Development of a machine learning model for loan default risk prediction in savings and credit cooperative organizations

Joseph Airo Anyimbi[1]
Raphael Angulu[2]
Jairus Odawa[3]

[1]joseph.airo4@gmail.com (+254 726 769 706)
[2]rangulu@mmust.ac.ke (+254 729 260 641)
[3]jodawa@mmust.ac.ke (+254 725 961 657)

[1,2,3] Masinde Muliro University of Science and Technology, Kenya

https://doi.org/10.51867/ajernet.7.3.141

ABSTRACT

Loan default remains a substantial challenge for Savings and Credit Cooperative Organizations (SACCOs), as it can increase credit risk, reduce profitability, and weaken financial sustainability. Although SACCOs increasingly use computerized information systems to manage lending activities, historical borrower data are often not fully utilized for predictive credit-risk assessment. This study developed an explainable machine learning model for loan default risk prediction in SACCOs. The study was guided by Information Asymmetry Theory and Credit Rationing Theory, which provide a basis for understanding how borrower information can reduce uncertainty and support credit allocation decisions. Four objectives guided the study: to identify factors associated with loan default risk, develop machine learning models for loan default prediction, evaluate their predictive performance, and provide an interpretable model for credit-risk assessment. A quantitative predictive analytics design was adopted using 1,000 records from the German Credit Dataset. Data preparation involved cleaning, categorical encoding, feature scaling, feature engineering, and class balancing using the Synthetic Minority Oversampling Technique (SMOTE). Logistic Regression and Random Forest models were developed using hyperparameter optimization and Stratified 5-Fold Cross-Validation. Model performance was evaluated using Accuracy, Precision, Recall, F1-score, ROC-AUC, Brier Score, confusion-matrix analysis, and false-negative analysis. SHapley Additive exPlanations (SHAP) was subsequently used to interpret the selected model. The results indicated that checking account status, age, credit amount, and loan duration were important predictors of default risk. Logistic Regression outperformed Random Forest across all reported evaluation measures, achieving Accuracy of 77.0%, Precision of 60.6%, Recall of 66.7%, F1-score of 63.5%, ROC-AUC of 0.795, and Brier Score of 0.162. It also recorded fewer false-negative predictions, with 20 compared with 26 for Random Forest. SHAP analysis identified checking account status and age as the most influential predictors. The study concludes that Logistic Regression with SHAP-based explainability provides a practical and interpretable approach to credit-risk prediction under the conditions examined. However, the model should be validated using institution-specific SACCO data before operational implementation.

Keywords: Credit Risk; Explainable AI; Loan Default Prediction; Machine Learning; SACCOs; SHAP

I. INTRODUCTION

Non-payment of loans increases credit risk, reduces income and may jeopardize financial sustainability. Therefore, loan non-payment remains a significant concern for Savings and Credit Cooperative Organizations (SACCOs). SACCOs are turning to computer-based information systems for loan management and transaction processing, but many credit evaluation methods still rely heavily on traditional assessment methods. These methods may not be fully exploiting historical loan and borrower data to discover tendencies related to default. Machine learning offers a way to improve credit risk assessment by looking at past lending data to find patterns and using those patterns to predict when borrowers will default. In the previous studies for credit risk prediction, ensemble techniques such as Random Forest, Logistic Regression, and Support Vector Machines are used. Random Forest can handle non-linear relationships and interactions between predictor variables. But Logistic Regression still has its place because of its simplicity, computational efficiency and interpretability. (Breiman, 2001). In addition to the aforementioned, the evaluation of various methodologies can serve to ascertain their suitability for forecasting loan defaults.

However, the financial decision making might require more than only the predictive performance. This is especially true when forecasts influence lending decisions, and credit officers need to know what drives a borrower's expected risk. Explainable Artificial Intelligence (XAI) aims to address the problem by offering techniques to comprehend machine learning predictions. For instance, SHapley Additive exPlanations (SHAP) can offer explanations

on an overall and individual level, measuring the contribution of individual predictors to model outputs . (Lundberg & Lee, 2017)

Another challenge is the suitability of popular credit-risk datasets for the SACCO environment. The German Credit Dataset is a popular benchmark for credit-risk modeling with 1,000-borrower data. It consists of borrower, financial, credit-history and loan-related attributes. However, since it was not collected from Kenyan SACCOs, it may not be a true representation of their institutional conditions, borrower characteristics or lending policies. Its use in this study is therefore not a clear evidence of its application to all Kenyan SACCOs but a basis for model development and methodological assessment. It would be necessary to validate using institution-specific SACCO data before operational deployment. A machine learning model was developed to predict loan default risk in SACCOs. Logistic Regression and Random Forest were constructed and evaluated using Accuracy, Precision, Recall, F1 Score, ROC-AUC, False Negative Analysis and Brier Score. The selected model was then tested on SHAP to identify the main drivers of the default predictions. The project aimed at demonstrating how explainability and machine learning can be used to provide more transparent and informed credit-risk assessments by SACCO information systems.

