CREDIT RISK ASSESSMENT FOR CUSTOMERS OF KENYAN COMMERCIAL BANKS: A CASE OF CO-OPERATIVE BANK

JAMES BOGONKO AREBA

Reg: 15/03067

A Research Project Submitted to the Faculty of Computing and Information Management in Partial Fulfillment for the Award of the Degree of Master of Science in Information System Management, KCA University

SEPTEMBER, 2018

DECLARATION

I declare that this research project is my original work and has not been presented for degree in any other university. No part of this project work may be reproduced without the prior written permission of the author and/or KCA University.

Sign: ……………………………………….. Date: …………………………

JAMES AREBA

Reg: 15/03067

This research project has been submitted for examination with our approval as university supervisors

Sign: …………………………………… Date: ………………………….

DR. MBURU LUCY,

KCA UNIVERSITY

DEDICATION

This project thesis is dedicated to my family and friends for their support during my studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

CBK	Central Bank of Kenya
CGAP	Consultative Group to Assist the Poor
CSMI	Credit Scoring Model for Individuals
ICT	Information and Communications Technology
MFIs	Microfinance Institutions
SACCOs	Savings and credit Cooperative Societies
SPSS	Statistical Package for Social Sciences
TAM	Technology Acceptance Model

ABSTRACT

Over the last two decades, credit assessments made by commercial banks have been evolving. Instead of the traditional assessment of the banks' credit experts which is subjective, increased credit risk means that comprehensive mathematical and statistical models must now be used. However, credit risk scoring applied for most commercial banks is not very effective, since a lot of defaulting exists which leaves the banks disadvantaged. The main purpose of this study is to conduct comprehensive modeling for effectively assessing credit risk. Primary data was collected through interviews with personnel related to credit control to establish the existing credit scoring methods and their viability. The target population included 50 staff members of Co-operative banks in Nairobi who comprised of credit officers, staff in risk management and staff ICT departments. The primary data was supplemented by secondary data gathered from bank records, company statistics, financial periodicals, books, journal articles and reports. The data was analyzed using descriptive and inferential statistics. The descriptive statistics include frequency distribution tables and measures of central tendency and measures of variability. Different inferential methods are tried and tested, leading to a conclusion that principal components analysis and logistic regression provide a suitable set of methods. Principal components analysis is used to identify significant variables among the many variables that can be used to assess credit risk. With fewer and effective measures of model performance, model development becomes a much more efficient process, the same goes for variable selection. Since the data used is only a small sample of the population, a resampling method is applied that is used to get stable estimation of credit scoring using a dataset of reasonable size. The developed model for credit risk scoring will inform management for decision making and provide predictive information on the potential for delinquency or default that may be used in the loan approval process and risk pricing. The study is expected to be of value to the various stakeholders who will include the management of commercial banks in Kenya and other financial institutions; to the CBK as the regulator, to the borrowers and to scholars and researchers.

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

The probability of defaulting i.e., customers failing to pay back the amount owed by them is an important variable in the day-to-day operation of any credit institute, and it is often used as a measure of credit risk (Cocheo, 2009). Credit risk assessment made by the bank is popularly known as *credit rating model.* In today's economy, many major lending institutions have abandoned the approach of basing their lending decisions on financial criteria only. Instead, this system has been replaced by credit rating models, where a computer program takes information provided by the applicant and other outside sources, and using a complex set of weighted variables, produces a single number by which to rate the applicant's credit risk (Kenneth, 2000). Due to the convenience of these systems, banks are relying more and more on computer credit scoring tools (Citron & Pasquale, 2014).

As late as the early 1990s, most lenders were still using a single "house rate" and relied on interview procedures to screen borrowers. As data storage and computing costs fell, and underwriting technology improved, lenders increasingly began to use estimates of default risk to price individual loans. Today, automated credit scoring has become a standard input into the pricing of mortgages, auto loans, and unsecured credit (Einav, Jenkins & Levin, 2013). Using data from the Survey of Consumer Finances, Edelberg (2006) documents the extent of this transformation. She finds that as a result the correlation between loan pricing and estimated and realized default risk has sharply increased. Grodzicki (2012) documents a similar pattern in the credit card industry and ties it specifically to lenders' investments in information technology. Data mining techniques especially have become very useful for credit scoring, since they involve a lot of data to make objective predictions (Huang et al., 2004; Yap et al, 2011).

1.1.1 Loan Repayment Performance

Loan default can be defined as the inability of a borrower to fulfill his or her loan obligation as at when due (Mwenje, 2006). According to Smirlok (2011) default is a risk threshold that describes the point in the borrower's repayment history where he or she missed at least three installments within a 24-month period.

Repayment performance thus serves as a positive signal for increasing the volume of credit availability to various sectors of the economy (Acquah & Addo, 2011). However, certain factors are considered before it is availed to the beneficiary and one of such factors is the beneficiaries' ability to repay the loan which in turn is also determined by many factors. According to Ugbomeh, Achoja, Ideh and Ofuoku (2008), credit repayment performance could be influenced by a myriad of factors such as interest rate, and the social relations and responsibilities of the borrower.

Kiiru (2007) found out that repayment performance is significantly affected by borrowers' characteristics, lender's characteristics and loan characteristics. The marginal effects of each set of characteristics are determined and analyzed. Repayment problems can be in the form of loan delinquency and default. Whatever the form however, the borrowers alone cannot be held responsible wherever problems arise as it is important to examine the extent to which both borrowers and leaders comply with the loan contract as well as the nature of the duties, responsibilities and obligations of both parties as reflected in the design of the Credit program rather than heaping blames only on the borrowers.

1.1.2 Commercial Banks in Kenya

In Kenya, the Banking Sector is composed of the Central Bank of Kenya, as the regulatory authority and the regulated; Commercial Banks, Non-Bank Financial Institutions and Forex Bureaus. As at December 2016, Kenya had a total of 42 commercial banks and 1 mortgage finance company with two banks; Chase bank and Imperial bank in receivership (CBK, 2016).

