# Machine Learning and Credit Rating in Financial Institution in Tanzania: A Literature Review Approach

**Neema Mfugale[1], Christopher Machibula[2]**

Assistant Lecturer
Economics
[1]Institute of Accountancy Arusha, [2]Mzumbe University
Tanzania.

**Abstract:**
**The aim of this study was to evaluate the effectiveness of Machine Learning models in credit rating within Tanzanian financial institutions, where data scarcity and informal financial systems limit the success of traditional models. Using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses methodology, a systematic review of 212 studies led to the inclusion of 36 high-quality paper in the qualitative synthesis, and 16 in the literature review for analysis. The findings reveal that ML models, especially Random Forest and Gradient Boosting, outperform traditional methods in predictive accuracy and adaptability, particularly in low-data environments. These models utilize alternative data such as mobile money transactions and utility payments, making them more inclusive for underserved populations. The study concludes that Machine Learning provides a viable solution to Tanzania's credit rating challenges and recommends adopting hybrid models and supportive regulatory frameworks to enhance credit access and financial inclusion**

**Keywords: Machine Learning, Credit Rating and Financial Institutions.**

## 1. INTRODUCTION

The advent of ML and AI technologies are revolutionising financial inclusion across the globe. To illustrate, ML-based solutions have been successfully applied in countries like China, India, and Kenya (Juma, 2022; Egorova, 2022). In China, ML powers digital lending platforms that offer microloans while fintech in Latin America leverage ML to detect fraud and extend credit to those without formal credit histories (Wu, 2021)**.** Such systems have been heralded as the solution to Tanzanian financial institutions, which struggle with a large informal financial sector and unbanked population (Barongo, 2024). The lack of structured data makes it difficult to accurately assess the creditworthiness of persons and small businesses. Currently, a large part of the Tanzania's financial system uses the traditional credit rating algorithms, which are based on formal credit histories and standardized scoring frameworks like FICO (Sathyanarayana, 2024). While the models are well supported by literature and have been used successfully for decades (Addy, 2024), they are essentially not well-suited to the Tanzanian context. Models such as FICO rely heavily on manual evaluation and basic statistical techniques and are, as such, insufficient in environments where financial information is scarce or outdated as is the case in parts of Tanzania (Addy, 2024).

The limitations of these traditional models have led to significant misalignments between financial institutions' offering and the client's demand. For instance, Sathyanarayana (2024) argues that there is a case of high rates of credit misclassification in markets where financial data is scarce. As a results, a huge number of potentially creditworthy individuals are denied loans while high-risk borrowers are approved. This shows that the lack of proper credit rating systems in Tanzania has contributed to financial exclusion especially of the marginalised populations. Such a situation was documents in Abdou et al. (2017) who argues that traditional credit rating systems have inherent limitations that lead to systemic financial exclusion (Abdou, 2017). Specifically, the systems use limited information and lack the capacity for real-time assessment and may be ineffective in a rapidly changing financial landscapes.

Machine Learning (ML) based systems have the potential overcome such limitations and the capacity to transform credit rating and credit risk assessment. In the context of credit risk assessment, ML models use a diverse range of customer data to develop dynamic assessment results. Using dynamic data enables ML based systems to outperform traditional statistical methods in predicting credit ratings and creditworthiness. ML based systems ability to synthesis complex, nonlinear data make the systems useful where credit history is limited or unavailable (Hlongwane, 2024). The models have inherent capability to address the many shortcomings of traditional models in use by most financial institutions in Tanzania. ML algorithms work by integrating and analysing a vast array of heterogeneous data sources to develop a more accurate, timely, and inclusive credit assessments (Magashi, 2024). Such data sources include mobile money usage, utility bill payments, and behavioural patterns among others.

Numerous studies have explored ML applications in credit scoring (Wu, 2021; Sun, 2024). However, most of these studies provide isolated case studies or qualitative insights. There is a lack of consolidated, quantitative evidence comparing the performance of ML models across different contexts (Hlongwane, 2024), data types, and algorithmic approaches (Addy, 2024). Furthermore, few studies focus specifically on low-data environments and developing countries with an equally developing finance market like Tanzania (Magashi, 2024). As such, there is a gap in understanding how well these models translate into practice in such settings. This study conducts a literature review on the application of Machine Learning (ML) model in credit rating, with a specific focus on financial institutions in Tanzania. The goal of the analysis is to provide a robust, evidence-based understanding of the effectiveness of machine learning in credit risk evaluation in data scarce environments in Tanzania. The analysis will offer practical insights for financial institutions looking to adopt more reliable, inclusive, and data-adaptive credit rating systems.

