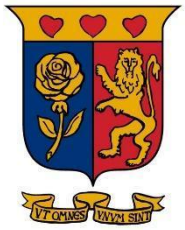


Strathmore
UNIVERSITY

Digital Loan Requests Assessment and Approval System for Registered Bodies and Informal Financial Institutions in Kenya.

Student Name: Mwangi Naomi Wambui

Admission No: 101426

An Information Systems Project Final Documentation Submitted to the Faculty of Information Technology in partial fulfilment of the requirements for the award of a Degree in Bachelor of Science in Informatics and Computer Science.

Declaration

I declare that this project has not been submitted to any other University for the award of a Degree in Bachelor of Science in Informatics and Computer Science.

Student Name: Mwangi Naomi Wambui

Student Admission: 101426

Signature: …… Date: …….

I certify that this work is being submitted for examination with my approval

Supervisor's Name: Julliet Kirui

Signature: …… Date: …….

Abstract

For various reasons, different credit providers deny credit to customers. The prime objective of financial institutions is to invest their assets in safe hands. Today, many banks/financial institutions approve loans after a shallow process of verification and validation but still, it is not chosen whether the applicant is deserving of the loan.

Credit risk prediction models seek to predict factors based on a qualitative aspect whether an individual will default on loan or not. This project aims to develop a system that can be used in SACCOs and small financial institutions to assess a loan candidate based on multiple factors that can be sourced from existing documents. This can be used to supplement the existing systems in place or to even have a stronger mechanism in place.

The proposed system will be an application that can predict whether a loan applicant is qualified for a loan or not. This application will be a simulation of a SACCO's portal that will enable the users to interact with the system to get the predicted results based on an average score provided by the machine learning model. The system will use multiple machine learning models to come up with a more accurate score to determine a borrower's fate. The targeted models are mostly non-parametric models such as K-Means, random forest, and gradient boosting algorithms to help the institutions have better models that can assess creditors.

Table of Contents

Table of Figures

Chapter 1: Introduction

1.1 Background of the Study

SACCOs in Kenya are classified as registered bodies and some smaller groups are registered as informal bodies according to the Financial Household survey 2019, conducted by the Central Bank of Kenya and Kenya National Bureau of Statistics. Informal finance is defined to refer to all transactions, loans and deposits that occur outside a central monetary authority's control and regulation, whereas the semi-informal sector has both informal and formal characteristics (Mwangi & Nairobi, 2011).

As of 2019, the use of microfinance remains significantly high at 30.1%, which means that informal groups in Kenyan households are still a key source of funding. This argument has been attributed mainly to commercial banks and other formal institutions struggling meet smallholders' credit needs primarily because of their terms and conditions for lending. This has made a major contribution to the notion of discouraged borrowers. According to (Kon & Storey, 2003), better lenders do not apply for bank loans because they fear they are going to be turned down. The discouragement stems from the bank's screening mistakes, the amount of transaction costs, and the degree to which interest rates vary from one lender to another. The application process is also said to discourage borrowers as a lot is required from them.

SACCOs emerge as a popular alternative to banks for access loans that are cheaper and faster (Tubei, 2019). Most of the time, Savings and Credit Co-operations usually loan their members three times the shares they had. How this works is by pooling savings from SACCO members and in turn, they are offered credit facilities at low fees and low interest to further their economic interests and general welfare. Even though the sector is much smaller in absolute terms than the banks, SACCOs form a vital part of the Kenyan financial system, their significance is far greater. They provide services to roughly three million Kenyans and provide services and facilities that cannot be found in other formal intuitions. This is evident as we see in rural areas, most farmers depend on their SACCOs for their credit and payment services. Being user-owned institutions or mutual they provide with an important alternative institutional form, SACCOs are playing an

increasingly important role in the cycle of financial intermediation in a highly competitive financial market in Kenya (Nkuru, 2015).