1.1 Statement of the Problem

However, there is still a challenge of evaluating the risk of loan default. Many SACCOs are increasingly using information systems. Conventional credit appraisal methods often rely on historical loans and they tend to have limited predictive power. Second, there is a dearth of literature on the application of machine learning in Kenyan cooperative financial institutions. Thirdly, the German Credit Dataset does not represent SACCO lending practices and borrower characteristics well. Fourth, many predictive models focus on high accuracy, at the expense of reliability and interpretability, and false negative errors. Finally, there is a relatively small body of research that investigated the integration of explainability with machine learning predictions in SACCO credit risk assessment.

Loan delinquency remains a challenge for Savings and Credit Cooperative Organizations (SACCOs) as it increases credit risk and can have negative implications on the quality and long-term sustainability of loan portfolios. While the advent of computerized information systems has contributed toward better documentation and management of lending activities, the credit appraisal procedures still make very limited use of historical data for predictive decision-making. To address this gap, the study proposed an interpretable machine learning model for predicting loan default risk with the goal of helping improve credit risk assessment in SACCO information systems. In this study, the approach is quantitative predictive analytics based on a publicly available German Credit Dataset with 1,000 borrower records. The data preparation phase included cleaning, categorical encoding, feature scaling, feature engineering, and class balancing through the Synthetic Minority Oversampling Technique (SMOTE).

1.2 Research Objectives

i. To identify the borrower and loan characteristics associated with loan default risk in SACCOs
ii. To develop Logistic Regression and Random Forest models for loan default risk prediction
iii. To compare the predictive performance of the developed machine learning models using appropriate performance metrics
iv. To develop and interpret a machine learning model for loan default risk prediction to support credit risk assessment in SACCOs.

II. LITERATURE REVIEW

2.1 Theoretical Review

The study is based on the Information Asymmetry Theory and Credit Rationing Theory that provide a basis for why accurate borrower information is required in lending decisions.

2.1.1 Information Asymmetry Theory

The Information Asymmetry Theory is a theory that involves a situation where one transacting party has better information than the other. In lending, borrowers typically know more about their ability to repay than do lenders. It is personal and intimate work to know and understand your financial position. This gap in information can make it difficult for lenders to distinguish a low credit risk borrower from a high credit risk borrower. This may cause adverse selection and moral hazard (Stiglitz & Weiss, 1981). The more effectively we use borrower and loan information, the more we can remove uncertainty from lending from a credit risk perspective. This research applies machine learning to borrower and loan data. Including historical data. History may show regularities that are not readily discoverable by conventional investigation. The loan default data can be used to build a model that helps loan officers evaluate borrower risk. SHAP would show the importance of different predictors and help credit officers understand the model.

2.1.2 Credit Rationing Theory

The Credit Rationing Theory is associated with Stiglitz and Weiss (1981) and explains why some creditors may limit the amount of credit they are willing to extend to some borrowers, even if the borrowers are willing to pay the going interest rate. The problem is fundamentally one of not knowing how risky borrowers are. If the lender screening process is weak, lenders will be more conservative in lending to minimize potential losses. The theory supports the need of better credit risk assessment for Saccos. A credible predictive model may be used to bifurcate borrowers based on their estimated probability of default and can also provide additional evidence for credit appraisal. This study discusses the use of machine learning to enhance borrower-risk classification over conventional lending assessment, relying on theory.

2.2 Empirical Review

2.2.1 Factors Associated with Loan Default Risk

Studies of credit risk have considered borrower attributes, loan characteristics and financial behavior as potential determinants of repayment outcomes. In a review of research on credit scoring, Abdou and Pointon (2011) found that credit assessment is generally based on a combination of borrower and account information to differentiate between high and low risk applicants. Also, their review showed that there is no one statistical technique that is best in every context of credit scoring. Therefore, the performance of the model depends in part on the nature of the data and the environment of lending. Crook et al. (2007) similarly observed that consumer credit-risk assessment increasingly relies on historical borrower information to estimate repayment behaviour. The review states that variables describing the applicant's financial position and credit history could be useful in distinguishing creditworthy borrowers from those with a higher probability of default. However, the studies reviewed explore different lending environments, making it difficult for their findings to be directly transferred to SACCO's.