The banking sector in Kenya has reported massive growth and development in recent years. This is attributable to the effective regulation and reforms effected by the central bank after many banks went into bankruptcy in the 1990s, much of the growth in the banking sector has been witnessed in branch network expansion, growth in capitalization and asset base and the expansion of some of the banks regionally. The banks have also been in the frontline of automating their functions to give their customers good service. Kenyan banks have engaged in product innovation whereby use of internet and mobile technology has taken root in various local banks (CBK, 2016).

However, there has been some concerns about the amount of non- performing loans (NPLs) in commercial banks in Kenya. Due to the alarming default statistics, CBK has introduced reforms and enhanced enforcement measures in the banking industry that are meant to enhance data reporting, transparency and corporate governance and review of business models of banks (CBK, 2016). However, this has not eradicated the problem, and research investigation is needed to intervene.

1.2 Problem Statement

Although banks gather a wide range of information from loan applicants to understand whether to grant or not to grant loans (Hanna & Boyson, 2013), the number of non-performing loans in commercial banks in Kenya is rising year by year. The loan defaults in the banking sector had touched a decade high. For instance, by 2016, the bad loans stood at eight per cent of the total loans issued by bank; this is way above acceptable limit of non-performing loans which is 4% of the gross loans (CBK, 2016). According to the recent Credit Survey Report by the Central Bank of Kenya, “The ratio of gross non-performing loans to gross loans increased from 9.1 percent in December 2016 to 9.5 percent in March 2017. The increase in gross non-performing loans was mainly attributable to a challenging business environment” (CBK, 2017). The amount of short-term liabilities had grown by 9.9% from the previous year and an increase in NLPs was expected in the second quarter (CBK, 2017). In this regard, the banks have been under increased pressure from CBK to adhere to set regulations on treatment of non-performing debt. The upsurge in NPLs begs the question, whether the credit scoring models or systems being used by the banks are effective to analyze and appraise borrowers.

With the increased use of automated credit scoring systems, there is need to establish how effective the systems are in relation to helping or enabling the bank gather a wide range of borrowers' information and use the same to approval or decline of the loan requests, and even show the possibility of a consumer to default (Huang et al., 2004; Einarsson, 2008; Yap et al, 2011; Bravo et al., 2017). This gap epitomizes this research study which seeks to examine the effects of automated credit scoring systems are affecting loan performance in commercial banks.

1.3 Objectives of the Study

The general objective is to establish a model of credit risk assessment on bank loan performance.

1.3.1 Specific Objectives

The study will be guided by the following research objectives:

i. To determine the significant variables for determining customers' credit worthiness and assessing the credit risk.

ii. To investigate data mining techniques that can be used for building a credit risk assessment model.

iii. To develop a model for credit risk assessment.

iv. To test and validate the model developed in iii above.

1.4 Research Questions

The study will seek to answer the following research questions:

i. Which variables can significantly determine default rates for credit customers of Co-operative Bank in Kenya?

ii. Which data mining techniques can be used for building a credit assessment model?

iii. Which model can more effectively measure credit risk assessment?

iv. Which validation techniques can be used to establish the model effectiveness for credit risk assessment?

1.5 Motivation

There has been a concern about the amount of non- performing loans (NPLs) in commercial banks in Kenya. According to the Bank Supervision Annual Report (2012), the ratio of non-performing loans to gross loans increased from 4.4 per cent in December 2011 to 4.7 per cent in December 2012; this has risen to 12.1 per cent in the year 2014 (Bank Supervision Annual Report, 2014).

In the period June 2010 - June 2016, major fluctuations have been observed in growth of NPLs. The key observation in the period is increased credit risks since the third quarter of 2015 as reflected in growth rates of NPLs. The key observation in the period is increased credit risks since the third quarter of 2015 as reflected in growth rates of NPLs.

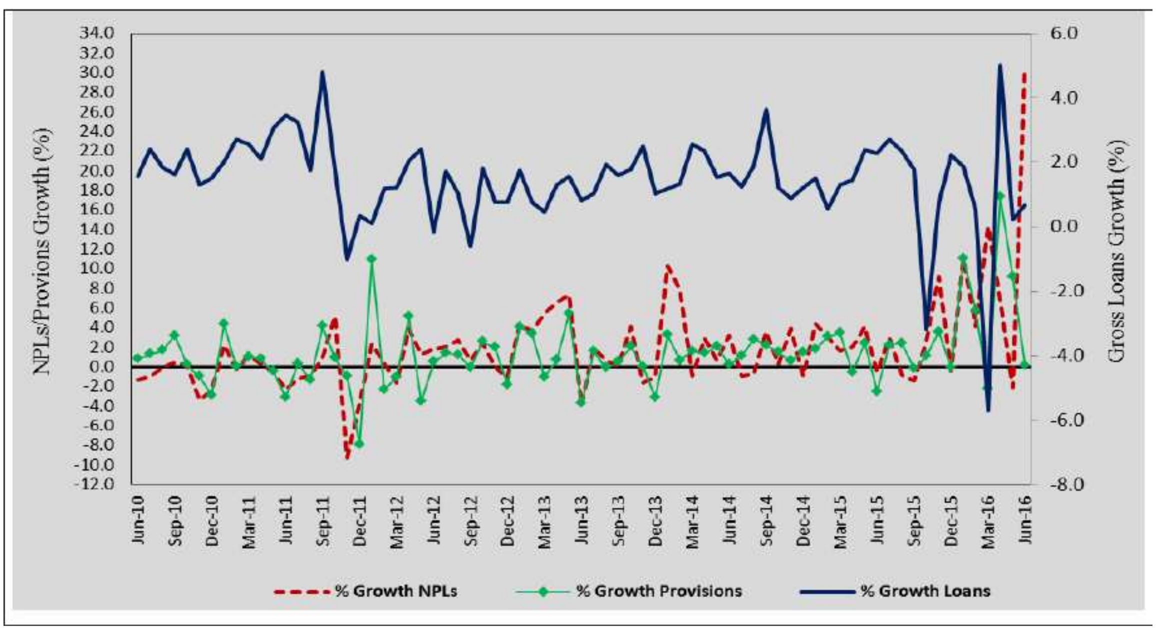

Figure 1.2: Monthly Trends in Loans, NPLs and Provisions (CBK, 2016).

