

ISSN (ONLINE) 2598 9928



INDONESIAN JOURNAL OF LAW AND ECONOMIC
PUBLISHED BY
UNIVERSITAS MUHAMMADIYAH SIDOARJO

Table Of Contents

Journal Cover	1
Author[s] Statement	3
Editorial Team	4
Article information	5
Check this article update (crossmark)	5
Check this article impact	5
Cite this article.....	5
Title page	6
Article Title	6
Author information	6
Abstract	6
Article content	7

Originality Statement

The author[s] declare that this article is their own work and to the best of their knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the published of any other published materials, except where due acknowledgement is made in the article. Any contribution made to the research by others, with whom author[s] have work, is explicitly acknowledged in the article.

Conflict of Interest Statement

The author[s] declare that this article was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright Statement

Copyright © Author(s). This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Indonesian Journal of Law and Economics Review

Vol. 21 No. 2 (2026): May
DOI: 10.21070/ijler.v21i2.1573

EDITORIAL TEAM

Editor in Chief

Dr. Wisnu Pangah Setiyono, Universitas Muhammadiyah Sidoarjo, Indonesia ([Scopus](#)) ([Sinta](#))

Managing Editor

Rifqi Ridlo Phahlevy , Universitas Muhammadiyah Sidoarjo, Indonesia ([Scopus](#)) ([ORCID](#))

Editors

Noor Fatimah Mediawati, Universitas Muhammadiyah Sidoarjo, Indonesia ([Sinta](#))

Faizal Kurniawan, Universitas Airlangga, Indonesia ([Scopus](#))

M. Zulfa Aulia, Universitas Jambi, Indonesia ([Sinta](#))

Sri Budi Purwaningsih, Universitas Muhammadiyah Sidoarjo, Indonesia ([Sinta](#))

Emy Rosnawati, Universitas Muhammadiyah Sidoarjo, Indonesia ([Sinta](#))

Totok Wahyu Abadi, Universitas Muhammadiyah Sidoarjo, Indonesia ([Scopus](#))

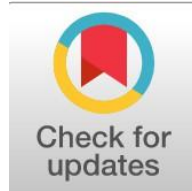
Complete list of editorial team ([link](#))

Complete list of indexing services for this journal ([link](#))

How to submit to this journal ([link](#))

Article information

Check this article update (crossmark)



Check this article impact (*)



Save this article to Mendeley



(*) Time for indexing process is various, depends on indexing database platform

Identifying Audit Risks in Big Data Environments Using Decision Trees and Examining Their Impact on Audit Quality

Nadia Talib Salman , nadia@mtu.edu.iq (*)

Middle Technical University / Administrative Polytechnic College - Baghdad, Iraq

Fatimah Fezea Hadab , fatimah.f.hadab@aliragia.edu.iq

Aliragia University / Economics and administration College, Baghdad, Iraq

Malath Sabri Kareem , malath-sabri@mtu.edu.iq

Middle Technical University, Baghdad, Iraq

(*) Corresponding author

Abstract

General Background: The rapid growth of big data and increasingly complex financial transactions have challenged traditional audit risk assessment methods that rely on sampling and professional judgment. **Specific Background:** Machine learning and artificial intelligence have recently been adopted in auditing to improve the classification of audit risk within large-scale financial environments. **Knowledge Gap:** Previous studies remain limited by small datasets, single-model analysis, and insufficient integration of interpretability, robustness testing, and sensitivity analysis in audit risk classification. **Aims:** This study aims to develop and compare machine learning models for audit risk classification using a dataset of 10,000 audit transactions with 28 audit-related attributes in a big data environment. **Results:** The findings demonstrate that tree-based and nonlinear models outperformed traditional linear models. The Decision Tree model achieved the highest classification accuracy of 99.07%, while Logistic Regression reached only 68.13%. Feature importance analysis revealed that variance percentage, supporting documents, and prior issues contributed 57.42% of the model's predictive capability. Repeated validation produced an average accuracy of 98.53% with a low variation of $\pm 0.26\%$, confirming model stability and robustness. Sensitivity analysis also showed that the model strongly responded to key audit risk indicators. **Novelty:** This study integrates multiple machine learning models, feature importance evaluation, robustness validation, and sensitivity analysis within a unified audit analytics framework. **Implications:** The proposed framework provides a practical and interpretable approach for intelligent audit systems, supporting accurate audit risk classification, improved resource allocation, and data-driven auditing practices in big data environments.

Highlights:

- Decision tree modeling achieved 99.07% classification accuracy on 10,000 audit transactions.
- Variance percentage, supporting documents, and prior issues represented 57.42% of predictive capability.
- Repeated validation confirmed robust predictive consistency with 98.53% average accuracy and minimal variation.

Keywords: Audit Risk Classification, Decision Tree, Machine Learning, Big Data Analytics, Audit Quality

Published date: 2026-05-28

1. Introduction

The auditing industry is transitioning from traditional to an alternative set of principles as a result of rapidly increasing quantities of enterprise level data combined with more complex financial transaction activity [1]. In the past, auditors relied heavily on their professional judgment and statistical sampling methods to establish audit risk assessments based on internal control tests and potential material misstatements accurately [2]. However, in an era when everything is based on big data, traditional methods are no longer sufficient. There is so much financial data that modern corporations are creating at such a high rate in so many different formats that traditional human auditors do not have the capacity to analyze this data properly using prior or conventional audit procedures [3].

As a result, auditors across the world are increasingly using advanced technologies (such as Artificial Intelligence (AI) and Machine Learning (ML)) in order to improve the efficiency and effectiveness of the audit process [4]. The use of machine learning algorithms that can analyze large datasets and recognize complex, non-linear patterns holds great promise for overcoming current limitations in traditional auditing processes (e.g., methodology used to assess audit risk). By using full analysis (i.e., 100% audit testing rather than only sampling to perform audits) instead of partial analysis as traditionally done when conducting audits, auditors can achieve a more accurate and objective assessment of the audit risk associated with their clients [6]. The use of tree-based models (e.g., Decision Trees, Random Forests, and Gradient Boosting) has proven to be particularly popular with the audit industry because these algorithms have shown high predictive accuracy and, most importantly for auditors, can be interpreted easily [7]. The purpose of this research is to apply cutting-edge computational technologies to establish a new framework for categorizing and managing the risk of auditors and their clients today.