## 2. OVERVIEW OF EXTANT LITERATURE

### 2.1. Machine Learning for Credit Scoring Systems

Machine learning models are powerful analytical tools with the capability to perform smarter analytics and create more adaptable credit scoring systems. Unlike statistical methods, ML algorithms use inherent machine learning capabilities to develop more flexible and data-driven models, effectively making ML systems better at spotting complex patterns. Studies by Barongo (2024) and Hlongwane (2024) show that techniques like Random Forests and Gradient Boosting outperform standard logistic regression in predicting creditworthiness. What sets ML models apart is their ability to pick up on nonlinear trends unlike conventional methods, which do not inherently have such capabilities (Faluyi, 2024). The influence of deep learning under ML models is well documented in extant literature. For instance, Faheem (2024) explored the added accuracy that deep learning has on the precision of credit rating systems. The study shows that ML models are more accurate compared to convention systems, although the results vary depending on the metrics used (Faheem, 2024; Abdou, 2017). A study by Barongo (2024) found that localized ML models deliver reliable result even without full access to international regulatory data like LCR and NSFR.

### ML and Financial Inclusion in Developing Contexts

As shown above, ML models have the potential to augment credit scoring systems in developing economies where formal credit information is scarce. In such context, evidence shows that ML offers unique opportunities to advance financial inclusion. The strength of ML models in scoring lies on the ability to use alternative data on customer behaviour, demographics, and public information to assess creditworthiness (Magashi, 2024; Juma, 2022). A study by Kumar (2020), which evaluated the application of ML based models in rural credit systems, also shows that ML can be used alongside traditional models to enhance the traditional assessment methods (Kumar, 2021). The studies shows that ML models support inclusive financing by enabling credit access for individuals without prior borrowing history or collateral, whether used unilaterally or as hybrid systems.

However, despite the potential of ML in credit modelling, several barriers exist to their application especially where customer data is scarce. For instance, complex multi-class ML systems require coordinated oversight and regular updates to remain compliant with evolving financial standards (Pamuk, 2023). Further, model performance is context-sensitive, and success depends on selecting appropriate algorithms for specific financial environments (Faheem, 2024). There are also challenges posed by incomplete local data collection. According to Barongo (2024), incomplete data undermines comprehensive credit risk modelling (Barongo, 2024).

Further, implementation of ML based credit rating systems in developing economies poses a number of serious risks. Firstly, there is the concern of potential algorithmic bias (Asongu, 2018). Algorithmic bias arises when models are trained on skewed or non-representative data. Such conditions would reinforce existing inequalities rather than alleviate them (Binns, 2018). Such risks are more likely in jurisdictions such as Tanzania where financial and demographic data is not standardised. Bias leads to gaps in accountability, consumer protection, and model validation. In addition, infrastructural challenges such as limited internet connectivity, inadequate data storage systems, and low digital literacy hinder the deployment and proper use of ML applications (Barongo, 2024). These limitations not only affect model accuracy and fairness but has the potential to erode trust among end-users. Effectively, lack of trust would impede the success or slow the adoption of innovative financial solutions.

## 3. MATERIALS AND METHODS

This review synthesizes current knowledge on the application of ML in credit rating, with a specific interest in Tanzanian financial system. A structured review of extant literature was conducted based on PRISMA guidelines. the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is a standardized framework designed to improve the transparency, consistency, and completeness of reporting in literature reviews, systematic reviews and meta-analyses (Swartz, 2021). It includes a 27-item checklist and a flow diagram to guide researchers in documenting the review process from identification to inclusion of studies (Swartz, 2021; Tugwell, 2021). The review process entailed an initial search, article selection, critical appraisal, and data synthesis.

### 3.1. Search Strategy and Databases

The preliminary stage involved a comprehensive search of relevant academic literature performed in three key databases. The databases used were Scopus, ScienceDirect, and African Journals Online (AJOL). Scopus and ScienceDirect have a diverse range of peer-reviewed journals in computer science, finance, and machine learning, which are relevant for the current study. On the other hand, AJOL was included for regional relevance to capture Africa-specific research. It was to gather studies with contextualized information for Tanzanian and Sub-Saharan financial systems. Snowball sampling was employed to identify additional relevant studies by scanning reference lists of selected articles and reviewing key author publications. In addition to the regular search, snowballing was used to capture grey literature and seminal works that may not appear in initial database searches but are useful for understanding emerging themes.

