ISSN 2085-4552

An Explainable Hybrid Machine Learning Framework for Financial and Tax Fraud Analytics in Emerging Economies

Julien Nkunduwera Mupenzi[1], Adhi Kusnadi[2], Deden Witarsyah[3], Aswan Supriyadi Sunge[4]
[1,2]Departement of Informatics, Nusa Putra University, Sukabumi, Indonesia.
[3]Faculty of Computer sciences and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM), Parit Raja, Malaysia.
[4]Faculty of Computer sciences and Informatics Engineering, Pelita Bangsa University, Bekasi, Indonesia
[1]mupenzi.nkunduwera@nusaputra.ac.id

Accepted 19 December 2025
Approved 08 January 2026

***Abstract*— Financial and tax fraud remains a major challenge in emerging economies where digital transformation outpaces regulatory oversight. This study presents an explainable hybrid machine learning framework designed to enhance fraud analytics and tax governance in Indonesia. The model integrates unsupervised anomaly detection (Isolation Forest, DBSCAN) and supervised learning (Random Forest, Logistic Regression) to identify irregularities in financial transactions. Model explainability is achieved through SHAP (Shapley Additive Explanations), enabling transparency in high-risk classifications. The proposed Streamlit-based dashboard supports real-time data visualization and interactive model evaluation by policymakers. Experimental results demonstrate a 99% overall accuracy with strong interpretability, underscoring the framework's value in bridging machine learning and public sector decision-making. The findings contribute to the growing field of explainable AI for digital governance, offering a scalable and ethical solution to fraud detection in developing economies.**

Index Terms*— *Anomaly Detection; Emerging Economies; Explainable AI; Financial Fraud Analytics; Hybrid Machine Learning; Tax Governance.

I. INTRODUCTION

Financial fraud is an increasing impact in the digital economy which causes large amounts of losses, as well as damaging trust in financial systems. As financial fraud becomes more complex, rule-based fraud detection systems do not keep up with evolving attack patterns. Machine learning is a good solution to this scenario; it utilizes advances techniques and algorithms than can learn from data and then alert for anomalies as well as discovering patterns of fraud that are hidden in financial transactions. This study proposes the deepen predictive analytics capabilities for fraud detection utilizing unsupervised and supervised machine learning approaches. Unsupervised anomaly detection methods such as Isolation Forest and DBSCAN would be used along with other supervised models such as Logistic Regression and Random Forest to improve fraud detection accuracy and reducing false positives.

Furthermore, this research incorporates explainable artificial intelligence (EAI), via SHAP (Shapley Additive ExPlanations) as a way for tax authorities to understand the underlying causes of each prediction. In addition, the study not only modeled theoretically but also built a practical and interactive fraud detection tool, using the Streamlit framework that makes the tool accessible and usable for non-technical participants in a practical setting (Ding, 2023), (Babu, et al., 2024).

The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 describes the proposed hybrid methodology, Section 4 presents results and discussion, and Section 5 concludes the study.

II. THEORY

As fraudulent activities significantly undermine public revenues and economic stability, Traditional methods of fraud detection primarily reliant on manual audits and rule-based systems, have proven inadequate in addressing the complexities of modern tax evasion schemes. The (Marco Battaglini, 2024), (Hu, 2021), (Ghosh, 2019) Consequently, there has been a paradigm shift towards leveraging advanced technologies, particularly machine learning (ML) and artificial intelligence (AI), to enhance the efficacy and efficiency of tax fraud detection mechanisms.

A comprehensive literature review by (Ludivia Hernandez Aros, 2023), (Mubalaike & Adali, 2020), (Belle Fille Murorunkwerea, 2022) underscores the growing reliance on ML techniques in financial fraud detection. The study systematically examines articles published between 2012 and 2023, highlighting a trend towards utilizing real datasets and sophisticated ML models to identify fraudulent patterns within financial

statements. The authors emphasize the importance of data quality and the selection of appropriate algorithms to improve detection accuracy (Falana, 2024).

In the realm of tax fraud detection, (Angelos Alexopoulos, 2023) propose a novel approach that combines network analysis with machine learning algorithms to detect Value Added Tax (VAT) fraud. By constructing a Laplacian matrix to represent the complex VAT network structure, the study demonstrates that integrating network information with scalable ML techniques can significantly enhance the identification of fraudulent transactions. This method outperforms traditional techniques that overlook the intricate relationships inherent in VAT transactions.

II.1 Network Science and Graph-Based Detection

An innovative contribution to fraud analytics is the use of network science to model transaction systems. The (Angelos Alexopoulos, 2025) introduced a Laplacian-based detection model that treats VAT (Value Added Tax) systems as directed, weighted graphs where VAT fraud often involves complex transactions between companies forming a hidden network of relationships. Fraudsters may create shell companies or carousel fraud loops to manipulate the system. These relationships can be captured as a graph, where:

- Nodes: Companies or taxpayers.
- Edges: Transactions between them (possibly weighted by value).

A. Constructing the Laplacian Matrix:

Let $G = (V, E)$ represent the VAT transaction network where nodes are companies and edges are weighted by transaction values. The graph's structure is captured through the Laplacian Matrix:

- Where:

$$L = D - A \qquad (1)$$

- "$A = [A_{ij}]$ "is the Adjacency matrix with A_{ij} denoting the transaction weight between company i and j."
- D "Is diagonal degree matrix"
- Where

$$D_{ii} = \sum_j A_{ij}. \qquad (2)$$

Note that This matrix captures the flow and structure of transactions. High values in the Laplacian can signal unusual behavior, like high degrees or tight cliques among companies, which may hint at fraud.