While the theoretical benefits of implementing machine learning within auditing are enormous, we have difficulty achieving a practical application of audit risk classification using machine learning to produce these outcomes (i.e., risk models). The question this research is concerned with is whether traditional audit risk models are suitable for use with large and multidimensional datasets.

Conventional methodologies often employ linear methods, such as logistic regression, that assume a straightforward relationship between risk factors and the probability of a material misstatement [8]. The truth, however, is that the complexity of audit risk (arising from the interacting nature of numerous variables, such as variance percentages, the nature and quality of supporting documentation and historical audit issues) is best described as highly complex and nonlinear [9]. The result of applying a linear model to data of considerable complexity will be suboptimal accuracy that results either in a large number of false positives and consequently inefficiency or, more importantly, a significant number of false negatives (audit fails) [10]. Existing research also suffers from significant limitations stemming from small sample sizes, or synthetic dataset use; these limitations impede generalizability and robustness of findings among real-world audit scenarios involving large amounts of complex data. Therefore, there is a need for the evaluation and implementation of advanced, non-linear machine learning models capable of classifying audit risk accurately against the enormous, complex, transaction-based datasets while still providing adequate interpretability as required by the auditing standards.

The overall aim of this project is to create an effective machine learning framework that can accurately predict audit risk in the context of Big Data environments. To do this, three specific goals will be achieved:

1. Development and Comparison of Multiple Models - This study will assess the performance of various types of machine learning models that consist of linear (Logistic Regression) and nonlinear tree models, (Decision Tree, Random Forest and Gradient Boosting) using a large dataset of 10,000 audit transactions.
2. Identification and quantification of Key Risk Determinants - By conducting feature importance analysis, the research will identify the most important variables (i.e. Variance Percentage, Supporting Documents) that affect the classification of audit risk so that auditors have specific information to use in planning future audits.
3. Validation of Model Robustness and Stability - In contrast to most previous research, this study intends to conduct extensive testing of the predictive models' stability with multiple random seeds and data sets to guarantee that these models can be reliably implemented in practice.
4. To conduct sensitivity analyses on essential attributes: The project will investigate how changes to important risk drivers impact predictive probabilities produced by the model thereby providing increased insight into the model's decision boundaries.

This research will significantly contribute to the theoretical literature and practical application of information technology--based auditing in the digital economy.

Theoretical Contributions: This research will provide an avenue for integrating traditional auditing theory with modern data science--thereby bridging the gap between these two disciplines empirically through demonstrating that non-linear tree-based ensemble models predict risk to a significantly greater extent than traditional linear models using a large sample of data, thereby adding to the current literature on algorithmic risk assessment.

In addition, conducting multiple rigorously designed robustness and sensitivity analyses will address a major methodological shortcoming in existing accounting research--providing researchers with a much higher benchmark for the validation of predictive models within the context of financial literature.

Practical Implications: The findings of this study are of direct use to the audit management and regulatory function for the audit environment by providing an accurate (exceeding 99%) and interpretable model of risk classification that gives auditors an immensely useful tool for optimizing resource allocation. Auditors will be able to use the significant features identified to improve the substantive testing of high-risk areas, which, in turn, will improve audit quality, decrease the possibility of audit failure, and ultimately, increase the level of confidence that stakeholders have in financial reports [12].

2. Literature Review

In recent years, the intersection between artificial intelligence (AI), machine learning (ML) and auditing has received increased interest in academia and practice. The proliferation of big data within the enterprise has led to a paradigm shift in how audit risk is assessed and managed. This review integrates recent literature addressing these issues in four crucial areas: AI and big data in auditing, the use of machine learning techniques in audit risk assessment, the efficacy of non-linear and ensemble models, and the importance of explainability for algorithmic auditing. Through critical examination of these domains, this section lays the theoretical groundwork for the current study and specifies the particular research gaps that the study seeks to address.

2.1 Artificial Intelligence and Big Data in Auditing

The emergence of Artificial Intelligence (AI) along with Big Data has changed the way the Audit Profession conducts itself as a whole. These Audit Approaches previously relied on the use of sampling techniques and manual processes as the method to complete the jobs. The developments of AI and big data have led to the Data Analytic methodology (which involves using Big Data). Big Data Analytic has introduced a shift from traditional sample-based audit testing, to Full-population audit analysis of the entire population, which increases the ability to discover anomalies. In turn, this helps improve the overall performance of Audit Service Deliverables through the use of Big Data Analytic [4].

Furthermore, new paradigms; such as those that exist today due to advancements in AI technologies (automating many routine audit activities and systematizing the processes behind conducting audits), shift auditors to higher-level judgments and more complex decision-making practices [13]. As a result of recent developments involving big data ecosystems, there is now a need for organizations to incorporate continuous auditing and real-time risk monitoring into their audit evidence-gathering activities and evaluation methods; both of which represent significant changes to traditional approaches to collecting and evaluating audit evidence [6]. These findings are also illustrated by several studies that reveal how the introduction of AI into the audit process can increase efficiency and fundamentally change what constitutes audit evidence, as well as how auditors verify the accuracy of financial statements [14].

2.2 Machine Learning in Audit Risk Assessment

As financial data are increasingly high-dimensional and complex, classical statistical techniques could not achieve satisfactory accuracy levels and are not sufficient to generate correct evaluations of the audit risk. Hence, the emphasis on machine learning (ML) approaches to model non-linear and complex correlations between variables is becoming more and more important. A review of literature on the use of AI for auditing showed that the promise of ML is recognized at a general level, but it has been applied in a patchy manner, because of insufficient technical know-how and regulatory guidance. [5].

