Journal of Economics, Finance and Management Studies

ISSN (print): 2644-0490, ISSN (online): 2644-0504

Volume 09 Issue 02 February 2026

Article DOI: 10.47191/jefms/v9-i2-33, Impact Factor: 8.317

Page No: 917-935

Financial Distress Prediction in Kenyan Saccos: A Comparative Analysis of Machine Learning Models

Dr. Dennis Mucee Ncurai, PhD*[1], Prof. Prashnant Priyadarshi, PhD[2]

[1,2]Gugliemo Marconi University

ABSTRACT: Financial distress prediction is crucial for assessing the economic health of any organization or nation; as a whole. This research focuses on developing a robust machine learning model to predict financial distress in Kenyan Savings and Credit Cooperative Organizations (SACCOs). Notably, this is one of the first studies to explore financial distress prediction specifically within the context of Kenyan SACCOs. The specific objectives included the assessment of the performance of machine learning predictive models using both financial and non-financial variables, the identification of predictors that significantly affect financial distress of Kenyan SACCOs, and determining which features are the most predictive. The study evaluates the effectiveness of various machine learning algorithms, including Decision Tree, K-Nearest Neighbors, Logistic Regression, Artificial Neural Networks, Naive Bayes Classifier, Random Forest, and Support Vector Machine, through nine performance evaluation metrics, with the ROC-AUC score as the primary measure. Results indicate that while K-Nearest Neighbors achieved an ROC-AUC score of 0.78, an ensemble model using Stochastic Gradient Boosting reached 0.81.

KEYWORDS: Financial distress, Artificial Intelligence, machine learning, predictive modelling, ensemble models, SACCOs, Kenya

1. INTRODUCTION

1.1 Background and Motivation

Financial distress is a common problem witnessed in companies in any country and industry. Alaminos and Ferna´ndez (2019) has referred to it as a global problem which has brought an element of control due to the various financial crises. This has triggered a myriad of reforms within various sectors and industries to establish early warning systems to predict financial distress of companies. Detection of early warning signals of financial risks is important to various stakeholders more so the relevant government agents (Loan and Tuan, 2019). The Financial distress prediction model is the model to identify the likelihood of being financially distressed of a company (Loan and Tuan, 2019).

For the management of such companies can relook at their strategies while the government can establish remedial measures to avoid or minimize financial losses to the economy. Accurate financial distress prediction supports the creditors and investors in making investing decisions to manage investment risks (Loan and Tuan, 2019). Left unnoticed then the related financial risks because of financial distress are catastrophic to the economy at large. In recent times, the global financial crisis in 2008 has made governments respond with sizeable investments to avoid exposure and shield companies to such occurrence. Of importance is that an accurate prediction of a company's financial situation will assist managers in initiating alternate solutions to prevent distress prior to it happening. It also assists investors in adjusting their investment strategies and lowering the expected loss on investment.

Savings and Credit Cooperative Organizations (SACCOs) are member owned organizations formed based on common and shared interests within a common bond. The SACCO sub sector is supervised, licensed and regulated by the Sacco Societies Regulatory Authority (SASRA) as the government's principal agency. Just like other organizations, financial distress has threatened the sustainability of SACCOs such that they have not been able to absorb their operational losses. According to Financial Sector Regulators (2022) and Kibue and Mang'ana (2022), some SACCOs have been grappling with fluctuating performance over recent years. The situation has been exacerbated with collapse of some (Kibue and Mang'ana (2022) and non-compliance with the regulatory frameworks. This has led to the losses being absorbed by members' savings and share capital which leads to their impairment. Due to this difficult situation faced by SACCOs, some of them have been shut down by the regulator with some receiving instructions to operate conditionally.

Prediction of financial distress plays a crucial role for SACCOs, other financial institutions, including banks, regulators, and investors (Huang, Wang, and Jiang 2024; Tsai et al. 2021). Early warning systems facilitate identifying potential problems within a company's financial health for proactive measures to be undertaken. Early anticipation of potential financial distress enables financial institutions to take appropriate actions before a situation spirals out of control. Late detection of financial distress poses the risk of significant financial losses, which can have a ripple effect on the entire financial system. Reliable early warning systems and proactive measures ensure stabilization of the entire financial system. Troubled institutions are detected early enough and supported timely with appropriate interventions. Zhang, Chi, Zhou and Liu (2025) opine that predicting financial distress is vital for banks, regulators, and investors to manage risks and make informed decisions. SACCOs and Financial institutions can use accurate financial distress prediction models to provide valuable insights for making informed decisions, risk assessment and effective management, better assessment of creditworthiness of potential borrowers and optimization of resources on areas where the risk of financial distress is highest.

1.2 Problem Statement

Prediction of financial distress has been a critical concern in the different sectors of the economy. Even though previous studies (Ooko, Jagongo and Mbewa, 2013; Olando, Mbewa and Jagongo 2012 and Kabaiya, 2012); have shown the presence of financial distress in SACCOs in Kenya, there are limited studies in this context. Evidently, the SACCO sub sector plays a significant role in the economy and households. However, their immense economic contributions are threatened considering multiple financial and governance issues. Previous literature outside SACCO contexts suggests that predictive models can be used to detect financial distress early enough for appropriate action to be taken. The literature on the financial distress prediction of Kenyan SACCOs seems quite economical. Huang *et al*. (2022) emphasizes the need for an accurate prediction of financial distress to allow the concerned institutions to build an early warning system and develop legal frameworks to avoid risk contagion. Looking at the SACCO context, there is need for a more robust predictive models that can combine both financial and non-financial variables as opined by Balasubramanian and Natarajan (2019), Situm (2023) and Zeng, Li, Yang and Li (2020).

1.3 Objectives of the Study

The main objective of the study was a comparative analysis of Machine Learning models in the Prediction of Financial Distress in Kenyan SACCOs. The specific objectives were:

1. To assess the performance of machine learning models.
2. To identify significant predictors of financial distress.
3. To determine the most predictive features.

2. LITERATURE REVIEW

2.1 Overview of Financial Distress Prediction

Financial Distress Prediction can be divided into two categories: statistical methods and machine learning (ML) methods. The traditional statistical and mathematical methods include linear discriminant analysis, multiple discriminant analysis, factor analysis, Logit regression, Probit regression, univariate analysis, and probabilistic modelling. Traditional statistical and mathematical methods begin appearing in the 1960's, with univariant data analysis to predict credit risk which was performed by Beaver (1966). Two years later in 1968, Altman pioneered the use of discriminant analysis in the field for prediction of financial distress which has elicited interest from researchers worldwide. Altman (1968) adopted a multivariate discriminant analysis in his famous Z-score model showed that the model had a significantly higher capacity of prediction in the year before bankruptcy than the univariate discriminant analysis models. Later in 1980, logistic regression was pioneered by Ohlson which was used as a way of overcoming restrictive assumptions of discriminant analysis, such as normality and equal covariance. The Probit model was used in 1984 by Zmijewski to predict financial distress.

These traditional statistical and mathematical methods were of significant influence on predictions of financial distress, though they have inherent limitations that make them restrictive. Khedr and El Bannany (2020) have observed that initial works in this field commonly used traditional statistical techniques namely discriminant analysis and the Logit model to predict business failure. However, these statistical techniques have limitations making their applications unreliable. Hence, Gregova, Valaskova, Adamko, Tumpach and Jaros (2020) among others, have considered the use of ML-based data mining techniques as an alternative due to their high accuracy and superior performance.

Advancement in Artificial Intelligence (AI) and Machine Learning (ML) has led to the development of more accurate predictive models. AI methods were first applied in 1990 by Bell, Ribar and Verichio when they predicted problems in the banking sector with their comparative study of neural networks and statistical models to demonstrate the superiority of artificial neural networks

(ANN). Since then, scholars and business practitioners have adopted different ML techniques such as Classification, Logic Regression, Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), and Artificial Neural Networks (ANN), and deep learning models for predictive modelling.

Some studies have shown that traditional techniques outperform modern techniques in predicting financial distress. Loan and Thuan (2019) found that Logit models outperform SVM models. Loan and Thuan (2019) called for further investigation on development of models using other ML classifiers. Other studies (Bell, Ribar and Verichio, 1990; Yousaf, Jebran and Wang, 2021; Tron, Dallocchio, Ferri and Colantoni, 2022) have shown ML techniques are superior accuracy as compared to traditional techniques. For instance, Li and Wang (2017) found that ML methods have superior accuracy to statistical methods in predicting bankruptcy in Chinese-listed companies. Gregova et al. (2020) found that neural network models yielded better results as compared to selected traditional and ML methods though all models demonstrated high discrimination accuracy and similar performance. Hence, it is concluded that ML methods are superior to statistical methods in predicting financial distress and earlier detection is important.

