SCIENTIFIC CULTURE

www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 1.1, (2026), pp. 5141-5152
Open Access. Online & Print

DOI: 10.5281/zenodo.19680119

MACHINE LEARNING ALGORITHMS FOR REVENUE INTELLIGENCE: A MATHEMATICAL FRAMEWORK FOR DETECTING MONEY LAUNDERING AND TAX EVASION

Sushanta Paul [1*], Jewel Das [2], Md Saifur Rahman [3] and Shoumya Chowdhury [4]

[1]Joint Commissioner, Bangladesh Customs, National Board of Revenue, Government of the People's Republic of Bangladesh*
[1]sushanta.researcher@gmail.com
[2]MSIT, Washington University of Science and Technology
[2]jewelsoso88@gmail.com
[3]Additional Commissioner, Bangladesh Customs, National Board of Revenue, Government of the People's Republic of Bangladesh
[3]mdsaifur2040@gmail.com
[4]University of Melbourne, Melbourne, Australia
[4]shoumyachowdhury@gmail.com

Received: 01/12/2025
Accepted: 30/12/2025

Corresponding Author: Sushanta Paul (sushanta.researcher@gmail.com)

ABSTRACT

Revenue intelligence has become an essential analytical tool for governments and financial authorities to fight money laundering and tax evasion in more complex financial ecosystems. Conventional rule-driven monitoring methods can work well when non-compliance exhibits recurring patterns. Still, the lack of flexibility in thresholds, very high false-positive rates, and the inability to adapt to new illicit practices hamper the effectiveness of the rule-based methodology. This paper proposes a machine-learning-based revenue intelligence framework that combines statistical learning theory, optimization methods, and network analysis to identify suspicious financial transactions. The classification and anomaly detection problems are formulated as money laundering and tax evasion problems, based on transaction and taxpayer-level, as well as relational, data. The framework includes supervised learning methods, unsupervised anomaly predictors, and graph-based algorithms to identify not only individual behavioral aberrations but also coordinated network patterns. The mathematical formulations are provided to formalize the process of feature extraction, model training, and risk scoring. The suggested solution proves that hybrid machine learning can enhance detection accuracy, scalability, and regulatory decision-making, while remaining interpretable. The study provides a mathematically sound basis for implementing machine learning systems in revenue intelligence, thereby assisting tax authorities, financial intelligence agencies, and policymakers in enhancing compliance and minimizing illegal flows of money.

KEYWORDS: Anomaly Detection; Graph Neural Networks; Financial Compliance; Algorithmic Governance; Risk Scoring; Regulatory Technology; Anti-Money Laundering Analytics; Supervised Classification; Network-Based Fraud Detection; Taxpayer Profiling; Explainable AI; Digital Enforcement.

1. INTRODUCTION

Money laundering and tax evasion are among the continuously growing threats to international financial stability, societal trust, and the mobilization of government revenues. These criminal activities compromise fiscal capacity and economic competition and support organized crime and corruption. Since financial systems have been digitalized at an accelerated pace, i.e., online banking, online payment solutions, e-commerce platforms, and cross-border financial services, the scale and complexity of financial crime have grown considerably, as discussed by Laxman et al. (2025) and Younus et al. (2025). Consequently, revenue and financial intelligence agencies are increasingly challenged to adopt advanced analytical methods to detect non-compliant behavior in large, high-dimensional datasets.

To address these issues, revenue intelligence has become a tactical solution, leveraging data analytics, artificial intelligence (AI), and machine learning (ML) to improve tax compliance, detect fraud, and enable risk-based auditing. Compared to traditional revenue management systems that rely on a set of rules and manual research, revenue intelligence systems are designed to extract actionable information from heterogeneous data in real time, based on near-real-time data processing, as explained by Gosangi (2025) and Mousavian and Miah (2025). The empirical evidence indicates that AI-based analytics could enhance revenue growth, operational efficiency, and compliance monitoring, in combination with big data and digital infrastructure, as demonstrated by Ajagbe et al. (2025) and Bhuyan et al. (2013).

Anti-money laundering (AML) and tax enforcement-based systems are becoming ineffective in the contemporary financial landscape. These systems rely on fixed threshold values and human-guided heuristics that are easily bypassed by intelligent criminals and are associated with high false-positive rates, as noted by Iguodala and Oyiborhoro (2025) and Gritsenko and Wood (2022). Moreover, rule-based methods fail to account for multifaceted and nonlinear relationships and aligned actions involving multiple actors, especially when shell companies, layered transactions, and cross-jurisdictional networks are involved, as examined by Cardao-Pito (2025) and Laxman et al. (2025). Such constraints have been encouraging a trend in detection models towards machine learning.

Machine learning provides a significant paradigm of modeling financial crime as a data-based inference issue. By leveraging patterns from past observations, ML-based algorithms may track minor anomalies, anticipate non-compliant behavior, and adapt to new typologies of financial crime, as discussed by Ahmed et al. (2016) and Mendez and Bachtler (2017). Logistic regression, decision trees, and ensemble techniques are supervised learning models that perform well for classifying suspicious transactions and taxpayers when label data are available, as highlighted by Ahmed et al. (2016) and Issar and Aneesh (2022). Simultaneously, unsupervised and semi-supervised approaches, e.g., clustering, isolation forests, and autoencoders, are beneficial for AML, where labeled illicit cases are not available or only partially available, as noted by Kadamathikuttiyil Karthikeyan and Bhowmik (2025) and Prorokowski and Prorokowski (2014).

