ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct097

Forecast Tax Non-Compliance in Rwanda Using Predictive Models

Dushime Yvonne[1]

[1]Master of Science in Big Data Analytics, Faculty of IT, Adventist University of Central Africa

Publication Date: 2025/10/15

Abstract: Rwanda's economic development, aligned with global trends, depends heavily on tax revenue to finance critical infrastructure and public services including education, healthcare, public safety and transportation networks. These services are vital for achieving Rwanda's Vision 2050 goals of sustainable growth and poverty reduction. However,tax compliance remains a significant challenge, with a substantial portion of the population, particularly among small-scale traders and rural taxpayers failing to file or pay taxes on time. This non-compliance limits the government's ability to fund essential services and hinders Rwanda's ambition to become middle-income economy.This study investigates the potential of machine learning models to predict tax non-compliance using historical taxpayer data from the Rwanda Revenue Authority(RRA) covering 2018-2023. By leveraging regression analysis and advanced predictive models such as Logistic Regression, Random Forest, XG-Boost and Decision tree,the study aims to identify individuals or businesses at high risks of failing to file or pay taxes on time. Additionally, it seeks to pinpoint key predictors of non-compliance such as income levels, business size, sector and geographic location.

Keywords: *Tax Non-Compliance, Machine Learning, Rwanda Revenue Authority, Corporate Income Tax (CIT), Personal Income Tax (PIT), Non-Compliance Detection, Risk-Based Auditing, Data-Driven Decision Making, Tax Administration.*

How to Cite: Dushime Yvonne (2025) Forecast Tax Non-Compliance in Rwanda Using Predictive Models. *International Journal of Innovative Science and Research Technology*, 10(10), 762-769. https://doi.org/10.38124/ijisrt/25oct097

I. INTRODUCTION

The economic development of any nation largely depends on the revenues achieved in form of developments of infrastructure and social services. To effectively provide these services, governments need adequate finances, and taxation is one of the most essential ways of providing the revenues needed. Taxes are compulsory taxes individuals and corporations pay to the state, whether or not they get particular goods or services in return. These monies are strictly required for the financing of public requirements and ensuring the government operates effectively. All governmental operations for economic stabilization and development are funded through taxes. Direct and indirect taxes are a very significant source of public revenue and economic development in Rwanda. The anticipated impacts have not occurred to the general extent because a large percentage of individuals do not remit their tax contribution. Compliance by force is an inevitable alternative for the state in order to finance primary services like education, welfare, security, and infrastructure. By enhanced tax compliance, the government will have more revenues to tap in a bid to finance public services without overburdening taxpayers. Wisely, for a nation which has seen massive economic change, taxation is one of the most contentious and contested topics in Rwandan politics. Taxation in Rwanda is a mirror of the overall challenges and advancement of the country. After the genocide, the state had the daunting task of reviving the economy and rebuilding people's trust in institutions. Governments have, in the past few years, introduced blended tax policies based on their different ideologies and agendas. For example, while some governments have aimed at augmenting tax revenue through expanding the base of taxation and increasing compliance, others have placed significant emphasis on tax incentives as a means to stimulate private investment and economic growth. These contradictory policies stem from varied worldviews and philosophies towards Rwanda's economic ambitions and possibilities. With this dissertation, I will set for myself the response to the question, "How are machine learning models using historical taxpayer information to predict taxpayers' probability of not paying or filing taxes on time, and what are the most important predictors of such behavior? To achieve this, I will utilize publically accessible data from Rwanda Revenue Authority to conduct a regression-based analysis of past economic trends from 2019 to 2023. This five-year time frame captures epochal years of economic boom, adversity, and resurgence, in different modes of governance". The study will span various regions of Rwanda, and the study will account for commonly accepted determinants of economic outcomes such as various types of taxes, government expenditure, and regime changes. With as many pieces of the economic jigsaw as human beings can put together, I would like to know how historical tax policy in Rwanda has shaped its economic destiny and if machine learning algorithms can forecast tax compliance behavior. Taxation has always been

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct097

the backbone of economic policy and governmental income. For Rwanda, a nation that has experienced enormous turbulence in its economic and political environment, accurate collection and administration of taxes is the backbone of its future development and growth. In the wake of the genocidal destruction of 1994, Rwanda embarked on a multidimensional initiative aimed at rebuilding its economy, institutions, and society. Tax policy and administration have also been at the core of this recovery and development process.