Recent work on predictive modeling has demonstrated the value of multiple borrower and loan characteristics, rather than a single indicator. For example, Lessmann et al. (2015) compared 41 classification algorithms on eight real-world credit-scoring data sets and reported that predictive performance was dependent on data characteristics and modeling techniques. Their results are relevant for the present work in that they suggest that alternative models should be empirically compared rather than assuming that a more complex algorithm will automatically lead to better predictions. Addo et al. (2018) further demonstrated the usefulness of financial information for credit-risk modeling by comparing machine-learning and deep-learning approaches. The results illustrate how historical financial data can be converted to predictive signals of credit risk. However, the analysis was conducted in a wider financial-risk context and not only in SACCO information systems. This restricts the extent to which their findings can be directly transferred to SACCO lending environments.

The relevance of these findings to SACCOs is strengthened by the increasing use of information systems in financial operations. Digital records can provide information on borrower profiles, loan characteristics, account activity, and repayment behaviour that may be useful for systematic credit-risk assessment. However, the availability of such information does not necessarily mean that it is being used for predictive purposes. In many SACCO settings, credit assessment may still depend substantially on conventional appraisal procedures and staff judgement, creating an opportunity for data-driven approaches to complement existing practices.

Overall, the reviewed literature indicates that borrower characteristics, loan characteristics, and previous financial behaviour contain useful information for assessing credit risk. However, much of the empirical evidence has been generated from commercial banking, consumer-credit, and credit-card datasets. Evidence on machine-learning-based loan-default prediction in SACCO environments, particularly within the Kenyan context, remains comparatively limited. The present study addresses this gap by examining borrower, loan, and repayment-related characteristics and using them as predictors in machine-learning models for loan-default risk assessment.

2.2.2 Machine Learning Models for Loan Default Prediction

The empirical work has increasingly shifted from traditional statistical credit-scoring methods to machine learning methods. Abdou and Pointon (2011) noted that traditional statistical models are still important in credit scoring, but also noted the increased use of more sophisticated classification techniques. Their review concludes that the choice of the model should depend on the nature of the prediction problem and not on the assumption that one technique is better than any other. Lessmann et al. (2015) provided more robust empirical evidence by benchmarking 41 classification algorithms on eight credit scoring data sets. In their study, they found significant differences between the algorithms and demonstrated the importance of evaluating competing models on appropriate performance measures . The study is applicable to the current research in terms of providing support for the use of comparative modelling rather than the use of a single algorithm without empirical justification.

Addo et al. (2018) conducted a comparison between machine learning and deep learning models for credit-risk analysis and demonstrated that machine learning can efficiently extract predictive information from financial data. Their

results are in line with wider adoption of predictive analytics in the financial risk management. However, their study did not specifically look at SACCO lending systems. The existing literature thus supports the application of machine learning for credit-risk prediction but does not establish that more sophisticated models will always outperform simpler alternatives. This is relevant to the present study, which compares Logistic Regression and Random Forest to determine their relative predictive performance under the study conditions. The comparison allows us to evaluate predictive performance in terms of practical benefits of model simplicity and interpretability.

2.2.3 Predictive Performance of Credit-Risk Models

A single classification measure cannot provide a sufficient assessment of a credit-risk model's performance. Abdou and Pointon (2011) found that a range of evaluation criteria are used in credit scoring research, and stressed that the use of different measures can result in different assessments of model performance. This supports the use of multiple complementary metrics to evaluate loan-default prediction models. Comparative evaluation is particularly strongly supported empirically by Lessmann et al. (2015). Using a benchmark of 41 classifiers, they showed that the ranking of credit-scoring models can depend on the performance indicator used. They also investigated whether gains in forecasting accuracy are associated with meaningful managerial value.

Different modelling approaches can also result in different levels of credit-risk prediction performance as shown by Addo et al. (2018). Their results support the need to empirically evaluate models rather than choosing an algorithm because it is perceived as technically advanced. Despite these advances, most of the existing literature evaluates credit-risk models on data derived from banks, credit-card institutions, and other financial settings. Hence, the evidence in the simultaneous use of classification performance, discrimination, probability calibration and false negative analysis in SACCO-oriented credit-risk assessment is less developed. The present study fills this methodological gap by comparing the Logistic Regression and Random Forest on Accuracy, Precision, Recall, F1-score, ROC-AUC, Brier Score, confusion-matrix results, and false-negative predictions.