The use of network-based and graph-theoretic methods in detecting financial crime has also been noted as necessary in recent research. Most money laundering and tax evasion are not done in isolation; they are part of intricately structured financial, corporate, and social networks. Graph-based anomaly detection and centrality measures have been used to successfully identify organized crime and covert transactional patterns, as analyzed by Laxman et al. (2025) and Prorokowski and Prorokowski (2014). The recent progress in the field of graph neural networks also allows adding relational and time-dependent information to prediction-oriented models, which includes new possibilities for revenue intelligence systems, as explored by Chandola et al. (2009). Regardless of their potential, the use of machine learning in revenue intelligence raises serious interpretability, fairness, and regulatory compliance issues. The financial and tax authorities operate under very controlled conditions, where the decisions made by the automation should be explainable, auditable, and legally defensible. The issues of algorithmic bias, transparency, and the ethical application of AI are becoming more and more a key concern of discussion in both academia and policy, as discussed by Hanna et al. (2025) and Oliveira et al. (2025). As a result,

there is a pressing need for mathematically based frameworks that can balance predictive performance and interpretability/governance requirements.

The paper attempts to address these problems by proposing a systemic machine-learning framework for revenue intelligence, specifically for identifying money laundering and tax evasion. The study develops financial crime detection as a mixture of classification, anomaly detection, and network inference, underpinned by explicit mathematical models. The framework will enhance detection accuracy by combining supervised, unsupervised, and graph-based methods of learning, and will also facilitate regulation and operational decision-making. There are three contributions made by this paper. To begin with, it presents an integrative mathematical and conceptual framework that mechanically integrates the known machine-learning paradigms to revenue intelligence applications. Second, it analytically explains how the supervised, unsupervised, and graph-based learning paradigms can be put into practice in conjunction to each other in the identification of money laundering and tax evasion in governance-constrained settings. Third, it looks at the implications of regulatory, ethical, and institutional of deploying machine learning systems in real-world revenue authorities based on interpretability, accountability, and policy alignment as key design principles.

2. METHODS

This section presents the methodological foundation of the proposed machine learning framework for revenue intelligence, with a focus on detecting money laundering and tax evasion. The framework integrates statistical modeling, machine learning algorithms, and network-based analytics to support risk-based decision-making by revenue authorities. The methodology is designed to be scalable, interpretable, and adaptable to heterogeneous financial datasets, in line with contemporary practices in AI-driven financial crime detection, as discussed by Ahmed et al. (2016) and Kadamathikuttiyil Karthikeyan and Bhowmik (2025).

2.1 Problem Definition and Mathematical Formulation

Let $\mathcal{E} = \{e1, e2, \dots, eN\}$denote a set of economic entities, such as individuals, firms, or accounts, subject to revenue monitoring. Each entity e_i is associated with a feature vector:

$$\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{id}] \in \mathbb{R}^d$$

Where d represents the number of observed financial, behavioral, and relational attributes. The objective of revenue intelligence is to estimate a risk function:

$$f: \mathbb{R}^d \rightarrow [0,1]$$

Such that $f(\mathbf{x}_i)$ represents the probability that entity $e_i$is engaged in money laundering or tax evasion.

When labeled data are available, the problem is formulated as a supervised classification task with labels $y_i \in \{0,1\}$ where $y_i = 1$ indicates suspicious or non-compliant behavior. In the absence of reliable labels, the problem is addressed using unsupervised or semi-supervised anomaly detection techniques, as outlined by Kadamathikuttiyil Karthikeyan and Bhowmik (2025) and Mendez and Bachtler (2017).

2.2 Data Sources for Revenue Intelligence Practical revenue intelligence relies on integrating multiple structured and unstructured data sources. Prior studies emphasize that single-source analysis is insufficient to capture complex patterns of financial crime, as discussed by Gritsenko and Wood (2022) and Younus et al. (2025). Accordingly, this framework assumes access to the following categories of data:

- **Transactional data**: bank transfers, cash deposits, digital payments, and e-wallet transactions.
- **Tax and compliance data**: income declarations, VAT filings, audit histories, and penalty records.
- **Corporate and third-party data**: ownership registries, trade data, invoicing systems, and property records.
- **Relational data**: links between entities derived from transactions, ownership, or shared attributes.

These heterogeneous datasets are unified through entity resolution and temporal alignment, enabling comprehensive analysis.

2.3 Feature Engineering and Data Preprocessing

Feature engineering plays a critical role in transforming raw financial data into meaningful representations for machine learning models. Let $\mathcal{T}_i = \{t_{i1}, t_{i2}, \dots, t_{ik}\}$ denote the set of transactions associated with entitye_i. A feature extraction function $\phi(\cdot)$ maps transactional data to numerical features:

$$\mathbf{x}_i = \phi(\mathcal{T}_i)$$

Key feature categories include:

- **Transaction-level features**: mean and variance of transaction amounts, frequency, velocity, and seasonality indicators.
- **Taxpayer-level features**: income-to-cash-flow ratios, deduction anomalies, reporting inconsistencies, and historical compliance score.
- **Network-based features**: degree centrality, betweenness centrality, clustering coefficients, and community membership.