In 2016, CBK reported that loan defaults in the banking sector had touched a decade high. In the first quarter of 2016, bad loans stood at eight per cent of the total loans issued by banks, up from 6.1 per cent in December 2015 and 4.6 per cent in June 2015. With a total loan book of Sh2.2 trillion, this means the bad loans are at Sh176 billion up from Sh139.4 billion in December, a Sh36.6 billion spike in the first three months of the year (CBK, 2016). By 2017, creditors from financial services were major defaulters as shown in Figure 1.3.

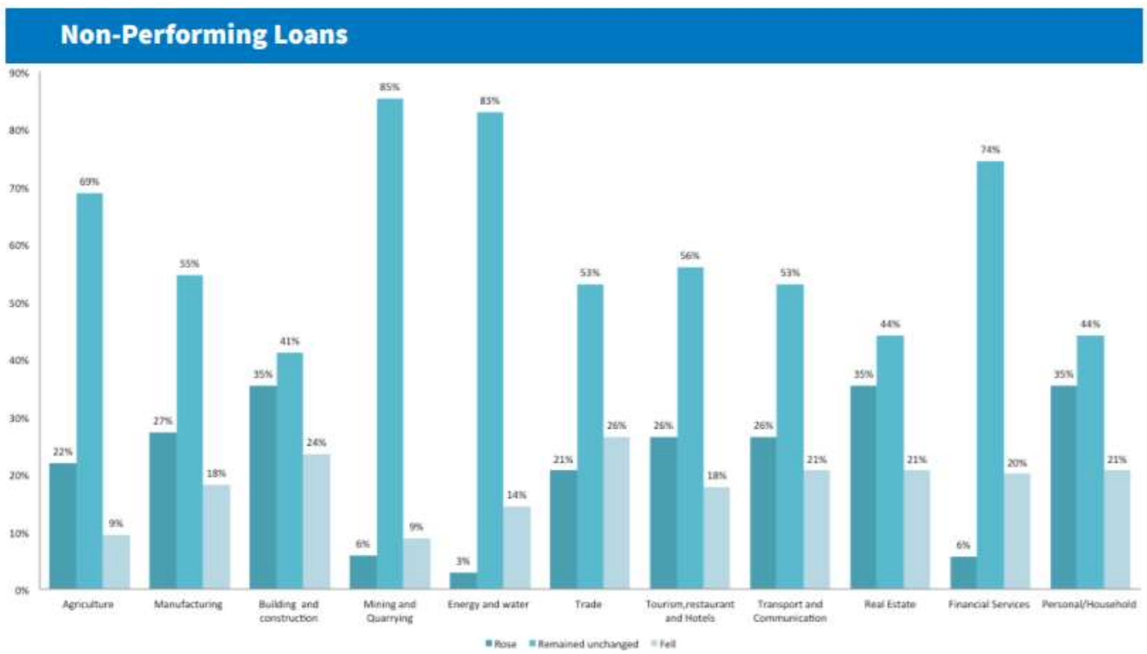

Figure 1.3: NPLs by Sector in the first quarter of 2017 (CBK, 2017 pg 11)

Due to the above alarming statistics, it is necessary to come up with a model that will enable effective credit scoring of potential borrowers and also determine the probability of defaulting.

1.6 Significance of the Study

The study is expected to be of value to the management of commercial banks in Kenya as they will get some insight on the effectiveness of credit scoring systems on loan performance in terms of loan repayment or default rate. Other financial institutions such as SACCOs and MFIs that are using such systems or considering using such systems will also benefit from this study.

This study is also expected to be of value to the borrowers. Most borrowers have limited financial experience and therefore are exploited by incompetent or unscrupulous lenders. This study will however enlighten the borrowers on how they are appraised using automated credit scoring systems to determine their credit worthiness; that is, how much a borrower can qualify and comfortably repay hence informing the banks' decision to issue out loans.

The study is also expected to be of value to scholars and researchers. The study will add value to the existing body of knowledge and act as a useful resource for those who would be undertaking research on credit scoring systems on loan performance. The study will also act as basis for further research.

1.7 The Study Scope

In the proposed study the researcher will examine the effect of automated credit scoring systems on loan performance in commercial banks in Kenya, with a focus on Co-operative Bank of Kenya. The study population will consist of staff in credit department, risk management department and ICT department. The choice of this study population is informed by the fact that these departments interact with automated credit scoring systems in the bank. The study will be limited to co-operative bank branches within Nairobi County.

1.8 Expected Challenges in the Study

The study anticipates that the respondents could be reluctant to provide the necessary data because the research study deals with quite internal business issues which may raise suspicion on the use of the data/information. Employees may also fear to give information about their company, as some may not be sure whether it is not allowed and would not want to be associated with such mistakes because they may be victimized by management. To overcome this challenge, the researcher will first seek permission from the management to collect data from the organization. The ethical requirements will be adhered to; confidentiality of the information will be upheld, and information will be used for study purposes only.

Another foreseen challenge involves getting accurate data and information from the respondents. This results from respondents not having adequate information on the matter under research, or respondents refusing to give the right information to the respondent for fear of that information getting to a third party. To overcome this problem; the researcher will make sure that he targets the right staff in the organization as respondents to the study. The researcher will personally administer the interviews to the respondents so that he can clarify and interpret the questions for the respondents to fully understand before their give information.

1.9 Organization of the Study

The study will be organized in five chapters. Chapter one lays out the background of the study, the problem statement, the research objectives, significance of the study, the scope and expected challenges of the study.

Chapter two will cover the literature review; it contains the empirical review to identify the knowledge gap. The empirical review is discussed based on the study objectives.

Chapter three presents the methodology to be adopted to achieve the objectives of the study. It describes the research design, the target population, sampling technique and sample size, instrument for data collection, data collection methods and finally the data analysis method.

