Prediction of SACCOS Failure in Tanzania using Machine Learning Models

Cosmas Magashi
Department of Information Technology System Development and Management (ITSDM), Nelson Mandela African Institution of Science and Technology (NM-AIST), Tanzania
magashi80@yahoo.com (corresponging author)

Johnson Agbinya
School of Information, Technology, and Engineering, Melbourne Institute of Technology (MIT), Australia
jagbinya@mit.edu.au

Anael Sam
Department of Information Technology System Development and Management (ITSDM), Nelson Mandela African Institution of Science and Technology (NM-AIST), Tanzania
anael.sam@nm-aist.ac.tz

Jimmy Mbelwa
Department of Computer Science and Engineering, University of Dar es Salaam, Tanzania
jimmymbelwa@gmail.com

Received: 15 June 2023 | Revised: 1 July 2023 and 2 December 2023 | Accepted: 13 December 2023

 DOI: https://doi.org/10.48084/etasr.6696

ABSTRACT

Savings and Credit Co-Operative Societies (SACCOS) are seen as viable opportunities to promote financial inclusion and overall socioeconomic development. Despite the positive outlook for socioeconomic progress, recent observations have highlighted instances of SACCOS failures. For example, the number of SACCOS decreased from 4,177 in 2018 to 3,714 in 2019, and the value of shares held by SACCOS members in Tanzania dropped from Tshs 57.06 billion to 53.63 billion in 2018. In particular, there is limited focus on predicting SACCOS failures in Tanzania using predictive models. In this study, data were collected using a questionnaire from 880 members of SACCOS, using a stratified random sampling technique. The collected data was analyzed using machine learning models, including Random Forest (RF), Logistic Regression (LR), K Nearest Neighbors (KNN), and Support Vector Machine (SVM). The results showed that RF was the most effective model to classify and predict failures, followed by LR and KNN, while the results of SVM were not satisfactory. The findings show that RF is the most suitable model to predict SACCOS failures in Tanzania, challenging the common use of regression models in microfinance institutions. Consequently, the RF model could be considered when formulating policies related to SACCOS performance evaluation.

Keywords-microfinance; SACCOS; machine learning; prediction; classification

I. INTRODUCTION

Savings and Credit Co-Operative Societies (SACCOS) have been providing great solution to people with economic challenges [1]. Several researchers have appreciated the importance of SACCOS on socioeconomic development at both individual and community levels around the world [2]. SACCOS serve as engines for enhancing financial and human resources which contribute much to the national development [3]. The ability of SACCOS to facilitate capital accumulation and provide loans for various social needs, such as housing, education, and community development, underscores their importance in enhancing the well-being of their members [4-8].

Despite the several advantages of SACCOS to the economic development, there is still however a growing concern about the high decrease in the number of SACCOS and the decrease in the value of shares among members [4]. This study addresses the issue of SACCOS failures which have been caused by challenges such as improper utilization of loans acquired, lack of financial management knowledge, poor loan repayment, weak leadership, financial constraints, and unreliable interest rates [5]. Many researches have shown financial and operational sustainability issues within SACCOS which indicates the need for comprehensive investigation and predictive models [6]. Although previous studies have explored failures in microfinance institutions, limited attention has been paid to SACCOS, and existing research focuses primarily on the current state rather than predicting failures. This study seeks to bridge this gap by employing a prediction model to anticipate SACCOS failures in Tanzania to contribute valuable information that can inform policies and strategies to prevent and address challenges within SACCOS, ensuring their continued positive impact on economic development and financial inclusion.

II. RELATED WORKS

SACCOS have been regarded as an eye towards financial inclusion in an economy where many people are facing financial challenges [7-8]. SACCOS have become triggers for mobilizing financial resources, human resources, and capital for national development [9]. SACCOS help their members accumulate capital and loans for social needs, such as building houses, buying clothes, paying school fees, hosting weddings, and other social activities [10]. They enable easy access to financial services, encourage savings, create employment opportunities, and directly support community development efforts, such as enabling community access to social services for stimulated growth [11-12]. SACCOS make finance more accessible to underprivileged members, limit financial exclusion, stimulate a thrift culture, help bring awareness to ordinary people of how to create assets, and generally improve the economic performance of developing countries [12].