The search focused on keywords including “machine learning,” “credit rating,” “credit scoring,” “financial institutions,” and “Tanzania.” To improve coverage, alternative terms and synonyms were used, such as “AI in finance,” “creditworthiness prediction,” “loan risk classification,” and “credit assessment using machine learning.” Only articles published in English from 2015 to 2024 were considered. This timeframe was selected to gather literature with most up to date information and reflect the rise and evolution of machine learning techniques in credit risk modelling. This is critical given that ML applications in finance gained momentum globally after over the last decade.

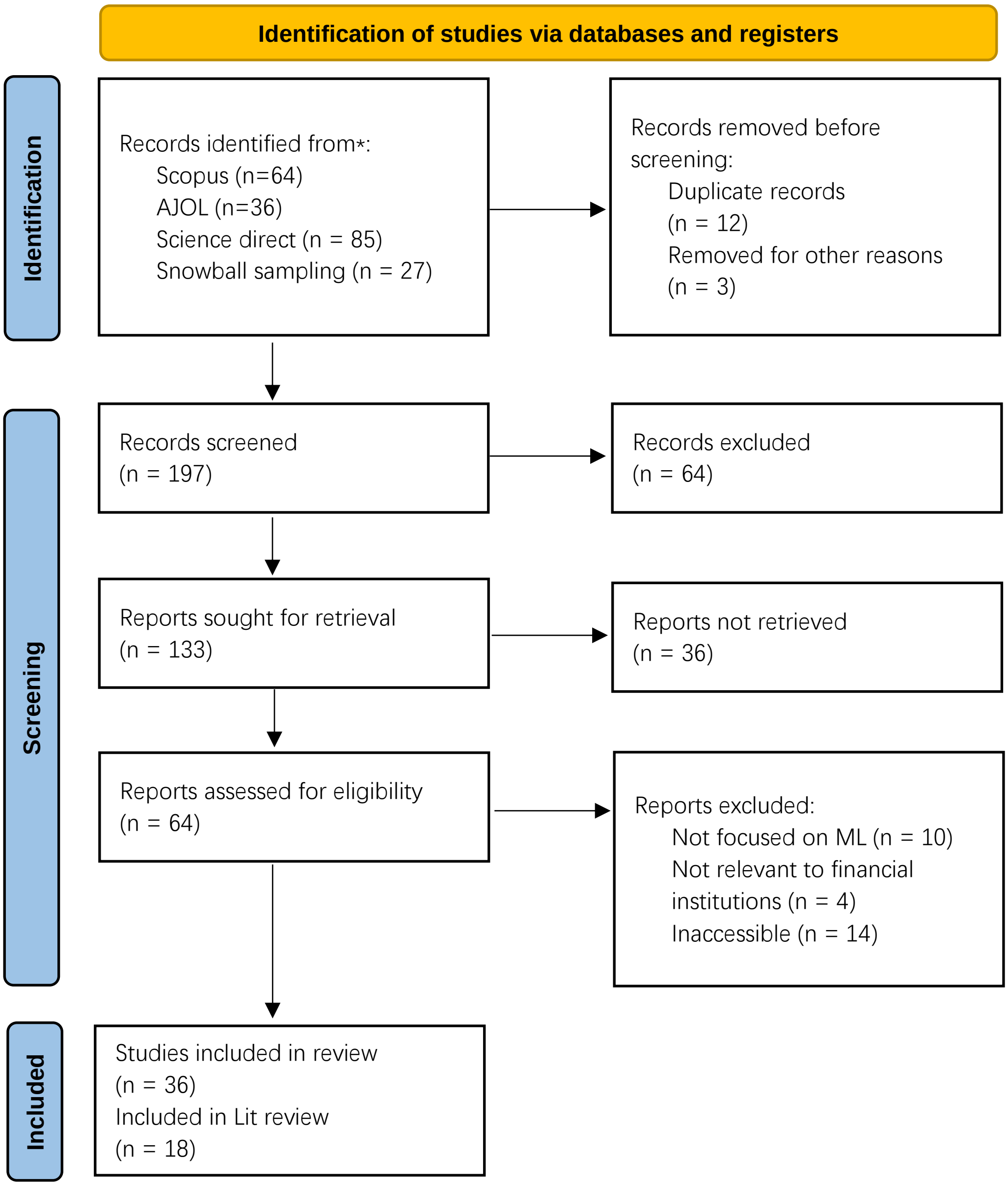
### 3.2. Eligibility Criteria

To be included in the review, a study had to be published in peer-reviewed journal or reputable conference and focus on the use of machine learning for credit rating or credit scoring in financial institutions. Additionally, articles were required to provide empirical or theoretical insights relevant to the Tanzanian context or to comparable developing economies. Only studies available in full text and published in English were considered for inclusion. Studies were excluded if they focused exclusively on traditional (non-machine learning) credit rating models or addressed machine learning in financial domains unrelated to credit assessment. Articles that were not accessible in full text or not published in English were also excluded.