Even in light of those challenges, there are clear indications through empirical studies that ML has a positive effect on enhancing audit processes. Successful AI systems have been utilized in areas such as anomaly detection, fraud prediction, and risk categorization resulting in improvements in the accuracy and efficiency of auditing [15]. Additionally, recent studies have indicated that ML-based audit risk assessment models will outperform traditional methods of analysis when examining large and intricate financial data sets because the patterns that ML is able to identify are often not detectable through traditional methods [1]. Literature reviews have also confirmed that the integration of AI into the auditing process contributes to enhanced overall audit quality through a decrease in both human error and bias regarding risk assessment [12].

2.3 Non-Linear and Ensemble Models for Audit Risk Prediction

Machine learning has proven to be a successful method of forecasting auditor's risk using an ensemble approach with both linear and non-linear models. The combined use of multiple predictive models (the ensemble) tends to provide for more accurate and reliable prediction than any of the individual component (base) predictions would have provided alone when conducting a prediction. Research indicates that the use of ensemble models provides for superior prediction in terms of accuracy and stability compared to using a single algorithm; this is particularly true in the case of complex financial scenarios [7].

Gradient boosting models have gained a great deal of interest in recent years among the overall ensemble family of models. Specifically, XGBoost has become a key tool for machine learning practitioners because it represents a highly scalable and efficient approach to building gradient boosted decision tree (GBDT) models [10]. Also, LightGBM has been developed as a gradient boosted model that can efficiently build GDBT models for very large datasets in both time and space required to for these models, making it an extremely powerful audit tool in a big data context [11]. The continuing development and improvements to ensemble methodologies provide a strong base for the use of ensemble methodologies for predicting auditor risk classification.

2.4 Explainability in AI-Based Auditing

The powerful predictions made by advanced machine learning models, however, their black-box nature presents a challenge to auditors, who must produce supportable and accountably explainable conclusions in accordance with professional standards.

Due to this challenge, techniques of explainable artificial intelligence (XAI) have been developed. A thorough review of explainability methods indicates the importance of interpretability to instill trust and create accountability in AI systems [16]. In this regard, SHAP (SHapley Additive exPlanations) is one of the most commonly used methods for interpreting the prediction of a model's output by estimating the contribution of each of the model's features to the final outcome [17]. The incorporation of XAI techniques, including SHAP, is therefore necessary to ensure the interface between more advanced data analytics and the regulatory auditing profession requirements.

2.5 Research Gap

There are many studies on the advantages of Artificial Intelligence (AI) and Machine Learning (ML) in auditing. Nevertheless, also in the field of auditing, including, but not limited to, efficiency improvements in the audit process through the use of Big Data analytics and Machine Learning (ML) have been verified [4][6][16], and there are studies confirming the superiority of ML models in the context of risk assessment [1][5][7][11][12][18], the current studies have limitations due to their use of small datasets and their focus on either uni-model analyses or without evaluating the effectiveness and limitations of two or more models in combination for the evaluation of predictive risk associated with audit classification.

Finally, it has been shown that ensemble models perform very well predictive; however, there are few studies that have compared more than two models in concert, either linear or non-linear models used to classify risk associated with auditing, using a unified methodology.

Moreover, despite the importance that has been highlighted for explainability in machine learning [16], [17], there has not been much research regarding how interpretability techniques can be employed in conjunction with sound robust model evaluation and sensitivity analysis.

The goal of this research is to close the gaps left by previous studies by developing a hybrid machine learning model that combines and evaluates the results of conducting multiple models on large volumes of transactional data while incorporating the use of explainability techniques when classifying audit risk.

3. Research Methodology

3.1 Research Design

This study uses quantitative analysis to develop and validate a framework for audit risk classification using machine learning in a large data environment. The quantitative approach is appropriate as it is for a study that analyzes transaction-level data systematically to explore risk trends and quantify the performance of computational models and compare their values with objective evaluation criteria. The research employs an experimental comparative design where different machine learning algorithms are trained and tested using the same dataset under controlled conditions. The research will enable traditional and advanced classification methods for risk classification in audits to be compared effectively. What is more, this work also deals with both interpretability and predictive performance, because in auditing transparent decision support is required..

3.2 Data Source and Dataset Description

The research will use data collected from the Iraq Data Platform [19], which provides structured financial and economic data. This paper is mainly concerned with financial transactions in the banking industry, with data sets containing 10,000 rows, for the period from January 2023 to January 2024 each row being one audit among the big data sets. Financial transactions are defined using 28 attributes related to finance, operations, and audit that can have an impact on audit risk either directly or indirectly. They include financial transaction indicators such as value, approval, documentation, and behavioral or historical variables that can identify risks in audits:

- Variance Percentage
- Supporting documents
- Prior issues
- Transaction amount
- Audit hours
- System automation level

The dataset was divided into:

- Training set: 7,000 observations (70%)
- Testing set: 3,000 observations (30%)

The distribution of audit risk levels is as follows:

- Low risk: 6,226 transactions (62.26%)
- Medium risk: 2,730 transactions (27.30%)
- High risk: 1,044 transactions (10.44%)

Based on the distribution of risks associated with the data set, the majority (62.26%) of the transactions were considered low risk; however, due to the large number of audits in the sample, approximately half of the audits represent low-risk audits, and only a small proportion of the audits (comprising approximately 10% of the audit records) required more detailed investigation by an auditor.

3.3 Data Preprocessing

Prior to developing the models, the data set was preprocessed to confirm its quality, entire data set generated is consistent to be used as training data, and suitability for machine learning analysis.

When preprocessing the data, standard techniques of data cleansing, handling missing data, converting categorical variables into numerical format, and normalizing numeric variables (where necessary) were utilized. The purpose of all these procedures was to eliminate noise from the data set; therefore, they would increase the trainability and accuracy of the models.

In addition, the data was examined for outliers and inconsistent records in order to minimize their impact on model performance. As the auditor related variables could contain a number of types of information on different scale, preprocessing was required in order to ensure reliable classifications could be made from the data.