Chapter four will chapter cover the presentation, analysis and interpretation of the results and findings of the study and lastly chapter five will present the summary of the findings, conclusions, policy recommendations as well as recommendations for further research.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter covers theoretical and empirical review to identify the knowledge gap. The theoretical review discusses the theories that inform the study; empirical review discusses past studies by other authors on the specific research objectives while the conceptual framework presents the schematic diagram that shows the interactions between the independent variables and the dependent variable. The chapter ends with research gaps identified.

2.2 Credit Scoring

Credit scoring is one of the most crucial processes in banks' credit management decisions. This process includes collecting, analyzing and classifying different credit elements and variables to assess the credit decisions (Abdou & Pointon, 2011). Credit scoring has been regarded as a core appraisal tool of different institutions during the last few decades, and has been widely investigated in different areas, such as finance and accounting. Different scoring techniques are being used in areas of classification and prediction, where statistical techniques have conventionally been used (Abdou & Pointon, 2011).

Credit risk assessment models are some of the most successful applications of research modeling in finance and banking, as reflected in the number of scoring analysts in the industry, which is continually increasing. However, credit scoring has been vital in allowing the phenomenal growth in consumer credit over the last five decades (Chuang & Lin, 2009). Lenders in developed countries analyze the creditworthiness of borrowers based on their credit histories taken from credit bureau and also check borrower's salary and experience before loan approval (Schreiner, 2010).

In developed countries, credit scoring is well established, and the number of applications is increasing because of excellent facilities and vast information being widely available, whilst in less developed or developing countries, less information and facilities are available. Advanced technologies, such as those used with credit scoring have helped credit analysts in different financial institutions to evaluate and subsequently assess the vast number of credit applications (Abdou & Pointon, 2011).

The preparation of inputs, the estimation of parameters and the selection of an appropriate statistical model all are based upon statistical valid procedures. The statistical models used by credit scores can be improved over time as additional data are collected (Wu, 2008).

2.3 Assessing a Customer's Creditworthiness

Using a set of criteria such as employment, income, age, assets, outstanding debt and history of repayment, banks can compute a score that could help determine the applicant's credit worthiness. The manual process is cumbersome and has several problems. First, the manual system required the lender to hire skilled operators to manually calculate these scores, resulting in excessive administrative costs. Second, the lenders who, due to costs, chose not to use these types of systems were forced to rely primarily upon the business judgment of their lending officers to approve loan applicants (Richardson, 2009).

West (2010) has stated that credit scoring is widely used by the financial industry, mainly to improve the credit collection process and analysis, including a reduction in credit analysts' cost; faster credit decision-making; and monitoring of existing customers. Also, around 97% of banks are using credit scoring for credit card applications, and around 82% of banks are using credit scoring to decide correctly who should be approved for credit card applications.

Lending institutions have started adopting the credit assessment models in evaluating loans to avoid large losses. Classification models in credit assessment analyze the characteristics of applicants such as age, income, marital status, payment history are used to classify new candidates into good or bad (Chen & Huang, 2013). According to Chijoriga (2011), Credit assessment models can be qualitative as well as quantitative in nature. Qualitative technique is judgmental and subjective; the disadvantage of qualitative method is that there is no objective base for deciding the default risk of an applicant. While, quantitative technique is a systematic method to categorize into performing or non- performing loans and it has removed the shortcomings of qualitative technique and proved to be more reliable and accurate model.

There are three internal components to the credit scoring model: there are inputs, which are obtained from the dataset of applicant companies or borrowers; there are also parameters, which are used to weight the inputs and to control the logic of the model; and lastly there is a well-defined statistical algorithm to combine the inputs and the parameters to create a score (Wu, 2008). Figure 1.1. below shows credit scoring model development.

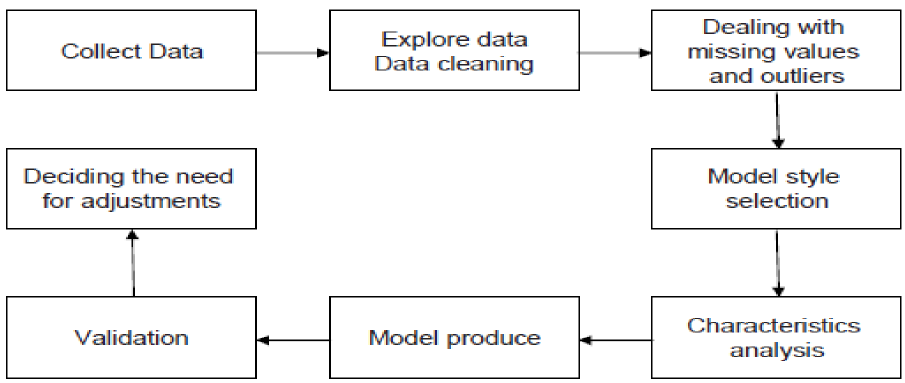

Figure 2.1: Credit Scoring Model Development.

2.3 Credit Assessment Techniques

Different credit assessment techniques are commonly used in practice to rate borrowers based on their credit worthiness. Three main techniques are heuristic evaluation, the use of causal techniques and the use of analytics (Datschetzky et al., 2015). In practice, combinations of heuristic and causal techniques, or heuristic and analytics can be applied thereby yielding a hybrid assessment model.

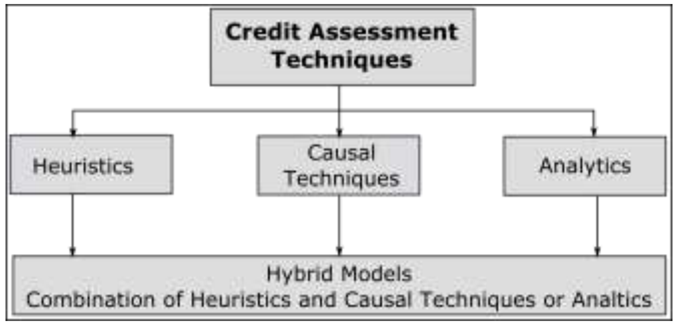

Figure 2.2: Overview of Credit Assessment Techniques

2.3.1 Heuristic Techniques

Heuristics employ past experience to assess the future credit worthiness of a potential borrower (Altman and Saunders, 1998). Several credit experts can select a number of relevant factors and their respective weights, based on their background experience. Such approaches include classic rating questionnaires that use factors such as gender, age, marital status, and income. Qualitative systems also belong to this category of techniques, but in contrasts to questionnaires, qualitative systems are not assigned a fixed value for each factor. Instead, an expert from the credit authority can evaluate the credit applicant for each factor and award a grade. The final assessment is based on weighted average of all grades. Expert systems and fuzzy logic are other examples of heuristic techniques.