### 3.3. Screening and Selection Process

All identified references were managed using Zotero reference management software, with duplicate entries systematically removed. The selection process followed a multi-stage approach. Initially, titles and abstracts were screened to assess their relevance. Articles that passed this stage underwent a full-text review to determine whether they met the predefined eligibility criteria. The final selection was based strictly on these criteria. A detailed PRISMA flow diagram outlining the selection process is presented in figure 1.

*Figure 1: PRISM Flow Diagram*




**Source: Author 2025**

A total of 212 records were identified. Of this, 185 articles were identified through database searches in Scopus, ScienceDirect, and AJOL). A further 27 articles were identified via snowball sampling. After removing duplicates, 197 records were screened. A hundred and thirty-three were excluded based on title and abstract. Twenty eight of the 64 full-text articles assessed for eligibility were excluded for not focusing on ML-based credit scoring, lacking relevance to Tanzania, or being inaccessible. Ultimately, 36 studies were included in the qualitative synthesis, and 16 in the literature review.

### 3.4. Data Extraction

A structured data extraction form was developed for this study. The form was created to guarantee consistency and thoroughness in data collection. The form included key details from each study, including the authors and year of publication, country and region of study, machine learning algorithms used, type and source of credit data, evaluation metrics including accuracy, precision, recall, and AUC, the type of financial institution involved, and the specific application context.

To assess the methodological quality of the studies included, a modified version of the 10-point Drummond checklist was employed. The Drummond checklist is a 10-point tool used to evaluate the methodological quality and validity of economic and social studies. The checklist is preferred because it ensures consistency in assessing aspects such as study design, data collection, and cost-effectiveness analysis. The modified tool comprised 10 questions evaluated using a 3-point rating system where:

- **0** where the criterion is not met at all,
- **0.5** indicates the criterion is partially met,
- **1** where the criterion is fully met.

The checklist is attached in Appendix 1. The Questions assessed relevance, methodological rigor, clarity of research objectives, transparency of data sources, and appropriateness of the evaluation metrics used. This scoring approach allows for a nuanced assessment of methodological quality across studies, ensuring partial fulfilment of criteria. The total possible score is 10, with higher scores indicating greater methodological rigor. After the rating, a further 20 studies were eliminated leaving 16 studies in the final literature review.

## 4. RESULTS

### Data Synthesis and Analysis

A qualitative synthesis of the included studies was conducted by systematically evaluating several thematic dimensions. These dimensions were selected to facilitate a comprehensive understanding of the application of machine learning (ML) in credit rating within financial institutions, particularly in Tanzania contexts. The first dimension evaluated was geographic emphasis. The dimension distinguished between studies conducted in developing nations like Tanzania and broader global contexts (Barongo, 2024). This classification enabled contextual comparisons by focusing on region-specific challenges such as data availability, infrastructure, and regulatory environments and their influence in the adoption and effectiveness of ML models in credit assessment (Juma, 2022; Magashi, 2024).

The second dimension identified a range of ML algorithms employed across studies. A summary of methodological application and suitability for different financial environments was noted. In addition, the type of credit product or focus area reported was noted and summarised. The key focus areas credit rating replication, financial strength prediction, liquidity risk classification, credit access for underserved groups (rural/micro-lending), and role of explainability (XAI) in model trust. Such focus enabled a clearer comparison of how ML performs across diverse financial functions and in solving context-specific challenges.

### 4.1. Summary of Results

#### 4.1.1. ML Model Used in Credit Scoring

The studies evaluated a variety of machine learning tools on credit risk assessment, albeit in different geographical locations. The studies evaluated various ML methods across different contexts. The following is a breakdown of the models evaluated in the various studies;