According to (Essendi, 2013), the provision of credit facilities is the core feature of each SACCO with a credit management function that encourages the efficient management and management of the SACCO loan portfolio with a view to ensuring a fair distribution of funds and promoting liquidity planning. Traditional statistical approaches have been employed mainly through logistics to build models for credit assessment. Models of credit risk are often used to mark a claimant as good or poor. Over the years, these methods have become popular especially with the development of machine learning and modern statistical models. Machine learning models are being used in many fields to help predict the possibility of a variety of risks and avoid unforeseen problems. Companies are using these models to plan for various outcomes based on the trend that is picked from data that is being learned over time. One of the famous fields used for machine learning is supervised algorithms using ensemble classifiers, where learning algorithms that construct a collection of classifiers classify new data by voting their predictions (Dietterich, 2000).

1.2 Problem Statement

As SACCOs continue to grow, their credit risks continue to grow significantly. Loans from the first two categories of financial groups in Kenya (Prudential and Non-prudential loans) are regulated by the Central Bank of Kenya while the registered bodies and informal sector are not monitored and therefore choose to outsource the assessment process or follow with traditional models. Studies carried out show a strong negative relationship between loan defaults and the profitability of SACCOs. Regression models are said to be a good fit for the independent variables to statistically predict the dependent variable though studies recommend that SACCOs should continuously review credit policies and establish characteristics of loan applicants (Keitany, 2011). The models should not only factor in quantitative aspects but also largely factor in qualitative aspects of the loan applicant. There has been a concern that insufficient attention is given to the ability of borrowers to repay their loans and this results in adverse impacts on the institutions. It is said that the financial institutional crisis is primarily due to external developments, and domestic factors are the main cause of banking distress. Non-performing loans are triggered by wrong-economic decisions and pure bad-luck (Waweru & Kalani, n.d.). An intelligent system is needed

in our local SACCOs to help rate a candidate borrower as either good or bad. This system should help ease the loan application process and provide a fast yet cheap solution.

The stages involved in applying for a loan are broken into application for loan, credit evaluation, loan review and repayment performance (Altman, 1980). Credit evaluation is normally obtained form credit scoring bodies in the country, and this is a major determinant in the repayment performance. If the loan review is not prioritized as well, then the credit score can as well become inefficient in determining the success of loan repayment.

1.3 Objectives

1.3.1 General Objective

The aim of the proposed project is to develop a machine learning model that will help micro-finance institutions such as SACCOs assess the chance of a borrower defaulting a loan or paying it back in good time.

1.3.2 Specific Objective

i. To investigate the existing methods and techniques used to assess a borrower in formal and informal financial institutions.
ii. To analyze the existing systems in order to determine existing loopholes that lead to loan defaulting in the various institutions.
iii. To find out techniques being used in successful institutions that have managed to apply efficient machine learning techniques that enable them to determine the chance of loan repayment.
iv. To develop and design a model that simulates an existing system for a micro-finance institute that will help them predict the chance of loan repayment through assessing various features tied to a borrower.

1.3.3 Research Questions

i. How do financial institutions assess a borrower before granting them a loan?
ii. What are the challenges the existing systems use to assess borrowers face?
iii. What are the techniques and models that are being used in successful financial institutions?
iv. What process is involved and required to design and develop the simulation of the system.

1.4 Justification

The Central Bank of Kenya regulates only prudential companies such as banks, deposit-taking cooperatives, and insurance. The reserved and informal financial institutions require a faster and easier system that can help them approve loan requests. The outcome of the proposed model will have a great significance to many institutions since the societies targeted have a big role to play as the country struggles to turnaround the economy. Of interest, the system will benefit the management of SACCOs who will use it to assess various borrowers as this will save the institution from making losses or struggling financially. Proper disbursement of loans will help in the growth of SACCOs within the various sectors in Kenya and this in turn will help them to concentrate on other financial aspects to gain competitive advantages. Further analysis can be done using the model to understand cooperative dynamics which will steer the social mission in the future.

The performance of the statistical models being used can be improved using powerful machine learning techniques which may lead to more powerful predictions. This will help strengthen the systems currently being used and provide a less tedious loan approval process.

1.5 Scope and Limitations.

This study focuses mainly on the process of requesting loans in SACCOs. It will cover mainly Mzima SACCO, which as a test case is at Strathmore University. They will use this system to evaluate different borrowers and determine their chance of defaulting on a loan. This project's success largely relies on data available to build a reliable and effective model. One of the limitations of this project is the amount of time it takes to create a detailed and accurate model for the SACCO. It however can be substituted with a model that will simulate the current system to demonstrate the functioning of the process to provide the expected results.