II. LITERATURE REVIEW

Several recent empirical studies have shaped the contemporary understanding of tax compliance and its predictors in developing contexts. Mascagni and Mengistu (2020) underscored the importance of ICT tools and public perception in Africa's tax systems, while Bird and Zolt (2019) addressed systemic weaknesses in tax systems of developing countries. Empirical evidence from Rwanda, such as Mascagni and Nell (2022) and Hakizimana and Santoro (2023), highlighted the impact of taxpayer education and technology evolution on compliance in the country's specific context.

Recent machine learning research has demonstrated significant potential for tax administration. Chen and Guestrin's (2016) XGBoost framework, though slightly older, remains foundational for current applications. More recent studies by Jayanti et al. (2024), Joseph et al. (2024), and Murorunkwere et al. (2023) illustrate the growing role of data-driven tools in detecting fraud and enhancing tax administration effectiveness. Modern analytical tools like Scikit-learn (2024) have become pivotal for implementing predictive analytics in tax compliance systems.

National perspectives from institutions like the Rwanda Revenue Authority (2023) provide essential operational context for understanding the current tax administration landscape. Collectively, these contemporary works reveal that taxpayer behavior is influenced by a combination of technological, institutional, and policy factors—and that machine learning offers promising new avenues for proactive, evidence-based tax enforcement in developing economies like Rwanda.

- *Theoretical Literature Review*

The theoretical foundation of tax non-compliance prediction within machine learning paradigms reveals significant gaps in context-specific applications, particularly regarding Rwanda's unique institutional framework. While contemporary literature demonstrates algorithmic efficacy across various jurisdictions—with recent contributions by Jayanti et al. (2024) establishing systematic frameworks for taxation applications, Joseph et al. (2024) validating enhanced revenue collection accuracy, and Murorunkwere et al.(2023) advancing supervised learning methodologies for African fraud detection—existing theoretical frameworks predominantly focus on developed economies, leaving substantial voids regarding country-specific implementations and comparative analyses of multiple algorithms (Logistic Regression,Random Forest, XGBoost and Decision Tree) within developing economies' tax administration systems. Addressing these theoretical gaps through a comprehensive comparative framework that integrates quantitative predictors (Business Income, Balance Due) and categorical variables (Taxpayer Type, Tax Category) within Rwanda's administrative context, the research achieves 88% predictive accuracy using XGBoost through advanced hyperparameter optimization, thereby extending existing theoretical boundaries and contributing empirical evidence that enhances understanding of machine learning efficacy in resource-constrained tax administration environments.

- *Empirical Literature Review*

Previous studies demonstrate machine learning's efficacy in tax compliance prediction globally, with recent research by Jayanti et al. (2024) providing systematic evidence of algorithmic applications across multiple jurisdictions, while Joseph et al. (2024) reported significant success in transforming compliance through approaches that reduce fraud and enhance revenue collection, and Murorunkwere et al. (2023) applied supervised learning in African contexts emphasizing financial indicators as critical predictors. Battaglini et al. (2024) reinforced these findings through machine learning applications to tax auditing policies, highlighting complexities involving taxpayer categories and administrative efficiency, while Chen and Liu (2023) noted behavioral complexities influenced by sentiment and administrative processes, and Zhang and Wang (2024) demonstrated how deep learning addresses financial indicators' variability affecting prediction accuracy. Studies by Antinyan and Asatryan (2023) showed effectiveness of data-driven approaches over traditional compliance methods, though Liu et al. (2023) focused on broader data mining techniques rather than country-specific algorithmic comparisons, revealing that few comparative analyses exist specifically evaluating multiple ML algorithms' efficiency in Rwanda's unique administrative and economic context.