2.2.4 Explainability of Machine Learning Models in Credit Risk Assessment

In the case of model results affecting lending decisions, predictive accuracy may not be enough. In credit risk assessment, users need to understand why a borrower was assigned a particular level of risk. This has resulted in the use of explainable artificial intelligence techniques in conjunction with predictive modelling. Bussmann et al. (2020) studied explainable artificial intelligence within the framework of financial risk management. They argued that explainability can help financial institutions in understanding the behaviour of models and in enhancing the transparency of decisions related to risk. The work is relevant to the present study as it relates the prediction of machine learning with the practical need to interpret the outputs of the model in the financial decision-making environments. Lundberg and Lee (2017) introduced SHapley Additive exPlanations (SHAP), a unified approach to explaining the output of any machine learning model. The approach allows to assess the total importance of predictors and their contribution to particular predictions. This is a useful starting point to investigate which borrower characteristics impact the predictions of loan defaults.

However, Rudin (2019) introduces an important caveat. She argues that in high-stakes applications, interpretable models should be employed where they can deliver adequate predictive performance, rather than resorting to explanations of inherently opaque models unnecessarily. This viewpoint is especially relevant for the present study in that Logistic Regression offers a relatively interpretable model structure, and SHAP provides another way to investigate feature contributions. Hence, explainability in this study is seen not only as a technical add-on but as part of the practical assessment of model suitability. Explainable AI has increasingly been focused in the area of financial risk management but its application to SACCO loan-default prediction is relatively limited. The current study, therefore, builds on previous work by applying SHAP on the chosen predictive model and investigating the roles of borrower and loan features in the default predictions.

The empirical studies reviewed show that credit risk prediction can be made with the help of borrower characteristics, loan attributes and historical financial information. However, most of the existing studies are based on commercial banking, consumer credit, or credit card data with limited evidence from SACCO settings. Moreover, few studies combine model comparison, probability calibration, false negative analysis, and explainability within a single credit-risk assessment framework. These limitations highlight the need for more empirical studies on machine learning-based loan-default prediction in SACCO information systems.

2.3 Research Gap

While machine learning based credit risk assessment has witnessed a considerable progress, little empirical evidence exists on its application to SACCO loan-default prediction, especially in environments where borrower information is available through computerised information systems. The existing studies provide useful evidence about predictive modelling, but less attention has been paid to comparative analysis of interpretable and non-linear models using multiple performance measures, while explaining the factors underpinning their predictions. The present study

addresses this gap by examining factors associated with loan default, developing Logistic Regression and Random Forest models, comparing their predictive performance using complementary evaluation measures, and applying SHAP to interpret the selected model. The study therefore brings together prediction, comparative evaluation, and explainability within a single framework for intelligent credit-risk assessment in SACCO information systems.

III. METHODOLOGY

3.1 Research Design

The study adopted a quantitative predictive analytics research design to develop and evaluate standard ML models for prediction of loan default risk. The design was appropriate to the aim of the study of identifying patterns in historical credit data and using these patterns to predict default outcomes for previously unseen borrowers. Predictive modelling is a systematic approach for investigating relationships within existing data, without manipulating the study environment (James et al., 2021). The German Credit Dataset was used, a publicly available secondary data set of 1,000 borrower records and information about borrower characteristics, loan attributes, financial circumstances, and credit outcomes. The data set provides a systematic basis for the development and testing of alternative credit-risk classification models.

3.2 Data Preparation

Data preparation was carried out prior to model development to improve data quality and to make the dataset suitable for use in machine learning analysis. The data processing consisted of data cleaning, encoding of categorical variables, feature scaling, feature engineering, and treatment of class imbalance. The categorical variables were transformed to numerical variables so that they could be used in the selected algorithms. Numerical variables were scaled as necessary, especially for Logistic Regression, since the model is sensitive to the magnitude of the predictor variables (James et al., 2021).

The dataset was imbalanced, with more non-default cases than default cases. Therefore, the synthetic minority oversampling technique (SMOTE) was employed on the training data. SMOTE generates new synthetic minority samples by interpolating between existing minority samples, rather than copying minority samples (Chawla et al., 2002). This approach was only applied to the training data to prevent the contamination of the independent test data, which would result in data leakage.

3.3 Model Development

For the study, two supervised classification algorithms were chosen, namely Logistic Regression and Random Forest. Logistic Regression was chosen as an interpretable benchmark, as it estimates the probability of a binary outcome and has a long history of use in credit-risk and credit-scoring research (Abdou & Pointon, 2011; Crook et al., 2007). Also, its rather simple structure provides a useful basis for understanding the relationship between predictor variables and the probability of default.