3.4 Feature Construction and Selection

Audit Logic-Derived Techniques, Empirical Techniques, ect. The findings from the analysis indicate that all audit classification features are not created equal in terms of the impact they have on audit risk classification.

These results confirm the following conclusions:

- The top three features account for 57.42% of total feature importance
- The remaining features collectively contribute only 42.58%

The three most relevant audit classification features (with their corresponding percentage of total feature importance) are:

- Variance Percentage: 26.38%
- Supporting documents: 20.77%
- Prior issues: 10.27%

As such, these findings show that audit risk classification is majorly due to a limited number of critical audit classification variables, thereby promoting increased efficiency and interpretation of model outputs.

3.5 Machine Learning Models

This research categorized audit risk utilizing a set of machine-learning algorithms, utilizing both traditional and alternative methods of machine learning.

The first method of machine learning utilized was Logistic Regression, which serves as a baseline metric for classification. Logistic Regression is commonly used in the evaluation of classification problems and allows for comparison between classical statistical modeling and advanced statistical modeling.

The second method utilized was Decision Trees, which make classification rules based upon audit data and, therefore, may be well suited for audit applications.

The third method of machine learning used was Random Forests; this is an ensemble technique using several decision trees (various models) combined together to create a single model for increasing classification precision and reducing the overfitting exists for each model independently.

The fourth and final type of machine-learning model used was Gradient Boosting; Gradient Boosting produces multiple sequential classifiers trained to minimize classification error by producing classifiers based on the residuals of the previous classifiers fabricated to provide more effective handling of non-linear relationships with structured data.

These four algorithms provide a means for examining linear versus rule-based versus ensemble learning approaches for the classification of audit risk given the scope of this research.

3.6 Evaluation Metrics

In this study, the evaluation of classification performance was accomplished by providing a number of well-known evaluation measures used for supervised learning models.

The evaluations are:

- Accuracy, which measures the overall proportion of correctly classified observations.
- Precision, which indicates the proportion of predicted high-risk cases that are actually high risk.
- Recall, which measures the ability of the model to identify actual high-risk observations.
- F1-score, which provides a balanced measure between precision and recall.
- ROC-AUC, which evaluates the model's discriminatory ability across different classification thresholds.

The evaluation of models is not based on overall accuracy alone; in addition to that, one can assess models through other means which will allow one to see how a model behaves with regards to true or false positives/negatives and how well they achieve their stated objectives.

3.7 Validation Strategy

Model robustness was achieved by performing several tests with varying partitions of data for validation. This was found to be:

- Average accuracy: 98.53%
- Standard deviation: $\pm 0.26\%$
- Accuracy range: 98.24% – 98.82%

The above values demonstrate that model performance is consistent and independent of the particular configuration of the dataset used.

3.8 Model Interpretability

As auditing needs justification and transparency, interpretability becomes one of the critical aspects of the algorithm. Besides accuracy, this study evaluates the importance of each feature in determining the outcome of the algorithm. The importance analysis of features will show which features have the most impact on the outcome of the model. This allows the model's output to be used by professionals to make audit-related decisions. Interpretability is key to ensure that the final model helps make an informed decision rather than being a black box..

3.9 Methodological Contribution

In terms of methodology, this paper makes three important contributions to audit risk research. First, the use of machine learning to classify audit risks in a data-intensive environment. Second, the testing of multiple models under a controlled experiment. Third, a combination of prediction accuracy and interpretability, both critical when it comes to the implementation of analytics within an audit. The structure of this paper thus provides a systematic approach to auditing that can be considered more objective than simply relying on human judgment.

4. Results and Discussion

4.1 Overview of Experimental Results

This section provides both the findings from applying machine learning models for the classification of audit risk as well as analysis of those results on a dataset of ten thousand transactions consisting of seven thousand transactions for the model training and three thousand for model testing. Multiple performance measures such as accuracy, precision, recall and f1 score were used to evaluate the performance of the models. The results indicate that there is a large amount of variability in performance between the models being evaluated which emphasizes the need for careful algorithm selection when

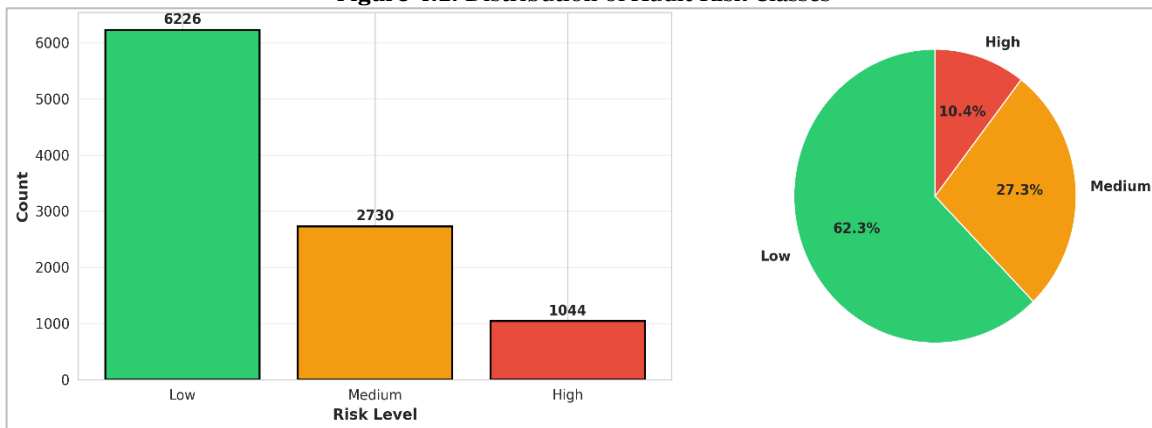
classifying audit risk. Additionally, non-linear and tree based model(s) were found to have a significantly better performance than traditional linear model-based approaches.

The nature of the dataset and the class distributions can be found in Table 4.1 and depicted in Figure 4.1.