Chapter 2: Literature Review

2.1 Introduction

This chapter discusses financial institutions' challenges during the appraisal process of loan applications. Additionally, the chapter will also explain the existing processes used to accept loans and weaknesses in current systems for both applicants and the institutions.

2.2 Challenges Encountered in Loan Application Processes.

People who borrow loans from SACCOs and then contribute their money with the intention of borrowing attribute the whole cycle of applying for loans to be tedious whereby you are expected to search for guarantors and their documents to decide your approval. In some large corporations, collateral is required to secure the loan and in the event where it is a smallholder or someone who is starting in the financial world as independent, they may face challenges getting collateral to secure the loan. The failure of customers to reveal sensitive details during the loan application process is regarded as the key customer-specific cause for bad loans, on the other hand this can be regulated by the policies and measures developed by the institutions. The challenge that comes with this is the lack of cheaper alternatives for many small institutions and therefore to get more clients, they may tend to reduce their interest rates and their regulations thus opening them up to greater risks.

For the institutions using traditional models such us Linear Discriminatory Analysis, they have a chance of improving their accuracy levels especially with emerging technology that will enable faster and efficient processing of data. Better performing and enhanced models are still yet to be implemented in existing SACCOs instead of manual and physical implementations. Automating these processes will provide a cheaper and better alternative to some of the existing systems being used.

2.3 Existing Loan Application and Approval System

2.3.1 Loan Application Assessment in Banks

The loan portfolio is considered to naturally be the one of the biggest assets and the largest source of revenue for banks (Kwambai & Wandera, 2013). Because of information asymmetry in banks' previous credit networks, the Central Bank of Kenya and the Kenya Bankers Association have joined forces to introduce credit information exchange in Kenya to cap the loophole exploited by serial defaulters. Sharing of credit information is a mechanism in which banks and other lenders send borrowers information to a credit database office so that it can be exchanged with other providers. This is what enables banks to know how an individual has been repaying loans and based on this score, the limit to which a certain borrower can borrow is determined. The credit scoring process involves statistical models that do regression based on qualitative data and based on this, the result is used to determine how likely it is for a certain candidate to repay their loans. Parametric models such as logit and Linear Discriminant Analysis are the most used models to achieve this. In banks variables such as income, amount of existing personal debt, number of accounts from other credit sources and credit history are used to determine an individual's credit score (Kwambai & Wandera, 2013).

2.3.2 Loan Application Process in SACCOs

For one to get a loan in most of the SACCOs, someone needs to have been a member for a specified period. The borrower also needs to ensure the loan is guaranteed by loanee's deposits and/or collateral or guarantor deposits. A maximum repayment period is given depending on the amount and certain policies like the maximum amount for the basic salary are given. Afterward, one is required to fill a form which is reviewed by a group of loan creditors who will then give a go-ahead or not. The amount being borrowed in a SACCO is highly connected to the amount the member has ever deposited in the SACCO's accounts. The main reason for this is for a member to enhance his/her creditworthiness (Mwangi & Nairobi, 2011).

2.3.3 Group Methodology in local SACCOs

The group-based model is used by some local SACCOs as a conduit for supplying financial services. Many institutions use the group approach with the assumption that they can minimize the cost of obtaining information about the borrower, increase the recovery of the loan by mutual

default, increase access to information about the nature of the borrower and assess the creditworthiness of potential borrower (Nkuru, 2015). Many SACCO distributions networks, whether community-based or business-oriented, require clients enter groups in order to obtain loans. The group is designed to guarantee the individual borrower and to ensure that the individual makes payments on time because the group is a stakeholder and does not obtain additional loans if the individual defaults.