- *Gaps in the Literature*

Current machine learning research on tax compliance prediction exhibits significant geographic and methodological limitations, with most studies focusing on developed economies while neglecting developing countries' unique administrative and cultural contexts. Rwanda-specific research remains particularly scarce, and the existing literature emphasizes qualitative factors such as taxpayer education rather than quantitative predictive modeling using advanced algorithms. The literature lacks comprehensive comparative analyses of multiple machine learning techniques (Logistic Regression, Random Forest, Decision Tree, XGBoost) within Rwanda's distinct socio-economic framework, and minimal empirical research leverages Rwanda Revenue Authority's historical data for compliance prediction. This study addresses these gaps by applying robust predictive models to historical RRA data, incorporating local variables such as taxpayer types, income levels, and tax liabilities, thereby providing Rwanda-specific empirical insights that facilitate targeted, evidence-based policy interventions and enhance tax compliance strategy

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct097

through effective resource allocation and strategic decision-making.

III. METHODOLOGY

This study used a quantitative approach to develop machine learning models that predict tax noncompliance in Rwanda. Historical data from the Rwanda Revenue Authority (2019–2023) covering CIT and PIT taxpayers was analyzed. Key variables included entity type, tax type, business income, and balance due. After data cleaning and pre-processing, four supervised learning algorithms were trained and evaluated: logistic regression, Decision Tree, Random Forest, and XGBoost using precision, precision, recall, and F1 score. The best-performing model was used to identify risk factors and support data-driven tax enforcement strategies.

IV. DATA ANALYSIS AND RESULTS

Table 1 Tax Non-Compliant Level on PIT

Late/Timely Count for Tax Non-Compliant Per Enterprise Description /PIT				
Enterprise Type	**Late**	**On-Time**	**Total**	**Percentage for Late Payment**
LOCAL NGOs	7	42	49	14.3%
COOPERATIVE	85	440	525	16.2%
INDIVIDUAL	86200	567394	653594	13.2%
JOINT VENTURE	3	4	7	42.9%
PRIVATE CORPORATION	435	1680	2115	20.6%
Total	**86781**	**569676**	**656457**	

Individuals represent the largest group of tax non-compliant entities, contributing over 86,000 late payments with a non-compliance rate of 13.2%, driven by their large population size. Private Corporations also show significant tax non-compliance, with a high rate of 20.6% across 2,115 entities, indicating a need for closer monitoring and enforcement. Cooperatives exhibit a 16.2% non-compliance rate, suggesting moderate risk that may benefit from targeted education or support. On the other hand, Companies show full compliance with a 0% non-compliance rate, while Local NGOs, Not Enterprises, and Joint Ventures, though having higher non-compliance percentages (up to 42.9%), account for very few cases and thus pose a lower overall risk.

Table 2 Tax Non-Compliant Level on CIT

Late/Timely Count for Tax Non-Compliant Per Enterprise Description /CIT				
Enterprise Type	**Late**	**On-Time**	**Total**	**Percentage for Late Payment**
LOCAL NGOs	285	1959	2244	12.7%
ASSOCIATION	7	64	71	9.9%
COOPERATIVE	2016	17429	19445	10.4%
GOVERNMENT BODY	10	59	69	14.5%
INDIVIDUAL	127	775	902	14.1%
INTERNATIONAL NGOs	10	37	47	21.3%
JOINT VENTURE	63	700	763	8.3%
NON GOV. ORG.	3	67	70	4.3%
OTHERS	2	21	23	8.7%
PARTNERSHIP		2	2	0.0%
PRIVATE CORPORATION	55936	467844	523780	10.7%
PUBLIC CORPORATION	3	4	7	42.9%
Total	**58463**	**488980**	**547443**	

The level of corporate Income Tax (CIT) non-compliance based on late payments across different enterprise types. The group with the highest tax non-compliance rate is Public Corporations, with 42.86% of their filings made late, though they represent only 7 cases overall—suggesting low volume but high risk. International NGOs (21.28%), Individuals (14.08%), and Government Bodies (14.49%) also display notable non-compliance rates and deserve targeted oversight, especially Individuals, who account for 902 cases, including 127 late payments.

Private Corporations, with a 10.68% non-compliance rate, are especially critical as they represent the largest volume of total records (523,780) and contribute the highest number of late cases (55,936). Although Cooperatives (10.37%) and Local NGOs (12.7%) have moderate non-compliance, they handle substantial volumes, warranting preventive engagement. In contrast, Companies (5.56%), Associations (9.86%), Joint Ventures (8.26%), and especially Not Enterprises and Partnerships (0.00%), show strong compliance and pose a lower risk.