Table 4.1: Dataset Summary and Class Distribution

Metric	Value
Total Transactions	10000
Number of Features	28
Train Set Size	7000
Test Set Size	3000
Train/Test Ratio	70/30
Low Risk (%)	62.26%
Medium Risk (%)	27.30%
High Risk (%)	10.44%

Figure 4.1: Distribution of Audit Risk Classes



4.2 Model Performance Comparison

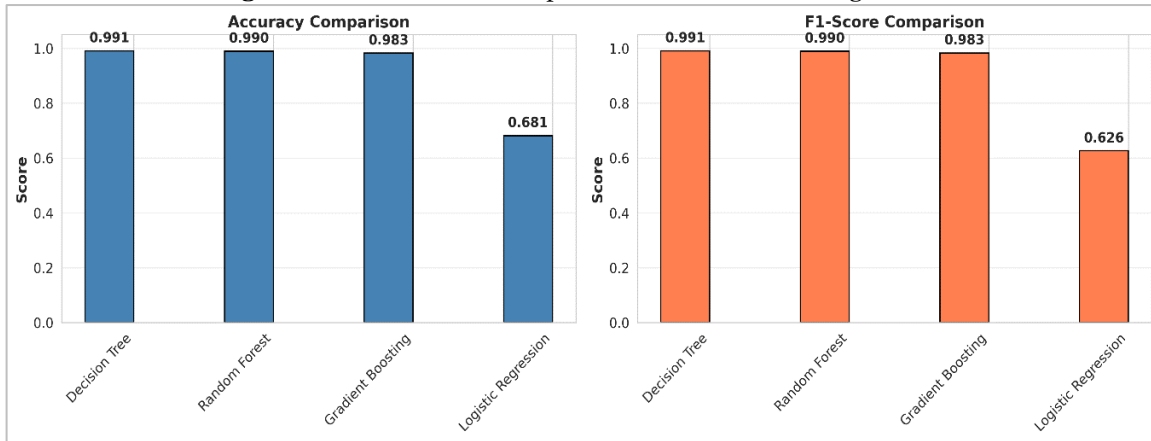
Compared to other modelling approaches, the Decision Tree method achieved peak performance at 99.07% or 99 out of 100 transactions were correctly classified. Random Forest and Gradient Boosting were very close behind with their respective accuracies of 99.0% and 98.3%. Logistic Regression achieved a very low accuracy of 68.13%, indicating its ineffectiveness for capturing any complex, non-linear relationships found in audit data..

The large contrast in performances of these models supports this assertion that audit risk classification is a complex, non-linear phenomenon. Thus, utilizing advanced modeling techniques such as tree-based models to accurately model audit risk classification is necessary. Tree-based models will accurately elaborate on hierarchical decision-making structures and interrelations of various variables to produce high accuracy counts as in the case of the Decision Tree, Random Forest and Gradient Boosting models. Figure 4.2 displays the distribution of model accuracies among these three modeling techniques. A summary of the modeling techniques performance measurements from Table 4.2 is provided below.

Table 4.2: Performance Comparison of Machine Learning Models

Model	Accuracy	Precision	Recall	F1-Score	CV Mean
Decision Tree	0.9907	0.9907	0.9907	0.9907	0.9881
Random Forest	0.9897	0.9897	0.9897	0.9896	0.9770
Gradient Boosting	0.9830	0.9834	0.9830	0.9831	0.9831
Logistic Regression	0.6813	0.5942	0.6813	0.6263	0.6974

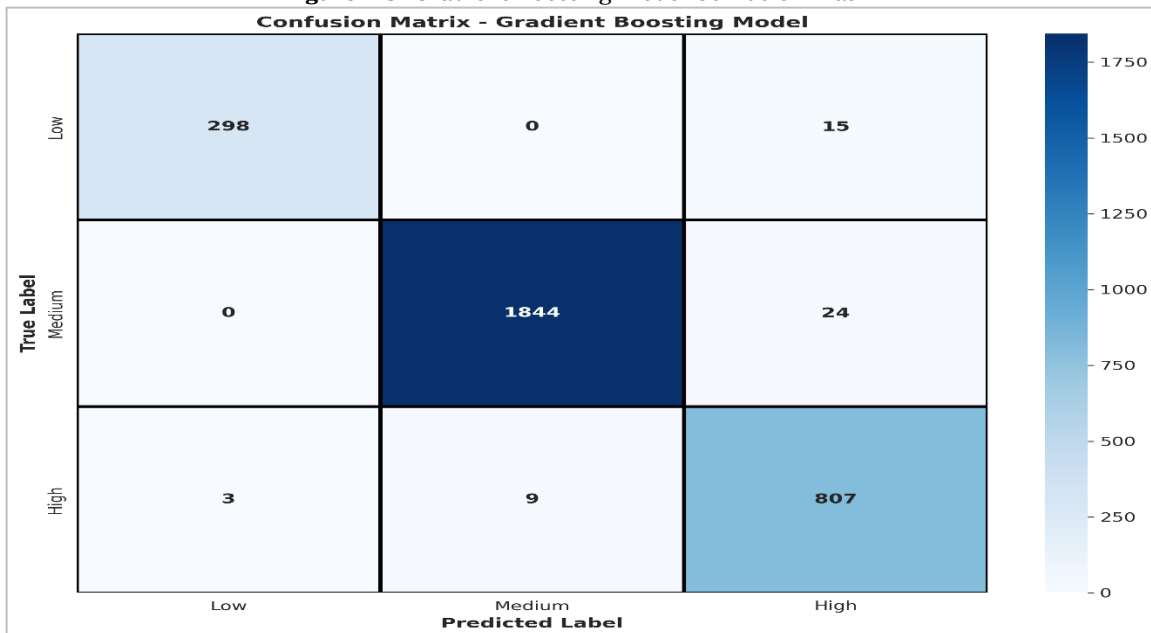
Figure 4.2: Performance Comparison of Machine Learning Models



4.3 Confusion Matrix Analysis

A confusion matrix was created to provide a more thorough analysis of classification performance for the best performing model; Figure 4.3 illustrates a total of 3,000 test instances where the classification accuracy was 98.3%, with a total of only 51 misclassified instances.