The expert systems attempt to recreate human problem-solving capabilities using data and selection rules by credit professionals in order to conduct and expert valuation. Altman and Saunders (1998) have proved that financial experts can be very pessimistic in their evaluation of credit risk, and that multi-variate credit scoring systems can therefore outperform such expert systems. Fuzzy logic systems are a special case of expert systems with the added capability of fuzzy logic. In this system, the specific values entered for credit worthiness criteria are not assigned to a single categorical estimate, e.g. “high” or “low”. Instead, they are given multiple values which are in line with human decision-making behavior.

2.3.2 Causal Techniques

Causal models as used for credit assessment apply financial theories to assess credit worthiness of lenders (Einarsson, 2008). Their approach is different from many other approaches because these models do not depend on empirical datasets. Methods here include option pricing models (OPM), cash flow models (CFM) and fixed income analysis (FIA).

OPMs originate from the Option Pricing Theory of Black and Scholes (1973). This theory was used to price options and was also applied to valuation of risk based on transactions. OPMs are based on the idea that credit defaulting happens when the economic value of the borrower's asset falls below the economic value of debt. Therebefore even though OPMs can be constructed without needing a comprehensive defaulting history, but it requires data on the economic value of assets, debt, equity and volatilities (Einarsson, 2008). The amount of data required makes it impossible to use OPMs in the public sector. E.g. it is impossible or very difficult to assess the economic value of assets in the public and corporate sector.

CFMs simulate future cash flow that results from the assets being financed. These models are therefore good for assessing defaulting rates within specialized lending institutions. CFMs rate the transaction itself, not the potential borrower. The result can therefore be referred to as a transaction rating. CFMs are viewed and variant of OPMs where the economic value of a borrower is calculated based in the expected future cash flow. The problem with CFMs is the focus on transactions, and therefore the inability to capture characteristics of the borrowers themselves, which are also significant for default.

FIA is from Portfolio Theory of Markowich in 1959, which has been common in the fixed-income area that involves corporate and government bond and banks' portfolio of loans. FIA can be applied to banks portfolio to price new loan applicants and their interest rates after determining their probability of defaulting. The basic objective of FIA is to maximize return for a given level of risk and to guide the pricing of risky assets. Even though FIA is a useful model, it does not have widespread application and its effectiveness is still unknown.

2.3.3 Analytics

Analytical models apply information technology and information systems processing on data suggested by credit professionals in order to predict creditworthiness. This approach is much better than the heuristics approach which relies purely on background experience of credit professionals which can be sometimes inaccurate and subjective.

According to Einarsson (2008), the most common types of analytics used in credit scoring include traditional methods such as discriminant analysis and the more advanced methods like statistical and machine learning methods.

2.3.3.1 Discriminant Analysis

This method was introduced in 1968 by Altman in a z-score formula for predicting bankruptcy. Altman's method was the first approach for predicting bankruptcy in the banking industry using financial ratios. A linear multi-variate analysis was applied in order to formulate the z score formula using data from 66 firms' half of which had gone bankrupt. The z-score formula is as follows (Frydman et al, 1985):

$$Z = 0.12X_1 + 0.14X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5$$

Where

- X_1 = Working Capital / Total Assets, used to measure net liquid assets in relation to the size of the company.
- X_2 = Retained Earnings / Total Assets, used to measure profitability to reflect the company's age
- X_3 = Earnings Before Interest and Taxes / Total Assets used to measure operating efficiency apart from tax and leveraging factors.
- X_4 = Market Value Equity / Book Value of Total Debt used to measure how much firms market value can decline before coming insolvent.

- X_5 = Sales / Total Assets used to measure turnover and varies greatly from industry to industry.

All the values except the Market Value Equity (X_4) can be found directly from financial statements of businesses or individuals seeking credit. Results of the discriminant analysis are summarized as:

- Z-score of 2.99 and above: low probability of defaulting
- Z-score between 1.81 and 2.98: intermediate probability of defaulting
- Z-score below 1.81: high probability of defaulting

The main advantage of discriminant analysis compared to other methods is that the individual weights show the contribution of each explanatory variable (Anolli et al., 2013). The result of the linear function is then also easy to interpret, as low Z-score is observed it represents a poor loan applicant. The downside to this approach is that it requires the explanatory variables to be normally distributed. The explanatory variables are also required to have the same variance for the groups to be discriminated. In practice this is however often thought to be less significant and thus often overlooked.

2.3.3.2 Machine Learning Methods

With advancements in computer programming, new machine learning approached have developed, e.g. Recursive Partitioning Algorithm (RPA), k-Nearest Neighbor Algorithm (kNN), Support Vector Machine (SVM) and Neural Networks (NN). More advanced analytics also exists, such as Hazard Regression Modeling.

RPA is a data mining approach that uses decision trees. The approach is also known as Classification and Regression Trees (CART). Frydman et al. (1985) found out that RPA is a much better approach that discriminant analysis. Since then RPA has been used in a variety of business and scientific applications (Anolli et al., 2013).

kNN is a non-parametric classification approach. It considers the average of the dependent variable for all *k* observations that are most similar to a new observation.

SVM is an optimization method that is very similar to discriminant analysis but is a little more complex because the optimal non-linear boundary must be constructed. SVM simultaneously minimizes the empirical classification error while maximizing the geometric margin of different datasets. It therefore results in dual functions for solving standard problems. Research by Yu et al (2007) and Hastie et al. (2009) shows that SVM outperforms methods like logistic regression.