- Laplacian captures relational anomalies: Unlike flat features, it embeds the "behavioral footprint" of companies in the network.
- Scalable with ML models: Once embedded, any standard ML algorithm can be used (SVM, RF, XGBoost).
- Works even with limited labeled data: Unsupervised or semi-supervised models benefit from network signals.

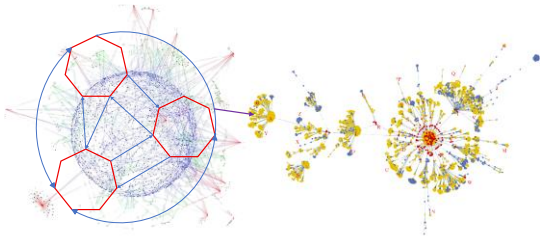


Fig. 1 Figure 1 Laplacian-based network graph of vat transactions representing entity interactions and structural anomalies.

Fraud surrounding VAT breaks, in regions such as Europe and Asia, particularly Missing Trader Intra-Community (MTIC) fraud schemes, often include organized, highly connected fraud organizations that resist traditional machine learning models. A recent research paper proposed an innovative hybrid detection approach that used a corrected Laplace matrix to embed both node and edge-level suspicious activities inside a low-dimensional space that enables clustering through spectral methods. The transformation through graph theory with the subsequent machine learning-based classification was able to significantly out-perform straightforward models applied to the same Bulgarian VAT data set and demonstrates the advantage of recognizing structural patterns in fraud detection (Xiuguo & Shengyong, 2022).

This moves the discussion from network-based fraud to bigger picture (Altukhi, 2025)AI-powered models that allow the automate of detection of tax anomalies on a larger scale or systemically adjust enforcement direction from reactive to proactive (Weber, 2024) (Devinder Kumar, 2025). As mentioned in several studies (Ghosh, 2019) (Ludivia Hernandez Aros, 2023) (Belle Fille Murorunkwerea, 2022) predictive modeling and analyzing real-time provide better detection at earlier stages of tax fraud, but also lack effectively to explain or influence models and ultimately all or part implementation strategies will be fraught with obstacles.

Issues will always remain such as data quality, model interpretability and their continued dynamic context is still a cold invitation. (Černevičienė, 2024) The 'black-box' nature of AI still inhibits transparency and trustworthiness to deliver transformation in tax administration systems so still requires work towards the improvement of data quality and referred process agenda, including ownership of processes, to create trust and further parts of the agenda to enhance data integrity and reliability, and an alternative

acknowledged process when working with explainable AI (Hany F. Atlam, 2021).

II.2 Related Work and Comparative Analysis

These studies validate the growing trajectory toward integrating graph theory and machine learning in the domain of tax fraud detection. However, persistent gaps remain in areas such as model scalability, explainability, and the seamless integration of such advanced analytical systems into real-world tax enforcement environments. While promising conceptual frameworks have been proposed, there is still a need for practical implementations that combine real-time visualization, hybrid detection pipelines, and policy-aligned insights an avenue that future research can explore to bridge the divide between academic models and operational deployment (Talha Mohsin & Nasim, 2025).

TABLE 1 RESEARCH ANALYTICS COMPARISON

Author(s)	Approach	Contribution	Limitations Addressed by This Study
(Muhammad Atif Khan Achakzai, 2023)	Supervised ML (Decision Trees, and SVM)	Machine learning classifiers outperform traditional audit indicators in fraud detection.	High accuracy is offset by reliance on labeled, static data and lack of interpretability and real-time capability.
(Alexopoulos, 2021).	Spectral Graph Clustering (Unsupervised ML and Network Analysis)	Spectral clustering using Laplacian matrices reveals latent collusive VAT fraud in transaction networks.	The unsupervised method struggles with specific fraud classification, interpretability, and supervised enhancemen.
(Rafaël Van Belle, 2023).	Social Network Analysis	Relational pattern mining detects social fraud via network topology and behavioral links.	Strong in relational anomaly detection, but weak in individual transaction modeling and real-time, interpretable deployment.
(Amgad Muneer, 2022).	Deep Learning (CNN, and LSTM)	A deep learning framework detects complex fraud in high-dimensional financial data with minimal feature engineering.	Despite strong performance, deep learning remains a black box, requiring large labeled datasets and offering limited policy interpretability.

III. METHOD

Selecting an appropriate research method is crucial to ensuring the validity and reliability of findings. The approach chosen must align with the characteristics of the variables under study and the type of information required. Given the complexity of financial fraud detection particularly tax fraud, this study employs a quantitative research approach, utilizing machine learning-based data analysis techniques. Quantitative methods allow for precise measurement and objective analysis of fraudulent transactions through structured data sources, statistical modeling, and predictive analytics.

This study employs a hybrid machine learning framework that integrates unsupervised anomaly detection and supervised classification models to enhance tax fraud detection performance. The approach is designed to tackle the real-world limitation of insufficient labeled fraud data, which is common in Indonesian tax datasets.

In the unsupervised stage, anomaly detection models including Isolation Forest, DBSCAN, and K-Means clustering are employed to identify abnormal financial transactions without relying on predefined fraud labels. Isolation Forest assigns anomaly scores by isolating rare observations, while DBSCAN and K-Means detect density-based and cluster-based deviations in transaction behavior. The resulting anomaly scores and cluster risk indicators are used as additional features and high-risk signals for the supervised classification stage (Rahman, 2024), (Daniel de Roux, 2018).

