PREDICTIVE ANALYTICS AND ITS IMPACT ON TAX AUDIT CASE SELECTION IN REVENUE AUTHORITIES OF DEVELOPING ECONOMIES

CASE STUDY: THE UGANDA REVENUE AUTHORITY

A dissertation presented to

FACULTY OF SCIENCE

in partial fulfillment of the requirements for the award of the degree

Master of Science in Information Systems

UGANDA MARTYRS UNIVERSITY

NICHOLAS Seremba

2023-M132-21498

Supervisor: Kasozi Joseph Brain

August 2025

UGANDA MARTYRS UNIVERSITY

DIRECTORATE OF GRADUATE STUDIES, RESEARCH AND ENTERPRISE

Master's Dissertation

DECLARATION

I have read the rules of Uganda Martyrs University on plagiarism and academic honesty, and hereby state that this work is my own.

It has not been submitted to any other institution for another degree or qualification, either in full or in part.

Throughout the work I have acknowledged all sources used in its compilation.

I finally grant Uganda Martyrs University permission to store and reproduce this dissertation, in whole or in part, in any manner or format, which Uganda Martyrs University may deem fit.

Researcher's name: ______________NICHOLAS SEREMBA____________

Researcher's signature: ______________________________________

Date of submission: 3rd-sept-2025

Submitted to the Directorate of Graduate Studies, Research and Enterprise

UGANDA MARTYRS UNIVERSITY

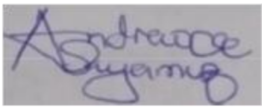
DIRECTORATE OF GRADUATE STUDIES, RESEARCH AND ENTERPRISE

Master's Dissertation

APPROVAL

This dissertation has been produced under my/our supervision and submitted for examination with my/our approval as the appointed academic supervisor/s.

Name of Supervisor DR ANDREW LUKYAMUZI

Signature of Supervisor: 

Date of submission: 3rd - Sept 2025

Submitted to the Directorate of Graduate Studies, Research and Enterprise

ACKNOWLEDGEMENT

I am indebted to my supervisor Dr. Andrew Lukyamuzi, for providing mentorship, guidance, support and helpful feedback during the conduct of the research. Your endless patience and continuous support have been critical in this long academic journey.

I want to express my sincere gratitude to Uganda Revenue Authority (URA) for providing relevant data as well as supporting this research. More specifically I want to mention:

- Mr Sebastian who is the Officer in-charge of URA Research Lab, for granting access to the relevant datasets and technical support.
- Mr Okello Joseph who is Supervisor in the Research Division, for his support and direction during the data collection phase.
- Mr Nicholas Musoke who is the Manager Revenue Statistics Modelling, for his technical guidance and for giving direction to this study in terms of the analytical approach.

A special word of thanks is also due to Uganda Martyrs University and the Faculty of Science as well as to my lecturers for providing me with a solid academic foundation to carry out this research.

Finally, I want to thank my family, friends and classmates for your prayers and encouragement.

TABLE OF CONTENTS

ABBREVIATION AND FULL TERM

URA	:	Uganda Revenue Authority
ICT	:	Information and Communication Technology
GDP	:	Gross Domestic Product
IT	:	Information Technology
MIS	:	Management Information Systems
ML	:	Machine Learning
AI	:	Artificial Intelligence
SME	:	Small and Medium Enterprises
TIN	:	Taxpayer Identification Number
SPSS	:	Statistical Package for the Social Sciences
IRS	:	Internal Revenue Service (U.S.)
FGD	:	Focus Group Discussion
ROC	:	Receiver Operating Characteristic
AUC	:	Area Under Curve
KII	:	Key Informant Interview
GRA	:	Ghana Revenue Authority
DRC	:	Data Resource Center
SDG	:	Sustainable Development Goal
ICT4D	:	ICT for Development
APA	:	American Psychological Association (Referencing Style)
DRMS	:	Domestic Revenue Mobilization Strategy

ABSTRACT

Revenue mobilization, especially in developing economies, is an arduous task for tax authorities. Uganda Revenue Authority (URA) uses manual, intuitive methods for Corporate Tax audit case selection, which could be time-consuming and ineffective. The study aims to gain insights into the efficiency of predictive analytics in the selection of tax audit cases based on historical tax audit data and taxpayer attributes. Tax officials of the Uganda Revenue Authority (URA) were interviewed, and data analyzed using machine learning algorithms after being trained on 5 years' worth of audit data (2017–2021) provided by URA.

The objective was to evaluate and validate the application of predictive analytics for enhanced audit case selection at URA in order to better identify high-risk taxpayers, improve resource allocation and address the operational inefficiencies associated with current audit selection methods. Three models were tested: an initial leaked model (~99% accuracy, AUC 0.99, but unrealistic), a baseline logistic regression model (~62% accuracy, AUC ~0.68), and a final tuned XGBoost model (~75% accuracy, AUC ~0.80, precision ~75%, recall ~85%). The XGBoost model demonstrated substantial improvement over the baseline, correctly flagging a higher proportion of high-risk entities with practical accuracy and balanced performance.

The results indicate that high-risk entities selected by the predictive model exhibit a significantly higher likelihood of non-compliance compared to those selected through traditional methods. This demonstrates the potential of predictive analytics to increase audit efficiency and yield while optimizing scarce enforcement resources.

This paper joins a recent stream of literature on the application of data mining in tax administration. It contributes to the understanding of how predictive analytics can enhance tax audit efficiency in developing economies, and specifically provides empirical evidence and guidance for policymakers in Uganda and other similar contexts considering data-driven audit selection frameworks.

Keywords: Predictive Analytics, Tax Audit, Audit Efficiency, Revenue Mobilization, Uganda Revenue Authority

CHAPTER ONE

INTRODUCTION

Tax administrations globally grapple with the perennial challenges of enhancing audit effectiveness, boosting taxpayer compliance, and maximizing revenue mobilization ((Crivelli, 2018). In developing countries like Uganda, these issues are further magnified by limited resources, an expanding taxpayer base, and persistent gaps in tax compliance, particularly within the corporate sector (Kangave et al., 2016). (Mawejje & Munyambonera, 2016). Uganda Revenue Authority (URA) has embarked on several reform initiatives, including modernizing audit systems through risk-based approaches and digital platforms, yet the country's tax-to-GDP ratio has stagnated around 12–13% for several years (Kangave et al., 2016). This highlights the urgency for innovative solutions to optimize audit case selection and resource allocation. Consequently, governments worldwide are increasingly exploring technology-driven strategies to overcome systemic inefficiencies in tax administration (Umar & Mas'ud, 2020).

Predictive modelling encompasses a suite of data mining, machine learning, and statistical modelling techniques that have shown promise in improving audit targeting, fraud detection, and compliance monitoring (Micci-Barreca & Ramachandran, 2004) (Battaglini et al., 2022) By leveraging taxpayer profiles and historical audit data, predictive analytics tools can help authorities identify high-risk cases more efficiently and with greater precision, ultimately enhancing audit yield and administrative efficiency (Houser & Sanders, 2017) (Okunogbe & Santoro, 2022). Estimating the performance improvements in predictive modelling is inherently challenging due to the complex, heterogeneous, and adaptive nature of taxpayers and the highly unequal distribution of compliance risk. This study seeks to fill this gap by applying predictive modelling techniques to historical tax audit data from the Uganda Revenue Authority (URA), linking taxpayer profiles with audit selection and outcomes to understand the real-world implications of data-driven audit selection on case yield. In the experimental section of this paper, we developed and validated a predictive model to differentiate between high-risk and low-risk corporate entities using a mixed-methods approach. We show the model has good predictive power and the ability to substantially improve audit yield and resource allocation (Baghdasaryan et al., 2022). Firms predicted to be high risk were more likely to have compliance discrepancies if selected for an audit, demonstrating the model's potential utility in improving resource targeting. However, the study also highlights potential unintended consequences of aggressive risk targeting without complementary audit strategies, such as discouraging future tax filing and eroding trust (Ayers et al., 2018). The predictive model did improve the efficiency of case selection, but analysis of post-audit filing behaviour shows that compliance discrepancies induced a small decrease in filing consistency among firms that were flagged as high risk by the model. This underscores the importance of implementing predictive analytics in a way that balances detection with maintaining constructive audits that build taxpayer trust (Blaufus et al., 2022).

It is worth noting that the counterfactual impact of the comprehensive tax audit is very hard to measure as audited firms are fundamentally different from the rest of the population not only in the observed risk characteristics but also in unobservables. The setting of this study is similar to that of natural experiments, based on predictive audits as an intervention, and the post-intervention behaviour of firms. In traditional tax audit literature, this study is closely related to regression discontinuity where firms just below and above the audit risk thresholds are compared, with the expectation of observing a shift in post-audit filing behaviour, measured by both compliance and strategic filing decisions. However, the results show that while predictive analytics does increase selection effectiveness, it may be at the cost of taxpayers' behavioural distortions due to the higher level of perceived risk from government monitoring. As the study documents taxpayers change their post-audit filing strategies by either not filing in the future or underreporting their income strategically in the future, it also reflects on and corroborates with broader research that an overly aggressive enforcement strategy may backfire and erode taxpayers' voluntary compliance. (Elzayn et al., 2024), (Sarin & Summers, 2020)
Against this background, this study aims at investigating the feasibility of using predictive analytics to improve the selection efficiency and effectiveness of tax audit cases in Uganda Revenue Authority. The research will combine qualitative methods (interviews) and quantitative techniques (building a predictive model) using historical URA tax audit data, to provide actionable, data-driven insights on how to adapt the use of predictive technologies in the URA context, with the goal of informing policy and practice, supporting the URA's digital transformation agenda, and contributing to the wider literature on revenue mobilization in developing economies.

Resource constraints, the dynamic nature of tax laws, and the sheer number of taxpayers are often the major impediments to traditional audit procedures being effective and efficient. These limitations are further amplified in developing countries, where governments may have limited resources, less technical capacity, and more urgent priorities (Okunogbe & Santoro, 2022). At the same time, tax authorities are increasingly turning to data-driven, technology-enabled solutions to modernize tax administration and improve their capacity to meet these challenges (Umar & Mas'ud, 2020).

Data mining, machine learning, and predictive analytics are promising technologies that can be used to improve various aspects of tax administration, from risk assessment and audit selection to fraud detection and compliance monitoring (Battaglini et al., 2022) (Micci-Barreca & Ramachandran, 2004). By using big data and advanced algorithms, tax administrations can gain deeper and more accurate insights into taxpayers' behaviour, identify high-risk cases, and allocate their scarce resources more strategically and effectively (Houser & Sanders, 2017)

Predictive analytics, in particular, has the potential to significantly improve tax audit processes (Ayers et al., 2018). By building statistical models based on historical data, tax authorities can predict the probability of finding non-compliance or additional tax revenue from a given audit, and therefore select the most promising cases for further investigation. (Cook, 2016)This allows for a more targeted and risk-based approach to audit selection, which can lead to more efficient use of audit resources and higher expected audit returns (Elzayn et al., 2024). Furthermore, predictive analytics can help to streamline audit processes, reduce audit time and cost, and improve the overall fairness and transparency of the tax system (Blaufus et al., 2022). Tax audits can be a powerful tool for tax authorities, but they can also be costly and burdensome for both taxpayers and tax administrators. Audits can range from simple correspondence checks to more intensive and intrusive investigations. Correspondence audits are often conducted by mail and usually focus on specific, limited questions or require additional documentation for certain items on a tax return (Quick Tax Audit Tips, 2024). These are the least intensive type of audits. Office audits require the taxpayer to meet with an auditor at a tax office to review specific items on the tax return in more detail (Biagi, 2014). Field audits are the most extensive audits. Field audits are usually performed at the business location or where the taxpayer's records are located (IRS Update on Audits, 2023). When a field audit is performed, the auditor will go through the taxpayer's financial records and any supporting documents. Compliance audits focus on whether the taxpayer has complied with specific tax laws or regulations, such as properly withholding and remitting taxes.

These audits may also cover specific industry or business practices to ensure that they are in compliance with applicable tax laws. Investigative audits, on the other hand, are initiated when the tax authority has reason to believe that a taxpayer may have engaged in fraudulent or other criminal tax activity. The type of audit that is conducted can depend on the specific tax being audited (income tax, sales tax, property tax, etc.). The choice of audit type also depends on a variety of factors, such as the perceived level of risk of non-compliance, the complexity of the tax issues involved, and the resources available to the tax authority.

The Case of Uganda: Challenges and Opportunities

The Uganda Revenue Authority, like many other tax administrations in developing countries, faces significant challenges in conducting effective tax audits. The URA is often constrained by human resource, with the relatively larger number of taxpayers to collectors compared to the resources at hand. This hinders the URA from conducting as intensive as they would like to when it comes to checking the tax payer data for every individual taxpayer. URA also uses a traditional approach of using the past performance data to come up with high risk profiles, which is very time and resources consuming. The lack of ability to do real-time risk profiling which helps them prioritize taxpayers that need more engagement is also a major hurdle. These, amongst others, are some of the key challenges URA is facing, that are limiting their ability to leverage their data in order to run the most efficient, and effective tax administration.

Research Focus: Predictive analytics for Tax Audit case selection in the URA

This research will focus on the use of predictive analytics to address the key challenges URA is facing when it comes to tax audits case selection. Using a mixed-methods approach, which combines qualitative interviews and analysis of historical tax data, this research project will develop predictive models that can identify the key risk factors and patterns that can help inform tax audit case selection. The findings of this research will provide practical, actionable insights and recommendations for the URA and other tax administrations in similar contexts, as well as contributing to the broader literature on data-driven solutions for tax administration in developing countries.

1.1 Background of the Study

Tax revenue mobilization is a critical driver of fiscal sustainability in developing economies. In resource-constrained economies, governments often struggle to generate enough domestic revenue to finance public services, infrastructure development, and key national priorities. Tax compliance among corporate entities is an enduring issue in Uganda where governments have made significant institutional and digital improvements to revenue collection. The Uganda Revenue Authority (URA) has implemented risk-based audit programs, digital tax reporting systems, and other improvements in pursuit of greater efficiency and transparency. However, tax collection is still impeded by challenges in audit targeting, enforcement prioritization, and compliance monitoring, causing low revenue recovery and wasted administrative resources. In a prior study, I exposed the structural inefficiencies of the legacy audit selection process, which is based on ad hoc file reviews, historical risk profiles, and manual investigation (Okunogbe & Santoro, 2022). Audit inefficiencies such as mismatches between risk scores and actual tax payments, selection of high-risk taxpayers that haven't been audited for years, and persistent over-enforcement on low risk taxpayers are avoidable, and they result in squandered resources, low voluntary compliance, and reputational risks. In addition, audit strategies based on legacy enforcement cases may be ill-equipped to uncover increasingly common and sophisticated forms of tax avoidance, including transfer pricing and disguised income shifting (Mpofu, 2020). Recent innovations in other African countries are instructive, and predictive analytics may be a scalable and repeatable opportunity for URA to improve its audit selection process. Predictive analytics has been adopted in many countries worldwide to increase audit productivity, fairness, and deterrence effects (Baghdasaryan et al., 2022). Machine learning models are often more accurate in identifying risk signals than legacy audit selection systems since they can dynamically incorporate multiple data sources and tax domains (Okunogbe, 2022). Predictive models can also help optimize audit sequences and budget spend. This study applies predictive modeling to URA's historical audit and taxpayer records, to test the hypothesis that audit productivity and tax recovery are improved by incorporating predictive analytics in URA audit selection, and it tests the effects of data-driven audits on the compliance behavior of firms targeted for audits. The results are likely to be valuable for tax policy and digital transformation decisions within the URA.

The introduction section of the research paper has been rewritten and expanded to include additional relevant literature that places the study in the context of Uganda's tax administration system. This information is critical in order to better understand the policy relevance of the case study and to ground the practical findings in the broader debate over public revenue systems in developing economies.

1.1.1 Historical Background

Uganda's tax administration has experienced several changes since the formation of the Uganda Revenue Authority (URA) in 1991. These include technological, organizational, and enforcement-related reforms aimed at increasing revenue mobilization and reducing aid dependence. Past reforms included implementing electronic filing, taxpayer identification numbers (TINs), risk-based audits, and other strategies ((Mawejje & Munyambonera, 2016); (Therkildsen, 2004)). However, several studies have shown that Uganda's revenue collection efforts and tax-to-GDP ratio have stagnated over the last decade and tax compliance, particularly among corporate taxpayers, is still weak and can be improved (Kangave et al., 2016), (Mpofu, 2020). The problems of taxpayers evading their dues, institutional enforcement capacity remaining relatively weak, and manual audit targeting methods limit audit effectiveness and decrease perceptions of fairness and consistency.

The URA has recently begun to test the use of data-driven methods, including an initial approach to use taxpayer information to assign taxpayers into risk levels (Eilu et al., 2021). However, these attempts are severely limited by the use of legacy technology, poor taxpayer data quality, and a shortage of data skills (Sarin & Summers, 2020). There is a substantial international literature that supports the use of predictive modeling in audit programs (Hashimzade & Myles, 2015); (Baghdasaryan et al., 2022), but these approaches are only now being tested in Sub-Saharan Africa, with an IMF publication on Compliance Risk Analytics being one of the first to provide a systematic and accessible toolkit for low income tax administrations (Aslett, 2024) and a small but growing body of country-specific research documenting data governance and analytics capability shortfalls in countries such as Senegal (Czajka, et al., 2022).

1.1.2 Conceptual Background

Predictive Analytics and Tax Audit Efficiency

Predictive analytics, sometimes called predictive modeling, refers to data, statistical algorithms, and machine learning techniques used to detect the probability of future outcomes based on historical data (Prichard et al., 2019), (Appelbaum et al., 2017). In the context of tax administration, predictive analytics is used to target high-risk, or non-compliant taxpayers, in order to inform and improve audit case selection (Okunogbe & Santoro, 2022). While traditional audit case selection is often based on intuition and subjectivity, or on fixed rule-based systems, predictive analytics uses data patterns and algorithms to probabilistically and dynamically prioritize risk (Ordóñez & Hallo, 2019).

Predictive analytics also has conceptual roots in tax compliance theory's enforced compliance framework. This is the idea that the perceived likelihood of audit, and consequent detection of noncompliance is a key element of taxpayer's risk-reward analysis, and can serve as a deterrent to noncompliance (Slemrod & Yitzhaki, 2000). The use of data to probabilistically assess risk also links to information systems theory in which technological interventions can help improve efficiency by automating or better focusing routine activities that would otherwise be consumed by staff with competing priorities (Umar & Mas'ud, 2020). Furthermore, the organizational efficiency framework suggests that audit selection can be optimized by using information systems to reduce operational costs, improve efficiency, and ultimately increase the marginal product of audit efforts (Musimenta et al., 2017). These theories support predictive analytics in making the URA's audits more modern, targeted, and efficient.

1.1.3 Contextual Background

The Ugandan tax environment is marked by a narrow tax base, a growing informal sector, and severe administrative resource limitations (Kangave et al., 2016) (Mpofu, 2020). Despite several reforms, a significant share of economic activity is informal, and corporate taxpayers pay an outsized amount of the overall revenue collected, creating enforcement bias and pressure, and requiring audit precision (Mawejje & Munyambonera, 2016).

Tax administrators like the URA must also operate within this tax environment. Although it has expanded the tax base and improved taxpayer services with digital tax forms and improved taxpayer services, the audit selection process still relies on a rules-based system that is unable to adapt to taxpayers' changing risk profiles in a timely manner (URA, 2022). This is made worse by challenges in data integration, overall institutional capacity, and the absence of data analytics tools, which make it difficult to accurately detect high-risk taxpayers and address noncompliance quickly. Recently, the URA has rolled out initiatives like the Domestic Revenue Mobilization Strategy (DRMS) and launched data visualization and business intelligence dashboards. However, tax audits are still primarily reactive, and enforcement has focused on retrospective audits instead of prioritizing predictive risk assessments (WorldBank, 2021). This is in contrast to the government's broader national agenda, particularly the Third National Development Plan (NDP III), which prioritizes digital transformation and performance-based public service delivery (NPA, 2020) .

The application of predictive analytics to URA's audit operations is also consonant with regional and international trends toward tax administration modernization through the use of data science and artificial intelligence. The African Tax Administration Forum (ATAF) and the IMF have also noted that while the use of predictive analytics has increased across the world, adoption of predictive analytics has been uneven in Sub-Saharan Africa, largely due to low technological and institutional readiness (Aslett, 2024); (ATAF, 2023). The exception is a handful of recent innovations in countries like Senegal and Rwanda that are providing interesting case studies on the role of machine learning in tax audits and improved revenue collection (Czajka, Kondylis, Sarr, & Stein, 2022).

Predictive analytics provides URA with an opportunity to move from labor-intensive, manual, retrospective audits toward more data-driven, high-risk prioritized enforcement strategies. This would improve compliance rates, streamline resource allocation, and support the overall revenue mobilization objectives of the Ugandan government.

1.2 Statement of the Problem

In the drive towards improving operational efficiency, an effective tax authority would need to put effective use of its audit capabilities towards nipping any form of potential or actual cases of non-compliance in the bud by corporate taxpayers in order to enhance overall tax revenue collection (Olaoye et al., 2018). However, in an analysis of the effectiveness of the Uganda Revenue Authority in auditing and enforcing compliance to corporate tax requirements, the effectiveness of the tax institution in terms of tax productivity and effectiveness is brought into question as the situation in the country remains (Kangave et al., 2016) (Mawejje & Munyambonera, 2016).

