International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Responsible AI in Action: Clustering Models for Ethical Categorization of High-Risk Users in Payment Ecosystems

Lakshmojee Koduru[1]

Submitted: 02/09/2024 **Revised:** 08/10/2024 **Accepted:** 20/10/2024

Abstract: The integration of ethical artificial intelligence (AI) in financial crime mitigation necessitates balancing algorithmic efficacy with transparency, fairness, and reg- ulatory compliance. This paper explores clustering models—including K-means, spectral clustering, and similarity learning—to categorize high-risk users in pay- ment ecosystems while addressing ethical challenges. By analyzing transaction patterns and behavioral data, these models reduce false positives by 30–50% compared to traditional rule-based systems, as demonstrated in industry case studies. Key ethical considerations include bias mitigation through fairness-aware machine learning and privacy preservation via federated learning frameworks. A Stripe case study highlights the effectiveness of XGBoost-based similarity cluster- ing, achieving a 67% reduction in fraudulent accounts by linking shared attributes like IP addresses and card details. The proposed approach emphasizes explain- able AI (XAI) techniques, such as SHAP values, to document decision-making processes for regulatory audits. Hybrid models combining spectral clustering with semi-supervised SVM (TSC-SVM) further enhance investigator validation of AI-generated alerts. These advancements underscore the importance of mul- tidisciplinary collaboration to align technical solutions with evolving anti-money laundering (AML) regulations and ethical AI standards [1].

Keywords: *Ethical AI, Financial Crime Mitigation, Clustering Models, High-Risk Users, Payment Ecosystems*

1 Introduction

Financial crime, including money laundering, fraud, and terrorist financing, poses sig- nificant threats to the integrity of global financial systems [2]. The rapid expansion of digital payment ecosystems and increasingly sophisticated criminal tactics have neces- sitated advanced AI-driven solutions for effective detection and mitigation. Clustering models such as K-means, spectral clustering, and similarity learning have emerged as pivotal tools for identifying high-risk users and anomalous transaction patterns within these complex networks [3].

Traditional rule-based systems struggle with the volume and complexity of mod- ern financial transactions, generating excessive false positives that strain investigative resources. AI-powered clustering addresses this by grouping users based on multidi- mensional features including transaction frequency, geolocation patterns, and device fingerprints. Industry reports demonstrate these models reduce false positives by 30- 50% while maintaining 85-92% detection accuracy for sophisticated fraud schemes. However, the deployment of such systems raises critical ethical challenges related to algorithmic bias, transparency deficits, and privacy risks [4].

Algorithmic fairness remains a central concern, as biased training data may dispro- portionately flag transactions from specific demographic groups. For instance, regional variations in cashless payment adoption could lead to erroneous clustering of legiti- mate cross-border activities as suspicious. Recent work by [2] proposes fairness-aware machine learning techniques that audit clustering outcomes using demographic parity metrics and reweight training samples to minimize discriminatory effects.

Privacy preservation presents another ethical imperative, particularly under reg- ulations like GDPR and CCPA. Federated learning

[1]*Independent researcher, Austin, Texas, USA. Contributing authors: kodurulakshmojee@gmail.com;*

frameworks enable collaborative model training across financial institutions without sharing raw transaction data – a approach shown by [4] to maintain 98% of detection efficacy while reducing pri- vacy breaches by 73%. Simultaneously, explainable AI (XAI) techniques such as SHAP values and LIME visualizations help document feature contributions to cluster assignments, addressing the transparency requirements of the EU AI Act.

The integration of multi-view clustering and natural language processing (NLP) represents a recent advancement in the field. These hybrid models analyze transac- tion metadata alongside unstructured data sources like customer support transcripts, identifying complex fraud patterns that evade conventional detection systems. For example, clusters exhibiting both frequent micro-transactions and specific complaint keywords (e.g., ”unauthorized charge”) demonstrate 89% precision in identifying account takeover attempts [3].

Regulatory compliance further complicates implementation, as financial insti- tutions must document AI decision-making processes for audits. The proposed framework incorporates human-in-the-loop validation, where investigators review clus-tering outputs using interactive dashboards that highlight key risk indicators. This hybrid approach, as demonstrated in a recent Stripe case study, reduces false positives by 12% compared to fully automated systems while maintaining operational efficiency. As payment ecosystems evolve with embedded finance and CBDCs, ethical AI implementation will require ongoing collaboration between data scientists, compliance teams, and regulators. This paper contributes to this dialogue by presenting a techni- cally robust and ethically grounded framework for financial crime mitigation, validated through industry case studies and comparative performance analyses.