2.3.4 Ensemble Classifiers

These are learning algorithms that come with a collection of classifiers that identify the new data points by taking a weighted vote of their predictions. (Dietterich, 2000). Ensemble methods are highly effective in reducing bias and variance in a machine learning model to help reach better predictions. A range of parameters are used to include the forecast of the expected class. Ensemble models combine the outputs of different models to predict the borrowing of the applicant. This helps to make better predictions and helps banks predict the future of the loan and its status depending on whether they can take action in the early days of the loan. (Al, 2019). How this comes about, is if given a set *P* of training algorithms, each algorithm will give a classifier as its output. The classifier will then become a proposed explanation about the true function *f*. Given a set of new input values, the ensemble classifier will predict the various outcomes which will form a set that will tie each classifier to the output of the individual classifiers. An ensemble classifier can use algorithms such as Logistic Regression, artificial neural networks, and random forest to include both parametric and non-parametric models. Ensemble classifiers are preferred over single classifier approaches due to their variety of classification models and their accuracy levels.

There are various ensemble techniques, this can be taking the mode of the results from the different classifiers to determine the overall result. Another method can be taking the mean of the results from the different classifiers. Lastly a weighted mean can be taken from the results where each classifier is assigned a weight that holds the relevance of each prediction model. There are other advanced techniques which are bagging and boosting which focus on manipulating the training data to give a prediction.

The diagram below shows the process of a single classifier goes through to give predictions in an ensemble technique.

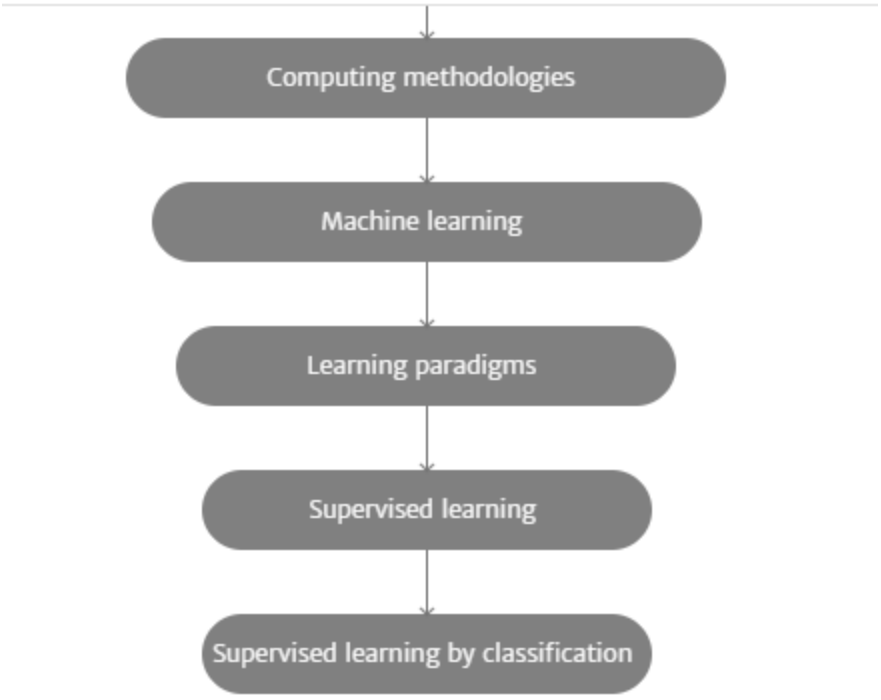


Figure 1: Machine learning classifier model (Twala, 2010)

2.4 Gaps in the existing systems

Banks in Kenya have been lent money to chronic defaulters, which is the result of banks having different credit details about borrowers, and this, in effect, results in borrowers using information asymmetry to borrow multiple loans from various Kenyan banks and, in the long run, results in non-performing aspects (Kwambai & Wandera, 2013). Credit scoring came to help solve this problem, but the problem comes where the traditional models being used to assess a candidate fail to assess the qualitative aspects of an individual. More information is needed especially one that can attribute to an individual's character to build more accurate models. Other than this, newer models have come into place which are more efficient and do not rely on parameters to classify the data. An experimental study was conducted to compare the accuracy between non-parametric model's vs parametric models to estimate saturated and nearly saturated hydraulic conductivities.