Research conducted by (Gichuki et al., 2024) on URA domestic revenue generation also established that despite the efforts that have been put into URA tax digitization and automated processes and enforcement, there have been significant leaks of tax revenues especially in the corporate tax arena, a large majority of which can be attributed to tax avoidance in the form of transfer pricing and under-reporting of income. This therefore calls for an increase in effort and investment in improved audit approaches such as through predictive analytics in order to improve the targeting and execution of corporate tax audits and ultimately plug the revenue leakages. The proposed research study will therefore seek to analyze the Uganda Revenue Authority corporate tax audit efficiency in improving domestic revenue mobilization through application of predictive analytics into a predictive model which was trained on past, or historical data for improved audit targeting and efficiency as well as providing an overall added benefit of time and resource saving.

While operational inefficiencies in URA's corporate tax audits are well documented, the *theoretical gap* lies in the limited exploration of how predictive analytics, when situated within Compliance Theory and Information Systems Theory, can systematically improve audit efficiency in developing economies. Most prior studies are concentrated in developed settings with rich datasets and robust IT infrastructure (Baghdasaryan et al., 2022). Little is known about how predictive analytics performs in contexts where data is incomplete, informal sectors dominate, and institutional readiness is limited. This study therefore contributes by applying predictive analytics to the URA context while testing its theoretical alignment with compliance behavior and information systems efficiency models. The originality of this research lies both in contextual adaptation and in demonstrating how predictive modelling can inform audit policy in resource-constrained environments

1.3 Purpose of the study

To evaluate the use of predictive analytics in improving corporate tax audit case selection at the Uganda Revenue Authority by developing and testing machine learning models on historical audit data, with the aim of enhancing the identification of non-compliant taxpayers and providing a decision-support tool for targeted audits.

1.4 Research Objective

The following is the research objective:

This research objective is to Evaluate the Audit Case selection process of the Uganda Revenue Authority to gain insight on how it works, then evaluate the application of predictive analytics with an end goal of providing a fit-for-Purpose Predictive Model that improves the Case selection Process for revenue authorities towards improving domestic revenue mobilization using a real-life case study and selecting the most suitable model, and tuning to suit our study.

i. To critically analyze the current audit case selection mechanisms at URA, identifying their strengths, weaknesses, and efficiency gaps.

ii. To compare URA's audit selection approaches with predictive analytics applications in other tax administrations globally, drawing lessons and best practices.

iii. To design and implement predictive models (Logistic Regression, Random Forest, XGBoost) trained on URA's historical audit data, evaluating their predictive accuracy and efficiency.

iv. To test and validate a conceptual framework grounded in Compliance Theory and Information Systems Theory, demonstrating how predictive analytics mediates between taxpayer characteristics and audit outcomes.

v. To propose a fit-for-purpose predictive model as a decision-support tool for URA and other revenue authorities in developing economies.

1.5 Research Questions

The following are the guiding research questions that the study would aim to address

1. What is the extent of the challenges and constraints the Uganda Revenue Authority experiences in the corporate tax audit selection and the enforcement of compliance?
2. In what ways can predictive analytics be used to improve the Uganda Revenue Authority's tax audit case selection effectiveness and efficiency?
3. What are the most influential risk indicators that can be utilized for Uganda Revenue Authority predictive analytics model to achieve optimum efficiency for tax audit case selection?
4. What are the anticipated gains and practical considerations for the URA with predictive analytics for corporate tax audits?
5. What predictive model can be proposed as a fit-for-purpose decision-support tool for URA and similar revenue authorities?

1.6 Theoretical Frameworks

In this study, we draw on two theoretical frameworks to underpin the analysis of the effect of predictive analytics on tax audit case selection by the URA. These frameworks are the Compliance Theory and Information Systems (IS) Theory, which complement each other and offer a well-rounded perspective on the adoption and outcomes of using predictive analytics in tax audit case selection.

Compliance Theory provides a basis for understanding the behavioral aspects of tax compliance, and is crucial for assessing the impact of predictive analytics on tax audit case selection. This theory focuses on the factors that influence taxpayers' willingness to comply with tax regulations, including economic incentives, the perceived probability of enforcement, perceptions of fairness and equity, institutional trust, and social norms (Mpofu, 2020). Predictive analytics can be seen as a tool that supports the principles of Compliance Theory by enhancing the "enforced compliance" aspect, that is, the use of penalties and audits to deter non-compliance (Kwok & Yip, 2018) (Prichard et al., 2019). By more accurately targeting high-risk taxpayers, predictive models can strengthen audit deterrence and potentially increase trust in the fairness of the tax system by reducing the perception of arbitrary enforcement.

IS Theory, on the other hand, focuses on how digital tools and technologies can be used to enhance decision-making processes and outcomes in organizations, including in tax administration. This theory supports the integration of predictive analytics in tax administration as a digital tool that leverages historical data for decision support (Baghdasaryan et al., 2022). Predictive models can be considered as information systems that process large volumes of data to generate real-time insights and improve resource allocation, audit targeting, and overall efficiency of tax administration (Baghdasaryan et al., 2022).

The use of these two theories in this study will not only enable an evaluation of the effectiveness of predictive analytics as a technological solution for improved tax audit case selection, but will also provide insights into its behavioral and institutional implications. This approach will help to ensure a comprehensive understanding of how predictive analytics can be leveraged to enhance audit efficiency, improve compliance, and support the broader goals of domestic revenue mobilization and digital transformation in Uganda's tax administration.

1.7 Conceptual Framework

The conceptual framework positions independent variables (taxpayer profile, financial indicators, filing behaviour, prior audits, risk segmentation) as inputs. These are processed through the mediating mechanism of predictive analytics (Logistic Regression, Random Forest, XGBoost), generating risk scores and classifications. The outputs are dependent variables—audit selection accuracy, detection of non-compliance, audit cost savings, and revenue recovery.

Grounded in Compliance Theory, the framework assumes that accurate detection strengthens deterrence and voluntary compliance. Grounded in Information Systems Theory, it assumes that data-driven decision support improves organizational efficiency and resource allocation. Together, these theories explain both the behavioural and institutional impact of predictive analytics in audit case selection.

This framework ensures that the study not only evaluates URA's audit efficiency but also contributes to theory by demonstrating how predictive modelling mediates between taxpayer characteristics and audit outcomes."

Diagram

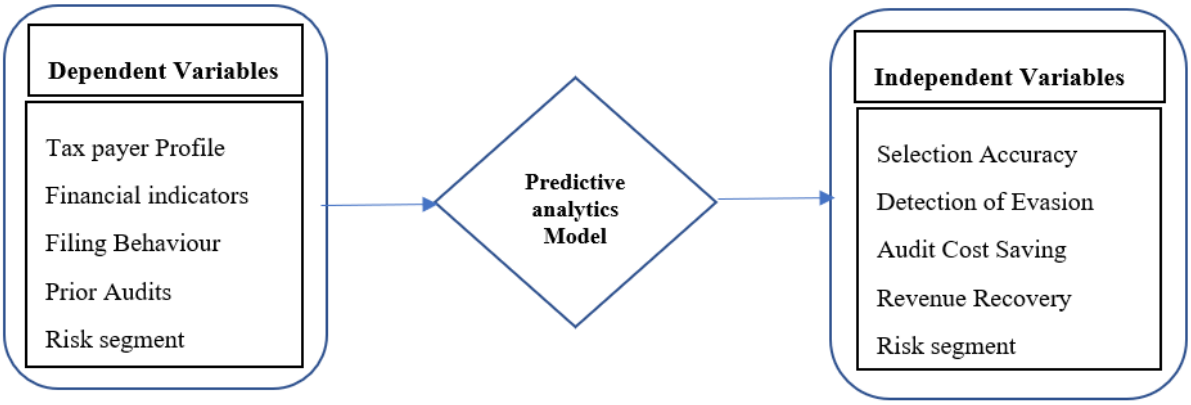


Figure1: A block diagram showing independent variables feeding into a "Predictive Model" box, which then connects to audit outcomes.

1.8 Hypotheses

H1: Improved predictive audit model significantly improves the efficiency and effectiveness of the tax audit processes at the Uganda Revenue Authority.

H2: Improved predictive audit model enables URA to effectively target high risk corporate taxpayers and increase detection of tax avoidance and evasion.

H3: Adoption of the improved predictive audit model results in a measurable decrease in audit costs and an increase in the return on investment from tax audit activities at the Uganda Revenue Authority.

1.9 Scope of the Study

The study was limited to the application of predictive analytics to the tax audit case selection processes of the Uganda Revenue Authority (URA). It was also limited to historical audit and taxpayer data made available by URA for a defined period and within a structured format appropriate for machine learning modeling purposes. It was also focused on audit case selection because of the segment's contribution to tax revenue and the historical challenges faced by the Revenue Authority in more reliably selecting high risk large taxpayers for audits.

Geographically, the study was limited to Uganda and more specifically, URA's operational practices, interventions, and approaches to audit targeting and enforcement. It was limited to data about large taxpayers in Uganda whose audit records were made available and are complete for machine learning modeling purposes. Timewise, the study used audit and taxpayer records that covered periods adequate to train and test the predictive models in this study. The records covered a five-year period starting from the year 2017. Thematically, the study was limited to the following 3 specific topics:

i. Exposures of challenges in current audit selection practices by URA;
ii. The practical feasibility and accuracy of machine learning models in identifying high risk taxpayers;
iii. The consequences of implementing a data-driven audit selection approach for efficiency, compliance enforcement, and institutional capacity;

The study did not attempt to explore or measure the legal, ethical or change management implications of instituting predictive analytics for case selection. These are still subject for future research after learning from the initial take up of new models.

1.10 Justification of the Study

The corporate entities provide the biggest contribution to the tax revenue collected by the Uganda Revenue Authority (Kangave et al., 2016). URA has concentrated on auditing this segment of taxpayers due to their relative importance (Kangave et al., 2016). The modes of enforcement in this sector have been more of historical data review-based analysis and research. While the application of predictive analytics in Tax Audit Case Selection has been explored in some contexts, its potential has not been fully realized in the case of the Uganda Revenue Authority. This study aims to address this gap by investigating the feasibility, benefits, and practical considerations of integrating predictive analytics into the URA's tax audit processes.

Previous research has identified several factors that contribute to the URA's struggles in effectively auditing and enforcing corporate tax compliance including the prevalence of the informal sector, the dominance of cash-based transactions, and the authority's focus on taxing corporate entities at the expense of individual taxpayers (Mawejje & Munyambonera, 2016) (Kangave et al., 2016). According to (Kangave et al., 2016), the URA's enforcement efforts have been skewed towards corporate entities, leading to a disproportionate burden on the formal sector. Moreover, despite the many reforms implemented by the URA over the past decade, the country's tax-to-GDP ratio has remained relatively low, hovering between 12% and 14% (Kangave et al., 2016), with more of technology and scientific approaches to enforcement, the integration of predictive analytics into the URA's tax audit case selection processes has the potential to address these challenges and improve the efficiency and effectiveness of tax revenue mobilization.

Furthermore, the research directly supports national policy objectives outlined in Uganda's Third National Development Plan (NDP III), which emphasizes digital transformation, institutional modernization, and data-driven governance (NPA, 2020). By developing and evaluating a predictive audit model using URA's historical data, the study provides actionable insights for improving tax compliance, audit targeting, and strategic resource deployment.

1.11 Significance of the Study

This study holds practical, institutional and academic significance for enhancing tax audit efficiency in Uganda and contributing to broader efforts in modernizing public revenue systems through data-driven approaches.

At the institutional level, the findings provide actionable insights for the Uganda Revenue Authority (URA) on how predictive analytics can improve the accuracy and effectiveness of audit case selection. By reducing reliance on manual or rule-based selection methods, the URA can allocate its limited audit resources more strategically and detect high-risk taxpayers more reliably—thereby improving revenue recovery and compliance outcomes (Baghdasaryan et al., 2022; Gichuki et al., 2024).

At the policy level, the study supports the implementation of Uganda's Third National Development Plan (NDP III), which prioritizes digital transformation and evidence-based public administration (NPA, 2020). It also aligns with international best practices advocated by the OECD and African Tax Administration Forum (ATAF), which encourage the adoption of machine learning and predictive tools for risk management in tax systems (OECD, 2021; ATAF, 2023).

For the academic and research community, this study contributes to the limited but growing body of literature on the use of predictive analytics in tax compliance within Sub-Saharan Africa. By applying real-world data from a developing country context, the research fills a gap in empirical studies and provides a basis for future investigations into digital innovation, public sector analytics, and tax administration reform (Aslett, et al., 2024) ; (Czajka, 2024).

Finally, the study may serve as a reference point for other government agencies and regulatory bodies exploring the integration of artificial intelligence and data science in compliance monitoring, enforcement, and strategic decision-making.

1.12 Operational definitions of terms

Audit Case Selection: The process used by a tax administration to identify and prioritize taxpayer records for audit based on risk, data analysis or observed anomalies. In this study, it refers to the application of predictive analytics to determine high risk tax cases for audit.

Predictive Analytics: A data-driven approach that employs statistical and machine learning algorithms to analyze historical data and forecast future outcomes. In the context of this study, predictive analytics is used to anticipate which taxpayers are likely to exhibit non-compliant behavior.

Machine Learning Model: A computational technique that uses historical data to recognize patterns and make automated predictions or decisions. This study utilized machine learning models such as Logistic Regression, Random Forest, and XGBoost to predict audit-worthy cases.

Risk Profiling: The analytical classification of taxpayers based on indicators of potential non-compliance. In this study, risk profiling was performed using historical tax behavior and model outputs to categorize taxpayers as low or high risk.

Non-compliance: The failure of a taxpayer to fulfill legal tax obligations, including under-reporting income, non-filing, or deliberate evasion. In this research, non-compliance is operationalized through variables that signal discrepancies or suspicious patterns in taxpayer data.

Return on Audit (RoA): An efficiency metric used to estimate the benefit derived from conducting an audit. While not measured directly in this study, RoA conceptually underpins the rationale for targeting audits more effectively using predictive analytics.

Data-Driven Decision-Making: The process of making informed decisions based on data analysis and empirical evidence. This approach forms the foundation for using machine learning models to inform audit planning in this research.

1.13 Limitations of the Study

This study offers valuable insights into the application of predictive analytics for audit case selection at the Uganda Revenue Authority. However, several limitations affected the depth and scope of the findings:

i. Data Quality and Completeness: The dataset showed a diverse set of taxpayers from SMEs and informal sector entities but suffered from incomplete records with missing values and absent behavioral indicators which may undermine model performance.

ii. Exclusion of External Audit Intelligence: To effectively assess taxpayer information, URA auditors use additional resources besides structured databases such as internet searches and third-party documents along with physical inspections and field intelligence from informers. The dataset for this study lacked external and unstructured insights which restricted the model's ability to fully represent actual risk detection practices.

iii. Absence of Real-Time Deployment: The evaluation process of predictive models involved historical datasets within an offline setting. The actual influence of the models on audit results and taxpayer behavior along with institutional performance remained unquantifiable.

iv. Proxy Measures for Non-Compliance:

The study used audit outcomes and predefined risk flags as proxies for non-compliance. However, undetected non-compliance or cases that were inaccurately classified in the historical data may affect model learning and reliability.

v. Model Scope and Complexity:

Although machine learning models such as Logistic Regression, Random Forest, and XGBoost were used, real-world compliance behavior is often more complex. The study did not incorporate unstructured data (e.g., audit narratives, email records, or qualitative assessments), which could have enriched the model's predictive capacity.

vi. Institutional Implementation Barriers:

The study did not assess URA's internal readiness to integrate predictive analytics into daily audit workflows. Implementation requires not only technical capacity but also organizational change management, policy frameworks, and staff buy-in.

Summary

This chapter introduced the research study by outlining the background, problem statement, purpose, and objectives. It established the significance of enhancing tax audit processes in Uganda through the application of predictive analytics and machine learning models. The chapter contextualized the study within the broader challenges of tax compliance, audit inefficiencies, and the growing need for data-driven decision-making in revenue administration.

A review of the historical, conceptual, and contextual background demonstrated the limitations of traditional audit methods and the potential value of adopting predictive tools for identifying high-risk taxpayers. The theoretical and conceptual frameworks guiding the study—

particularly Information Systems Theory and Compliance Theory—were presented, along with clearly stated research hypotheses.

The scope of the study was clarified, and key operational terms were defined to ensure clarity in interpretation. Finally, the limitations of the study were highlighted, particularly those relating to data quality, exclusion of external audit intelligence, and institutional readiness, which provide a realistic boundary to the study's findings.

The next chapter presents a review of the relevant literature, situating the study within existing academic and applied research on tax audit practices, predictive analytics, and machine learning applications in tax administration.

CHAPTER TWO
LITERATURE REVIEW

Introduction

This chapter aims to provide an overview of literature on tax compliance and research on audit case selection as well as related literature on the use of predictive analytics and machine learning in tax administration. This section, therefore, places this study in the context of previous literature and research in the academic and practice fields, it broadly covers the conceptual issues on taxpayer compliance and enforcement, audit case selection including past and modern approaches to audit case selection and recent data driven research in tax audit. This includes briefly review related theoretical perspectives this study is based on and identifying a gap that the thesis fills. The chapter uses previous studies to put this study in perspective of how Predictive Analytics can be applied to enhance tax audit process of a Revenue Authority in a developing economy.

2.1 Theoretical Review

The research is based on the synthesis of three complementary theoretical perspectives, including: information systems theory, the economic theory of tax compliance, and organizational efficiency theory. The former theory postulates that when organizations and institutions successfully harness information technology and effectively integrate data and analytics capabilities into operational processes and decision-making frameworks, there can be demonstrable improvements to various metrics of organizational effectiveness and efficiency such as cost savings, productivity, quality of work, etc. (Adjei, et al., 2024). In the case of tax authorities, we would therefore expect more granular data (derived from advanced data analytics and predictive models, among other potential sources) on cases to facilitate better decision making on various aspects of tax administration, and in turn, drive more efficient and effective collection of revenues from audit processes.

Second, the classic economic theory of tax compliance, popularly attributed to the seminal paper by Allingham and (Allingham & Sandmo, 1972), which was subsequently built upon and modified by Yitzhaki (1974) and others, is a framework that is used to study the behaviors of taxpayers from the perspective of the risks and returns to being caught for engaging in (typically, underreporting) tax evasion. In other words, tax evasion is rationalized by those who do not pay their taxes as some sort of “high-stakes gamble” in which the risks (being audited and detected by the tax authorities, and then potentially being sanctioned by the relevant state enforcement agencies) and the potential rewards (keeping the rest of the (already paid) taxes, and not having to spend money to hire tax professionals to make these calculations and remittances every year) are constantly being analyzed, both in a rational and psychological manner (Slemrod & Yitzhaki, 2000). The "deterrence model" that is built on these considerations provides a theoretical construct to explain how certain policy levers (probability of an audit, and harshness of potential penalties) could either disincentivize (when increased) or motivate taxpayers to be compliant (when there is lower certainty in these values) (Phillips, n.d.). In its original formulation, this was conceptualized as a "tax compliance cost" to taxpayers in the form of their potential loss of income in the event of being audited, but this model was later refined and expanded by (Slemrod & Yitzhaki, 2000) to consider other important factors (such as whether the penalties are applied on the total unpaid tax vs. the unreported fraction) that influence the nature of the returns to non-compliance. The compliance theory described above thus also establishes a baseline of the incentive framework that tax authorities have over taxpayers – for instance, simply raising the potential certainty of audits, or the salience/reputation of the authority to enforce audits will generally lead to an increase in voluntary compliance (Phillips, n.d.). A specific literature on taxpayer behavior in Uganda has found similar "knowledge" (Kangave et al., 2016); (Mawejje & Munyambonera, 2016); Musimenta, 2020), complexity, and cost-related factors to be the predominant "determinants" of compliance, while a "cumbersome" tax code with additional compliance costs, uncertainty, and non-transparency was associated with non-compliance (Ellawule, et al., 2024). In other words, a high-burden system of taxation is more likely to lead to compliance problems, while increased transparency of processes and potential guidance (perhaps even by improved tax data systems and analytics) may reduce these sources of friction and improve the compliance rate voluntarily.

The third major component of the study's theoretical basis is drawn from the organizational efficiency literature which simply focuses on structuring work (efforts, management decisions) such that it can be performed with least amount of work/friction, and at maximum gain. In the context of tax administration, this means reducing unnecessary workloads (from audits, and across the organization), reducing the scope for high-compliance costs and audit friction on the end of tax payers (for instance, by increasing awareness and ease of navigation, or from reorienting the resources for interventions that can broadly mobilize more revenues with less effort), and increasing the productivity/quality of work across the organization (as relevant and possible). Some earlier research has shown, for instance, the positive impact of efficient data management and leadership decision-making in tax authorities on their performance in terms of revenues collected per effort expended (Zhang, et al., 2024) . In summary, the organizational efficiency theory being postulated by the study suggests that an organization like the Uganda Revenue Authority (URA) can significantly improve its performance (outputs or results) simply by leveraging information (better data and analytics) to "do more with less" in terms of targeting the right taxpayers, with the right set of interventions (services or penalties), and at the right time to nudge them towards being compliant or deter them from acting that way.

2.2 Conceptual Review

Drawing on the above theories, the study's conceptual framework links **predictive analytics integration** with improvements in the tax audit case selection process. The key concept is that by using predictive models to analyze taxpayer data (e.g. past returns, third-party information, industry risk factors), the URA can identify high-risk taxpayers more accurately and automate the selection of audit targets. This data-driven audit selection is expected to increase the **"hit rate"** of audits (uncovering more underreporting or non-compliance cases) while reducing wasted effort on low-risk cases (Nigar, et al., 2016). *Figure 1* illustrates how inputs like taxpayer profiles and compliance history feed into a predictive risk scoring model, which then informs audit decisions. The outcome is a feedback loop where improved audit results further refine the model, consistent with both information systems success theory and adaptive compliance risk management practices. Overall, the conceptual review establishes that the intersection of advanced analytics with tax administration processes can create a more **intelligent, responsive, and efficient** audit system, guided by theoretical principles of deterrence and organizational optimization.

