ARTIFICIAL INTELLIGENCE AND TAX FRAUD DETECTION: EMPIRICAL EVIDENCE FROM EDO STATE INTERNAL REVENUE SERVICE, NIGERIA

Dr. Tina Ashafoke
Department of Accounting
Faculty of Management Sciences
University of Benin
Benin City, Edo State, Nigeria.
oghenekome.ashafoke@uniben.edu

Elugwu George Ugochukwu
Department of Accounting
Faculty of Management Sciences
University of Benin
Benin City, Edo State, Nigeria.
george.elugwu@mgtsci.uniben.edu

Obianke Osaretin Joy
Department of Accounting
Faculty of Management Sciences
University of Benin
Benin City, Edo State, Nigeria.
joyosaretin3@gmail.com

Corresponding Author: oghenekome.ashafoke@uniben.edu
DOI: https://doi.org/10.57233/ijamer.v2i3.010

Article History: Received: 15 August 2026 | **Revised:** 28 August 2026 | **Accepted:** 3 September 2026 | **Published:** 25 September 2026

How to Cite This Article:
Ashafoke, T., Elugwu, G. U., & Obianke, O. J. (2026). Artificial intelligence and tax fraud detection: Empirical evidence from Edo State Internal Revenue Service, Nigeria. *International Journal of Accounting, Management and Economic Review, 2*(3). https://doi.org/10.57233/ijamer.v2i3.010

Abstract
Tax fraud remains a major challenge to effective tax administration in Nigeria, as fraudulent reporting, concealment of taxable income and other forms of non-compliance can reduce government revenue and undermine the effectiveness of the tax system. Existing tax fraud detection mechanisms often rely on conventional audit procedures and manual examination of taxpayer records, which may be limited in their ability to process large volumes of data, identify complex fraud patterns and detect emerging forms of tax evasion. Against this background, this study examined the relationship between Artificial Intelligence (AI) and tax fraud detection in the Edo State Internal Revenue Service (EIRS), Nigeria. Specifically, the study examined the association between machine learning, predictive analytics, data mining and tax fraud detection. The study adopted a quantitative survey research design and obtained primary data from selected tax officers in EIRS. The data collected were analysed using multiple regression analysis. The findings revealed that machine learning had a positive and significant association with tax fraud detection (B = 0.276, t = 3.958, p = 0.000). Predictive analytics also had a positive and significant association with tax fraud detection (B = 0.259, t = 3.700, p = 0.000), while data mining had a positive and significant association with tax fraud detection (B = 0.312, t = 4.216, p = 0.000). The regression model produced an R of 0.842 and an R^2 of 0.709, indicating that 70.9% of the variation in tax fraud detection was jointly explained by machine learning, predictive analytics and data mining. The findings suggest that AI tools can enhance the efficiency of detecting abnormalities, strengthen audit effectiveness, and improve compliance monitoring and risk assessment. The study therefore recommends increased investment in AI-enabled tax administration technologies, continuous

training of tax officers in the application of AI tools, and greater integration of machine learning, predictive analytics and data mining into tax fraud detection processes to improve tax compliance and revenue administration in Edo State.

Keywords: Artificial Intelligence; Machine Learning; Predictive Analytics; Data Mining; Tax Fraud Detection; Tax Administration; Nigeria.

1.0 Introduction

The world of public administration is constantly changing due to the quick advance of technology, and tax administration is one of the fields that is greatly affected by the digital transformation. Governments worldwide depend on taxes for essential services such as health, education, security, infrastructure and so forth. With the government increasingly adopting new technology in public administration to improve the public sector's effectiveness, tax administration becomes one of the core areas for the digital innovation (OECD, 2025a; Mergel et al., 2019). Tax fraud is a serious problem that could be represented as intentionally misstating the revenue, concealing of revenues via commercial company or making misrepresentation of fiscal items in order to escape from the taxation. The amount of revenue lost due to tax evasion globally is as high as USD 492 billion each year (Tax Justice Network, 2024). In Nigeria, tax fraud remains a concern for revenue administration, particularly where conventional detection mechanisms rely heavily on manual examination of taxpayer records, routine audits and retrospective investigations. These approaches may be inadequate for processing large volumes of taxpayer information and identifying concealed income, unusual transactions and complex patterns of fraudulent activities promptly. At the sub-national level, these limitations are relevant to the Edo State Internal Revenue Service (EIRS), where difficulties in efficiently identifying tax irregularities and high-risk taxpayers may constrain effective tax fraud detection, delay compliance enforcement and contribute to potential revenue leakages. The limitations of existing detection mechanisms therefore highlight the need for more advanced, data-driven approaches capable of analysing complex taxpayer information and improving the identification of potential tax fraud.Most revenue losses arise from the multinational corporations and the private entities engaging aggressive tax avoidance and evasion; however low and middle income countries disproportionately feel its impacts due to its heavier reliance on taxation in funding public expenditure and lack of fiscal capacity. Annual losses of revenue will stand at approximately USD 4.92 trillion in the coming decade at present trends and the inequality will likely widen, investment in public services constrained, economic development slowed (Tax Justice Network, 2024). Traditional auditing, inspection, and tax compliance tools based on rules have always helped tax authorities to enforce tax, yet those were costly, time consuming and less efficient in analyzing large amount of complicated financial information in digital economy and detecting complex fraudulent cases (OECD, 2025a; IMF, 2024).

The aforementioned traditional methods are unreliable to detect sophisticated tax fraud cases hidden within a massive volume of data resulting in delays in compliance enforcement and loss of revenue. Therefore, technological tools could be utilized by the tax administration for automating the data-handling processes within these areas, and analyze huge amount of transaction data to uncover previously unknown patterns and identify fraud candidates faster and more efficient (OECD, 2025). Artificial intelligence is defined as 'the extent to which a computer system can perform the task requiring human intelligence, human language ability: learning from data, patterns, prediction, support of complex decision making (Russell and Norvig, 2021). Artificial intelligence has been extensively applied to various sectors such as tax by machine learning, predictive analytics, data mining, natural language processing. The applications help

tax departments to reach a higher level of tax compliance enforcement and fraud detection capabilities (Edupuganti, 2024). Based on empirical results, AI can efficiently improve tax fraud detection by enabling the tax authorities to uncover complex tax fraud cases, assess taxpayer risk, and automate tax compliance monitoring. For instance, de Souza et al. (2024) found that the machine learning algorithms can detect fraudulent tax transactions in large and complicated datasets more effectively than traditional audit tools. Similar to this, Verma and Kumar (2024) demonstrate that predictive analytics enhanced the selection of high-risk taxpayers, decreased the mis-selection of auditors, and led to positive effects in tax compliance. Moreover, more than 75% of OECD member countries adopted the AI tools for tax fraud detection, risk assessment and compliance monitoring which implies the application of AI in tax administration has been widely explored and embraced by tax authorities (OECD, 2025b). AI can analyze both structured and unstructured data such as social media, invoices and financial statements to reveal behaviors potentially indicative of tax non-compliance or fraud (OECD, 2025). AI solutions have also shown their power to largely increase the detection rate: it has been demonstrated that AI-assisted tax compliance increases the detection rate and case evasion detection is improved from 14.7% in 2021 to 55% in 2024 (Gidisu et al., 2025), it has also shortened average audit time from 30 days to 15 days and increased the prediction precision of detecting fraudulent filing from 85% to 95%. In this paper, tax fraud detection is defined as the dependent variable in the assessment of effectiveness of fraud detection, prevention and management. The results can be measured by percentage of detected fraud cases, amount of reduced audit time and recovered revenue from fraud, accuracy in prediction model and false positive and false negative rate. The value of AI has added more by real time detection of unusual patterns and prediction of risk that can hardly be done by usual systems (Gidisu et al., 2025). However, there are drawbacks related to the use of all the advantages in the tax administration systems such as data privacy, insufficient digital infrastructure, costliness, algorithmic bias, lack of sufficient trained human resources (OECD, 2025), and few countries lack appropriate ethical/governance frameworks for implementation of AI in tax system (UNESCO, 2021). The ability of AI to transform tax administration and make tax fraud detection and management smarter will result in increase of revenue and tax compliance; nevertheless its effective use is dependent on quality data, digital infrastructure, trained personnel, organizational readiness, and governance framework which ensure integrity, transparency, accountability, and ethical behavior in the AI system (OECD, 2025; UNESCO, 2021). Here, we review the application of some AI technologies, machine learning, predictive analytics and data mining, in the area of tax fraud detection.