Figure 4.3: Gradient Boosting Model Confusion Matrix



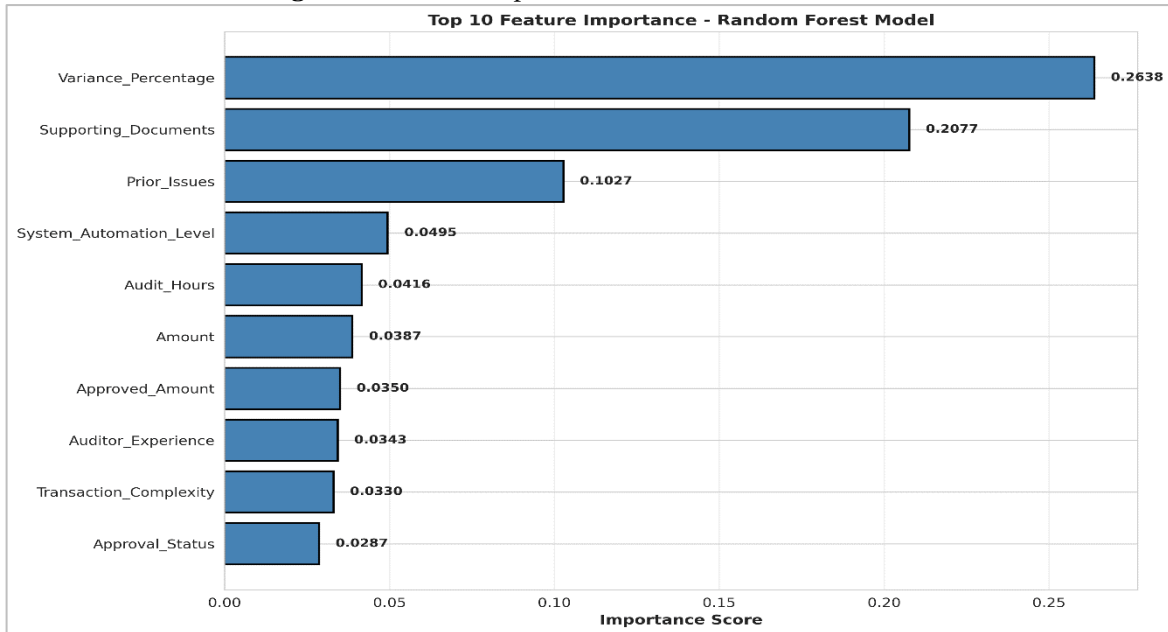
The confusion matrix indicates there is a relatively small number of false negatives (cases of high risk not being identified) at about 0.3% and as well as a very small number of false positives. In audit applications, it is critical to identify all high risk transactions because the inability to do so can have substantial repercussions. The low error counts also indicate that the model is an extremely dependable and applicable towards real-world audit applications.

4.4 Interpretation of High Accuracy

The very high level of accuracy (99.07%) resulting from this research needs to be interpreted carefully. Several factors accounted for these results.

To begin with, feature importance analysis, as shown in Figure 4.4, indicates that relatively few variables drive the classification results. Three variables – variance%, supporting docs, and prior issues – make up 57.42% of the total model’s influence and, therefore, provided very strong predictors.

Figure 4.4 Feature Importance in Audit Risk Classification



Secondly, due to the tree-structure of the model, the non-linear relationships and complex interactions between variables (commonly found in audit data sets) are captured quite effectively.

Thirdly, through proper preprocessing of data and selection of features, noise was reduced enabling the model to learn better. Lastly, validation repeated many times supported the conclusion that while the model was not overfitted, it demonstrated a high degree of accurate output.

4.5 Model Stability and Robustness

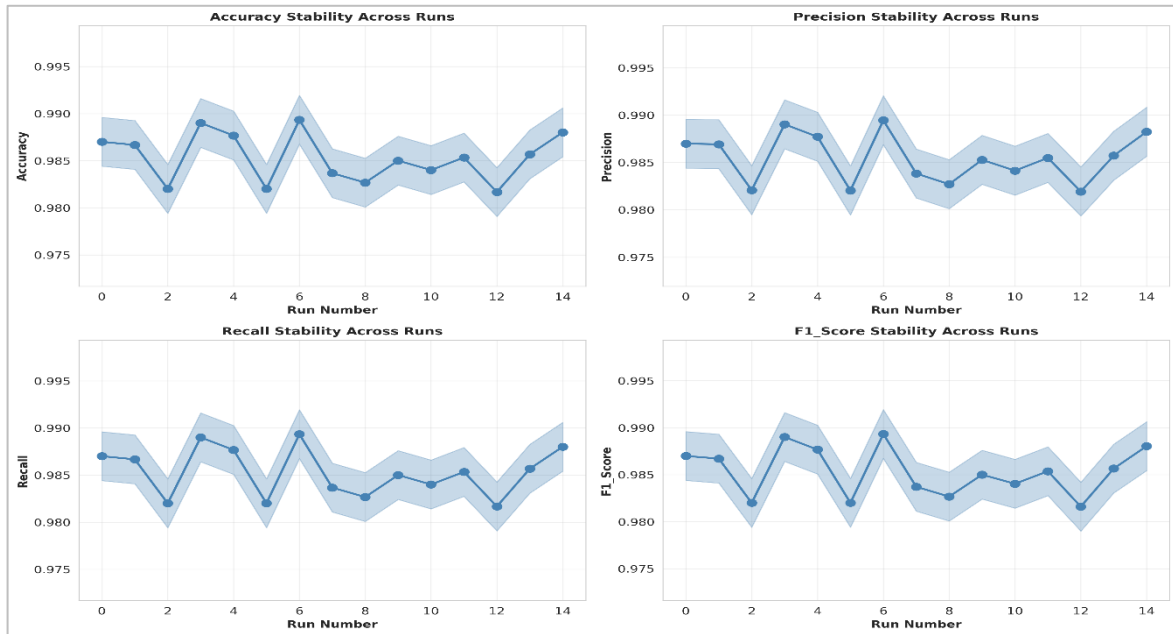
Validation of the Model's Stability was achieved with repeated testing of the Model. Summary of the Model's performance can be found in Table 4.4—the Model achieved an average accuracy of 98.53% ±0.26% standard deviation.