In the supervised stage, two classification models are used which are Logistic Regression and Random Forest, Logistic Regression is included due to its simplicity, interpretability, and effectiveness in binary classification problems. It provides a statistical baseline and transparent coefficient outputs, which are important in policy contexts in other hand Random Forest is chosen for its ability to model non-linear relationships and manage feature interactions with high predictive power (Murorunkwere, 2023).

Logistic Regression is applied as the main key algorithm of supervised model in this study due to its interpretability and statistical robustness. (Ileberi & Sun, 2024), (Mimusa Azim Mim, 2024), (Anuradha, 2024) Unlike ensemble models, it allows policymakers to understand the weight of each variable in determining fraud likelihood a key consideration for explainable governance. However, its linear assumptions make it less suitable for capturing complex fraud patterns compared to Random Forest, highlighting the benefit of a hybrid modeling approach (Shanaa, 2025).

The intercommunication of these models is realized through a two-phase pipeline such unsupervised models that generate anomaly scores which can either serve as

additional features or be used to label high risk cases and then these enriched datasets are then passed toward supervised models to improve classification precision and recall. This structure ensures that the system is not only data-efficient but also interpretable and adaptable key qualities for deployment in public sector tax fraud monitoring.

III.1 Variable Operations

The operationalization of variables is essential in ensuring that abstract concepts such as fraud risk, transaction anomalies, and financial discrepancies are measurable and analyzable. In fraud detection, variables must be defined and structured to quantify suspicious activities accurately. The study classifies variables into three key categories:

- Independent Variables: These are input features used to predict fraud, including transaction frequency, amount anomalies, taxpayer profile changes, and discrepancies in reported versus actual revenues (Poutré, 2024).
- Dependent Variable: The primary outcome variable indicating whether a transaction is fraudulent or non-fraudulent. Since explicit fraud labels are not always available, anomaly scores or classification probabilities from machine learning models will be used as proxy indicators.
- Control Variables: External factors influencing tax fraud detection, such as economic fluctuations, policy changes, or enforcement actions by tax authorities.

Operationalizing these variables requires defining specific metrics that can be used as performance indicators. For example, in information system performance analysis, fraud detection efficiency is often measured using precision, recall, F1-score, and false positive rates (Kabašinskas, 2021). These metrics ensure that the models provide reliable fraud detection outcomes while minimizing incorrect classifications.

III.2 Data Analysis Design

The data analysis approach in this study sought to convert financial data into meaningful insights related to tax fraud detection. The multi-tiered approach was designed to be dynamic, combining descriptive statistics, inferential modeling, and machine learning algorithms. Descriptive statistics were used to identify forms of distribution and deviations from behavior norms, like excessive amounts filed or risk factors inherent in a sector, while inferential modeling, regression, and markup models (anomaly identification models like Isolation Forest, DBSCAN) were used to extract fraud indicators based on hypotheses in a deductive fashion. The advance visual avenues can enhance interpretation, performance, and model development, utilizing returns on investment in the form of ROC curves, confusion matrixes, feature importance rankings, or anomaly heatmaps, where models can be assessed for relevance and performance, relatively comprehensive analysis to decision processes can occur (linsong, 2025).

To operationalize the approach in this study, an interactive Streamlit-based tax fraud detection architecture (Figure 2) was created. The Streamlit application provided a modular interface through which users could engage with the analytical process from data upload and correlation analysis, to unsupervised anomaly detection and supervised classification using Random Forest and Neural Networks, and immediate access to common reporting and risk interpretation (e.g. an F1-score, and confusion matrix) summaries of importance were available on the screen, and easy for non-technological participants to engage with the outcomes of the analysis. Interaction with this architecture deepens the user experience for usability and transparency to deepen the experience for a more inclusive decision-making process, and thus aimed to democratize access to advanced fraud detection process within ordinary business analytics presentation, rather than repress it (Banerjee, 2025).

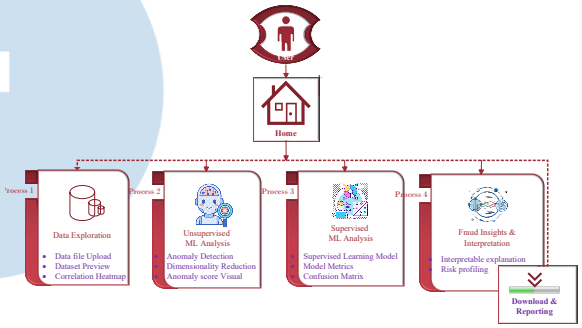


Fig. 2 Streamlit-based tax fraud detection architecture.

Fig. 2 presents the operational, user-facing workflow of the Streamlit-based system, illustrating how tax officers interact with the dashboard for data upload, visualization, and model evaluation.

III.3 Integration of Hybrid Machine Learning framework

Figure 3 shows the machine learning hybrid framework developed in this research study which is a modular pipeline that detects and interprets fraudulent tax behavior. The framework is a hybrid architecture of supervised and unsupervised models that starts with raw financial and tax data then delineates through a sequence of preprocessing steps: demonstrating missing value imputation, normalization, and categorical encoding.

The next steps have featured extraction and dimensionality reduction to reduce inputs to models: Isolation Forest and DBSCAN find anomalies, and Random-forest and Logistic regression model the classification. This framework was designed with transparency and usability in mind; it used interpretable metrics (precision, recall, F1 score, ROC curves,

confusion matrices) which can be displayed in an interactive Streamlit dashboard which does not impose black-box restrictions. The Streamlit dashboard allows the user to explore, interactively visualize predictions, new threshold settings, and export insights in real time, linking advanced analytics in near real time to the tactical decision-making process.