Even with the development of anti-corruption efforts, Tax Fraud has continued to be a prevailing and harmful aspect of economic activities across nations and is becoming one of the most serious threat to the revenues of countries and for government development of any countries. Developing nations are still susceptible and at high risk to the phenomenon. The costs of fraud and evasion amounts to over 20 percent of the potential annual tax revenue of developing countries (IMF, 2024). The traditional tax fraud methods (manual audit, rule-based tests, random inspections) fail to handle the enormous amount of financial data, leading to higher opportunity for more sophisticated fraud and delays in enforcing law (OECD, 2025).

Many researches have investigated the possibilities of using machine learning and data mining techniques to identify and reduce tax fraud (e.g.

Edupuganti, 2024; Russell & Norvig, 2021; Gidisu, Mensah & Opoku, 2025; Kure-Bisa & Abey-Jonas, 2024). However, most of these studies were from developed countries' platforms

(especially financial organizations and corporate fraud data), which make them less suitable to develop country environments due to technological and institutional constraints (Gidisu, Mensah & Opoku, 2025). Most studies were theoretical with limited empirical investigation on the actual impacts of AI on tax compliance outcomes in the public sector, and in developing countries particularly. These gaps require more researches to verify the feasibility and effectiveness of AI to detection tax fraud, particularly in developing contexts with technological, financial and operational constraints.

The objective of the present research was therefore to enhance knowledge of the latter so as to confirm its impact on tax fraud effectiveness and feasibility in Nigeria with the view to providing a set of insights on the application of AI, its' advantages, its' disadvantages, and improvement opportunities to reach better tax collection.

The aim of this study was to answer the following questions:

i. How does machine learning enhance the identification of tax fraud in Edo State Internal Revenue Service?
ii. How does predictive analytics impact the identification of fraudulent tax activities at Edo State Internal Revenue Service?
iii. How does the use of data mining enhance the efficiency of tax fraud detection in Edo State Internal Revenue Service?

2. 0 Literature Review

Tax Fraud Detection (TFD) Tax Fraud Detection (TFD) refers to the processes and techniques used by tax authorities to identify, investigate and respond to activities through which individuals or organisations deliberately evade or reduce their tax obligations. Tax fraud may involve underreporting or concealing income and assets, overstating deductible expenses, falsifying accounting records, concealing taxable operations and providing inaccurate information in tax returns. Effective tax fraud detection is therefore an important component of tax administration because it enables tax authorities to identify potential non-compliance, protect government revenue and promote compliance with tax obligations. Beyond identifying fraudulent activities, effective TFD provides information that enables tax authorities to assess risk, prioritise audit cases and allocate limited enforcement resources to areas where the likelihood of fraud is higher.

The effectiveness of tax fraud detection depends largely on the ability of tax authorities to identify irregularities accurately and at an early stage. Traditionally, tax authorities have relied on manual examination of records, financial statement analysis, routine audits, rule-based systems and scoring techniques to identify suspicious tax activities. Although these approaches can detect certain forms of non-compliance, their effectiveness becomes constrained when tax authorities are required to process large volumes of structured and unstructured information. Manual procedures are often time-consuming, resource-intensive and costly, while rule-based approaches may be less capable of identifying fraud patterns that do not conform to predetermined indicators. The increasing sophistication of tax fraud through electronic transactions, cross-border financial flows and digitally enabled business activities further exposes these limitations.

The growing complexity and volume of taxpayer information therefore create a need for detection mechanisms that can process data more rapidly and identify relationships and patterns

that may not be readily observable through conventional approaches. The Organisation for Economic Co-operation and Development (OECD, 2025) observes that the increasing complexity, mobility and digitisation of transactions have made the identification of sophisticated and evolving fraud schemes more challenging for tax administrations. This limitation provides the conceptual basis for applying Artificial Intelligence (AI) to tax fraud detection. Unlike conventional approaches that largely depend on predefined rules or manual review, AI-based systems can analyse large datasets, recognise patterns, identify anomalies and generate risk indicators from taxpayer information. Consequently, the value of AI for TFD lies not simply in automating existing procedures but in improving the capacity of tax authorities to detect relationships and behavioural patterns associated with potential fraud.

The relationship between AI and TFD can be further understood through the specific technologies examined in this study: machine learning, predictive analytics and data mining. Machine learning contributes to TFD by enabling systems to learn from existing taxpayer and transaction data and identify patterns associated with fraudulent or abnormal activities. Through classification and anomaly-detection techniques, machine learning can assist tax authorities in distinguishing potentially suspicious transactions or taxpayers from normal compliance behaviour. This can improve the identification of unusual activities that may not be detected through conventional rule-based procedures. Predictive analytics, in contrast, focuses on using historical and current data to estimate the likelihood of future tax non-compliance or fraud. Its contribution to TFD is therefore particularly relevant to risk assessment and audit prioritisation, as tax authorities can use predictive models to identify taxpayers or transactions that require closer examination. Data mining complements these technologies by extracting previously unknown patterns, relationships and anomalies from large and complex datasets. Its relevance to TFD arises from its ability to uncover hidden relationships among taxpayer records and transactions that may provide indicators of fraudulent behaviour.

The three technologies consequently address different but interconnected aspects of tax fraud detection. Machine learning enhances the ability to recognise and classify suspicious patterns; predictive analytics supports the assessment and prioritisation of potential fraud risks; and data mining facilitates the discovery of hidden patterns and relationships within large datasets. Together, these capabilities can shift tax fraud detection from a predominantly reactive process towards a more proactive and data-driven approach. AI-driven tools can support continuous transaction monitoring, automated anomaly detection, risk profiling and audit case selection, thereby improving the speed and efficiency of tax fraud detection (Edupuganti, 2024). AI may also assist in identifying complex schemes, including fictitious invoices, shell companies and other forms of concealed or sophisticated tax fraud that may be difficult to detect through standard procedures (Nguyen et al., 2025). Empirical evidence further suggests that AI applications in tax administration can improve fraud detection accuracy, reduce audit processing time and strengthen tax compliance and revenue collection (Gidisu, Mensah, & Opoku, 2025).