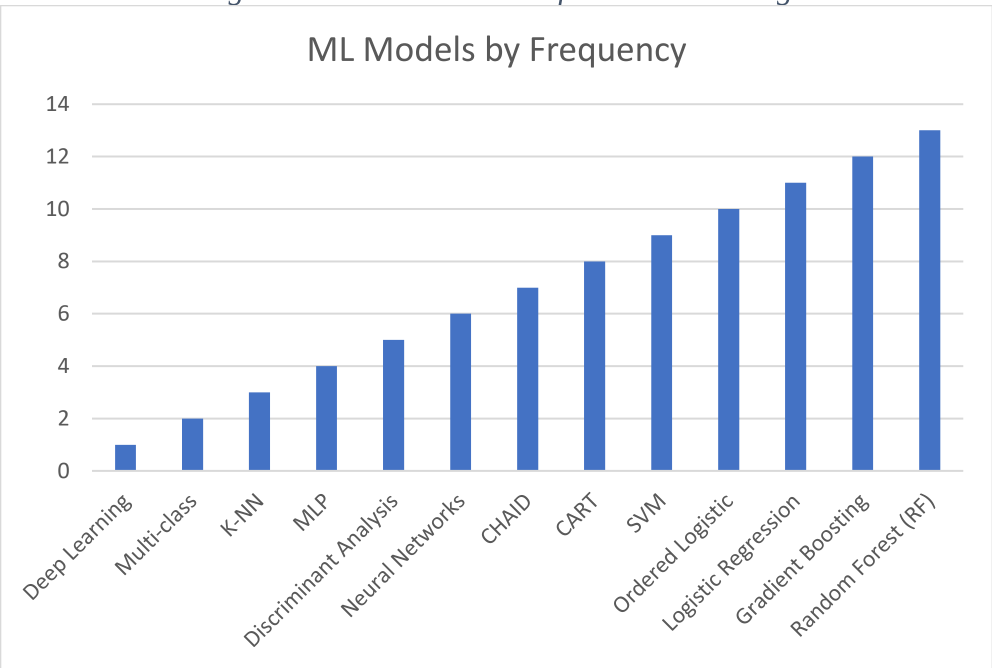
*Table 1: ML models Evaluated*

| ML Model | Count | Studies |
|---|---|---|
| Random Forest (RF) | 6 | Egorova (2023), Wu & Pan (2023), Amoupountlas (2023), Barongo et al., Mwende, Liu (2024) |
| Gradient Boosting (GB) | 3 | Egorova (2023), Mwende, Liu (2024) |
| Neural Networks | 2 | Hussein et al. (2017), Liu (2024) |
| Ordered Logistic Regression (OLR) | 2 | Egorova (2023), Hashimoto & Miura (2022) |

| | | |
|---|---|---|
| Support Vector Machine (SVM) | 1 | Wu & Pan (2023) |
| Logistic Regression | 2 | Wu & Pan (2023), Egorova (2023) |
| CART (Classification and Regression Trees) | 1 | Hussein et al. (2017) |
| CHAID | 1 | Hussein et al. (2017) |
| Discriminant Analysis (DA) | 1 | Hussein et al. (2017) |
| K-Nearest Neighbors (K-NN) | 1 | Mwende |
| Multi-class Classifiers | 1 | Amoupountlas (2023) |
| General ML (unspecified) | 3 | Liu (2024), Pamuck (2021), Kumar |

The results show that ML models outperformed traditional approaches in accuracy and adaptability. This is especially more profound where there is limited or alternative data (Barongo, 2024). Of the models evaluated, Random Forest and Gradient Boosting were the most frequently used and consistently showed strong results as shown above. A few studies also incorporated explainable AI to improve model transparency (Hashimoto, 2023; Egorova, 2022; Liu, 2024). In all cases, there is evidence showing that the ML models evaluated performed better than the conventional methods explored. The following is a summary of the ML models used and the frequency in the studies explored:

*Figure 2: ML Models Used for Credit Scoring*




**Source: Author**

The results show a preference for ensemble tree-based models. As shown above, Random Forest is the most popular among the models explored in all the studies included in the review. Its frequent application and consequent success are attributed to its robustness and effectiveness in diverse credit evaluation contexts (Ampountolas, 2021; Wu, 2021; Egorova, 2022). Gradient Boosting and Logistic Regression follow as common choices and reflect their balance between predictive power and interpretability (Egorova, 2022; Wu, 2021).

#### 4.1.2. ML Model Focus and Success Compared to Traditional Models

The analysis further shows that Random Forest (RF) models are predominantly applied in diverse credit risk and financial evaluation contexts. They are extensively used for credit rating replication (Egorova, 2022), personal and micro-lending credit scoring (Wu, 2021; Ampountolas, 2021), and bank liquidity risk classification (Barongo, 2024) The studies show that RF consistently demonstrates strong predictive accuracy and often outperforms other models. On the other hand, Gradient Boosting (GB) is commonly paired with RF

in corporate credit forecasting and comparative model analyses (Mwende, 2022; Egorova, 2022). GB models show competitive accuracy but generally trail RF models. Logistic regression remains a baseline and is favoured for interpretability and scorecard generation in personal credit scoring (Wu, 2021).
SVM, CART, CHAID, Neural Networks, and discriminant analysis address specialized tasks like financial strength ratings and rural credit scoring (Kumar, 2021). They reflect varied regional and data-driven requirements. Emerging techniques like deep learning and explainable AI are gaining traction but remain less widespread (Hashimoto, 2023; Liu, 2024).

