FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 10, October, 2025, pp 5 – 11
DOI: https://doi.org/10.33003/fjs-2025-0910-4004

FRAUD DETECTION IN NIGERIAN INVESTMENT ADVISORY SECTOR USING MACHINE LEARNING ALGORITHMS

***[1]Fatokun Johnson O., [2]Mustapha Sikiru, [2]Funmilola Balogun and [3]Okorie Daberechi D.**

[1]Department of Mathematics, Anchor University Lagos, Nigeria
[2]Department of Mathematics, Faculty of Physical Sciences, Federal University Dutsin-Ma, Katsina, Nigeria
[3]Department of Computing, Faculty of Science, Anchor University, Lagos, Nigeria

*Corresponding authors' email: jfatokun@aul.edu.ng

ABSTRACT

The increasing rate of financial fraud within Nigeria's investment advisory sector presents a pressing challenge, particularly as traditional detection methods struggle to keep pace with evolving fraudulent behaviors. With the growing reliance on technology in the financial ecosystem, machine learning offers a promising solution for detecting anomalies and minimizing risk. This study explores the application of three machine learning algorithms: Logistic Regression, Random Forest, and XGBoost to predict fraudulent transactions in the Nigerian investment advisory landscape. To address class imbalance which is a common issue in fraud datasets, hybrid resampling using SMOTE and Tomek Links was implemented. The Random Forest model emerged as the most robust, maintaining consistent performance across key metrics even after resampling. This research also emphasizes the importance of integrating cost-sensitive learning, regular model retraining, and cross-validation to adapt to Nigeria's dynamic fraud landscape. The study presents a scalable approach to enhancing fraud detection systems and safeguarding investor trust in the sector.

Keywords: Fraud detection, Nigerian investment, Random Forest, Machine learning, Class imbalance

INTRODUCTION

The financial sector in Nigeria has witnessed significant growth over the past decade, driven by the increasing participation of both local and international investors (AGYA et al., 2018). Investment advisory services have become a critical component of this sector, offering expert guidance to individuals and institutions on asset allocation, risk management, and portfolio diversification. However, the rise in investment activities has also led to a corresponding increase in fraudulent schemes targeting unsuspecting investors. Financial fraud is the act of gaining financial benefits by using illegal and fraudulent methods (Ashtiani & Raahemi, 2021). These fraudulent activities range from Ponzi schemes and insider trading to misrepresentation of investment products and outright scams (Čunderlík, 2021; Gui et al., 2024). Investment advisors in Nigeria play a pivotal role in safeguarding the financial interests of their clients. They are responsible for conducting thorough due diligence on investment opportunities and ensuring that their clients' funds are allocated to legitimate and profitable ventures. However, the traditional methods of fraud detection employed by many advisors are increasingly proving inadequate in the face of sophisticated and rapidly evolving fraudulent schemes. These methods, often reliant on manual checks and intuition, are not only time-consuming but also prone to human error, leading to significant financial losses for investors (Bello & Olufemi, 2024; Hilal et al., 2022; Martins & Fonkem, 2024). Thus, the adoption of more advanced and efficient fraud prediction tools is necessary.

Machine learning, a subset of artificial intelligence, offers promising solutions to this challenge. By analyzing large volumes of data and identifying patterns that may be indicative of fraudulent behavior, machine learning models can provide investment advisors with a powerful tool for early fraud detection. These models can be trained to recognize subtle signs of fraud that may be missed by human analysts, thereby enhancing the overall effectiveness of fraud prevention efforts (Lokanan & Sharma, 2022). In the Nigerian context, the need for such advanced tools is particularly pressing. The country's financial markets are characterized by high volatility, limited regulatory oversight, and a significant level of informal investment activities, all of which create a conducive environment for fraudulent schemes (Akinninyi et al., 2025). As Nigeria continues to attract both domestic and foreign investment, the ability of investment advisors to effectively detect and prevent fraud will be crucial in maintaining investor confidence and ensuring the long-term stability of the financial sector. Future advancements in artificial intelligence are likely to yield more advanced algorithms that can detect not only existing forms of fraud but also identify new and emerging patterns. These advancements will be driven by the continuous growth in the volume and diversity of data available for training and refining machine learning models (Odufisan et al., 2025).