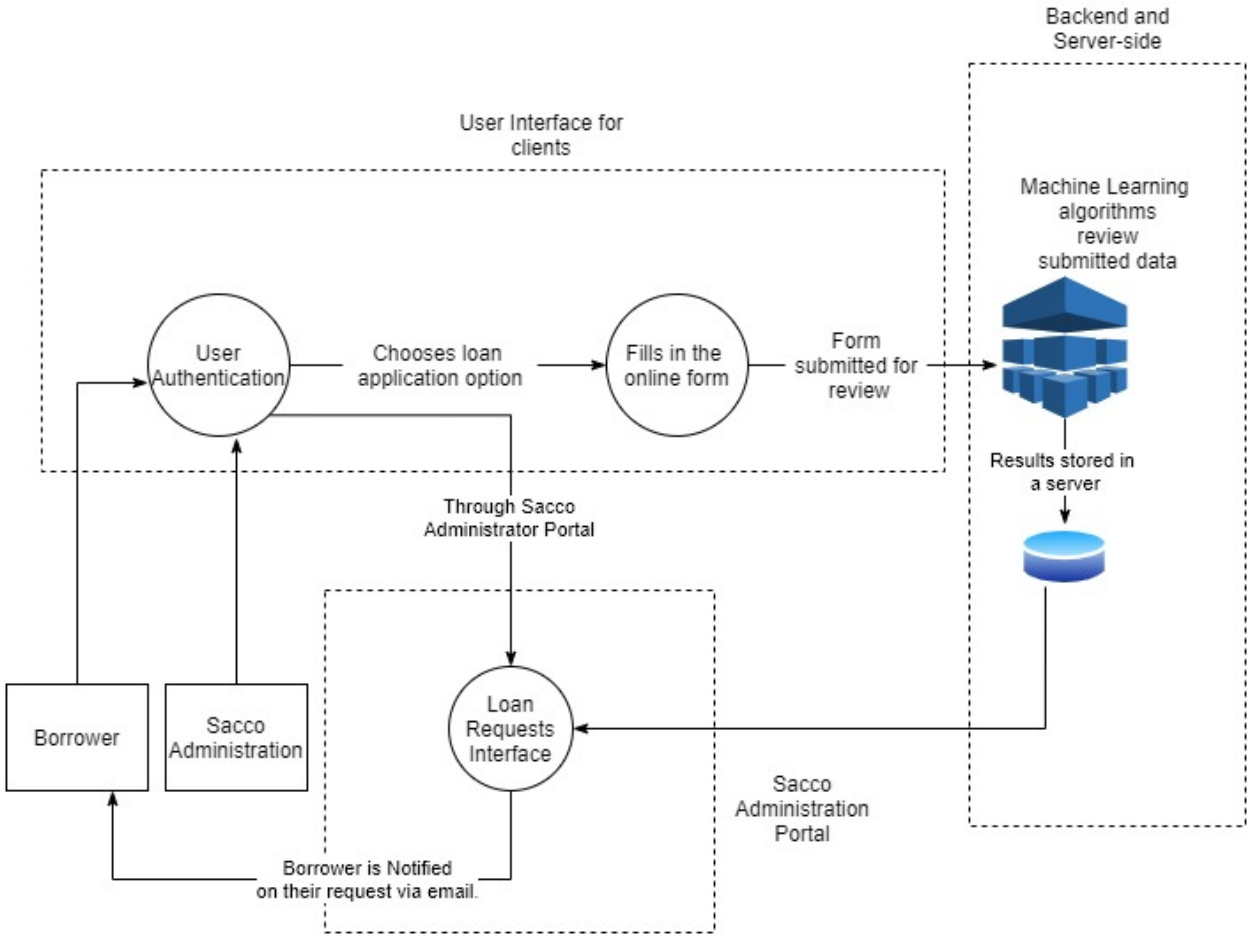
Non-parametric supervised models tested and trained with similar data sets significantly improved accuracy (Kotlar et al., 2019).

The loan approval process, where the SACCO uses monthly meetings to approve member loans, is considered time-consuming and keeps customers waiting while commercial banks have reduced processing times. The other alternative was to establish collateral, namely community guarantees. If the group is unstable and the group leverage (collateral) decreases, this raises the credit risk of the SACCO and most of the time leads to default on the loan (Nkuru, 2015). Groups usually become vulnerable when mutual responsibility is compromised, and a default occurs. Good customers typically object to the application of this collective liability because they feel that they are overburdened by other people's debts over and above their own, in particular in cases where the person has incurred additional collateral other than group liability. Others may even opt out of the party at the end of their loans.