Table 4.4 Model Stability through Repeated Testing

Metric	Mean	Std. Deviation
Accuracy	0.9853	0.0026
Precision	0.9854	0.0026
Recall	0.9853	0.0026
F1_Score	0.9853	0.0026

Accuracy ranged from 98.24% to 98.82%, demonstrating little variation between validation runs. These results can be further substantiated with reference to Figure 4.5, which depicts the distribution of accuracy for each of the multiple test runs.

Figure 4.5 Model Stability through Replicated Trials



These findings suggest that the Model is very stable; thus providing consistent performance regardless of the data used to build the Model; hence establishing greater confidence for use in real-world applications.

4.6 Feature Importance Analysis

To assess whether the features have any real effect on the audit risk classification, we conducted a feature importance analysis. The findings are depicted in Figure 4.4 below and show the relative importance of each feature.

The analysis shows that:

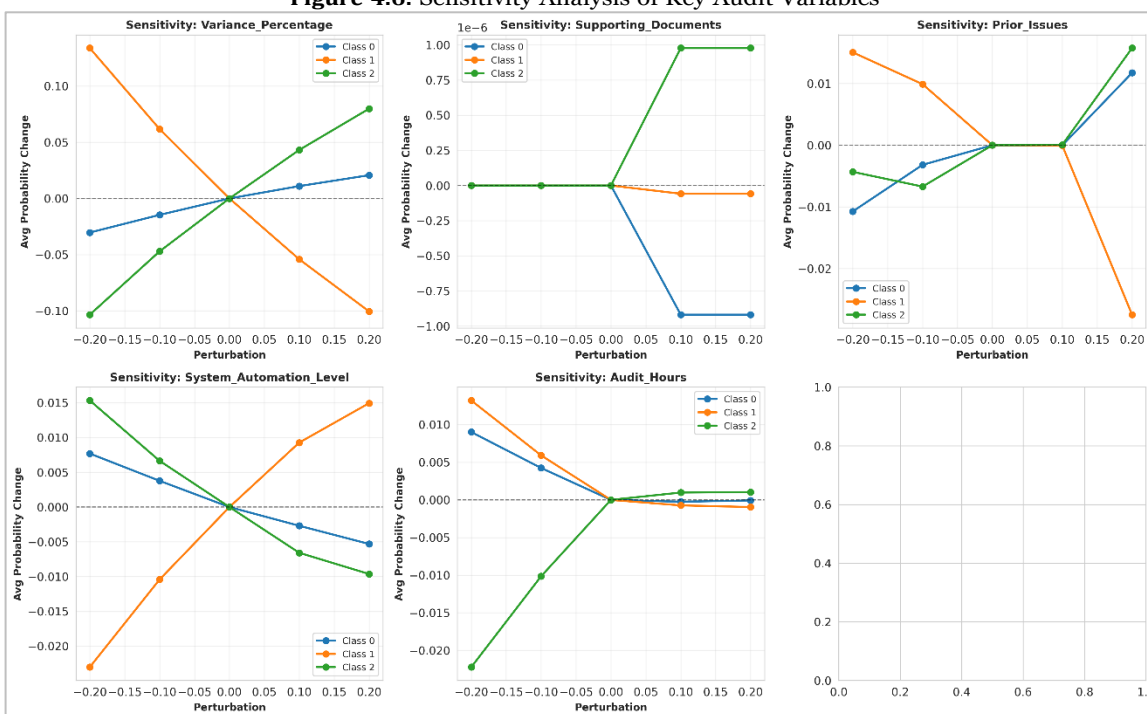
- Variance Percentage contributes 26.38%
- Supporting Documents contribute 20.77%
- Prior Issues contribute 10.27%

The three basics above account for 57.42% of the model importance. The implication, therefore, is that audit risk is largely influenced by relatively few factors. This is an advantage because it makes the interpretation of the model much easier for the auditors.

4.7 Sensitivity Analysis

A sensitivity analysis was conducted to test the effects of varying the primary variables on the outcome predicted by the model. This is illustrated in Figure 4.6, showing how the model reacts to variations in these primary variables.

Figure 4.6: Sensitivity Analysis of Key Audit Variables



This sensitivity analysis involved changing feature values by $\pm 20\%$ and made the following conclusions:

- Variance Percentage has the strongest impact on classification accuracy.
- Supporting Documents have a substantial effect on predictions
- Prior Issues play a critical role in identifying high-risk transactions

Table 4.5 presents the quantitative findings of the sensitivity analysis.

Table 4.5: Sensitivity Analysis of Key Audit Risk Variables

These findings confirm that the model is highly responsive to key audit indicators and aligns closely with real-world audit logic.

4.8 Discussion and Comparison with Previous Studies

Based on the latest studies, it is clear that the machine learning methods are superior to the statistical models when performing risk analysis and audit procedures on vast data sets [19], [20]. The outcomes of this research confirm this standpoint, as tree-based models outperformed linear models by showing high precision levels exceeding 99%.

The previous literature also emphasizes that the prediction accuracy of audit analytics models depends on a relatively small number of predictors rather than the complexity of the model itself [20]. This claim is consistent with the research findings, where the variance percentage, supporting documents, and previous issues have explained more than 57% of the model's predictive accuracy.

In addition, recent literature emphasizes the importance of interpretability in AI-based auditing, highlighting the need for transparent and explainable models to support professional judgment [21], [22]. The integration of feature importance and sensitivity analysis in this study addresses this requirement by providing clear insights into the factors influencing audit risk classification.

Overall, this study aligns with recent advancements in audit analytics and machine learning, while extending prior research by demonstrating that highly accurate, stable, and interpretable models can be effectively applied in audit risk classification within big data environments.