Fraud detection using machine learning significantly enhances the accuracy and efficiency of detecting fraudulent activities across various industries. Machine learning algorithms can process vast amounts of data and identify complex patterns that traditional rule-based systems may overlook, leading to improved real-time detection and a substantial reduction in false positives. This not only helps in minimizing financial losses but also strengthens customer trust by providing robust security measures (Alonge et al., 2021; Mohamed, 2025). Moreover, the adaptability of machine learning models to evolving fraudulent tactics ensures continuous protection against new threats, making this research approach crucial for maintaining the integrity of financial systems and other vulnerable sectors.

Nigerian financial sector has experienced a significant rise in fraudulent activities targeting investors, particularly within the realm of investment advisory services. The inability of existing fraud detection techniques to accurately and efficiently identify fraudulent investment schemes poses a substantial risk to investors, leading to financial losses and eroding trust in the financial markets (Alonge et al., 2021). Furthermore, the high prevalence of fraud undermines the effectiveness of investment advisors, who are responsible for

safeguarding their clients' assets and ensuring the integrity of investment decisions.

There is a pressing need for advanced, data-driven solutions that can enhance the fraud detection capabilities of Nigerian investment advisors. Machine learning, with its ability to analyze large datasets and identify patterns indicative of fraud, offers a promising approach to address this challenge (Mohamed, 2025). However, the successful application of machine learning in this context requires the development of models that are specifically tailored to the unique characteristics of the Nigerian financial market and the types of fraud prevalent within it.

The remainder of this paper is thus: Section 2 provides an overview of existing literature on fraud detection, the application of machine learning in financial systems, as well as other related works. Section 3 discusses the methodology of the research, detailing the hardware requirements, software requirements, system requirements, and working procedures in this project. Section 4 presents and discusses the results obtained from the implemented models, while section 5 concludes the paper.

Overview of Literature

A brief background regarding previous works on investment fraud, the application of machine learning in the detection of fraud in financial systems, and related works is presented here.

According to the Nigeria Inter-Bank Settlement System (NIBSS) 2023 Annual Fraud Landscape Report, the Nigerian financial sector recorded a loss of ₦17.6 billion (approximately $11.2 million) to fraudulent activities in 2023. This represents a staggering 496% increase compared to losses recorded in 2019, which stood at ₦2.96 billion ($1.8 million). The data further reveals a consistent year-on-year rise in fraud-related losses over the past five years, with a sharp escalation observed between successive years. By 2022, the total loss had reached ₦11.61 billion ($9 million). Moreover, independent investigations based on publicly reported cases indicate that just six high-profile fraud incidents within the same year accounted for an estimated ₦82.4 billion ($52.6 million) in losses, suggesting that the actual scale of financial fraud in the country may be significantly underreported (NIBSS, 2024).

A study by Balogun et al. (2024) investigated perceptions of online investment fraud among Nigerian internet users. The authors surveyed 164 participants revealing that most respondents were young, well-educated individuals whose limited financial experience and attraction to quick returns made them particularly vulnerable to fraudulent schemes. Over half of the sample reported engaging in online investments, with many experiencing victimization through Ponzi/pyramid schemes, cryptocurrency scams, and forex fraud. Figure 1 illustrates data from the Financial Institutions Training Centre (FITC) detailing amounts lost to fraud (2019-2023).

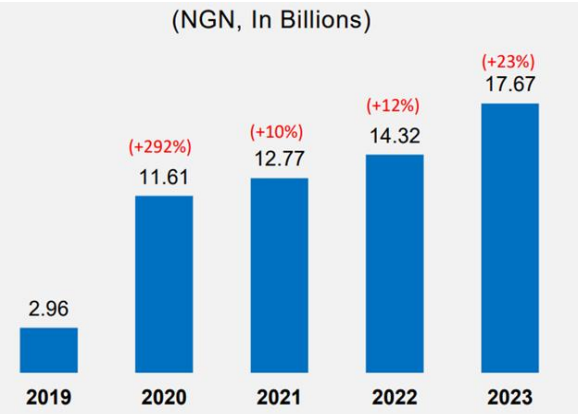

Figure 1: Amount lost to fraud in Nigeria (2019-2023)