SACCOS in Tanzania have a history dating back to 1930. The total of loans issued by SACCOS in Tanzania increased to 927.66 billion in 2019 from 833.89 billion, indicating increased credit to finance various economic activities. Additionally, total loans issued increased by 15.21 billion in the last quarter of 2018. The total savings of SACCOS members increased to 200.21 billion in 2019 compared to 176.48 billion in 2018. In this period, SACOOS members total deposits increased to 51.67 billion compared to 39.53 billion in December 2018. The total number of members, share values, value of deposits, savings, loans issued, and outstanding loans were found to have increased in 2019. Regardless of such socioeconomic development and contributions, some unexpected issues were observed. For example, the number of SACCOS decreased from 4,177 at the end of 2018 to 3,714 at the end of 2019. The value of shares of SACCOS members in the country decreased to 53.63 billion in 2019 compared to 57.06 billion in 2018. The above-given statistics signal SACCOS failures. In [12], SACCOS failures were attributed to various challenges such as improper utilization of loans taken, lack of knowledge in financial management, and poor repayment of loans taken by members. In Tanzania, SACCOS fail to deliver adequate socioeconomic benefits to members due to both internal and external factors, which may generally include financial constraints, weak leadership, unreliable interest rates of loans, and poor record-keeping systems. Out of 103 SACCOS in Tanzania, 61% of them fail to be sustainable operationally and 51% of them fail financially.

SACCOS failure has previously been studied in [9-10, 12, 13]. Most studies have investigated the existing situation of SACCOS and not the prediction of the failure of given SACCOS. Other studies have used prediction models but not in SACCOS. This study sheds light on the pressing issue of SACCOS failures, emphasizing the need for proactive measures to ensure their sustained socioeconomic impact. This study used machine learning models to provide a valuable tool for predicting potential failures, offering a strategic approach to address the challenges faced by SACCOS. The experimental results demonstrate the effectiveness in classifying datasets, further underscoring its potential practical applications. As SACCOS continue to be crucial to financial inclusion and community development, the insights derived contribute to a broader discourse on mitigating operational and financial risks within cooperative societies. Moving forward, policymakers, regulatory bodies, and stakeholders can take advantage of these findings to implement informed strategies to strengthen the resilience of SACCOS and sustain their positive impact on economic growth and social well-being.

III. THE PROPOSED ALGORITHM

This study used a quantitative approach aligned with its objective of focusing on causal-effect relationships utilizing statistical data to optimize efficiency and resource utilization. Employing an explanatory cross-sectional survey research design, the study analyzes each targeted SACCOS as a unit of analysis across diverse regions in Tanzania. This design allows for a comprehensive examination of the failure information among the SACCOS surveyed. The research spans five key regions: Dar es Salaam, Arusha, Mbeya, Mwanza, and Dodoma, which not only host a significant concentration of SACCOS but also represent major zones in Tanzania. In particular, the high-risk nature of SACCOS in these regions is acknowledged by the Tanzania Cooperative Development Commission (TCDC). The study strategically involved SACCOS members as primary informants due to their deep knowledge of key figures, such as staff, management, and TCDC representatives, to retrieve comprehensive information on SACCOS performance. Units of analysis predominantly focus on SACCOS, given the relatively limited exploration of their performance compared to other Microfinance Institutions (MFIs). The sample size, determined using (1) [14-15], involved 880 SACCOS members in the specified regions, allowing for a nuanced examination of factors that contribute to their success or failure. This research design encompasses geographic diversity and key informant perspectives and enables the study to offer valuable information on the dynamics of SACCOS failures in Tanzania.

$$n = \frac{Z^2 pqN}{e^2(N-1) + Z^2 pq} \tag{1}$$

In (1), n is the sample size for a finite population, N is the population size, p is the population reliability or the estimated frequency for a sample of size n, which was 0.5 as used for the population of all developing countries, $p + q = 1$, e is the margin of error considered 3%, and Z is the normal reduced variable, which at the 0.05 level of significance is 1.96 giving a sample size of 880.