NN apply information technology to simulate the complicated manner in which human beings process information. It applies a multiple-stage processing of information where each stage has hidden correlation among the dependent variables. This makes the process a Blackbox model, where the internal structure of the model is not viewable and is difficult to understand and interpret. NNs can process many forms of information which makes them good at formulating rating models. By combining Blackbox modeling and a large set of information, NNs generally have a lot of discriminatory power. However, the Blackbox nature of NNs results in many acceptance problems (Einarsson, 2008). Altman et al. (2004) concluded that NNs do not significantly improve upon the linear discriminant structure.

2.2.3.3 Statistical Analytics

One of the most popular methods in this group is the multivariate regression model which considers the multiple variables that contribute to failure, i.e. defaulting in the case of credit modeling. Lando (2013) refers to multivariate regression as the most natural statistical framework for analyzing the probability of credit risk.

Multi-variate regression can be used alongside other descriptive techniques to analyze the factors that increase of decrease the possibility of defaulting among customers.

Logistic Regression

This is one of the most effective multivariate regression approaches for credit assessment. It uses the dependent variable as a binary (0 or 1) variable that is usually 1 if a borrower defaulted in the observation period and 0 if he didn’t default. The independent variables are all potentially relevant parameters to credit risk. A logistic regression formula is usually represented using the logit link function as follows:

$$p(X) = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k)]}$$

Where ***p(X)*** is the probability of defaulting given the *k* input variables **X**. Logistic regression has a number of benefits over discriminant analysis. First, it does not require the input variables to be normally distribute and therefore qualitative creditworthiness variables can be put into the models. Second, unlike with the z-scores which are a bit difficult to interpret, the results of logistic regression can be interpreted directly as the probability of defaulting (Anolli et al, 2013).

2.4 Empirical Review

A review of the existing literature shows that, Einav, Jenkins and Levin (2013) conducted a study on the impact of credit scoring on consumer lending. They investigated the adoption of automated credit scoring at a large auto finance company and the changes it enabled in lending practices. The study found out that the adoption of credit scoring technology led to a large increase in profitability and improved repayment. Lending to the highest-risk applicants contracted due to more stringent down payment requirements, and lending to lower-risk borrowers expanded. The study also identified two distinct benefits of risk classification through automated credit scoring: the ability to screen high-risk borrowers and the ability to target more generous loans to lower-risk borrowers.

In Pakistan Samreen (2012) conducted a study to evaluate credit risk in commercial banks of Pakistan using automated credit assessment models. A credit scoring model was developed known as the Credit Scoring Model for Individuals (CSMI), which can be used by commercial banks to determine the creditworthiness of individual borrowers requesting for personal loans. The results of the developed credit scoring model were compared with the other statistical credit scoring techniques known as logistics regression and discriminant analysis. Type I and type II errors had been calculated for all the credit scoring models used. The results show that the proposed model "CSMI" had more accuracy rate with no errors as compared to other manual methods.

Thun (2011) reported that there are several standard sources of information required for assessing the borrowers default. In Figure 1 below the essential information required for company credit scoring/rating is shown. These are industry risk and competitive in the market, management, ownership and corporate governance and financial soundness which includes profitability, liquidity and financial flexibility, and capitalization. Moreover, the same information will be required for score the financial institutions except bank account data that can be substituted with relation with other financial institutions.

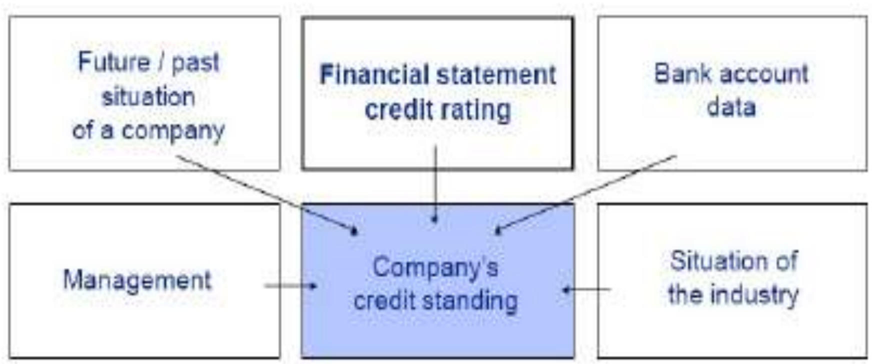

Figure 2.3: Sources of Information to Assess Borrowers Default Risk. Source: Thun (2011)

In Iran, Dastoori and Mansouri (2013) conducted a study to establish the effectiveness of automated credit scoring model for Iranian banking customers in forecasting creditworthiness of borrowers. The creditworthiness of the borrowers was determined through using financial information obtained, from 2006 to 2011. The results of the study indicated that the two fitted automated models were reasonable reliable in predicting the creditworthiness of banking customers.

Owizy (2013) evaluated the impact of credit management on financial performance of Nigerian banks, with particular reference to UBA Plc. Financial ratios as measures of bank performance and credit indicators were the data collected from secondary sources mainly the annual reports and accounts of sampled banks from 2004 - 2008. Descriptive, correlation and regression techniques were used in the analysis. The findings revealed that credit management has a significant impact on the profitability of Nigeria banks.

Byusa and Nkusi (2012) investigated effects of credit policy on bank performance in selected Rwandan Commercial banks. The aim of this study was to investigate the effects of credit policy on bank performance using data on selected Commercial Banks. The results obtained indicated that the Rwanda's commercial banks increased their accounts, increased customer base and improved their financial indices, thereby maximizing their profits. However, inadequate competition in the banking system led to high spreads. Banks have unusually high and increasing average interest rate spreads and interest rate margins showing both highly poor competition and inefficiency.

Ntiamoah, Diana and Kwamega (2014) carried out a study on assessment of the relationship between credit management practices and loan performance using some selected microfinance institutions in the Greater Accra region of Ghana as a case study. Results of the study indicated that there was high positive correlation between the credit terms and policy, lending, credit analysis and appraisal, and credit risk control and loan performance.