However, the relationship between AI and effective tax fraud detection is not automatic. The ability of machine learning, predictive analytics and data mining to generate reliable detection outcomes depends on the quality and availability of taxpayer data, digital infrastructure, technical expertise and institutional capacity. Poor-quality or incomplete data may produce inaccurate predictions and misleading fraud indicators, while inadequate infrastructure and limited technical skills may restrict the effective implementation of AI systems. In addition, tax

administrations must address concerns relating to data privacy, cybersecurity, algorithmic bias, transparency and ethical use of taxpayer information (Kapoor & Singh, 2025; OECD, 2025). Thus, although AI provides important opportunities for strengthening TFD, its effectiveness depends on the availability of reliable data, appropriate technological infrastructure, skilled personnel and adequate governance frameworks.

Conceptually, therefore, TFD represents the outcome of a detection process, while machine learning, predictive analytics and data mining represent technological mechanisms through which the identification of potential tax fraud can be strengthened. The expected relationship is that greater application of these AI technologies improves the capacity of tax authorities to identify anomalies, assess fraud risks and uncover hidden patterns, thereby enhancing tax fraud detection. This conceptual relationship provides the basis for examining the individual contributions of machine learning, predictive analytics and data mining to tax fraud detection in the Edo State Internal Revenue Service.

Artificial Intelligence (AI)

Artificial Intelligence (AI) refers to the capability of computer systems to perform tasks that ordinarily require human intelligence, including learning, reasoning, problem-solving and pattern recognition (Russell & Norvig, 2021). In public administration, AI supports decision-making by enabling the analysis of large volumes of data and the generation of insights for predicting outcomes. In tax administration, its relevance lies in its ability to process complex taxpayer and transaction data, identify unusual patterns and support risk-based decision-making. Thus, AI provides tax authorities with capabilities that can strengthen tax fraud detection beyond conventional manual and rule-based approaches.

AI applications in tax administration include data analytics, risk profiling, anomaly detection and fraud prediction. These applications can help tax authorities manage large volumes of taxpayer records, identify suspicious taxpayer behaviour and improve compliance monitoring (Rodríguez Valencia et al., 2025). Specifically, machine learning and deep learning can be used to analyse tax documents and transaction data to identify patterns associated with fraudulent activities (OECD, 2023). In this regard, machine learning, predictive analytics and data mining provide complementary mechanisms through which AI can contribute to tax fraud detection by identifying patterns, predicting risks and uncovering hidden relationships in taxpayer data.

Despite these benefits, the effectiveness of AI-based tax fraud detection depends on the availability of reliable data, adequate digital infrastructure and skilled personnel. Issues relating to data privacy, algorithmic bias, cybersecurity, transparency and governance may also affect the responsible use of AI in tax administration (OECD, 2025). Therefore, while AI has the potential to improve the efficiency and accuracy of tax fraud detection, its effectiveness depends on the technological and institutional capacity of tax authorities to implement and govern AI systems appropriately.

Machine Learning (ML)

Machine Learning (ML) is a field within Artificial Intelligence that enables computer systems to learn from experience, identify patterns from data and make predictions or decisions without being explicitly programmed for every task (Jordan & Mitchell, 2023). Its relevance to tax fraud detection lies in its ability to analyse large and complex taxpayer data and identify patterns associated with potentially fraudulent behaviour. Unlike traditional rule-based methods, ML can

learn from changing data and recognise emerging fraud patterns, making it useful for detecting sophisticated and evolving forms of tax fraud.

The contribution of ML to tax fraud detection occurs through anomaly detection, fraud risk scoring and the identification of suspicious taxpayer behaviour based on income, tax filings and transaction records. ML models may be trained on previous fraud cases through supervised learning to identify characteristics associated with fraud, while unsupervised approaches can identify unusual behaviour even where previously labelled fraud cases are unavailable. This enables ML to support the identification of both known and emerging fraud patterns and improve the targeting of high-risk cases for further investigation (Nguyen et al., 2025).

However, the effectiveness of ML depends on the quality of the data and the design of the models used. Issues such as biased data, overfitting and limited transparency may produce inaccurate outcomes or undermine fairness and accountability in tax administration (OECD, 2025a). Therefore, ML can strengthen tax fraud detection through improved pattern recognition and risk identification, provided appropriate data quality and governance mechanisms are in place.

Predictive Analytics (PA)

Predictive Analytics (PA) uses historical and real-time data, statistical techniques and machine learning algorithms to estimate future behaviours and events. In tax administration, its relevance to tax fraud detection lies in its ability to assess the likelihood of taxpayer non-compliance and support risk-based compliance monitoring. By analysing tax returns, payment records and other financial information, predictive models can generate risk scores that help tax authorities identify taxpayers or transactions with a higher probability of non-compliance (Shmueli & Koppius, 2024).

The analytical value of PA is therefore its ability to shift tax fraud detection from a largely reactive process towards proactive risk identification. Techniques such as logistic regression, gradient boosting and ensemble models can support the selection and prioritisation of tax cases for audit, potentially improving detection efficiency and the allocation of audit resources (Lee et al., 2025). However, the effectiveness of PA depends on the availability of clean and reliable data, explainable models and strong data governance. These limitations may be particularly important in developing tax administrations where data quality and technological capacity remain constrained (Attah & Eze, 2025).

Data Mining (DM)

Data Mining (DM) refers to the process of discovering patterns, trends and relationships within large datasets using statistical and machine learning techniques (Han, Pei, & Kamber, 2023). In tax administration, its relevance to tax fraud detection lies in its ability to analyse structured and unstructured information, including tax returns, financial statements, invoices and supporting documents. Techniques such as association rules, clustering and anomaly detection can reveal unusual transactions and taxpayer behaviours that may not be identified through conventional audits (Rana & Sharma, 2025).

The relationship between DM and tax fraud detection is therefore based on its capacity to uncover hidden relationships and patterns that provide indicators of potential fraud. When integrated with machine learning models, DM can support fraud identification, audit targeting

and efficient allocation of tax administration resources, thereby improving the effectiveness of tax collection (Nguyen et al., 2025). However, its effectiveness depends on accurate and reliable data, interpretable analytical models and appropriate privacy and ethical safeguards (OECD, 2025).

Machine Learning Techniques and Tax Fraud Detection

Li, Chen, and Zhang (2019) investigated machine learning approaches to tax evasion detection among 2,350 Chinese SMEs between 2014 and 2018, using decision trees, Support Vector Machine (SVM) and random forest. They found that random forest produced the highest prediction accuracy and recommended ensemble ML models for tax audits. The use of multiple algorithms allowed direct comparison of model performance; however, the study focused on Chinese SMEs, limiting its applicability to other tax environments. Compared with Li, Chen, and Zhang (2019), Kou, Gan, and Xu (2020) used a substantially larger dataset of over 1.2 million EU VAT transactions and applied neural networks and logistic regression. Their finding that neural networks improved fraud detection and reduced false positives supports the earlier evidence that ML improves detection, although differences in dataset size, tax systems and model types make direct comparison of performance difficult.

Adebayo and Oladipupo (2021) examined ML in Nigeria's federal tax administration using gradient boosting and data from only 150 corporate taxpayers. Although the sample was considerably smaller than those used by Li, Chen, and Zhang (2019) and Kou, Gan, and Xu (2020), the Nigerian setting provides greater contextual relevance. The study found improved detection of underreported income and fraudulent expense claims, but identified data integrity and staff capacity as limitations. This contrasts with the larger international studies, where greater data availability and technological capacity may have supported stronger model performance.