Concerning the predictive model for SACCOS failure in Tanzania, the focus encompasses two primary aspects: the formulation of a suitable model to predict SACCOS failure and the subsequent validation of its accuracy and efficacy. These tasks were accomplished using machine-learning models [16-19]. Throughout this process, the dependent variable was identified as SACCOS Failure (*SP*), while the independent variables were Return on Assets (*RA*), SACCOS Size (*SS*), SACCOS Members (*SM*), SACCOS Assets (*SA*), Deposits Mobilization (*DM*), Interest Rate (*IR*), and Age of SACCOS (*AS*). The resulting model is expressed by:

$$SP = \beta_1 RA + \beta_2 SS + \beta_3 SM + \beta_4 SA + \beta_5 DM + \beta_6 IR + \beta_7 AS \tag{2}$$

where β_1, β_2, β_3, β_4, β_5, β_6, and β_7 are estimated regression coefficients that represent the change in *SP* relative to one unit in change in the respective independent variable. The values of the regression coefficients were estimated from the dataset used in the study. The resulting values of each regression coefficient were: β_1= -153.8069, β_2 = 153.9002, β_3= 154.6836, β_4= -153.8151, β_5 = 0.00001215, β_6 = 0.0009, and β_7 = 0.0495. The dependent variable was computed using:

$$SF = \frac{SACCOS\ Operating\ Income\ (SOI)}{SACCOS\ OPERATING\ Expenses\ (SOE)} \tag{3}$$

The descriptions of the independent variables were made as follows: RA is the ratio of the total SACCOS income to the average total SACCOS assets (4), where the average total SACCOS assets is the ratio of the total assets to the total members of *SS* expressed by (5):

$$RA = \frac{Income}{Average\ Total\ Asset} \tag{4}$$

$$SS = \ln\left(Total\ SACCOSS\ Income\right) \tag{5}$$

$$DM = \frac{Deposits}{Loans} \tag{6}$$

IV. EXPERIMENTAL RESULTS

The sample size of this study comprised 880 individuals drawn from SACCOS members across diverse regions of Tanzania. To develop and validate the prediction model for SACCOS failure, the dataset was partitioned into training data (80%) and testing data (20%). The training data, serve as the foundation for instructing machine learning models, including KNN, SVM, LR, and RF. KNN operates by classifying data points based on the majority class within their proximity. SVM constructs a hyperplane to segregate classes, aiming to maximize the margin between them. LR models the probability of an event occurring, and RF employs an ensemble of decision trees for classification. Each method presents distinct advantages and disadvantages. KNN is simple and effective but can be computationally expensive. SVM is robust in high-dimensional spaces but may struggle with larger datasets. LR is interpretable and efficient but assumes linear relationships. RF excels in handling complex datasets but may overfit with noisy data. The performance of these methods was evaluated through the training and testing phases, considering their respective strengths and limitations in predicting SACCOS failure. Table I shows the performance of each algorithm based on the sample size.

TABLE I. MACHINE LEARNING MODELS FOR SACCOS FAILURE PREDICTION

Index	SVM	LR	KNN	RF
Accuracy	0.6498	0.9663	0.8232	1.0000
Recall	0.0000	0.9999	0.8043	1.0000
AUC	0.7722	0.9999	0.9012	1.0000
Precision	0.0000	0.9159	0.7208	1.0000

