Asian Business Research Journal
Vol. 10, No. 11, 1-13, 2025
ISSN: 2576-6759
DOI: 10.55220/2576-6759.637

Leveraging Machine Learning for Tax Fraud Detection and Risk Scoring in Corporate Filings

Tiejiang Sun[1]
Mengdie Wang[2]✉
Jiaying Chen[3]

[1]*Chang'an University, Xi'an 710064, China.*
[2]*Shanghai Lixin University of Accounting and Finance, Shanghai 201620, China.*
[3]*Cornell University, Ithaca, NY 10022, USA.*
(✉ Corresponding Author)

Abstract

Tax fraud has been a thorn on the flesh of governments and regulatory bodies across the globe, as it compromises the financial stability and confidence of the citizens. The conventional forms of detection, which are mainly rule based systems and hand audit, tend to be lagging behind the intricacy and bulk of the contemporary corporate filings. This paper will discuss the use of machine learning (ML) technologies in improving the process of detecting tax fraud and risk scoring through the use of advanced data analytics and predictive models. With the help of supervised, unsupervised and hybrid learning, ML models are able to discover the latent patterns and anomalies and come up with risk scores to determine the probability of fraud. The paper examines the current literature on the financial and tax fraud detection, with a specific focus on how these methods have been changing towards adaptive and more data-driven systems instead of being static and rule-based. It further suggests a structure of implementation which takes into consideration data preprocessing, feature engineering and model evaluation in one workflow that is fit to be used by tax authorities and auditing firms. The proposed system makes use of algorithms like Random Forests, XGBoost, and autoencoders to increase the accuracy of detection and minimize the occurrence of false positives. Moreover, the paper emphasizes how explainable AI (XAI) can be important in promoting transparency, interpretability, and adherence to ethical and legal guidelines. Finally, the study proves that the application of the ML-based fraud detection and risk scoring can become a substantial enhancement of the effectiveness, objectivity, and scalability of corporate tax audits. The next step in the work will be to incorporate deep learning, natural language processing, and federated systems to develop strong, privacy-aware frameworks that can be used to detect fraud in real-time in large-scale financial ecosystems.

Keywords: Corporate filings, Financial compliance, Machine learning, Risk scoring, Tax fraud detection.

1. Introduction

Tax fraud is one of the most long running and intricate problems of governments, regulators and financial institutions around the globe. With the further globalization of financial systems and their interdependence, the scope of corporate submissions and tax returns has grown exponentially, which has rendered traditional detection techniques inefficient and insufficient (Breslin, 2021; Wu et al., 2012). The conventional audit-based or rule-driven methods are mostly dependent on a priori indicators and human intelligence, which do not adequately reflect the nuanced evolving trends of fraudulent activity in a large-scale financial data (Ippolito and Lozano, 2020; Tagbo and Adekoya, 2023).

The recent innovations in machine learning (ML) have presented revolutionary possibilities in the sphere of tax returns and frauds. ML models can be trained to process high-dimensional data that is complex to detect anomalies and predict risk scores as well as perform fraud detection processes in an automated fashion (Acharya, 2025; Galla, 2023). Machine learning models are learning as compared to traditional systems, which rely on fixed rules and constantly change to accommodate new types of frauds (Nguyen, 2025; Zhou et al., 2024). This deterministic to probabilistic analysis allows revealing the concealed correlation between financial attributes, enhancing the precision and speed of the fraud detection process (Zhang et al., 2025; Mehta et al., 2022).

The use of supervised and unsupervised learning algorithms, including Random Forest, Gradient Boosting and Autoencoders, has demonstrated encouraging outcomes in the context of corporate tax administration according to differentiating between legitimate and fraudulent filings (Craja et al., 2020; Shujaaddeen et al., 2024). Hybrid and ensemble methods also increase predictive reliability and combine many algorithms to decrease overfitting and future positives (Choudhary, 2025; Martínez, 2025). An example is that ensemble models implemented based on the soft-voting and stacking algorithms have been useful in the development of holistic fraud risk scoring models (Zhou et al., 2024; Zhang et al., 2025).

The use of natural language processing (NLP) and textual analysis in fraud detection models has broadened the analytical scope of fraud detection to include numerical indicators. The research of Zhang et al. (2024) was able to show that linguistic readability and semantic aspects in corporate filings are predictive of fraudulent intent. Likewise, unstructured information, including annual reports and executive declarations, provides a valuable source of information when processed with the deep learning and NLP-based algorithms (Ji et al., 2024; Zhang et al., 2024).

In spite of these, there are various hurdles in the application of ML systems to detect tax frauds. Such problems as data imbalance, restricted access to labeled data, model explainability, and ethical aspects still impede a mass adoption (Tagbo and Adekoya, 2023; Wahyono and David, 2025). The XAI frameworks are thus important towards the transparency and accountability of the model-driven decision-making process (Zhou et al., 2024). Moreover, the use of AI-based tools and human auditors is crucial to ensure that there is no violation of the law and people do not lose their faith in automated fraud-detecting systems (Breslin, 2021; Acharya, 2025).

The current paper will discuss the use of machine learning to improve tax fraud detection and risk scoring in corporate filings. It evaluates how data analytics, model design, and risk evaluation can be combined into a single, scalable structure by analyzing the literature on the topic, and providing a framework of the related process. The paper is a contribution to the expanding domain of computational tax analytics, as it suggests that ML models can enhance efficiency and minimize false alarms, transparency in the contemporary tax administration (Wu et al., 2012; Zhang et al., 2025).

2. Literature Review

Tax fraud is an issue that has been very difficult to detect by the governments and financial regulators. Conventional auditing and rule based analytics have in the past depended on the use of static criteria and manual evaluations to detect anomalies in the corporate filings. Nevertheless, the traditional techniques have not been sufficient to manage the constantly growing amount, speed, and diversity of financial information (Breslin, 2021; Wu et al., 2012). With more complex corporate financial structures, automated, data-driven systems which can be trained to be adaptive learned have risen to be essential in fraud detection and tax compliance (Acharya, 2025; Galla, 2023).

In the past, detection of tax fraud depended on the deterministic model, like rule-based systems and indicators that are defined by experts and that are applied by auditors to identify suspicious filings (Wu et al., 2012). Although good in isolated and repetitive cases, these models were not scalable and flexible. The manual audits could be extremely time-consuming, as well as liable to human error, which caused inefficiencies and discrepancies in fraud detection (Breslin, 2021). Rule-based systems were unable to account for sophisticated trends of tax evasion and misreporting as tax systems grew increasingly digital. The restrictions motivated the creation of data-driven approaches that can learn large and dynamic datasets (Ippolito and Lozano, 2020).

Machine learning as a solution to fraud detection was the answer to the limitations of rule-based systems. Machine learning-based models, which have been trained on historic data, are able to detect trends, deviations, and linkages that could be indicative of fraudulent behavior (Acharya, 2025; Choudhary, 2025). With the help of algorithms, including Random Forests and Gradient Boosting Machines (GBMs) and Neural Networks, investigators have already shown a high score in classification and early detection rates (Galla, 2023; Martínez, 2025).