Machine Learning Algorithms

Machine learning algorithms apply advanced mathematical techniques, including statistical analysis, linear algebra, and optimization algorithms, to model and detect fraudulent behaviors (Singh, 2023). The fraud detection model explores various machine learning algorithms, trained and tested on financial data specific to Nigeria, to ensure the model's relevance and accuracy. Predictive models such as Logistic Regression and ensemble learning algorithms like Random Forest and XGBoost have built-in features to deal with high-dimensional data and categorical variables. They also have features to handle overfitting problems and minimize the loss function during model training. Random forest is a machine learning ensemble algorithm that combines multiple independent decision trees to provide more precise predictions and decrease bias and variance in the model (Barrett Dr et al., 2020; Kotsiantis & Pintelas, 2004). The algorithm works by constructing a series of decision trees and then combining the predictions of all the trees to make a final prediction. It is easily adaptable to new datasets and fraud types, making it a valuable tool for fraud detection.

Related Works

Olushola and Mart (2024) conducted a comprehensive investigation into machine-learning–based fraud detection, comparing traditional algorithms such as Logistic Regression, Decision Trees, and Support Vector Machines with ensemble methods like Random Forest, Gradient Boosting, and XGBoost. Their analysis showed that ensemble models like random forest generally outperformed traditional classifiers (Ogwueleka, 2011), particularly when dealing with highly imbalanced datasets, while Logistic Regression still performed well in identifying specific types of fraud (Adewumi & Akinyelu, 2017). One of the pioneering works in this area was the application of logistic regression and decision trees to credit card fraud detection. These methods laid the groundwork for more complex algorithms (Olushola & Mart, 2024). Olushola and Mart (2024) cautioned against the "black-box" nature of some models, which can impede interpretability and raise bias concerns. To balance accuracy and transparency, the authors recommend hybrid approaches that combine ensemble performance with explainable models.

Bello and Olufemi (2024) reviewed supervised models (Logistic Regression, Decision Trees, Support Vector Machines), ensemble methods (Random Forest, Gradient Boosting, XGBoost), and unsupervised algorithms (clustering and anomaly detection) for financial fraud detection. A financial institution utilized a random forest algorithm to examine transaction data and detect patterns linked to fraudulent behavior. This method enhanced detection accuracy and lowered the false positive rate, resulting in substantial cost savings. Moreover, the paper highlighted innovative deep learning approaches, such as convolutional neural networks (Okorie et al., 2024) and recurrent neural networks, which effectively analyze sequential transaction patterns in real time. Natural Language Processing (NLP) techniques were also explored to flag suspicious transaction descriptions or communications (Lerma, 2022). Natural Language Processing (NLP) is widely applied in fraud detection and prevention, especially within financial transactions. Financial institutions leverage NLP to examine transaction descriptions, uncovering patterns and specific terms that may indicate fraudulent behavior (Bello & Olufemi, 2024; Fanni et al., 2023).

Consequently, Kumari and Mittal (2024) developed a fraud detection system applied to financial transaction data, leveraging a two-phase process of feature engineering and model training. In the first phase, they extracted key features such as transaction amount, location, time, historical customer behaviors, IP geolocation, and blacklist status to build a robust dataset. In the second phase, various machine learning classifiers were evaluated, including Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, and Gradient Boosting (XGBoost). The only two models with unhinged dataset and have veritably high Recall and F1-Score were MLP regressor and random forest classifier only. Random forest algorithm works by constructing a series of decision trees and then combining the predictions of all the trees to make a final prediction. It is easily adaptable to high dimensional datasets and fraud types, making it a valuable tool for fraud detection (Rokach, 2010). Specifically, this study adopted the Kitchenham approach, which follows structured protocols to extract, integrate, and present findings systematically.

Mohammed et al. (2024) developed a fraud detection framework targeting Nigerian banks, utilizing Logistic Regression and Random Forest to analyze transaction data. They implemented key preprocessing steps; extracting features like transaction type, amount, balance fluctuations, and timestamps before splitting the data into training and testing sets. Their results indicate that the random forest model outperformed logistic regression, delivering higher accuracy and better handling complex data relationships (Ceriani & Verme, 2012; Mohammed et al., 2024; Schonlau & Zou, 2020). To manage skewed fraud-to-non fraud distributions, they employed oversampling techniques such as SMOTE ((Synthetic nonage Oversampling fashion) and advanced preprocessing methods, which enhanced model sensitivity without inflating false positives (Kumari & Mittal, 2024). The study also integrated visualization tools like Matplotlib to present prediction probabilities, improving understanding of model behavior and decision boundaries (Cao et al., 2021).