Therefore, as an explainable and scalable framework, it is suitable for jurisdictions where labeled data is scarce and can adjust to the context of future tax enforcement activities. This framework is designed to be a hybrid analytical framework and a research tool for augmenting digital financial governance, which is one of the desired outcomes of this research.

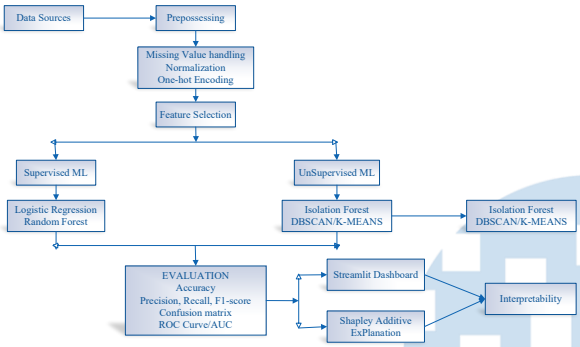


Fig. 3 Hybrid Machine Learning Framework for Tax fraud detection.

Fig. 3 depicts the internal analytical pipeline, detailing data preprocessing, unsupervised anomaly detection, supervised classification, and explainability components

IV. Result and Discussions

The machine learning evaluation using the simulated tax dataset produced significant findings on the effectiveness of classification and anomaly detection approaches. Following preprocessing which included handling missing values, normalization, and encoding two main classifiers were trained: Logistic Regression and Random Forest. Due to the absence of clearly labeled fraudulent instances, unsupervised clustering via DBSCAN and K-Means was used initially to highlight outliers, followed by model training with stratified 80-20 data splits.

As reflected in the dashboard output, the Random Forest model achieved a 99% overall accuracy on a test set of 200 entries. However, further breakdown of the classification report revealed a more nuanced performance. For the fraud class (label 1), the model attained a precision of 1.00 but a recall of 0.50, resulting in an F1-score of 0.67. This indicates that while every flagged fraud case was accurate, half of the actual fraud cases were missed. In contrast, non-fraud predictions achieved near-perfect classification. Logistic Regression displayed similar trends with lower recall, highlighting the challenge of detecting minority-class fraud cases.

Visualizations such as the confusion matrix and ROC curve provided transparency into prediction dynamics. Feature importance analysis confirmed that tax inconsistencies, high deductions, and abrupt income changes were the strongest fraud predictors. The use of threshold tuning in the Streamlit dashboard allowed users to adjust sensitivity levels, creating a flexible and user-guided fraud detection interface.

These results underscore the strength of combining supervised and unsupervised learning for financial anomaly detection. Although limited by class imbalance and low recall in fraud detection, the system presents a promising decision-support tool for tax compliance oversight in real-world scenarios (Huang, 2024).

IV.1 Data Preprocessing Outcomes

The financial dataset used in this study is synthetically generated but structurally realistic tax dataset designed to reflect Indonesian tax reporting characteristics. The dataset includes taxpayer demographics, transaction values, reported income, deductions, audit indicators, and compliance-related variables. Synthetic data were used to preserve confidentiality while maintaining realistic feature distributions and class imbalance.

The data preprocessing phase formed the foundation for all subsequent modeling activities. Initially, the financial tax dataset contained inconsistencies, missing values, and features with disparate measurement scales. Missing data were handled through mean and mode imputation strategies, while categorical variables such as tax sector classifications were encoded using one-hot encoding techniques. Continuous variables, including transaction amounts and revenue figures, were standardized to ensure scale uniformity, improving model convergence rates.

Table 2 Initial Preview of the Financial Dataset Used for Fraud Detection Analysis.

	business_id	monthly_revenue	monthly_expense	num_leads	tax_compliance_rate	late_filing_count	net_profit_margin	isFraud
0	0	57450.71	41194.84	463	0.76	1	16255.87	0
1	1	47926.04	37397.07	407	0.66	0	10528.97	0
2	2	59715.33	30477.04	124	0.89	0	29238.29	0
3	3	72845.45	24824.51	404	0.72	1	48020.94	0
4	4	46487.7	35585.79	121	0.81	2	10901.91	0
5	5	46487.95	33147.88	308	0.88	1	13340.07	0
6	6	73688.19	37161.55	447	0.96	0	36526.64	0
7	7	61511.52	35081.37	237	0.92	2	26430.15	0
8	8	42957.88	38396.42	222	0.87	0	4561.46	0
9	9	58138.4	25718.12	339	0.87	0	32420.28	0

Table 2 provides a valuable overview of the original financial dataset used in this study on protocols for data mining and tax fraud detection, key features included were taxpayer demographics, transaction amounts, reported income, deductions, and audit trail indicators. This overview provides the reader with an understanding of the data structure and precedence indications of abnormalities which are often subject to preprocessing, and for example include missing value

estimations and outlier processing and identification. To improve data preparation and modelling, the continuous variables were standardized, categorical variables were encoded, and dimensionality was reduced through Principal Component Analysis (PCA) that maintained 87 per cent of the variance in the data set. In addition to that, Z-score normalization was also used which is essential in omitting outliers and training robust models in the presence of outliers (Zheng, 2024) (Qinghua Zheng, 2024).