A. Model Building

➢ *Model 1: Logistic Regression*

Logistic regression estimates the probability that a taxpayer is compliant by applying the logistic (sigmoid) function to a linear combination of the independent variables:

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct097

- Purpose: Predict categorical outcomes (mostly binary, like 0 or 1). But for our case it's either compliant or non-compliant

- *Formula:*

$$P(y=1)=\frac{1}{1+e^{-(\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_nx_n)}}$$

- Explanation: The formula gives a probability (between 0 and 1) that the output is class 1. The expression inside the exponent is the same linear combination used in linear regression. The logistic (sigmoid) function squashes the output to the range [0, 1].

➢ *Decision Tree*

A Decision Tree splits the data into subsets based on the value of input features. It uses criteria like Gini impurity or entropy to determine the best split at each node. The process continues recursively until a stopping condition is met (e.g., maximum depth or minimum number of samples per leaf).

$$\text{Gini} = 1 - \Sigma_{i=1}^{c} (p_i^2)$$

Where p_i is the proportion of class i at a node, and C is the total number of classes. The Gini impurity measures how often a randomly chosen element from the set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset. A Gini value of 0 indicates perfect purity (only one class is present), while higher values indicate more impurity.

➢ *Random Forest*

Random Forest builds on the concept of decision trees, but instead of relying on just one tree, it creates an entire "forest" of them. Each tree sees a different slice of the data and makes its own prediction, and then the forest takes a vote to decide the final result. This approach improves accuracy and reduces the risk of overfitting—a problem where a single decision tree might become too tailored to the data it was trained on. For the Rwanda Revenue Authority, this means getting a more reliable prediction about taxpayer compliance, even with varied taxpayer profiles and tax types. In this project, Random Forest proved valuable for detecting patterns that aren't immediately obvious, and for handling the diversity in Rwanda's taxpayer base

$$\hat{y} = \text{mode}\ \{h_1(x), h_2(x), ..., h_t(x)\}$$

Where $h_t(x)$ is the predicted class label by the t-th decision tree, and T is the total number of trees in the forest.

➢ *XGBoost*

XGBoost stands for "Extreme Gradient Boosting," and while the name might sound technical, the idea is simple: it builds a series of small, smart decision trees—each one learning from the mistakes of the last. Over time, it gets better and better at making accurate predictions. XGBoost is known for its high performance and speed, especially with large datasets like the one used in this study. Among all the models tested, XGBoost delivered the highest accuracy when predicting whether taxpayers were likely to comply or not. Its ability to capture complex patterns makes it ideal for helping the RRA prioritize audits, send reminders, or design targeted interventions. In short, XGBoost doesn't just predict who might fail to file or pay—it helps explain why, and does so efficiently.

$$L = \Sigma_{i=1}^{n} l(y_i, \hat{y}_i) + \Sigma_{k=1}^{K} \Omega(f_k)$$

Where:

$l(y_i, \hat{y}_i)$: is the loss function (e.g., logistic loss, squared error)

$\Omega(f_k)$: is the regularization term to prevent overfitting

Where:

$$\Omega(f) = \gamma T + (1/2)\lambda\ \Sigma_{j=1}^{T} w_j^2$$

Where:

T: number of leaves in the tree

w_j: weight (score) of leaf j

γ, λ: regularization parameters controlling model complexity

- *Accuracy*

$$\text{Accuracy}=\frac{\text{TP + TN}}{\text{TP + TN + FP + FN}}$$

Accuracy measures the overall correctness of the model by evaluating how many predictions were correct out of all predictions made. It is most useful when classes are balanced.

- *Precision*

$$Precision=\frac{TP}{\text{TP + FP}}$$

Precision evaluates the accuracy of the positive predictions. It tells us how many of the predicted positive cases were actually positive.

- *Recall*

$$Recall=\frac{\text{TP}}{\text{TP + FN}}$$

Recall measures the model's ability to detect all actual positive cases. It is crucial when missing a positive instance is more critical than having false positives.

- *F1 Score*

$$F1\ score=\frac{2*(\text{Precision}*\text{Recall})}{(\text{Precision}+\text{Recall})}$$

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct097

F1 Score is the harmonic mean of precision and recall. It is useful when you need a balance between precision and recall, especially in cases of uneven class distribution.