Table I presents an evaluation of different machine learning models for predicting SACCOS failure based on training and test datasets. The SVM model showed suboptimal performance with recall and precision scores of 0%, an Area Under the Curve (AUC) score of approximately 77%, and an accuracy score of approximately 65%. These results indicate that SVM is not suitable for predicting SACCOS failure in Tanzania. In contrast, the LR model exhibited superior performance with approximately 99% recall, 91.59% precision, 100% AUC, and 99.63% accuracy. Similarly, the KNN model presented significant index scores, including approximately 80% recall, 72% precision, 90% AUC, and 82% accuracy. Finally, the RF model demonstrated exceptional performance with perfect scores in all indices, indicating that it is the most effective model for classifying SACCOS failure compared to the other models tested in this study. These findings reveal distinctive results for each algorithm, as reflected in their respective confusion matrices. For SVM, the recall and precision scores both registered 0%, underscoring its limitations in correctly identifying true positives and minimizing false positives. LR, on the other hand, exhibited remarkable performance, with high recall, precision, AUC, and accuracy scores, indicating its efficacy in capturing true positive instances while maintaining precision. KNN demonstrated significant recall and precision scores, emphasizing its ability to correctly classify positive instances with a balanced precision rate. RF presented perfect scores on all indices of the confusion matrix, showcasing its unparalleled ability to correctly identify true positives, minimize false positives and negatives, and achieve overall accuracy. The confusion matrices shown in Figures 1-4 offer insight into the performance of the models.

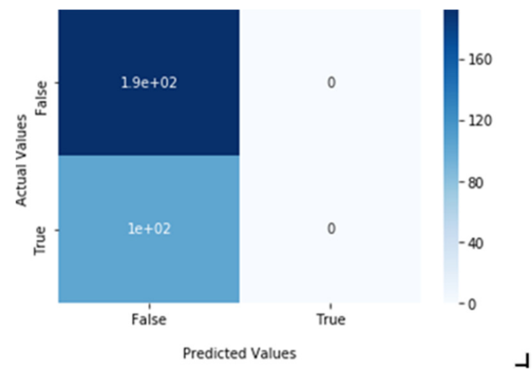

Fig. 1. Confusion matrix for SACCOS failure prediction using SVM.

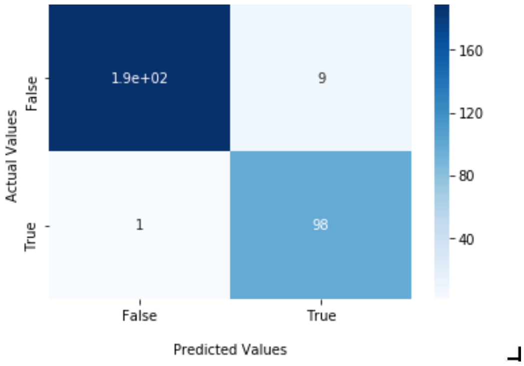

Fig. 2. Confusion matrix for SACCOS failure prediction using KNN.

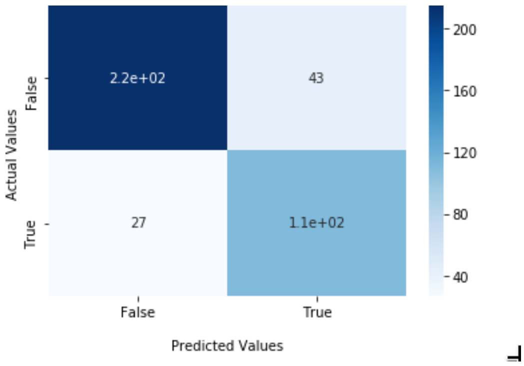

Fig. 3. Confusion matrix for SACCOS failure prediction using LR.

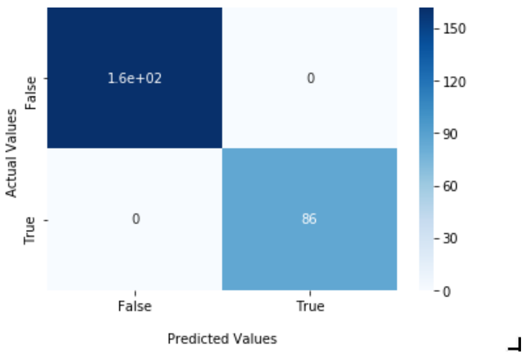

Fig. 4. Confusion matrix for SACCOS failure prediction using RF.