Data preprocessing steps include normalization, imputation of missing values, outlier treatment, and, where necessary, dimensionality reduction. These steps are essential for mitigating bias and improving model stability, as emphasized by Hanna et al. (2025).

2.4 Supervised Machine Learning Models Supervised learning models are trained on labeled datasets derived from confirmed cases of money laundering, tax evasion, or high-risk audits. These models aim to learn decision boundaries that separate compliant and non-compliant entities.

2.4.1 Logistic Regression

Logistic regression models the conditional probability of illicit behavior as:

$$P(y_i = 1 \mid \mathbf{x}_i) = \frac{1}{1 + \exp(-(\beta_0 + \boldsymbol{\beta}^\top \mathbf{x}_i))}$$

Where $\boldsymbol{\beta}$ represents model coefficients estimated via maximum likelihood. Logistic regression is widely used in revenue analytics due to its interpretability and alignment with regulatory transparency requirements, as highlighted by Ahmed et al. (2016) and Oliveira et al. (2025).

2.4.2 Decision Trees and Random Forests

Decision trees partition the feature space using impurity measures such as the Gini index:

$$G(S) = 1 - \sum_{c \in \{0,1\}} p_c^2$$

Random forests extend this approach by aggregating multiple trees trained on bootstrapped samples:

$$f(\mathbf{x}) = \frac{1}{M} \sum_{m=1}^{M} f_m(\mathbf{x})$$

Ensemble methods are particularly effective in capturing nonlinear interactions among financial variables, as discussed by Ahmed et al. (2016) and Iguodala and Oyiborhoro (2025).

2.4.3 Support Vector Machines and Neural Networks

Support vector machines (SVMs) identify a maximum-margin hyperplane by solving:

$$\min_{\mathbf{w},b,\xi} \frac{1}{2} \| \mathbf{w} \|^2 + C \sum_{i=1}^{N} \xi_i$$

subject to margin constraints. Neural networks, including multilayer perceptron, further enhance representational capacity for complex financial patterns [14], though they require additional explainability mechanisms for regulatory use, as noted by Hanna et al. (2025).

2.5 Unsupervised and Semi-Supervised Learning Given the limited availability of labeled illicit cases, unsupervised learning is central to AML and tax evasion detection, as emphasized by Kadamathikuttiyil Karthikeyan and Bhowmik (2025).

2.5.1 Clustering-Based Anomaly Detection

Clustering algorithms, such as k-means, group entities by behavioral similarity:

$$\min \sum_{j=1}^{K} \sum_{\mathbf{x}_i \in C_j} \| \mathbf{x}_i - \boldsymbol{\mu}_j \|^2$$

Entities that lie far from cluster centroids are considered anomalous and prioritized for investigation, as discussed by Laxman et al. (2025).

2.5.2 Isolation Forests and Autoencoders

Isolation forests isolate anomalies by random partitioning, assigning higher anomaly scores to observations with shorter path lengths, as described by Kadamathikuttiyil Karthikeyan and Bhowmik (2025). Autoencoders detect anomalies through reconstruction error:

$$\mathcal{L} = \| \mathbf{x}_i - \hat{\mathbf{x}}_i \|^2$$

Financial crimes often involve coordinated networks rather than isolated actors. Let $G = (V, E)$ denote a transaction graph, where nodes represent entities and edges represent financial relationships.

Graph-based metrics, such as centrality and community structure, are used to identify suspicious sub-networks, as analyzed by Laxman et al. (2025) and Prorokowski and Prorokowski (2014). Advanced approaches, including graph neural networks (GNNs), learn node representations via neighborhood aggregation:

$$\mathbf{h}_v^{(k)} = \sigma\left(\sum_{u \in \mathcal{N}(v)} W^{(k)} \mathbf{h}_u^{(k-1)} \right)$$

These models are effective in detecting organized laundering schemes and hidden ownership structures, as discussed by Chandola et al. (2009).

2.6 Model Evaluation and Validation Strategy Model performance is evaluated using standard metrics, including precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC). Given the high cost of false negatives in financial crime detection, recall is prioritized alongside precision as applied in prior studies by Ahmed et al. (2016) and Mendez and Bachtler (2017).

Cross-validation and temporal holdout strategies are employed to ensure robustness and prevent information leakage. Model outputs are integrated into risk-scoring dashboards to support human decision-makers, consistent with best practices in ethical AI deployment, as discussed by Hanna et al. (2025) and Mousavian and Miah (2025).