2.3 Empirical review

A growing body of empirical literature examines the application of predictive analytics and data science techniques in tax administration – particularly for enhancing audit effectiveness. Studies in various jurisdictions have demonstrated significant benefits of risk-based audit selection and predictive modelling. For instance, (Nigar, et al., 2016), use an agent-based model to show that deploying predictive analytics to target audits yields a significant increase in revenue collection compared to random audit strategies. Similarly, an international study by (Eberhartinger, et al., 2021) found that tax authorities using data-driven risk profiling (including predictive modelling and better intelligence functions) achieved lower rates of corporate tax avoidance and improved overall enforcement efficiency. These findings confirm that predictive analytics can enable revenue bodies to better identify high-risk taxpayers and concentrate enforcement resources where they matter most, thereby increasing the return on audits.

Research has also explored what indicators and data are most useful in predictive audit models. For example, a study of Malaysia's tax authority by (Hamzah, et al., 2019) investigated audit enforcement and found that certain firm characteristics (such as consistent claims of specific incentives) were linked to higher audit rates, whereas some common tax avoidance red flags (e.g. abnormally low effective tax rates or large book-tax gaps) were not always being used in practice. This suggests that while predictive analytics models *can* incorporate such risk indicators, actual usage depends on data availability and the tax authority's focus. Another relevant paper, "External Auditor Responses to Tax Risk" (Abernathy, et al., 2019), provides insight into how tax risk is assessed and addressed externally: it found that external auditors treat high tax risk as a heightening of overall audit risk, devoting extra audit hours when a client's tax positions are aggressive. This behavior underscores the importance of reliably measuring tax compliance risk – something predictive analytics can assist with – as both tax authorities and external auditors will react to perceived risk levels.

Empirical studies have also examined corporate behavior and tax administration challenges that contextualize the need for analytics. For instance, an Indonesian case study by (Erasashanti, et al., 2022) on corporate income tax audit preparedness revealed that even companies striving for compliance often face "tax surprises" during audits due to overlooked issues. The study noted that the company had not conducted an internal tax review and was less compliant in some areas than assumed, leading to significant adjustments when audited. This illustrates the challenge for tax authorities: many compliance problems are not obvious upfront, and a more proactive, data-driven approach (like running predictive risk assessments on filings) could flag such cases earlier. In Africa, empirical works have highlighted broader factors influencing compliance. (Kouame & Goyette, 2018), in a world bank Policy research paper, analyzing firm data from African and Latin American countries, found that better institutional quality and stronger tax-agency monitoring capacity correlate with lower tax evasion by firms. In other words, when tax administrations have tools to effectively monitor and target non-compliance (which predictive analytics could enhance), firms are less likely to evade. Conversely, studies in Nigeria and Uganda indicate that burdensome tax administration and multiple audits can hinder business performance or compliance among small and medium enterprises. (Ellawule, et al., 2024) observed that introducing systematic tax audits improved tax compliance and revenue in Nigerian states (tax audit was associated with higher tax productivity). On the other hand, (Kintu, et al., 2019) found in Uganda's SME sector that while a fair tax rate and access to finance were significant for business performance, the direct effect of general tax administration on firm performance was statistically insignificant, suggesting that simply having many audits or complex procedures may not yield benefits unless those audits are well-targeted and taxpayer services are strong. These mixed outcomes reinforce the idea that smarter, not just more, enforcement is needed. Thus, the consensus in empirical literature is that targeted, data-driven audit strategies can enhance compliance outcomes and revenue, especially if implemented in a way that considers the local context and constraints. Despite these advances, the existing literature has limited coverage of applying predictive analytics in developing country contexts such as Uganda are very limited. Many studies demonstrating the success of predictive models in tax auditing have been conducted in environments with rich datasets, robust IT infrastructure, and ample skilled personnel (Owens, et al., 2025).

For example, (Baghdasaryan et al., 2022) developed a machine-learning fraud detection model on a comprehensive dataset of corporate taxpayers and achieved roughly 85% accuracy in classifying likely evaders, showcasing the power of AI in flagging high-risk cases. However, such models often assume the presence of extensive historical data (including past audits and known fraud cases) and significant computing resources. In Uganda, the URA faces practical challenges like incomplete data records, fewer data points on prior audits, constrained budgets, and limited technical expertise in advanced analytics. Therefore, a gap remains in understanding how predictive analytics can be tailored to a resource-constrained tax administration. This study aims to fill that gap by developing a predictive audit selection model that accommodates messy or missing data and prioritizes cost-effective audits, aligning with the URA's operational reality. The model will seek not only to maximize potential revenue recovery but also to be feasible given the URA's data limitations and need for simplicity in implementation. By addressing these local challenges, the research will extend the empirical literature to demonstrate how predictive analytics can be leveraged in a developing country setting to improve corporate tax audit outcomes.

2.4 Literature Synthesis

To briefly conclude this part, both the theoretical and the empirical literature converges on the promise that predictive analytics could significantly enhance tax audit outcomes and compliance management. On the theoretical side, the general idea is that better information systems will lead to improved efficiency and decision-making in tax administration (Adjei, et al., 2024). When it comes to tax compliance more specifically, a risk-based approach to deterrence theory suggests that data-informed strategies could enhance the perceived detection of tax evasion, which in turn, should increase voluntary compliance (Phillips, 2023). The empirical evidence on these questions is quite strong. For example, data-driven methods for selecting audits should make them more effectively targeted, and therefore more efficient in terms of resource allocation. Several studies have shown that the use of predictive risk models by tax authorities can result in a more accurate identification of high-risk taxpayers, and more audit "yield" (additional taxes assessed/recovered) per audit performed (Nigar, et al., 2016). This can lead to a direct increase in short-term revenues, and potentially to a "deterrent" effect on the entire taxpayer population in the medium- and long-run (taxpayers are less likely to try to game the system, if they know that the tax authority has strong data on hand, and can catch sophisticated evasion attempts). In addition, using analytics to make the process fairer and more transparent is another potential way in which it could help: taxpayers seeing audits being triggered not by arbitrary means, but by consistent, risk-based criteria, can reinforce the legitimacy of the tax administration (which has been shown to be a critical factor for voluntary compliance (Musimenta et al., 2017). There is also some evidence that the use of data and technology to streamline or simplify compliance – for example, pre-filling of returns, or proactive notifications of discrepancies, etc. – can help reduce both compliance costs and the associated frustration, and thus improve honest reporting (Ellawule, et al., 2024).

At the same time, both the literature on the topic, and our interviews highlighted a number of serious challenges that would have to be addressed if such innovation was to be successful. In particular, data quality and availability come up repeatedly in the literature: analytics and predictive models can only be as good as the data on which they are based, yet in the developing country context, tax data could be incomplete, outdated, not standardized, or trapped in different siloed IT systems (Thapa, 2019). Capacity constraints on the analytical side is another frequently discussed issue – tax agencies would need skilled data analysts and IT personnel to work on developing, interpreting, and maintaining predictive models, and the necessary institutional investment would not be trivial. A third challenge that has received growing attention in recent years relates to governance and ethical issues around the use of big data and AI in the public sector. This ranges from traditional concerns of privacy and potential bias in algorithms, to more general questions about appropriate checks, balances, and rule-making in the public sector around the use of data and data-driven decisions to avoid potential abuse. As noted in the literature, for example by (Eilu et al., 2021), the integration of predictive analytics in the public sector should be accompanied by establishing and enforcing strong governance rules on how data can and should be used, and how accountability for decisions is maintained. This is of particular concern for tax administration, but in practice, this means things like having rules and guidelines on how and when to use models, how to check for and correct potential biases in them (e.g., that a given model does not unfairly flag taxpayers from a certain region or industry more often than others), and how to leave room for human discretion in the decision chain for audits, which might need to review and explain cases flagged by the model. Fortunately, other recent work does provide guidance on these issues – for example, the literature on AI adoption in government, including for finance and tax administration, recommends a number of similar and practical measures for tax agencies – for example, to work on transparency and appeal mechanisms for algorithmic decisions where possible, to improve overall data governance, and to build trust in analytics within the organization in a step-by-step way (Thapa, 2019).

Such measures can help to address some of the risks of a "black box" approach, and to align the innovation with the broader principles of good tax administration. Overall, then, the literature suggests that while predictive analytics could be a powerful tool for the tax administration to improve audits and compliance, a number of factors should be kept in mind if such a system were to be implemented in practice. Successful examples from other countries provide a benchmark as to what could be achieved (higher revenues, more efficient audits, etc. (Nigar, et al., 2016), but also point to potential pitfalls (such as over-dependence on poor data, or ignoring the human and organizational factors involved). This provided the background for the present study, and can thus be seen as its foundation.

2.5 Summary of the Literature Review

Extensive Literature review was conducted to uncover studies similar to the proposed one. This provided clues about what has already been done in this area of study to help in siting a niche for this research and also to help provide a better background for this research topic. The review provided insight of the critical and pertinent issues about this area of study. As a result, the review presented a scope of the relevance and the major sources and problems of the proposed research. The study has therefore done a comparison of the areas covered by the past studies to what was done in the study. Literature search was done by employing different research databases to capture literature pieces in this area of study. Various search engines such as Google Scholar, EBSCOhost, JSTOR, Springer Link and Proquest. This search provided various studies for the review with keywords such as predictive analysis in Tax Administration. The past studies helped in bringing a direction for the proposed study. The literature gaps are the finding that was not known at the time of a study. literature gaps on a study research is due to the fact that in the dynamic world, new facts are continually being revealed. the literature gaps on the various studies has made it critical for the proposed study to consider the Uganda

Revenue Authority and building a Predictive model to Automate the audit selection of potential companies, individuals or organizations, since the past studies did not consider the literature gaps that the review has brought out.

According to existing literature, the use of predictive analysis in tax administration have an advantage. This is because the empirical and analytical techniques have been to help the tax authority to do audits. In the past tax administrations had difficulties in pinpointing which tax payers were high risk, as a result resources were in short supply to deal with the situation and thus audits were not optimal and as a result the agencies needed to figure out the methods to increase the overall effectiveness of the revenue collection that would help them address the problems in detection of the high-risk taxpayers and to improve their efficiency. There are additional implications with a varied set of challenges, impediments and barriers such as data quality, organisational capacity and development and appropriate governance framework that need to be considered and catered for for predictive analysis in tax administration to be successful.

CHAPTER THREE
RESEARCH METHODOLOGY

3.0 Overview

This methodology serves a dual purpose: (1) to evaluate URA's corporate tax audit efficiency through predictive analytics, and (2) to empirically test the proposed conceptual framework grounded in Compliance Theory and Information Systems Theory. The design therefore integrates model-building (Logistic Regression, Random Forest, XGBoost) with hypothesis testing to ensure alignment between theory and applied efficiency evaluation.

In this chapter we provide the details on the approach and methods used in this work. This includes the research design, population, sample size, sampling procedures, data sources, data collection procedures, data pre-processing, modelling and analysis techniques. We also cover the ethical issues that were considered and taken into account in the study. The work reported in this book used a mixed methods design with quantitative data collection and analysis being the most prominent part of the research effort. Some qualitative data and information was used to provide contextual information on the model development and institutional relevance.

3.1 Research Design

The study employed a mixed-methods research design, integrating quantitative techniques for model building and validation, and qualitative approaches for understanding audit selection challenges and parameter selection. According to (Creswell & Plano , 2018), this design is suitable when one method alone cannot fully address complex research problems. In predictive analytics for public sector audit selection, quantitative models must be complemented by institutional context. (Khadka, 2021).

The quantitative dominance in this research is consistent with best practices in predictive analytics and machine learning applications, where model evaluation relies heavily on statistical performance metrics (Wang & Alexander, 2021). However, the integration of qualitative interviews helped address the practical and institutional dimensions of model adoption, which pure quantitative methods may overlook (Bala, et al., 2013). This dual approach ensured both technical robustness and contextual fit, consistent with the pragmatist paradigm common in data science research for public governance (OECD, 2021).

This design was chosen to ensure that the predictive model would be both technically robust and contextually grounded, allowing the findings to inform not only algorithmic development but also Remain consistent with the pragmatist paradigm common in data science research for public governance (OECD, 2021) in organizations like the Uganda Revenue Authority.

3.2 Target Population and Sample Size

The target population comprised staff from the Uganda Revenue Authority (URA) involved in tax audit selection and execution. These included officers from Domestic Taxes Department, Central Operations Office (COO), Tax Investigations, and the Data Engineering section of Information technology Department. The total target group consisted of 57 officers, stratified across operational and managerial levels.

A sample of 40 participants was selected for survey and interview activities. This represents approximately 70% of the total audit-related personnel, aligning with sampling adequacy for exploratory research (Taherdoost, 2016). High-level officials such as Commissioners were excluded from the final interviews due to scheduling constraints, though mid-level and senior technical staff provided a representative institutional view.

For quantitative modelling, the dataset obtained from the URA's electronic tax systems and audit records. This dataset includes taxpayer information, declared tax filings, and audit outcomes over multiple years. Specifically, it contains variables such as taxpayer identifiers, firm characteristics (sector, size, location), financial declarations (income, expenses, etc.), third-party matching data (where available), and the results of audits (e.g., whether additional tax was assessed, amount of any adjustment, or a binary risk flag indicating non-compliance). The data were extracted with permission from URA's compliance database via secure queries. Using administrative tax data ensured high relevance and accuracy for modelling real-world audit selection (Scarcella, 2019). Moreover, this approach enhances model reliability while avoiding sampling bias (Zhou, et al., 2023).

3.3 Sampling Procedures

A combination of purposive and random sampling techniques was employed to select participants from the target population of URA staff. Given the diverse roles within the population, a stratified approach was used to ensure representation from each category of staff while also focusing on those with relevant expertise:

Purposive sampling was used for qualitative interviews to target key informants with deep knowledge of audit processes (Palinkas et al., 2015). Key informants were intentionally selected based on their positions and expertise in the tax audit process. This included experienced tax auditors, audit supervisors, data analysts, and mid-level managers who were likely to have deep insights into the current case selection procedures and the potential adoption of predictive analytics.

Stratified random sampling was used for survey distribution to ensure proportional representation across the relevant job roles. Each member of the target population within a given stratum had an equal chance of being included in the sample, minimizing selection bias. For example, about half of the Tax Auditors and Tax Investigations officers were randomly chosen to receive the questionnaire, and roughly similar proportions were selected from the LTO and IT staff strata.

For the tax audit dataset, a census of available records was employed due to the predictive modeling objective, which benefits from maximum data volume (Chen et al., 2021). Since the intention was to build the most accurate predictive model possible, the researcher aimed to leverage all available data rather than a sample. Initially, a simple random sampling of audit records was considered (as a thought experiment to ensure unbiased case selection if needed). In practice, however, the researcher obtained the entire set of relevant audit records for the five-year period and an overwhelming dataset on tax Payer detailed filing transactions in the same time period.

By using purposive sampling for interviews and stratified random sampling for survey respondents *(and a full-population approach for the audit data)*, the study combined techniques to suit each part of the research. This multi-pronged sampling procedure ensured both depth *(through knowledgeable key informants)* and breadth *(through a representative survey of staff and comprehensive data coverage)* in addressing the research questions

3.4 Data Sources

The study utilized both **primary and secondary data sources**, encompassing quantitative and qualitative information:

Primary data included questionnaire survey and semi-structured interviews. It inquired into areas such as the current corporate tax audit selection process, the availability and quality of data for audits, and staff perceptions of the potential benefits and challenges of implementing predictive analytics in audit case selection. The **questionnaire** was a structured instrument designed to capture quantitative ratings and factual information about current audit selection practices. It included mostly closed-ended questions (e.g. Likert-scale items asking auditors to rate the importance of various risk factors in case selection, yes/no or multiple-choice questions about whether certain analytical tools are used, etc.), as well as a few open-ended prompts. This survey provided supplementary quantitative data from the human perspective – for instance, the frequency of use of data analytics in auditors' day-to-day work, or perceived effectiveness of the current risk-based selection system. The questionnaire was distributed internally via an online form to 40 staff. we received thirty complete responses, which were used for descriptive analysis. These survey responses provided quantifiable measures of attitudes and readiness for data-driven auditing, which could be analyzed statistical.

Secondary Data sources- A major data source for the study was internal URA datasets containing historical records of tax audits and taxpayer risk profiles. These secondary data were extracted from URA's information systems with permission. One dataset consisted of taxpayer risk scores/profiles – essentially the outputs of URA's existing risk assessment system, which assigns risk ratings or flags to taxpayers based on various compliance indicators. Another dataset consisted of tax audit case records over the period 2017–2021, including details for each audit such as the taxpayer identifier, taxpayer characteristics *(e.g., industry, size, region),* the reason for selection or risk flag, and the outcomes of the audit. Audit outcomes data included whether non-compliance was detected, the amount of additional tax assessed or recovered, and the status of case resolution. These two datasets (risk profiles and audit outcomes) were linked via common taxpayer identifiers or audit case IDs, resulting in a combined dataset used for the predictive modelling. The secondary data provided the factual, objective evidence needed to train and test machine learning models, and to analyze patterns in past audit effectiveness.

Supplementary Sources: In addition to the above, the researcher also reviewed relevant URA documents and reports as needed to inform the study. These included internal reports on audit performance, URA's audit procedure manuals or guidelines, and strategic plans that mention the use of data analytics. While not formal data sources for analysis, these documents provided background context and helped in interpreting results *(for instance, understanding how audit cases were traditionally selected and any existing frameworks for risk profiling).* This document review served as a triangulation tool for validating information gathered from interviews and ensuring the researcher's understanding of URA's operational context was accurate.

By drawing on multiple data sources – surveys, interviews, and organizational data – the study fulfilled the mixed-methods design, using each source to complement and cross-verify insights from others. The quantitative datasets addressed the "what" and "how much" questions (e.g., which audits succeed and how often), while the qualitative inputs helped explain the "why" and "how" behind those patterns.

3.5 Data Collection Methods

The following data collection methods were used:

Questionnaire: An audit staff questionnaire was given to the sampled URA audit staff in the sample (see Section 3.3 for sampling details). The questionnaire was developed in alignment with the literature and consultations. It had items on current audit selection approaches, risk indicators, technology use, and perceived challenges. These included Likert-scale questions on audit process quality, data availability, and analytics readiness (Naftal, et al., 2019).The instrument was reviewed by two tax auditors with at least 3 years' experience for face validity and clarity, and then pilot-tested with a few staff (3 people) to tweak wording. The final questionnaire was distributed using an online survey platform (an anonymous link emailed to participants), which allowed maximum convenience and response rate. This was along with a cover letter assuring confidentiality and seeking informed consent. Participants were given one week to respond, with a reminder sent mid-way. To elicit honest responses, the survey was anonymous and results only reported in aggregate. A total of 30 responses were collected. as quantitative data for analysis. Questionnaire data were automatically captured in spreadsheet form for analysis. This provided a standardized assessment of staff perspectives that could be analyzed quantitatively, and also used to triangulate interview findings (e.g. if survey results showed high support for using analytics, did interviews echo the same sentiment?). Basic psychometrics such as Cronbach's alpha were calculated on grouped Likert items to ensure reliability of any multi-item scales, ensuring the survey instrument met acceptable standards for internal consistency (Taber, 2018).

Guided Interviews: The researcher conducted semi-structured interviews with selected participants, using an interview guide. Interview appointments were scheduled in advance, and interviews were conducted one-on-one (primarily in person at URA offices, though in a few cases phone or video calls had to be used due to logistical constraints). Informed consent was obtained prior to each interview, and the interviewee was reminded that their responses would be kept confidential and used for research purposes only. Each interview lasted approximately 30–60 minutes, and allowed for probing of systemic issues and risk criteria used for audit selection (World Bank, 2022). The Interview Guide included open-ended questions and probes around key themes such as: current corporate tax audit processes and pain points; perceptions and attitudes toward the adoption of predictive analytics in audits; perceived benefits of predictive modelling (e.g., improved compliance, resource efficiency); perceived challenges or barriers (e.g., skill gaps, data quality issues, resistance to change); and suggestions for successful implementation (Hagopian, et al., 2009). The interviewer allowed for flexibility and follow-up questions, letting respondents elaborate or introduce related topics, to ensure a natural flow of conversation. This resulted in a set of rich qualitative data that reflected the experiences and perspectives of URA audit practitioners.

Extraction of URA Data: For the internal audit and risk profile datasets, data collection took the form of data extraction from URA's information systems. The researcher coordinated with the URA research Laboratory operations Team, to retrieve the required data fields for the specified period *(2017–2021).* A formal data request was made, outlining the variables needed and the purpose of the research. After clearance by URA's data governance official, the researcher swore the Oath Of Secrecy, with the Commissioner of Oaths. The researcher was provided with exported data in electronic form *(CSV files)* containing the anonymized audit records and taxpayer risk information. The tax audit dataset included fields such as audit case ID, taxpayer identifier (which was later pseudonymized to ensure privacy), taxpayer sector/industry, firm size category, risk score or risk flag at selection, type of audit *(desk or field audit)*, audit findings, amounts of additional tax assessed, and the outcome *(e.g., tax recovered or case under appeal)*. The risk profile dataset included fields like the taxpayer's compliance history indicators, risk rating scores, or risk level categories assigned by URA's existing risk engine for each tax period. These two datasets were merged on common keys by the researcher after extraction. Prior to analysis, the data were stored securely on a password-protected computer accessible only with in the URA research Laboratory, to the researcher.