Ayodele, Thomas, Raphael and Ajayi (2014) carried out a study on impact of credit policy on the performance of Nigerian Commercial Banks using Zenith Bank Plc as case study. Primary data was collected through questionnaires served on sixty (60) respondents of the bank. The findings from the study showed that having a good credit policy in place goes a long way in minimizing the incidence of bad debts.

In Kenya, Simiyu (2008) investigated on the techniques used by micro finance institutions in the management of credit risk in Kenya, and to examine the main challenges facing the micro finance institutions operating in Kenya in the management of credit risk. The study established that most microfinance institutions use 6C (6C's are Character, Capacity, Capital, Collateral, Conditions and Control), techniques of credit risk management, the study also revealed that understanding the organizations exposure to the customers is treated as critical by the micro finance institutions. The study established that majority of the institutions used credit matrix to measure the credit migration and default risk.

In another study, Moti, Masinde, Mugenda and Sindani (2012) studied loan performance in Micro Finance Sector in Kenya to understand the effectiveness of credit management system used. The study was informed by the high levels of non-performing loans in micro finance institutions. The study found out that high involvement of credit officers and use of borrower's information in formulating credit terms was found to have a significant relationship with loan performance. The study recommended that the 5C's (character, capacity, capital, collateral and conditions) model of client appraisal was important when appraising clients, therefore microfinance institutions should take a greater consideration on character of the client, capacity of the customer to repay, collateral attached as security, history of repayment, need assessment and size of the business.

A review of the existing literature on automated credit scoring systems and loan performance shows that this study area has attracted many researchers in the recent past. Several researchers have shown that various credit assessment models are used to categorize borrowers into either high-risk or lower-risk borrowers. Some of the studies have gone further to evaluate the accuracy rate of the credit assessment models/ systems as compared to other manual methods. However, most of these evidences have are from studies conducted in developed countries or in more developed economies than Kenya and therefore it becomes difficult to generalize the findings. There is little empirical literature in Kenyan context to show the extent of use of credit assessment models in commercial banks and how this use affects loan performance in commercial banks in Kenya. This is the gap that this study will be seeking to fill.

2.5 Conceptual Framework

As guided by the literature review, the following conceptual framework was identified. It illustrates the interaction between independent variables and the dependent variables in the study. In this study, the independent variables include client's financial information, client appraisal information, credit terms and credit risk control measures which measure the probability of defaulting as well as the 5Cs (Character, Capacity, Collateral, Capital, Condition) which measure the customer's credit worthiness.

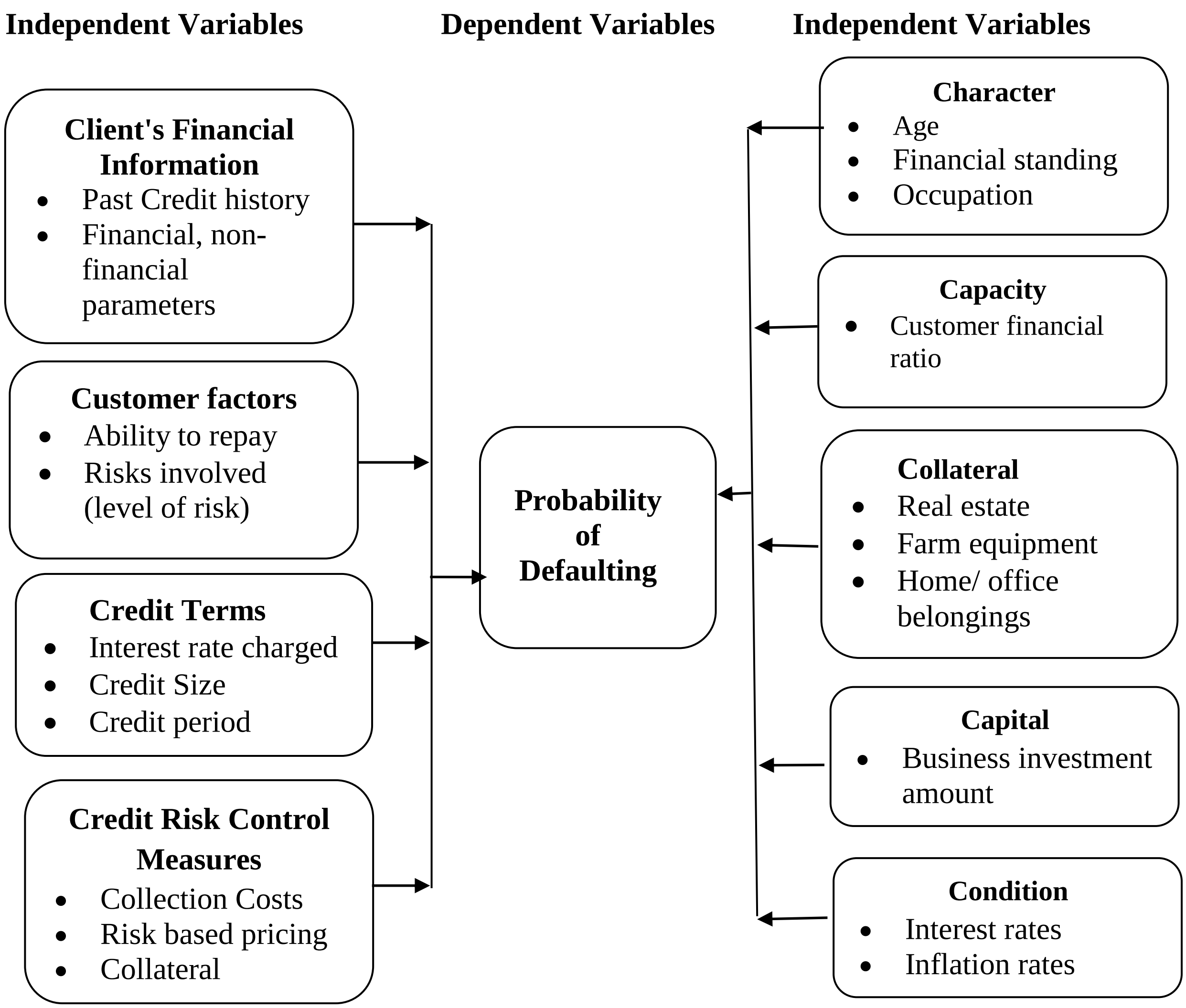

Figure 2.4: Conceptual Framework. Source: Author (2018)

CHAPTER THREE

METHODOLOGY

3.1 Introduction

The chapter will look at the research methods that will be employed in the study in order to achieve the objectives of the study. This chapter covers the research design to be adopted, target population of study, sample size and sampling technique, data collection instrument, data collection procedures and data analysis methods.