2.2 Machine Learning in Financial Distress Prediction

Artificial Intelligence (AI) was founded as an academic discipline in 1956 after experiencing several waves of optimism (Crevier, 1993). Machine learning (ML) is a subfield of AI. The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people (Tagliaferri, 2017). Arthur Samuel in 1959 and later, Tom Mitchell in 1997 have opined that ML is a field where computers can learn from experience or without being explicitly programmed and which improves with time. ML techniques were developed to model cognitive processes directly from previous experience. Each technique consists of two steps: learning and inference phases. Machine Learning has gained popularity by being used every day to make critical decisions in predictive modelling amongst other applications.

Traditional machine learning encompasses various algorithms that learn from data to make predictions or decisions. Machine learning techniques are broadly classified as supervised, unsupervised, semi-supervised and reinforcement learning depending on the business problem at hand (Villani, 2018). Supervised learning uses labeled data to train models. Supervised learning algorithms are generally categorized into Classification and Regression algorithms, each with different algorithms suited to different types of problems. The key difference between Classification and Regression algorithms lies in the nature of the output. Regression deals with continuous values while classification deals with categorical values.

This study was a classification problem since it aimed at developing a financial distress prediction model. It can be either a binary classification problem or a multi-class problem too (Waseem, 2023). Specifically, this study was of binary classification nature with two classes: financially distressed or not. Classification algorithms fall under two types: Linear Models (Logistic Regression and SVM) and Non-Linear Models (KNN, Kernel SVM, Naïve Bayes, Decision Tree Classification and RF Classification).

2.2.1 Logistic Regression Algorithm

Logistic Regression Algorithm is a classification algorithm that uses one or more independent variables to determine an outcome. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more independent variables. The outcome is measured with a dichotomous variable meaning it will have only two possible categorical or discrete values. It quantitatively explains the factors leading to classification. It gives the probabilistic values which lie between 0 and 1. The curve from the logistic function indicates the likelihood of something, for instance in the prediction of financial distress, it will indicate whether a given organization is financially distressed or not.

2.2.2 Support Vector Machine Algorithm

Support Vector Machine (SVM) is a predictive analysis algorithm, widely used in classification and regression problems. SVM classifier represents the training data as points in space separated into categories by a gap as wide as possible. Then, a Hyper-Plane is found which best segregates the classes in N-dimensional space. New points are then added to space by predicting which category they fall into and which space they will belong to. The SVM has been applied as is or in some varied form to design FDP models with superior performance as compared to other competing models.

2.2.3 Random Forest Algorithm

Random forest is an ensemble learning method in classification and regression problems. Dwivedi (2021) found that Random Forest classifier is at the top when it comes to classification tasks. The superior performance of Random Forest Algorithm could be attributed to the reason it comes from ensemble methods that are combinations of different or the same algorithms that are used in classification tasks. It operates by constructing a multitude of decision trees at training time and outputs the class that is the mode of the classes or classification or mean prediction(regression) of the individual trees. A random forest is a meta-estimator that fits several trees on various subsamples of data sets and then uses an average to improve the accuracy in the model's

predictive nature. The sub-sample size is always the same as that of the original input size, but the samples are often drawn with replacements.

2.2.4 Artificial Neural Networks Algorithm

An Artificial Neural Network learning algorithm is a computational learning system that uses a network of functions to understand and translate a data input of one form into a desired output, usually in another form. A neural network consists of neurons that are arranged in layers, they take some input vectors and convert it into an output. The process involves each neuron taking input and applying a function which is often a non-linear function to it and then passing the output to the next layer. Weighing is applied to the signals passing from one layer to the other, and these are the weighing that are tuned in the training phase to adapt a neural network for any problem statement.

2.2.5 Decision Tree Algorithm

Decision tree is used in prediction tasks, clustering, and classification (Sreedharan, Khedr and El Bannany, 2020). Decision Tree Algorithm works by the data being continuously divided into different categories according to certain parameters. It builds the classification model in the form of a tree structure. It utilizes the if-then rules which are equally exhaustive and mutually exclusive in classification. The process goes on with breaking down the data into smaller structures and eventually associating it with an incremental decision tree. The final structure looks like a tree with nodes and leaves. The rules are learned sequentially using the training data one at a time. Each time a rule is learned, the tuples covering the rules are removed. The process continues the training set until the termination point is met. The tree is constructed in a top-down recursive divide and conquer approach. A decision node will have two or more branches, and a leaf represents a classification or decision. The topmost node in the decision tree that corresponds to the best predictor is called the root node, and the best thing about a decision tree is that it can handle both categorical and numerical data.

2.2.6 K-Nearest Neighbor Algorithm

K-Nearest Neighbors (KNN) algorithm is a popular Machine Learning algorithm that uses the Supervised Learning technique for classification and regression tasks, though it is mostly used for classification problems (Srivastava, 2024). It uses proximity to make classifications or predictions about the grouping of an individual data point with similar labels or values. The KNN algorithm is a non-parametric algorithm, hence makes no assumptions about the underlying data. KNN algorithm has been considered as a 'lazy learner' as learning does not take place immediately. It stores training dataset and performs an action on it during classification upon receiving new data, it classifies it into a category that is very similar to the new data. To determine the closet nearest points for a query points, KNN algorithm uses distance metrics with the most common ones used in ML problems includes Euclidean Distance, Manhattan Distance, Hamming Distance, and Minkowski Distance.

2.2.7 Ensemble Learning Technique

Ensemble learning combines multiple simple models to create a stronger and smarter model. There are importantly two types of ensemble learning: Bagging that combines multiple models trained independently, and boosting that builds models sequentially, each correcting the errors of the previous one. Previous studies (Sankhwar et al., 2020; Uthayakumar et al., 2020) have shown that ensemble learning outperforms individual algorithms. However, Sreedharan, Khedr and El Bannany (2020), Gregova et al. (2020) and Zhu et al. (2021) seem to suggest that individual models together with deep learning models are the most efficient machine learning techniques in producing predictive models. El Bannany et al. (2021) compared the output of various single machine learning classifiers and ensemble techniques to predict financial distress and found that the ensemble technique has significantly greater prediction accuracy.

2.3 Empirical Studies on Financial Distress Prediction

Ledhem (2022) applied LASSO regression, RF, ANN and KNN models for predicting the financial performance of Islamic banking in Indonesia. Ledhem (2022) found that KNN technique and closely followed by RF had the best-predicting accuracy and gives the optimal knowledge from the financial performance of Islamic banking determinants in Indonesia. Earlier, Muñoz-Izquierdo et al. (2020) developed a Logit prediction model to predict corporate financial distress of 1,821 Spanish distressed private firms by combining accounting and auditing data. They found that only accounting models register a classification accuracy of 77%, but when combined models of accounting and auditing data exhibit considerably higher accuracy of 87%.

Halim, Shuhidan and Sanusi (2021) investigated the performance of deep learning models in predicting the financial distress among the Malaysian public listed corporation over the time-series data. In addition, the study compared the performance of logistic regression, support vector machine, neural network, decision tree and the deep learning models on single-year data. Halim, Shuhidan and Sanusi (2021) found that Deep learning models achieved above 90% accuracy and above with long short-term memory and gated recurrent unit getting 93% accuracy. Further, deep learning models were found to have a consistent good performance compared to the other models over single-year data.

Gregova *et al.* (2020) sought to compare selected traditional and machine learning methods (logistic regression, random forest and neural network) to identify a predictive model that can work in industrial enterprises operating in the specific Slovak environment. Study findings show that all models demonstrated high discrimination accuracy and similar performance, with neural network models yielding better results. Paule-Vianez, Gutiérrez-Fernández and Coca-Pérez (2020) sets to build the first model of ANN to predict the financial distress in the Spanish banking system. ANN was used because of their greater effectiveness in predicting stressful situations. Study findings show that a prediction model can predict short-term financial distress with an overall accuracy of more than 97% for predicting all cases in which an entity has short-term financial problems.

Crespí-Cladera, Martín-Oliver and Pascual-Fuster (2021) combined both the stress test methodology and the probabilistic model of failure to detect which firms could face financial distress situations among Spanish hospitality firms during the Covid-19 disaster. The study relied on parameters estimated with data from the Great Recession (2008–2013) to predict which firms would face financial distress. Crespí-Cladera, Martín-Oliver and Pascual-Fuster (2021) found that both methods coincide to predict that 25% of these firms will face a financial distress situation if revenues drop 60%, with this forecast raising up to 32% of firms if revenues drop 80%.

Zhu, Zhu and Lin (2023) sought to address the gap concerning few studies on the dynamic prediction of financial distress in the financial sharing environment. Zhu, Zhu and Lin (2023) asserted that the dynamic prediction of financial distress can monitor the financial status of an enterprise in real time and provide evidence for financial analysts. The study employed the probabilistic neural network to construct a dynamic prediction model of financial distress and used experimental results to verify the effectiveness and feasibility of the constructed model. The model proved to be effective in giving dynamic predictions on the financial distress in the financial sharing environment.

Considering the limited number of studies on the early detection of crises and insolvencies in the tourism industry (hotels and restaurants), Situm (2023) sought to address this research gap by analyzing the influence of financial and non-financial variables on the financial distress of the tourism businesses in Austria from 2005 to 2015 inclusive. Study findings show that variables describing the endogenous unsystematic risk of tourism businesses make the greatest contribution to explaining financial distress, while exogenous unsystematic and exogenous systematic risk variables show little or no relevance.