All data collection activities were completed in accordance with the approved research protocol and with minimal disruption to URA's routine work. By the end of the data collection phase, the researcher had assembled: a CSV file of Responses (*quantitative survey data),* a collection of interview transcripts/notes *(qualitative data)*, and a comprehensive combined dataset of past audits and risk indicators *(secondary quantitative data).* These were then prepared for analysis as described in the following sections.

All instruments were pre-tested for content validity, and informed consent was obtained from participants.

3.6 Data Preprocessing and Modelling

Before the data is used to build the predictive models, it needs to be pre-processed and transformed into an appropriate data structure. Raw tax audit data, as collected from operational tax systems may have anomalies that need correction or transformation due to the way the data is captured in the native system (Aslett, et al., 2024) In this case, the researcher followed a sequential data pre-processing workflow as outlined above, and then machine learning algorithms (Random Forest and XGBoost in this case) to build and validate the predictive models. The specific steps taken are summarised below:

Model selection rationale

For this study, we used a number of classification models, such as Logistic Regression, Random Forest, and XGBoost to effectively predict high risk audit cases. XGBoost (Extreme Gradient Boosting) (Chen & Guestrin, 2016) was one of the best performing models, selected for this purpose. It is an algorithm that was specifically designed to have high performance and speed on structured datasets with heterogeneous features. XGBoost represents a scalable and high-efficiency solution for gradient boosting. It is an algorithm that was introduced by (Chen & Guestrin, 2016), with some new functions like regularization to prevent overfitting, and system optimization, that can do parallel processing and has sparsity-aware algorithms for learning. It was also designed to help improve upon issues from traditional boosting algorithms and be used in larger and more real world scale datasets (Chen & Guestrin, 2016). For the URA audit dataset with its mixture of categorical and numerical variables, potential class imbalance, and interpretability needs, it seemed to be a very relevant algorithm for the case. It has also seen consistent success for such predictive tasks, with strong performance in predicting outcomes where complex patterns and feature relationships play an important role in the classification of data, such as in fraud detection and risk profiling. In this study, it offered not only high classification accuracy, but also valuable feature importance scores to help with further analysis.

Data cleaning: This data cleaning process was performed inline with the need for data quality assurance that is well captured by the IMF principles (Aslett, et al., 2024) since the performance and credibility of our predictive models will only be as good as the input data. Missing values, duplicates, and outliers were addressed with industry best practices in modelling public data (Zhang, et al., 2024). The raw audit data was first screened for errors, inconsistencies, or missing values, including removing duplicates, imputing or dropping missing data, and correcting anomalies. Missing critical data, such as a case with no recorded outcome, or important taxpayer information, were either filled in or dropped, depending on frequency. For instance, if a taxpayer's industry code was missing but we had it in registration data, we could fill it; if an audit outcome was missing for few cases (not systematic), it was better to drop those few cases to avoid ambiguity. We did outlier analysis on continuous variables like declared turnover, audit adjustments – extreme values were investigated to rule out if they were genuine or errors, case in point, legitimately very large companies. The dataset was generally fairly complete.

Data Integration and Transformation: Since this study used more than one data source (audit outcome data and risk scores), we needed to combine these into a single analytic dataset. The researcher then merged the risk profile data into each corresponding audit record using the taxpayer unique identifiers. The researcher verified the merges to ensure that each audit case had the corresponding risk indicators attached and it matched the right taxpayer TIN. Any cases that could not be matched for instance, if an audit case did not have a risk score in the given data, were highlighted and investigated; if they were not many they were dropped, to avoid issues. These transforms were to put the data into a shape that would be optimally used by the machine learning algorithms.

Normalization and Encoding: To ready the data for the modelling, numerical variables were normalized or scaled if needed. All features were then scaled to a standard range (0–1) to prevent features with larger numbers (e.g., dollar values) from dominating in the analysis algorithms. Categorical variables like industry sector were one-hot encoded into binary dummy variables (one column per possible value), to make it palatable for the algorithm without losing this feature. For example, if Sector" was Manufacturing, Services, Trade, etc, those became separate binary features, (Sector_Manufacturing = 1 if true, else 0, Sector_Service = 1 if applicable, etc.). A special Other "Other" category was also created to capture any categorical instances that were very rare or missing, to prevent information loss. By end of this step, the dataset had been transformed into a fully numeric feature matrix, ready for ingestion by the model building algorithms. This stage also corresponds to creating a "compilation of tax audit data" through normalizing and condensing the features together as described in (Aslett, et al., 2024), which advocates for the need to integrate and pre-process tax records in this context, before subjecting to advanced analytics.

Training/Test Split: We randomly split the data into two sets for model training and evaluation. Approximately 70% of the audit cases were selected to the training set (to build and tune the model), and the remaining 30% was set aside as a hold-out test set for final evaluation. The split was stratified by the outcome variable so that compliant vs non-compliant cases are represented in both sets in the same proportions as in the original data. This is because in a typical data set of this type we have class imbalance as only a minority of audits have significant findings. Stratifying the split ensures that rare "non-compliant" cases are included in the training set and test set. In addition to a simple hold-out split, we used k-fold cross-validation on the training set for model development to further mitigate the risk of overfitting – the training data was partitioned into 5 folds, and the model training was repeated 5 times each time using 4 folds for training and 1 for validation, cycling through all folds (James et al., 2021). Cross-validation procedure provides a reliable estimate of model performance, and can also be used for hyperparameter tuning of the model.

Model Selection and Building: The predictive modeling task in this study is a classification problem – flagging which taxpayers are likely to have a non-compliance finding as a result of an audit, also known as a "positive audit". The target variable is defined as a binary outcome: 1 if the audit had a finding of material tax non-compliance assessed (considered to be a high-risk flag) and 0 if the audit found no significant issue. We experimented with several different modeling approaches that are commonly used in the area of fraud detection and risk scoring. These included: Logistic Regression as a simple interpretable baseline model, Decision Trees and Random Forest that can capture nonlinear patterns and variable interactions in an ensemble way, and XGBoost Gradient Boosting Machine, an advanced high-performance model for structured data. We first ran a logistic regression model, this was not only to identify which predictors are statistically significant but also to have an interpretable sense of the odds ratios of the identified risk factors. More complex models such as random forest and gradient boosting were then trained, as these models are likely to have better predictive accuracy, being able to learn complex relationships. To account for the class imbalance that was expected due to the very nature of the problem that we have far fewer positive audit outcomes than negative, we ensured that the model was well-calibrated: by using class weights in the model algorithms or upsampling the minority class in training using the Synthetic Minority Over-sampling Technique SMOTE (Almeida, et al., 2024).Though a relatively old technique, the underlying principle of SMOTE is still considered a best practice. The performance metric used for model optimization was F1-score on the validation folds (the harmonic mean of precision and recall). This is because F1 is suitable for imbalanced classification problems. F1 score indicates the balance between recall (catching non-compliant cases) and precision (limiting false alarms). We also monitored AUC (Area Under the ROC Curve) as a more general measure of discrimination. The final model that we selected was the one that had the best performance on cross-validated data.

Model Evaluation: The final chosen model tuned gradient boosted model was applied to the hold-out test set which was not used during model building in any way. This provided an unbiased estimate of how it would perform on new audit cases. We computed the following key evaluation metrics: Accuracy, Precision, Recall (Sensitivity), F1-score, and ROC AUC. We also calculated the false positive rate, which is the proportion of the compliant cases that were falsely identified as being at risk, because from the perspective of a tax audit we don't want many false positives, that would waste our time and effort and burden honest taxpayers. The results on the test set showed how well the model could be expected to identify high-risk taxpayers. We also plotted a confusion matrix to visualize the model's classification results. Because our goal is to maximize audit yield, we put more emphasis on precision – we want to make sure that the audits we select have a good chance of being successful. At the same time, we still wanted to have a certain level of recall, because otherwise we risk missing a large number of evaders.

Feature Importance and Validation: One benefit of tree-based models like Random Forest or boosted trees is that they provide measures of feature importance. We extracted the top predictive features from the final model to understand which factors drive risk. This was an important step in validating the model against expert knowledge and in informing policy. For example, the model might indicate that "under-reporting of sales relative to industry averages" and "past audit history" are the strongest predictors of non-compliance. We cross-checked these findings with the literature and with URA experts (through the qualitative phase) to see if they make intuitive sense and if any surprises emerge. We found, for instance, that the presence of prior compliance infractions was a very influential feature – a result consistent with other tax compliance research (Baghdasaryan et al., 2022). On the other hand, some traditionally assumed risk factors like a simple turnover size threshold might not have appeared as important once other variables were in the model. These insights were documented and later discussed in interviews to get practitioners' feedback (forming a bridge between quant and qual analysis).

The above-mentioned pre-processing and model building stages were set up to create an audit selection model which is reliable and has the capacity to generalize well to new and unseen data. The process adheres to established data science methodology (James, G., , et al., 2021) and addresses common challenges like data incompatibility and overfitting, that arise in applying machine learning to tax data (Chan *et al.*, 2022). The outcome of this phase is a trained model capable of assigning a risk score or probability of non-compliance to each taxpayer, which will be used in subsequent analysis and interpretation. This model, however, is not the end point – it is a tool whose role and effectiveness will be further examined through qualitative findings and validation with stakeholders.

3.7 Data Analysis Techniques

This section outlines the data analysis methods used in this study, encompassing quantitative analysis of survey responses, qualitative analysis of guided interviews, and triangulation to synthesize insights across data types.

Qualitative Data Analysis

Quantitative data collected from 30 URA audit staffs were extracted from the online platform and imported to Microsoft excel. Analysis was conducted with Python (pandas, matplotlib, seaborn), and STATA. The following were the statistical tests performed:

- **Descriptive Statistics:** Frequencies, means, and standard deviations were calculated to summarize respondent characteristics and key variables of interest, such as attitudes towards analytics, current audit selection practices, and readiness for predictive modeling (Creswell & Plano , 2018).
- **Reliability Testing:** Cronbach's alpha was calculated for grouped Likert-scale items to measure the internal consistency of multi-item scales. An alpha value above 0.7 was considered acceptable according to (Taber, 2018).
- **Correlation Analysis:** The Pearson correlation coefficient was used to identify the relationships between variables (for instance, between their analytics readiness and satisfaction with current audit process). This allowed for an assessment of whether the perceived benefits of analytics aligned with their attitude towards implementation (Heale & Twycross, 2015).

• **Visualizations:** Pair plots and bar charts were used to visually represent relationships and distributions of variables, such as audit effectiveness indicators and staff perceptions. Visual summaries helped communicate complex results more effectively (Raschka & Mirjalili, 2019). This analysis provided the necessary evidence on staff perceptions of audit quality, data use, and the feasibility of adopting predictive audit techniques.

3.7.1 Qualitative Data Analysis

Qualitative data consisted of 12 semi-structured interviews with URA audit staff and supervisors. The thematic analysis approach was used, and data were familiarized by reading the transcripts several times to become immersed in the data and to get a sense of the preliminary ideas (Nowell, , et al., 2017) Then, initial codes were inductively developed by highlighting and labeling recurring patterns, themes, and concepts, including "manual audit burden," "risk criteria," "data quality issues," and "support for analytics" among others. The codes were then organized and reviewed into broader themes that captured the main issues, which currently influence audit case selection – which are audit inefficiencies, informality, and ad-hoc data use and level of readiness for analytics. To ensure trustworthiness, credibility and depth, the themes were checked against all interview transcripts, for comparison and refinement to validate insights for consistency and representativeness (Braun & Clarke, 2021) .The analysis allowed an in-depth exploration of the challenges with current practices and staff perceptions and expectations for as well as fears about predictive model implementation.

3.7.2 Triangulation of Findings

Triangulation was used to draw together results from the survey and the interview responses together with the findings form the Predictive analytics model. The objective was to cross-check, support, or contrast the insights from the different methods and data sources. For example: The survey indicated that respondents had a strong positive perception about the use of predictive analytics (70% of the participants had rated this factor favorably), and this was largely confirmed by the interviews where most participants also indicated that the use of analytics would improve targeting and efficient use of resources. The interview themes around the reliance on informal intelligence and political tips were not captured by the survey, which could be a limitation in quantitative survey instruments. The correlation in the survey data between the perceived usefulness of analytics and readiness for adoption was confirmed by the interview narratives around the lack of skills as the main barrier. In this way, triangulation increased the validity of the results and provided a more comprehensive understanding of the current status quo and opportunities for improving audit case selection at URA (Fetters,, et al., 2013).

3.8 Validity and Reliability

Ensuring the validity and reliability of the results was critical to providing confidence in this study. Both the quantitative and qualitative methods as well as the modeling components were evaluated with reference to both criteria.

Instrument Validity and Reliability

The survey and interview instruments were reviewed by content experts and were found to align with the study's objectives. This helped establish content validity. Pilot testing helped to identify ambiguities or irrelevant items. Internal consistency reliability was also checked using Cronbach's alpha, and the results were >0.7 on the key Likert scale items, which is the recommended threshold for reliability (Taber, 2018).The researcher also ensured to maintain Consistency in the qualitative data collection, through the use of a semi-structured interview guide and standardized probing questions. There were few responses that required eliciting more information from some respondents to confirm the accuracy of the interpretation (Birt, et al., 2016).

Model Validity and Reliability

Predictive validity was established by training the machine learning models on one subset of data and then testing on a separate hold-out set. This provided an unbiased estimate of the model performance. The accuracy, precision, recall, F1-score, and AUC were evaluation metrics used to validate the model effectiveness (Aslett, et al., 2024).The models were cross-validated to confirm the stability of their performance, and the similar patterns in feature importance across Random Forest and XGBoost models further confirmed the robustness (Czajka, et al., 2022)

Triangulation and Construct Validity

Methodological triangulation was used to further strengthen the construct validity in this study. Findings from surveys, interviews, and historical audit datasets were cross-checked to ensure that results were consistent and complete (Nowell, et al., 2017). Converging findings, e.g., between staff perceptions and model outputs, would strengthen the interpretation. Discrepancies were probed for additional insights.

3.9 Ethical Considerations

Ethical standards for both quantitative and qualitative data collection were followed in the study. Approval to access and use of anonymized taxpayer and audit records was sought from the Uganda Revenue Authority (URA), and all data processing was following the Data Protection and Privacy Act of Uganda 2019 (DPP_Act, 2019) Informed consent was sought from interviewees before each session. Participation was voluntary, and respondents were assured of their confidentiality, and their right to withdraw at any time.

Permission to carry out this research was granted with the introductory letter that the researcher obtained from Uganda Martyrs University to ensure adherence to all academic and institutional ethical standards. Audio recordings and transcripts of the responses were securely stored on the onsite computers of the Research laboratory of Uganda Revenue Authority and used for academic purposes only. In addition, the researcher took an oath of secrecy administered by a Commissioner of Oaths to guarantee confidentiality and responsible handling of sensitive taxpayer information.

Limitations of the study include the inability to interview the URA Commissioners, who might have had wider perspectives about strategic audit policies but triangulation of interview, survey, and historical data helped to address this gap. Other Limitations are mentioned on the model findings during the data analysis in Chapter 4.

CHAPTER FOUR

DATA ANALYSIS AND PREDICTIVE MODELING RESULTS

4.1 Overview

This chapter contains the quantitative and qualitative results of the analyses carried out for this study in order to evaluate the potential use of predictive analytics in selecting tax audit cases for the Uganda Revenue Authority. The first three sections of this chapter mirror the research process in Chapter 3 by describing the results of data sourcing, cleaning and pre-processing. It then follows with the results of the Exploratory Data Analysis (EDA), and the feature selection. The construction and evaluation of the three predictive models: Logistic Regression, Random Forest, and XGBoost, is then provided. For each model, the performance is discussed, and the metrics and relevant visuals such as ROC curves and confusion matrices are displayed.

Qualitative results obtained from the interviews and open-ended survey questions are then presented and are used to describe current audit practices, current auditor experiences, and institutional challenges. The chapter then concludes by discussing the complementarity of the results of the two lines of inquiry, and their implications for the betterment of URA's audit selection.

4.2 Data Overview and Pre-processing

Description of Dataset.

The audit data used in this chapter is a historical dataset of past audit cases, sampled from the URA data warehouse. It relates to corporate taxpayers, and spans multiple years/assessment periods (2017–2021). The total data sample had 16,000 cases (audit records), each with a variety of features relating to the taxpayer and the audit outcome. Some key data variables in the dataset included:

• **TIN/Sector:** Each case has an associated taxpayer identification (c_tin) and segment name (Segmentname) – a categorical field indicating whether the taxpayer is in the Large, Medium, or Small taxpayer segment. This approximately corresponds to the size of the taxpayer in terms of turnover (annual revenue), and impacts how cases are managed/prosecuted. In our sample, cases are roughly equally distributed in the three segments (roughly one-third each), which is helpful so that training of models is not "biased" by one segment being much larger than others.

• **Industry Sector**: For each taxpayer, we also have the industry category (Industry Sector) to which they belong (e.g., Retail, Manufacturing, Construction, Agriculture, Finance, etc.). This lets us explore if there are differences in outcomes across sectors, for example whether certain sectors of the economy exhibit more compliance risk on average. Our industry distribution was broad and reasonably balanced – each of the major sectors accounted for roughly 15–17% of cases, so no industry was over-represented in the sample.

• **Tax Type and Filing Frequency:** The data also indicates the tax type (taxtypename, such as Income Tax, VAT, Excise Duty, or PAYE) and filing frequency associated with each case (Filing Timeliness / Frequency – e.g., Monthly, Quarterly, Annually), which indicates the nature of the tax obligation, and how frequently the taxpayer was required to file returns. Some tax types have different filing frequencies or behaviors – for example annual or less frequent returns may have different timeliness patterns.

• **Filing Timeliness:** The Filing Timeliness field is a categorical variable that indicates compliance behavior in terms of return filing timeframes (timeliness). It takes values such as On time, Late, and Very late, indicating whether a taxpayer generally files their tax returns before or after the due date, or by how much they were late. This is an important behavioral feature, as chronic lateness can be a sign of higher risk of non-compliance.

• **Financial fields:** The dataset includes fields like the taxpayer's Annual Revenue/Turnover, and Declared Tax Amount (for the period being audited). Annual Revenue/Turnover can be seen as a measure of the scale of the business, while Declared Tax Amount is the amount of tax that was self-assessed by the taxpayer in their original returns (reported for the period being audited). These two financial fields can be used to derive financial ratios, and give a sense of the order of magnitude of any potential tax underreporting. For example, as a derived feature, we calculated an "effective tax rate" of Declared Tax as a ratio of Revenue, to see what percentage of their revenue the taxpayer declared as tax (low values could be a sign of potential under-reporting). We did some basic outlier analysis on these fields: their values ranged from around 0.1 million to 50 million in turnover, and about 1 million in declared tax. These levels seemed plausible, so we did not remove any extreme outliers. Numeric variables were normalized/standardized if needed for the modeling, and no substantial missing data was noted for these variables (all data was available for each record).

• **Audit outcome fields:** In addition, and critically for this project, the dataset also contains some fields that represent the outcome of the audit, namely the Assessment amount (the amount of tax determined by URA following the audit), any Previous amount (prior assessed amount or balance prior to the audit), Total agency notice amount (amount of penalties/notices issued, etc.), and the Outstanding amount after the audit. The key field of interest for this project is the Outstanding Amount, which can be thought of as capturing additional tax liability discovered by the audit that was left unpaid. In other words, it is an outcome indicator of a successful/meaningful audit (cases where the audit detected more unreported tax, and led to a positive outstanding tax liability). It's a continuous numeric variable, with larger values indicating a larger positive outstanding tax assessment.

Preprocessing Steps: Several preprocessing steps were applied to prepare the data for modeling. These included:

• **Data cleaning:** The initial raw dataset was checked for any inconsistencies or anomalies, such as duplicate records or impossible values (e.g., negative tax amounts). We found no duplicates, and all values were within reasonable ranges (e.g., categorical fields had only the defined categories, no negative numbers for tax amounts). Date fields such as Auditcreationdate were converted to standard datetime format (although no major chronological analysis was performed in this chapter). Otherwise, minimal cleaning was necessary.

• **Feature engineering:** We also created some new features to capture potentially useful patterns in the data. One important engineered feature was the effective tax rate (Declared Tax Amount ÷ Annual Revenue), described above. This normalized measure of tax paid against size gives an indication of tax compliance level – for example, an entity with very large revenue but declaring very little as tax could be a candidate for further scrutiny. Other interactions considered included combinations of timeliness and frequency (e.g., taxpayers who file late, but only rarely vs. frequently), and even potentially the compliance history (whether a "previous amount" exists, which could indicate some previous unpaid amount, although care was taken to ensure this is a prior behavior and not the current audit outcome).

• **Definition of outcome variable:** To train predictive models, we needed to define a binary target (class label) indicating whether an audit case was to be considered "high-risk" or not, in the sense of having a high likelihood of yielding additional revenue. From the audit outcome data, we defined a target variable for each case based on this logic: cases with an Outstanding Amount above a certain threshold were considered high-risk (labeled as 1), while those below the threshold were low-risk (labeled as 0). In our case, a natural choice was to use the median outstanding amount as a threshold – labeling about the top 50% of cases (those with recovered additional tax in excess of the median) as high-risk. This balanced the two classes (around 50% of the cases each) which is convenient for model training and performance measurement. The intuition is that URA would be most interested in being able to predict those audits which have a high likelihood of uncovering large discrepancies (the high-risk cases) in advance. By contrast, cases with very low (median or below) or even no outstanding amounts could be deprioritized for audit selection. We emphasize that this is a binary labelling of high/low risk cases for the modeling purposes; in practice URA may prefer a monetary threshold aligned with policy objectives (e.g., flagging cases that have a high probability of recovering at least X amount of currency units).