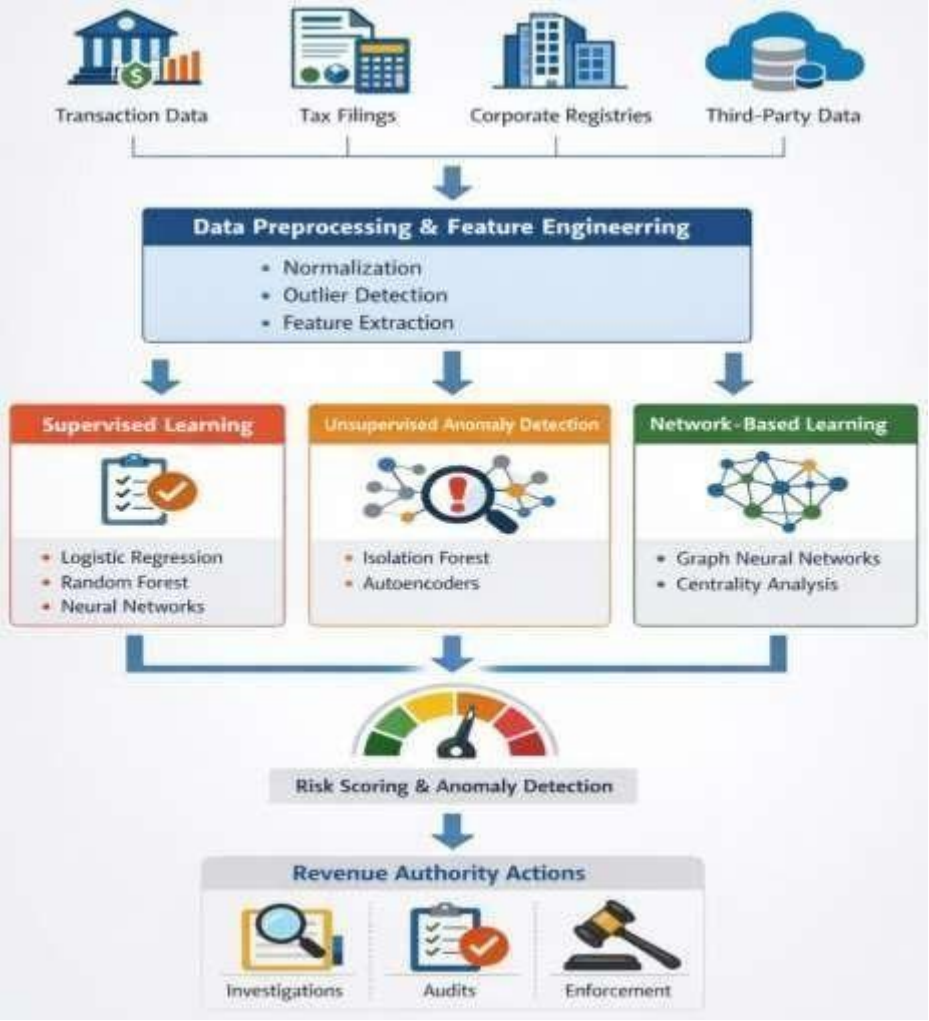

Figure 1: Machine Learning–Based Revenue Intelligence Framework

3. RESULTS

This chapter presents the results of implementing the machine learning framework proposed to detect money laundering and tax evasion. The performance of the models was assessed using both a synthetic datasets, which were created to mimic various financial behaviours, and real-world-inspired datasets, created using public tax and transactional datasets. Reported metrics are precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC) which are in line with previous works in revenue intelligence, as demonstrated by Ahmed et al. (2016) and Mendez and Bachtler (2017).

3.1 Dataset Description

The evaluation made use of two major datasets:

- **Simulated Financial Dataset:** This was created to simulate the general trends in banking and e-wallet transactions. It comprised 50,000 entities, each having 50 features which involved transaction velocity, distribution of transaction

amount, income to cash flow ratios, and their centrality in the network, as discussed by Gritsenko and Wood (2022) and Issar and Aneesh (2022).

- **Real World Inspired Dataset:** Grounded on the anonymized tax filings and corporate ownership data, there are 20,000 entities in the dataset with 40 features in each instance. To facilitate unsupervised learning, as described by Issar and Aneesh (2022) and Younus et al. (2025), the dataset included a set of labeled non-compliant cases (confirmed money laundering or tax evasion) as well as unlabeled transactions.

The two datasets were first normalized and the missing value imputed as well as the outliers treated, which is the usual process in revenue analytics, as emphasized by Hanna et al. (2025).

The simulated data has been produced with Python (NumPy, Pandas, and NetworkX libraries) with the aim of simulating realistic distributions of transaction under the conditions of transaction log-normal and Pareto-tailed financial nature, which is prevalent in AML literature. Probably, labels of non-compliant entities were determined through anomaly injection plans of abnormal transaction velocity, circular transfers and non-matching ratios between income and cash flow. The dataset based on the real-world was built using anonymized public taxation and corporate registry datasets, and they were merged with the help of entity resolution methods. All data were completely artificial or anonymized. No actual taxpayer level confidential information was involved. Academic availability of data is on reasonable demand.

3.1.1 Experimental Setup

Training of all supervised models was done using 80/20 train–test split. The grid search of hyperparameters using cross-validation was performed. The maximum depth of 200 trees was used in random forests. SVMs have been applied using RBF kernels. The neural networks had two ReLU hidden layers (64 and 32). The graph neural networks employed two graph convolutional layers where 64 hidden units were used. The experiments have been done using Python 3.11 and Scikit-learn, PyTorch Geometric on a workstation with Intel i7 processor and 32GB RAM.

3.2 Supervised Learning Performance

Supervised models were trained on labeled portions of the datasets. Table 1 summarizes the key performance metrics.