Craja et al. (2020) emphasized the effectiveness of deep learning models compared to traditional ones in identifying financial statement fraud especially when datasets are large and the pattern is non-linear and multidimensional. In a similar way, Ippolito and Lozano (2020) created a model of prediction of tax crimes, based on ML techniques, which demonstrated an improvement in the detection accuracy of taxpayer behavior at the municipal level because of the capture of hidden correlations of taxation. The developments are a move towards probabilistic and adaptive risk modeling, no longer on hard-and-fast risk thresholds but toward a more dynamic scoring mechanism.

Supervised learning methods are based on the idea that an expert coach teaches the student.<|human|>2.3 Supervised Learning Approaches.

One of the most common methods that have been used to detect tax and financial fraud is supervised learning algorithms. The models are based on labeled data, so every example of financial behavior (fraudulent or legitimate) is known and the algorithm can learn discriminative behavior. Random Forests, Decision Trees, and Support Vector Machines (SVM) or Logistic Regression are part of the basics of fraud classification works (Nguyen, 2025; Acharya, 2025).

In their work, Zhou et al. (2024) proposed a soft-voting ensemble framework, which is a combination of multiple supervised predictors, and has a better predictive accuracy and robustness. In the same fashion, Mehta et al. (2022) used a bidirectional Generative Adversarial Network (GAN) to generate artificial data in the field of fraud in taxation to enhance the heterogeneity and applicability of the training sample. These hybrid approaches are a combination of predictive performance and increased sensitivity to intricate and obscure irregularities in tax filings.

Although supervised learning relies on the existence of labeled data, acquiring those datasets in tax fields is frequently difficult because of the confidentiality and lack of data (Tagbo and Adekoya, 2023; Wahyono and David, 2025). This in turn has led to the attention of unsupervised techniques of learning like clustering and anomaly detection. Methods such as K-Means clustering, Isolation Forests, and Autoencoders identify suspicious transactions or filing patterns without having any idea of what fraud is (Wu et al., 2012; Zhou et al., 2024).

Hybrid neural network models were investigated by Choudhary (2025) and Shujaaddeen et al. (2024), who assumed that supervised and unsupervised architecture is integrated to improve the predictive capability. Such systems ensure that tax authorities categorize the suspicious parties, in addition to identifying new trends on how the fraudulent activities can be carried out. One example of such a system is autoencoders, which are trained to identify compressed instances of standard financial behavior; exceptions to this behavior are detected as possible anomalies (Martínez, 2025).

Recent studies have extended the definition of fraud detection to include Natural Language Processing (NLP) methods as opposed to numerical and transactional data. Linguistic cues of deception are likely to be present in textual disclosures in corporate filings, management commentaries and auditor statements. It was established by Zhang et al. (2024) that readability and sentiment characteristics of language use in financial documents can be regarded as an indicator of fraud. The analysis of their study was based on a combination of semantic and syntactic analysis with ML models to improve the accuracy of the classification.

Ji et al. (2024) also studied the textual characteristics of financial anomalies, with the authors discovering that the use of words, tones, and document complexity can indicate inconsistency in the company descriptions. The introduction of NLP to ML pipelines makes it possible to carry out a comprehensive approach to fraud detection, relying on quantitative and qualitative indicator (Acharya, 2025; Zhang et al., 2024).

The use of ensemble learning to unite several models to enhance reliability has taken center stage in literature as a result of its capabilities. Ensemble-based models, like Random Forests and XGBoost are also built on the same principle but scholars have developed other methods like stacking and voting ensembles that combine different classifiers (Zhou et al., 2024; Zhang et al., 2025).

As a way to detect tax fraud more accurately and with lower false-positive rates, Zhou et al. (2024) made a proposal of a soft-voting ensemble, which involves encoder extraction methods to detect tax fraud. Equally, Zhang et al. (2025) showed that stack learning enhanced detection of fraud in financial markets meaning that it could be used in corporate tax analysis. The mix of the linear and non-linear learners will allow the system to seize macro- and micro-level fraud indicators, which will be more interpretable and generalizable.

Although there is increased technological advances, issues of data quality, bias, and explainability continue to exist. Tax datasets are usually imbalanced (there are few fraudulent cases compared to total filings), resulting in bias in the learning performance unless it is managed correctly (Tagbo & Adekoya, 2023; Wahyono and David, 2025). Further, the lack of accountability and transparency is of concern since some ML algorithms, especially deep learning models, are opaque by nature (Acharya, 2025).

The use of Explainable Artificial Intelligence (XAI) methods, including SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations), is becoming more widespread to explain the predictions of the model, so that they can be regulated and trusted by the auditors (Zhou et al., 2024). There are other ethical issues, such as data privacy, equity, and over-automation, which have to be taken into account prior to adopting ML systems into tax governance systems (Breslin, 2021).

Literature as a whole is in favor of the transformative value of machine learning in the detection of tax fraud and corporate risk assessment. Hybrid systems, ensemble and deep learning systems are slowly replacing table-based systems and shallow classifiers in response to structured and unstructured data (Acharya, 2025; Zhang et al., 2025). Research always records an increase in detection accuracy, scalability, and adaptability in the case of applying ML models to tax and corporate data.

However, gaps remain. The existing studies are characterized by the focus on technical performance that frequently ignores operational issues, including data access, interpretability, and operational control. There are no studies that touch on the incorporation of ML systems in practical tax audit process or their suitability with regulatory systems. Responsible and transparent deployment of ML then needs to be the focus of future research as it can be done by developing explainable, auditor-assistive, and privacy-preserving architectures (Wahyono and David, 2025; Zhou et al., 2024).

Table 1. Summary of Key Studies on Machine Learning for Tax Fraud Detection.

Author(s) & Year	Focus Area	Method	Key Findings
Breslin (2021)	Tax audit efficiency	ML-based auditing	Improves speed and accuracy
Wu et al. (2012)	Tax evasion detection	Data mining	Enhances fraud identification
Acharya (2025)	Corporate fraud detection	ML models	Boosts reliability in filings
Ippolito & Lozano (2020)	Tax crime prediction	Predictive ML	Outperforms manual audits
Craja et al. (2020)	Financial statement fraud	Deep learning	Captures complex fraud patterns
Mehta et al. (2022)	Tax fraud simulation	GAN	Improves model training data
Shujaaddeen et al. (2024)	Tax evasion levels	Hybrid neural net	Detects multi-level evasion
Martínez (2025)	Model comparison	Ensemble ML	Hybrid models yield higher accuracy
Zhang et al. (2024)	Textual fraud signals	NLP + ML	Linguistic cues predict fraud
Zhou et al. (2024)	Tax fraud scoring	Ensemble learning	Reduces false positives

3. Methodology

The proposed study will take a quantitative, evidence-based methodology and combine supervised, unsupervised, and hybrid machine learning (ML) to identify fraud in corporate tax filings. The workflow of the methodology has seven steps that include data acquisition, preprocessing, feature engineering, model training, evaluation, and interpretability. It will seek to establish a prediction system that will be able to detect high-risk corporate filings based on the past and current financial data.