Similarly, Dichev et al. (2025) proposed a robust fraud detection system tailored for the banking sector, deploying Classification and Regression Trees (CART), Gradient Boosting, and XGBoost, with Logistic Regression used as a baseline. Their framework was applied to a synthetic dataset of nearly 1.5 million banking transactions, sourced from J.P. Morgan. Logistic Regression has been a fundamental tool in the field of fraud detection due to its simplicity and interpretability in binary classification problems. The model predicts the probability of a binary outcome, making it suitable for identifying fraudulent transactions (Kou et al., 2004). However, its linear nature can be a limitation when dealing with complex fraud patterns. Adewumi and Akinyelu (2017) explored Logistic Regression for fraud detection in banking transactions, showing moderate success. However, they also highlighted the model's vulnerability to imbalanced datasets, which often leads to a higher rate of false positives. Despite these challenges, Logistic Regression remains a popular choice due to its ease of implementation and explainability.

Though Random Forest has repeatedly emerged as a strong candidate in past works on fraud detection for its robustness, ability to handle imbalanced datasets, and superior performance over traditional models like Logistic Regression (Bello & Olufemi, 2024; Kumari & Mittal, 2024; Mohammed et al., 2024; Olushola & Mart, 2024); there is still limited understanding of how such ensemble methods can be specifically adapted or optimized for localized fraud patterns, particularly within the Nigerian investment ecosystem.

MATERIALS AND METHODS

This section discusses the methodology of the research, stating and describing the machine learning techniques. The proposed system applies various machine learning models that are trained, tested, and evaluated using real-world financial data to predict fraudulent activities in the Nigerian investment advisory sector.

Research Methods

Dataset Description

The dataset used for this study labelled "Investment Fraud data of Nigerian" contains data on fraud complaints received by the Economic and Financial Crimes Commission (EFCC) and the Securities and Exchange Commission of Nigeria (SEC). The data includes information about the type of fraud, the victim's demographics, the method of solicitation, and the financial loss incurred. This data is used to understand the trends in fraud committed against Nigerians abroad, identify common fraud schemes, and develop strategies to prevent and combat fraud. Table 1 below shows the variables with their data type and description.

Table 1: Variables with their data type and description

Variable	Data Type	Description
Number_ID	int64	Unique identifier for each transaction.
Date_Received	datetime64	Date when the complaint was received
Complaint_Received_Type	object	The source of the complaint
Country	Object	Country of the complaint (all entries are "Nigeria").
Province_State	object	Nigerian state where the complaint originated
Fraud_and_Cybercrime_Thematic_Categories	object	Type of fraud or cybercrime
Solicitation_Method	object	Method used to solicit the victim
Gender	object	Gender of the victim
Language of Correspondence	object	Language used in the complaint
Victim_Age_Range	object	Age range of the victim
Complaint_Type	object	Whether the victim was an "Attempt" or "Victim."
Number_of_Victims	int64	Count of victims involved.
Financial_Loss	float64	Financial loss reported

Handling Class Imbalance

The distribution in Figure 2 below depicts a noticeable imbalance between the classes before resampling. Non-fraudulent transactions significantly outnumber fraudulent ones, which is a common challenge in fraud detection datasets. Such an imbalance can lead to predictive models being biased toward the majority class, often resulting in a higher rate of misclassification for the minority (fraudulent) class. Resampling helps fix this by either increasing the number of minority class samples (oversampling) or reducing the majority class (undersampling).

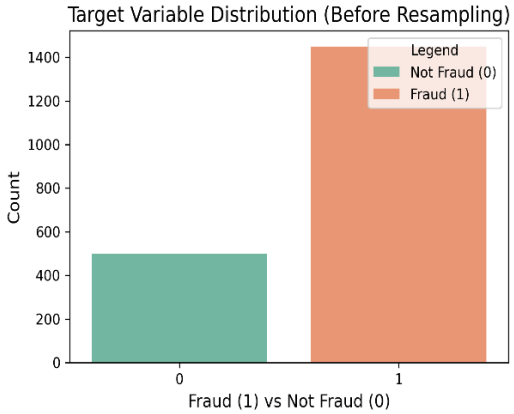

Figure 2: Target variable distribution (Before resampling) Fraud (1) vs Not Fraud (0)

Figure 3 below represents the distribution after resampling showing a balanced dataset where the counts of fraudulent and non-fraudulent transactions are nearly equal. This balance was achieved using advanced resampling methods such as SMOTE combined with Tomek links removal. Balancing the dataset helps enhance model performance by improving sensitivity to the minority class, thereby aiding in more accurate fraud detection.