Tron *et al.* (2022) analyzed the relationships between corporate governance characteristics and financial distress status. They used a unique set of non-listed Italian Unlikely to Pay positions to compare the predictive performance of a Logit and a RF model using the ROC curve. Tron *et al.* (2022) found that the use of corporate governance variables increases the accuracy of the RF technique and influences the success of the turnaround process. Further, RF technique was confirmed as a significant ability to outperform the Logit model in terms of accuracy. The study concluded that financial variables are the most important indicator in predicting the financial default of a company.

Yousaf, Jebran and Wang (2021) applied different board diversity attributes in predicting the financial distress of Chinese A-listed companies during 2007–2016. Yousaf, Jebran and Wang (2021) found that board diversity attributes can significantly predict the financial distress of firms. Though machine learning models performed better, the best model in terms of Type I error and accuracy is RF. The study being first of its kind shows how different attributes of diversity can predict financial distress in Chinese firms.

Zeng *et al*. (2020) applied feature selection of sparse algorithm to explore the financial distress prediction of 377 listed companies in Shanghai and Shenzhen stock exchanges of China from 2015 to 2019 based on grouping sparsity. Zeng *et al*. (2020) proposed a corporate financial distress prediction model that is better in line with the practice and combines the grouping sparse principal component analysis of financial data, corporate governance characteristics, and market transaction data with support vector machine. Study findings show that this method can improve the prediction efficiency of financial distress with fewer characteristic variables.

2.4 Research Gap

The research incorporated all the traditional machine learning models to build predictive models. Different models have varying assumptions, strengths and limitations making their applications unreliable. Traditional machine learning models making predictions based on the underlying data structures and relationships. The underlying assumptions need to be met for high accuracies. However, this is not the case due to the underlying data structures and relationships. Traditional machine learning models often make different types of errors. Utilization of all the traditional machine learning models addresses the individual errors. Utilization of ensembles aid combining and maximizing on various strengths for a more robust and accurate model. Ensembles are helpful in attaining the bias-variance tradeoff, improving accuracy and robustness. Hence, the research aimed at exploring and evaluating the effectiveness of different machine learning techniques for improved prediction of financial distress. By applying all the traditional machine learning models on the given dataset, the study can compare and evaluate the most accurate, reliable model with robust prediction of financial distress. Further, a review of empirical studies shows no predictive

models on financial distress of Kenyan Saccos using different traditional machine learning models. Hence, the study sought to conduct a comparative analysis of all the traditional machine learning models to build the most accurate and robust predictive models on financial distress.

3. METHODOLOGY

3.1 Data Collection

3.1.1 Description of Dataset.

Data was collected from the SACCO Supervision annual Report (available at https://www.sasra.go.ke/) of 186 Deposit Taking Saccos in Kenya. The study used data from SACCOs that were licensed by SASRA and were in business operation between 2011 to 2021. The data contains financial (asset base, deposits, losses, and income) and non-financial (licensing status, registration status, affiliations, membership base, legal compliance status, and business scope – geographical and operations) features. Table 1 presents the description of dataset features.

Table 1: Description of Dataset Features

Feature	Type	Description
Total assets	Continuous	Total entities owned by Saccos
Total deposits	Continuous	Total non-withdrawable savings and used as collateral to secure and guarantee loans.
Gross loans	Continuous	Total amount owed by members
Total income	Continuous	Income from loans and investments
Membership	Continuous	Number of owners of Saccos
Licensing status	Categorical	Licensed to undertake deposit taking SACCO business or awaiting.
Registration status	Categorical	Year of registration
Affiliations	Categorical	Type of ownership based on common and shared interests: Government-based, Teacher-based, Private sector-based, community-based and Agriculture-based.
Compliance status	Categorical	Status of compliance and adherence to existing laws and regulations, categorized as yes or no.
Sacco size	Categorical	Saccos are classified into 3 categories based on total deposits and assets: • Large sized-Saccos – Above Ksh. 5 billion • Medium sized-Saccos – Between Ksh.1 billion and Ksh. 5 billion • Small sized-Saccos – Below Ksh. 1 billion
Financial distress	Categorical	The target variable represents SACCO classification, either distressed or healthy status.

3.1.2 Variable Selection

FDP studies lack a theoretical model that guides selection of features. According to Balasubramanian and Natarajan (2019) and Laitinen and Laitinen (2000), the selection of variables in the FDP modelling is usually built from a series of trial-and-error processes. Balasubramanian and Natarajan (2019) alluded that in FDP models, the dichotomous qualitative dependent variable is used. In the works of Ross, Westerfield and Jaffe (as cited in Balasubramanian and Natarajan, 2019) summarized prior studies and concluded that financial distress consists of the following four conditions: Business Failure, Legal Bankruptcy, Technical Bankruptcy and Accounting Bankruptcy.

SACCOs, with a significant decline in their total assets, deposits and gross loans, were considered financially distressed. The total assets, deposits and gross loans are some of the metrics for measuring the growth and performance of financial institutions engaged in mobilization of deposits and issuance of credit facilities (SASRA, 2022). Additionally, the assets and deposits are the principal revenue streams. This implies that a significant decline in those parameters affects their stability and sustainability. Poor compliance with legal licensing requirements was considered a sign of financial distress. The study elected a single dependent variable, namely, financial distress (sickness). Financial distress is a binary variable taking a dummy value of 1 if the SACCO was declared sick between 2014 and 2022, or the value of 0 otherwise. Upon subjecting the 186 SACCOs to the above financial distress criteria, 56 were classified as distressed, while 130 as being in healthy status.

3.2 Data Pre-Processing

3.2.1 Missing Data Analysis and Imputation

Many statistical methods cannot be applied directly to data with missing values, though some machine learning methods (for example random forests, and Naïve Bayes Classifier) can accommodate the presence of missing values. The dataset was checked for the presence of missing values which were found in both continuous (Total Assets, Total Deposits, Gloss Loss, number of Sacco branches and Total Income) and categorical (size of saccos) variables as shown in Figure 1. Omission of the observations with missing values was not an option for this study. The dataset was small. Hence, mean and mode imputation was preferred for the continuous and categorical variables respectively. Missing data analysis imputation was conducted on the dataset. In the case of continuous variables, mean imputation was preferred for this study using the mice package. In the case of categorical variables, the missMDA package which is based on multiple correspondence analysis for categorical variables was used for multiple imputations. Figure 1 and 2 shows the missing data visualization and complete values after the imputation of missing values.

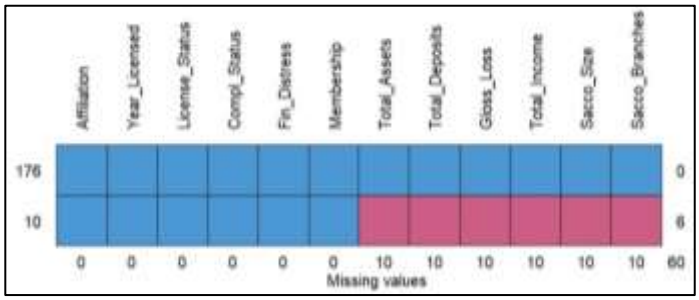

Figure 1: Presence pf missing data values

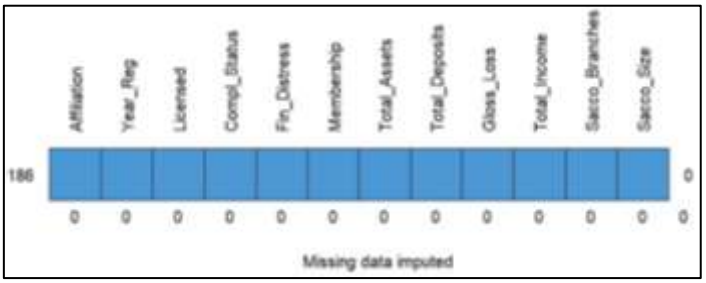

Figure 2: Missing data imputated

3.2.2 Outliers Detection and Treatment

The dataset was checked for outliers or observations that are substantially different from all other observations which can make a large difference in the final model results. Basically, an outlier is an observation with large residual. Outliers can skew the data, distort statistical analysis, and affect machine learning models. Some machine learning algorithms, for instance Logistic Regression, and KNN are sensitive to outliers. Outliers occur because of data entry errors, inaccuracies with data collection instruments and experimental errors.

Potential outliers were checked and visualized using boxplot as shown in Figure 3, where all the features had the presence of outliers. The explanation for this occurrence is that Saccos are classified in 3 tiers based on their total assets and deposits as described in Table 1. The variations between different classifications contribute to the unusual observation. Unusual observations caused by natural variations can still represent legitimate aspects of the data (McBean and Rovers, 1992). The high values which are considered as outliers are true representative and not because of data entry errors or sampling. Ignoring the outliers is likely to affect some machine learning algorithms, while removing them will make the dataset so small. Outliers were imputed using the interquartile range method. Figure 4 shows that extreme outliers have been controlled.