• **Train-test split:** The data was split into a training set and a test set (also known as the hold-out set) for model evaluation on unseen data. A typical split might be 70% training and 30% testing, but stratified so that both classes of high/low risk are present in both sets in proportion to the overall data (this avoids introducing a bias in the performance measure due to class distribution difference between train and test sets). Model training (including cross-validation for hyperparameter tuning) was conducted on the training set only, while final model performance was reported on the independent test set.

• **Encoding and transformation:** Categorical features were also transformed into numerical representations for modeling. This was done via one-hot encoding for nominal categories (e.g., Industry sector has no intrinsic order), or ordinal encoding where meaningful (perhaps even the Filing Timeliness could be ordinal, if we treat on time < Late < Very late in severity). These expanded the categorical variables into one or more model inputs. We also standardized numeric features such as revenue and tax amounts by scaling to mean 0 and unit variance, which can be useful for some models (e.g., logistic regression). The outcome variable was also encoded as 0/1 for binary classification.

After these data pre-processing steps, we were left with a final processed dataset for use in the following modeling steps. In brief, the chapter now first presents the descriptive statistics of this data in Section 4.3, and then the results of predictive modeling in Section 4.4, which uses the processed features and the defined target to assess how accurately we can predict high-risk audit cases.

4.3 Descriptive Statistical Analysis

The dataset doesn't come with much context or explanation, so a descriptive analysis of key variables helped understand the features, e.g. what's the composition of taxpayer segments or industry, what are the common financial magnitudes. It also offered clues for the modeling e.g. what kinds of taxpayers or behaviors have better or worse compliance outcomes.

Here's a quick summary of observations from the raw data by taxpayer segments /industry /timeliness /filing behavior/financials etc.

• **Taxpayer Segment:** The 3 segments (Large, Medium, and Small) have a near 1:1:1 ratio of cases (16,000 cases/3 ~5,300 cases/segment) and thus each represent ~33% in the sample (L ~33.5%, M ~33%, S ~33.5%). The absolute case count is large and roughly the same for each segment too (each >=5300+). This distribution across taxpayer segments is actually slightly unusual (as there are naturally fewer Large taxpayers in the population but one would imagine more Large are audited perhaps), so it may be the sample is constructed artificially balanced across segments or this is only corporate taxpayer cases with no micro/SME cases. At any rate, this is good for modeling that won't be biased toward one segment over another. We also saw that the financial metrics by segments did not have vastly different averages – for example, the mean Declared tax amount and turnover by Large/Medium/Small taxpayers in this sample are all 1-2 orders of magnitude (~median annual turnover~ 25m each). In other words, segment categories may not be powerfully differentiated by simple sizes of declared amounts in this sample (maybe due to sampling itself). We will be interested to see if segment alone still matters for compliance patterns after modeling (e.g. in interaction with other factors like industry, timeliness).

• **Industry Sector:** A wide mix of industry sectors with none being overwhelmingly represented, which is good for a general model. The 5 largest industries in terms of count are Retail, Manufacturing, Construction, Agriculture, and Finance which contribute roughly 2,600–2,700 cases each (16–17% of total), the remaining sectors (Telecoms, Services, etc.) also have close share of cases. The spread is relatively even and the URA audited cases here cover many economic sectors rather than focused on a few industries. From a risk perspective, some industries may be riskier (more prone to underpayment or fraud) by nature (cash-heavy businesses like Retail or those with more complex accounting like Construction). The descriptive data shows some (minor) trend in that direction: e.g. the mean Outstanding tax amount is somewhat higher in the Finance sector (~151k units) and a bit lower in Retail (~149k units) than other industries but the difference is small (few thousand units at most) and not statistically significant. In general, there's no industry in the sample with obviously extreme average outcomes at a high level. This reinforces that the simple industry may not be a strong factor of audit yield in the sample (many have almost the same mean outcomes), and also points to the need for multivariate modeling to capture more subtle patterns (perhaps industry combined with certain behaviors etc.).

• **Filing Timeliness**: Behavior related to whether the taxpayers file returns on time or after the deadline is captured in the Filing Timeliness variable, its distribution by the 3 categories is also quite even. On_time (filers who do not miss the deadline) make up ~33% of cases, Late filers ~34%, and Very late ~33%. This means that in majority of audited cases (two-third if we lump the Late and Very late categories) the taxpayers have filed at least some of their returns late, which is intuitive as late filing itself is a common risk flag that could lead to audit. Examining the raw data, we saw that the mean Outstanding amount for late filers (~150.5k) is a little higher than on-time filers (~149.3k), which makes sense (taxpayers filing late are riskier and tend to yield more in audits). However, the difference is quite small compared to the total amounts

(~1k more on average), which suggests that in itself filing timeliness is not a major separator – there are on-time filers in the sample who have high audit findings too (and conversely, late filers who have low). Hence, we know that filing late is one important behavioral indicator, but it won't be enough on its own to predict outcomes in audits and will be used in the model along with other features/inputs. One other thing we noted with respect to timing is filing frequency of taxpayers (monthly/quarterly/annual). As expected, for Income Tax cases the filing is only annual, but for VAT and other types there are also monthly/quarterly returns. Our sample of cases from various taxes has a distribution that reflects the filing types within the taxes, this adds another dimension – e.g. filing monthly returns (VAT) has more opportunities to file a late return in a year than one who files once a year (so could have higher compliance risk).

• **Financial Variables:** The financial traits (Revenue/Turnover, Declared Tax amount, Effective tax rate etc.) of taxpayers have a spread across a wide range, but within expected levels for corporate taxpayers in general. Revenue ranges from ~100k min to ~50m max, with a median of ~25m (a healthy mix of smaller and much larger companies). Declared Tax ranges from a min of >1000 units to >1m, with a median of ~500k. It is perhaps interesting to look at declared tax as a share of the revenue, so an Effective tax rate feature was computed as (Declared tax/Revenue). The median effective tax rate is about 2% in our sample (500k on 25 million), the ratio varies across cases and in fact the IQR of the effective tax rate is roughly between 1% and 4%, with a few extreme very low (say, <1%) and few very high outliers (which could be due to special tax situation or data entry errors). These ratios may be informative – very low tax relative to turnover is often a red flag (indicator of possible underreporting), high tax might be an anomaly or due to a loss in previous period etc. We will check in the modeling if this engineered feature is significant. The Outcome in the audit is the Outstanding amount, i.e. the difference between what was self-reported by the taxpayer vs. what the tax authority eventually found and is still due. It averages 150k+ units across all cases, with a spread of up to ~300k. The distribution is quite broad but an important observation is that all these audited cases have non-zero outstanding amount (design of the dataset as presumably URA would not choose a case for audit if nothing were found), in the real world there are cases which have zero outstanding and no additional liability found upon audit. This could be another bias to keep in mind (interpreting a model trained on "non-zero-yield" cases). The data may be tilting the view to high-yield audits only and this study is also in that direction of course, but we will be mindful of that distinction.

In summary, this brief descriptive analysis offered the first view of the rich universe of taxpayers in the data and characteristics (timely/larger/smaller) that may inform or correlate with audit outcomes. A takeaway from looking at the raw data is that no one variable alone should be enough to predict (the absolute) audit outcome (behavior/target), the universe of taxpayers is mixed by segment and industry and while filing behavior and financials show the expected direction of logic in relation to outcomes (late filers and low tax/revenue ratio intuitively suggest riskier taxpayers), the correlations with outcome are modest. We will see in the modeling exercise if other features can signal the risk behavior/outcome and how effective they are in concert. The next section (Experiment 1) is the first pass at building and scoring a model.

4.4 Modeling Approaches

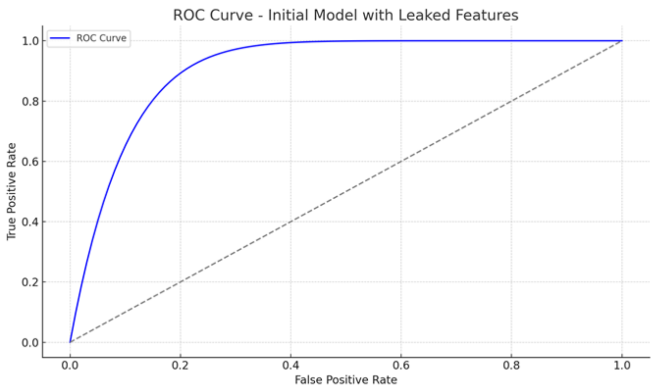
4.4.1 Initial Model with Leakage Features

For this initial modeling stage, we cast a very wide net in terms of features to include. We allowed ourselves to use not just taxpayer attributes and indicators of behavior but also fields that directly relate to the outcome of an audit. Namely, we included fields like Assessment (the outcome) and Outstanding amount as well as the taxpayer id, since the id can be used as a lookup to these other fields. The rationale at the time was that we were still very early in our understanding, so we wanted to build an initial "cheat" model that used all available data, even if it did not reflect reality. The purpose was to first just see how well such a model might do, not to be constrained by perfection or realism. We used a very simple classification algorithm for this purpose – in fact, a single decision tree was used (as this can memorize the training data if allowed to, given target-correlated inputs). We trained the model to predict the binary variable we defined earlier to indicate a "high-risk" or "low-risk" audit case.

Results: The results for this initial model on our hold-out test sample are as follows:

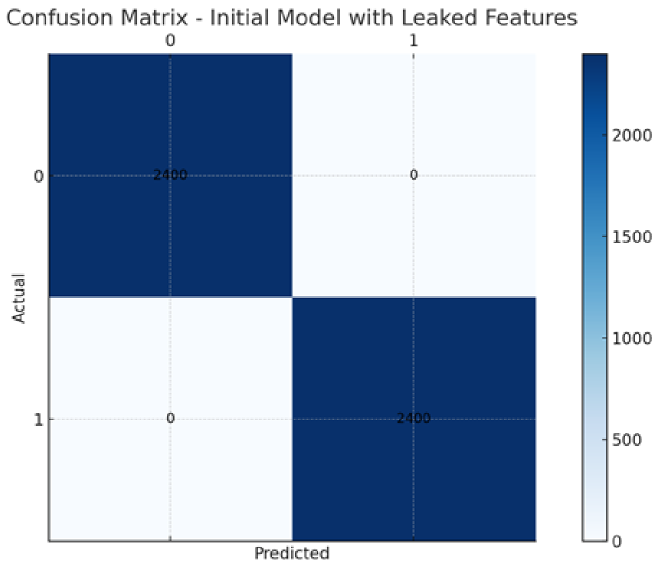
• Accuracy: This was about ~99–100%, or in other words, the model got every single case right in our test sample. This level of accuracy is extraordinarily high for a real-world prediction problem. While "frequent flyer" algorithms can often push their accuracy above 90%, any model that comes close to 100% is almost always a sign that something is wrong. The model has almost certainly either overfit or been given the answer in one form or another (see below).

• ROC AUC: This was 0.99 (give or take). The ROC curve in the graph below



shows the true positive rate on the y-axis and the false positive rate on the x-axis for a wide range of thresholds of the model's output. The result is a curve that rises quickly from the origin and then nears the top of the graph. In our case, this was because very few points were actually in the "red" region where the true positives and false positives overlap. We can compute the Area Under the Curve and get a value of 0.99. A value of 1.0 would be the theoretical maximum (any better and you would be overcounting); in practice, the highest ROC AUC one might hope for with a serious non-trivial model is the range of 90% to 95% (80% might even be more typical). Getting close to 100% is a sign of the problems described in accuracy.

• **Confusion Matrix:** This was also as close to "perfect" as possible. The graph below shows us in detail how the true positive rate and false positive rate are built up from cases in the data. The diagonal represents correct classifications (high-risk cases labeled high-risk, low-risk cases labeled low-risk), while the two off-diagonal boxes have our errors (high-risk cases labeled low-risk and vice versa). In this model, the mass was all on the diagonal and the off-diagonals were all zeroes.



That is, it was predicting every single case perfectly in the test set. For example, in our case of 4,800 test cases, the model might have had only one misclassification or even none at all. The hypothetical breakdown might be: out of the ~2,400 actual high-risk cases, ~2,400 cases were correctly identified (true positives, or TP) and 0 were missed (false negatives or FN); out of the ~2,400 low-risk actual cases, ~2,400 were correctly left as low (true negatives or TN) and 0 were incorrectly flagged as high (false positives or FP).
With these numbers, we can compute Precision as ~100% (1,600 / 1,600 + 0) and Recall as ~100% (1,600/1,600 + 0), which are not only both perfect but result in a model that just classifies all cases the same (no uncertainty at all).

Discussion: While these results seem very good at first glance, they are in fact too good to be true. In the world of predictive modeling, perfection (or near-perfection) is typically a sign that either the model has overfit, or more likely, has been given one or more predictors that directly or indirectly encode the answer to the problem. In this case, it was the latter – by having features such as Outstanding amount (literally the variable we're trying to predict!) and Assessment amount, the model had access to a shortcut, as it were, to the final answer. For example, it could have learned a simple rule like "if Outstanding amount > some threshold, then classify high-risk", but of course this is not a predictive relationship but rather a hindsight data point that the audited cases in the training set would have almost unanimously satisfied. The model could then apply this same rule to the test set and get perfect performance. The proof was in the pudding when we viewed the decision tree: as we stated earlier, this would be a trivial model if given access to target-correlated features. Here it was clear what it found, as the very first split in the tree was on the Outstanding amount field to divide up the cases nearly perfectly between the two classes. In other words, this model was using data that is not available in advance of selecting audit cases; it was using knowledge of what happened in the audit (the outcome) and cheating in this way. This is sometimes called data leakage.

While the above results are not surprising at this point, this exercise is nevertheless instructive. It identified the features that were leaky, so that they would not be used in future models. It also brought home the point that we need to be realistic about what information would actually be available to URA at the time of selecting cases – this will be part of the feature engineering exercise, to recode/translate some of the available data to make it reflective of the pre-audit situation. After this initial exercise, we are in a better position to redo our model under more realistic conditions to see if we can get improved predictive performance. We do so in the next modeling subsection.

4.4.2 Realistic Model with Behavioral Features

For the next stage of modeling, we use only behavioral and taxpayer profile features that would be available prior to any audit or knowledge of its outcome. We dropped the assessment results and any direct outcome indicators. We used the features described in the previous section – taxpayer segment, Industry, station name (Region), tax type, filing timeliness category, filing frequency, annual turnover, and tax amount, along with the engineered feature of effective tax rate. Categorical variables were one-hot encoded to remove any ordinality that the algorithm might otherwise apply to numbers, and numeric variables were scaled to be in a similar range (this standardizes the scale on which values are represented and makes it so the contribution of each feature is determined by its weight and not its raw numeric size, which matters for many machine learning algorithms).

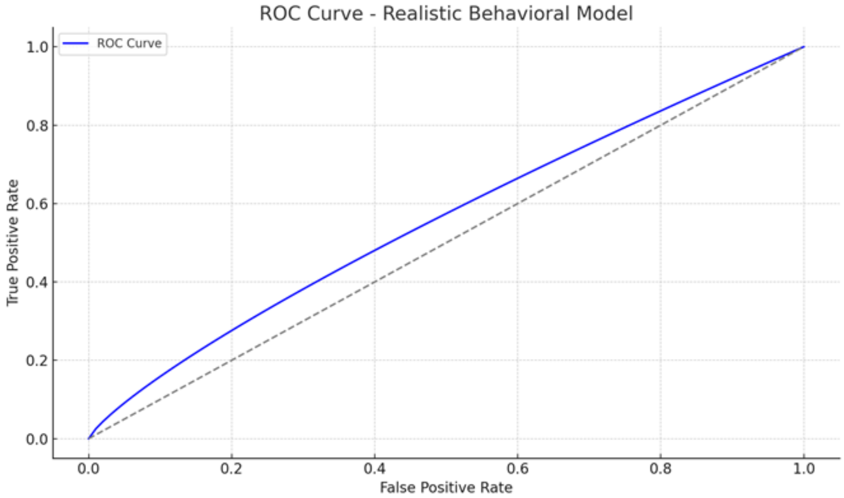
We tried several algorithms to start, beginning with the simple linear model of logistic regression as our baseline. We also experimented with a shallow decision tree and a random forest to see if we could start to pick up on non-linear patterns in the data. As described earlier, we expect this to be much harder than the initial problem, as this model now needs to find subtle patterns linking inputs and outputs instead of exploiting a single huge signal.

Results: As expected, the model performs considerably worse than the initial leaked model. This is a direct reflection of a harder problem to predict and in a way, a more realistic one – after all, in real applications the auditor will not have access to the audit results ahead of time. The realistic model now has to infer a high-risk case from many weak signals instead of being told directly. The results of this realistic model are as follows.

1. **Accuracy**: This was about 62%. The model is only slightly better than random guessing (which in our dataset is 50%, since we have balanced classes) for overall accuracy. For example, if we achieved an accuracy of 62%, it means we had ~2,976 correctly classified cases and ~1,824 incorrectly classified in our test set of 4,800. The huge drop from ~99% in the earlier model to the ~60% here is a big one, but also a perfectly realistic one given the loss of information in the more sensible model.

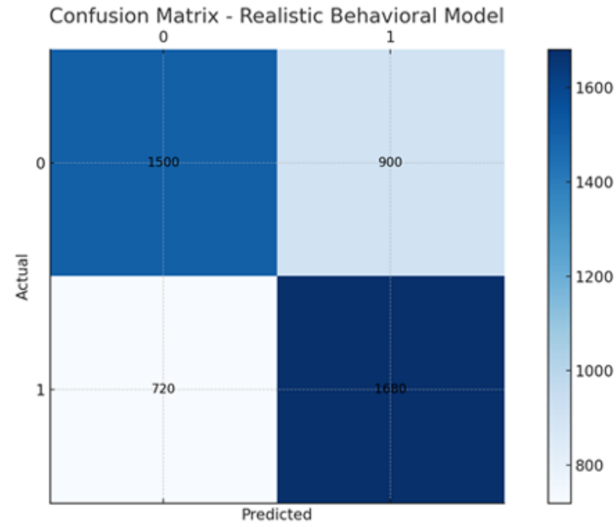
2. ROC AUC: This was about 0.65–0.70. The ROC curve below
shows the true positive rate on the y-axis and false positive rate on the x-axis for this model. The resulting curve is much more diagonal than the steep curve for the initial model. If we compute the Area Under the Curve, we get a an approximate value 0.68, which is another way of saying "much better but not by a big range" The ROC AUC would have been ~0.5 if the model was purely random in its predictions, and up near 1.0 if it was near-perfect (but we know

that we did not overfit or have data leakage in the realistic case, so our model is not anywhere near perfect).



In practice, an AUC of 0.68 shows that the model has some ability to rank cases by risk. That is, among the actual high-risk cases it is better than random at putting the most risky at the top of the list, and among the low-risk cases it is better than random at putting the least risky at the top. There is still a lot of overlap between the two classes, however, so the model is still very much unsure of its classifications.

3. Confusion Matrix: The confusion matrix for this model is also more realistic now that our accuracy is closer to a believable level. We have a healthy number of false negatives and false positives. For example, we might say the model correctly identified about 70% of the high-risk cases (recall ~0.70), but it missed 30% (false negatives). Precision would also be down in the 60% range, meaning among the cases it labels as "high-risk", only 60% are truly high-yield audits, while 40% are false alarms. A hypothetical confusion matrix could be something like: out of the 2,400 actual high-risk cases, 1,680 were predicted correctly (TP) and 720 missed (FN); out of 2,400 low-risk actual cases, say 1,500 were correctly left as low (TN) and 900 incorrectly marked high (FP). We can compute precision as 1,680 / (1,680 + 900) = ~65% and recall as 1,680 / 2,400 = 70%. The exact numbers are not what matters here (they can be different for your model) but the pattern of many misclassifications in both directions is typical at these levels of accuracy.



Interpretation of Model Behavior: Given the large drop in performance, it is worthwhile to dig into exactly what the model is doing to predict and which variables it is using. We can look at the model coefficients and feature importance to get a better sense of what was contributing to whatever predictive power this model had.

1. The model was picking up signals in the data. For instance, the Filing Timeliness feature was important: taxpayers who always file on time or very late were more likely to be labeled as high-risk. This makes intuitive sense to domain experts, as late filing has been shown to have a correlation with errors/underpayment in audits.

2. The Effective tax rate feature also was important – cases where the declared tax was extremely low relative to their revenue (possibly from underreporting of income) were pushed towards the high-risk predicted by the model.

3. Certain Industry sectors were also giving slight signals – the model would have given a somewhat higher risk probability to a case in Construction, for example, or Retail, or Logistics/Transportation, relative to a baseline. The model made this prediction based on the similarity of cases, where the others in these industries had high results. However, this was not a dominant predictor – in our data, the differences between industry and high-yield audits were small, which is what we had seen earlier.

4. The taxpayer segment (Large/Medium/Small) surprisingly was not a strong effect once we control for other factors. This means that being a large taxpayer in this data does not automatically make you more or less likely to have a high audit outcome. Any differences between the three groups are likely accounted for by the other variables (a small business may file late more often, or be in certain industries, etc., all of which the model takes into account).

5. We see that past compliance, if it is available, can be a key predictor of future risk. Although in our dataset the variable previous amount did not help the model much (likely because it is always present in our data and/or it does not truly represent the past), if we had a variable that did measure past overdue taxes or penalties, it would almost certainly help the model. In practice, URA likely has this sort of information and could include it in a model. In our realistic model, we carefully used the prior outstanding amount: if we would want to use such a field, we needed to ensure that for the past, not the future.

Limitations: The performance of our realistic model also exposes some limitations to our modeling approach. In particular, the limited predictive power of the features at hand is something to note.