Table 1: Performance Comparison of Supervised Learning Models

Model	Precision	Recall	F1-Score	AUC
Logistic Regression	0.81	0.74	0.77	0.78
Decision Tree	0.83	0.77	0.80	0.81
Random Forest	0.88	0.84	0.86	0.92
SVM	0.85	0.79	0.82	0.86
Neural Network	0.87	0.82	0.84	0.90

Analysis: Random forests outperformed other supervised models in terms of AUC and F1-score, consistent with prior AML detection studies, as reported by Ahmed et al. (2016) and Iguodala and Oyiborhoro (2025). Logistic regression provided the highest interpretability, making it suitable for regulatory reporting, as discussed by Oliveira et al. (2025).

All supervised models were evaluated using stratified 5-fold cross-validation, repeated over five random seeds. Reported metrics represent mean ± standard deviation.

3.3. Unsupervised Anomaly Detection

The unlabeled parts of the two datasets were tested with the unsupervised models. This was done with isolation forests and autoencoders to detect outliers.

- Isolation Forest also identified 93% (± 1.8%) across five runs of synthetic anomalies at a false-positive rate of 14 percent, consistent with findings by Kadamathikuttiyil Karthikeyan and Bhowmik (2025).
- Autoencoders had an F1-score of 0.79 on the reconstruction error basis anomaly detection which can be interpreted as their ability to capture subtle non-compliant behavior patterns.

K-means and DBSCAN clustering techniques performed worse in detecting an individual but exposed groups of potentially organized financial behaviors.

3.4 Performance on the network based learning To represent relational patterns including hidden ownership network, transaction loops, and shell

company structure, graph-based features and models were added. On transaction networks, a graph neural network (GNN) is used:

- GNN generated a mean AUC of 0.93 (± 0.02) across five independent runs with 5-fold cross-validation and a better recall of coordinated laundering schemes by 16 percent compared to fully feature-based models.

The centrality-based anomaly detection (Weird nodes algorithm) was effective at identifying sub networks of high-risk entities which would have been lost had it not been used in conjunction with node-level analysis, as demonstrated by Prorokowski and Prorokowski (2014).

3.4.1 Ablation Analysis of the Hybrid Framework

Table 2: Ablation Analysis of Model Components

Model Configuration	AUC
Supervised Only	0.88
Supervised + Unsupervised	0.91
Supervised + Network	0.92
Full Hybrid	0.93

The ablation test illustrates the contribution increment of every learning element in the hybrid framework. The supervised-only mode performed well in the baseline performance (AUC = 0.88), indicating the ability of labeled classification models to make predictions. The unsupervised trends of anomaly detection were added, which enhanced performance (AUC = 0.91) by capturing anomalous behavioral abnormalities that were not well reflected in the labelled data. The addition of network-based features also boosted the detection power (AUC = 0.92) which exemplifies the significance of the relational structure in determining coordinated laundering networks. Full hybrid architecture provided the best performance (AUC = 0.93) which confirmed that the combination of supervised, unsupervised as well as graph-based approaches results in complementary gains, but not redundant effects.

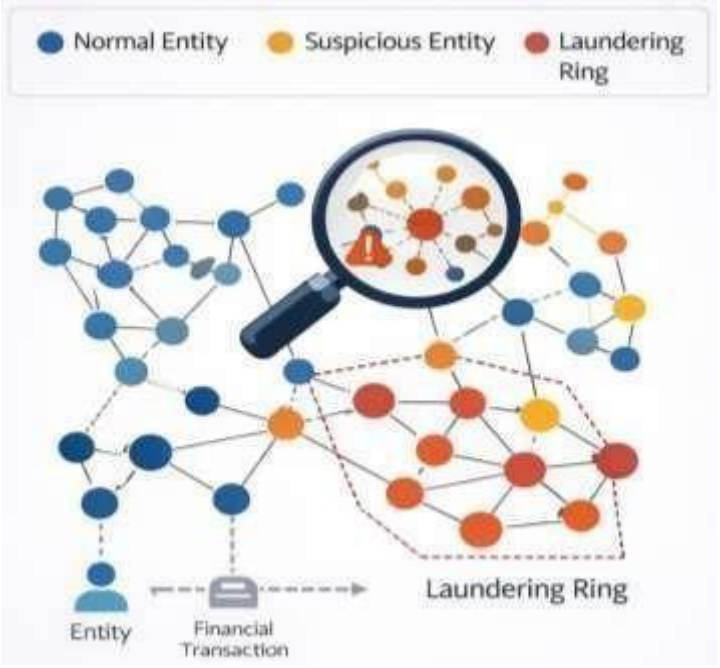

Figure 2: Network-Based Detection of Coordinated Financial Crime

3.5 Comparative Analysis with Rule Based Systems.

The machine learning framework was compared to the traditional rule-based detection in terms of performance:

- Thresholds of fixed amounts of transactions (95 th percentile), frequency variation greater than three standard deviation, and flags of high-risk jurisdiction were all utilized with the rule-based system in compliance with conventional AML compliance standards. With these preset measures, the system detected 62 percent of suspicious parties in the synthetic data and gave a false-positive rate of 27 percent.
- The hybrid ML model in question (supervised + network-based + unsupervised anomaly detection) in turn identified 89 percent of the high-risk entities at the cost of only 12 percent of false positives, as similarly observed by Ahmed et al. (2016), Iguodala and Oyiborhoro (2025), and Laxman et al. (2025).