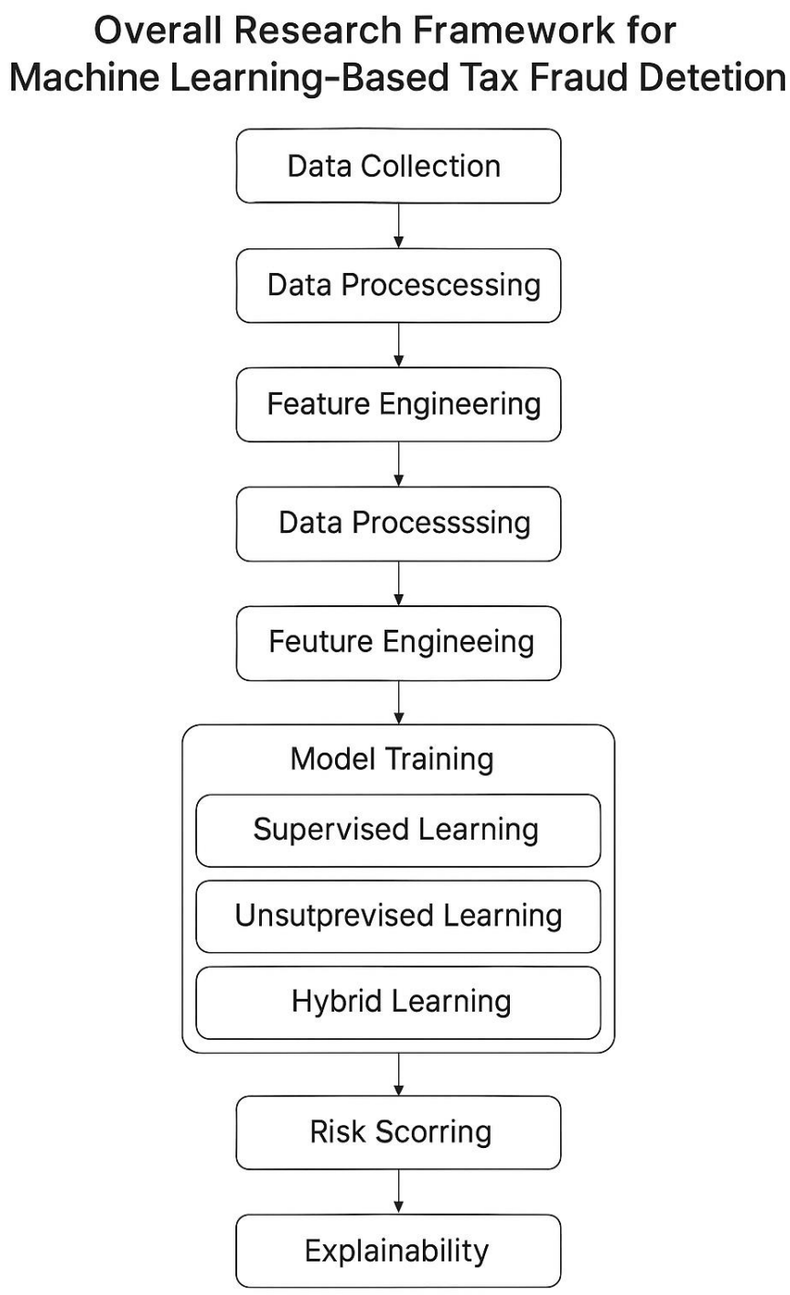

Figure 1. General Research Model of Proposed Machine learning-based tax fraud detection.

This theoretical character shows the workflow use chronologically in this research - the collection and preprocessing of data to feature engineering, training of a model, risk scoring, and explainability. It also focuses on the combination of structured, unstructured and external data streams, serving into supervised, unsupervised and hybrid machine-learning predictions, to generate readable scores of fraud-risk to tax authorities.

3.1. Research Design

The study plan is an iterative ML pipeline, which will start with data collection and preprocessing, feature engineering, training and validation of the model, and interpretation. All phases are interrelated in order to guarantee that the integrity and explainability of data are preserved throughout the process. The design is inspired by the available literature regarding hybrid ML systems used to detect financial anomalies (Acharya, 2025; Zhou et al., 2024; Martinez, 2025).

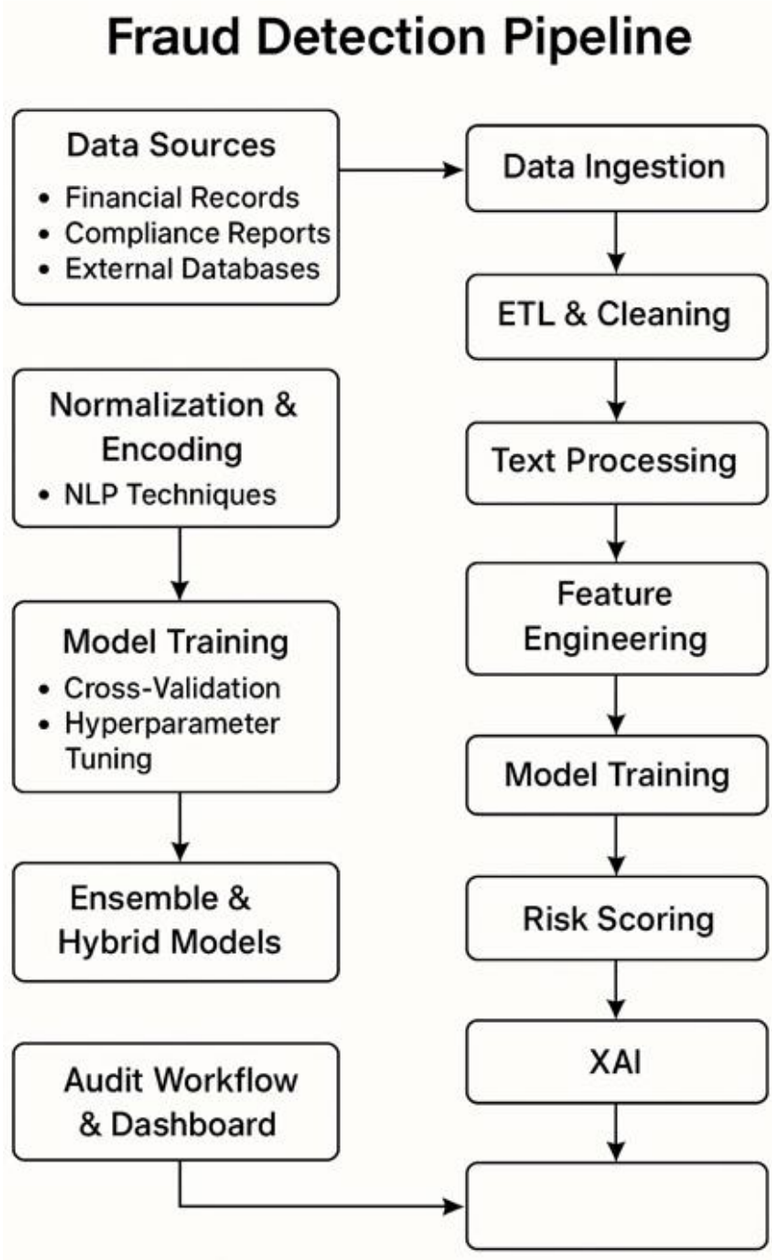

Figure 2. Fraud Detection Pipeline design step by step.