Bnrescu and Dinu (2022) investigated 420 Romanian firms using deep learning and anomaly detection and found that ML identified complex fraud patterns that manual audits could miss. Methodologically, the use of deep learning and anomaly detection allowed the study to focus on less obvious fraudulent patterns, unlike Adebayo and Oladipupo (2021), which concentrated on specific forms of underreporting and fraudulent claims. However, the Romanian technological environment differs from Nigeria's, making the findings less directly transferable to Nigerian tax administration. Nguyen, Le, and Tran (2023) similarly used K-Nearest Neighbour (KNN) and SVM among 680 Southeast Asian manufacturing firms and reported improved fraud detection and faster audits. While this supports the findings of the earlier studies, the focus on manufacturing firms limits generalisation across taxpayers and tax sectors.

Okafor and Eze (2024) analysed data from more than 3,000 taxpayers across Nigeria, Kenya and South Africa and found that ML improved tax fraud detection and voluntary compliance. The larger and multi-country sample provides broader evidence than Adebayo and Oladipupo (2021), while the inclusion of Nigeria improves the relevance of the findings to the present study. However, the multi-country design also introduces differences in tax laws, digital infrastructure and institutional capacity that may influence the results. Taken together with the earlier studies, the evidence supports ML as a useful fraud detection technique, but differences in samples,

algorithms, data environments and research settings leave limited evidence on its application within sub-national tax administration in Nigeria, particularly the Edo State Internal Revenue Service (EIRS), and its specific contribution when examined alongside predictive analytics and data mining.

Predictive Analytics and Tax Fraud Detection

Lee, Park, and Kim (2025) investigated predictive analytics within South Korea's tax system between 2019 and 2024 using data from 1,850 companies and applying neural networks and decision trees. Unlike Almeida and Santos (2025), who combined logistic regression with gradient boosting, this study incorporated a neural network capable of modelling more complex relationships alongside the relatively interpretable decision tree technique. The study reported a 28% increase in fraud detection rates and shorter audit durations, which is consistent with Almeida and Santos (2025) in demonstrating that predictive analytics can improve both risk identification and audit efficiency. However, the smaller sample and concentration on companies within one highly digitalised tax system limit the extent to which the findings can be transferred to a less technologically developed sub-national tax administration.

Rodríguez Valencia et al. (2025) examined predictive analytics across Latin American tax administrations using data from more than 5,000 taxpayers. Unlike Almeida and Santos (2025) and Lee, Park, and Kim (2025), who focused on individual national tax systems, the study adopted a broader regional approach and concentrated on audit targeting and revenue recovery. The findings of improved audit targeting and increased revenue recovery support the earlier evidence that predictive analytics can strengthen tax enforcement. However, the study's emphasis on data quality and inter-agency cooperation reveals an important limitation: the effectiveness of predictive analytics is influenced not only by the predictive model but also by the availability, quality and integration of tax data. This limitation is particularly relevant to developing tax administrations where data-sharing systems may be less established.

Attah and Eze (2025) examined predictive analytics and tax fraud detection in Nigeria and Ghana between 2018 and 2023 using survey responses from 420 tax officials and regression analysis. Unlike the previous studies that relied primarily on administrative or transactional datasets and computational prediction models, this study measured the relationship through the perceptions of tax officials. The finding of a statistically significant positive relationship between predictive analytics and tax fraud detection is consistent with the findings of Almeida and Santos (2025), Lee, Park, and Kim (2025), and Rodríguez Valencia et al. (2025). However, the use of perceptual survey data represents a methodological limitation because reported improvements in fraud detection may not necessarily reflect actual predictive accuracy or changes in detected fraud cases. The infrastructure challenges identified by the study further suggest that technological capacity may condition the effectiveness of predictive analytics in developing tax administrations.

Almeida and Santos (2024) examined AI-powered tools for tax compliance in Brazil using administrative data from 30,000 taxpayers between 2017 and 2022. The large administrative dataset provides stronger empirical evidence of actual tax administration outcomes than studies based solely on officials' perceptions. The study found improved audit accuracy and fewer false positives, complementing the findings of Almeida and Santos (2025) and Lee, Park, and Kim (2025) on improved audit effectiveness. However, because the study examined AI-powered tax compliance tools generally rather than predictive analytics alone, it is difficult to attribute the reported improvements specifically to predictive analytics. This limits its direct relevance to studies examining predictive analytics as an independent technology.

Zhao and Cheng (2024) examined predictive analytics models among 2,600 Chinese firms using logistic regression, random forest and gradient boosting between 2016 and 2021. Compared with Almeida and Santos (2025), who used logistic regression and gradient boosting, this study introduced random forest and directly compared different predictive techniques. Its finding that ensemble models outperformed individual models strengthens the argument that model selection can influence predictive performance. However, the study's focus on firms rather than tax administrations and its Chinese institutional context limit the generalisability of the findings. In addition, its recommendation for explainable AI indicates that higher predictive accuracy may create a trade-off with transparency and accountability in tax decision-making.

The OECD (2025) conducted a comparative analysis of predictive analytics across 24 OECD tax administrations between 2020 and 2024. Unlike the individual-country studies, the OECD analysis provides cross-country evidence and reported 25–35% higher early detection of high-risk taxpayers alongside improved audit efficiency. These findings reinforce the positive results reported by the other empirical studies, particularly regarding risk identification and audit targeting. However, the OECD also emphasised the need for explainable predictive models, showing that effective predictive analytics should not be assessed solely by detection accuracy but also by transparency, accountability and the ability of tax officials to understand model-generated risk classifications. More importantly, the evidence largely reflects OECD tax administrations with stronger technological and institutional capacities, leaving limited evidence on how predictive analytics performs within sub-national tax administrations in developing countries such as Nigeria. This limitation creates a specific empirical gap concerning the relationship between predictive analytics and tax fraud detection within the Edo State Internal Revenue Service (EIRS).

Data Mining and Tax Fraud Detection

Edupuganti (2024) examined tax records of 120,000 Indian taxpayers between2018 and 2022 using clustering, association rules and outlier detection. The use of a large administrative dataset and multiple data mining techniques provided a broad basis for assessing fraud patterns and produced a 30–40% improvement in fraud detection. However, the study was based on Indian tax records and focused mainly on the technical accuracy of detection, with limited consideration of the institutional conditions required for implementation. In comparison, Rodríguez Valencia, Morales and Hernández (2025) used data from over 15,000 Latin American companies between 2019 and 2023 and applied decision trees and anomaly detection. Their finding that audit workload fell by 20% while detection rates remained stable suggests that data mining may improve not only detection but also operational efficiency. However, unlike Edupuganti (2024), the study focused more on audit workload than on the extent to which data mining identifies previously concealed fraud.

Nguyen, Patel, and Kaur (2025) analysed 9,500 SMEs across Vietnam, India and Malaysia using classification, clustering and text mining between 2020 and 2024. Unlike the single-country approach of Edupuganti (2024), the multi-country design provides broader comparative evidence, while the inclusion of text mining allows information beyond conventional numerical tax records to be analysed. The study found improved identification of irregular operations and recommended continuous model retraining and inter-agency data integration. However, differences in tax systems across the three countries may make it difficult to determine whether the findings are attributable to data mining itself or to differences in institutional and regulatory environments. This limitation is important when considering its application to a specific sub-national tax authority such as EIRS.