Overall, ensemble tree-based models, particularly Random Forest and Gradient Boosting, consistently demonstrate better performance than traditional credit scoring methods such as logistic regression and discriminant analysis across various global contexts. Studies like Egorova (2023) and Hashimoto & Miura (2022) show that ML methods improve predictive accuracy and offer enhanced interpretability when combined with explainable AI tools (Hashimoto, 2023). While traditional models remain useful in some settings, ML approaches adapt better to complex, nonlinear financial data, providing more reliable credit risk assessments. The following table summarises the comparison between ML and classical models.

| Study | ML Methods Used | Traditional Methods Used | Key Finding (Comparison) |
|---|---|---|---|
| **Egorova (2023)** | Random Forest (RF), Gradient Boosting (GB) | Ordered Logistic Regression | RF and GB outperformed logistic regression in replicating Moody's ratings |
| **Hashimoto & Miura (2022)** | ML + SHAP, PDP | Ordered Logistic Regression (OLR) | ML models showed better accuracy; traditional methods enhanced with XAI for interpretation |
| **Hussein Et Al. (2017)** | CART, CHAID, Neural Networks | Discriminant Analysis (DA) | DA performed best, but ML models were competitive |
| **Wu & Pan (2023)** | Random Forest, SVM | Logistic Regression | Logistic regression preferred due to interpretability and performance with feature reduction |
| **Liu (2024)** | ML, Deep Learning | Statistical models | Reviews trade-offs; traditional models still favoured by some stakeholders for transparency |
| **Mwende (2024)** | RF, Gradient Boosting, K-NN | Traditional credit scoring (unspecified) | ML models outperformed traditional scoring by leveraging diverse borrower data |

### 4.1.3. Effects of ML in Data Scarce Environments

In the Tanzanian and other developing country contexts, ML models have shown especially strong results. Barongo et al. reported high accuracy (90–96%) with a Random Forest–MLP hybrid for bank liquidity risk classification (Barongo, 2024), while Mwende found Random Forest significantly outperformed other ML algorithms like K-Nearest Neighbours and Gradient Boosting in credit scoring by incorporating diverse borrower attributes (Mwende, 2022). Additionally, ML's ability to tap into alternative data sources—especially where formal credit histories are missing—can significantly improve credit access for underserved communities (Ampountolas, 2021; Mhlanga, 2021). This highlights how machine learning models can offer a meaningful upgrade to credit scoring systems, particularly in countries like Tanzania where traditional models struggle due to limited data.

## 5. CONCLUSION AND RECOMMENDATIONS

From the review conducted, Machine learning (ML) has the potential to improve credit rating systems in Tanzania. Traditional credit scoring methods predominantly used in Tanzania today are inadequate largely due to Tanzania's financial landscape, which is characterised by informal activity, minimal credit history information, and mottled data. ML models offer better prediction accuracy, are more dynamic and flexible and create dynamic systems that support inclusion (Asongu, 2018; Munkhdalai, 2019). ML systems ability to draw insights from alternative data like mobile money usage and behavioural patterns is especially useful in reaching underbanked populations ( (Barongo, 2024; Binns, 2018).

However, Tanzanian policymakers and financial institutions should adopt a multifaceted approach to ML implementation. The following are a few practical changes to enable the Tanzania financial system to take advantage of the potential that ML portend:

**Adopt Hybrid Models**

One of the ways Tanzania can adopt is to combine traditional tools like logistic regression or expert scorecards with modern ML techniques like Random Forest or Support Vector Machines. A hybrid approach will help strike a balance between the transparency and familiarity of old methods and the accuracy and adaptability of ML based systems. This hybrid setup offers a practical way forward in a country where regulators are cautious and data is often fragmented.

**Enhance regulatory support**

Enhancing regulatory support is also key to the successful and responsible integration of AI and machine learning in Tanzania's financial sector. As shown in the review, there are a number of ethical and information security concern associated with the adoption of ML models. As such, policymakers should develop comprehensive regulatory frameworks that clearly define acceptable practices for the use of these technologies. These guidelines must address key issues such as ethical use, data privacy, and model explainability. This would ensure that ML algorithms do not perpetuate bias or discrimination, safeguarding customer information in compliance with national and international standards and that providers are transparent about ai supported decisions making. Such frameworks will foster innovation while protecting consumers and maintaining trust in the financial system.