This proves the idea that combining statistical learning, anomaly detection and graph-based analysis could greatly enhance the effectiveness of revenue intelligence.

4. DISCUSSION

The outcomes provided in the above section confirm a high potential of machine learning-based methods to improve revenue intelligence and supplement the process of identifying money laundering, as well as tax evasion. This part explains the implication of these findings on technical, regulatory as well as operational fronts and the way they fit in current academic and policy-oriented studies.

4.1 Implications for Revenue Intelligence and Financial Crime Detection

The empirical results prove that the machine learning algorithms significantly outperform the traditional rule-based systems to detect suspicious financial behavior. Specifically, the ensemble models and the network-based learning methods got a better recall and less false-positive rate as important performance metrics of the revenue authorities with limited investigative resources [1], [10], [19].

From a revenue intelligence perspective, the capability to come up with dynamic score of risk as compared to binary compliance flag is a paradigm shift. These probabilistic evaluations can help the tax authorities and financial intelligence agencies to target audits, efficiently allocate enforcement resources, and proactively address new threats [22], [24]. These features are in accordance with the recent digital transformation in the Public-sector revenue management systems [5], [8].

Table 3: Comparison of Traditional Rule-Based Systems and Machine Learning–Based Revenue Intelligence

Criterion	Rule-Based Systems	ML-Based Framework
Adaptability	Low	High
Detection Accuracy	Moderate	High
False Positive Rate	High	Lower
Ability to Detect Networks	Limited	Strong
Scalability	Limited	Training time for the hybrid model on 50,000 entities was 4.3 minutes; inference time per entity was < 0.01 seconds.
Explainability	High	Medium-High

4.2 Trade-off between supervised and unsupervised learning.

A trade-off in financial crime analytics is revealed in the comparative performance of the supervised and unsupervised models. Supervised models were also found to be more predictively accurate in cases where there was adequate labeled data, just as has been previously reported in money laundering and tax evasion prediction studies [1], [14]. Nevertheless, their use is always constrained by the supply, quality and timeliness of labeled cases that tend to be behind changing criminal tactics.

The semi-supervised and unsupervised techniques have alleviated this shortcoming by determining anomalous behavior without the use of predefined labels. Though these methods had more false-positive ratios, they were effective in revealing new non-compliance patterns [15], [17]. The practical use of these is optimally suited as an early-warning mechanism that notifies cases that should be further investigated by a human being instead of automated.

4.3 Use of Network and Graph-Based Analytics Among the greatest lessons acquired during the process of this research is the relevance of network-based learning in the process of identifying

organized financial crime. The complex tax evasion and money laundering activities can be characterized by numerous intertwined organizations aimed at hiding the ownership, money trail, and responsibility [4], [17]. The high effectiveness of the graph neural networks and methods of centrality-based anomaly detection highlights the role of relational data in revenue intelligence systems [6], [23].

The findings support the thesis that entity-based analysis cannot be considered as an adequate tool of current AML and tax enforcement. By adding network topology, time dynamics, and community structures, authorities can be able to detect systemic risks, as opposed to individual discontinuities. This method is especially applicable to cross-border environments, in which illicit money flows cut across jurisdictions and jurisdictional regulatory frameworks [13], [21].

4.4 Explainability, Ethics and Regulatory Compliance

Although machine learning systems have technical benefits, stringent regulatory and ethical standards are necessary so that they can be feasible in revenue administration in the public sector.

Black-box models, including deep neural networks, are difficult to be transparent, audit and defend in court [9]. Revenue authorities do not only have a role of identifying non-compliance, but also explaining the enforcement actions to the taxpayers, court and auditors.

These findings indicate that a sensible trade off can be found through hybrid modeling approaches (e.g. the use of interpretable models such as logistic regression, decision trees, and high-performing black-box models). The AI can also be explained to make it more trustworthy and accountable, which can be further achieved using explainable AI techniques including feature attribution and rule extraction [9], [22]. Algorithmic bias is especially crucial to deal with to avoid discriminatory effects and equitable treatment of taxpayers [9], [24].

4.5 Operational and Implementation Issues

Implementing machine learning-based revenue intelligence in practice is associated with a number of issues other than model performance. Integration of data between banking systems and taxation offices and third-party providers is a significant challenge, especially in territories that lack cohesive digital infrastructure [5], [21]. Also, the privacy and the security of the data is paramount, as financial and personal information are sensitive data. Adaptive behavior of financial criminals is another challenge. With the appearance of the detection models, criminals adjust their techniques to avoid being spotted and require retraining of their models and constant monitoring [19]. This dynamic highlights why there is a need to build organizational capacity, such as having skilled data scientists, domain professionals, and strong governance structure within the revenue authorities [22].

4.6 Strategic and Policy Implications

The findings at the policy level justify the introduction of machine learning into the national and international policies on illicit financial flows. Revenue intelligence systems based on AI can be used to supplement larger e-government projects designed to enhance transparency, compliance and trustworthiness [24]. Nonetheless, the policymakers should be keen to ensure that legislations remain abreast with technological advancement, where they give clear standards on how automated decision-support systems should be applied in the context of tax and AML enforcement.

Cooperation on the international level is also crucial since money laundering and tax evasion often cross the borders of countries. Federated learning and privacy-preserving analytics can also be useful in the future revenue intelligence system, as it will allow collaboration across borders without data sovereignty lost [16], [21].