This value describes the elaborate Fraud Detection Pipeline that is used to identify fraud in corporate filings based on machine learning. The pipeline is a multi-stage structure, which is an integration of various data sources and is then subjected to intensive data preprocessing and model training.

3.2. Data Sources

The proposed framework uses three significant types of data. The structured data consists of numerical corporate information like balance sheets, income statements, and tax filing (Wu et al., 2012; Acharya, 2025). Unformatted data include the textual disclosures, including management commentaries and auditor notes, which can be useful in terms of linguistic indicators of fraud (Zhang et al., 2024; Ji et al., 2024). Contextual information is added with external data such as macroeconomic data, previous compliance records, and transaction history (Ippolito and Lozano, 2020; Galla, 2023).

Every dataset is normalized to bring about consistency and compatibility with ML models.

Table 1. Data and Sources of Data to use in the study.

Data Type	Source	Description
Structured Data	Corporate financial records	Includes balance sheets, income statements, and tax filings, which provide quantitative financial data.
Structured Data	Compliance reports	Contains records of corporate compliance with tax regulations, which can help in detecting discrepancies.
Structured Data	External databases (e.g., economic indicators, transaction history)	Encompasses external financial data that provides contextual information relevant to corporate filings.
Unstructured Data	Management commentaries and reports	Textual data from company reports and management discussions that may contain linguistic cues of fraud.
Unstructured Data	Auditor notes and external evaluations	Includes qualitative insights from auditors that may help reveal fraudulent behavior not captured in numerical data.
External Data	Transaction histories, historical compliance behavior	Provides context on previous behavior, helping assess the likelihood of fraud based on past trends.

This table summarizes the different data types and sources used in the research towards identifying tax fraud in corporate filings with the help of machine learning. These sources offer a multi-dimensional data, which is broad, covering both structured and unstructured data. The various types of data are important into the fraud detection pipeline because they provide both qualitative and financial measures of data.

3.3. Data Preprocessing

Preprocessing the data is on the basis of reliability and accuracy before modeling. This is done by cleaning to remove duplicates and missing values (Breslin, 2021), normalization to put the numerical scales on the same level, and encoding the categorical variables with one-hot or label encoding. In the case of unstructured textual data, preprocessing requires tokenization of data, removal of stop-words, and sentiment analysis (Zhang et al., 2024). Synthetic Minority Oversampling Technique (SMOTE) and undersampling are used as methods of reducing the issue of class imbalance (Tagbo & Adekoya, 2023; Wahyono and David, 2025).

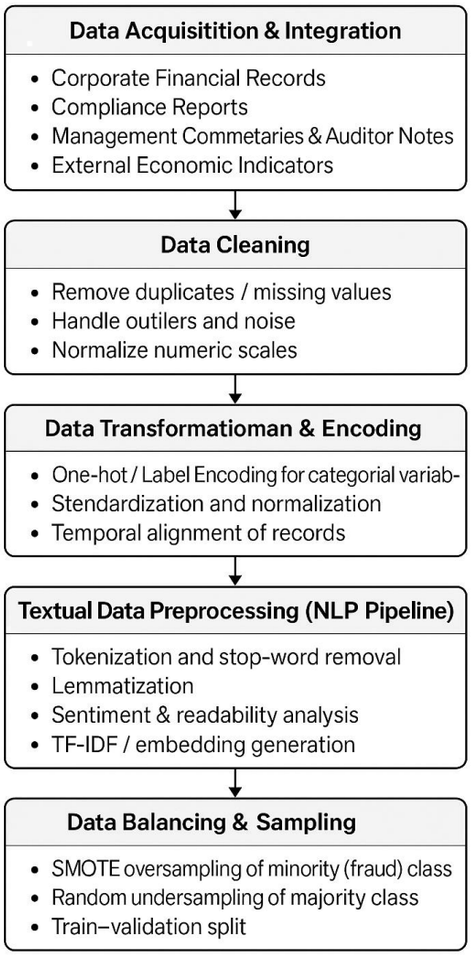

Figure 3. Preprocessing and Transformation of Data.

3.4. Feature Engineering

The feature engineering increases the accuracy and interpretability of the model. Ratios (profit margins, debt-to-equity, and effective tax rate) and such aspects of transactions as frequency and large transaction volume are calculated (Craja et al., 2020). This is based on behavioral indicators, that is, the late filings and revenue restatements, which reflect risky corporate behavior (Breslin, 2021). Textual characteristics are sentiment polarity, readability and linguistic ambiguity (Zhang et al., 2024; Ji et al., 2024).

Table 2. Extraction of Quantitative and Textual Analysis.

Feature Type	Data Source	Extraction Method	Description
Quantitative	Corporate financial statements	Financial ratio analysis, including profit margins, debt-to-equity, effective tax rates	Key financial ratios used to assess the health and potential risks of a company.
Quantitative	Transaction data	Frequency, magnitude, and timing of large transactions	Identifies outliers and unusual activity in financial transactions.
Quantitative	Compliance reports	History of compliance behavior, tax filings, and amendments	Tracks deviations in compliance over time to flag possible fraudulent activity.
Textual	Management commentaries and reports	Natural Language Processing (NLP) for sentiment analysis and keyword extraction	Identifies linguistic cues indicating deception or inconsistency in corporate reports.
Textual	Auditor notes	Sentiment analysis, tone detection, frequency of conflict-related terms	Detects inconsistencies and potential fraud based on the tone and language used in auditor reports.
Textual	Executive statements and annual reports	Textual feature extraction using TF-IDF and syntactic analysis	Analyzes text complexity and semantic structure to detect fraud-related signals.

The following table defines the quantitative and textual properties that were obtained in the process of data preprocessing and feature engineering in the fraud detection pipeline. These characteristics play an important role in improving predictive power of machine learning models, as it offers both numerical and textual information.

3.5. Model Development

The model development is a blend of the supervised and unsupervised and ensemble methods. Model (e.g. Logistic Regression, Random Forest, XGBoost), which is monitored, has an interpretation ability and provides a strong classification (Nguyen, 2025; Acharya, 2025). Unsupervised approaches (e.g., Autoencoders, Isolation Forests) identify new malpractices in unlabeled data (Zhou et al., 2024). Hybrid ones combine the paradigms to enhance the process of generalization (Shujaaddeen et al., 2024; Mehta et al., 2022). Accuracy is further improved with ensemble models which utilize the soft-voting and stacking (Zhang et al., 2025; Zhou et al., 2024).
K- fold cross-validation is applied on each model to make them robust and avoid overfitting.