A comparison was carried out using ROC curves to evaluate the four considered machine-learning models. Figure 5 illustrates that the RF model's curve ranks highest with an AUC of 1.00, followed by the LR model with an AUC of 1.00, the KNN model with an AUC of 0.90, and the SVM model at the bottom with an AUC of 0.77.

This comparative analysis indicates that RF excels as the premier model for classifying the dataset, followed by LR and KNN, while SVM exhibits comparatively inferior performance. These findings align consistently with the individual model evaluations discussed earlier. In summary, although KNN and LR demonstrate commendable performance in classifying the dataset, RF emerges as the optimal model in this study for predicting SACCOS failure in Tanzania. Furthermore, with a 100% accuracy rate, RF proves to be an exceptionally effective model in anticipating SACCOS failure in the Tanzanian context, as shown through the Receiver Operating Characteristic (ROC) curve in Figure 5. Recommendations include integrating predictive analytics into SACCOS management, fostering financial literacy programs, and implementing stringent loan utilization monitoring. Such measures can fortify SACCOS resilience, ensuring sustained socio-economic contributions and mitigating operational risks.

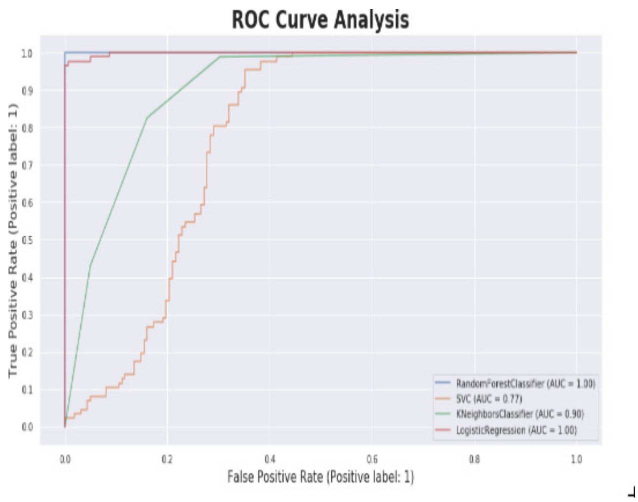

Fig. 5. The ROC curve for comparison of the machine learning models.

The ROC curves for the four machine-learning algorithms indicate distinct performances. The RF model demonstrated the highest performance, with an AUC of 1.00, followed by LR with an AUC of 1.00, KNN with an AUC of 0.90, and SVM with the lowest AUC of 0.77. This comparison underscores RF as the superior model in classifying the dataset, with LR and KNN following closely. The analysis of the ROC curve reaffirms the findings discussed above, highlighting the predictive efficacy of RF in anticipating SACCOS failure.

V. CONCLUSION

This study investigated the critical issue of SACCOS failure, employing a quantitative approach and an explanatory cross-sectional survey research design. The study collected and analyzed data from 880 SACCOS members in five key regions in Tanzania. Examination of four machine-learning models, SVM, LR, KNN, and RF, revealed significant disparities in predictive efficacy. The RF model emerged as the optimal choice, boasting a 100% accuracy rate, making it the most effective for anticipating SACCOS failure. Recommendations include integrating predictive analytics into SACCOS management practices, improving financial literacy initiatives, and implementing robust monitoring mechanisms for loan utilization. These measures aim to strengthen the resilience of SACCOS, ensuring sustained socioeconomic contributions and mitigating operational risks in the microfinance sector.

REFERENCES

[1] M. J. Zikalala, "The role of savings and credit cooperatives in promoting access to credit in Swaziland," MSc Thesis, University of Pretoria, South Africa, 2016.

[2] "The role of SACCOS' microcredits in the empowerment of female-headed households in the Njombe Region, Tanzania," *African Journal of Applied Research*, vol. 8, no. 1, pp. 200-212, Aug. 2022, https://doi.org/10.26437/ajar.03.2022.14.