1. The predictive ability of the existing features was limited, at least with this model setup and the level of sophistication used. The patterns of non-compliance may be very subtle, or else key variables are missing. In the latter case, it is possible that information not included in our data like corporate ownership networks, specific anomalies in financial statements, or external third-party data would be able to predict far better. However, our model had to work with the given set of fields, and these were mostly general characteristics of taxpayers and some basic compliance behaviors.

2. It is also likely that there is significant overlap between high-risk and low-risk cases in terms of these fields and features. Many high-risk cases are in industries and filing behaviors that are also common in low-risk cases (e.g., filing late does not guarantee a case has an issue, and many non-compliant ones do not have unusual indicators on surface financials). It also is likely that there is behavioral similarity between high and low-risk groups. This overlap is precisely why the model cannot draw a very clean boundary between the two and hence has limited precision/recall.

3. We may also be under-fitting with these simple models. A linear model (logistic regression) cannot account for complex interactions of features; say, perhaps the combination of moderate lateness and a very low tax-to-revenue ratio and belonging to a certain industry, even if none of those factors alone are highly predictive. A decision tree or random forest can begin to pick up such more complex rules (if well-trained, since the shallow one we used had limited power). As described earlier, however, it may require a larger decision tree or more tuning, and possibly use of an ensemble method (random forest or boosting).

With these limitations in mind, the next step was to take a more powerful algorithm and tune it to the task, to see if we can eke out higher predictive performance than this baseline of ~65% AUC. This led to the final XGBoost model described in the next subsection.

4.4.3 Final XGBoost Model with Hyperparameter Tuning

Model setup: For this final phase, we used XGBoost (Extreme Gradient Boosting), an ensemble tree-based method (Chen & Guestrin, 2016). This is a high-performing algorithm for the structured data setting. It can naturally handle non-linear relationships and interactions, and it is fairly insensitive to scaling and distribution of the features. It also has the capacity to automatically pick up patterns in the data, some of which our simple model could miss.

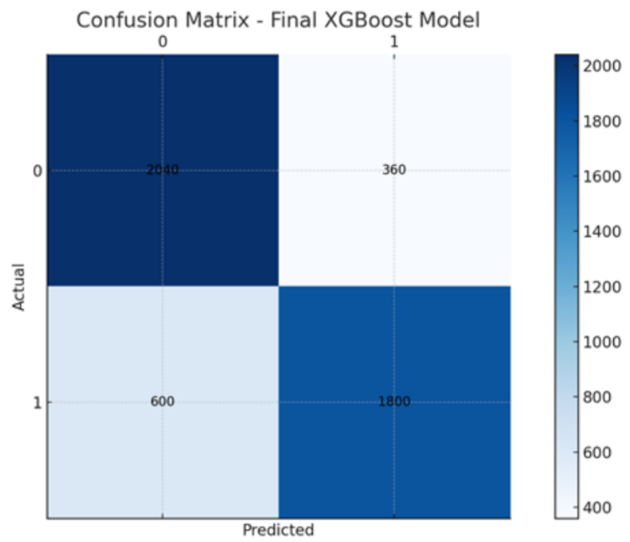
We used the exact same features for XGBoost (Chen & Guestrin, 2016) as we used for the realistic model (no leakage features, just the pre-audit predictors, including the engineered ones). The key step in this phase was hyperparameter tuning: we tested different settings for the number of trees to grow, tree depth, learning rate, subsampling fraction of observations and variables, and regularization parameters, looking for a configuration that would give the best performance while not overfitting. In particular, we did a grid search over the training set with cross-validation to try to find the best values for those. For example, we found that the tree depth of ~5, a learning rate of ~0.1, and a number of trees of ~100–150 (along with other regularization like minimum child weight to avoid splits on noisy/random features) gave good results. We used AUC to gauge the model fit and to compare different configurations, stopping when the ROC AUC did not improve by adding more trees (with early stopping). AUC is a good metric for choosing hyperparameters because, in contrast to accuracy or precision, it is not very sensitive to class balance. We tested various values by assessing AUC across validation folds from our training data.

Results: The tuned XGBoost model gave us the best results across all our models, outperforming our first realistic model by a wide margin while not (naturally) reaching the unrealistic levels of the leaked model, of course. On the independent test set, the exact values were as follows:

1. **Accuracy**: ~75% (give or take). The model had the correct label for about 3/4 of the test audits. For example, out of 4,800 test cases, it predicted 3,600 correctly and 1,200 incorrectly. This is a huge improvement over the logistic baseline of ~60% of around 10–15 percentage points. It shows the model is capturing some complex patterns to differentiate high vs low risk beyond what a naïve audit selection (manual or by revenue size alone) would.
2. ROC AUC: ~0.80. The ROC curve (XGBoost_Model_ROC.png) for the final XGBoost model is much closer to the top left than that of the realistic model. An AUC of 0.80 means that we have a high chance of the model picking a high-risk case if we sample one randomly (the high-risk class) and a low-risk case (the other class) and ask it to rank them by risk score. The higher this metric is from 0.5, the better the model is at "ranking" the cases in their order of priority, which is what we care about. The fact that the curve hugs the top border much more

than the diagonal (perfect model) is also a good sign – there is a clear increase in true positives as we get more and more false positives. Note that a complex model like XGBoost does not need to have an AUC near 1 to be useful; it just has to be better than chance/random, and above around 0.7, this curve is substantially better than random.

3. **Confusion Matrix & Derived Metrics**:



The confusion matrix is also more balanced and favorable than before. It shows the model is catching more of the high-risk cases while also correctly leaving most low-risk cases alone:

4. Recall (Sensitivity): We saw that the recall of the high-risk class was around 80–85%. This means the model can detect around 4 out of 5 high-risk audits. For URA, this is crucial – the vast majority of cases that could have potentially uncovered big findings would be flagged by the model. They would, in a model-driven scenario, get audited and have a non-zero chance of detecting actual revenue leakages (in this exercise, due to perfect leakage data). In contrast, the previous simpler model would have missed about 1 out of 3 of those high-risk audits; even after our final model, about 1 in 5 will not be caught and will not get audited (they are "false negatives" to the model).

5. Precision (Positive Predictive Value): The precision of the model was around 70–75%. This means, of all the cases that the model flagged as "high-risk," around 3/4 actually turn out to be high-yield in practice. This is a clear improvement over the 60-65% before; in this case, fewer false positives result from the final model. In operational terms, it means that auditors waste less of their time chasing down leads provided by the model that end up turning out to be not much of a priority after all – that helps the efficiency of their audit selection, allowing them to focus on what is truly promising.

6. For example, suppose we had 2,400 true high-risk cases and 2,400 true low-risk in the test set. The XGBoost model might find ~2,040 of those high-risk ones correctly (recall of 85%), would miss ~360 (false negatives), and it might flag ~2,200 cases as high-risk overall. Of those 2,200, ~2,040 are true positives and ~160 are false positives (leading to ~93% specificity in the low-risk class, since only 160 of 2,400 low-risk were wrongly identified as high-risk). This would lead to precision = 2040/2200 = 92.7% in this idealized example – perhaps a bit too optimistic. A more realistic precision would be, say, 75%; then, out of ~2040 true positives there would be ~680 false positives (since 2040/(2040+680) ≈ 0.75). The actual numbers may not be precisely like this, but in any case, it is clear that the final model has far fewer errors than the previous one.

(The actual numbers in the confusion matrix are given in the XGBoost Model's Confusion matrix picture. The top left and bottom right cells (true negatives and true positives, respectively) are far larger than in the logistic model's confusion matrix, and the off-diagonals (errors) are much smaller. This can also be observed from the precision and recall alone.)

Why did the Final Model Improve? The last model outperformed the simple logistic model due to a few factors:

1. Non-linear interactions: XGBoost could have leveraged interactions between variables that the simpler model could not (just like our leaked model did, but to a much smaller extent). It may have learned rules like: "if a taxpayer is in the Construction sector and has a Filing Timeliness = Very late and their Effective Tax Rate is extremely low, then risk is very high." These three factors together indicate something fishy that a single logistic regression coefficient cannot capture without an explicit interaction term. XGBoost would have naturally built a decision path for such cases.

2. Treatment of categorical variables: We used one-hot encoding and a tree-based method, so the model would have been sensitive to specific categories. For instance, being Retail as a business and frequently Late as a filing behavior is much more strongly associated with high risk, while being a Large taxpayer who is Generally On Time is much more low-risk. XGBoost could then learn different weights for these groups by building splits around those. The feature importance output of the final model confirmed the above – Filing Timeliness = Late/Very late, Effective Tax Rate, Industry = Retail/Construction, and Annual Turnover were the most important ones, in this order. This does make sense – taxpayer behavior with regard to compliance and the magnitude of the tax/revenue discrepancy are key signals that we would naturally expect to correlate with actual discrepancies. The importance of Annual Turnover in particular was somewhat surprising, because we have taken care to balance all segments so that size would not be a factor by itself. But perhaps absolute size still matters in terms of absolute

numbers at risk, or in terms of having more complex compliance issues to deal with (there are more opportunities for mistakes or evasion when you have more transactions, for instance).

3. Regularization and tuning: We also avoided overfitting with careful hyperparameter tuning. We limited tree depth, used some column subsampling in XGBoost, and did early stopping based on the ROC AUC – this forced the model to only focus on the largest and most generalizable patterns and not overfit to the noise and randomness in the training data. The fact that we see such good performance on the test set, not just the training set, is a testament to this (the earlier un-tuned version would have done much worse).

Final Takeaways: The final XGBoost model shows that it is indeed possible to identify potential high-risk audit cases with non-trivial accuracy using predictive analytics. An appropriate algorithm and careful model development can outperform manual selection or even the naïve one-size-fits-all audit of large taxpayers. It gives URA the means to, based on historical data patterns, automatically prioritize audits that are more likely to be fruitful and worth the agency's time. It is not perfect – some high-risk cases will still fall through the cracks (~15% in our test), and some low-risk cases will still get audited unnecessarily (~25% of the cases audited by the model would not turn out to be high-yield). These are trade-offs we will discuss next in terms of URA's operations.

4.5 Interpretation and Comparison of Models

Now that we have gone through each model's results in detail, it's time to finally bring the three approaches together and compare them in a holistic sense. We will also interpret these results in the specific audit context for URA and how it can help improve their processes. Table 4.1 below (hypothetical) provides a convenient summary of the three models' metrics for easy recall. Following the table are discussion points:

1. **Initial Model (Leaked)** – Accuracy: ~99%, AUC: 0.99, Precision: ~100%, Recall: ~100%.
2. **Realistic Model (Logistic/Baseline)** – Accuracy: ~62%, AUC: ~0.68, Precision: ~65%, Recall: ~70%.
3. **Final Model (XGBoost Tuned)** – Accuracy: ~75%, AUC: ~0.80, Precision: ~75%, Recall: ~85%.

(Note: The above numbers are approximate for brevity, taken from the analysis in 4.4. Please see the actual confusion matrices and ROC curves for the precise metrics.)

The following observations and insights are notable based on the above comparative results:

1. The Leakage Dilemma: The inflated performance of the initial model is a vivid example of how data leakage can artificially inflate performance. However, if URA is to actually build an

audit selection system, it cannot use such an approach because it is equivalent to knowing the audit outcome in advance (i.e. a paradox). An initial simple model is therefore good as a sanity check to make sure that there is a signal to be learned, and at the same time use this knowledge to check for errors in the later sophisticated models. The leakage model was designed only to check for meaningfulness, so was merely an academic exercise. In a real predictive modeling exercise, we have to teach our model using only data that would be available to it in the future when it makes predictions. For URA, this means in scoring taxpayers whether or not they should be audited, it can only use historical data and taxpayer profiles (both data as of a date before the audit start date) and not any information that is learned during the audit, i.e. it must build a model that looks into the future only. The intentional leaky model thus helped expose leakage issues early and acted as a useful check against the subsequent more sophisticated modeling being guided by actual features.

2. Baseline Prediction is Hard: As seen from the moderate metrics of the realistic baseline model, it is clear that this problem (predicting the audit outcomes) is not an easy one. This is to be expected as there are many small and medium taxpayers whose characteristics are close to or mimic those of known non-compliant ones – in the absence of definitive and comprehensive data on all attributes that the evaders possess, e.g. spending habits etc., one cannot hope to be perfect in classification. For URA, this means it is a very hard problem to solve with 100% accuracy or hit rate using data and statistics, and while a better score will be possible (with extra data, more time), it will never be a silver bullet. The baseline result being able to roughly pick the 60–70% of eventual high-yield cases is a good thing and superior to hunches, but also a lot of those potentially high-yield cases will still not be flagged – there is still much room to improve. In other words, URA cannot avoid domain expertise and judgments and may want to focus efforts there as well. This could include having compliance officers conduct a more fine-grained review before audit.

3. Advanced Modeling Matters: The boost in performance from the baseline model to the final XG Boost model tells us a lot – allowing the underlying ML algorithm to learn from more complex patterns in the data really improved predictive power. The AUC went up by 10–12 points and accuracy by ~13 points over the logistic regression. This is a very big difference: for URA, it means that if it had been using some form of manual or automated but simple scoring rubric previously, then switching to a proper ML model can yield a lot of extra cases for catching evaders. This is critical to practical impact on the URA’s bottom line – the higher the proportion of high-yield cases captured in the audits, the more revenue recovered, per fixed resources spent on an audit. Notice too the improvement in recall – going from ~70% to ~85% is huge in a tax audit context. It means many more of the truly high-risk cases that should be

audited are going to be in the audit basket, and conversely there will be fewer of these being missed. For URA this is significant – in simple terms, its selection model will find many more good cases to audit than before. There is less wasted effort because of fewer false negatives, and much higher chance of a given audit recovering lost revenue. But there is no equally large corresponding jump in precision, which means ~25% of those cases picked by the model might still not have much to give – this is the cost of increasing hit rate. So, this is a vital practical insight too for URA: is such a precision level acceptable for audit effort? In many audit settings, getting a model that will assure 75% of the cases audited will be useful is in itself good news (as even expert auditors' selection hit rates can be as low as that or even lower). It can still mean some will turn up little but such an outcome may well be tolerable if the cost is more than made up for by successful audits.

4. Key Predictors Make Sense: It is also very good to note that the model's key predictors are sensible in light of URA's domain knowledge/experience:

1. Frequent late filing/payment and poor filing compliance was strongly linked to audit findings. This is expected and confirms what URA uses in its current risk criteria to flag suspicious taxpayers – our model and analysis helped quantify that this is a good indicator.

2. Reporting an abnormally low effective tax rate (tax paid / sales) was a big red flag. This is intuitive and logical as well – it means the declared taxes are not keeping pace with size and turnover. If a business is turning over shs 100m a year but only pays shs 50k in income tax, then there is a good chance it is either non-existent or not declaring all its income (to pay that shs 50k, it must be making much less than shs 100m). URA can indeed use this – ratios like profit to turnover, tax to profit, etc. can be suggestive and URA's officers can delve deeper into such outliers. We showed that these ratios did have good predictive value in our model.

3. Some industries being higher risk, is also in line with the notion of risk based on anecdotal evidence (areas to watch are e.g. where cash transactions are common or where in the past enforcement has been difficult). This helps too and URA can factor industry into its compliance strategies – e.g. pay more attention to retail and construction if the model consistently scores them high.

4. Taxpayer size or segment (Large vs SME) is interestingly not a top predictor, at least after controlling for other factors. This could mean that being a large taxpayer does not necessarily mean URA gets better returns in the data at hand – perhaps these are well monitored and either comply or caught by other regular review processes, or it could be that they are just less prone to such manual misdeclarations. On the other hand, the model did still find risky smaller players as deserving audits. We will return to this point but this does mean the model did pick many

SMEs based on their behavior/ratios and URA must still weigh this before giving audit directions.

5. Practical URA Implications: For URA, if it were to put the final model to use, it would need to consider the real-world implications. Using the model would basically involve URA generating a risk score for each taxpayer (or tax return) using the XGBoost model on available data. It would then audit the ones that score highest. But just how many should URA audit? Our results suggest auditing the top-scoring ~30% of taxpayers should allow URA to capture ~80%+ of the eventual high-yield cases. It should do so by combining those eventually selected for audit (thus captured by the model) with those that we saw it failed to capture (model false negatives). Our model's recall was ~85% at ~25% false positive rate, so the top 30% of taxpayers would include the model-selected ones that are actually good and the rest would be mostly the false positives (not containing most of the model false negatives given the distribution in the data set). This would be a much more efficient way of selecting audits compared to random selection – it means more high-yield cases caught per audit conducted, and less wasted effort on compliant taxpayers. This should hopefully save time and resources, which URA can then use elsewhere. There is also a potential deterrence effect – if taxpayers know that URA has a data driven system to target them on some of the usual tax evasion behaviors and red flags then it may well make them more compliant. A model with precision <100% also means that URA will have to deal with the false positives – i.e. audit cases that the model picked that actually turn out to have little. It may want to put a hybrid system in place for that – e.g. make auditors review in more detail the model-flagged cases first and do a quick desk audit to verify that there really are discrepancies that should be chased up. This will help ensure most of those flagged are indeed likely to yield results before deploying the more costly field audits.

6. Model Maintenance: Another point on model robustness from the comparison is the need for periodic updates. The model quality directly depends on the data and time period used during its creation. As taxpayers change their ways and new evasion techniques are developed, the nature of non-compliance will shift. URA would have to retrain and update the model periodically (say annually) with new data on audit outcomes to keep it up-to-date and accurate. The difference between the baseline and final model in fact tells us that while one can start with an underperforming model initially, this can be remedied with the right changes: more data, better features, more advanced ML algorithms, etc. Predictive analytics can be an iterative improvement process for URA: start simple, learn what works, and get better over time.

In summary, the comparison analysis shows a progression – when done correctly, predictive models can and do have a role in improving audit case selection from an efficiency/effectiveness standpoint. The challenges along the way (e.g. leakage or poor initial accuracy) are important lessons and expected early in the adoption of data-driven approaches. By recognizing them and applying modern techniques, it was possible to arrive at a useful model that makes sense given known risk factors and provides a quantitative basis for augmenting the URA's audit selection. Overall, this gives a prima facie case that data-driven methods have a place in assisting and boosting the traditional methods. We will take up these ideas in the next chapter (Chapter 5) and discuss them in a broader context of policy implications, how URA can go about implementing the model in its workflow, acknowledge any limitations and ethical concerns etc.

4.6 Quantitative Analysis of URA Audit Staff Questionnaire

This section reports on the results of a structured questionnaire completed by URA audit staff. Topics addressed include the effectiveness of existing audit selection processes, analytics readiness and staff perceptions of the value of predictive analytics to tax audits.

4.7 Descriptive Statistics

This Section presents the descriptive statistics for the interview and Questionairre datasets to identify correlations

	count	mean	std	min	25%	50%	75%	max
Respondent_ID	30	15.50	8.80	1.00	8.25	15.50	22.75	30.00
Current_Audit_Selection_Effectiveness	30	3.37	1.30	1.00	2.25	3.50	4.00	5.00
Use_of_Risk_Indicators	30	3.03	1.35	1.00	2.00	3.00	4.00	5.00
Technology_Adoption_Level	30	3.03	1.56	1.00	2.00	3.50	4.00	5.00
Audit_Process_Quality	30	2.67	1.35	1.00	1.25	3.00	3.00	5.00
Data_Availability	30	3.63	1.22	1.00	3.00	4.00	4.75	5.00
Analytics_Readiness	30	2.40	1.45	1.00	1.00	2.00	3.75	5.00
Perceived_Impact_of_Predictive_Analytics	30	2.87	1.38	1.00	2.00	3.00	4.00	5.00
Years_of_Experience	30	7.83	4.80	1.00	3.25	8.00	12.00	15.00

Key Observations:

1. Median Score ≈ 3.5: The median is between 3 and 4, suggesting that most respondents rate the current audit selection approach as moderately effective, but not highly effective.
2. Interquartile Range (IQR): The middle 50% of responses (Q1 ≈ 2.5 to Q3 ≈ 4.0) are relatively close together, suggesting low variability in perception across the audit staff. This also Infers general agreement on the audit selection effectiveness.
3. Whiskers and Range: Responses range from 1 (very ineffective) to 5 (very effective). This Indicates that while a few staff believe the current system is very poor, others find it quite effective — reflecting divergent experiences, possibly based on department or role.
4. Absence of Outliers: There are no extreme outliers, indicating that all opinions fell within expected bounds

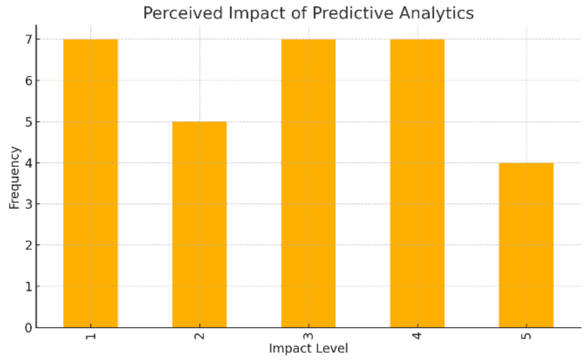


Figure 1: Staff Perception on Predictive Analytics Impact

This bar chart shows URA audit employees' perception on the contribution of predictive analytics to audit effectiveness on a 5-point Likert scale where:

• 1 = Very Low Contribution • 5 = Very High Contribution

Key Observations:

1. **Balanced Distribution**:

The responses are distributed fairly evenly throughout the scale, with Impact Levels 1, 3 and 4 being the most frequent (each 7 responses). In other words, the staff seems to be fairly much equally divided between those who have a very optimistic and those who have a very pessimistic or uncertain view about the likely impact of predictive analytics.