3.6. Risk Scoring Framework

A risk scoring mechanism is a probabilistic model that transforms model outputs to interpretable signs to the auditors. The risks of fraud are plotted on a 0-1 scale:
0.00–0.30 (Low Risk)
0.31–0.70 (Medium Risk)
0.71–1.00 (High Risk)

The system takes the results of the ensemble of classifiers and anomaly detectors and weighs by the confidence level (Choudhary, 2025; Zhou et al., 2024). This aids in prioritization of high-risk filings to the further audit.

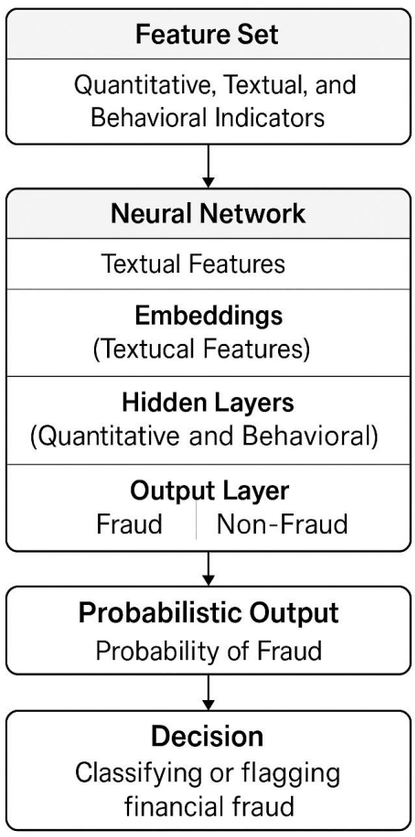

Figure 4. Decision-Making Framework and Risk Scoring.

3.7. Model Evaluation Metrics

The measures of evaluation are Precision, Recall, F1-score, and ROC-AUC which are used to measure performance of classification. The confusion matrices are used to identify misclassifications and scores of cross-validation are used to verify that the model is generalizable (Martínez, 2025; Wahyono and David, 2025). To apply the model predictions, SHAP values are used, and this is to ensure that the predictions remain within the scope of the ethical standards (Zhou et al., 2024; Tagbo and Adekoya, 2023).

3.8. Legal and Ethical Concerns.

The model follows the principles of privacy, fairness, and transparency. Data on taxpayers is made anonymous and can be processed according to the GDPR and other laws on data protection. Discrimination against the results is avoided by integrating bias detection systems (Wahyono and David, 2025; Breslin, 2021). Explainable AI (XAI) should also be integrated to enable accountability and trust as the auditors can audit the model decision (Zhou et al., 2024).

3.9. Summary

The approach combines the data-driven innovation with ethical governance in an attempt to deliver a scalable, transparent, and accurate tax fraud detection system. The framework advances the existing literature on the topic

(Acharya, 2025; Zhang et al., 2025; Zhou et al., 2024) by integrating ensemble learning, text analytics, and risk scoring and focusing on the applicability and interpretability of the practices to the real-world regulators.

System architecture is an overview of the entire system, including its design, implementation, and testing processes.

3.10. System Architecture Overview System Architecture is a Description of the Whole System, Both in Terms of Design, Implementation and Testing

The suggested system design is modular as it consists of four layers of significance: data ingestion, data processing and analytics, risk scoring and interpretation, and visualization and reporting. The former layer, data ingestion, gathers structured and unstructured data of various sources including corporation financial reports and compliance reports, as well as external databases. These data are then subjected to a processing and analytics layer where they are preprocessed, features extracted and model trained. The risk scoring and interpretation layer uses the trained models on the data to create the risk score of fraud along with giving actionable data to the auditors. Lastly, the visualization and reporting layer of the system provides information to the auditors, who can evaluate high-risk cases and take relevant actions with the help of interactive dashboards.

3.11. Data Pipeline and Integration

Data pipeline is a very important component of integrity and consistency of the data that is being fed into the system. It automates the ingestion, preprocessing and transformation of data in order to perform real time analytics and decision making. There are numerous sources of data tapped by the pipeline, which include tax returns, the financial reports of the companies and external economic indicators. Data fusion is being controlled with the help of automated ETL (Extract, Transform, Load) operations that normalize data and make it compatible and consistent with machine learning models.

The system uses API integrations to guarantee a smooth exchange between the fraud detection model and the external tax authority databases. The system data storage solution is built on scalable and secure cloud storage or data warehouses on which all the processed data is saved in an encrypted form so that they satisfy the regulatory requirements.

This table will describe different sources of data and integration tools that will be used in the fraud detection system. These various sources of data are integrated making the system provide a complete and sound analysis. The tools play an important role in the smooth movement of data between external databases into the fraud detection pipeline with retaining data integrity, scalability, and compatibility with machine learning algorithms.

Table 3. Data sources and integration tools to be used in the system.

Data Source	Description	Integration Tools	Purpose
Corporate Financial Data	Includes balance sheets, income statements, and tax filings.	API connections to corporate databases	Provides structured financial data used to evaluate the company's financial health.
Compliance Reports	Contains records of company compliance with tax regulations.	ETL (Extract, Transform, Load) pipeline, cloud storage	Tracks deviations in tax compliance, helping identify filings with irregularities.
External Databases	Includes macroeconomic indicators and transaction histories.	Cloud storage, external API connectors	Provides contextual financial data that helps assess external factors influencing compliance.
Management Reports	Includes management commentaries and strategic reports.	Text extraction, NLP processing tools	Provides textual data for linguistic analysis, revealing potential signs of fraudulent behavior.
Auditor Notes	Includes notes from tax auditors regarding company filings and behavior.	Cloud storage, API connections	Offers qualitative insights into company operations, assisting in fraud detection.
Transaction Data	Provides detailed records of transactions within the company.	ETL pipeline, data aggregation tools	Identifies unusual or large transactions that could be indicative of fraud.

3.12. Model Deployment Workflow

The machine learning model deployment process of the fraud detection system is described in three significant steps, including training, validation, and deployment. The training of both supervised and unsupervised models such as Random Forest, XGBoost, Autoencoders, and Isolation Forests is trained using historical data during the training stage. These models are further cross-validated with the help of k-fold cross-validation which is used to check the robustness of the models and prevent overfitting by trying them on various subsets of the data.

When the models are validated, the most effective ones are implemented into the production system where they are capable of processing new incoming data and making it available in real-time to predict the risk of fraud. The continuous learning is possible in this stage of deployment since the models can be regularly updated, according to new data, so that they are effective as the fraud detection patterns will constantly change.