Kou, Xu, and Gan (2021) investigated VAT fraud among 25,000 companies in European Union countries between 2016 and 2020 using association rule mining and network analysis. While the study differs from Nguyen, Patel, and Kaur (2025) in its focus on VAT fraud networks rather than general irregular operations, both demonstrate the ability of data mining to identify relationships that conventional audit procedures may overlook. Kou, Xu, and Gan (2021), however, relied on a cross-border European setting and recommended stronger cooperation among tax authorities. This suggests that the effectiveness of data mining for complex fraud may depend on access to interconnected datasets, which may be less developed in some Nigerian tax administration environments.

Li, Chen, and Zhang (2020) examined 4,800 Chinese manufacturing companies from 2015 to 2019 using logistic regression, clustering and outlier detection. The study reported higher fraud detection rates and fewer false positives and recommended combining data mining with expert judgement. Compared with Kou, Xu, and Gan (2021), who concentrated on fraudulent networks, Li, Chen, and Zhang (2020) adopted a more firm-level approach to identifying abnormal financial behaviour. Their findings therefore suggest that combining computational techniques with professional judgement may improve the practical interpretation of detected anomalies. However, the study's concentration on manufacturing companies limits the generalisability of its findings to tax administration involving taxpayers across different sectors.

Okafor and Eze (2023) examined 3,200 Nigerian companies between 2017 and 2022 using classification and anomaly detection to analyse tax data. Their finding that data mining improved the identification of omitted income and false deductions provides stronger contextual relevance to the present study because the research was conducted within Nigeria. Nevertheless, the study focused on companies at a broader national level and recommended improved data mining infrastructure and personnel training. This indicates that the availability of technology alone may not guarantee effective fraud detection and that institutional and human capacity can influence the usefulness of data mining.

Deloitte (2022) analysed AI-based tax systems across 30 countries in Europe, North America and Asia between 2019 and 2021 and found that integrated AI and data mining systems could improve fraud detection by up to 50% and reduce audit time. In contrast to the individual-country studies, its multi-regional approach provides broader evidence of the operational benefits of integrating data mining with other AI technologies. However, as a comparative industry analysis rather than a study focused specifically on one tax authority, its findings may not adequately capture the institutional, infrastructural and human-resource constraints faced by sub-national tax authorities in Nigeria.

OECD (2025) examined 35 tax administrations between 2020 and 2024 and found that AI and data mining improved real-time fraud detection and audit targeting. Unlike the studies that concentrated mainly on particular data mining techniques, the OECD analysis considered implementation across multiple tax administrations and identified data quality, human capacity and institutional capacity as important conditions for successful implementation. This suggests that differences in outcomes across the reviewed studies may not depend solely on the technical capability of data mining but also on the quality of data and the capacity of tax authorities to use the technology effectively. Despite evidence from Nigeria and other countries, there remains limited empirical evidence on the specific relationship between data mining and tax fraud detection within sub-national tax administration in Nigeria, particularly the Edo State Internal Revenue Service. This unresolved contextual and empirical issue provides the basis for the present study.

Theoretical Review

This study is anchored on the Technology Acceptance Model (TAM), supported byDeterrence Theory and the Resource-Based View (RBV). The three theories jointly explain why the application of Artificial Intelligence technologies can improve tax fraud detection. TAM explains the adoption and use of AI technologies by tax officials; Deterrence Theory explains how the increased probability of detecting fraudulent activities can discourage tax fraud; while RBV explains the organisational resources and capabilities required to effectively deploy AI technologies. On this basis, the theories provide the theoretical foundation for examining the effects of Machine Learning (ML), Predictive Analytics (PA), and Data Mining (DM) on Tax Fraud Detection (TFD).

Technology Acceptance Model (TAM)

The Technology Acceptance Model developed by Davis (1989) explains that the acceptanceand use of technology are influenced primarily by perceived usefulness and perceived ease of use. In tax administration, tax officials are more likely to adopt ML, PA and DM when these technologies are perceived as useful for identifying fraudulent activities and sufficiently easy to use. The more effectively these technologies are accepted and utilised, the greater their potential to improve the processing of taxpayer information, identification of anomalies and detection of suspicious transactions. Thus, TAM provides the adoption mechanism linking the three independent variables to TFD: greater utilisation of ML is expected to improve TFD, greater utilisation of PA is expected to improve TFD, and greater utilisation of DM is expected to improve TFD. This provides theoretical support for $H0_1$, $H0_2$ and $H0_3$.

Deterrence Theory

Deterrence Theory proposes that individuals are less likely to engage in fraudulent behaviour when the probability of detection and punishment is sufficiently high (Gibbs, 2023). Applied to tax administration, ML can increase the probability of detecting abnormal taxpayer behaviour, PA can identify taxpayers with a higher likelihood of non-compliance, while DM can uncover hidden patterns and relationships associated with fraudulent activities. These capabilities make tax fraud more detectable and can consequently strengthen tax compliance and fraud control. Therefore, Deterrence Theory provides the behavioural mechanism through which ML, PA and DM are expected to enhance TFD, supporting the three hypotheses that predict significant relationships between each AI technology and tax fraud detection.

Resource-Based View (RBV)

The Resource-Based View argues that organisational performance depends on the possession and effective utilisation of valuable resources and capabilities (Barney, 2023). In tax administration, ML, PA and DM represent technological resources, while reliable taxpayer data, digital infrastructure and skilled personnel constitute complementary organisational capabilities. The availability and effective utilisation of these resources determine the extent to which AI technologies can improve fraud detection. Consequently, RBV explains why the application of ML, PA and DM should be associated with differences in TFD, provided that EIRS possesses the data, infrastructure and human capabilities required to deploy them effectively. This supports the three proposed regression relationships between ML and TFD, PA and TFD, and DM and TFD.

Summary of Identified Gaps

Existing studies show four main gaps. First, most research is conducted in developed countries with advanced digital systems, limiting applicability to developing economies like Nigeria. Second, many studies rely on secondary or simulated data rather than real tax administration experiences. Third, AI is often treated as a single concept without separating machine learning, predictive analytics, and data mining, limiting understanding of their individual effects. Finally, few studies focus on sub-national tax authorities in Nigeria.

To address these gaps, this study examines the effect of AI specifically machine learning, predictive analytics, and data mining on tax fraud detection in Edo State, Nigeria, using primary data from tax officials.

3.0 Methodology

This study adopts a quantitative survey research design to examine the effect of Artificial Intelligence, specifically machine learning, predictive analytics and data mining, on tax fraud detection in Edo State, Nigeria. The population consists of 350 staff of the Edo State Internal Revenue Service (EIRS), covering audit, compliance, ICT and enforcement units. A sample size of 187 was determined using the Taro Yamane (1967) formula, and 187 questionnaires were distributed to the selected respondents. Of these, 175 questionnaires were correctly completed and returned, representing the valid responses used for the analysis. Primary data were collected through a structured questionnaire designed on a 5-point Likert scale. Data were analysed using descriptive statistics, including frequencies, percentages, mean and standard deviation, and inferential statistics using multiple regression analysis through SPSS. The hypotheses were tested at a 5% significance level using the p-value decision rule, with $p < 0.05$ indicating statistical significance.