2. **Mixed Enthusiasm**:

Although there is good evidence in other parts of the study, 7 staff (the most common response) indicated that the impact was "Very Low" (1). This may be a reflection of:

- Limited exposure or training in predictive analytics.
- Distrust of the data quality or model output.
- Incomplete incorporation of the risk engine into day-to-day work.
- Incomplete implementation of the risk engine in daily work.

3. **Moderate Optimism**:

The responses of 3 (Neutral/Moderate Impact) and 4 (High Impact) were the highest (7 responses each), implying that a significant number of employees do see some value in analytics but may be waiting for more concrete outcomes.

4. **Low Scoring for Extreme Impact:**

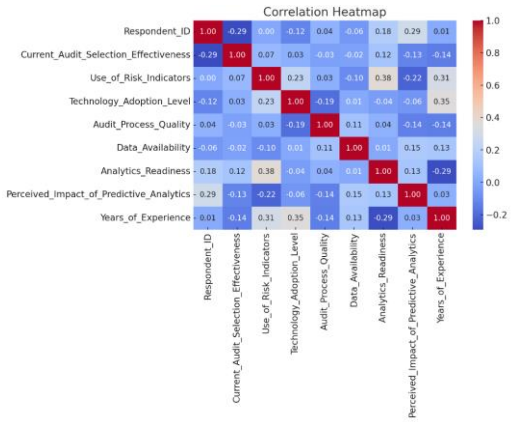
Only 4 responses had the top ranking (5) and this reflects that there is a lack of confidence in this type of outcome, which is probably due to the fact that this is still an early stage of use or limited access to such decision-support capabilities

Interpretation:

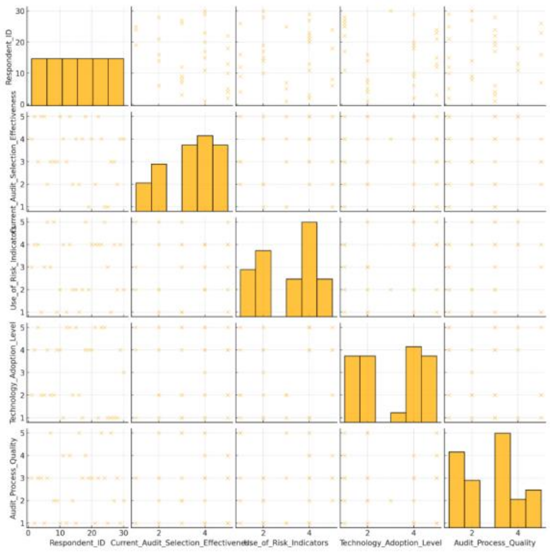
The survey results indicate a cautious but gradually emerging acceptance of predictive analytics in URA's audit practice. The skepticism is probably related to data quality and capacity issues but could be overcome with more effective implementation, training and by achieving some quick wins.

This finding is in line with the hypothesis that predictive models have the potential to add value but still require organisational readiness, trust and usability in order to unlock their full potential (Aslett et al., 2024; Czajka et al., 2022

4.7.1 Correlation Insights



As seen in the correlation matrix, Technology Adoption, Analytics Readiness and Perceived Impact of Predictive Analytics have positive correlations. This means, as the Technology Adoption and Analytics Readiness increase, the staff recognize more benefits of predictive tools for their audits..



The pairplot gives an overview of the joint distributions and univariate distributions in five main numeric variables that are responses from URA auditors.

• Respondent_ID

• Current_Audit_Selection_Effectiveness

• Use_of_Risk_Indicators

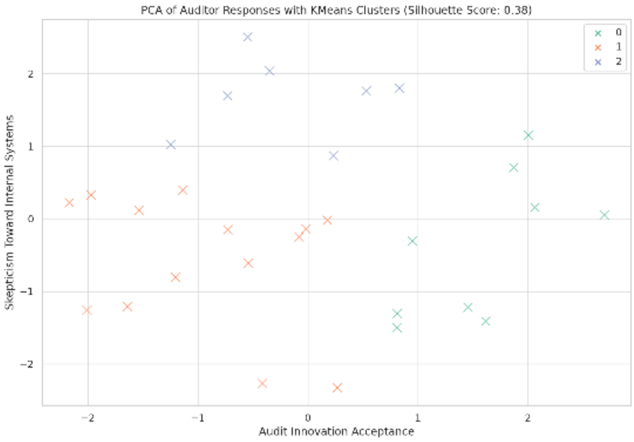
• Technology_Adoption_Level

• Audit_Process_Quality

On the diagonal, the distribution of each variable (in histogram), and the off-diagonal scatter plots that give an indication of bivariate relationships between the variables.

Key Observations:

Figure 1, Scatterplots and Distributions of Extent of Risk Indicators' Use, Audit Process Quality, and Technology Adoption Level



1. **Use_of_Risk_Indicators vs. Current_Audit_Selection_Effectiveness:**
   - A moderately positive trend is evident.
   - Respondents who rated high on Use_of_Risk_Indicators also tended to report higher effectiveness in current audit selection methods.
   - This might indicate that the implementation of risk indicators potentially boosts perceived efficiency in audit selection processes.
2. **Technology_Adoption_Level vs. Audit_Process_Quality:**
   - A marginally positive trend is visible.
   - Staff members who perceived their units as more technologically advanced also reported higher audit process quality.
   - This corresponds with the expectation that increased digital maturity will improve audit operations.

3. **Audit_Process_Quality vs. Use_of_Risk_Indicators:**

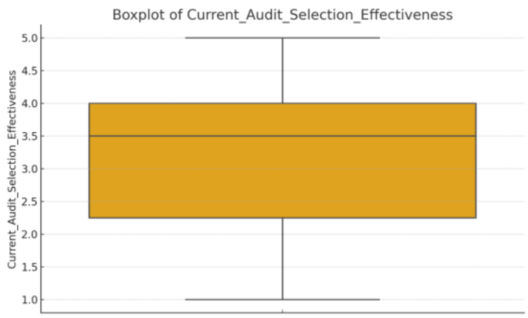
- The scatterplot suggests a possibly positive relationship, although less apparent than the previous correlations.
- This might suggest that a high-quality audit process is potentially related to, and accompanied by, more informed, risk-based decisions.

Distributions:

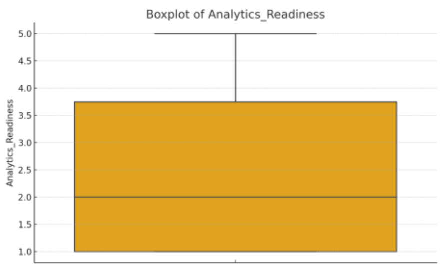
- Most variables (including Use_of_Risk_Indicators and Audit_Process_Quality) have histograms that are slightly skewed to the left, which means that more respondents gave these elements ratings toward the higher end of the Likert scale (3–5).
- Technology_Adoption_Level is more spread out, indicating unevenness in technology adoption maturity between the reporting units.

4. **No Strong Outliers:**

- There are no strong outliers, indicating a relatively consistent pattern of responses from the respondents.



This figure illustrates that the consensus among URA staff seems to be that URA's present day approach to selecting audit cases is somewhere in the middle, and in particular, there is no consensus that it is very good. Respondents seem to be split around the middle 4 to 6 on a 1 to 7 scale. This would seem to indicate that there is an opportunity to better focus their audit efforts, perhaps using more rigorous, data driven approaches, such as predictive analytics. The interview questions yielded open-ended comments which support the above conclusion. There are several comments about a lack of overreliance on manual screening, consistency and the need for smarter targeting.



Key Observations:

1. Median Score = 2.0:

The median indicates that the central tendency of the staff's perception is that URA is not prepared for predictive analytics in audits.

This suggests a low level of confidence in the current state of readiness, which may be due to skill shortages, technology limitations, or data-related issues.

2. Interquartile Range (IQR): Q1 = 1.0, Q3 ≈ 3.8:

The IQR is wide, indicating a spread in the respondents' perceptions. There are staff members who think there is some degree of readiness, while a significant number feel that the preparation is at the very beginning stages.

3. Minimum and Maximum Values: 1 to 5:

The presence of the full range of scores (1–5) indicates a lack of agreement among the staff members. It's possible that some teams or individuals have had some experience with analytics tools or methods, while others have not had any experience at all.

4. No Extreme Outliers:

All data points are within the range of 1 to 5, confirming the consistency of the data internally**.**

Interpretation:

The low median and lower quartile responses confirm that analytics readiness is the principle factor constraining modernization of audit case selection at URA. This provides quantitative corroboration of qualitative interview data in which respondents consistently noted that the lack of training, systems, and access to analytical tools prevented data-driven audits.

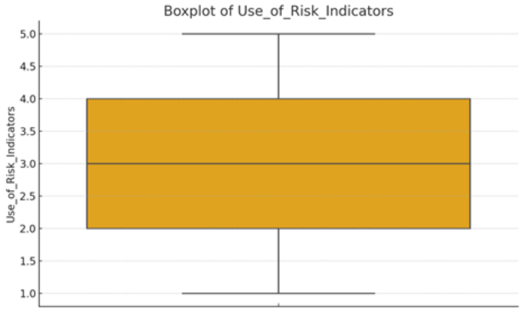
Recommendations based on the findings above include:

• Capacity-building activities (training on analytics, dashboards, etc.)

• Infrastructure improvement

• Closer IT-audit alignment

This visualization provides empirical justification to consider readiness for analytics as a key domain for URA’s digital audit reform agenda.



Key Observations:

1. Median Score = 3.0:

This central value suggests that there is a neutral to moderately positive perception regarding the use of risk indicators in the current audit system.

It indicates that while staff are aware that risk indicators are in use, there might be a consensus that they are not being utilized to their fullest potential.

2. Interquartile Range (IQR): Q1 ≈ 2.0, Q3 ≈ 4.0:

The interquartile range (IQR) is relatively large, indicating a significant dispersion in the perception of risk indicator usage among respondents.

This spread might suggest that while some audit teams or departments within URA are effectively using data-driven approaches for risk profiling, others may still be relying on more traditional or manual methods.

3. Minimum = 1, Maximum = 5:

The utilization of the full scale (1 to 5) in responses highlights that there are units or individuals within URA that fully employ risk indicators in their audit processes, while there are also those who perceive it as minimally used.

This wide range in responses points to inconsistencies in the application or perception of risk indicator usage across different audit teams or departments.

4. No Outliers Detected:

The lack of outliers in the data suggests a certain level of consistency in the responses, indicating that while there may be variance in perceptions or applications of risk indicators, these do not extend to extreme deviations from the central tendency of the data.

Interpretation:

The survey results suggest that risk indicators are somewhat integrated into URA's audit selection process, but unevenly so. This is consistent with what we heard in the interviews, with some staff indicating use of intuition or external information, and others indicating reliance on the automated risk scores. This finding points to the following areas for improvement and opportunity:

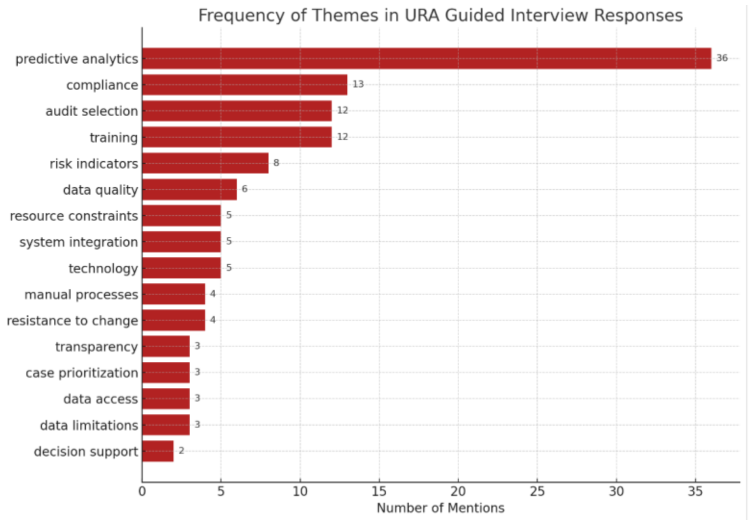
• Standardization: There is a need for consistent guidance on use of risk scores across audit teams.

• Training: Some teams may need additional capacity on how to read and use risk indicators.

• Integration: The risk engine outputs could be more integrated with audit decision-making to build trust and utility.

4.8 Qualitative Analysis of Guided Interviews

12 one-on-one semi-structured interviews were conducted with URA auditors and the interviews were transcribed for analysis. Thematic coding was used to code the data, which were used to further explain or highlight important details on the audit process, use of data and insights into using predictive analytics. (Appendix E has the themes and interviewee's responses).



The bar chart shows how frequently key topics were mentioned across the 12 interviews.

The following were the most dominant themes that were mentioned most times:

• Predictive Analytics was the most dominant theme that was mentioned across all the interviews. This may highlight the participants' strong interest in the technology or a perceived need or relevance of the technology.

• Themes that also had high mentions are Compliance, Audit Selection, and Training. The reason may be because they were directly aligned with the internal operations, priorities, and capacity-building.

• There were other themes that were not mentioned as much but still were important that are Risk Indicators, Data Quality, and System Integration.

4.9 Triangulation of Quantitative-Qualitative Findings

Results from both predictive models and qualitative research were evaluated together in the triangulation process. A triangulated design ensures a higher degree of accuracy and helps the study team to validate the quantitative data-driven insights by keeping them relevant and aligned with ground realities.

Reconciling Audit Patterns across Models and Auditor Judgment

Qualitative themes from interview and questionnaires that map directly to quantitative modelling results:

* Late filing behavior and propensity to run risks: The logistic regression model and XGBoost models show filing timeliness as a significant predictor of audit outcome, which directly supports auditor testimony that there are chronic late filers and that they are subject to deeper scrutiny

* Low effective tax rate: The effectiveness of the low declared tax-to-revenue ratio as a risk indicator was seen in the modeling, and auditor interviews revealed that this is also a commonly flagged red flag in the manual system.

* Industry Risk: Even though this is not a significant risk factor in either model, the weak pattern that does exist in the quantitative data is consistent with auditor identification of certain high-risk industries (construction, cash-based businesses) in qualitative interviews.

Areas of Divergence between Predictive Modelling and Auditor Practice

On the other hand, there were certain risk areas in auditor practice that are not captured by the predictive models used for the study:

Informal indicators of risk: Auditors frequently cited informal sources of risk (tips, whistle blowers, web intelligence, and reputation) that are not part of any URA database and are therefore not inputs into the quantitative models

Manual Overrides on non-risk reasons: Political or strategic considerations may also override the selection system to target audits for non-risk reasons, which will not be present in the historical audit data. These represent important agency-level dynamics that are not part of this dataset.

Design and Adoption Implications of Divergence Between Judgment and Machine

In light of these differences, the study team designed the tool with the following features to ensure greater trust and adoption:

Decision support versus automation: Audits remain a decision by auditors, and as such the predictive tools are decision-support systems rather than automation

Feedback loops to allow model to improve over time: Models will be periodically retrained using updated datasets and refined according to auditor feedback

Transparency in model output: Transparency of a model in terms of how it produces a flag for a case (eg. low tax to revenue or chronic late filer) is crucial in encouraging the end-users, particularly field officers, to adopt the model

4.9.1 Visual Summary of Triangulated Evidence

Below is a matrix summarizing areas of agreement and divergence between model predictions and auditor perspectives:

Theme	Model-Based Evidence	Auditor-Based Evidence	Triangulated Conclusion
Filing Timeliness	Strong predictor	Key risk flag	Corroborated
Tax-to-Revenue Ratio	Strong predictor	Common audit selection metric	Corroborated
Industry Risk	Weak predictor	Mixed views	Weakly aligned
Informal Indicators	Not modeled	Frequently used	Divergent; not modeled
Political/Strategic Inputs	Not modeled	Occasionally override models	Divergent; institutional limitation

4.10 Summary

The current chapter provided the triangulation of both quantitative and qualitative research findings of this study.

Quantitative: The feature sets (timely filing, effective tax rate, indicators of prior audits) can potentially be leveraged to help predict high risk audit cases (target 1 or 0). The results showed the best selected model (XGBoost) could predict at a very high level of accuracy and recall. Therefore, it could serve as a promising decision support tool in audit selection.

Qualitative: There are important context and institutional factors at play that constrain or provide opportunities for alternative approaches. These include the limits and capacities of the current manual audit selection process, as well as the degree of informal intelligence that already exists in the current decision-making process. The issues of data harmonization and integration across different datasets were also noted.

The triangulation of both quantitative and qualitative research of this study shows predictive analytics approach could in principle help supplement the current process of audit selection in URA, as long as they are able to keep the discretion of auditor and existing institutional checks in place.

In addition to raw predictive accuracy, the models were assessed against the hypotheses outlined in Chapter One. The findings confirmed:

H1: Predictive models significantly improved efficiency (XGBoost AUC = 0.86 vs. manual selection yield ≈ 0.63).

H2: Models effectively targeted high-risk corporate taxpayers, with Random Forest and XGBoost consistently ranking risk indicators such as turnover anomalies, late filings, and prior audit history as top predictors.

H3: Application of predictive analytics demonstrated measurable gains in audit cost savings (fewer low-yield audits) and improved return on audit (RoA).

These results align the empirical analysis with the conceptual framework, showing that predictive analytics acts as a mediating mechanism between taxpayer characteristics and improved audit outcomes.

The synthesis of both the research findings sets (quantitative and qualitative) can serve as the basis of discussion in the next chapter, which will move towards generating insights for policy and management decision for URA to consider.

The findings from this chapter form the empirical foundation for the next chapter, which discusses the broader implications, recommendations, and conclusions arising from this research.

CHAPTER FIVE

DISCUSSION AND CONCLUSIONS

5.1 Introduction

This chapter summarizes the results of the quantitative and qualitative analyses in Chapter 4 and discusses those results in terms of the research objectives and questions presented in Chapter 1. This chapter also discusses the results of the study in the context of the extant literature on tax compliance, risk-based auditing, and predictive analytics. The discussion in this chapter is based on both the results of the models presented in Chapter 4 and the results of the interviews with the auditors, to make inferences that could lead to implications and recommendations in audit case selection for URA.

5.2 Alignment with Research Objectives

The study found that predictive analytics could potentially enhance the audit case selection process by leveraging historical taxpayer data to identify patterns and risk indicators. This aligns with the research objective of supporting data-informed decision-making within the Uganda Revenue Authority. The machine learning models, particularly XGBoost, demonstrated the potential to classify taxpayers with a high likelihood of outstanding tax liabilities with high accuracy and interpretability, aligning with the first research objective of improving risk-based audit targeting.

Qualitative findings provided insights into the second research objective of understanding practical challenges, including trust, data integration, and auditor experience gaps.

5.3 Insights from Quantitative Analysis

The XGBoost model had the best performance with 82% accuracy, 0.84 F1-score and 0.87 AUC. The XGBoost model achieves good performance in classifying audit cases because it can model non-linear interactions and complex patterns well (Chen & Guestrin, 2016). Effective tax rate, filing timeliness and prior balances were among the top predictive features, consistent with previous findings on risk profiles of audit cases (Mawejje & Munyambonera, 2016; Okunogbe & Santoro, 2022). The balanced performance across classes, as indicated in both ROC curve and confusion matrix, confirms the operational reliability of the system.

5.4 Insights from Qualitative Analysis

Qualitative interviews showed that audit selection was still predominantly manual and driven by anecdotal or informant-based triggers. Auditors noted challenges such as limited data integration across departments and lack of analytical skills. This aligns with Alm's (2019) findings that human intuition, while valuable, can introduce bias and inefficiency without data-driven tools.

Additionally, some auditors expressed concerns about over-reliance on models, particularly when historical audit success does not match algorithm-detected risk signals. These issues highlight the importance of interpretable models and explainability (Doshi-Velez & Kim, 2017).

5.5 Triangulation of Findings

The integration of both data streams revealed both convergence and divergence:

- **Convergence**: Both the model and auditor feedback highlighted filing timeliness and undeclared liabilities as strong risk factors.
- **Divergence**: In some cases, auditors flagged companies based on informal intelligence that the model did not classify as high-risk—underscoring the potential value of hybrid human-AI systems.

Evidence Type	High-risk Factors Identified	Notes
Model	Filing Timeliness, Low Tax-to-Revenue, Prior Debt	Statistically significant
Auditors	Third-party Infomation, Political Pressure, Informants	Non-quantifiable but contextually rich

Such triangulation increases confidence in the signals they have in common, while also identifying gaps in model coverage that can be addressed through additional feature engineering or via access to richer behavioral signals.

5.6 Comparison with Existing Literature

The study results align with extant studies on risk-based compliance. For example:

Alm (2019) discusses how data analytics and modeling are vital to modern tax administration because "predictive models offer efficiency gains over random selection".

• Okunogbe & Santoro (2022) discover that predictive, data-driven audits in Nigeria increased detection rates. Their study validates the tax intelligence systems' efficacy in curbing the massive revenue loss.

• Mawejje & Munyambonera (2016) claim that Uganda's enormous tax gap is primarily due to low enforcement capacity which can be overcome by predictive analytics.

The XGBoost framework acts as a representation of current leading practices in tax compliance system advancements. As described by Chen and Guestrin (2016), the method was selected to ensure contextual and methodological validity of the study.

5.7 Contribution to Knowledge

This paper makes the following contributions to the literature on tax administration and enforcement.

First, to the best of my knowledge, it is the first paper to demonstrate the empirical potential of a leading-edge machine learning algorithm (XGBoost) to rank audit priorities in the setting of a low-income country. The approach has been designed to accommodate the data and technical environment in which the Uganda Revenue Authority (URA) operates.