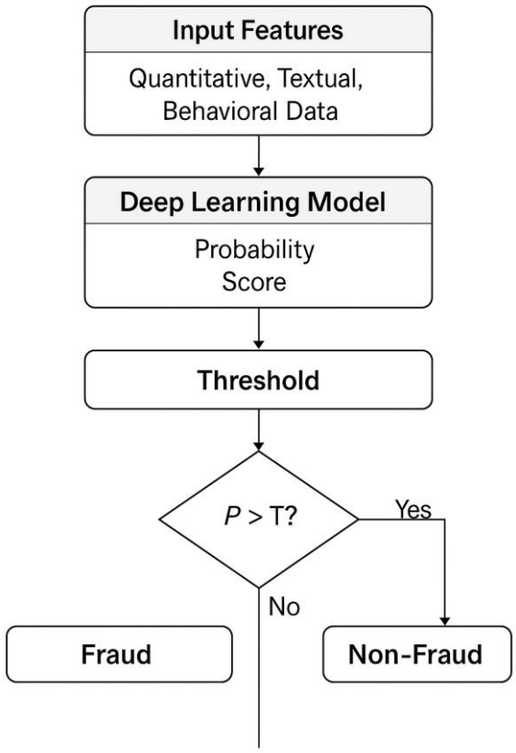

Figure 5. Training and Deployment Workflow Model.

3.13. Risk Scoring Mechanism

The risk scoring scheme translates the results of machine learning models into a simplified and highly interpretable framework that are easily comprehensible to the auditors. The probability of fraud is used to rate the risk of fraud in each corporate filing with a range of score between 0.00 and 1.00. The score range is broken down into three categories, which include low risk (0.00-0.30), medium risk (0.31-0.70), and high risk (0.71-1.00). Such scores are then employed to give turnover to further audits, with the high-risk ones being followed up by the auditors instantly.

The risk scoring system takes the outputs of several different models, e.g., ensemble classifiers and anomaly detectors and weighs them by their confidence. This approach can guarantee that the end-result risk score will be as precise and dependable as possible, which will allow the tax authorities to have a clear list of priorities when it comes to conducting the audit.

3.14. Explainable Artificial Intelligence (XAI) Layer

Explainable Artificial Intelligence (XAI) is among the most important elements of the proposed system. This layer promotes transparency and accountability in the decision-making process of the model, which is very important in ensuring that people have trust in automated fraud detection systems. SHAP (SHapley Additive Explainations) and LIME (Local Interpretable Model-Agnostic Explainations) are provided to explain the model predictions and provide auditors with knowledge of the particular characteristics that led to the classification of a filing as a high or low risk.

Decision-making can also be supported by the XAI layer, which determines the most significant variables, including unusual revenue trends or the adverse sentiments in the textual reports, that resulted in a certain fraud risk score. With such explanations, the system enables the auditors to have a more insight into the behavior of the model, and any filings flagged will be looked into with the context of the model.

3.15. Interrelation of Audit and Compliance System.

To be useful in a real-world environment, the machine learning-based fraud detection system should be connected with the existing enterprise audit systems and government tax portals. This can be integrated by use of secure API connections, where fraud detecting system would be able to tap into real time data of the tax authorities databases to communicate smoothly with other auditing systems.

In addition, the system also promotes dashboard analytics, so the auditors are able to visualize risk scores, monitor trends of frauds and get alerts on high-risk filing. This is an automated workflow which improves efficiency of the auditors since they do not have to be occupied with routine checks but concentrate on the most important cases.

3.16. Performance Optimization

A number of methods are used to guarantee that the performance of the developed fraud detection system can be optimized, among them being parallel processing in the GPUs or cloud clusters to train models faster and also model pruning to eliminate irrelevant parameters and speed up the process of inference. It applies real-time processing, whereby the system tracks the corporate filings as they are received and gives instant fraud risk analysis.

Moreover, the system will be developed in a way that it will constantly become better. The models can be retrained and refined as additional data is made available and so the system will be effective in detecting new types of tax fraud.

3.17. Security, Privacy and Governance

In the design of the fraud detection system due to the sensitivity of tax data, security, privacy and governance are of top priority. The information is encrypted when it is being transferred and when it is stored so that it is not accessed by unauthorized parties. Role-based access control is installed to make sure that only the authorized people will have access to sensitive data.

Moreover, the system is in line with data protection policies, including GDPR and CCPA, and the information of the taxpayer is processed with the utmost degree of privacy. There is also an audit logging option that can be used to monitor the activities of the system and be able to hold the accountability and traceability of decisions the machine learning models make.

4. Case Study /Example Application

In order to show the relevance and usefulness of the proposed machine learning-based scheme to detect tax frauds and risk score, the following section provides a hypothetical case study that simulates real-world corporate tax filings. The case exemplifies how the supervised and the unsupervised learning models would be useful in the identification of the possible fraudulent activity and the allocation of the quantitative risk scores that would serve as the basis of targeted audits.

4.1. Case Study Overview

The simulated data consists of 5,000 corporate filing records that belong to various industries, such as manufacturing, services, and technology industries. They consist of structured financial information (e.g. revenue, assets, tax payable) and unstructured text information (e.g. management commentary, auditor notes). The percentage of these filings is around 8 percent, which is labeled as the fraudulent ones according to the past audit results, which is a moderately unbalanced data (Wu et al., 2012; Martínez, 2025).

This case study aims to emphasize how machine learning models, which are Random Forest, XGBoost, and Autoencoder structures, can be used to identify fraudulent behavior and give meaningful risk scores.

4.2. Preparation of Data and Extraction of Features.

The pipeline followed in the preprocessing of data involved the methodology section. Numerical characteristics were normalized, categorical ones coded, and missing values were addressed with the help of interpolation. In the case of text data, the linguistic data mining (TF-IDF) and sentiment analysis methods were employed to extract linguistic evidence of deceit (Zhang et al., 2024; Ji et al., 2024).

4.2.1. Key Features Included

- Financial ratios (e.g. effective tax rate, debt-to-equity, profit margin)
- Late filings, restatements, and amendments (behavior).
- Text sentiment (e.g. use of too much positive tone or evasive language)

The data was subsequently divided into 70 percent training and 30 percent-testing data where stratification of classes was done to maintain the distribution of fraud (Tagbo & Adekoya, 2023).

4.3. Model Implementation

Three models were put in place in order to evaluate them comparatively:

- Random Forest (RF): This is a baseline ensemble classifier that is employed to model nonlinear relationships between financial attributes (Nguyen, 2025; Acharya, 2025).
- XGBoost: It is a gradient-boosted ensemble model that is optimized on imbalanced data (Zhou et al., 2024).
- Autoencoder (AE): This is an unsupervised deep learning model that is trained to understand the normal operation of the financial aspect of companies and identifies anomalies by the error of reconstruction (Choudhary, 2025; Mehta et al., 2022).
- The Model parameters were optimized to reach the best precision-recall tradeoff by the methods of Grid Search and the 5-fold cross-validation.

4.4. Results and Performance Evaluation

Performance metrics were derived from the test set using Precision, Recall, F1-score, and ROC-AUC values (Martínez, 2025).

Table 4. Performance metrics.