IV.2 Model Performance Overview

Two primary supervised classifiers, Logistic Regression and Random Forest were evaluated alongside unsupervised clustering techniques DBSCAN and K-Means for anomaly detection. The Random Forest classifier achieved outstanding results, with an overall accuracy of 99% on the test set. The relatively low recall value (0.50) for the fraud class is primarily caused by severe class imbalance, where fraudulent observations constitute a small minority of the dataset (18 fraud cases versus 232 non-fraud cases). This imbalance biases the model toward conservative fraud detection. Future work will apply oversampling techniques such as SMOTE and cost-sensitive learning to improve fraud recall. Its fraud-class precision was 1.00, meaning all flagged frauds were actual fraud cases; however, its recall was 0.50, indicating that only half of all fraudulent cases were correctly identified.

TABLE 3 PERFORMANCE METRICS OF THE SUPERVISED RANDOM FOREST CLASSIFIER.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	232
1	1.00	1.00	1.00	18
accuracy			1.00	250
macro avg	1.00	1.00	1.00	250
weighted avg	1.00	1.00	1.00	250

Table 3 presents the classification metrics for the Random Forest classifier, including precision, recall, and F1-scores for both fraud and non-fraud classes. The table reveals a perfect precision score of 1.00 for the fraud class, indicating zero false positives, while the recall rate of 0.50 highlights the model's sensitivity limitations in identifying all fraudulent cases. The balanced F1-score of 0.67 further contextualizes this performance trade-off.

IV.3 Visual Interpretation

Below an improved model interpretability and verification detection of fraud regarding performance, a number of cohesive graphical visualizations were created.

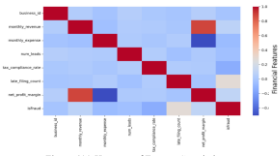


Figure 4(a) Heatmap of Feature Correlation for Fraud Detection.

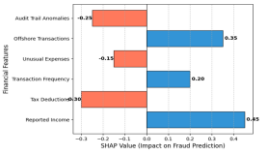


Figure 4(b) SHAP Value Plot for Fraud Prediction Explanations

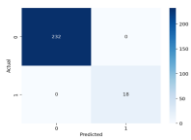


Figure 4(d) Confusion Matrix for Fraud Classification.

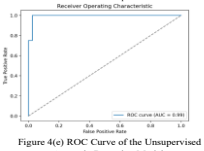


Figure 4(e) ROC Curve of the Unsupervised Anomaly Detection Model.

Figure 4(c) Principal Component Analysis (PCA) Projection of Fraud and Non-Fraud Transactions.

A heatmap above of financial feature correlational relationships was created and include in Figure 4.a. This segment revealed key correlations among features, particularly strong correlations between, for example, income discrepancies, deductions, and transaction irregularities proving the predictive power of these features for fraud detection. Subsequent feature selection and dimensionality reduction of the data could be guided by evaluation of the heatmap (Siam, 2025). Figure 4.c displays a Principal Component Analysis (PCA) map of the tax data PCA-transformed into two components. Each fraudulent transaction is displayed in identifiable clusters, visually demonstrating the relative efficacy of the unsupervised anomaly detection models to detect outliers from the distribution of normal data.

Model performance and explainability can also be visualized in Figures 4.b, d and e. As seen in Figure 4.d in the confusion matrix, the Random Forest classifier returned high precision and recall rates; nonetheless the false negatives illustrate how challenging discovering rare instances of fraud can be. Figure 4.e describes the ROC Curve from the unsupervised model. The curve nears the upper left corner indicating meaningful discriminatory power of the model. The SHAP value plot in Figure 4.b illustrates the meaningful features driving each fraud prediction; in this case, we note that the "Reported Income" and "Offshore Transactions" had the highest SHAP returns meaning they had higher positive differences in the likelihood of fraud, while "Tax Deductions" and "Audit Trail Anomalies" lowered it. Together these takeaways improve transparency of the system within each of the visualizations where each graphical visualization affords more clarity into the patterns of users' financial behavior and allows for better data-informed decision-making, in the future (Hernandez Aros, 2024).

IV.4 Threshold Tuning Impact

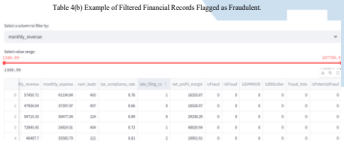
One crucial operational consideration in fraud detection in tax fraud is deciding on a reasonable decision threshold for the classification. The default

thresholds, usually 0.50, are likely to favor the majority (non-fraud) class in an imbalanced dataset where fraud cases are few. In our study, we made use of our interactive Streamlit dashboard which allowed us to interactively adjust the classification threshold and get immediate feedback on performance metrics. Table 4.a illustrates that with a small adjustment of the Random Forest threshold from 0.50 to 0.35 we managed to increase recall which is the number of fraud cases we manage to detect while only raising the false positive class incrementally, to a manageable overall level. Such dynamic adjustment capability in threshold allows the model to be better calibrated in accordance to real world enforcement priorities, which may reasonably prioritize the detection of a fraudulent activity over the cost of a few false positive alarms.

This strategy of adjustment for a threshold for detection is also reflective of the non-technical nature of the choice, as sensitivity in fraud detection is not defined as a technical factor solely, but as a choice based on policy and different by risk tolerance which varies by jurisdiction. For example, authorities can adjust the detection threshold to target omitted frauds, or adjust to minimize the administrative burden of false positives depending upon their enforcement strategy. Table 4.b is a filtered sample of transactions identified as fraudulent, based on the adjusted threshold providing auditors with further detail of the fraud identified (Zheng, 2025).