The study followed the multiple regression model specification of Ogunsola (2023) to investigate the connection between artificial intelligence tools and tax fraud detection. According to Ogunsola (2023), the functional model specification was stated as:

$$TFD = f(ML, PA, DM)$$

The econometric model was expressed as:

$$TFD = \beta_0 + \beta_1 ML + \beta_2 PA + \beta_3 DM + \varepsilon$$

Where:

TFD = Tax Fraud Detection

ML = Machine Learning

PA = Predictive Analytics

DM = Data Mining
β_0 = intercept
β_1–β_3 = regression coefficients
ε = error term

Table 1:
Operationalization of Variables

S/N	Variable	Type	Operational Definition	Measurement Indicators	Measurement Procedure/Scale	Source of Data
1	Machine Learning (ML)	Independent	The application of algorithm-based systems to automatically identify patterns and anomalies in tax data for fraud detection.	Automated risk scoring; anomaly detection; classification accuracy; pattern recognition	Five-point Likert scale (1–5). Composite ML score obtained by calculating the mean of all valid ML questionnaire items.	Structured Questionnaire
2	Predictive Analytics (PA)	Independent	The use of statistical and AI-driven models to forecast potential tax fraud and identify high-risk taxpayers before audits.	Risk prediction; audit prioritisation; fraud forecasting; early warning signals	Five-point Likert scale (1–5). Composite PA score obtained by calculating the mean of all valid PA questionnaire items.	Structured Questionnaire
3	Data Mining (DM)	Independent	The process of extracting hidden patterns and relationships from large tax databases to detect fraudulent activities.	Clustering; association rule mining; outlier detection; pattern discovery	Five-point Likert scale (1–5). Composite DM score obtained by calculating the mean of all valid DM questionnaire items.	Structured Questionnaire
4	Tax Fraud Detection (TFD)	Dependent	The effectiveness of tax authorities in identifying, preventing and reducing fraudulent tax activities.	Detection accuracy; reduction in tax evasion; audit efficiency; compliance improvement	Five-point Likert scale (1–5). Composite TFD score obtained by calculating the mean of all valid TFD questionnaire items.	Structured Questionnaire

Source: Author's Compilations, 2026.

Pilot Testing, Validity and Reliability of the Instrument

The questionnaire was subjected to a pilot test before the main data collection to assess the clarity, consistency and suitability of the measurement items. The pilot test was conducted among a small group of respondents with characteristics similar to those of the study population but who were excluded from the final sample. The instrument was also reviewed by experts in accounting, taxation and research methodology to establish its content and face validity.

Construct validity was assessed by ensuring that the questionnaire items adequately represented the dimensions of Machine Learning (ML), Predictive Analytics (PA), Data Mining (DM) and Tax Fraud Detection (TFD). The reliability of the instrument was assessed using Cronbach's alpha in SPSS. A coefficient of 0.70 or above was considered acceptable for the study.

Table 2:
Reliability Test of the Research Instrument

Variable	No. of Items	Cronbach's Alpha	Decision
Machine Learning (ML)	5	0.841	Reliable
Predictive Analytics (PA)	5	0.856	Reliable
Data Mining (DM)	5	0.872	Reliable
Tax Fraud Detection (TFD)	5	0.864	Reliable
Overall Instrument	20	0.858	Reliable

Source: Survey, 2026.

4.0 Data Presentation and Analysis

Data were analysed using descriptive and inferential statistics. The descriptive analysis presents the demographic characteristics of the respondents and their responses to the study variables, namely Machine Learning (ML), Predictive Analytics (PA), Data Mining (DM), and Tax Fraud Detection (TFD). Frequencies, percentages, means and standard deviations were used to summarise the responses.

The inferential analysis employed multiple regression to examine the relationships between Machine Learning, Predictive Analytics, Data Mining and Tax Fraud Detection in Edo State. The hypotheses were tested at the 5% level of significance using the p-value decision rule, where a p-value less than 0.05 indicates a statistically significant relationship.

Data Presentation

The descriptive results indicate generally positive perceptions of the use of Machine Learning, Predictive Analytics and Data Mining in tax administration. The mean scores suggest that respondents generally agreed that these AI technologies are useful for identifying unusual transactions, prioritising audits, recognising fraud patterns and improving the efficiency of tax fraud detection. However, these descriptive results represent respondents' perceptions and do not, by themselves, establish statistical significance or causality. The statistical significance of the relationships is therefore assessed through the multiple regression analysis.

The regression analysis examines the extent to which Machine Learning, Predictive Analytics and Data Mining are statistically associated with Tax Fraud Detection in Edo State. Before interpreting the regression results, relevant diagnostic tests were conducted to assess whether the assumptions underlying multiple regression were satisfied, including multicollinearity, heteroskedasticity, autocorrelation and residual diagnostics.

Table 3 summarizes the characteristics of respondents including Gender, age range, profession, years of experience and economic sectors in which they work.

Table 3:
Demographic Characteristics of Respondents

Variable	Category	Frequency	Percentage (%)
Gender	Male	98	56.0
	Female	77	44.0
	Total	175	100
Age Bracket	Below 25 years	20	11.4
	25-35 years	75	42.9
	36-45 years	55	31.4
	Above 45 years	25	14.3
	Total	175	100
Professional Background	Tax Officer	50	28.6
	Auditor	30	17.1
	Compliance Officer	35	20.0
	ICT/Data Analyst	28	16.0
	Internal Control Officer	22	12.6
	Others	10	5.7
	Total	175	100
Years of Experience	Less than 5 years	40	22.9
	5-10 years	65	37.1
	11-15 years	45	25.7
	Above 15 years	25	14.3
	Total	175	100
Sector of Employment	Public Sector	165	94.3
	Others	10	5.7
	Total	175	100

Source: Survey, 2026.

Demographic Characteristics of Respondents

The demographic profile of respondents provides useful context for interpreting findings on the impact of Artificial Intelligence tools on tax fraud detection in Edo State.

Gender Distribution:

Male respondents constituted 56.0%, while females accounted for 44.0%, indicating a slight male dominance but balanced female participation in tax administration roles.

Age Distribution:
Respondents aged 25-35 years were 42.9%, 36-45 years were 31.4%, above 45 years were 14.3%, and below 25 years were 11.4%. This shows a predominantly active working-age population familiar with emerging technologies.

Professional Background:
Tax officers (28.6%), compliance officers (20.0%), auditors (17.1%), ICT/data analysts (16.0%), internal control officers (12.6%), and others (5.7%) formed the sample. This confirms that respondents are directly involved in tax administration and fraud detection processes.

Years of Experience:
Respondents with 5-10 years' experience were 37.1%, 11-15 years were 25.7%, less than 5 years were 22.9%, and above 15 years were 14.3%, indicating adequate professional exposure for informed responses.

Sector of Employment:
The majority (94.3%) were from the public sector, while 5.7% were from other sectors, reflecting the government-driven nature of tax administration in Edo State.

Machine Learning (ML)
Table 4:
Responses on Machine Learning (ML)

S/N	Statement	SD	D	U	A	SA	Mean	Std. Dev.	Decision
1	Machine learning is used to detect tax fraud in my organization.	10 (5.7%)	18 (10.3%)	30 (17.1%)	72 (41.1%)	45 (25.7%)	3.71	0.95	Agree
2	Automated systems help identify high-risk taxpayers.	8 (4.6%)	15 (8.6%)	25 (14.3%)	80 (45.7%)	47 (26.9%)	3.82	0.90	Agree
3	Machine learning improves accuracy in fraud detection.	6 (3.4%)	12 (6.9%)	20 (11.4%)	85 (48.6%)	52 (29.7%)	4.00	0.85	Agree
4	Pattern recognition helps identify fraudulent tax activities.	7 (4.0%)	14 (8.0%)	28 (16.0%)	78 (44.6%)	48 (27.4%)	3.83	0.88	Agree
5	Machine learning reduces human errors in tax fraud detection.	9 (5.1%)	13 (7.4%)	24 (13.7%)	82 (46.9%)	47 (26.9%)	3.83	0.89	Agree
	Overall Mean						**3.84**	**0.89**	**Agree**

Source: Survey, 2026.