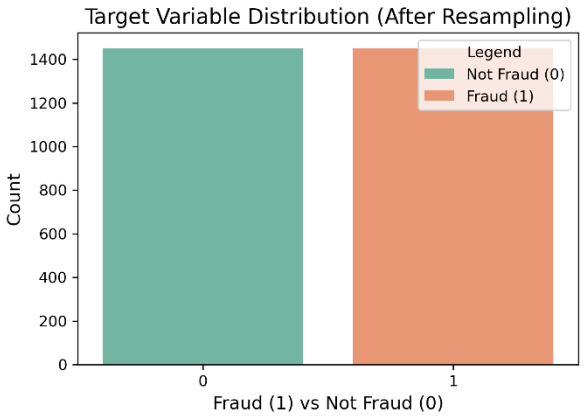

Figure 3: Target variable distribution (After resampling) Fraud (0) vs Not Fraud (1)

Logistic Regression (before and after resampling)
The performance of the Logistic Regression model before applying any resampling techniques demonstrates good predictive accuracy as misclassifications were minimal, with only 13 false positives and 8 false negatives. After resampling, the model demonstrated strong performance, achieving a precision of 0.90 and recall of 0.97 for non-fraud cases. For fraud detection, it attained a precision of 0.99 and recall of 0.96. The F1-scores were 0.93 and 0.98 for non-fraud and fraud respectively, with overall averages of 0.97 (precision), 0.96 (recall), and 0.97 (F1-score), indicating high accuracy and balanced predictive capability following resampling.

Random Forest (before and after resampling)
The Random Forest model's performance before resampling was evaluated to determine its effectiveness in detecting fraudulent transactions within the dataset. By analyzing the confusion matrix, classification report, and AUC-ROC curve, the goal was to assess the model's precision, recall, and overall accuracy. The results suggest a minimal rate of misclassification. After resampling, the model showed strong performance across both classes. For non-fraud (class 0), it achieved a precision of 0.93, recall of 0.94, and F1-score of 0.94. For fraud (class 1), precision reached 0.98, recall 0.97, and F1-score 0.98. The overall average for precision, recall, and F1-score was 0.97, indicating high reliability in detecting both fraudulent and legitimate transactions.

XGBoost (before and after resampling)
The XGBoost model's performance before applying any resampling techniques has been evaluated using a confusion matrix, an AUC-ROC curve, and classification metrics. The model's ability to distinguish between fraudulent and non-fraudulent cases is crucial, given the class imbalance typically present in fraud detection datasets. There were a few misclassifications and the slight imbalance in recall for the non-fraud class suggests a need for resampling techniques or further hyperparameter tuning to enhance performance across both classes. After resampling, The XGBoost model achieved a precision of 0.91, recall of 0.94, and F1-score of 0.93 for the non-fraud class (0), indicating strong accuracy and low false positives. For the fraud class (1), it recorded a precision of 0.98, recall of 0.97, and F1-score of 0.97, highlighting its effectiveness in detecting fraudulent cases. The weighted averages of 0.96 across all metrics reflect the model's overall balanced and robust performance on the resampled dataset.

RESULTS AND DISCUSSION

The performance of three classification models Logistic Regression, Random Forest, and XGBoost was evaluated on a fraud detection task using key performance metrics such as Accuracy, Precision, Recall, F1-Score, and AUC-ROC. The models were tested under two scenarios: before resampling and after resampling using SMOTE combined with Tomek Links. Each model exhibited distinct performance patterns, reflecting how they handle class imbalance and the effect of resampling techniques.