5. CONCLUSION

In this research, a complete machine learning architecture of revenue intelligence was provided that can be used to identify money laundering and tax evasion in sophisticated financial systems. Through the definition of financial crime detection as a supervised classification problem, which combines unsupervised anomaly detection problem and network-based inference problem, the study showed that superior analytical methods can greatly help in improving the efficiency of conventional revenue monitoring systems.

The results show that machine learning algorithms, especially ensemble models and graph based models are more effective in detection of high-risk entities with a lower false-positive rate in comparison to traditional rule-based systems. The predictive performance of supervised-learning models was high in cases where there were a number of labeled data, and otherwise, the unsupervised and semi-supervised approaches became vital in the discovery of new and changing trends of non-compliance. The integration of network and relationship

functionalities also enhanced detection skills by exposing well-coordinated laundering patterns and sophisticated evasion plans to taxes which could have not been detected.

Practically, the suggested framework assists in transitioning towards risk-based and intelligence-led enforcement policies in the revenue authorities. Machine learning systems can help audit firms, investigators, and policymakers prioritize cases, allocate resources in the most effective way, and achieve better compliance results in general by computing probabilistic risk scores and interpretable insights. Notably, the paper also highlighted that a balance needs to be struck between predictive accuracy and explainability, fairness and regulatory accountability which are major requirements in the implementation of AI systems in financial regulation in the public sector.

This study has some limitations although it has made its contributions. The analysis was based on simulated and anonymized data, which might not be representative of the diversity and complexity of the real-world financial systems. Moreover, the dynamic and changeable character of financial crime presents persistent problems in the generalization of models and long-lasting efficiency. To mitigate these challenges, future studies ought to be done on realtime detection, privacy-preserving learning methods, and cross-jurisdictional collaborations of data.

Ethical Considerations

Synthetic and anonymous datasets only were used in this study. None of the real taxpayer or secrets on finances were accessed. Thus, institutional ethics approval was not needed.

Acknowledgements

The authors thank their respective institutions for academic and professional support during the preparation of this manuscript.

REFERENCES

Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19-31. https://doi.org/10.1016/j.jnca.2015.11.016

Ajagbe, S. A., Majola, S., &Mudali, P. (2025). Comparative analysis of machine learning algorithms for money laundering detection. Discover Artificial Intelligence, 5(1), 144. https://doi.org/10.1007/s44163-025-00397-4

Al-Shabandar, R., Lightbody, G., Browne, F., Liu, J., Wang, H., & Zheng, H. (2019, October). The application of artificial intelligence in financial compliance management. In Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (pp. 1-6).https://doi.org/10.1145/3358331.3358339

Ardito, L., Filieri, R., Raguseo, E., &Vitari, C. (2025). Artificial intelligence adoption and revenue growth in European SMEs: synergies with IoT and big data analytics. Internet Research, 35(4), 1508-1534. https://doi.org/10.1108/INTR-02-2024-0195

Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2013). Network anomaly detection: methods, systems and tools. Ieee communications surveys & tutorials, 16(1), 303-336. https://doi.org/10.1109/SURV.2013.052213.00046

Cardao-Pito, T. (2025). Fair value accounting and untraceable financial crime. Journal of Financial Crime, 32(3), 661-680. https://doi.org/10.1108/JFC-01-2024-0033

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM computing surveys (CSUR), 41(3), 1-58. https://doi.org/10.1145/3789256

Edwards, J., & Wolfe, S. (2005). Compliance: A review. Journal of Financial Regulation and Compliance, 13(1), 48-59. https://doi.org/10.1108/13581980510622018

Ge, W., De Silva, R., Fan, Y., Sisson, S. A., & Stenzel, M. H. (2025). Machine learning in polymer research. Advanced Materials, 37(11), 2413695. https://doi.org/10.1002/adma.202413695

Gosangi, S. R. (2025). Architecting intelligent invoicing platforms: Leveraging Oracle EBS customization for high-volume revenue management in the public sector. International Journal of Research Publications

in Engineering, Technology and Management (IJRPETM), 8(1), 11798-11809. https://doi.org/10.15662/IJRPETM.2025.0801006

Gritsenko, D., & Wood, M. (2022). Algorithmic governance: A modes of governance approach. Regulation & Governance, 16(1), 45-62. https://doi.org/10.1111/rego.12367

Hanna, M. G., Pantanowitz, L., Jackson, B., Palmer, O., Visweswaran, S., Pantanowitz, J.,..... & Rashidi, H. H. (2025). Ethical and bias considerations in artificial intelligence/machine learning. Modern Pathology, 38(3), 100686. https://doi.org/10.1016/j.modpat.2024.100686

Iguodala, O., &Oyiborhoro, A. (2025). AI-Powered Anti-Money Laundering (AML) and fraud detection-enhancing financial security through intelligent fraud detection. World Journal of Advanced Research and Reviews, 26, 3702-3714. https://doi.org/10.30574/wjarr.2025.26.2.0637

Issar, S., & Aneesh, A. (2022). What is algorithmic governance?. Sociology Compass, 16(1), e12955. https://doi.org/10.1111/soc4.12955