## REFERENCES:


1. Abdou, H. M. (2017). Prediction of financial strength ratings using machine learning and conventional techniques. Investment Management and Financial Innovations, 14(4), 194–211. https://doi.org/10.21511.
2. Addy, e. a. (2024). AI in credit scoring: A comprehensive review of models and predictive analytics. Global Journal of Engineering and Technology Advances, 18(02), 118-129.
3. Ampountolas, A. N. (2021). A Machine Learning Approach for Micro-Credit Scoring. Risks, 9(3), 50. https://doi.org/10.3390/risks9030050.
4. Asongu, S. A. (2018). A literature survey on the use of machine learning algorithms in credit scoring for developing countries. . Research in International Business and Finance, 45, 243–258.
5. Barocas, S. H. (2019). Fairness and Machine Learning. http://fairmlbook.org.
6. Barongo, R. I. (2024). Using machine learning for detecting liquidity risk in banks. Machine Learning with Applications, 15, 100511. https://doi.org/10.1016/j.mlwa.2023.100511.
7. Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. Proceedings of the 2018 Conference on Fairness. Accountability and Transparency (FAT), 149–159.
8. Brown, I. &. (2012). An experimental comparison of classification algorithms for imbalanced.
9. Egorova. (2022). Comparative Analysis of the Predictive Power of Machine Learning Models for Forecasting the Credit Ratings of Machine-Building Companies. Journal of Corporate Finance Research / Корпоративные Финансы | ISSN: 2073-0438,.
10. Faheem, M. A. (2024). Enhancing financial forecasting accuracy through AI-driven predictive analytics models.
11. Faluyi, S. G. (2024). Comparative analysis of machine learning models for predicting online loan defaults in Nigeria. International Journal of Artificial Intelligence and Soft Computing, 8(3), , 238-256.
12. Hashimoto, R. M. (2023). Application of Machine Learning to a Credit Rating Classification Model: Techniques for Improving the Explainability of Machine Learning.

13. Hlongwane, R. R. (2024). Enhancing credit scoring accuracy with a comprehensive evaluation of alternative data. Plos one, 19(5), e0303566.
14. Jemai, A. A. (2020). Machine learning in credit scoring: A survey. . Journal of Decision Systems, 29(1), 17–33.
15. Juma, S. O. (2022). Relevance of alternative data and machine learning in predicting default in a non-deposit taking SACCO in Kenya.
16. Kumar, A. S. (2021). Machine Learning (ML) Technologies for Digital Credit Scoring in Rural Finance: A Literature Review. Risks, 9(11), 192. https://doi.org/10.3390/risks9110192.
17. Lessmann, S. B. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. . European Journal of Operational Research, 247(1), 124–136.
18. Liu, S. (2024). Application Analysis of Machine Learning Models in Credit Risk. Highlights in Science, Engineering and Technology, 107, 76–81. https://doi.org/10.54097/6qm0dk77.
19. Magashi, C. A. (2024). Prediction of SACCOS failure in Tanzania using machine learning models. . Engineering, Technology & Applied Science Research, 14(1), , 12887-12891.
20. Makabe, Y. M. (2023). Estimating Pipeline Pressures in New Keynesian Phillips Curves: A Bayesian VAR-GMM Approach.
21. Mhlanga, D. (2021). Financial Inclusion in Emerging Economies: The Application of Machine Learning and Artificial Intelligence in Credit Risk Assessment. International Journal of Financial Studies, 9(3), 39. https://doi.org/10.3390/ijfs9030039.
22. Munkhdalai, L. M.-E. (2019). An Empirical Comparison of Machine-Learning Methods on Bank Client Credit Assessments. Sustainability, 11(3), 699. https://doi.org/10.3390/su11030699.
23. Mwende. (2022). Effect of Machine Learning in Early Prediction of Personal Loan Defaulting on Commercial Bank in Tanzania (2009-2020). AFRICAN CENTRE OF EXCELLENCE IN DATA SCIENCE.
24. Pamuk, M. &. (2023). Opening a New Era with Machine Learning in Financial Services? Forecasting Corporate Credit Ratings Based on Annual Financial Statements. . International Journal of Financial Studies, 11(3), 96. https://doi.org/10.3390/ijfs.
25. Sathyanarayana. (2024). The Unified Lending Interface-Innovations in Lending: Conceptual Insights into the Future of Financial Services.
26. Sun, J. (2024). Research on risk detection and management based on machine learning. . Applied and Computational Engineering, 92(1), 142–146. https://doi.org/10.54254/2755-2721/92/20241678.
27. Swartz, M. K. (2021). PRISMA 2020: an update. Journal of Pediatric Health Care, 35(4)., 351.
28. Tugwell, P. &. (2021). PRISMA 2020. Journal of Clinical Epidemiology, 134, A5-A6.
29. Wu, Y. &Pan. (2021). Application Analysis of Credit Scoring of Financial Institutions Based on Machine Learning Model. Complexity, 2021(1), 9222617. https://doi.org/10.1155/2021/9222617.
30. Yuichiro, I. Y. (2023). Corporate Pension Funds’ Investment Strategies and Financial Stability: Lessons from the Turmoil in the UK Gilt Market.
31. Zhang, D. L. (2020). Credit risk evaluation using ensemble learning in peer-to-peer lending. . Journal of the Association for Information Science and Technology, 71(4), , 413–426.