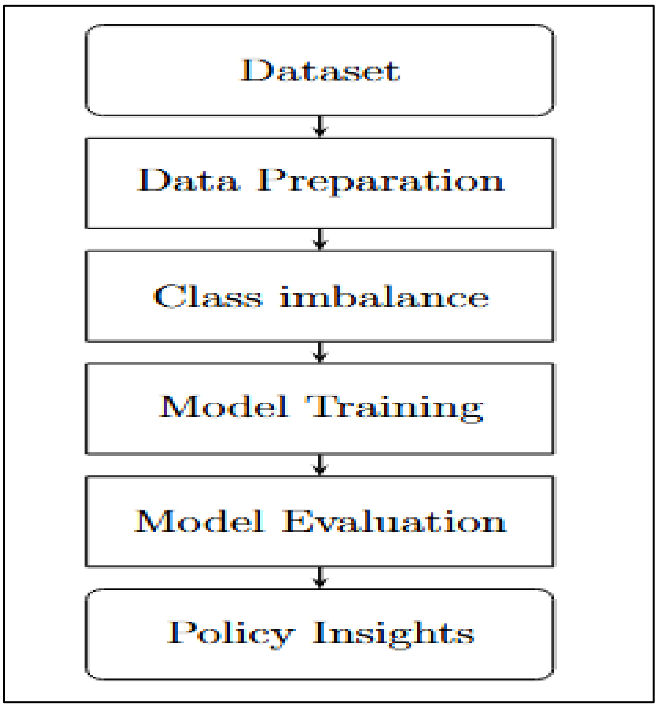

Fig 1 Simplified Conceptual Framework for Employment Prediction

B. Model Results

This section presents the results obtained from applying various machine learning models to the dataset. The models evaluated include Logistic Regression, Random Forest, XGBoost, and Decision Tree classifiers. Each model was trained and tested on the same dataset, and performance was measured using standard classification metrics: Accuracy, Precision, Recall, F1 Score, and the Area Under the Curve (AUC). The goal is to identify the model that offers the most reliable and balanced predictive performance for the tax compliance classification task.