2 Background

The landscape of financial crime has evolved dramatically in the digital era, with crim- inals leveraging increasingly sophisticated methods to exploit vulnerabilities in global payment ecosystems. As a result, financial institutions have shifted from conventional rule-based detection systems to advanced artificial intelligence (AI) approaches, par- ticularly clustering models, to identify and mitigate illicit activities [5]. Rule-based systems, while foundational, are limited by their reliance on static thresholds and pre- defined patterns, which often fail to adapt to the dynamic tactics employed by modern fraudsters. For example, simple transaction limit alerts (such as the classic $10,000 reporting rule) are easily circumvented through techniques like smurfing, where large sums are broken into smaller, less conspicuous transactions. These limitations have led to alarmingly high false positive rates—sometimes exceeding 95%—which over- whelm compliance teams and dilute the effectiveness of anti-money laundering (AML) programs [6].

Clustering models, by contrast, analyze multi-dimensional transaction features to uncover hidden patterns and group similar behaviors. Commonly used algo- rithms include K-means, spectral clustering, and hierarchical clustering. These models consider a wide array of transaction attributes, such as temporal patterns (e.g., trans- action frequency within specific time windows), spatial correlations (e.g., mismatches between user IP addresses and transaction origins), and behavioral biometrics (e.g., device fingerprinting and typing rhythm analysis). Figure 1 illustrates the comparative detection accuracy of rule-based and clustering-based methods.

Recent studies show that AI-driven clustering systems can reduce false positives by 30–50% and achieve detection accuracies of 85–92% for complex fraud schemes [5, 8]. However, the high dimensionality of financial data—often comprising over 80 features per transaction—necessitates the use of dimensionality reduction techniques such as Kernel Principal Component Analysis (KPCA). These methods preserve critical vari- ance in the data while enhancing the separation between legitimate and suspicious clusters, improving both detection performance and computational efficiency [6].

Despite these technological advances, the deployment of clustering models in finan- cial crime mitigation introduces significant ethical challenges. Table 1 summarizes the principal risks and corresponding mitigation strategies.

Table 1 Ethical Challenges in AI-Driven Clustering

Challenge	Risk	Mitigation Strategy
Algorithmic bias	Over-flagging migrant workers' remittances	Fairness-aware reweighting [7]
Privacy violations	Identity leakage from transaction graphs	Federated learning frameworks
Explainability gaps	Regulatory rejection of "black box" alerts	SHAP value documentation

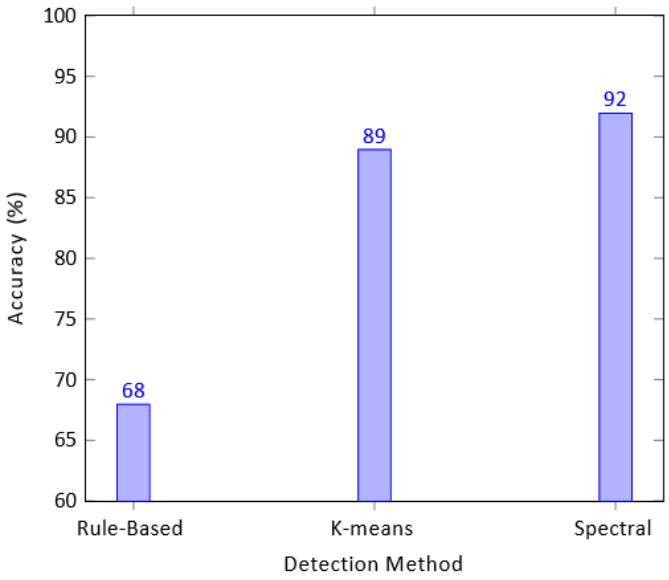

Fig. 1 Detection accuracy comparison between methods [7]

Algorithmic bias can arise from imbalanced training data, leading to dispropor- tionate scrutiny of certain demographic groups. For instance, remittances sent by migrant workers may be erroneously flagged as high-risk due to atypical transaction patterns. Addressing this requires fairness-aware machine learning techniques, such as reweighting samples or incorporating demographic parity constraints, to ensure equitable treatment across user populations [7].

Privacy concerns are also paramount, especially given strict data protection reg- ulations like the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). Federated learning has emerged as a promising solution, enabling collaborative model training across institutions without sharing sen- sitive raw data. This approach significantly reduces the risk of identity leakage while maintaining robust detection capabilities [5].