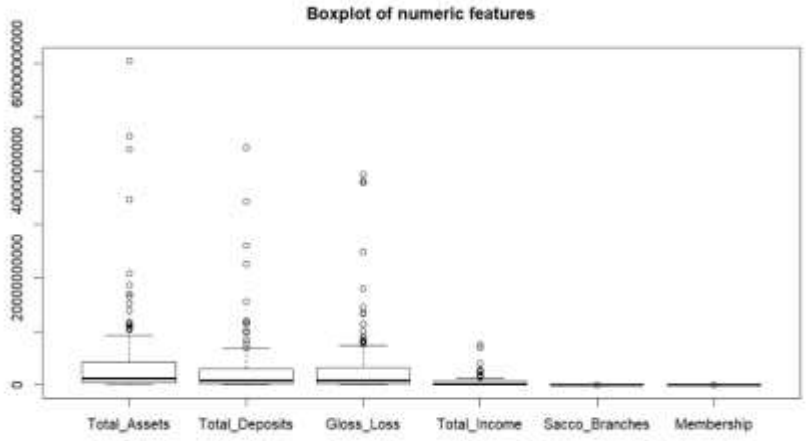

Figure 3: Boxplot of numeric features with outliers

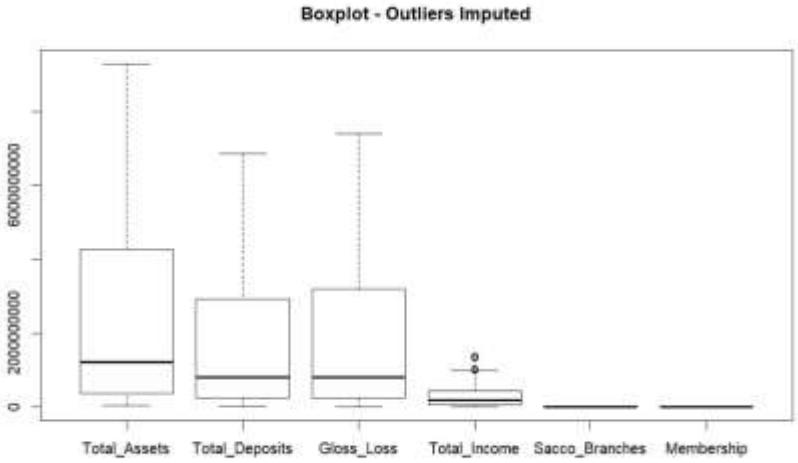

Figure 4: Boxplot - Outliers Imputed

3.2.3 Dummy Variable Creation

Dummy variables are binary variables used to represent categorical data in numerical form. Dummy variables are essential in statistical models and machine learning algorithms because most algorithms require numerical input. A salient problem in machine learning is transforming categorical variables into efficient numerical features (Kunanbayev, Temirbek and Zollanvari, 2021). By converting categories into binary values, dummy variables allow these models to process and analyze categorical features effectively. Dummy variables are "proxy" variables for categorical data in statistical models and machine learning algorithms. The dataset has described in Section 3.1.1, Table 1 shows that categorical variables were licensing status, registration status, affiliations, compliance status and business scope. The target variable (financial distress) and year of registration feature were of numeric type and they were converted into factors before variable encoding.

New features were created to reflect these binary values. The *fastDummies* function from the r package *fastDummies* was used to create dummy variables to incorporate nominal variables into models. These dummy variables were created with a one-hot encoding procedure in which each attribute had a value of either 0 or 1, representing the presence or absence of that attribute. One-hot encoding prevents algorithms from assuming a numerical order among categories. Kunanbayev, Temirbek and Zollanvari (2021) argued that one-hot encoding can become computationally inefficient as they create high-dimensional data matrices. This concern was addressed through feature scaling and centering of the numeric features to have them all to be in the same range as discussed in detail in Section 3.2.5. Further, to avoid the dummy variable trap, one dummy variable was excluded during the analysis. This was done to avoid the presence of highly correlated (Multicollinear) variables.

3.2.4 Correlation Analysis

Highly correlated features may affect performance of the ML algorithms. Further, they present multicollinearity problem which can undermines the statistical significance of an independent variable. The dataset was checked for highly correlated features.

Figure 5 shows the presence of highly correlated features namely membership, total assets, total deposits and licensing status. They were higlhy correlated with an absolute correlation of 0.75 or higher and hence removed from the dataset. Figure 6 vizualization of the dataset after highly correlated features were removed.

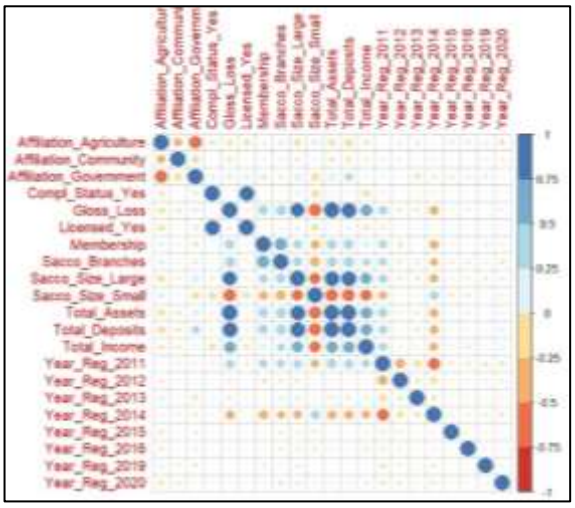

Figure 5: Presence of highly correlated features

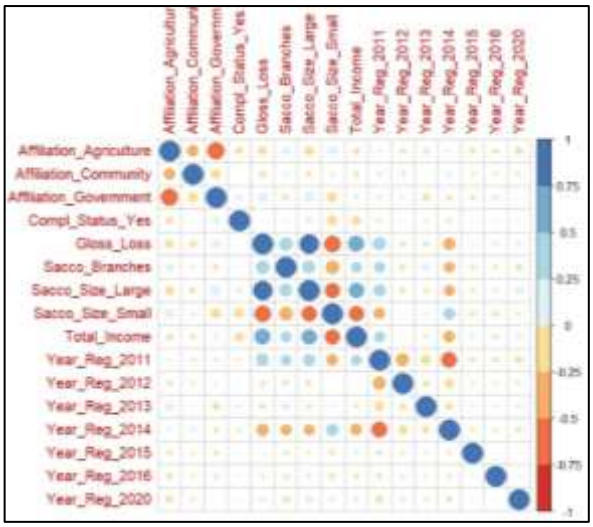

Figure 6: Highly correlated features removed

3.2.5 Feature Scaling and Selection

To ensure an accurate machine learning model was developed, it was important to select optimal features since having a lot of them can increase model complexity or lead to other problems such as multicollinearity and overfitting. Specifically, SVM, Logistic Regression, KNN, and neural networks are sensitive to features in different ranges. Several features on different scales makes machine learning problems hard to handle. Optimal features ensure a retention of a set of features for future predictions. The dataset has features with varying ranges hence need for feature scaling and centering of the numeric features to have them all to be in the same range, centered around 0 or in the range (0,1).

Recursive Feature Elimination (RFE) algorithm was used for selecting features that are most relevant in predicting. The packages caret and randomForest were used for implementing RFE for feature selection. Random forest selection function was used in defining the control, while the *rfe* function from the caret package was used to implement RFE. Random forest in the caret package was used considering its computing feature importance. The repeated 10-fold cross-validation with 5 repeats were used to improve the performance of feature selection with RFE. Dataset was randomly split into training (70%) and test (30%) datasets. In the rfe function, sizes = c (1:10) was used so that the function tries all possible solutions. Table 2 shows the output returned from the rfe function.

Table 2: Result of Recursive feature selection

Recursive feature selection **Outer resampling method:** Cross-Validated (10 fold, repeated 5 times) **Resampling performance over subset size:**					
Variables	**Accuracy**	**Kappa**	**AccuracySD**	**KappaSD**	**selected**
1	0.6804	0.1661	0.06456	0.1808	
2	0.7854	0.4672	0.08112	0.2164	
3	0.8177	0.5516	0.07508	0.2016	
4	**0.8309**	**0.5995**	**0.07898**	**0.2018**	*
5	0.8252	0.585	0.08423	0.2123	
6	0.8187	0.5708	0.09197	0.2221	
7	0.809	0.5491	0.08589	0.2115	
8	0.7935	0.5142	0.07965	0.2012	
9	0.7887	0.5137	0.08522	0.2004	
10	0.8053	0.5509	0.08615	0.2035	
The top 4 variables (out of 4): Compl_Status_Yes, Gloss_Loss, Total_Income, Sacco_Size_Small					

The output of Table 2 indicates that RFE recommends 4 features namely compliance status, Gloss loss, Total Income and Small sized Saccos. Both accuracy and Kappa reach the maximum level when 4 features are retained in the model. The asterisk mark (*) in Table 2 represents the optimal solution (4 features). Figure 7 shows that compliance status was the most important feature in predicting the target variable, followed by gloss loss, total income and small sized Sacco.