[3] R. A. Effiom, "Impact of cooperative societies in national development and the Nigerian economy," *Global Journal of Social Sciences*, vol. 13, no. 1, pp. 19–29, 2014, https://doi.org/10.4314/gjss.v13i1.2.

[4] M. D. Kwai and J. K. Urassa, "The contribution of savings and credit cooperative societies to income poverty reduction: A case study of Mbozi District, Tanzania," *Journal of African Studies and Development*, vol. 7, no. 4, pp. 99–111, Apr. 2015.

[5] A. Addae-Korankye, "Causes and control of loan default/delinquency in microfinance institutions in Ghana," *American International Journal of Contemporary Research*, vol. 4, no. 12, pp. 36–45, 2014.

[6] M. Semaw Henock, "Financial sustainability and outreach performance of saving and credit cooperatives: The case of Eastern Ethiopia," *Asia Pacific Management Review*, vol. 24, no. 1, pp. 1–9, Mar. 2019, https://doi.org/10.1016/j.apmrv.2018.08.001.

[7] N. Marwa and M. Aziakpono, "Technical and Scale Efficiency of Tanzanian Saving and Credit Cooperatives," *The Journal of Developing Areas*, vol. 50, no. 1, pp. 29–46, 2016.

[8] M. T. Pasara, A. Makochekanwa, and S. H. Dunga, "The Role of Savings and Credit Cooperatives (SACCOs) on Financial Inclusion in Zimbabwe," *Eurasian Journal of Business and Management*, vol. 9, no. 1, pp. 47–60, 2021, https://doi.org/10.15604/ejbm.2021.09.01.004.

[9] Mjatta, G.T, Akarro, and R.R.J, "Traits Associated with the Success or Failure of Emerging SACCOS in Tanzania : A Case of Four Regions in Tanzania Mainland," *Global Journal of Commerce & Management Perspective*, vol. 5, no. 4, pp. 29–38, 2016.

[10] P. Anania and A. Gikuri, "SACCOS and Members' Expectations: Factors Affecting SACCOS Capacity to Meet Members' Expectations," presented at the Co-operative Research Workshop - Moshi Co-operative University, Tanzania, Mar. 2015.

[11] J. Anderson, D. Hopkins, and M. Valenzuela, "The role of financial services in youth education and employment.," Consultative Group to Assist the Poor, Washington, DC, USA, Working Paper, Jun. 2019.

[12] C. Magashi, A. Sam, J. Agbinya, and J. Mbelwa, "Indicative Factors for SACCOs Failure in Tanzania," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11177–11181, Aug. 2023, https://doi.org/10.48084/etasr.5989.

[13] A. Gikuri and M. B. Sanka, "Status of SACCO's growth before and during JK Billion Fund In Tanzania Mainland," *International Journal of Community and Cooperative Studies*, vol. 6, no. 2, pp. 8–18, Jun. 2018.

[14] T. M. Mallya, "The Role of SACCOS on Improving Livelihood of Clients in Moshi Rural District.," MSc Thesis, The Open University of Tanzania, 2020.

[15] A. Bryman, *Social Research Methods 4th ed.* Oxford University Press, 2001.

[16] C. R. Kothari, *Research Methodology: Methods and Techniques*. New Delhi, India: New Age International, 2004.

[17] A. Rahman and M. N. A. Khan, "A Classification Based Model to Assess Customer Behavior in Banking Sector," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 2949–2953, Jun. 2018, https://doi.org/10.48084/etasr.1917.

[18] S. Joseph, N. Mduma, and D. Nyambo, "A Deep Learning Model for Predicting Stock Prices in Tanzania," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10517–10522, Apr. 2023, https://doi.org/10.48084/etasr.5710.

[19] T. S. Mian, "Evaluation of Stock Closing Prices using Transformer Learning," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11635–11642, Oct. 2023, https://doi.org/10.48084/etasr.6017.