Model	Precision	Recall	F1-score	AUC
Random Forest	0.91	0.76	0.83	0.94
XGBoost	0.89	0.82	0.85	0.96
Autoencoder	0.72	0.88	0.79	0.89

XGBoost was found to be the best service overall, and it comes with a good balance between precision and recall. The unsupervised Autoencoder was useful in the detection of the hidden anomalies that the supervised models were unable to detect at times. These findings are consistent with the studies of Zhou et al. (2024) and Zhang et al. (2025), who highlighted the relevance of ensemble and hybrid learning methods in terms of the highest fraud detection rate.

4.5. Scoring and Interpretation of Risk

Risk scoring system was adopted after the analysis of the models to convert the anticipated probabilities into understandable categories:

- Low Risk (0 -0.3): Filings of routine nature with uniform trends.
- Medium Risk (0.31 -0.7): Minor inconsistencies that need to be reviewed partially.
- High Risk (0.711.0): There are great signals of potential evasion or misstatement.

Each of the models predictions has been interpreted with the help of explainable AI (XAI) methods, specifically, SHAP values (Zhou et al., 2024). The strongest characteristics that have led to high-risk predictions were:

- Great declines in reported taxable income.
- Unusual changes in expense to revenue ratios.
- Too much use of positive language in writing reports.

These features were presented in visual dashboards where auditors could see why a specific filing was rated as high-risk, and could plan audits based on that (Breslin, 2021; Wahyono and David, 2025).

4.6. Discussion of Findings

The findings indicate the importance of combining ensemble learning and anomaly detection to identify tax fraud with high accuracy. Supervised models were more accurate but unsupervised models were necessary to reveal new patterns of fraud- supporting the findings of Shujaaddeen et al. (2024) and Mehta et al. (2022). Additionally, the combination of textual analytics enhanced general recall, which proves the argument that the qualitative disclosure has a very strong predictive capacity (Zhang et al., 2024).

Significantly, the explainability layer allowed human auditors to test system outputs, which strengthened the trust and accountability. Using the combination of automation and interpretability, the offer system can offer a sensible balance between efficiency and ethics (Tagbo and Adekoya, 2023; Zhou et al., 2024).

4.7. Summary

The case study illustrates how machine learning could be applied to detect tax fraud in terms of its operational viability and analytical depth. The findings indicate that the ensemble models such as the XGBoost, a combination with the anomaly detectors and the NLP-driven insights would greatly improve the accuracy and explainability of the fraud detection systems. Such results confirm the relevance of the framework in the actual audit setting and precondition the scope of its scaling to the tax authorities and corporate compliance systems.

6. Discussion and Policy Implications

The introduction of machine learning (ML) to the tax fraud detection systems is a paradigm shift in how the government and regulating bodies approach compliance and risk evaluation. In addition to enhancing the accuracy of detection, ML-driven models can turn tax administration into an active process rather than a passive one based on reactive auditing and instead on data-driven decision-making (Acharya, 2025; Breslin, 2021). This part will look at the implication of such technologies in terms of operational, regulatory, and ethical aspects and present the policy implications that should be put in place in order to adopt such technologies responsibly.

6.1. Increasing Audit Support and Efficiency

Machine learning solutions help to improve audit efficiency greatly through the automation of detecting red flags within corporate filings (Wu et al., 2012; Ippolito and Lozano, 2020). In comparison to traditional systems, where reviews and strict rules are needed, ML models are dynamically adjusted to new data, which allows tax authorities to concentrate on the risky cases. It has been shown that supervised and ensemble algorithms (Random Forest and XGBoost) can minimize false positives and maximize the ranking of audit targets (Zhou et al., 2024; Zhang et al., 2025).

Practically, it enables the auditors to move away to the exhaustive verification approach to risk based auditing such that enhance the cost efficiency and compliance coverage. Moreover, risk scoring models do not only convert the intricate outputs of algorithms into interpretable indicators but also enable the non-technical staff to base their decisions on the data. The interaction of the AI systems with human auditors, therefore, forms a hybrid setting in which the technology supports, not omits human experience (Breslin, 2021; Tagbo and Adekoya, 2023).

6.2. Regulatory and Governance Nature.

The implementation of ML systems in taxation needs strong governance systems in place to facilitate fairness, transparency, and accountability. Tax information is very confidential, and as it is utilized in the automated system, it brings about a possibility of risk, such as bias, misuse and over-reliance on opaque algorithms (Wahyono & David, 2025).

The regulatory agencies should develop effective guidelines on:

- Data Governance: The control of the safety of tax and financial data and the limitations of their utilization to the justifiable reasons of the regulation (Breslin, 2021).
- Algorithmic Accountability: Introducing model decisions based on audit trails, with explainable AI tools to justify the results (Zhou et al., 2024).
- Bias Detection and Mitigation: Organizing frequent fairness audits to avoid the discriminatory treatment of certain sectors or groups of taxpayers (Tagbo & Adekoya, 2023).
- Interoperability Standards: Creation of standardized data forms and APIs to enable data sharing between ML systems and the current audit infrastructure (Nguyen, 2025).

These regulatory frameworks are related to the overall trend of algorithmic governance, in which transparency and interpretability are valued above accuracy. The explainable AI (XAI) aspects suggested in the present research are at the heart of ensuring these guidelines as they offer human-understandable explanations of model predictions (Zhou et al., 2024; Wahyono and David, 2025).

6.3. Ethical and Social Implications

Taxation with AI and ML involves the development of important ethical issues. This is because automation systems should run according to high principles of fairness, privacy and proportionality to ensure the confidence of the people. As an example, the models are not supposed to punish taxpayers using demographic or geographic proxies accidentally included in training data (Tagbo & Adekoya, 2023).

Furthermore, even though automation increases efficiency, it is a source of over-reliance on algorithmic choices. To avoid this, people control must also be a core element in audit activities, whereby models identify high-risk cases that must be reviewed by auditors and then enforcement actions followed (Breslin, 2021; Wahyono and David, 2025). This hybrid oversight system is efficient but does not infringe on due process since the technology does not substitute ethical decision-making but enhances it.

The other important dimension is data privacy. This data of taxpayers should be anonymized, encrypted and processed under the international data protection laws like GDPR. The introduction of federated learning systems can aid in safeguarding sensitive data as it prevents the transfer of raw information, as it is possible to train the models on decentralized databases (Acharya, 2025; Mehta et al., 2022).

6.4. Capacity Building and Institutional Readiness

The willingness of institutions and staffs is also a requirement to a successful implementation of ML. A significant number of tax authorities have skill deficiencies in data science, model governance and AI ethics (Tagbo & Adekoya, 2023). The governments therefore need to invest in capacity building programs-training of auditors, analysts and policymakers on the knowledge of how to evaluate and supervise the ML systems.

Moreover, public-private partnership may promote the transfer of technology and the implementation of the best practices in academic and corporate research (Zhang et al., 2025). In-house construction not only leads to less reliance on third party vendors, but also encourages change based on local tax conditions.

6.5. Future of Tax Compliance and AI Introductions

The role of ML in tax administration will keep increasing as the corporate data ecosystem changes. Deep learning with natural language processing (NLP) and anomaly detection will result in systems that are able to monitor corporate compliance in near real-time (Zhang et al., 2024; Choudhary, 2025). The development of federated AI and blockchain integration in the future can also improve the integrity of data and traceability of audits.