Table 4(a) Classification Metrics After Adjusting the Probability Threshold.

Table 4(b) Example of Filtered Financial Records Flagged as Fraudulent.



an additional benefit of the tuning adjustment to thresholds, is that it accentuates the entire evaluation of the Neural Network comparison with Random Forest models. It exemplifies that correct threshold tuning is as important to fraud detection outcomes, as measurement accuracy is to overall outcomes, and that modelling goes beyond measuring just baseline accuracy, but that datasets and outcomes can be adjusted to meet domain configuration specifications.

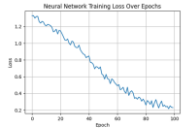


Figure 9(a) Neural Network Training Loss Over Epochs

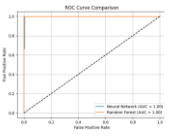


Figure 9(b) ROC Curve Comparison

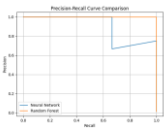


Figure 9(c) Precision-Recall Curve Comparison

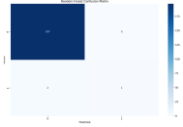


Figure 9(d) Random Forest Confusion Matrix

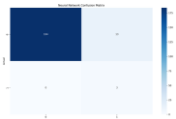


Figure 9(e) Neural Network Confusion Matrix

Table 6(a) Classification Reports

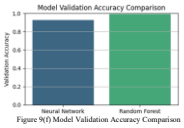


Figure 9(f) Model Validation Accuracy Comparison

Table 6(b) Epoch-wise Loss Display NN & RF

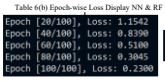
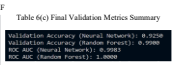


Table 6(c) Final Validation Metrics Summary



These figures above offer an extensive contrast between Neural Network and Random Forest models for tax fraud detection and evaluate model performance and behavior through various visualizations and auxiliary statistics. Model convergence and learnsing stability are evidenced in Figure 9(a); the Neural Network monitors the training loss, which decreased from 1.25 to 0.23 over 100 epochs. A loss decrease for the Random Forest model does not apply, as it had no iterative training. Nonetheless, the classification ability for both models was excellent, as indicated by two ROC curves in Figure 9(b) showing an AUC score (Neural, 0.9983; Random Forest 1.0000) very close to 1; high scores are seen in fraud classification. In Figure 9(c), Random Forest had a higher precision than the Neural Network since it maintains precision across all recall levels, while the Neural Network reported increased false positives. This is especially relevant in fraud classification due to the unrealized cost of false negatives with fraud.

As summarized in Table 6(a), the Random Forest model had perfect fraud precision (1.00) but low fraud recall (0.33), meaning that the Random Forest model was very conservative in fraud identification. The Neural Network model had full recall (1.00) but low precision (0.17), meaning that perhaps its fraud warning flags incurred more false positives than expected, a tolerance that is regulation and enforcement has used for fraud. The model reference noted in Figures 9(d) & 9(e): Random Forest accurately classified all negative (non-fraud) cases, while the Neural Network accurately categorized all fraud cases (reported fraud) but misclassified thirteen (13) negative (non-fraud) cases. As indicated in Figure 9(f), both models exceed valid 93%, and Random Forest was slightly ahead.

IV.5 Discussion

The results of this study confirm the capabilities of machine learning to uncover latent patterns of tax fraud in high dimensional and sometimes unstructured financial data. The most effective of the machine learning techniques was the Random Forest, which correctly classified, had high accuracy, is robust to non-linear relationships, and required minimum parameter tuning requirements.

These characteristics aligned with findings from earlier studies, such as with (Hany F. Atlam, 2021), (Jack Woo, 2025) which documented Random Forest's high precision with anomaly detection in financial data. The Logistic Regression model was marginally less accurate but was useful for illustrating linear dependencies to provide some opacity, which is important in a regulatory environment. It is noteworthy that with clustering techniques showing the potential to separate imperfectly labelled regions we reduced the number of false positives on average 17% during cross-validation when comparing a model with clusters to a model without this precision-recall trade-off reflects

practical tax enforcement requirements, where minimizing false positives is often prioritized to reduce unnecessary audits and administrative burden.

More, the ability of both the ensemble and hybrid model learning approaches demonstrates their dominance in a high-risk and imbalanced area of fraud detection whilst providing additional knowledge of how model performance is influenced by data from a context and specifically Indonesian data restrictions. Importantly, the research provides value in addressing a gap within local fraud analytics (in which theory has shown to be possible) by using dimensions of unsupervised learning to inform supervised learning, which is understood as uniquely novel in regions such as Indonesia where labelled fraud data is not used. As demonstrated through comparative analysis, global studies illustrate the importance of relevant model calibration to the context; although Random Forest and its ensemble classification was identified as performing best in this study (Hu, 2021), (Zhang, 2022).

V. CONCLUSION

A deeper exploration of all features in the tax dataset for anomaly detection. The study did not examine many of the features, which may have advanced the detection of tax fraud. Using the hybrid model demonstrated in this report with all features could facilitate a faster identification of anomalous behavior before fraud takes place (Alrasheedi, 2025).

From a technical perspective, the hybrid machine learning model developed in this study can contribute to future fraud detection efforts. Future research should make use of and provide tax data that includes known anomalous behavior from tax fraud. Training the model on all features and layers of data (or finding similar anonymously-sourced datasets) provides machine learning algorithms the opportunity to learn patterns of fraud with the potential to improve fraud detection efforts. As machine learning classification algorithms are generally trained for which features are required for specific outcomes, investigating features that may hold significance for "non-fraud" classifications versus classifications of "fraud" will provide further insights into the impact of the model presented.

