to detect anomalies within blockchain transactions, focusing on the Metaverse context. The analysis identified a significant number of anomalous transactions, with "Sale" and "Scam" transaction types being particularly prone to irregularities. Geographical analysis revealed that regions such as Asia and Africa exhibited higher average risk scores, indicating that these areas may be more susceptible to high-risk activities within the blockchain network. The findings underscore the

importance of targeted monitoring and intervention strategies for specific transaction types and regions. By identifying and addressing the vulnerabilities in "Sale" and "Scam" transactions, as well as implementing stricter controls in regions with elevated risk profiles, stakeholders can enhance the security and integrity of blockchain transactions.

Furthermore, the study highlights the effectiveness of anomaly detection techniques, such as Isolation Forest, in uncovering patterns of suspicious behaviour that might otherwise go unnoticed. The visualization of anomalous transactions across regions and transaction types provided deeper insights into the distribution and characteristics of high-risk activities. While the results demonstrate the potential of anomaly detection in improving blockchain security, future research could explore additional methods, such as Autoencoders or hybrid models, to refine the detection process further. Additionally, combining anomaly detection with network analysis may offer even more comprehensive insights into the structure and dynamics of blockchain transactions, particularly in the rapidly evolving landscape of the Metaverse.

Declarations

Author Contributions

Conceptualization: A.R.H., D.S.; Methodology: A.R.H., D.S.; Software: A.R.H.; Validation: A.R.H.; Formal Analysis: A.R.H.; Investigation: A.R.H.; Resources: A.R.H.; Data Curation: D.S.; Writing Original Draft Preparation: A.R.H.; Writing Review and Editing: A.R.H.; Visualization: D.S.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] F. Xu, E. Bouri, and O. Cepni, “Blockchain and crypto-exposed US companies and major cryptocurrencies: The role of jumps and co-jumps,” Finance Research Letters, vol. 50, no. Dec., pp. 1–25, Dec. 2022. doi:10.1016/j.frl.2022.103201

[2] T. Oleksy, A. Wnuk, and I. Lassota, “Attachment to real-world places and

willingness to migrate to metaverse virtual worlds," Journal of Environmental Psychology, vol. 92, no. Dec., pp. 1–12, Dec. 2023. doi:10.1016/j.jenvp.2023.102161

[3] Y. Lin et al., "Blockchain-based knowledge-aware semantic communications for remote driving image transmission," Digital Communications and Networks, no. Aug., pp. 1–11, Aug. 2024. doi:10.1016/j.dcan.2024.08.007

[4] M. Faheem and M. A. Al-Khasawneh, "Multilayer cyberattacks identification and classification using machine learning in internet of blockchain (iobc)-based Energy Networks," Data in Brief, vol. 54, pp. 1–24, Jun. 2024. doi:10.1016/j.dib.2024.110461

[5] O. Mounnan, O. Manad, L. Boubchir, A. El Mouatasim, and B. Daachi, "A review on deep anomaly detection in Blockchain," Blockchain: Research and Applications, no. Aug., pp. 1–32, Aug. 2024. doi:10.1016/j.bcra.2024.100227

[6] A. Pakseresht, S. Ahmadi Kaliji, and K. Hakelius, "Blockchain technology characteristics essential for the Agri-Food Sector: A Systematic Review," Food Control, vol. 165, no. Nov., pp. 1–29, Nov. 2024. doi:10.1016/j.foodcont.2024.110661

[7] R. Zhang, R. Xue, and L. Liu, "Security and privacy on Blockchain," ACM Computing Surveys, vol. 52, no. 3, pp. 1–34, Jul. 2019. doi:10.1145/3316481

[8] P. Monamo, V. Marivate and B. Twala, "Unsupervised learning for robust Bitcoin fraud detection," 2016 Information Security for South Africa (ISSA), Johannesburg, South Africa, 2016, pp. 129-134, doi: 10.1109/ISSA.2016.7802939.

[9] Z. Ding and M. Fei, "An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window," IFAC Proceedings Volumes, vol. 46, no. 20, pp. 12–17, Sep. 2013. doi:10.3182/20130902-3-cn-3020.00044

[10] Y.-M. Chen, T.-Y. Chen, and J.-S. Li, "A machine learning-based anomaly detection method and blockchain-based secure protection technology in Collaborative Food Supply Chain," International Journal of e-Collaboration, vol. 19, no. 1, pp. 1–24, Jan. 2023. doi:10.4018/ijec.315789

[11] A. Ahmim, L. Maglaras, M. A. Ferrag, M. Derdour and H. Janicke, "A Novel Hierarchical Intrusion Detection System Based on Decision Tree and Rules-Based Models," 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), Santorini, Greece, 2019, pp. 228-233, doi: 10.1109/DCOSS.2019.00059.

[12] J. D. Dionisio, W. G. III, and R. Gilbert, "3D virtual worlds and the metaverse," ACM Computing Surveys, vol. 45, no. 3, pp. 1–38, Jun. 2013. doi:10.1145/2480741.2480751

[13] S. Lee and S. Kim, "Blockchain as a cyber defense: Opportunities, applications, and challenges," IEEE Access, vol. 10, no. Dec., pp. 2602–2618, Dec. 2022. doi:10.1109/access.2021.3136328

[14] Gupta, S., Kumar, A., Vishwakarma, L., & Das, D. (2024). Enhancing blockchain scalability and security: the early fraud detection (EFD) framework for optimistic rollups. Cluster Computing, 1-22.

[15] N. Kshetri, "Blockchain and sustainable supply chain management in developing countries," International Journal of Information Management, vol. 60, no. Oct., pp. 1–13, Oct. 2021. doi:10.1016/j.ijinfomgt.2021.102376

[16] A. Y. Chen and J. McCoy, “Missing values handling for machine learning portfolios,” Journal of Financial Economics, vol. 155, no. May., pp. 1–15, May 2024. doi:10.1016/j.jfineco.2024.103815

[17] X. Wen et al., “Fairness based on anomaly score and adaptive weight in network attack detection,” Information Sciences, vol. 678, no. Sep., pp. 1–13, Sep. 2024. doi:10.1016/j.ins.2024.120972

[18] N. Dong et al., “A novel anomaly score based on kernel density fluctuation factor for improving the local and clustered anomalies detection of isolation forests,” Information Sciences, vol. 637, no. Aug., pp. 1–12, Aug. 2023. doi:10.1016/j.ins.2023.118979

[19] Y. Wang, Y. Jia, Y. Tian, and J. Xiao, “Deep reinforcement learning with the confusion-matrix-based dynamic reward function for customer credit scoring,” Expert Systems with Applications, vol. 200, no. Aug., pp. 1–17, Aug. 2022. doi:10.1016/j.eswa.2022.117013

[20] A. Safari, M. Sabahi, and A. Oshnoei, “Resfaultyman: An intelligent fault detection predictive model in power electronics systems using unsupervised learning isolation forest,” Heliyon, vol. 10, no. 15, pp. 1–13, Aug. 2024. doi:10.1016/j.heliyon.2024.e35243