Table 2: Summary of results of each classifier

Classifier	Resampling Technique	Accuracy (Fraud)	Precision (Fraud)	Recall (Fraud)	F1-Score (Fraud)	AUC-ROC
Logistic Regression	No Resampling	0.964587	0.970721	0.981777	0.976217	0.993728
	SMOTE and Tomek Links removal	0.964587	0.990610	0.961276	0.975723	0.993403
Random Forest	No Resampling	0.967960	0.977273	0.979499	0.978385	0.987656
	SMOTE and Tomek Links removal	0.966273	0.979405	0.974943	0.977169	0.983337
XGBoost	No Resampling	0.962901	0.968539	0.981777	0.975113	0.993832
	SMOTE and Tomek Links removal	0.961214	0.979263	0.968109	0.973654	0.993063

Discussion
The resampling technique appears to have a positive impact, particularly on the XGBoost and Random Forest models, with the former consistently delivering the best accuracy. The Logistic Regression model, while stable, did not benefit significantly from resampling, indicating that ensemble methods like XGBoost and Random Forest are more effective in leveraging the balanced dataset.
Logistic Regression achieved 96.46% accuracy, with strong precision (97.07%), recall (98.18%), and an AUC-ROC of 0.9937. After applying SMOTE and Tomek Links, precision improved to 99.06% while recall slightly dropped to 96.13%, showing a shift toward reducing false positives at the cost of missing a few fraud cases. Random Forest slightly outperformed Logistic Regression pre-resampling with 96.80% accuracy, 97.73% precision, and 97.95% recall. Post-resampling, it remained stable, with marginal gains in precision (97.94%) and minor drops in recall (97.49%), reflecting its resilience to class imbalance. XGBoost had the lowest pre-resampling accuracy (96.29%) but the highest AUC-ROC (0.9938), with strong precision (96.85%) and recall (98.18%). Resampling raised its precision to 97.93% but reduced recall to 96.81%, indicating greater caution at the expense of missing some frauds. Overall, all models performed well, with Logistic Regression benefiting most in precision post-resampling, Random Forest showing consistency, and XGBoost excelling in class separation.

Study Limitations
While this study contributes meaningful insights into fraud detection using machine learning, several limitations remain. The evaluation was limited to three classification algorithms, which, although effective and widely applied, may not fully capture the potential of more advanced methods such as deep learning or ensemble techniques. Furthermore, despite the application of resampling strategies, the dataset may not adequately represent the dynamic and complex nature of real-world fraud cases, which could affect model generalizability. Future research should therefore consider larger, more diverse, and up-to-date datasets, alongside a broader set of algorithms, to enhance robustness. Continuous monitoring and retraining of models are also crucial to sustain predictive accuracy and mitigate overfitting, with approaches such as 10-fold cross-validation offering performance stability. Since fraud patterns evolve over time, frequent model updates are imperative to ensure adaptability to emerging threats.

CONCLUSION

The main goal of this paper was to evaluate a more balanced and accurate fraud prediction framework using machine learning, with a focus on enhancing real-time fraud identification and reducing financial risks. The proposed system has shown strong potential and may be scaled further with more advanced algorithms and up-to-date datasets. In practice, such systems can be integrated into Nigerian financial institutions to support better decision-making, reduce losses, and improve trust in automated fraud detection processes.

By applying the hybrid resampling approach, specifically the combination of SMOTE and Tomek Links in addressing class imbalance within fraud detection datasets the predictive accuracy and reliability of the models were significantly enhanced. Among the evaluated algorithms, the Random Forest model emerged as the most robust, showing superior performance across key metrics both before and after resampling. Its resilience to overfitting and ability to manage complex datasets make it a strong candidate for deployment in Nigerian-based fraud detection systems.

REFERENCES

Adewumi, A. O., & Akinyelu, A. A. (2017). A survey of machine-learning and nature-inspired based credit card fraud detection techniques. *International Journal of System Assurance Engineering and Management*, *8*(Suppl 2), 937-953.

AGYA, A. A., OJIYA, E. A., & SAMUEL, P. A. (2018). The effect of remittances on economic growth in Nigeria: An autoregressive distributed lag approach. *Journal of Economics, Management & Social Science, April Special Edition*, 139-152.

Akinninyi, P. E., Akpan, D. C., & Umoren, A. O. (2025). Forensic accounting and financial integrity in the Nigerian public sector. *Journal of Accounting and Financial Management*, *11*(3), 122-145.

Alonge, E. O., Eyo-Udo, N. L., Ubanadu, B. C., Daraojimba, A. I., Balogun, E. D., & Ogunsola, K. O. (2021). Enhancing data security with machine learning: A study on fraud detection algorithms. *Journal of Data Security and Fraud Prevention*, *7*(2), 105-118.

Ashtiani, M. N., & Raahemi, B. (2021). Intelligent fraud detection in financial statements using machine learning and data mining: a systematic literature review. *Ieee Access*, *10*, 72504-72525.