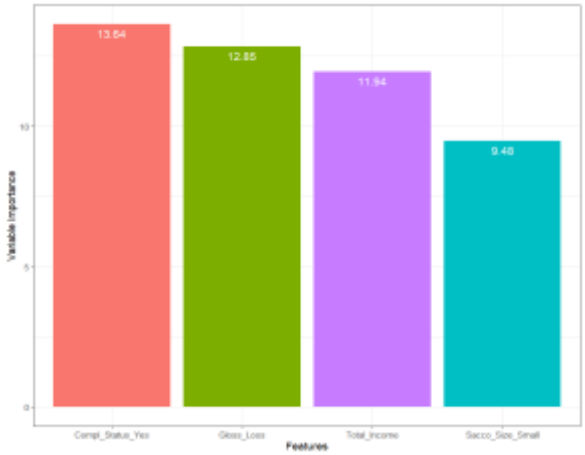

Figure 7: Variable Importance of Optimal Features

Table 3 and figure 8 shows the model performance using the test dataset as compared to training dataset. Post prediction results shows that both accuracy and Kappa values appear a narrow variation to those obtained from the training dataset.

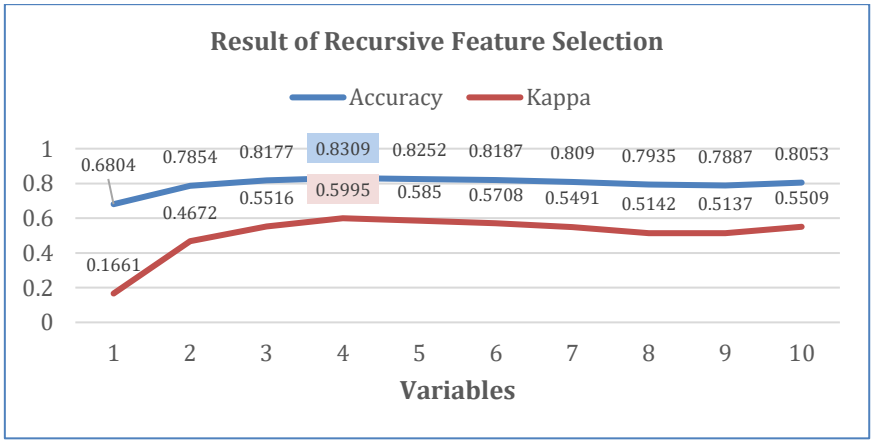

Figure 8: Result of Recursive Feature Selection

Table 3: Feature Selection Model Performance

	Training dataset	Test dataset
Accuracy	0.8309	0.8108
kappa	0.5995	0.5448

3.3 Diagnostics Tests

Data was subjected to diagnostics tests which included assessment of the binary nature of the dependent variable, independence of observations, multicollinearity, adequacy of sample size, class imbalance classification problem, normality test, and linearity test. This was necessary considering violations of these assumptions can lead to erroneous analytical results.

3.4 Machine Learning Algorithms

Table 4 provides a summary of the Algorithms used in this study for model building, validation and performance evaluation to address the classification problem of this study.

Table 4: Summary of Algorithms used for model building, validation and evaluation

Algorithm	Description
Logistic Regression	• Predicts a binary outcome using logistic function based on one or more predictor variables.
K-Nearest Neighbors	• Uses proximity to make classifications about the grouping of an individual data point with similar labels or values
Decision Tree	• Build the classification model in the form of a tree structure based on feature values
Support Vector Machine	• Finds the optimal hyperplane that separates classes in a high-dimensional space in a linear equation
Naive Bayes Classifier	• Applies Bayes' theorem with the assumption of independence between features
Random Forest	• Ensemble method that uses multiple decision trees to improve accuracy
Artificial Neural Networks	• Mimics the human brain to identify patterns and make predictions through layers.
Greedy Algorithm	• Type of optimization algorithm that makes locally optimal choices, though it doesn't always produce the optimal solution
Recursive Feature Elimination	• Effective at selecting features that are more or most relevant in predicting the target variable. • Irrelevant input features can mislead some ML algorithms resulting in worse predictive performance; it requires feature scaling.
Gradient Boosting	• Ensemble algorithm that combines several weak learners into strong learners

4 PERFORMANCE EVALUATION METRICS

4.1 ROC-AUC score as primary measure

The Area Under the Receiver Operating Characteristic curve (AUC-ROC curve) is a graphical representation of the performance of a binary classification model at various classification thresholds. AUC-ROC curve consists of two terms: ROC (Receiver Operating Characteristics) and AUC (Area Under the Curve). ROC curve is the graphical representation of the effectiveness of the binary classification model, which plots the True Positive Rate (TPR, also known as sensitivity or recall) against the False Positive Rate (FPR, or 1 - specificity) at various threshold settings. The AUC curve represents the area under the ROC curve. It measures the overall performance of the binary classification model. As both TPR and FPR range between 0 to 1, So, the area will always lie between 0 and 1, and A greater value of AUC denotes better model performance. Though TPR and FPR are associated with confusion matrix, Acharya (2024) has argued that they are not enough to assess model performance. ROC-AUC score was the primary measure adopted in this classification problem to assess the ability of a model to distinguish between two classes, typically the positive class (distressed Saccos) and the negative class (healthy Saccos). ROC-AUC score provides a broader view of the model's ability to differentiate between positive and negative classes since it considers the entire range of possible thresholds (Acharya, 2024). The ROC curve helps in visualization of the trade-off between sensitivity (True Positive Rate) and specificity (1 - False Positive Rate) for various threshold values.

4.2 The Confusion Matrix

The ROC-AUC score was the primary measure for this study. However, the metrics associated with Confusion Matrix (True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN)) were also referred to in the computation of model's recall(sensitivity or True positive rate), accuracy, error rate, precision, False positive rate, False negative rate, F1 score, and Specificity (True negative rate) and overall assess the classification model performance. These are the most popular evaluation metrics used for evaluating classifiers (Swalin, 2018; Srivastava, 2024) and they aid in the understanding if the model can predict using new dataset to tackle real-world tasks effectively (Prasanna, 2024).

5 STUDY FINDINGS

5.1 Model Building

The target variable was converted to a factor to facilitate the running of the algorithms, after data preprocessing stage. The observations were shuffled which was followed by ordering the dataset randomly. This was important to avoid bias in the train-test split and make sure there is a representative sample of the whole dataset. The dataset was split into the 70/30 train-test split considering the dataset is very small, so a 30% split gives a more reliable test set.

To ensure consistent and reproducible model evaluation the hyperparameter optimization (hyperparameter tuning or model tuning) was conducted. Model tuning is important to ensure an equal evaluation and comparison of performance models. The process involved tuning grids to be used in the training of Logistic Regression, SVM, KNN, Naïve Bayes, Decision Tree and RF models. The custom indices were generated using the createFolds () function for the purpose of K-fold cross-validation using the caret package. The target variable acted as the input to ensures that the folds are stratified or representative. The number of folds to be created was set to k = 10. Thereafter, a reusable trainControl object for comparing models was created by specifying the resampling method to be "cross-validation". The number of folds for cross-validation was set to 10 just like in createFolds, to ensure consistency. The summaryFunction was set to twoClassSummary argument to calculate the performance metrics (sensitivity, specificity, and ROC area under the curve) during resampling. The AUC metric was preferred to rank the performance of models. The twoClassSummary is specifically designed for binary classification problems. For twoClassSummary to work correctly, both classProbs and verboseIter were set to TRUE. Setting the classProbs to TRUE, ensures that the model will output class probabilities in addition to class predictions for model evaluation or for generating ROC curves. verboseIter was set to TRUE for monitoring of the training procedures. Finally, index was set to myFolds to ensure the model uses the custom folds created with createFolds.

After the model and tuning parameter values were defined, the type of resampling was specified. The trainControl function was used to specify 10-fold cross-validation (method = "cv" and number = 10). The verboseIter = TRUE argument means that the function will print out information about each resampling iteration. The savePredictions = TRUE argument is used to specify that the predictions for each resampling iteration should be saved. Finally, the index = myFolds argument is used to specify the resampling indices that should be used, which in this case are the 10-fold CV indices stored in the myFolds object. The result of the trainControl () function is stored in the myControl object, which can then be used as an argument to the train () function in caret package when fitting different models. The models were fitted by simply changing the method argument in the train function for each model. The method values 'ranger', 'svmPoly', 'nnet', 'nb', 'knn', 'glm' and 'rpart' for the random forest modelling, support vector matrix modelling, Artificial Neural Networks modelling, Naibes Bayer modelling, K nearest neighbor modelling, logistic regression modelling and decision tree modelling respectively. This specified the various methods in which the models were to be trained on.