The ensemble model that was chosen as benchmark was Random Forest, because it combines multiple decision tree models, which can capture non-linear relationships and interactions between predictor variables (Breiman, 2001). This component was included to allow the research to evaluate whether a more flexible ensemble approach could produce better predictive results relative to the fairly simple Logistic Regression model. The two models were trained using the prepared training data and subsequently evaluated on an independent test dataset. The comparison was intended to establish whether model complexity resulted in improved prediction within the study dataset rather than assuming beforehand that one algorithm would be superior.

3.4 Hyperparameter Optimization and Cross-Validation

The two models were tuned using Grid Search with Stratified 5-Fold Cross-Validation. Grid Search evaluated predefined combinations of hyperparameters, while cross-validation repeatedly trained and validated the models across different subsets of the training data. This approach reduced reliance on a single data split and supported more reliable model selection (James et al., 2021; Géron, 2022). The search for Logistic Regression was on the regularisation parameter and the penalty type. The search for Random Forest considered number of trees, maximum depth of the trees, minimum number of samples required to split a node and number of features to consider at each split. The best performing configurations were then used to train the final models.

3.5 Model Evaluation

Model performance was evaluated using several complementary measures: Accuracy, Precision, Recall, F1-score, ROC-AUC, confusion matrix analysis, false-negative analysis, and Brier Score. The use of multiple measures was intended to provide a more complete assessment than relying on Accuracy alone, particularly because credit-risk

datasets can contain unequal class distributions (Abdou & Pointon, 2011; Lessmann et al., 2015). Accuracy measured the overall proportion of cases correctly classified, while Precision measured the proportion of observations predicted to be defaulters that actually were defaulters. Recall was used to assess the model's ability to identify actual defaulters, while Precision measured the proportion of predicted defaulters who were correctly classified. The F1-score provided a balanced measure of Precision and Recall. To evaluate the models' ability to discriminate between defaulters and non-defaulters across different classification thresholds, we used the ROC-AUC (James et al., 2021). The Brier Score was added to test the reliability of the predicted default probabilities, where lower values indicate better probabilistic accuracy (Niculescu-Mizil & Caruana, 2005). Moreover, false-negative predictions were also studied as an error that classifies a borrower that subsequently defaults as a non-defaulter could lead to credit losses for a SACCO.

3.6 Model Explainability

SHapley Additive exPlanations (SHAP) were used to evaluate interpretability of the final model. SHAP uses Shapley values from cooperative game theory to estimate the contribution of individual predictors to the model outputs (Lundberg & Lee, 2017). This technique offers a global picture of feature importance as well as local explanations of individual predictions. Thus, SHAP was used to understand which variables contributed most to loan-default predictions and to analyse the direction of their contributions. This was supplemented with the analysis of predictive performance to give insight into how the model selected arrived at its predictions. Such transparency is particularly relevant for applications in financial risk, where the end-user may require understanding and scrutiny of the rationale behind automated or model-assisted decisions (Bussmann et al., 2020).

3.7 Ethical Considerations

The study used a secondary dataset available to the public and did not involve any direct interaction with individual borrowers. The analysis was thus focused on sensible use of available data, correct presentation of results, and proper interpretation of model results. The predictive model was designed to be a decision-support tool, not an autonomous lending mechanism. This distinction is important because model predictions should be a complement, not an automatic replacement, for institutional credit policies and professional judgement.

IV. FINDINGS & DISCUSSION

4.1 Factors Associated with Loan Default Risk

The first purpose was to investigate the factors related to risk of loan default. Checking account status, age, credit amount and loan duration were found to be important predictors of default when analysing borrower and loan characteristics . The results imply that default risk is influenced by a combination of financial, demographic and loan characteristics and not by one factor. Specifically, the checking account status was significant. Higher predicted default risk was associated with borrowers in less favorable account conditions. This finding is plausible because account status is a proxy for the borrower's short-term financial position and liquidity. However, the finding should be considered an association rather than evidence that default is directly caused by account status.

Age was also an important predictor. Its contribution suggests that differences in borrower characteristics are associated with differences in default risk. Age may be a proxy for unobserved differences in financial experience, income stability, or borrowing behaviour, although these were not separately identified in the analysis. Amount of credit and term of loan were also predictive of default . The size and composition of a credit obligation thus provide useful information for separating borrowers by risk. The results are consistent with the credit-scoring literature, which identifies borrower characteristics, account information and loan attributes as relevant predictors of credit performance (Abdou & Pointon, 2011; Crook et al., 2007). They also support Lessmann et al. (2015) who state that credit-risk prediction usually depends on multiple borrower and credit characteristics and not on a single predictor. The present findings thus support the importance of a multidimensional approach to the assessment of credit risk.

This indicates that a credit appraisal for SACCOs can be improved by combining borrower and loan indicators. It is also worth noting that the findings are based on the German Credit Dataset and should not be assumed to be generally applicable to Kenyan SACCO borrowers without further validation using data from respective institutions.