C. Model Comparison

➤ *Exploration Analysis*

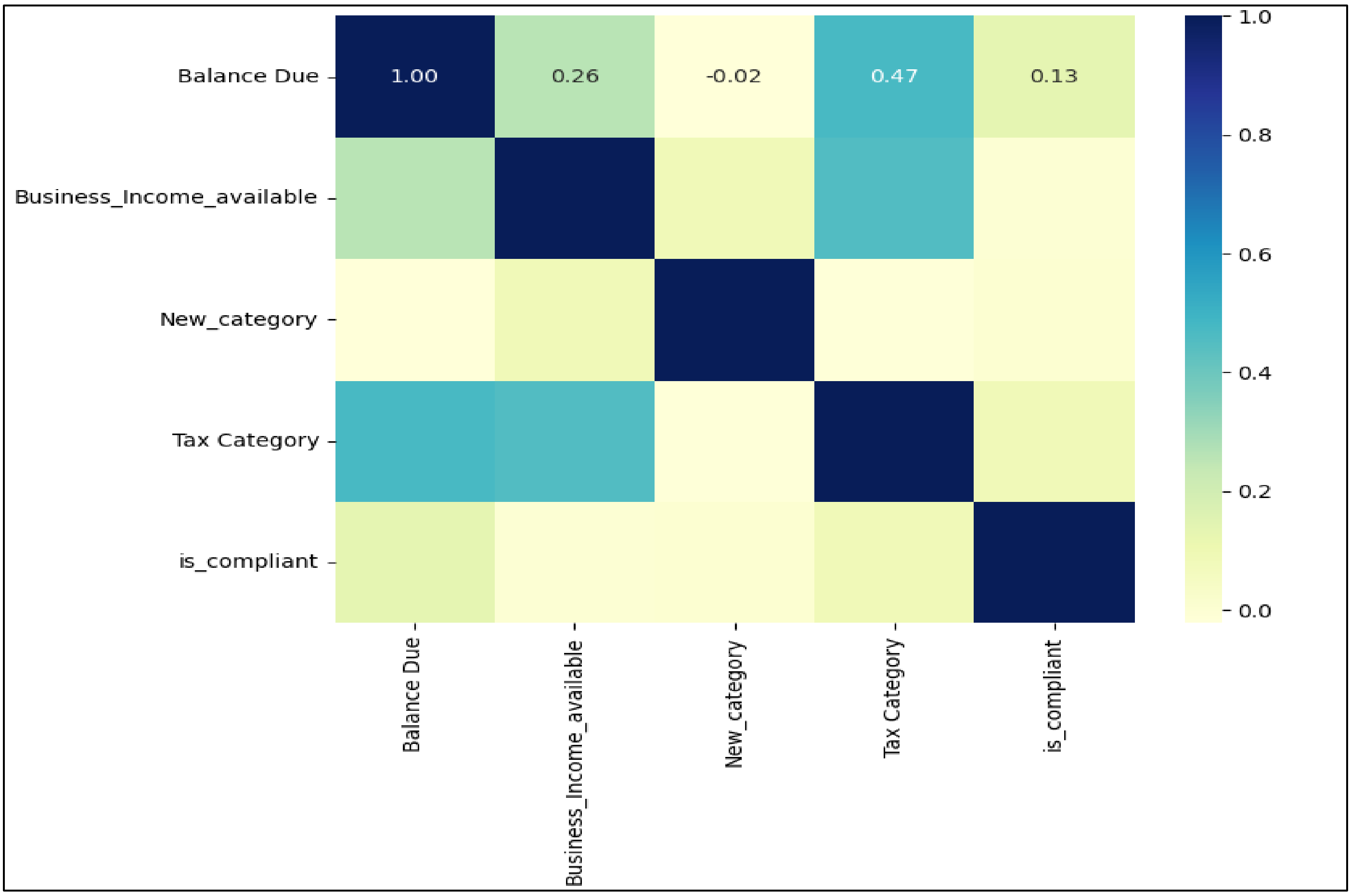

Fig 2 Correlation Matrix

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct097

The correlation matrix provides insights into the linear relationships between the variables in the dataset. Each cell represents the Pearson correlation coefficient between two variables, ranging from -1 to 1. A value close to 1 indicates a strong positive correlation, while a value near -1 indicates a strong negative correlation. Values around 0 suggest little to no linear relationship. As expected, each variable shows a perfect correlation (1.00) with itself along the diagonal. The variable "Balance Due" shows a moderate positive correlation (0.47) with "Tax Category", implying that higher balances are somewhat associated with certain tax categories. It also has a weak correlation (0.26) with "Business_Income_available", suggesting a slight relationship between income levels and outstanding tax balances. "Business_Income_available" and "Tax Category" also display a moderate correlation (around 0.42), possibly reflecting structural differences between taxpayer types. However, the variable "New_category" exhibits negligible correlations with all other variables, indicating it may not play a significant role in predicting compliance. Most importantly, "is_compliant" has very weak correlations with all other variables, with the highest being only 0.13 with "Balance Due". This implies that no single feature strongly predicts compliance on its own. Such low individual correlations support the use of machine learning models to capture complex, non-linear interactions between multiple variables to better predict taxpayer compliance behavior.

Table 3 Model Performance Comparison

Model	Accuracy	Precision	Recall	F1 Score	AUC
Logistic Regression	0.84	0.82	0.86	0.84	0.89
Random Forest	0.87	0.85	0.89	0.87	0.91
XGBoost	0.88	0.86	0.90	0.88	0.92
Decision Tree	0.83	0.80	0.85	0.82	0.88

To evaluate the effectiveness of the classification models used in this study, several performance metrics were computed, including Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). These metrics provide a comprehensive view of each model's predictive capability, especially in the context of class imbalance and overall classification quality. A summary of the performance results is presented in Table