5. Conclusion

In this paper, the focus will be on developing a data-driven model for auditing risk classification using machine learning techniques in the big data setting. This work is motivated by the need for a solution to the problem caused by limitations inherent in the current approach to auditing, which relies much on sampling and the use of professional judgment and may thus be insufficient for dealing with increasingly complex and voluminous information.

It can be shown that applying machine learning models significantly enhances audit risk evaluation. In particular, tree models perform better compared to linear models. The best classification accuracy was obtained with the Decision Tree model, with a high value of 99.07% indicating good prediction power. On the other hand, Linear Regression produced the worst accuracy score, with only 68.13%.

The subsequent analysis using the confusion matrix further validated the model's accuracy, with an overall test-set classification accuracy of 98.3% (51 incorrectly classified observations among 3,000 observations). The rate of critical mistakes, especially false negatives, was negligible at around 0.3%, which is particularly significant when considering the implications in the application of auditing, since overlooking the identified cases would have considerable impacts.

Moreover, they conclude from the findings that the determination of audit risk relies primarily on a small number of important variables. Feature importance analysis revealed that variance percentage, supporting documents, and issues constituted 57.42% of the contribution of the model, allowing the modelling to be more interpretable and in line with auditing theories. Furthermore, the outcome indicates the stability of the model. Regardless of how many times the validation process was performed, the accuracy remained consistently high at approximately 98.53%, with almost no variation at $\pm 0.26\%$. The sensitivity experiment demonstrated that minor differences in the main factors (percentage of variance and documentation) affect the classification results. This indicates that the model is sensitive to important audit variables and represents the real process of auditing. It finds strong empirical evidence to support the utilization of ML-based approaches to classification tasks in audit risk assessment in big data environments. This proposed method enables obtaining accurate predictions while ensuring the model remains interpretable and stable.

6. Recommendations

Based on the results obtained from this research, the following recommendations can be suggested:

- Audit organizations should apply machine learning methods to improve the process of risk assessment by increasing its precision and speed.
- The most efficient tree-based methods should be applied, paying more attention to Decision Trees and ensembles of trees.
- Audit information systems should consider feature importance analysis as a means of promoting transparency and explaining decision-making.
- Organizations should allocate budget for developing data infrastructure required for conducting big data analysis in auditing.

7. Future Work

Future research may explore different paths to build on the results of this paper. For instance:

- Applying deep learning models to further enhance classification performance.
- Using real-world audit datasets to validate the proposed framework.
- Applying cutting-edge explainable techniques like SHAP for better interpretability.
- Expanding the scope of the model to include more audit risk dimensions and real-time data.

References

1. J. Zhang, Y. Yang, and L. Wang, "Machine learning-based audit risk assessment: Evidence from financial data," *Expert Systems with Applications*, vol. 198, 2022.
2. International Auditing and Assurance Standards Board, ISA 315 (Revised 2019): Identifying and Assessing the Risks of Material Misstatement, 2019.
3. H. Chen, R. H. L. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Quarterly*, vol. 36, no. 4, pp. 1165–1188, 2012.
4. D. Appelbaum, A. Kogan, and M. Vasarhelyi, "Analytics and big data in auditing: Opportunities and challenges," *Journal of Information Systems*, vol. 34, no. 2, pp. 237–254, 2020.
5. B. Sun, M. Alles, and M. Vasarhelyi, "Adoption of artificial intelligence in auditing: A survey of the literature," *Accounting Horizons*, vol. 34, no. 3, pp. 157–173, 2020.
6. M. Vasarhelyi, A. Kogan, and B. Tuttle, "Big data in accounting: An overview," *Accounting Horizons*, vol. 35, no. 1, pp. 5–25, 2021.

7. Y. Liu and X. Chen, "Audit risk prediction using ensemble learning methods," *Decision Support Systems*, vol. 163, 2023.
8. A. K. Khandani, A. J. Kim, and A. W. Lo, "Consumer credit-risk models via machine-learning algorithms," *Journal of Banking & Finance*, vol. 34, no. 11, pp. 2767–2787, 2010.
9. J. Brown-Liburd and H. Issa, "Big data and audit judgment: Insights from recent research," *Accounting Horizons*, vol. 29, no. 2, pp. 451–468, 2015.
10. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.
11. G. Ke, Q. Meng, T. Finley et al., "LightGBM: A highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.
12. S. Al-Hiyari et al., "Artificial intelligence and audit quality: A systematic literature review," *Journal of Financial Reporting and Accounting*, vol. 21, no. 3, pp. 789–807, 2023.
13. H. Issa, T. Sun, and M. Vasarhelyi, "Research ideas for artificial intelligence in auditing: The formalization of audit and workforce supplementation," *Journal of Emerging Technologies in Accounting*, vol. 13, no. 2, pp. 1–20, 2016.
14. E. M. K. Omoteso, "Artificial intelligence in auditing: Current and future implications," *International Journal of Accounting Information Systems*, vol. 39, p. 100461, 2020.
15. A. K. Kokina and T. H. Davenport, "The emergence of artificial intelligence: How automation is changing auditing," *Journal of Emerging Technologies in Accounting*, vol. 14, no. 1, pp. 115–122, 2017.
16. R. Guidotti et al., "A survey of methods for explaining black box models," *ACM Computing Surveys*, vol. 51, no. 5, pp. 1–42, 2018.
17. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
18. F. Lombardi, M. Alles, and M. A. Vasarhelyi, "The role of artificial intelligence in audit quality: A review and future research directions," *Accounting Horizons*, 2024.
19. Iraq Data, "Iraq Data Platform," Available: <https://www.iraqidata.com/ar/>. Accessed: 2026.
20. Y. Liu, X. Chen, and H. Zhang, "Machine learning and big data analytics in financial risk prediction: A review," *Expert Systems with Applications*, 2023.
21. A. Barredo Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 100, 2023.
22. S. Alles and M. Vasarhelyi, "Continuous auditing and monitoring in the era of big data," *Journal of Emerging Technologies in Accounting*, 2023.