Kadamathikuttiyil Karthikeyan, G., & Bhowmik, B. (2025). Intelligent money laundering detection approaches in banking and E-wallets: a comprehensive survey. Journal of Computational Social Science, 8(4), 91. https://doi.org/10.1007/s42001-025-00421-8

Kothawade, S., & Zellar, I. (2025). An Intelligent Ride-Sharing Algorithm with Integrated Advertising Exposure for Enhanced MaaS Efficiency. International Journal on Advanced Computer Theory and Engineering, 14(1), 46-52. https://doi.org/10.65521/ijacte.v14i1.210

Laxman, V., Ramesh, N., Prakash, S. K. J., &Aluvala, R. (2025). Emerging Threats in Digital Payment and Financial Crime: A Bibliometric Review. Journal of Digital Economy. https://doi.org/10.1016/j.jdec.2025.04.002

Mahesar, A. J., Wighio, A. A., Imtiaz, N., Jamali, A., Nawaz, Y., & Urooj, U. (2025). Predicting tax evasion using machine learning: A study of e-commerce transactions. Spectrum of Engineering Sciences, 3(4), 840-852. https://thesesjournal.com/index.php/1/article/view/311

Mendez, C., & Bachtler, J. (2017). Financial compliance in the European Union: A Cross-National Assessment of financial correction patterns and causes in cohesion policy. JCMS: Journal of Common Market Studies, 55(3), 569-592. https://doi.org/10.1111/jcms.12502

Mousavian, S., & Miah, S. J. (2025). Review of artificial intelligence-based applications for money laundering detection. Intelligent Systems with Applications, 200572. https://doi.org/10.1016/j.iswa.2025.200572

Nayak, S. (2025). Synergizing AI and Quantum Computing to Revolutionize Financial Crime Detection. World Journal of Advanced Research and Reviews, 28(1), xx–xx.. https://doi.org/10.30574/wjarr.2025.28.1.3637

Oliveira, R. M. A., Sant'Anna, A. M. O., & Ferreira, P. H. (2025). Complex networks-based anomaly detection for financial transactions in anti-money laundering. Forensic Science International: Digital Investigation, 55, 302005. https://doi.org/10.1016/j.fsidi.2025.302005

Pang, G., Shen, C., Cao, L., & Hengel, A. V. D. (2021). Deep learning for anomaly detection: A review. ACM computing surveys (CSUR), 54(2), 1-38. https://doi.org/10.1145/3439950

Prorokowski, L., & Prorokowski, H. (2014). Organisation of compliance across financial institutions. Journal of Investment Compliance, 15(1), 65-76.https://doi.org/10.1108/JOIC-12-2013-0041

Ramadhan, S. (2025). Harnessing machine learning for money laundering detection: a criminological theory-centric approach. Journal of Money Laundering Control, 28(1), 184-201. https://doi.org/10.1108/JMLC-04-2024-0083

Saidur, M. J. I. (2025). AI-enhanced business intelligence dashboards for predictive market strategy in US enterprises. International Journal of Business and Economics Insights, 5(3), 603-648. https://doi.org/10.63125/8cvgn369

Shin, D., Fotiadis, A., & Yu, H. (2019). Prospectus and limitations of algorithmic governance: an ecological evaluation of algorithmic trends. Digital Policy, Regulation and Governance, 21(4), 369-383. https://doi.org/10.1108/DPRG-03-2019-0017

Vilella, S., Capozzi, A., Fornasiero, M., Moncalvo, D., Ricci, V., Ronchiadin, S., & Ruffo, G. (2025). Weirdnodes: centrality based anomaly detection on temporal networks for the anti-financial crime domain. Applied Network Science, 10(1), 1-29. https://doi.org/10.1007/s41109-025-00702-1

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020). A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems, 32(1), 4-24. https://doi.org/10.1109/TNNLS.2020.2978386

Younus, M., Manaf, H. A., Nurmandi, A., Mutiarin, D., Sohsan, I., Rehman, A., ... & Minhas, R. (2025). The Role of E-Government in Mitigating Tax Evasion Through Behavioral Profiling of Non-Compliant Taxpayers. In Modeling and Profiling Taxpayer Behavior and Compliance (pp. 271-304). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-0422-9.ch012

Youssef, M. I., Kada, K., Abushanab, Y., Alnawafah, Q., Amano, R. S., & Khasawneh, A. (2025). Utilizing Artificial Neural Networks to Correlate Energy Consumption and Intensity in Metal Industries for the Midwest States. Spectrum of Mechanical Engineering and Operational Research, 2(1), 154-171. https://doi.org/10.31181/smeor21202541

Zenati, H., Romain, M., Foo, C. S., Lecouat, B., & Chandrasekhar, V. (2018, November). Adversarially learned anomaly detection. In 2018 IEEE International conference on data mining (ICDM) (pp. 727-736). IEEE. https://doi.org/10.1109/ICDM.2018.00088

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., ... & Sun, M. (2020). Graph neural networks: A review of methods and applications. AI open, 1, 57-81. https://doi.org/10.1016/j.aiopen.2021.01.001

Zook, M. A.,& Blankenship, J. (2018). New spaces of disruption? Thefailures of Bitcoin andthe rhetorical power of algorithmic governance. Geoforum, 96, 248-255. https://doi.org/10.1016/j.geoforum.2018.08.023