**Appendix 1: Drummond 10-point Check list**

| No. | Criterion | Description | Score (0–1) |
|---|---|---|---|
| 1 | **Clearly stated objective** | The study clearly outlines its aim, focusing on ML application in credit risk assessment. | |
| 2 | **Appropriate comparators used** | Relevant models or benchmarks (e.g., traditional scoring methods) are used for comparison. | |
| 3 | **Well-described data sources** | Data sources, sample size, and selection criteria are clearly described and appropriate. | |
| 4 | **Justified and reproducible ML methods** | ML methods are explained, justified, and presented in a way that allows reproducibility. | |
| 5 | **Robust model evaluation metrics** | Performance metrics (e.g., accuracy, AUC, F1-score) are appropriate and clearly reported. | |
| 6 | **Handling of class imbalance or bias** | The study addresses common data issues like class imbalance or sampling bias. | |
| 7 | **Interpretability of results** | The results are presented in a way that facilitates understanding and decision-making. | |
| 8 | **Discussion of limitations** | Limitations of the study and model constraints are clearly acknowledged. | |
| 9 | **Policy/practical implications discussed** | The study offers practical recommendations for financial institutions or regulatory agencies. | |
| 10 | **Ethical or data privacy concerns noted** | Ethical, privacy, or fairness concerns related to ML implementation are acknowledged. | |
| | **Total Score** | | **/10** |

**Appendix 2: Literature Review Summary**

| Study | Focus Area | ML Methods Used | Region/Context | Key Findings | Gaps/Limitations |
|---|---|---|---|---|---|
| **Egorova (2023)** | Moody's credit rating replication | Ordered Logistic, RF, GB | Global (18 countries) | RF and GB ~50% accuracy, outperform logistic regression; macro vars not helpful | Moderate sample size; sector-specific |
| **Hashimoto & Miura (2022)** | Credit rating with XAI | ML + SHAP, PDP, OLR | Japan/global | ML better accuracy; ICR < 2 critical; XAI enhances interpretability | XAI requires domain expertise |
| **Pamuck (2021)** | Corporate credit forecasting | Various ML + sampling strategies | Europe | Sampling improves ML performance; regulatory compliance discussed | Model update standardization needed |
| **Hussein et al. (2017)** | Financial Strength Ratings | CART, CHAID, Neural Networks, DA | Middle East | DA performed best, but ML methods competitive | Mixed results across methods |
| **Barongo et al.** | Bank liquidity risk classification | RF–MLP hybrid | Tanzania | High accuracy (90–96%) with extended liquidity metrics | Limited liquidity ratio data; operational tests needed |

| | | | | | |
|---|---|---|---|---|---|
| **Mhlanga (2022)** | AI & financial inclusion | ML with alternative data | Emerging Markets | Alternative data improves credit access for underserved populations | Data ethics and privacy concerns |
| **Kumar** | Rural credit scoring | ML algorithm survey | Global (rural focus) | ML enables rural financial inclusion but faces technical and ethical challenges | Legacy system integration challenges |
| **Liu (2024)** | Overview of credit risk ML models | Statistical, ML, Deep Learning | Global | Deep learning promising; different stakeholders prefer different models; diverse evaluation metrics | Data quality and interpretability challenges |
| **Wu & Pan (2023)** | Personal credit scoring big data | Random Forest, SVM, Logistic Regressio n | Global (Lending Club dataset) | Feature reduction crucial; logistic regression preferred; final scorecard output | Dataset-specific findings; model generalizability unknown |
| **Amoupountlas (2023)** | Micro-lending credit evaluation | Random Forest, multi-class classifiers | Developing countries | RF effective with limited borrower data; enables fair credit evaluation without credit history | Small dataset; limited feature variety |
| **Mwende** | **K-Nearest Neighbors (K-NN), Gradient Boosting, and Random Forest (RF)—were trained and tested** | | Developing countries/TZ | ML models outperform traditional credit scoring methods by incorporating diverse borrower attributes. | **Random Forest significantly outperformed K-NN and Gradient Boosting in all evaluation metrics.** |