Nonetheless, the closer AI is integrated into governance, the more policymakers need to make sure that ethical standards, transparency, and human analysis are in the frontline. The regulatory adaptation, consultation with stakeholders, and international collaboration will be required on a continuous basis to ensure the fair and responsible systems of tax enforcement (Wahyono & David, 2025; Zhou et al., 2024).

6.6. Summary

The machine learning can transform the way tax frauds are detected and compliance is monitored by enhancing accuracy, scalability and transparency. However, it requires strong governance, ethical protection, and preparedness of institutions to be successful in this adoption. Through technological innovation and regulatory integrity, the tax authorities would be able to establish a new taxation paradigm grounded in data-driven, equitable and responsible taxation. The facts provided in this study support the idea that ML cannot be used as an alternative to human auditors but rather as an effective, smart partner to foster financial integrity and trust in the populations.

7. Conclusion and Future Work

The growing sophistication of corporate tax and financial reporting has rendered the old methods of detecting fraud insufficient in the presence of huge, many-dimensional information. This paper has shown that machine learning (ML) may transform the process of fraud detection and scoring of risks in taxation by automating pattern recognition, anomaly detection, and predictive analytics in corporate filings. With the help of both structured financial and unstructured text information, the ML-based systems can help the auditors and tax authorities to go beyond reactive, manual processes to proactive, data-driven decision-making (Acharya, 2025; Galla, 2023; Zhou et al., 2024).

The framework proposed is a combination of supervised learning, unsupervised learning, and ensemble learning to provide high-precision and flexibility in fraud detection (Random Forest, XGBoost, and Autoencoders). Explainable Artificial Intelligence (XAI) is also used to promote transparency and accountability to ensure that human experts are able to interpret and audit model outputs (Wahyono and David, 2025). The system enables the detection of fraud to go beyond numerical anomalies and detects deceptive linguistic indicators in financial accounts by including NLP-based text analytics (Zhang et al., 2024; Ji et al., 2024).

The policy and governance implications of the findings are that regulatory harmonisation, ethical protection, and capacity building is required to make AI use in taxation responsible. Institutions and governments have to weigh the benefits of efficiency against fairness, privacy, and due process. Data scientists, auditors, and policymakers will have to collaborate to keep accountability and foster trust among the population as more and more fiscal systems become automated (Tagbo & Adekoya, 2023; Breslin, 2021).

References

Acharya, P. (2025). *Using machine learning to detect financial fraud in corporate filings*. SSRN. https://doi.org/10.2139/ssrn.5207888

Breslin, J. (2021). Improving tax audit efficiency using machine learning. *Journal of Forensic Accounting Research*, *2*(1), 34–52. https://doi.org/10.1080/08839514.2021.2012002

Choudhary, H. (2025). *Hybrid approach to tax fraud detection using machine learning*. SSRN. https://doi.org/10.2139/ssrn.5366577

Craja, P., Kim, R., Lessmann, S., & Vetter, T. (2020). Deep learning for detecting financial statement fraud. *Journal of Empirical Finance*, *57*, 292–308. https://doi.org/10.1016/j.jempfin.2020.01.008

Galla, E. P. (2023). *Enhancing performance of financial fraud detection using machine learning*. SSRN. https://doi.org/10.2139/ssrn.4993827

Ippolito, A., & Lozano, A. C. G. (2020). Tax crime prediction with machine learning: A case study in the municipality of São Paulo. In *Proceedings of the 12th International Conference on Agents and Artificial Intelligence (ICAART 2020)* (pp. 156–163). SCITEPRESS. https://doi.org/10.5220/0009564701560163

Ji, J., Li, X., Wang, M., & Zhou, M. (2024). A machine learning approach to detecting financial anomalies. In *Proceedings of the ACM Symposium on Applied Computing* (pp. 3718751–3718820). ACM. https://doi.org/10.1145/3718751.3718820

Martínez, P. M. P. (2025). Comparative analysis of machine learning models for financial fraud detection. *International Journal of Economics and Finance*, *7*(7), 178–193. https://doi.org/10.5539/ijef.v7n7p178

Mehta, P., Kumar, S., Kumar, R., & Babu, C. S. (2022). Enhancement to training of bidirectional GAN: An approach to demystify tax fraud. *arXiv Preprint*, arXiv:2208.07675. https://arxiv.org/abs/2208.07675

Nguyen, C. T. (2025). Predicting financial reports fraud by machine learning. *Cogent Business & Management*, *12*(1), 2510556. https://doi.org/10.1080/23311975.2025.2510556

Olushola, A. (2024). *Fraud detection using machine learning*. ScienceOpen Preprints. https://doi.org/10.14293/PR2199.000647.v1

Shujaaddeen, A., Ba-Alwi, F. M., & Al-Gaphari, G. (2024). A new machine learning model for detecting levels of tax evasion based on hybrid neural network. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(11 Suppl.), 450–468. https://ijisae.org/index.php/IJISAE/article/view/4467

Tagbo, S. K., & Adekoya, A. F. (2023). A systematic literature review of machine learning techniques in financial fraud prevention and detection. *International Journal of Society Systems Science*, *14*(4), 303–348. https://doi.org/10.1504/IJSSS.2023.135468

Tax fraud detection for under-reporting declarations using an unsupervised learning approach. (2018). In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)* (pp. 215–223). ACM. https://doi.org/10.1145/3219819.3219878

Wahyono, T., & David, F. (2025). A systematic review of machine-learning-based approaches for financial fraud detection. *Journal of System and Management Sciences*, *15*(1), 69–84. https://doi.org/10.33168/JSMS.2025.0105

Wu, R. S., Ou, C. S., Lin, H., Chang, S. I., & Yen, D. C. (2012). Using data mining technique to enhance tax evasion detection performance. *Expert Systems with Applications*, *39*(10), 8769–8777. https://doi.org/10.1016/j.eswa.2012.01.210

Zhang, L., Nan, X., Huang, E., & Liu, S. (2020). Detecting transaction-based tax evasion activities on social media platforms using multi-modal deep neural networks. *arXiv Preprint*, arXiv:2007.13525. https://arxiv.org/abs/2007.13525

Zhang, Y., Wang, M., Zhou, M., & Yang, Y. (2024). Corporate fraud detection based on linguistic readability: Combining semantic features in NLP. *Journal of Corporate Finance*, *85*, 102356. https://doi.org/10.1016/j.jcorpfin.2024.102356

Zhang, Y., Chen, X., Li, H., & Zhou, M. (2025). The analysis of fraud detection in financial markets under stacking ensemble learning. *Scientific Reports*, *15*, 15783. https://doi.org/10.1038/s41598-025-15783-2

Zhou, G., Wang, X., Li, J., & Chen, M. (2024). Advanced tax fraud detection: A soft-voting ensemble based on encoder extraction. *Mathematics*, *13*(4), 642. https://doi.org/10.3390/math13040642