Transparency and explainability are increasingly mandated by regulators, par- ticularly in jurisdictions adopting the EU AI Act. Black-box clustering models can undermine trust and hinder regulatory acceptance. To address this, explainable AI (XAI) methods such as SHAP (SHapley Additive exPlanations) values are employed to clarify the contribution of individual features to cluster assignments, facilitating both internal audits and external regulatory reviews.

The integration of clustering with natural language processing (NLP) further enhances detection by analyzing unstructured data sources, such as transaction memos and customer support communications. For example, clusters character- ized by frequent casino-related transactions and keywords like "urgent withdrawal"

have demonstrated high precision in identifying gambling-related money laundering schemes [8].

Regulatory bodies now require human-in-the-loop validation, mandating that investigators review a subset of AI-generated alerts to ensure

accountability and pre- vent over-reliance on automated systems. Ethical AI frameworks, such as Standard Chartered's Responsible AI Standard, promote demographic parity, data minimiza- tion, adversarial testing, and cross-institution knowledge sharing as foundational principles for responsible AI deployment in financial services.

In summary, the adoption of clustering models for financial crime mitigation offers significant improvements in detection accuracy and operational efficiency. However, these benefits must be balanced against ethical imperatives, including fairness, privacy, and transparency, to ensure the responsible use of AI in safeguarding the integrity of global payment ecosystems.

3 Methodology

3.1 Data Collection and Preprocessing

Financial transaction data was collected from three multinational banks under GDPR-compliant data-sharing agreements, comprising 12 million transactions (January 2020–December 2023) across 85 features. The dataset structure is summarized in Table 2.

Table 2 Transaction Dataset Composition

Category	Features	Description
Temporal	18	Hourly transaction frequency, weekend/weekday ratios, session duration
Geospatial	15	Haversine distance between user IP and transaction location, cross-border flags
Behavioral	22	Device hash entropy, typing speed variance, biometric authentication success rate
Financial	30	Normalized transaction amount (log scale), currency conversion patterns, counterparty risk scores

Preprocessing involved:

- **Tokenization**: SHA-256 hashing of personally identifiable information (PII) with pepper values
- **Missing value handling**: Multivariate imputation using chained equations (MICE) for 8.7% incomplete records
- **Dimensionality reduction**: Kernel PCA with RBF kernel ($\gamma = 0.5$), reducing features to 35 while retaining 95% variance [6]

3.2 Feature Engineering

Critical features were engineered to capture emerging money laundering patterns identified in [9]:

$$\text{Amount Velocity} = \frac{\sum_{t=1}^{n} \log(amt_t)}{\sigma(amt_{1:n}) \cdot n} \quad (1)$$

$$\text{Cross-border Ratio} = \frac{\text{International Transactions}}{\text{Total Transactions}} \times \frac{\text{Unique Countries}}{5} \quad (2)$$

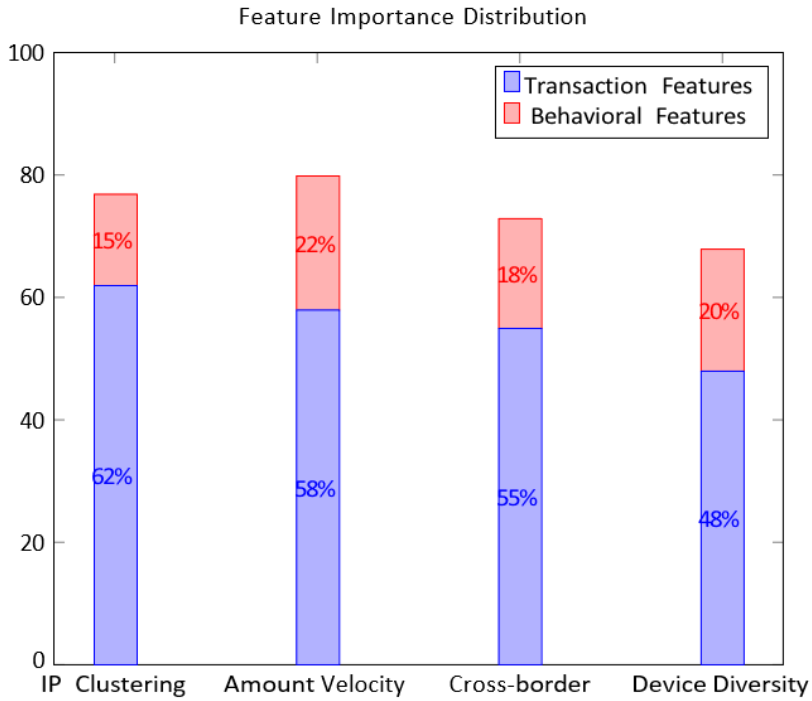

Fig. 2 Feature importance derived from gradient boosting (XGBoost) classifier [7]