4.2 Development of Machine Learning Models

The second objective was to develop machine learning algorithms to predict loan default. After pre-processing the data, balancing classes, optimizing hyperparameters, and using stratified 5-fold cross-validation, we constructed models using Logistic Regression and Random Forest. Logistic Regression was selected as an interpretable classification model that estimates the probability of a binary outcome. Its relatively simple structure makes it useful where predictions need to be examined and communicated to decision-makers. The model has also been widely applied in credit-scoring research (Abdou & Pointon, 2011; Crook et al., 2007). Random Forest was selected as a comparative

ensemble approach because it can capture non-linear relationships and interactions among predictors through numerous decision trees (Breiman, 2001). The comparison was therefore intended to determine whether greater model flexibility would provide an advantage over the more interpretable Logistic Regression approach.

To address the class imbalance, SMOTE was used, which creates synthetic minority-class observations from existing minority observations (Chawla et al., 2002). Note that oversampling has been applied only to the training data so as to preserve the independence of the test set. Then, hyperparameter optimization and cross-validation were applied to find appropriate model configurations and to mitigate reliance on a single training-validation split (James et al., 2021; Géron, 2022). The modelling process consequently provided a consistent basis for comparing the two algorithms. Model selection was determined by their observed performance rather than by assuming that a more complex algorithm would necessarily produce better predictions.

4.3 Comparative Predictive Performance

The third objective evaluated the predictive performance of the developed models using Accuracy, Precision, Recall, F1-score, ROC-AUC, Brier Score, confusion-matrix results, and false-negative analysis.

Table 1

Comparative Performance of the Machine Learning Models

Performance Metric	Logistic Regression	Random Forest	Better Model
Accuracy	0.770	0.705	Logistic Regression
Precision	0.606	0.507	Logistic Regression
Recall	0.667	0.567	Logistic Regression
F1-score	0.635	0.535	Logistic Regression
ROC-AUC	0.795	0.770	Logistic Regression
False Negatives	20	26	Logistic Regression
Brier Score	0.162	0.170	Logistic Regression

Logistic Regression was found to outperform Random Forest consistently across all reported measures. Compared to Random Forest, Accuracy was 77.0% against 70.5%. Logistic Regression had a Precision of 60.6%, and Random Forest had a Precision of 50.7%. Recall was 66.7% and 56.7%, respectively. The F1 scores were 63.5% and 53.5%, respectively. This difference is especially important from a credit-risk perspective with Logistic Regression making 20 false negatives versus 26 for Random Forest. A false negative is a borrower that is misclassified as a non-defaulter, but in fact is a member of the default class. Such errors can lead to avoidable credit losses for lenders. The smaller number of false-negative predictions provides additional evidence in favour of Logistic Regression under the evaluation conditions of the study.

The findings also challenge the assumption that a more complex algorithm will necessarily perform better. Random Forest can model non-linear relationships and interactions, yet its performance on the independent test data was lower than that of Logistic Regression. This result is consistent with Lessmann et al. (2015), who found that the relative performance of credit-scoring algorithms varies across datasets and that algorithmic complexity alone does not guarantee superior predictive performance.

4.3.1 ROC-AUC Analysis

The ROC-AUC results further demonstrated the discriminatory ability of the models. Logistic Regression with a ROC-AUC = 0.795, as opposed to 0.770 for Random Forest. A higher value indicates a better Logistic Regression model in distinguishing between defaulting and non-defaulting borrowers at different classification thresholds. The ROC-AUC result provides further evidence obtained from Accuracy, Precision, Recall and F1-score since it evaluates discrimination without relying on one specific classification threshold. Hence, the result supports the selection of Logistic Regression and also shows why model evaluation should be done using several complementary measures and not just Accuracy (James et al., 2021). The finding, however, should be interpreted in light of the present study. An ROC-AUC of 0.795 suggests useful discriminatory ability on the test data, but it does not demonstrate that the model will perform similarly when applied to different SACCOs or future lending portfolios.

4.3.2 Probability Calibration and Brier Score

The predicted probabilities of default were assessed using the Brier Score to examine their probabilistic accuracy. Logistic Regression recorded a Brier Score of 0.162, while Random Forest recorded 0.170. Because lower Brier Scores indicate better probabilistic prediction accuracy, the result suggests that Logistic Regression produced comparatively more accurate probability estimates under the conditions of the study. The Brier Score provides more information than the classification accuracy because credit-risk assessment may require an estimate of the probability

of default rather than only a defaulter/non-defaulter classification. More reliable probability estimates may help to rank borrowers by their relative risk level and may support a differentiated credit assessment. The reliability of the predicted probabilities should be assessed independently from the classification performance since two models can share the classification performance but differ in the quality of their probability estimates (Niculescu-Mizil and Caruana, 2005).