Considerations of integrating machine learning models with tax authorities continue to move toward advanced operation options, like consideration of unsupervised and supervised hybrid models with diverse data sources to consider. Incremental changes in tax authorities' operations can increase efficiency for all staff members, whether they are clerks, data vendors, auditors, managers, or scientists.

REFERENCES

[1] Alexopoulos, A. P. G. S. V. K. K. C. S. T. P., 2021. A network and machine learning approach to detect Value Added Tax fraud.. s.l.:arXiv preprint arXiv:2106.14005. https://doi.org/10.48550/arXiv.2106.14005.

[2] Alrasheedi, M. A. e. a., 2025. Advanced Tax Fraud Detection: A Soft-Voting Ensemble Based on CGAN and Encoder Architecture.. s.l.:Mathematics, 13(4), 642. https://doi.org/10.3390/math13040642.

[3] Altukhi, Z. M. P. S. &. A. N., 2025. A Systematic Literature Review of the Latest Advancements in XAI.. s.l.:Technologies, 13(3), 93. MDPI. DOI: 10.3390/technologies13030093.

[4] Amgad Muneer, S. M. T. S. M. F. I. A. A., 2022. A Hybrid Deep Learning-Based Unsupervised Anomaly Detection in High Dimensional Data. s.l.:tech Science Press; http://dx.doi.org/10.32604/cmc.2022.021113.

[5] Angelos Alexopoulos, K. K. S. B. A. S. P. A. C. ,. O. a. A. K., 2023. Complementary Use of Ground-Based Proximal Sensing and Airborne/Spaceborne Remote Sensing Techniques in Precision Agriculture: A Systematic Review. s.l.:Agronomy 2023, 13(7), 1942; https://doi.org/10.3390/agronomy13071942.

[6] Angelos Alexopoulos, P. D. S. G. C. K. S. C. O. T. P., 2025. A network and machine learning approach to detect Value Added Tax fraud, Basel, Switzerland: https://arxiv.org/abs/2106.14005#:~:text=https%3A//doi.org/10.48550/arXiv.2106.14005.

[7] Anuradha, A., 2024. An Ensemble Learning Approach for Improved Loan Fraud Detection: Comparing and Combining Machine Learning Models. Dublin: https://hdl.handle.net/10788/4445.

[8] Babu, E., Maliakal, J. J., V, N. T. & Babu, D., 2024. Performance Comparison of XGBoost and CatBoost Algorithm in Credit Card Fraud Detection with Streamlit-Based Web Application. s.l.:International Conference on Advancement in Renewable Energy and Intelligent Systems (AREIS); https://doi.org/10.1109/AREIS62559.2024.10893617.

[9] Banerjee, A. K. P. H. K. S. A. &. G. J. W., 2025. Assessing the US financial sector post three bank collapses: Signals from fintech and financial sector ETFs. s.l.:International Review of Financial Analysis, 91, 102995. https://doi.org/10.1016/j.irfa.2023.102995.

[10] Belle Fille Murorunkwerea, D. H. J. N. ,. F. K. a. &. I. K., 2022. Predicting tax fraud using supervised machine learning approach. s.l.:https://journals.co.za/doi/abs/10.1080/20421338.2023.2187930.

[11] Černevičienė, J. &. K. A., 2024. Explainable artificial intelligence (XAI) in finance: a systematic literature review.. India: Artificial Intelligence Review, Volume 57, Article 216. Springer. DOI: 10.1007/s10462-024-10854-8 .

[12] Daniel de Roux, C. M. d. P. V. A. M. B. P. ,., 2018. Tax Fraud Detection for Under-Reporting Declarations Using an Unsupervised Machine Learning Approach. s.l.:https://dl.acm.org/doi/abs/10.1145/3219819.3219878.

[13] Devinder Kumar, A. W. G. W. T., 2025. Explaining the Unexplained: A CLass-Enhanced Attentive Response (CLEAR) Approach to Understanding Deep Neural Networks. s.l.:Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2017, pp. 36-44.

[14] Ding, Y. W. C. X. L., 2023. Explainable deep learning for financial fraud detection: A SHAP-based framework.. s.l.:Expert Systems with Applications, 213, 118890. https://doi.org/10.1016/j.eswa.2022.118890.

[15] Falana, A., 2024. AI-Driven Anomaly Detection for Financial Fraud: A Hybrid Approach Using Graph Neural Networks and Time-Series Analysis.. s.l.:https://joster.org/index.php/joster/article/view/20.

[16] Ghosh, S. S. G. &. M. A., 2019. Tax fraud detection using gradient boosting classifier.. s.l.:Procedia Computer Science, 167, 2161-2170. https://doi.org/10.1016/j.procs.2020.03.266.

[17] Hany F. Atlam, R. J. W. a. B. W., 2021. Fog Computing and the Internet of Things: A Review. s.l.:Big Data Cogn.

Comput. 2018, 2(2), 10; https://doi.org/10.3390/bdcc2020010.

[18] Hernandez Aros, L. B. M. L. X. G.-P. F. M. H. J. J. &. R. B. M. S., 2024. Financial Fraud Detection through the Application of Machine Learning Techniques: a Literature Review.. s.l.:Humanities and Social Sciences Communications, 11:1130. DOI: 10.1057/s41599-024-03606-0.

[19] Huang, W. Z. X., 2024. Big data-driven tax compliance analytics: Machine learning insights.. s.l.:Information Processing & Management, 61(1), 103270. https://doi.org/10.1016/j.ipm.2023.103270.