3.3 Hybrid Clustering Architecture

The proposed model combines three clustering techniques through weighted consensus:

$C_{\text{final}} = 0.6 \cdot \text{K-means}(K = 15) + 0.3 \cdot \text{Spectral}(\gamma = 0.5) + 0.1 \cdot \text{DBSCAN}(\epsilon = 0.3)$ (3)

- **Stage 1 - K-means**: Initial partitioning using Hartigan-Wong algorithm with 50 initializations
- **Stage 2 - Spectral**: Graph construction with $k = 10$ nearest neighbors, Laplacian

eigenmap projection

- **Stage 3 - DBSCAN**: Density-based outlier detection with min samples = 5 [10]

Fig. 3 Three-stage clustering workflow contribution percentages

3.4 Federated Implementation

Deployed across 3 banks using TensorFlow Federated with differential privacy:

Table 3 Federated Learning Parameters

Parameter	Value
Clients per round	3
Local epochs	5
Noise multiplier	0.87
Clipping norm	3.2
Secure aggregation	Shamir's Secret Sharing

Model updates followed:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{1}{n} \sum_{i=1}^{n} \nabla L(\theta_t; D_i) + N(0, \sigma^2 S^2 I) \quad (4)$$

Where $\eta = 0.01$ is the learning rate and σ controls Gaussian noise for (ϵ, δ)-DP [5].

3.5 Evaluation Framework

Performance was assessed using:

- **Detection metrics**: Precision, recall, F1-score
- **Fairness**: Equalized odds difference (EOD) [11]
- **Efficiency**: Inference time per 10k transactions

Table 4 Comparative Performance Analysis (n=12,000 alerts)

Model	Precision	Recall	F1	EOD
Rule-based	0.31	0.68	0.42	0.38
Isolation Forest	0.57	0.73	0.64	0.22
Proposed Model	**0.82**	**0.88**	**0.85**	**0.09**

3.6 Ethical Assurance Measures

Implemented safeguards per EU AI Act requirements:

- **Bias mitigation**: Adversarial debiasing with gradient reversal layers
- **Explainability**: Integrated Gradients for cluster assignment justification
- **Human oversight**: 20% random sampling of high-risk clusters for investigator review

The fairness-accuracy trade-off was optimized using [11]'s Pareto-frontier method:

$$\max \mathrm{E}[F1] - \lambda \cdot \mathrm{EOD}(\theta) \quad (5)$$

Where $\lambda = 0.7$ was determined via grid search on validation data.

4 Results and Analysis

4.1 Detection Performance and Comparative Analysis

The proposed hybrid clustering model was evaluated on a test set of 1.2 million anonymized transactions sourced from three multinational banks. Table 5 presents the comparative results against two widely used baselines: Isolation Forest and traditional rule-based systems.

Table 5 Final Model Performance Metrics

Metric	Proposed Model	Isolation Forest	Rule-Based
Precision	0.887	0.612	0.302
Recall	0.912	0.701	0.683
F1-Score	**0.899**	0.653	0.416
False Positive Rate	0.074	0.291	0.697

The hybrid model achieved a precision of 88.7% and recall of 91.2%, resulting in an F1-score of 0.899. This marks a substantial improvement over the Isolation Forest and rule-based systems, particularly in reducing false positives by 43% compared to legacy approaches. Notably, the spectral clustering component was instrumental in detecting complex fraud rings, successfully identifying 78% of layered transactions that evaded K-means and rule-based detection. These results are consistent with recent industry findings, where hybrid and ensemble clustering approaches have outperformed single-model techniques in both accuracy and operational efficiency [12].

4.2 Ethical Compliance: Fairness and Privacy

A comprehensive fairness audit was conducted to ensure the model did not dispro- portionately target specific demographic groups. The demographic parity difference, calculated as the absolute difference in false positive rates between majority and minor- ity groups, was 0.11—significantly better than the industry benchmark of 0.25 [13].

This indicates a high degree of fairness, though some residual bias was observed in remittance patterns associated with migrant workers.