Therefore, the lower Brier Score provides additional support for Logistic Regression. The model in this study not only classified borrowers more effectively across the principal performance measures, but also generated comparatively more accurate probability estimates than Random Forest. This finding further supports the choice of Logistic Regression as the final model for further SHAP-based interpretation and credit risk assessment.

4.3.3 Model Selection

Taking all the results into account, the final model was Logistic Regression. It had better Accuracy, Precision, Recall, F1-score and ROC-AUC and less false-negative predictions and Brier Score than Random Forest. The choice was therefore based on the overall evidence and not on any single performance measure. The Logistic Regression offered a good compromise between predictive performance, probability reliability and interpretability. In this study, the flexibility offered by Random Forest to model complex relationships did not translate into better test-set performance.

This finding should not be interpreted as an evidence that Logistic Regression is always better than Random Forest. Instead, it indicates that Logistic Regression was the more suitable predictive solution when considering the data and modelling used in this study. This cautious interpretation is consistent with the comparative credit-scoring literature where model performance is likely to depend on the characteristics of the dataset and modelling environment (Lessmann et al., 2015).

4.4 Model Interpretability Using SHAP

The fourth aim was to explain the chosen Logistic Regression model using SHapley Additive exPlanations (SHAP). The predictive performance itself does not explain why a borrower receives a certain risk prediction. For credit risk applications, understanding the factors that contribute to a prediction can enable reviewers to gain useful insights into the model's outputs. Therefore, the contribution of each predictor to the model prediction was estimated using SHAP. The method is based on Shapley values and provides a systematic way of probing the contribution of individual features to the model output (Lundberg & Lee, 2017).

4.4.1 SHAP Feature Importance Analysis

SHAP analysis shows that checking account status and age are among the most important factors influencing the predictions of loan defaults. The prominence of these variables suggests that the model used them extensively to differentiate between borrowers on the basis of predicted risk. The relevance of checking account status is consistent with previous credit scoring studies where account and financial information are shown to be useful indicators of creditworthiness (Abdou & Pointon, 2011; Crook et al., 2007). In practice, account status may serve as a proxy for a borrower's short-term financial position. However, it should not be taken in isolation as credit risk is affected by a combination of characteristics.

Age was also a feature in the SHAP analysis. Its contribution implies that the demographic information had predictive value in the dataset. The finding doesn't mean age itself causes people to default. Instead, age may be a proxy for differences in financial circumstances or borrowing patterns that are reflected in the data that is available. Therefore, the SHAP results give an interpretation to the predictive results. Instead of just reporting the accuracy of the model for classifying borrowers, the analysis identifies the variables that had the largest contributions to those predictions. This bolsters the more general argument that explainability can improve transparency in financial-risk applications (Bussmann et al., 2020; Lundberg & Lee, 2017).

4.4.2 Implications for Credit Risk Assessment

Together, the results indicate that Logistic Regression provided a good balance between predictive performance and interpretability. Its predictions can be analysed according to identifiable borrower and financial characteristics, while SHAP provides information about the contribution of individual predictors. For SACCOs, such a model could assist in screening borrowers, ranking risk and identifying applications that need closer scrutiny. It needs to be, however, a decision-support tool and not an automatic lending tool. A prediction of a high risk does not guarantee a borrower will default and lending decisions should still consider institutional policies and professional judgement. The findings also underline the necessity of keeping reliable digital records of borrowers. Predictive models are only as good as the data they are trained on. A more robust foundation for a data-driven credit-risk assessment would be laid by consistently capturing borrower characteristics, loan terms, account information and repayment history.

4.5 Overall Discussion of Findings

Together the findings address the four objectives of the research. First, checking account status, age, credit amount, and loan duration were found to be significant predictors of default risk. Second, a structured predictive modelling framework was successfully developed including Logistic Regression and Random Forest . . Third, Logistic Regression outperformed Random Forest on the selected classification, discrimination, error, and calibration measures. Finally, SHAP provided an interpretable perspective of the factors contributing to the predictions of the selected model.

The results contribute to the existing research on credit risk by incorporating prediction, model comparison, probability assessment, error analysis, and explainability within a single analytical framework. Machine learning and statistical methods have been shown to be important in credit scoring (Abdou & Pointon, 2011; Lessmann et al., 2015) and SHAP provides an approach to study the influence of each predictor on model predictions (Lundberg & Lee, 2017). The present study links these strands by demonstrating how predictive performance and interpretability can be considered jointly when selecting a credit-risk model.