From Table 4 above majority of the respondents responded that they felt that machine learning (ML) plays an important role in the upgrading of tax fraud detection in Edo state, a mean of 3.84 means most agreed on either agree somewhat or agree slightly or agree strongly to an average. The most mean score of 4.00 indicating most of the respondents believe ML is important in the detection accuracy of fraud, Similarly, machine learning is also very important for an high mean of 3.82 on the issue that automated system help in identifying high risk taxpayers, as ML minimize human error and improves the capacity to recognise patterns of tax fraudulent activities.

This means that it is more useful and effective, the result indicates that machine learning is of superior and better performance when compare with the other tax detection process in detecting tax fraud. This agrees with research carried out on this matter that says; The ML based fraud detection systems help enhance the efficiency of anomaly detection, risk Assessment, and decision making.

Predictive Analytics (PA)

Table 5:

Responses on Predictive Analytics (PA)

S/N	Statement	SD	D	U	A	SA	Mean	Std. Dev.	Decision
1	Predictive analytics is used to forecast potential tax fraud.	7 (4.0%)	13 (7.4%)	28 (16.0%)	82 (46.9%)	45 (25.7%)	3.83	0.88	Agree
2	It helps prioritize tax audits effectively.	6 (3.4%)	12 (6.9%)	24 (13.7%)	85 (48.6%)	48 (27.4%)	3.90	0.85	Agree
3	Predictive tools provide early warning signals of fraud.	8 (4.6%)	14 (8.0%)	26 (14.9%)	80 (45.7%)	47 (26.9%)	3.82	0.89	Agree
4	Risk forecasting improves fraud detection efficiency.	5 (2.9%)	11 (6.3%)	22 (12.6%)	88 (50.3%)	49 (28.0%)	3.94	0.83	Agree
5	Predictive analytics enhances decision-making in tax administration.	6 (3.4%)	10 (5.7%)	20 (11.4%)	90 (51.4%)	49 (28.0%)	3.95	0.81	Agree
	Overall Mean						**3.89**	**0.85**	**Agree**

Source: Survey, 2026.

Table 5 indicates that the result revealed that a large proportion of the respondents stated that PA is highly beneficial in tax fraud detection in Edo State. Total average score of 3.89 represents an evidence that there is a significant high agreement among the respondents. The prediction tool result show a high value 3.95 indicate that predictive analytics has a great effect on decision making concerning the administration of tax while, as previously indicated 3.94 prediction tool risk forecasting is very effective to fraud detection process.

Furthermore, prediction tool helps in prioritisation of tax auditing correctly(3.90).

The tool predictive model indicate presence of fraudulent conduct (3.82) and used in identification of suspected fraudulent activities (3.83) The prediction tool assist in early prediction of possible fraudulent risk instead of taking actions at the time fraud occurred. Therefore, the implication is that prediction tools could help make use of forecasting, risk and forecasting as significant in the enhancement of tax administration process in terms of effectiveness in identifying and prevent tax fraud in Edo State.

Data Mining (DM)

Table 6:

Responses on Data Mining (DM)

S/N	Statement	SD	D	U	A	SA	Mean	Std. Dev.	Decision
1	Data mining techniques help identify unusual tax transactions.	6 (3.4%)	12 (6.9%)	25 (14.3%)	84 (48.0%)	48 (27.4%)	3.89	0.86	Agree
2	Data mining supports early detection of tax fraud.	7 (4.0%)	11 (6.3%)	23 (13.1%)	86 (49.1%)	48 (27.4%)	3.90	0.85	Agree
3	Pattern discovery improves fraud risk management.	5 (2.9%)	10 (5.7%)	22 (12.6%)	90 (51.4%)	48 (27.4%)	3.95	0.82	Agree
4	Data mining strengthens fraud prevention strategies.	6 (3.4%)	9 (5.1%)	24 (13.7%)	88 (50.3%)	48 (27.4%)	3.93	0.83	Agree
5	Data mining enhances detection of concealed fraud activities.	7 (4.0%)	10 (5.7%)	20 (11.4%)	90 (51.4%)	48 (27.4%)	3.93	0.84	Agree
	Overall Mean						**3.92**	**0.84**	**Agree**

Source: Survey, 2026.

As clearly shown in Table 6, it can be noted that respondents largely agreed on DM in improving detection of tax fraud greatly in Edo state. This obtained overall mean is 3.92, reflecting that there is generally high agreement on it among all participants. Among these items, it was found that pattern detection assists in managing fraud risk effectively and it achieved a mean of 3.95. Similarly, data mining improved fraud prevention techniques and strengthens detection of unknown tax fraud, this achieved mean was 3.93.

It is very common across participants that it is beneficial to trap tax fraud at an earlier stage; thus enabling to detect unusual transactions in tax; in turn data mining can find patterns and anomalies present in large volumes of records in an unprecedented way and extremely useful to tax administration in terms of improving such strategy like fraud detection, risk and prevention; It is powerful tool by means of large amount of information, finding the hidden structure which can provide valuable insight in cases of tax fraud in Edo State.

Tax Fraud Detection (TFD)

Table 7:

Responses on Tax Fraud Detection (TFD)

S/N	Statement	SD	D	U	A	SA	Mean	Std. Dev.	Decision
1	Tax fraud detection systems improve identification of fraud cases.	5 (2.9%)	10 (5.7%)	22 (12.6%)	90 (51.4%)	48 (27.4%)	3.95	0.82	Agree
2	AI tools improve accuracy in detecting tax	6 (3.4%)	9 (5.1%)	20 (11.4%)	92 (52.6%)	48 (27.4%)	3.96	0.81	Agree

	fraud.								
3	Early detection of fraud is enhanced through AI technologies.	5 (2.9%)	11 (6.3%)	21 (12.0%)	88 (50.3%)	50 (28.6%)	3.96	0.83	Agree
4	Tax fraud detection improves transparency in tax administration.	7 (4.0%)	10 (5.7%)	24 (13.7%)	86 (49.1%)	48 (27.4%)	3.90	0.86	Agree
5	AI-based systems enhance overall tax fraud detection effectiveness.	6 (3.4%)	9 (5.1%)	18 (10.3%)	92 (52.6%)	50 (28.6%)	3.98	0.80	Agree
	Overall Mean						**3.95**	**0.82**	**Agree**

Source: Survey, 2026.

Table 7 clearly indicates that the highest of respondents indicated that they agreed with the fact that a tax fraud detection system could successfully detect tax fraud especially when complemented with an artificial intelligent tool. It could also be noticed that there is overall consensus among respondents. As per table, The overall average for this survey is 3.95 indicating high overall consensus by the survey respondents that AI systems boost the general effectiveness of tax fraud systems.

The mean average of the data set of 3.98 also strongly confirms that AI systems increase the general efficiency of the systems in fraud detection and also at an earlier stage (mean = 3.96).

Furthermore, AI systems improve the number of fraud cases reported (mean = 3.95), and increase the level of transparency within the tax administrative systems (mean=3.90). The results show that artificial intelligence tool contribute immensely to enhancing the detection of fraud in terms of their accuracy, speed, transparency, accountability and so on, and it would thus be said that the ability to detect tax fraud is critically determined by the introduction and adoption of new-generation technologies in data science such as the use of machine learning, predictive analysis and data mining in the taxation arena.