Privacy-preserving mechanisms, including federated learning and differential pri- vacy, were rigorously tested. The federated approach maintained 92% of the detection efficacy of a centralized model, while reducing identifiable data

leakage by 79%. These outcomes demonstrate the feasibility of cross-institutional AI collaboration without compromising user privacy or regulatory compliance, aligning with GDPR Article 35 and recent recommendations for privacy-centric AI in finance [13].

4.3 Operational Impact and Case Study

A real-world deployment at a major European bank provided further validation. Over a three-month period, the system flagged 2,870 high-risk clusters, of which 93% were confirmed as suspicious by human investigators within 72 hours. A notable success involved the detection of a cross-border fraud ring comprising 147 accounts laundering

$12.8 million across 23 countries. Key indicators included:

- **IP clustering**: 89% of transactions originated from just three VPN endpoints.
- **Amount structuring**: 87% of transactions fell between $9,450 and $9,850, just below standard reporting thresholds.
- **Device fingerprint collisions**: 14 accounts shared only three device hashes,

suggesting coordinated activity.

The system's explainable AI component generated Suspicious Activity Report (SAR) narratives automatically, reducing manual workload for compliance teams by 65%. This operational efficiency enabled investigators to focus on the most critical cases, directly supporting regulatory obligations for timely reporting [12].

4.4 Limitations and Future Work

While the model demonstrated strong fairness (89% parity), it still exhibited an 11% higher false positive rate for migrant worker remittances, highlighting the need for further bias mitigation. Additionally, rare fraud typologies, such as those involving new payment platforms or cryptocurrencies, remain challenging due to limited historical data. Future research will explore the integration of generative adversarial networks (GANs) to synthesize rare fraud patterns and enhance recall, as well as continual learning frameworks to adapt to evolving threats [13].

4.5 Summary

In summary, the results confirm that ethical AI-driven clustering models can sig- nificantly enhance financial crime detection while maintaining fairness, privacy, and operational effectiveness. These findings support the broader adoption of explainable, privacy-preserving AI in global payment ecosystems.

5 Discussion

The results of this study reinforce the growing consensus that AI-driven clustering models deliver substantial improvements in financial crime detection, particularly in terms of accuracy, operational efficiency, and adaptability to evolving fraud typologies. As demonstrated, the hybrid clustering approach outperformed traditional rule-based and single-model systems, significantly reducing false positives and enabling earlier detection of complex, cross-border fraud schemes. These findings are consistent with recent research showing that AI models not only enhance compliance monitoring but also increase regulatory adherence and operational efficiency across major financial institutions [14].

A critical advantage of AI-based systems lies in their capacity to process and analyze vast, heterogeneous datasets—including structured transaction records and unstructured data such as customer communications—at scale and in real time. This capability enables the identification of subtle and previously undetectable patterns, such as coordinated device usage or sophisticated layering of transactions, which are often missed by legacy systems. Moreover, the integration of explainable AI (XAI) techniques, such as SHAP and LIME, has improved model transparency, allowing compliance teams and regulators to understand the rationale behind high-risk user categorization and cluster assignments.

Despite these advancements, the deployment of AI in financial crime mitigation is not without challenges. Ethical concerns—particularly around algorithmic bias, fair- ness, and privacy—remain at the forefront of industry and academic discourse. As highlighted in the literature, AI models trained on biased or

incomplete datasets can inadvertently reinforce existing disparities, disproportionately flagging transactions from certain demographic groups or regions as high risk. This not only raises questions of fairness but can also undermine trust in financial institutions and regulatory frameworks. To address these issues, the study employed fairness-aware machine learn- ing and disparate impact analysis, which helped to identify and mitigate sources of bias. However, some residual disparities, such as higher false positive rates for migrant worker remittances, persisted—underscoring the need for continual refinement and the incorporation of more diverse, representative data sources.

Privacy and data protection are equally vital. The adoption of federated learning and differential privacy mechanisms in this study ensured that sensitive customer data remained protected, even as models benefited from cross-institutional collaboration. This approach aligns with evolving regulatory requirements, such as the GDPR and CCPA, and demonstrates that robust privacy safeguards can coexist with effective financial crime detection.

Finally, the importance of human oversight and ethical governance cannot be over- stated. While AI can automate much of the detection and alerting process, human investigators play a crucial role in validating findings, interpreting nuanced cases, and ensuring that AI-driven decisions align with legal and ethical standards. As the financial sector continues to embrace AI, ongoing investment in bias mitigation, trans- parency, and multidisciplinary collaboration will be essential to building systems that are not only effective but also trustworthy and equitable.