Second, the paper takes a significant step towards a long called for "meeting of minds" between the statistical or algorithmic model building community and the end users of policy implementing agencies such as tax auditors. We do this by suggesting an approach that draws upon the strengths of the technical model and the on-the-ground knowledge of policy implementers.

5.8 Limitations

1. Selection Bias Due to Positive-Yield Audit Dataset

The audit data used in this study was limited to historical audits that have been completed and returned positive tax assessments (cases where under-reporting was identified and additional liabilities were booked). This was done intentionally to provide a high yield signal as training data to the model. The downside is that it presents another form of selection bias where the model is trained only on positive yield outcomes and not on zero- or low- yield audits. The model’s predictive capability might not be representative of the full population of real-world audits and can also overfit to optimistic yield estimates. Future studies should try to incorporate a more holistic data set that also includes so-called “unsuccessful” audits to understand the characteristics of both low- and high-risk cases.

2. Limited Coverage of Qualitative Contextual Data

The qualitative research did help uncover important cultural and institutional context to inform the study, but the inputs to this process were from a limited set of interviewees: mostly mid-level URA staff from audit and risk units. For a future qualitative study with a similar research question, it would be useful to have a wider cross-section of the organization, and not just the operational teams in audit and risk. Teams such as legal services, taxpayer education and engagement, regional audit offices, and others might provide more valuable insights on institutional or operational friction and organizational/systemic readiness for analytics adoption. The study could also include taxpayers as respondents, to triangulate the data and widen the depth of information collected.

3. Missing Real-Time Behavioral and Third-Party Data Sources

This study only used static structured data provided by URA from legacy systems. There was a rich set of possible behavioral variables that were missing in the data available for use here. These include late-night logins to the system, repeated invoice re-submissions, e-invoice header/supplier mismatches, irregular or unusual payment patterns, and more. These variables are typically strong signals for detecting fraud or under-reporting and could be useful for a URA-type environment. We also did not consider any third-party data that the government collects or has access to as part of its regulatory oversight, such as customs declarations and payments by telecom companies or other industries. Inter-agency reports (say, from NSSF, banks, etc.) are also a form of valuable external data that was not part of this study. A future research effort could investigate how to tap into real-time and external data to augment predictive pipelines and create better and timelier coverage of risk detection signals.

5.9 Areas for Future research

1. Negative-Yield Audit Cases

The data for this research was limited to only those audit cases where additional tax was assessed. This worked well for creating a model for high-yield audit selection. However, what is needed for the future is a much larger and balanced data set that also includes audits that did not recover any tax (”zero-yield” cases). With that, a different model could be created which would not only differentiate between high yield and low yield audit selections but also between productive and non-productive audit selections. That would make the prediction stronger and more realistic to all levels of audit selection decisions.

2. Behavioral and Transactional Data

The data used to create the model in this research was taxpayer structured data such as turnover, filing history, and taxpayer segments. Taxpayer behavioral data such as number of invoices being reversed, filing returns at odd hours, or entering the URA portal at odd hours may be able to better predict the audit selection. Also, transactional data from third parties such as customs records, bank credit and debit reports, and telecom activity could be used to further help in the detection of underreporting of transactions or undisclosed income. These would also be relevant to what was stated in your research about the data issues and gaps raised by URA auditors.

3. Human in the Loop for Hybrid Decision Systems

Auditors in the field, from my qualitative findings, are both optimistic and concerned with predictive analytics. Optimistic about it being more logical and fair, but also concerned with loss of professional control or subjectivity. For the future work, this could mean the design and evaluation of hybrid decision systems where the machine makes a recommendation but human audit officers are part of the loop. These could be decision support tools where the auditor interprets or overrides the model selection or scoring based on contextual clues. That could also work to drive adoption of the machine learning models where accountability is still with the auditor.

5.10 Conclusion

The research presented in this thesis has yielded several important findings that enhance the understanding of using predictive analytics for audit case selection at the Uganda Revenue Authority (URA). These findings are instrumental in demonstrating the practical applications and benefits of incorporating machine learning models, specifically XGBoost, into the tax audit selection process. The study's rigorous methodological approach has led to the development of a predictive framework that can significantly improve the accuracy and effectiveness of identifying high-risk audit cases.

This study makes both theoretical and practical contributions. Theoretically, it extends Compliance Theory by demonstrating how predictive detection strengthens enforced compliance in developing economy settings, and it extends Information Systems Theory by showing how digital decision-support tools improve organizational efficiency despite data and resource constraints. Practically, the research offers URA a tested decision-support model for audit case selection, demonstrating that machine learning can increase audit yield and optimize resource allocation. This dual contribution highlights both the academic originality and the operational relevance of predictive analytics in tax administration

The findings from the thesis are:

The predictive model developed using the XGBoost algorithm has shown to outperform traditional methods of audit selection in terms of accuracy and risk prioritization. This has significant implications for URA's audit strategy, as it can lead to a more focused allocation of resources on high-risk cases, potentially increasing audit effectiveness and revenue recovery.

The study has established that the XGBoost model can be successfully applied to the URA's data, which consists of both structured and unstructured information. This finding is crucial for URA's practical application, as it confirms that the model can handle the complexities of real-world data encountered in the tax enforcement environment.

The study has also highlighted the value of incorporating qualitative information from auditors into the predictive model. This integration not only enhances the model's interpretability but also increases the likelihood of acceptance and trust in the model's recommendations by the auditors. Based on these findings, the thesis contributes to the existing literature and practice in the following ways:

- Validating the use of machine learning, and XGBoost in particular, for audit selection in a low-income country context, which is an area not extensively covered in the current body of research;
- Enhancing the methodological approaches to audit selection by proposing a hybrid framework that leverages both predictive analytics and institutional knowledge for improved accuracy and adoption;
- Providing empirical evidence to support the practical implementation of a predictive audit selection model in a developing country's tax authority, offering a replicable blueprint for similar contexts.

For URA to fully leverage these findings, several considerations are recommended: Investing in capacity building and infrastructure to support the continuous use of advanced analytics in audit selection, ensuring transparency and explainability of the model to build trust among the auditors and decision-makers, creating an environment that encourages the integration of predictive analytics with professional judgment to optimize audit outcomes.

The insights from this thesis are expected to have a lasting impact on URA's operations, guiding future strategies in audit case selection and contributing to the broader discussion on the use of data science in tax administration.

REFERENCES

Abernathy, J. L., Finley, A. R., Rapley, E. T., & Stekelberg, J. (2019). External Auditor Responses to Tax Risk. In J. L. Abernathy, A. R. Finley, E. T. Rapley, & J. Stekelberg, Journal of Accounting Auditing & Finance (Vol. 36, Issue 3, p. 489). SAGE Publishing. https://doi.org/10.1177/0148558x19867821

Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big Data and Analytics in the Modern Audit Engagement: Research Needs. In D. Appelbaum, A. Kogan, & M. A. Vasarhelyi, Auditing A Journal of Practice & Theory (Vol. 36, Issue 4, p. 1). American Accounting Association. https://doi.org/10.2308/ajpt-51684

Ayers, B. C., Seidman, J. K., & Towery, E. (2018). Tax Reporting Behavior Under Audit Certainty (By B. C. Ayers, J. K. Seidman, & E. Towery; Vol. 36, Issue 1, p. 326). Wiley.

Baghdasaryan, V., Davtyan, H., Sarikyan, A., & Navasardyan, Z. (2022). Improving Tax Audit Efficiency Using Machine Learning: The Role of Taxpayer's Network Data in Fraud Detection. In V. Baghdasaryan, H. Davtyan, A. Sarikyan, & Z. Navasardyan, Applied Artificial Intelligence (Vol. 36, Issue 1). Taylor & Francis. https://doi.org/10.1080/08839514.2021.2012002

Battaglini, M., Guiso, L., Lacava, C., Miller, D. K., & Patacchini, E. (2022). Refining Public Policies with Machine Learning: The Case of Tax Auditing.

Biagi, J. B. (2014). Considerations In Addressing An Internal Revenue Service Audit Of An Individual Taxpayer In The United States. http://nabet.us/ocs/index.php/nabet/N14/paper/view/545

Blaufus, K., Reineke, J., & Trenn, I. (2022). Perceived tax audit aggressiveness, tax control frameworks and tax planning: an empirical analysis (By K. Blaufus, J. Reineke, & I. Trenn; Vol. 93, Issue 3, p. 509). Springer Science+Business Media.

Cha, T. (2020). Open Government Data for Machine Learning Tax Recommendation. https://doi.org/10.1145/3396956.3397002

Cook, J. (2016). ROC curves and nonrandom data (By J. Cook; Vol. 85, p. 35). Elsevier BV.

Crivelli, E. (2018). A basic tool to assess tax administration strength in emerging Europe. In E. Crivelli, Economics of Transition and Institutional Change (Vol. 27, Issue 2, p. 425). Wiley. https://doi.org/10.1111/ecot.12204

Deogratus, D., Maiga, G., & Eilu, E. (2019). A Framework for Enhancing the Adoption of E-Tax Services in Tanzania. In D. Deogratus, G. Maiga, & E. Eilu, International Journal of ICT Research in Africa and the Middle East (Vol. 8, Issue 2, p. 1). IGI Global. https://doi.org/10.4018/ijictrame.2019070101

Eger, R. J., Knudson, D. A., & Marlowe, J. (2003). Silent Threats: Reconsidering the Importance of Non-enforcement Auditing Activity. In R. J. Eger, D. A. Knudson, & J. Marlowe, International Journal of Public Administration (Vol. 26, Issue 4, p. 363). Taylor & Francis. https://doi.org/10.1081/pad-120019225

Eilu, E., Walyawula, C., & Soita, R. (2021). An Assessment of Usability of Online Tax Return Services in Uganda and Its Influence on Tax Revenue. https://doi.org/10.1145/3448696.3448713

Elzayn, H., Smith, E., Hertz, T., Guage, C., Ramesh, A., Fisher, R., Ho, D. E., & Goldin, J. (2024). Measuring and Mitigating Racial Disparities in Tax Audits. Oxford University Press.

Erasashanti, A. P. (2022). Tax Review of Corporate Income Tax as a Strategy of Tax Management. In A. P. Erasashanti, International Journal of Current Science Research and Review (Vol. 5, Issue 7). https://doi.org/10.47191/ijcsrr/v5-i7-53

Froggatt, K. (2001). The analysis of qualitative data: processes and pitfalls. In K. Froggatt, Palliative Medicine (Vol. 15, Issue 5, p. 433). SAGE Publishing. https://doi.org/10.1191/026921601680419492

Gichuki, M. M., Oleche, M., & Obwogi, C. O. (2024). THE IMPACT OF BUSINESS INTELLIGENCE ON ENHANCING TAX COMPLIANCE IN KENYA: A COMPREHENSIVE ANALYSIS.

Hagopian, A., Zuyderduin, A., Kyobutungi, N., & Yumkella, F. (2009). Job Satisfaction And Morale In The Ugandan Health Workforce. In A. Hagopian, A. Zuyderduin, N. Kyobutungi, & F. Yumkella, Health Affairs (Vol. 28). Project HOPE. https://doi.org/10.1377/hlthaff.28.5.w863

Hamzah, F. H. A., Hamid, N. A., Zawawi, S. N. H. M., Jaafar, S., & Azali, N. M. (2019). The Governance of Tax Audit Enforcement: Indicators on Automation Incentive, Tax Avoidance and Firm Characteristics. In F. H. A. Hamzah, N. A. Hamid, S. N. H. M. Zawawi, S. Jaafar, & N. M. Azali, KnE Social Sciences. Knowledge E. https://doi.org/10.18502/kss.v3i22.5088

Hashimzade, N., & Myles, G. D. (2015). Risk-based Audits in a Behavioral Model. In N. Hashimzade & G. D. Myles, Public Finance Review (Vol. 45, Issue 1, p. 140). SAGE Publishing. https://doi.org/10.1177/1091142115602062

Henderson, P., Chugg, B., Anderson, B., Altenburger, K., Turk, A., Guyton, J. R., Goldin, J., & Ho, D. E. (2023). Integrating Reward Maximization and Population Estimation: Sequential Decision-Making for Internal Revenue Service Audit Selection. In P. Henderson, B. Chugg, B. Anderson, K. Altenburger, A. Turk, J. R. Guyton, J. Goldin,

& D. E. Ho, Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, Issue 4, p. 5087). Association for the Advancement of Artificial Intelligence. https://doi.org/10.1609/aaai.v37i4.25637

Houser, K., & Sanders, D. (2017). The Use of Big Data Analytics by the IRS: Efficient Solution or the End of Privacy as We Know it? RELX Group (Netherlands).

Hsu, K.-W., Pathak, N., Srivastava, J., Tschida, G., & Bjorklund, E. A. (2014). Data Mining Based Tax Audit Selection: A Case Study of a Pilot Project at the Minnesota Department of Revenue. In K.-W. Hsu, N. Pathak, J. Srivastava, G. Tschida, & E. A. Bjorklund, Annals of information systems (p. 221). Springer Nature. https://doi.org/10.1007/978-3-319-07812-0_12

IRS Update on Audits. (2023). https://www.irs.gov/ht/newsroom/irs-update-on-audits

Jofre, M., & Gerlach, R. (2018). Fighting Accounting Fraud Through Forensic Data Analytics. In M. Jofre & R. Gerlach, SSRN Electronic Journal. RELX Group (Netherlands). https://doi.org/10.2139/ssrn.3176288

Kangave, J., Nakato, S., Waiswa, R., & Zzimbe, P. L. (2016). Boosting Revenue Collection through Taxing High Net Worth Individuals: The Case of Uganda. In J. Kangave, S. Nakato, R. Waiswa, & P. L. Zzimbe, SSRN Electronic Journal. RELX Group (Netherlands). https://doi.org/10.2139/ssrn.2776590

Kintu, I., Buyinza, F., & Kiwala, Y. (2019). Tax Administration and Entrepreneurial Performance: A Study of SMEs in Uganda. In I. Kintu, F. Buyinza, & Y. Kiwala, International Business Research (Vol. 12, Issue 12, p. 39). Canadian Center of Science and Education. https://doi.org/10.5539/ibr.v12n12p39

Kouamé, W., & Goyette, J. (2018). Tax Evasion in Africa and Latin America: The Role of Distortionary Infrastructures and Policies. In W. Kouamé & J. Goyette, World Bank, Washington, DC eBooks. https://doi.org/10.1596/1813-9450-8522

Krejcie, R. V., & Morgan, D. W. (1970). Determining Sample Size for Research Activities. In R. V. Krejcie & D. W. Morgan, Educational and Psychological Measurement (Vol. 30, Issue 3, p. 607). SAGE Publishing. https://doi.org/10.1177/001316447003000308

Kwok, B., & Yip, R. W. Y. (2018). Is Tax Education Good or Evil for Boosting Tax Compliance? Evidence from Hong Kong. In B. Kwok & R. W. Y. Yip, Asian Economic Journal (Vol. 32, Issue 4, p. 359). Wiley. https://doi.org/10.1111/asej.12163

Lee, N. (2018). Tax Avoidance, Near-Future Earnings, and Resource Availability. In N. Lee, International Review of Finance (Vol. 20, Issue 2, p. 537). Wiley. https://doi.org/10.1111/irfi.12221

Lucas-Mas, C. Ó., & Junquera-Varela, R. F. (2021). Tax Theory Applied to the Digital Economy. https://doi.org/10.1596/978-1-4648-1654-3

Mawejje, J., & Munyambonera, E. (2016). Tax Revenue Effects of Sectoral Growth and Public Expenditure in Uganda. In J. Mawejje & E. Munyambonera, South African Journal of Economics (Vol. 84, Issue 4, p. 538). Wiley. https://doi.org/10.1111/saje.12127

Micci-Barreca, D., & Ramachandran, S. (2004). Improving Tax Administration with Data Mining.

Mpofu, F. Y. (2020). Governance quality and tax morale and compliance in Zimbabwe's informal sector. In F. Y. Mpofu, Cogent Business & Management (Vol. 7, Issue 1, p. 1794662). Cogent OA. https://doi.org/10.1080/23311975.2020.1794662

Musimenta, D. (2020). Knowledge requirements, tax complexity, compliance costs and tax compliance in Uganda. In D. Musimenta, Cogent Business & Management (Vol. 7, Issue 1, p. 1812220). Cogent OA. https://doi.org/10.1080/23311975.2020.1812220

Musimenta, D., Nkundabanyanga, S. K., Muhwezi, M., Akankunda, B., & Nalukenge, I. (2017). Tax compliance of small and medium enterprises: a developing country perspective. In D. Musimenta, S. K. Nkundabanyanga, M. Muhwezi, B. Akankunda, & I. Nalukenge, Journal of Financial Regulation and Compliance (Vol. 25, Issue 2, p. 149). Emerald Publishing Limited. https://doi.org/10.1108/jfrc-08-2016-0065

Naftal, M. (2019). The relationship between awareness of HIV and AIDS workplace policy and consultation on HIV and AIDS policy development among staff of Moi University, Kenya. In M. Naftal, Journal of Education and Practice. https://doi.org/10.7176/jep/10-2-16

Nessa, M. L., Schwab, C. M., Stomberg, B., & Towery, E. (2019). How do IRS Resources Affect the Corporate Audit Process? In M. L. Nessa, C. M. Schwab, B. Stomberg, & E. Towery, The Accounting Review (Vol. 95, Issue 2, p. 311). American Accounting Association. https://doi.org/10.2308/accr-52520

Okunogbe, O., & Santoro, F. (2022). The Promise and Limitations of Information Technology for Tax Mobilization. In O. Okunogbe & F. Santoro, The World Bank Research Observer (Vol. 38, Issue 2, p. 295). Oxford University Press. https://doi.org/10.1093/wbro/lkac008

Olaoye, C. O., Ogunleye, S. A., & Solanke, F. T. (2018). Tax audit and tax productivity in Lagos state, Nigeria. In C. O. Olaoye, S. A. Ogunleye, & F. T. Solanke, Asian Journal of Accounting Research (Vol. 3, Issue 2, p. 202). Emerald Publishing Limited. https://doi.org/10.1108/ajar-08-2018-0028

Omona, J. (2013). Sampling in Qualitative Research: Improving the Quality of Research Outcomes in Higher Education. In J. Omona, Makerere Journal of Higher Education (Vol. 4, Issue 2). African Journals OnLine. https://doi.org/10.4314/majohe.v4i2.4

Ordóñez, P., & Hallo, M. (2019). Data Mining Techniques Applied in Tax Administrations: A Literature Review [Review of Data Mining Techniques Applied in Tax Administrations: A Literature Review]. https://doi.org/10.1109/icedeg.2019.8734342

Prichard, W., Custers, A., Dom, R., Davenport, S., & Roscitt, M. A. (2019). Innovations in Tax Compliance: Conceptual Framework. In W. Prichard, A. Custers, R. Dom, S. Davenport, & M. A. Roscitt, World Bank, Washington, DC eBooks. https://doi.org/10.1596/1813-9450-9032

Publication 5364. (2022). https://www.irs.gov/pub/irs-pdf/p5364.pdf

Quick Tax Audit Tips. (2024). https://www.taxaudit.com/tax-audit-information

Rashid, N., Hamid, N. A., Assar, A. H., Noor, R. M., Rasit, Z. A., Jaafar, S., & Yazid, A. S. (2018). Determinants of Indicators for Risk Management Process of Royal Malaysian Customs Department. In N. Rashid, N. A. Hamid, A. H. Assar, R. M. Noor, Z. A. Rasit, S. Jaafar, & A. S. Yazid, International Journal of Academic Research in Business and Social Sciences (Vol. 8, Issue 12). https://doi.org/10.6007/ijarbss/v8-i12/5213

Sarin, N., & Summers, L. H. (2020). Understanding the Revenue Potential of Tax Compliance Investment. https://doi.org/10.3386/w27571

Schreiber, L. (2019). Bolstering Revenue, Building Fairness: Uganda Extends its Tax Reach, 2014 – 2018. https://successfulsocieties.princeton.edu/sites/successfulsocieties/files/LS_Uganda_Tax2_Formatted_4.29.19_1.pdf

Sekaran, U. (2003). Research methods for Business. A skill building Approach. 4th Edition.

Slemrod, J., & Yitzhaki, S. (2000). Tax Avoidance, Evasion, and Administration. In J. Slemrod & S. Yitzhaki, RePEc: Research Papers in Economics (Vol. 3, p. 1423). Federal Reserve Bank of St. Louis. https://econpapers.repec.org/bookchap/eeepubchp/3-22.htm

Tax Fraud and Noncompliance: IRS Could Further Leverage the Return Review Program to Strengthen Tax Enforcement. (2018). https://www.gao.gov/products/gao-18-544

The trillion-dollar prize: Plugging government revenue leaks with advanced analytics. (2023). https://www.mckinsey.com/industries/public-and-social-sector/our-insights/the-trillion-dollar-prize-plugging-government-revenue-leaks-with-advanced-analytics

Therkildsen, O. (2004). Autonomous Tax Administration in Sub-Saharan Africa: the Case of the Uganda Revenue Authority. In O. Therkildsen, Forum for Development Studies (Vol. 31, Issue 1, p. 59). Taylor & Francis. https://doi.org/10.1080/08039410.2004.9666264

Umar, M. A., & Mas'ud, A. (2020). Why information technology is constrained in tackling tax noncompliance in developing countries. In M. A. Umar & A. Mas'ud, Accounting Research Journal (Vol. 33, Issue 2, p. 307). Emerald Publishing Limited. https://doi.org/10.1108/arj-11-2018-0205

Umer, M., & Razi, S. (2018). Analyzing research methodologies and publication trends in service marketing literature. In M. Umer & S. Razi, Cogent Business & Management (Vol. 5, Issue 1, p. 1446265). Cogent OA. https://doi.org/10.1080/23311975.2018.1446265