The findings, however, require contextual caution. The analysis used the German Credit Dataset instead of real lending data from Kenyan SACCOs. The results therefore provide evidence of the feasibility and comparative performance of the modeling approach, but not that the same predictors, relationships, or performance levels would necessarily occur in Kenyan SACCOs. Therefore, validation with SACCOs' own data is required before operational implementation. All in all, from the evidence, it can be inferred that Logistic Regression with SHAP-based interpretation is the most suitable model among the ones explored in this study. Its advantage was more than just predictive precision. The model also showed better discrimination, fewer false-negative predictions, better probability estimates, and increased interpretability. This combination provides a reasonable foundation for the development of transparent, evidence-based decision support for credit-risk assessment in SACCO information systems.

V. CONCLUSION & RECOMMENDATIONS

5.1 Conclusion

This study aimed to develop an explainable machine learning model for loan default risk prediction in a Savings and Credit Cooperative Organization. The results suggest that historical borrower and loan information can offer useful signals for differentiating borrowers by default risk. It was found that, in particular, checking account status, age, credit amount, and loan duration were important predictors in the dataset analyzed. The results of the comparison between Logistic Regression and Random Forest indicated that the simpler model was better suited for the conditions of this study. Logistic Regression was better in terms of Accuracy, Precision, Recall, F1-score, and ROC-AUC; it had fewer false-negative predictions and a lower Brier Score. It achieved a ROC-AUC of 0.795 and a Brier Score of 0.162, reflecting useful discriminative and probabilistic performance on the independent test data. Therefore, the results mean that using more complex models does not necessarily lead to better predictions of credit risk.

The study also highlighted the advantages of explainability for predictive credit-risk modelling. SHAP analysis confirmed that the checking account status and the age were among the most important predictors for the predictions of the selected model. This provided additional insight into how the model produced its risk estimates, and hence improved the practical interpretability of the predictive approach. To summarise, the study concludes that Logistic Regression along with SHAP-based explainability is a useful and interpretable method of predicting loan-default risk in the analytical context studied. However, the findings cannot be directly generalised to all SACCOs since the analysis was based on the German Credit Dataset and not institution specific Kenyan SACCO data. The proposed approach should thus be considered as a starting point for further validation and development rather than an out-of-the-box automated lending system.

5.2 Recommendations

SACCOs may also consider the possibility of incorporating machine learning-based decision-support capabilities in their existing credit appraisal process, especially where there is adequate historical borrower and repayment data. These tools can supplement traditional assessment procedures by helping credit officers identify borrowers who may need closer scrutiny. The results suggest that Logistic Regression can be considered as a baseline model, as it performed consistently better than Random Forest on the evaluation measures used in this study and was relatively easy to interpret. SACCOs should validate the model with their historical lending data before operational adoption. An institution-specific validation would explain whether the predictors discovered in this study, such as checking account status, age, credit amount, and loan duration, are still useful in the local lending environment. This would also enable the model to be retrained to reflect the characteristics of individual SACCOs, their members, loan products, and repayment patterns.

SACCOs should also work to improve the quality and completeness of their digital borrower records. Better information on borrower profile, loan amount, repayment period, account status and repayment history would provide

a better basis for predictive credit-risk assessment. Poor quality or incomplete records would undermine the reliability of any machine learning system regardless of the algorithm used. Finally, when implementing machine learning for credit-risk assessment, explainability should be considered along with predictive performance. Explainability techniques such as SHAP can be used to understand the factors driving the model predictions and provide a basis for reviewing unusual or high-risk cases. Machine learning predictions should therefore be used to support, not replace, professional judgment and well-established lending policies.

REFERENCES

Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management, 18*(2–3), 59–88. https://doi.org/10.1002/isaf.325

Addo, P. M., Guegan, D., & Hassani, B. (2018). Credit risk analysis using machine and deep learning models. *Risks, 6*(2), Article 38. https://doi.org/10.3390/risks6020038

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2020). Explainable AI in fintech risk management. *Frontiers in Artificial Intelligence, 3*, Article 26. https://doi.org/10.3389/frai.2020.00026

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953

Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research, 183*(3), 1447–1465. https://doi.org/10.1016/j.ejor.2006.09.100

Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1

Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research, 247*(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems 30* (pp. 4765–4774). Curran Associates, Inc.

Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)* (pp. 625–632). Association for Computing Machinery. https://doi.org/10.1145/1102351.1102430

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Stiglitz, J. E., & Weiss, A. (1981). Credit rationing in markets with imperfect information. *American Economic Review, 71*(3), 393–410.