Balogun, O. S., Akangbe, T. A., Fagbamila, O. D., & Aigbovbioisa, F. O. (2024). INTERNET USERS'PERCEPTION OF THE PREVALENCE OF ONLINE INVESTMENT FRAUD AND VICTIMISATION IN NIGERIA.

Barrett Dr, S., Gray Dr, G., & McGuinness Dr, C. (2020). Comparing Variable Importance in Prediction of Silence Behaviours between Random Forest and Conditional Inference Forest Models.

Bello, O. A., & Olufemi, K. (2024). Artificial intelligence in fraud prevention: Exploring techniques and applications challenges and opportunities. *Computer science & IT research journal*, *5*(6), 1505-1520.

Cao, S., Zeng, Y., Yang, S., & Cao, S. (2021). Research on Python data visualization technology. Journal of physics: Conference series,

Ceriani, L., & Verme, P. (2012). The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. *The Journal of Economic Inequality*, *10*(3), 421-443.

Čunderlík, Ľ. (2021). Fraudulent schemes in the financial market (financial pyramids)–detection and prevention. *Financial Law Review*, *21*(1), 16-30.

Dichev, A., Zarkova, S., & Angelov, P. (2025). Machine learning as a tool for assessment and management of fraud risk in banking transactions. *Journal of Risk and Financial Management*, *18*(3), 130.

Fanni, S. C., Febi, M., Aghakhanyan, G., & Neri, E. (2023). Natural language processing. In *Introduction to artificial intelligence* (pp. 87-99). Springer.

Gui, Z., Huang, Y., & Zhao, X. (2024). Financial fraud and investor awareness. *Journal of Economic Behavior & Organization*, *219*, 104-123.

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: a review of anomaly detection techniques and recent advances. *Expert systems With applications*, *193*, 116429.

Kotsiantis, S., & Pintelas, P. (2004). Combining bagging and boosting. *International Journal of Computational Intelligence*, *1*(4), 324-333.

Kou, Y., Lu, C.-T., Sirwongwattana, S., & Huang, Y.-P. (2004). Survey of fraud detection techniques. IEEE international conference on networking, sensing and control, 2004,

Kumari, P., & Mittal, S. (2024). Fraud Detection System for Financial System Using Machine Learning Techniques: A Review. 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO),

Lerma, L. (2022). Comparative analysis of natural language processing and gradient boosting trees approaches for fraud detection.

Lokanan, M. E., & Sharma, K. (2022). Fraud prediction using machine learning: The case of investment advisors in Canada. *Machine Learning with Applications*, *8*(100269), 1-12.

Martins, O., & Fonkem, B. (2024). Leveraging big data analytics to combat emerging financial fraud schemes in the USA: a literature review and practical implications. *World J Adv Res Reviews*, *24*, 17-43.

Mohamed, N. (2025). Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms. *Knowledge and Information Systems*, 1-87.

Mohammed, U., Wajiga, G., & SAIDU, H. A. (2024). Financial Fraud Detection in Nigerian Banks: Data Mining Approach.

NIBSS. (2024). Data and Insights Archives. *https://nibss-plc.com.ng/category/insights/*.

Odufisan, O. I., Abhulimen, O. V., & Ogunti, E. O. (2025). Harnessing Artificial Intelligence and Machine Learning for Fraud Detection and Prevention in Nigeria. *Journal of Economic Criminology*, 100127.

Ogwueleka, F. N. (2011). Data mining application in credit card fraud detection system. *Journal of Engineering Science and Technology*, *6*(3), 311-322.

Okorie, D. D., Fatokun, J. O., & Umoren, U. M. (2024). Applying Convolutional Neural Networks for Enhanced Digital Image Steganalysis. In *Practical Statistical Learning and Data Science Methods: Case Studies from LISA 2020 Global Network, USA* (pp. 697-718). Springer.

Olushola, A., & Mart, J. (2024). Fraud detection using machine learning. *ScienceOpen Preprints*.

Rokach, L. (2010). Ensemble-based classifiers. *Artificial intelligence review*, *33*(1), 1-39.

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, *20*(1), 3-29.

Singh, K. (2023). The Role of Mathematics in Artificial Intelligence and Machine Learning. *Int. J. Res. Publ. Semin*, *14*, 186-197.