This study aimed to develop and evaluate machine learning models to predict taxpayer compliance behavior using historical tax data obtained from the Rwanda Revenue Authority (RRA). The research focused on identifying key factors influencing whether taxpayers file or pay taxes on time, and determining which predictive algorithms can best support early identification of non-compliant taxpayers. The dataset covered a five-year period from 2019 to 2023, including records from both Corporate Income Tax (CIT) and Personal Income Tax (PIT) categories. Variables such as Entity Type, Tax Type, Category, Business Income, Balance Due, and Late/Timely Flag were extracted, cleaned, and transformed for analysis. Data preprocessing included encoding categorical values, handling missing entries, and standardizing numerical features. Four machine learning models—Logistic Regression, Random Forest Classifier, XGBoost Classifier, and Decision Tree Classifier—were applied to the dataset. The models were evaluated using classification metrics including accuracy, precision, recall, F1-score, and confusion matrices. Among them, Logistic Regression and XGBoost achieved the highest testing accuracy (87.02The analysis revealed that Business Income, Balance Due, and Taxpayer Category were the most influential predictors of compliance behavior. The findings aligned with a priori expectations and supported economic theories such as Deterrence Theory, which posits that taxpayers are more likely to comply when potential penalties outweigh the benefits of evasion. The study contributes to the growing body of knowledge on data-driven tax administration by demonstrating how machine learning can enhance compliance monitoring in resource-constrained environments. The results offer practical insights for RRA, suggesting that predictive tools can help prioritize audits, improve risk assessment, and ultimately increase revenue collection without increasing the burden on compliant taxpayers.

V. SUMMARY

Machine learning models were developed and evaluated to predict taxpayer compliance using five years (2019-2023) of Rwanda Revenue Authority data covering Corporate and Personal Income Tax records. After preprocessing variables including Entity Type, Business Income, Balance Due, and compliance status, four algorithms were applied: Logistic Regression, Random Forest, XGBoost, and Decision Tree, with performance evaluated through accuracy, precision, recall, and F1-score metrics. XGBoost and Logistic Regression achieved the highest accuracy (87.02%), with Business Income, Balance Due, and Taxpayer Category identified as the most influential predictors of compliance behavior. The findings support Deterrence Theory and demonstrate machine learning's potential to enhance tax administration in resource-constrained environments by enabling the Rwanda Revenue Authority to prioritize audits, improve risk assessment, and increase revenue collection while reducing burden on compliant taxpayers.

VI. CONCLUSION

Machine learning models were developed to predict tax compliance behavior using five years (2019-2023) of Rwanda Revenue Authority data, addressing Rwanda's persistent compliance challenges caused by unpredictable taxpayer behavior and limited enforcement capacity. Analysis of corporate and personal income tax records identified key compliance predictors including entity type, business income, and balance due, with four algorithms tested for taxpayer classification based on compliance likelihood. XGBoost

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct097

emerged as the most accurate model, achieving 88% accuracy in a Rwanda-specific context where few comparative studies exist. The research integrates economic and behavioral theories to explain compliance motivations, demonstrating how data science combined with understanding of trust, fairness, and technology can enable tax authorities to make better decisions, target resource efficiently, and improve overall compliance. The findings provide practical insights for creating a more responsive and effective tax system in Rwanda, with future research opportunities in behavioral data integration and model expansion to other regions or tax types.

RECOMMENDATIONS

The Rwanda Revenue Authority should adopt predictive analytics for tax compliance monitoring, integrating high-accuracy models like Logistic Regression and XGBoost into risk management systems to shift from reactive to proactive enforcement approaches. Key recommendations include developing a real-time compliance dashboard for visual insights into filing patterns and risk profiles, prioritizing compliance support for taxpayers with high balances due and low business income, and expanding datasets with behavioral, sectoral, or geographic variables to improve model performance. Implementation requires investing in staff training for data science and machine learning capabilities to ensure system sustainability and scalability. Ethical guidelines must govern predictive analytics use, emphasizing transparency in data usage and protection against biased treatment. These measures would create a more efficient, targeted, and modern tax administration system in Rwanda through optimized audit efforts, resource allocation, and targeted interventions.

REFERENCES

[1]. Bird, R. M., & Zolt, E. M. (2019). Taxation and development: The weakest link? Essays in International Taxation. https://doi.org/10.2139/ssrn.3292406

[2]. Mascagni, G., & Mengistu, A. (2020). Can ICTs Increase Tax Compliance? Evidence on Citizen Engagement from a Randomised Controlled Trial in Ethiopia. ICTD Working Paper 109. https://www.ictd.ac/publication/ict-tax-compliance-ethiopia/

[3]. Mascagni, G., & Nell, C. (2022). Teach to comply? Evidence from a taxpayer education programme in Rwanda. International Centre for Tax and Development. https://www.ictd.ac/publication/teach-comply-evidence-taxpayer-education-programme-rwanda/