Table 8:
Relationship Between Artificial Intelligence Tools and Tax Fraud Detection in Edo State

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.842[a]	0.709	0.704	2.510
ANOVA[a]				
Model	Sum of Squares	Df	Mean Square	F	Sig.
Regression	2456.318	3	818.773	129.845	0.000[b]
Residual	1008.462	171	5.897		

Total	3464.780	174			
Coefficients[a]					
Model	Unstandardized Coefficients	Standardized Coefficients		T	
	B	Std. Error	Beta		Sig.
(Constant)	1.102	0.228		4.833	0.000
ML (Machine Learning)	0.276	0.072	0.312	3.958	0.000
PA (Predictive Analytics)	0.256	0.070	0.284	3.700	0.000
DM (Data Mining)	0.312	0.074	0.329	4.216	0.000

Source: SPSS Output, 2026.

Hypotheses Testing and Results

The multiple regression results indicate a strong positive association between the three Artificial Intelligence tools and Tax Fraud Detection in Edo State. The model produced an R-value of 0.842 and an R^2 of 0.709, indicating that Machine Learning, Predictive Analytics and Data Mining jointly explain 70.9% of the observed variation in Tax Fraud Detection among the respondents. The remaining 29.1% is explained by other factors not included in the model. The ANOVA result shows that the overall regression model is statistically significant ($F = 129.845$, $p < 0.001$).

H_{01}: Machine Learning has no significant relationship with Tax Fraud Detection in Edo State.

The regression result shows that Machine Learning has a positive and statistically significant relationship with Tax Fraud Detection ($B = 0.276$, $\beta = 0.312$, $t = 3.958$, $p < 0.001$). Since the p-value is less than 0.05, the null hypothesis is rejected. This indicates that higher reported use of Machine Learning is significantly associated with higher reported Tax Fraud Detection.

H_{02}: Predictive Analytics has no significant relationship with Tax Fraud Detection in Edo State.
PredictiveAnalytics has a positive and statistically significant relationship with Tax Fraud Detection ($B = 0.256$, $\beta = 0.284$, $t = 3.700$, $p < 0.001$). Therefore, the null hypothesis is rejected. The result indicates that greater reported use of Predictive Analytics is significantly associated with higher reported Tax Fraud Detection.

Discussion of Findings

The findings confirm that Machine Learning, Predictive Analytics, and Data Mining all have significant positive effects on tax fraud detection in Edo State.

Machine Learning:

The result aligns with Appelbaum et al. (2020) and Kokina & Davenport (2017), who found that ML improves audit accuracy and fraud identification. It also contrasts Chukwuma & Okafor

(2021), who noted implementation challenges in developing countries. This study shows ML is effective in Edo State despite such challenges.

Predictive Analytics:

The result supports Li & Zhang (2022) and Deloitte (2021), who found that predictive models improve risk assessment and enable proactive auditing. Although Adeyemi & Salami (2020) highlighted data and infrastructure challenges, the present study confirms its effectiveness in Edo State.

Data Mining:

The findings agree with Ngai et al. (2011) and Chen & Lin (2018), who stated that data mining improves fraud detection through pattern recognition and anomaly detection. While Adewale & Ojo (2019) noted limitations in Nigeria, this study confirms its strong positive impact in Edo State.

5.0 Conclusion

Based on the findings of the study, Machine Learning, Predictive Analytics and Data Mining have significant positive relationships with Tax Fraud Detection in Edo State, Nigeria. The regression results indicate that the three AI technologies jointly explain 70.9% of the observed variation in Tax Fraud Detection, while each technology individually shows a significant positive relationship with the dependent variable.

The findings therefore suggest that the use of these AI technologies is associated with improved capabilities for identifying fraudulent activities, recognising unusual patterns and supporting tax fraud detection processes in Edo State. However, the findings are based on respondents' perceptions obtained through a cross-sectional survey and should therefore be interpreted as evidence of association rather than proof of causal effects. Consequently, effective use of AI for tax fraud detection should be supported by appropriate data quality, technical infrastructure and skilled personnel.

Recommendations

The following recommendations are drawn from this research.

i. The Edo State Internal Revenue Service (EIRS) needs to invest more in machine learning technologies to enhance automated fraud detection in anomaly detection, pattern recognition and risk analysis through tax activities.

ii. Tax authorities can adopt and improve the use of predictive analytics systems that will help them to spot the earlier risk of tax fraud, improve forecasting capability, and prioritize audits more proactively for fraud.

iii. The use of data mining methods in tax administration should also be expanded to better analyses of large tax datasets and to be able to detect hidden structures, deviant transactions, and hidden fraudulent behavior.

References

Adewale, O. A., & Ojo, M. O. (2019). Data mining applications and fraud detection in Nigerian public institutions. *Journal of Accounting and Taxation Research, 11*(1), 33-47.

Adeyemi, T. O., & Salami, A. A. (2020). Challenges of predictive analytics adoption in developing economies: Evidence from public sector data systems. *African Journal of Information Systems, 12*(2), 45-61.

Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2020). Big data and analytics in auditing: An overview. *Accounting Horizons, 34*(3), 1-18.

Chen, H., & Lin, C. (2018). Big data analytics for fraud detection: A review. *International Journal of Accounting Information Systems, 29*, 1-19.

Chukwuma, I. C., & Okafor, G. N. (2021). Technological readiness and adoption of artificial intelligence in Nigerian public sector tax systems. *Journal of African Public Administration, 9*(2), 88-103.

Deloitte. (2021). *Using predictive analytics in tax compliance and risk management*. Deloitte Insights.

de Souza, R. C., Silva, M. A., & Oliveira, P. H. (2024). Machine learning applications in tax fraud detection: Evidence from digital tax administration. *Journal of Financial Crime, 31*(4), 1256-1274.

International Monetary Fund. (2024). *Tax compliance and domestic revenue mobilization in developing countries*. International Monetary Fund.

Kokina, J., & Davenport, T. H. (2017). The emergence of artificial intelligence: How automation is changing auditing. *Journal of Emerging Technologies in Accounting, 14*(1), 115-122. https://doi.org/10.2308/jeta-51730

Li, X., & Zhang, Y. (2022). Predictive analytics and tax compliance: Evidence from digital tax systems. *Journal of Financial Regulation and Compliance, 30*(4), 512-528.

Mergel, I., Edelmann, N., & Haug, N. (2019). Defining digital transformation: Results from expert interviews. *Government Information Quarterly, 36*(4), 101385. https://doi.org/10.1016/j.giq.2019.06.002

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework. *Decision Support Systems, 50*(3), 559-569. https://doi.org/10.1016/j.dss.2010.08.006

Organisation for Economic Co-operation and Development. (2025a). *Tax administration 2025: Comparative information on OECD and other advanced and emerging economies*. OECD Publishing.

Organisation for Economic Co-operation and Development. (2025b). *Artificial intelligence in tax administration: Opportunities, challenges and governance*. OECD Publishing.

Tax Justice Network. (2024). *The State of Tax Justice 2024*. Tax Justice Network.

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Verma, S., & Kumar, R. (2024). Predictive analytics and artificial intelligence for tax compliance and fraud detection. *International Journal of Accounting Information Systems, 53*, 100706.