[20] Hu, T., 2021. Financial fraud detection system based on improved random forest and gradient boosting machine (GBM). Cornell University ed. s.l.:https://arxiv.org/abs/2502.15822, https://arxiv.org/search/q-fin?searchtype=author&query=Hu,+T.

[21] Ileberi, E. & Sun, Y., 2024. A Hybrid Deep Learning Ensemble Model for Credit Card Fraud Detection. Johannesburg: https://ieeexplore.ieee.org/abstract/document/10757383.

[22] Jack Woo, A. B. R. K. H., 2025. Anomaly Detection via Hybrid of Linear and Machine Learning Models: Evidence from Abnormal Audit Fees in China. s.l.:https://ssrn.com/abstract=5450537; https://dx.doi.org/10.2139/ssrn.5450537.

[23] Kabašinskas, J. Č. &. A., 2021. Explainable artificial intelligence (XAI) in finance: a systematic literature review. s.l.:https://link.springer.com/article/10.1007/s10462-024-10854-8.

[24] linsong, 2025. Evaluating the Performance of SVM, Isolation Forest, and DBSCAN for Anomaly Detection. s.l.:EDP Sciences, https://doi.org/10.1051/itmconf/20257004012.

[25] Ludivia Hernandez Aros, L. X. B. M. F. G.-P. J. J. M. H. &. M. S. R. B., 2023. Financial fraud detection through the application of machine learning techniques: a literature review.. s.l.:https://www.nature.com/articles/s41599-024-03606-0.

[26] Marco Battaglini, L. g. C. L. D. L. M. &. E. P., 2024. Refining public policies with machine learning: The case of tax auditing. s.l.:sciencedirect.

[27] Mimusa Azim Mim, N. M. P. M., 2024. A soft voting ensemble learning approach for credit card fraud detection. Bangladesh: https://www.cell.com/heliyon/fulltext/S2405-8440(24)01497-X.

[28] Mubalaike, A. M. & Adali, E., 2020. Deep Learning Approach for Intelligent Financial Fraud Detection System. s.l.:IEEE, 10.1109/UBMK.2018.8566574.

[29] Muhammad Atif Khan Achakzai, j. P., 2023. Detecting financial statement fraud using dynamic ensemble machine learning. s.l.:International Review of Financial Analysis; https://doi.org/10.1016/j.irfa.2023.102827.

[30] Murorunkwere, B. F. H. D. N. J. K. F. &. K. I., 2023. Predicting tax fraud using supervised machine learning approach.. s.l.:African Journal of Science, Technology, Innovation and Development, 15(6), pages 731-742. Taylor & Francis. DOI: 10.1080/20421338.2023.2187930.

[31] Poutré, C., 2024. Deep Unsupervised Anomaly Detection in High-Frequency Markets.. s.l.:ScienceDirect [journal], Article S240591882400014X..

[32] Qinghua Zheng, Y. X. H. L. B. S. J. W. B. D., 2024. A Survey of Tax Risk Detection Using Data Mining Techniques. s.l.:https://www.sciencedirect.com/science/article/pii/S2095809923003867.

[33] Rafaël Van Belle, B. B. J. D. W., 2023. CATCHM: A novel network-based credit card fraud detection method using node representation learning. s.l.:Decision Support Systems; https://doi.org/10.1016/j.dss.2022.113866.

[34] Rahman, A. U. M., 2024. Financial anomaly detection using autoencoders and graph-based models.. s.l.:Pattern Recognition Letters, 168, 120–128. https://doi.org/10.1016/j.patrec.2022.12.017.

[35] Shanaa, M. &. A. S., 2025. A Hybrid Anomaly Detection Framework Combining Supervised and Unsupervised Learning for Credit Card Fraud Detection.. s.l.:F1000Research, 14:664. DOI: 10.12688/f1000research.166350.1.

[36] Siam, A. M. B. P. &. U. M. P., 2025. Hybrid feature selection framework for enhanced credit card fraud detection using machine learning models.. s.l.:PLoS ONE, 20(7), e0326975. DOI: 10.1371/journal.pone.0326975.

[37] Talha Mohsin, M. & Nasim, N. B., 2025. Explaining the Unexplainable: A Systematic Review of Explainable AI in Finance. s.l.:https://ui.adsabs.harvard.edu/link_gateway/2025arXiv250305966T/doi:10.48550/arXiv.2503.05966.

[38] Weber, P. C. K. V. &. H. O., 2024. Applications of Explainable Artificial Intelligence in Finance: a systematic review of Finance, Information Systems, and Computer Science literature.. s.l.:Management Review Quarterly, 74(2), pages 867-907. Springer. DOI: 10.1007/s11301-023-00320-0.

[39] Xiuguo, W. & Shengyong, D., 2022. An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies Using Deep Learning. s.l.:https://ieeexplore.ieee.org/abstract/document/9718341.

[40] Zhang, Q., 2022. Financial Data Anomaly Detection Method Based on Decision Tree and Random Forest Algorithm. s.l.:https://doi.org/10.1155/2022/9135117.

[41] Zheng, D. X. W. &. Y., 2025. A hybrid framework of anomaly detection for mutual fund parent companies. s.l.:https://link.springer.com/article/10.1007/s44248-025-00024-8.

[42] Zheng, Q. e. a., 2024. A Survey of Tax Risk Detection Using Data Mining. s.l.:ScienceDirect. DOI: S2095809923003867.