[4]. Hakizimana, N., & Santoro, F. (2023). Technology evolution and tax compliance: Evidence from Rwanda. International Centre for Tax and Development. https://www.ictd.ac/publication/technology-evolution-tax-compliance-evidence-rwanda/

[5]. Murorunkwere, B. F., Haughton, D., Nzabanita, J., Kipkogei, F., & Kabano, I. (2023). Predicting tax fraud using supervised machine learning approach. African Journal of Science, Technology, Innovation and Development. https://doi.org/10.1080/20421338.2023.2187930

[6]. Chen, Y., & Liu, W. (2023). The sentiment attitude of Weibo users towards annual individual income tax return: Based on natural language processing and machine learning methods. Proceedings of the 6th International Conference on Big Data and Artificial Intelligence, 67–72. https://doi.org/10.1109/BDAI59165.2023.10256913

[7]. Jayanti, D., Sulistyo, S., & Santosa, P. I. (2024). Application of machine learning in taxation: A systematic literature review. 2024 International Seminar on Intelligent Technology and Its Applications (ISITIA). https://doi.org/10.1109/ISITIA63062.2024.10668268

[8]. Joseph, A., Olabanji, O., & Wibowo, A. (2024). Transforming tax compliance with machine learning: Reducing fraud and enhancing revenue collection. Asian Journal of Economics, Business and Accounting, 24(11), 503–513.

[9]. Rwanda Revenue Authority. (2023). Official RRA Website. https://www.rra.gov.rw

[10]. Scikit-learn. (2024). Machine Learning in Python. https://scikit-learn.org/stable/

[11]. Battaglini, M., Guiso, L., Lacava, C., Miller, D. L., & Patacchini, E. (2024). Refining public policies with machine learning: The case of tax auditing. Journal of Econometrics. https://www.sciencedirect.com/science/article/pii/S0304407624001921

[12]. Zhang, J., & Wang, L. (2024). Using machine deep learning AI to improve forecasting of tax payments for corporations. Journal of Risk and Financial Management, 6(4), 48. https://www.mdpi.com/2571-9394/6/4/48

[13]. Liu, Y., Chen, X., & Zhang, K. (2023). A survey of tax risk detection using data mining techniques. Neurocomputing, 545, 126308. https://www.sciencedirect.com/science/article/pii/S2095809923003867

[14]. International Monetary Fund. (2024). Understanding artificial intelligence in tax and customs administration. Technical Notes and Manuals, 2024(006). https://www.elibrary.imf.org/view/journals/005/2024/006/article-A001-en.xml

[15]. Okoye, P. V., & Ezejiofor, R. A. (2024). The role of artificial intelligence in enhancing tax compliance and financial regulation. ResearchGate. https://www.researchgate.net/publication/378475910_THE_ROLE_OF_ARTIFICIAL_INTELLIGENCE_IN_ENHANCING_TAX_COMPLIANCE_AND_FINANCIAL_REGULATION

[16]. Zheng, S., Trott, A., Srinivasa, S., Parkes, D. C., & Socher, R. (2022). The AI economist: Taxation policy design via two-level deep multiagent reinforcement learning. Science Advances, 8(6). https://www.science.org/doi/10.1126/sciadv.abk2607

[17]. Antinyan, A., & Asatryan, Z. (2023). Nudging for tax compliance: A meta-analysis. The Economic Journal,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25oct097

135(668), 1033-1056. https://academic.oup.com/ej/article/135/668/1033/7810274
[18]. Chen, L., Liu, X., & Wang, S. (2024). Information nudges and tax compliance: Evidence from a field experiment in China. Journal of Economic Behavior & Organization, 227, 106543. https://www.sciencedirect.com/science/article/abs/pii/S0167268124003937
[19]. Martinez, A., Rodriguez, B., & Silva, C. (2023). The $100 million nudge: Increasing tax compliance of firms using a natural field experiment. Journal of Public Economics, 216, 104785. https://www.sciencedirect.com/science/article/abs/pii/S0047272722001815
[20]. World Bank Group. (2024). Nudging under pressure: Behavioral insights for tax compliance during a global pandemic. World Bank Blogs. https://blogs.worldbank.org/en/governance/nudging-under-pressure-behavioral-insights-tax-compliance-during-global-pandemic