5.2 Comparing Resampling Distributions of all models

5.2.1 Resampling Distributions of Models

Table 5 shows the summary statistics of resampling distributions of all models grouped by ROC, Sensitivity and Specificity. In summary, result from Table 5 show that SVM model had the highest ROC value of 0.87, while DC model had the lowest ROC value of 0.73. SVM model had the lowest value of Sensitivity at 0.28, while Logistic regression model had the highest value of Sensitivity at 0.74. KNN model had the highest value of 0.94, while ANN model had the lowest value of Sensitivity at 0.75.

Table 5: Resampling Distributions of Models using train data

Summary(resamps1) Call: Summary.resamples(object = resamps1) Models: rf, svm, ann, nbc, knn, lg, dc Number of resamples: 10							
ROC							
	Min.	**1st Q**	**Median**	**Mean**	**3rd Q**	**Max.**	**NA's**
rf	0.675	0.787500	0.83125	0.83500	0.896875	0.9500	0
svm	0.775	0.806250	0.86250	0.86500	0.918750	1.oooo	0
ann	0.575	0.800000	0.90000	0.86000	0.925000	1.oooo	0
nbc	0.475	0.712500	0.77500	0.76500	0.881250	0.9500	0
knn	0.725	0.740625	0.84375	0.83750	0.896875	0.9875	0
lg	0.650	0.756250	0.88750	0.84750	0.918750	1.oooo	0
dc	0.500	0.662500	0.72500	0.73375	0.809375	0.9000	0
Sensitivity							
rf	0.4	0.60	o .7	0.70	o.80	1.0	0
svm	0.0	0.00	0.2	0.28	0.40	0.8	0
ann	0.4	0.60	0.6	0.72	0.95	1.0	0
nbc	0.0	0.45	O.6	0.48	0.60	0.6	0
knn	0.4	0.40	0.6	0.60	o.80	0.8	0
lg	0.6	0.60	0.8	0.74	o.80	1.0	0
dc	0.0	0.40	0.6	0.52	0.75	1.0	0
Specificity							
rf	0.750	0.75000	0.8750	0.8750	1.ooooo	1.000	0
svm	0.625	1.ooooo	1.0000	0.9250	1.ooooo	1.000	0
ann	0.500	0.53125	0.8125	0.7500	0.87500	1.000	0
nbc	0.625	0.75000	0.8750	0.8375	0.87500	1.000	0
knn	0.750	0.90625	1.0000	0.9375	1.ooooo	1.000	0
lg	0.625	0.75000	0.7500	0.7750	0.84375	0.875	0
dc	0.500	0.78125	0.8750	0.8500	0.96875	1.000	0

5.2.2 Resampling Results with Final Tuning Parameters

Table 6 shows the resampling results of Models across tuning parameters. ROC was used to select the optimal model using the largest value of the tuning parameter. SVM model had the highest performance of ROC at 0.87, Sensitivity at 0.28 and Specificity at 0.93 using the largest values of the tuning parameters. The final tuning parameters values for the optimal SVM model were degree = 1, scale = 0.001 and C = 0.25. ANN model came second with a model performance of ROC at 0.86, Sensitivity at 0.72 and Specificity at 0.75. The final tuning parameters values for the optimal ANN model were size = 1 and decay = 0.1. Logistic Regression, Random Forest and KNN followed closely with ROC values of 0.85. 0.84 and 0.81 respectively. Decision Tree had the lowest ROC value of 0.73, with the final tuning parameter value for the optimal model being cp = 0.01.

Table 6: Resampling results with final tuning parameters using train data

Model	Final values used for the model	ROC	Sens	Spec
Decision Tree	cp = 0.01	0.73	0.52	0.85
K-Nearest Neighbors	k = 5	0.81	0.64	0.93
Logistic Regression	-	0.85	0.74	0.78
Artificial Neural Networks	size = 1, decay = 0.1	0.86	0.72	0.75
Naibes Bayer	fL = 0, usekernel = TRUE, adjust = 1	0.77	0.48	0.84
Random Forest	mtry = 2, splitrule = gini, mini.node.size =1.	0.84	0.7	0.88
Support Vector Matrix	degree = 1, scale = 0.001, C = 0.25.	0.87	0.28	0.93

5.3 Model Validation and Accuracy

The trained models were validated and predictions made using the test data. Table 7 presents a comparison of the performance of different models using multiple metrics which are the most popular evaluation metrics used for evaluating classifier (Swalin, 2018; Srivastava, 2024).

Table 7: Comparing Model Performance using test data

Evaluation metric/model	rf	logit	nbc	knn	dt	ann	svm
Precision value	0.88	0.68	0.8	0.88	0.98	0.66	0.71
Accuracy	0.79	0.64	0.66	0.75	0.82	0.62	0.66
Error rate	0.21	0.36	0.34	0.25	0.18	0.38	0.34
Recall value	0.38	0.62	0.67	0.45	0.14	0.64	0.6
False positive rate	0.38	0.62	0.67	0.45	0.14	0.64	0.6
False negative rate	0.16	0.2	0.25	0.2	0.18	0.21	0.19
F1 score	0.53	0.65	0.73	0.6	0.25	0.65	0.65
Specificity	0.84	0.8	0.75	0.8	0.82	0.79	0.81
ROC AUC value	0.713	0.754	0.65	0.784	0.69	0.757	0.752

The precision value which is the number of items correctly identified as positive out of total items identified as positive. Table 7 shows that DC model had the highest precision value at 98%, while both RF and KNN models achieved 88%, while NBC and SVM models achieved precision value of 68% and 71% respectively, while LOGIT and ANN models had the lowest precision value at 68% and 66% respectively. The Accuracy is an evaluation metric that shows the percentage of total items classified correctly. According to Table 7, DT model had the highest accuracy at 82%, closely followed by random forest model at 79%, then KNN model at 75%, with both SVM and NBC models having an accuracy of 66%. Logit and ANN models had an accuracy of 64% and 62% respectively. The error rate is computed by finding a difference between 1 with the accuracy value. Results in Table 7 shows that DT model had the lowest error rate by wrongly predicting 18% of the values, while ANN model had the highest error rate by wrongly predicting 38% of the values. Both NBC and SVM had wrongly predicted 34% of the values, while RF, KNN and LOGIT models had wrongly predicted 21%, 25% and 36% of the values respectively. The Recall or Sensitivity (True Positive Rate) provides the number of items correctly identified as positive out of total true positives. Table 7 shows that NBC model registered a recall value of 0.67, while ANN, logit, SVM, KNN, RF and DC models had recall values of 0.64, 0.62, 0.6, 0.45, 0.38 and 0.14 respectively.

The False Positive Rate or Type I Error shows the number of items wrongly identified as positive out of total true negatives. Table 7 shows that NBC model registered a recall value of 0.67, while ANN, logit, SVM, KNN, RF and DC models had recall values of 0.64, 0.62, 0.6, 0.45, 0.38 and 0.14 respectively. The False Negative Rate or Type II Error refers to the number of items wrongly identified as negative out of total true positives. Table 7 shows that NBC model registered False negative rate of 0.25, closely followed by ANN model at 0.21, while both logit and KNN models which had False negative rate of 0.2, SVM, DT and RF models had False negative rates of 0.19, 0.18 and 0.16 respectively. The F1 Score is a harmonic mean of precision and recall. Table 7 shows NBC model had F1 score of 0.73, closely followed by both Logit, ANN and SVM models which had F1 score of 0.65. KNN model registered a F1 score of 0.6, while RF and DT models had F1 Score of 0.53 and 0.25 respectively.

The Specificity (True Negative Rate) is the number of items correctly identified as negative out of total negatives. Table 7 shows that RF model had a Specificity of 0.84, closely followed by DC and SVM models at 0.82 and 0.81 respectively. Both Logit and KNN models had specificity of 0.8, while ANN and NBC models achieved the True negative rate of 0.79 and 0.75 respectively.

The ROC AUC score summarizes the classifier's performance across all possible classification thresholds by measuring the area under the ROC curve. ROC AUC score shows how well the classifier distinguishes positive and negative classes. It can take values from 0 to 1, a model with higher ROC AUC indicates better performance. Dash (2022) suggests that the accuracy of tests with AUCs between 0.50 and 0.70 is low; between 0.70 and 0.90, the accuracy is moderate; and it is high for AUCs over 0.90. Hence, a perfect model would have an AUC of 1, while a random model would have an AUC of 0.5. Table 7 shows that KNN model had ROC AUC score of 0.784, closely followed by ANN, SVM, Logit, RF, DT and NBC models with ROC AUC scores of 0.757, 0.752, 0.754, 0.713, 0.69 and 0.65 respectively. This implies that DT and NBC models can be considered to have low accuracy since their ROC AUC scores falls between 0.50 to 0.70, while KNN, ANN, SVM, Logit and RF models have moderate accuracy since their ROC AUC scores are between 0.70 to 0.90.