SACCOs would need to tailor the size of the loan to the needs of the borrower, to incorporate certain forms of loans, for example individual loans, in addition to community guaranteed loans, and to expand the variety of goods and services beyond credit. Manual loan approval by a committee can be biased. The loans can be granted based on certain aspects such as favouritism and this in turn will lead to mismanagement of loans. The process also takes so much time as the group must debate based on the aspects and judgment towards a certain person. For new members, due to little or no experience, they are less likely to be granted the loans.

2.5 Conceptual Framework for the Proposed Solution

The proposed system will include user authentication, where users will log into the system and access the loan application portal. This portal will have an online form that will enable the users to fill in their details, whereupon submission of the form, the various machine learning algorithms will analyse the input, and based on this a prediction will be given on whether the creditor is approved or not. This information will then be stored in an MYSQL database that will enable retrieval from the SACCO administrator’s portal and based on this, the administrator will be able to initiate a notification that will go to the client to inform them on their results. This will include various reports that will help the borrower be accountable for why their loan was rejected and probably some advice on whether they should start at a lesser limit.


Backend and
Server-side
User Interface for
clients
Machine Learning
algorithms
review
submitted data
User
Authentication
Chooses loan
application option
Fills in the
online form
Form
submitted for
review
Results stored in
a server
Through Sacco
Administrator Portal
Loan
Requests
Interface
Borrower
Sacco
Administration
Sacco
Administration
Portal
Borrower is Notified
on their request via email.

REFERENCES

Al, N. M. et. (2019). LOAN PREDICTION ANALYSIS USING DECISION TREE. *Journal of the Gujarat Research Society*, *21*(14s), 214–221.

Altman, E. I. (1980). Commercial Bank Lending: Process, Credit Scoring, and Costs of Errors in Lending. *The Journal of Financial and Quantitative Analysis*, *15*(4), 813–832. JSTOR. https://doi.org/10.2307/2330559

Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. In G. Goos, J. Hartmanis, & J. van Leeuwen (Eds.), *Multiple Classifier Systems* (Vol. 1857, pp. 1–15). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45014-9_1

Essendi, L. K. (2013). *The Effect of Credit Risk Management on Loans Portfolio Among Saccos in Kenya*. 51.

Keitany, N. J. (2011). *The Relationship Between Loan Default And The Financial Performance Of Saccos In Kenya*. 63.

Kon, Y., & Storey, D. J. (2003). A Theory of Discouraged Borrowers. *Small Business Economics*, *21*(1), 37–49. https://doi.org/10.1023/A:1024447603600

Kotlar, A. M., Iversen, B. V., & Lier, Q. de J. van. (2019). Evaluation of Parametric and Nonparametric Machine-Learning Techniques for Prediction of Saturated and Near-Saturated Hydraulic Conductivity. *Vadose Zone Journal*, *18*(1), 180141. https://doi.org/10.2136/vzj2018.07.0141

Kwambai, K. D., & Wandera, M. (2013). *EFFECTS OF CREDIT INFORMATION SHARING ON NONPERFORMING LOANS: THE CASE OF KENYA COMMERCIAL BANK KENYA*. 26.

Mwangi, J. M., & Nairobi, U. O. (2011). *The role of savings and credit cooperative societies (saccos) in financial intermediation in Nairobi county*. 108.

Nkuru, F. N. (2015). *FACTORS AFFECTING GROWTH OF SACCOs WITHIN THE AGRICULTURAL SECTOR IN KENYA: A CASE OF MERU FARMERS SACCOs*. 12.

Tubei, G. (2019, February 15). *Here are 5 reasons why Kenyans borrow SACCO loans and how they spend it [ARTICLE]—Pulse Live Kenya*. Business Insider. https://www.pulselive.co.ke/bi/strategy/here-are-5-reasons-why-kenyans-borrow-sacco-loans-and-how-they-spend-it/fshqs51

Twala, B. (2010). Multiple classifier application to credit risk assessment. *Expert Systems with Applications*, *37*(4), 3326–3336. https://doi.org/10.1016/j.eswa.2009.10.018

Waweru, N. M., & Kalani, V. M. (n.d.). *Commercial Banking Crises in Kenya: Causes and Remedies*. *3*(3), 21.