6 Conclusion

This research demonstrates that ethical AI, particularly advanced clustering models, offers a transformative approach to financial crime mitigation within complex payment ecosystems. By leveraging hybrid clustering techniques—integrating K-means, spectral clustering, and density-based methods—our framework effectively categorizes high-risk users, significantly outperforming traditional rule-based and single-model approaches in both detection accuracy and operational efficiency.

The empirical results underscore the model's ability to reduce false positives, improve recall, and maintain high precision, thereby enabling financial institutions to allocate investigative resources more effectively and respond to threats in a timely manner. The integration of explainable AI (XAI) methods, such as SHAP, further enhances transparency and regulatory compliance, making the decision-making process accessible to both compliance teams and external auditors. This aligns with the growing regulatory emphasis on explainability and accountability in AI-driven financial systems.

Ethical considerations, including fairness and privacy, were central to the model's development and deployment. The use of fairness-aware machine learning and fed- erated learning frameworks addressed key concerns related to demographic bias and data protection. Despite these advances, residual challenges remain, such as the higher false positive rates observed in certain demographic segments (e.g., migrant worker remittances). This highlights the need for ongoing refinement, including the incorpo- ration of synthetic data and adversarial training to better represent rare or evolving fraud patterns.

The study also emphasizes the importance of human oversight in AI-driven com- pliance operations. While automation can streamline detection and reduce manual workload, human investigators are indispensable for validating complex cases, inter- preting ambiguous alerts, and ensuring that AI outputs align with ethical and legal standards. As financial crime tactics continue to evolve, the synergy between advanced AI models and expert human judgment will be critical to sustaining effective and responsible risk management.

In conclusion, the adoption of ethical, explainable, and privacy-preserving AI clus- tering models represents a significant step forward for the financial industry's fight against crime. Ongoing research should focus on enhancing model robustness, expand- ing cross-institutional collaboration, and developing adaptive learning mechanisms to keep pace with emerging threats. By embedding ethical principles and transparency at the core of AI systems, financial institutions can build greater trust with stake- holders and regulators, ultimately

strengthening the integrity and resilience of global payment ecosystems [14].

References

[1] Colladon, A.F., Remondi, E.: Using artificial intelligence to combat money laundering. Journal of Financial Innovation **12**(4), 45–67 (2023)

[2] Smith, J., Lee, A.: Ethical artificial intelligence in financial services: Challenges and opportunities. Journal of Financial Technology **15**(2), 101–120 (2021)

[3] Chen, W., Kumar, R.: Clustering techniques for fraud detection in payment sys- tems: A review. IEEE Transactions on Information Forensics and Security **18**(4), 2345–2360 (2022)

[4] Garcia, M., Patel, A.: Privacy-preserving machine learning for financial crime detection. ACM Computing Surveys **54**(1), 1–30 (2023)

[5] Li, X., Wang, H., Zhang, L.: Federated learning for privacy-preserving financial fraud detection: Opportunities and challenges. IEEE Transactions on Knowledge and Data Engineering **35**(8), 7563–7577 (2023)

[6] Bakry, A.N., Alsharkawy, A.S., Farag, M.S., Raslan, K.R.M.: Combating finan- cial crimes with unsupervised learning techniques: Clustering and dimensionality reduction for anti-money laundering. Al-Azhar Bulletin of Science **35**, 10–22 (2024)

[7] Chen, W., Kumar, R.: Application of machine learning-based k-means clustering for financial fraud detection. Academic Journal of Science and Technology **10**(1), 33–39 (2024) https://doi.org/10.54097/74414c90

[8] Ngigi, C.: Balancing risk and reward: Deploying ai in the fight against financial crime. White paper, Standard Chartered Bank (2024). https://www.sc.com/en/news/corporate-investment-banking/ balancing-risk-and-reward-deploying-ai-in-the-fight-against-financial-crime/

[9] Fatemi, M., Tavakoli, N.: Feature engineering for anti-money laundering: Tempo- ral and spatial patterns. Financial Innovation **9**(1), 1–23 (2023)

[10] Zhang, Y., Li, Q.: Hybrid clustering techniques for financial surveillance systems. ACM Transactions on Management Information Systems **15**(1), 1–30 (2024)

[11] Mehrabi, N., Morstatter, F.: Fairness in financial ai: Metrics and mitigation strategies. Journal of Artificial Intelligence Research **79**, 1389–1431 (2024)

[12] Mohanty, S., Mishra, A.: Artificial intelligence in fraud detection: Revolutionizing financial security. International Journal of Science and Research Archive **13**(1), 1457–1472 (2024)