5.4 Ensemble Techniques

The study elected to use ensemble techniques by combining predictions of multiple base models to improve and boast the accuracy of the machine learning models. The study used caretEnsemble package for making ensembles. This package extends the functionality of the caret package by providing tools to create and analyse ensemble models. caretList was used to build lists of 7 caret models (glm', 'svmPoly', 'ranger',' nnet', 'nb', 'knn', 'rpart') on the same training data, with the same re-sampling parameters (Mayer, 2023). Thereafter, caretEnsemble and caretStack were used to create ensemble models from the caret model list using the training set resamples and validation conducted on the test set.

5.4.1 Performance of Ensemble Techniques

According to Mayer (2023), caretEnsemble is preferred in building models ensemble as it ensures the resampling indexes are identical across all models. It makes a robust linear combination of models using a GLM. caretEnsemble uses the linear greedy optimization on AUC. Table 8 shows that the ensemble's AUC on the test set resamples is 0.81, which is about 3% higher than the best individual model. Mayer (2023) observes that caretStack allows us to move beyond simple blends of models to using "meta-models" to ensemble collections of predictive models. Generalized Linear Model and e were adopted to train and for validation. Table 24 shows that the ensemble's AUC on the test set resamples is 0.81, which is about 3% higher than the best individual model. Mayer (2023) observes that the glm-weighted model weights and test-set AUCs are extremely similar to the caretEnsemble greedy optimization. This explains why the ensemble results in Table 8 are similar.

Table 8: Performance of Ensemble Techniques

ensemble	r package	AUC
Linear Greedy Optimization on AUC	caretEnsemble	0.81
Generalized Linear Model on AUC	caretStack	0.81
Stochastic Gradient Boosting on AUC	caretStack	0.81

5.4.2 Model Relative influence using Stochastic Gradient Boosting

Table 9 shows models relative influence using Stochastic Gradient Boosting, with 'ranger' model scoring model relative influence of about 42.5%, followed by 'knn' and 'nnet' models which scores 23.6% and 15.7% respectively. 'rpart', 'glm', 'svmPoly' and 'nb' models had a model relative influence of 9.3%, 5.2%, 2.7% and 1% respectively.

Table 9: Model Relative Influence using Stochastic Gradient Boosting

Model	var	Models Relative influence
ranger	ranger	42.49
knn	knn	23.56
nnet	nnet	15.72
rpart	rpart	9.34
glm	glm	5.17
svmpoly	svmpoly	2.66
nb	nb	1.04

5.5 Summary of Model Performance Comparison

Table 10 shows a comparison of the model performance, including the ensemble model using Sensitivity, Specificity and ROC as evaluation metrics. The ensemble model had higher performance as compared to the individual models. The individual model with the highest ROC value at 0.78 was KNN, however, Stochastic Gradient Boosting ensemble model (gbm) had higher performance of about 4%. In the case of specificity, gbm ensemble model had higher performance of 0.88, as compared to rf, a single model with the highest specificity at 0.84. For sensitivity values, logit regression had 0.67 compared to the ensemble model which had 0.64. Considering the three ensemble techniques resulted in a similar AUC score as presented in Table 10, only Stochastic Gradient Boosting was shown in Table 9 as a representative.

Table 10: Summary of Models' Performance

Evaluation metric/model	rf	logit	nbc	knn	dt	ann	svm	gbm
Sensitivity	0.38	0.62	0.67	0.45	0.14	0.64	0.60	0.64
Specificity	0.84	0.80	0.75	0.80	0.82	0.79	0.81	0.88
ROC	0.71	0.75	0.65	0.78	0.69	0.76	0.75	0.81

6 DISCUSSION OF RESULTS

6.5 Interpretation and Insights from model performance

The study applied different ML predictive models using publicly available data to predict financial distress of the Kenyan SACCOOs. The different ML techniques included Classification, Logistic Regression, SVM, RF, ANN and Decision Tree. The performance of predictive models was evaluated using ROC-AUC score and other metrics associated with Confusion Matrix. The ROC-AUC score was the primary measure of model performance. ROC-AUC score was preferred as it plots the true positive rate against the false positive rate. Generally, the predictive models achieved a ROC-AUC score in the range of 65% to 78%. KNN had the highest ROC-AUC score of 0.784, closely followed by ANN, SVM, Logistic regression, RF, Decision Trees and Naive Bayes classifier models with ROC-AUC scores of 0.757, 0.752, 0.754, 0.713, 0.69 and 0.65 respectively. The study implication is that Decision Trees and Naive Bayes classifier models can be considered to have low accuracy since their ROC-AUC scores falls between 0.50 to 0.70, while KNN, ANN, SVM, Logistic regression and RF models have moderate accuracy since their ROC-AUC scores are between 0.70 to 0.90.

Findings from previous study (Ledhem, 2022) found that KNN technique had the best-predicting accuracy. The better performance of ANN model could be attributed to the algorithm being one of the highly preferred data mining classifier models for financial distress prediction (Sreedharan, Khedr & El Bannany, 2020).

The ensemble model had higher performance as compared to the individual models. The individual model with the highest ROC value at 0.78 was KNN, however, Stochastic Gradient Boosting ensemble model had higher performance of about 4%. This finding is supported by El Bannany, Khedr, Sreedharan and Kanakkayil (2021) and Sreedharan, Khedr and El Bannany (2020) who found that use of ensemble techniques significantly increases prediction accuracy than individual ML algorithms. The models relative influence analyses using Stochastic Gradient Boosting revealed that 'ranger' model had the highest model relative influence, closely followed by 'knn' and 'nnet' models.

Feature selection through RFE algorithm revealed that compliance status was the most important feature in predicting the target variable, followed by gloss loss, total income and SACCO size. Previous studies (Zeng et al., 2020) have shown that use of feature selection algorithms improve classification accuracy because many redundant and irrelevant features degrade performance, and a good feature selection algorithm would reduce computational complexity. The study implies that a combination of financial and non-finncial variables can be used to predict the financial distress of SACCOs. Although, Tron et al. (2022) concluded that financial variables are the most important indicator in predicting the financial default of a company. Yousaf, Jebran and Wang (2021) found that board diversity attributes can significantly predict the financial distress of firms. Balasubramanian and Natarajan (2019) concluded that combining both financial and non-financial variables increases predictive ability. Generally, the feature selection and importance analyses suggest that financial related features were the most important predictors of financial distress situations of the Kenyan SACCOs. Nevertheless, the study show that financial distress can be predicted from a combination of financial and non-financial data.

6.6 The study implications

The study implications include:

1. SACCO management can use crucial data to build robust predictive models to assess their financial stability, growth and performance.
2. The regulators can build early warning systems using predictive models to have a general perspective of the SACCO sub sector to avert bankruptcy, risk assessment and governance related challenges so that an appropriate action can be taken well in time.
3. Potential investors can use predictive models in making investment decisions.

6.7 The study limitations

The study methodology relied on limited and small dataset which might lead to overfitting of the ML models. Further, the study relied on publicly available financial and non-financial data to develop the ML models, this is limiting considering the unavailability of core data that might provide a deeper insight on the financial healthy of the organisations.

7 CONCLUSION

The study concludes that the finacial distress of Kenyan SACCOs can be predicted by application of different Machine Learning models using publicly available data. Combination of financial and non-financial data can be used in the prediction of financial distress. Developing accurate and reliable predictive models is important as they provide early warnings to SACCOs that are distressed. This provides an indication to the direction the organisations are moving into and allows for appropriate action to be taken out of necessity. KNN had the highest ROC-AUC score of 78% while the ensemble models scored 81%.

8 FUTURE RESEARCH DIRECTIONS

The study suggests the following future research direction:

1. Predict financial distress of SACCOs using internal data for purposes of triangulation and development of early warning systems.
2. Explore application of ML models in predicting financial distress using financial and non-financial data individually or combined.
3. Comparative analysis of different ML models against ensemble techniques in predicting the financial distress in different contexts.

9 CONTRIBUTION TO THE STUDY OF KNOWLEDGE

The study avails knowledge in the predictive analytics on applicability of data mining techniques in predicting the financial distress. The study demonstrates the need to develop ML models using both financial and non-financial features. The methodology developed in this study can be used in similar organizations with publicly available data as an approach to get an insight into the financial health of such organizations. The study findings supports the arguments that ensemble techniques are superior to individual ML models.

REFERENCES

1) Acharya, N. 2024. *Why ROC AUC and Log-Loss Are Your Model's Secret Superpowers: Beyond F1 Score, Recall, and Precision.* Available at: https://medium.com/@nirajan.acharya777/why-roc-auc-and-log-loss-are-your-models-secret-superpowers-beyond-f1-score-recall-and-5373998deda2 (Accessed 20 August 2025).
2) Alaminos, D. and Fernández, M.Á., 2019. Why do football clubs fail financially? A financial distress prediction model for European professional football industry. *PloS one, 14*(12), p.e0225989.
3) Balasubramanian, S.A., GS, R. and Natarajan, T., 2019. Modeling corporate financial distress using financial and non-financial variables: The case of Indian listed companies. *International Journal of Law and Management, 61*(3/4), pp.457-484.
4) Beaver, W.H. 1966. Financial ratios as predictors of failure. *Journal of accounting research,* pp.71-111.
5) Beaver, W.H., Correia, M. and McNichols, M.F., 2011. Financial statement analysis and the prediction of financial distress. *Foundations and Trends® in Accounting, 5*(2), pp.99-173.
6) Bell, T.B., Ribar, G.S. and Verichio, J. 1990. Neural nets versus logistic regression: A comparison of each model's ability to predict commercial bank failures.
7) Crevier, D. 1993. *AI: The Tumultuous Search for Artificial Intelligence*. New York, NY.
8) Crespí-Cladera, R., Martín-Oliver, A. and Pascual-Fuster, B., 2021. Financial distress in the hospitality industry during the Covid-19 disaster. *Tourism Management, 85*, p.104301.
9) Dwivedi, R. 2021. *How to use the Random Forest classifier in Machine learning?* https://www.analyticssteps.com/blogs/how-use-random-forest-classifier-machine-learning.
10) El Bannany, M., Khedr, A.M., Sreedharan, M. and Kanakkayil, S., 2021. Financial Distress Prediction based on Multi-Layer Perceptron with Parameter Optimization. *IAENG International Journal of Computer Science, 48*(3).
11) Financial Sector Regulators, 2022. Kenya Financial Stability Report, 2020. Issue No. 13. 2021 FinAccess Household Survey | CBK. (n.d.). Www.centralbank.go.ke. https://www.centralbank.go.ke/2021/12/17/full-2021-finaccess-household-survey/
12) Gregova, E., Valaskova, K., Adamko, P., Tumpach, M. and Jaros, J., 2020. Predicting financial distress of slovak enterprises: Comparison of selected traditional and learning algorithms methods. *Sustainability, 12*(10), p.3954.
13) Halim, Z., Shuhidan, S.M. and Sanusi, Z.M., 2021. Corporation financial distress prediction with deep learning: analysis of public listed companies in Malaysia. *Business process management journal*.
14) Huang, Y., Wang, Z. and Jiang, C., 2024. Diagnosis with incomplete multi-view data: A variational deep financial distress prediction method. *Technological Forecasting and Social Change*, *201*, p.123269.
15) Huang, B., Yao, X., Luo, Y. and Li, J., 2022. Improving financial distress prediction using textual sentiment of annual reports. *Annals of Operations Research*, pp.1-28.
16) Kabaiya, F. M. 2012. The relationship between corporate governance practices and financial performance, C l OF, (October).

17) Kibue, V.M.W. and Mang'ana, R., 2022. Influence of corporate governance practices on performance of savings and credit cooperative societies in the central region of Kenya. *International Academic Journal of Human Resource and Business Administration, 4*(1), pp.268-273.
18) Kunanbayev, K., Temirbek, I. and Zollanvari, A., 2021, July. Complex encoding. In *2021 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-6). IEEE.
19) Ledhem, M.A., 2022. Data mining techniques for predicting the financial performance of Islamic banking in Indonesia. *Journal of Modelling in Management, 17*(3), pp.896-915.
20) Li, Y. and Wang, Y. 2017. Machine learning methods of bankruptcy prediction using accounting ratios. *Open Journal of Business and Management, 6*(1), pp.1-20.
21) Laitinen, E.K. and Laitinen, T., 2000. Bankruptcy prediction: Application of Taylor's expansion in logistic regression. *International review of financial analysis, 9*(4), pp.327-349.
22) Loan, V.T. and Tuan, H.M., 2019. Using market variables in financial distress prediction for Vietnamese listed companies. *Religación: Revista de Ciencias Sociales y Humanidades, 4*(14), pp.341-352.
23) Mayer, Z. 2023. *A Brief Introduction to caretEnsemble.* Retrieved March, 30, 2024, from https://cran.r-project.org/web/packages/caretEnsemble/vignettes/caretEnsemble-intro.html
24) McBean, E.A, and Rovers, F.A. 1992. Estimation of the Probability of Exceedance of Contaminant Concentrations. *Ground Water Monitoring Review Winter*, pp. 115–119.
25) Muñoz-Izquierdo, N., Laitinen, E.K., Camacho-Miñano, M.D.M. and Pascual-Ezama, D., 2020. Does audit report information improve financial distress prediction over Altman's traditional Z-Score model?. *Journal of international financial management & accounting, 31*(1), pp.65-97.
26) Olando, C.O., Mbewa, M.O. and Jagongo, A., 2012. Financial Practice as a Determinant of Growth of Savings and Credit Co-Operative Societies' Wealth (A Pointer to Overcoming Poverty Challenges in Kenya and the Region). *International journal of Business and social science*, *3*(24).
27) Ooko, C., O., Jagongo, A., & Mbewa, M. O. 2013. The Contribution of Sacco Financial Stewardship to Growth of Saccos in Kenya PhD in Finance Kenyatta University Kenya. *International Journal of Humanities and Social Science, 3*(17), 112–137.
28) Paule-Vianez, J., Gutiérrez-Fernández, M. and Coca-Pérez, J.L., 2020. Prediction of financial distress in the Spanish banking system: An application using artificial neural networks. *Applied Economic Analysis, 28*(82), pp.69-87.
29) Situm, M., 2023. Financial distress in the Austrian tourism industry: hotels and restaurants analysis. *European Journal of Tourism Research, 34*, pp.3411-3411.
30) Swalin, A. 2018. *Choosing the Right Metric for Evaluating Machine Learning Models*. Retrieved: https://medium.com/usf-msds/choosing-the-right-metric-for-evaluating-machine-learning-models-part-2-86d5649a5428
31) Tagliaferri, L. 2017. *An Introduction to Machine Learning | DigitalOcean. [online]* Available at: https://www.digitalocean.com/community/tutorials/an-introduction-to-machine-learning#machine-learning-methods.
32) Tsai, C.F., Sue, K.L., Hu, Y.H. and Chiu, A., 2021. Combining feature selection, instance selection, and ensemble classification techniques for improved financial distress prediction. *Journal of Business Research*, *130*, pp.200-209.
33) Tron, A., Dallocchio, M., Ferri, S. and Colantoni, F., 2022. Corporate governance and financial distress: lessons learned from an unconventional approach. *Journal of Management and Governance*, pp.1-32.
34) Sankhwar, S., Gupta, D., Ramya, K.C., Sheeba Rani, S., Shankar, K. and Lakshmanaprabu, S.K. 2020. Improved grey wolf optimization-based feature subset selection with fuzzy neural classifier for financial crisis prediction. *Soft Computing, 24,* pp.101-110.
35) SASRA, 2012-2022. SACCO Supervision Annual Report, 2011-2021 (Deposit-taking SACCOs).
36) Sreedharan, M., Khedr, A.M. and El Bannany, M. 2020. A comparative analysis of machine learning classifiers and ensemble techniques in financial distress prediction. *In 2020 17th International Multi-Conference on Systems, Signals & Devices (SSD)* (pp. 653-657). IEEE
37) Srivastava, T. 2024. *12 Important Model Evaluation Metrics for Machine Learning Everyone Should Know (Updated 2023).* Retrieved: https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/
38) Uthayakumar, J., Metawa, N., Shankar, K. and Lakshmanaprabu, S.K. 2020. Intelligent hybrid model for financial crisis prediction using machine learning techniques. *Information Systems and e-Business Management, 18*, pp.617-645
39) Villani, C. 2018. *What Is Artificial Intelligence? Villani mission on artificial intelligence*.
40) Waseem, M. 2023. *How To Implement Classification in Machine Learning?* Retrieved: https://www.edureka.co/blog/classification-in-machine-learning/#algo

41) Yousaf, U.B., Jebran, K. and Wang, M., 2021. Can board diversity predict the risk of financial distress? *Corporate Governance: The International Journal of Business in Society, 21*(4), pp.663-684.
42) Zhang, Z., Chi, G., Zhou, Y., & Liu, Y. (2025). Predicting financial distress: the power of sentiment words in business plans. *The European Journal of Finance*, 1–20. https://doi.org/10.1080/1351847X.2025.2476538
43) Zeng, S., Li, Y., Yang, W. and Li, Y., 2020. A financial distress prediction model based on sparse algorithm and support vector machine. *Mathematical Problems in Engineering,* pp.1-11.
44) Zhu, J., Zhu, H. and Lin, N., 2023. A Dynamic Prediction Model of Financial Distress in the Financial Sharing Environment. *Discrete Dynamics in Nature and Society,* 2023.
45) Zhu, F.J., Zhou, L.J., Zhou, M. and Pei, F., 2021. Financial distress prediction: A novel data segmentation research on Chinese listed companies. *Technological and Economic Development of Economy*, 27(6), pp.1413-1446.

There is an Open Access article, distributed under the term of the Creative Commons Attribution – Non Commercial 4.0 International (CC BY-NC 4.0) (https://creativecommons.org/licenses/by-nc/4.0/), which permits remixing, adapting and building upon the work for non-commercial use, provided the original work is properly cited.