3.2 Research Design

The study will adopt a descriptive research design. Descriptive research design is one of the best methods for conducting research in human contexts, because of portraying accurate current facts through data collection for testing hypothesis or answering questions to conclude the study. A descriptive study is concerned with finding out the what, where and how of a phenomenon (Creswell, 2014). The descriptive design will therefore be appropriate for this study since it will help in collecting data in order to answer the questions of the current status on how the automated credit scoring systems are affecting loan performance in commercial banks in Kenya.

3.2.1 Target Population

The target population will be bank staff in the credit department, risk management and ICT departments in Co-operative bank, Nairobi. Since the population is small, a census study will be adopted whereby the entire population will be considered for the study. According to Cooper and Schindler (2011) a census is feasible when the population is small and necessary when the elements are quite different from each other. When the population is small and variable, any sample we draw may not be representative of the population from which it is drawn. Therefore, a census study was deemed to be appropriate for study since the sampling frame is small; therefore, all the 50 respondents formed the sample size for the study.

Table 3.1: Study Population

Population Categorization	Number
Staff in credit department	20
Staff in risk management department	20
Staff in ICT department	10
Total	**50**

3.2.2 Data Collection Instrument

In this study, the credit risk is assumed to be a function of actual and expected loss, generalized by the following equation:

Credit risk = max {Actual Loss – Expected Loss, 0}

where the actual loss is the observed financial loss to a lending institution. Credit risk is therefore the risk that the actual loss will be greater than the expected loss. Expected Loss is an estimate that can be divided further into the following components according to Ong (2007):

Expected Loss = Probability of Default x Exposure at Default x Loss Given Default

where Probability of default (PD) is the probability that a borrower will default on a debt before it reaches maturity. Exposure at default (EAD) is the amount that the borrower legally owes the bank, which is not necessarily what was borrowed. Loss given default (LGD) is a percentage of the actual loss of EAD that the bank will suffer. Many banks tend to collect collateral such as land title deed, while others hold bilateral contracts to protect themselves against liability.

In order to model the Expected Loss, the credit rating model will combine values of PDs, EADs and LGDs. A graphical representation is as follows:

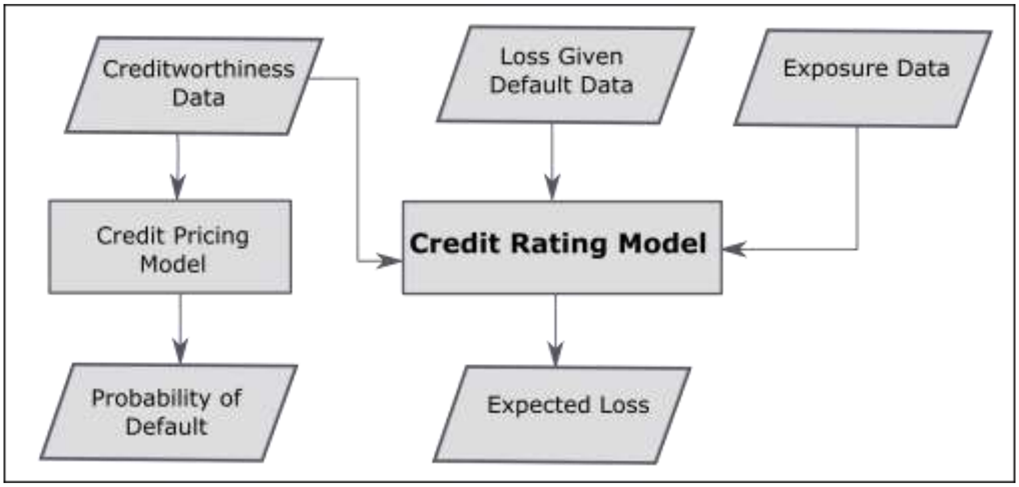

Figure 3.1: Typical Credit Assessment Environment

PDs are derived from credit risk variables of borrowing clients, the EAD can easily be estimated as the current risk exposure. An estimation of LGD can be obtained by collecting past historical data of previous LGDs and displaying these in the form of a histogram.

Credit risk variables from borrowing clients were identified by liaising with the credit section of Co-operative bank to conduct interviews on certain critical study-related aspects about financial standing, lending and repayment over five years (2013 to 2017). The interview sessions had both closed and open ended questions. The closed ended questions helped the researcher to collect quantitative data while open ended questions made up the qualitative data. Interviews were considered as the appropriate data collection instrument for this study since interview can provide a high degree of data standardization, it is relatively quick to collect information from people in a non-threatening way and it is cheap to administer. Interviews can also elicit detailed answers to complex problems (Kombo & Tromp, 2009) and to guide the selection of secondary data. Secondary data was obtained by randomly sampling the credit database as well as other sources such as statistics of companies, financial periodicals, books, journal articles and reports.

3.3 Data Processing and Modeling Procedure

The researcher sought permission and consent to collect data from the database and various representatives in management (the credit/loan officers, risk management officers and IT officers). Consents were sought through use of a letter for data collection which was obtained from the University. After permission was granted, appointments were made with the respective respondents for interviews and permissions to collect selected records of interest. Personal data collection gave the researcher a chance to interpret and clarify aspects that were not clear. This ensured maximum cooperation from the respondents and a successful data collection process.

Data records were obtained from the Co-operative bank's credit database. The data underwent a cleansing process as follows.

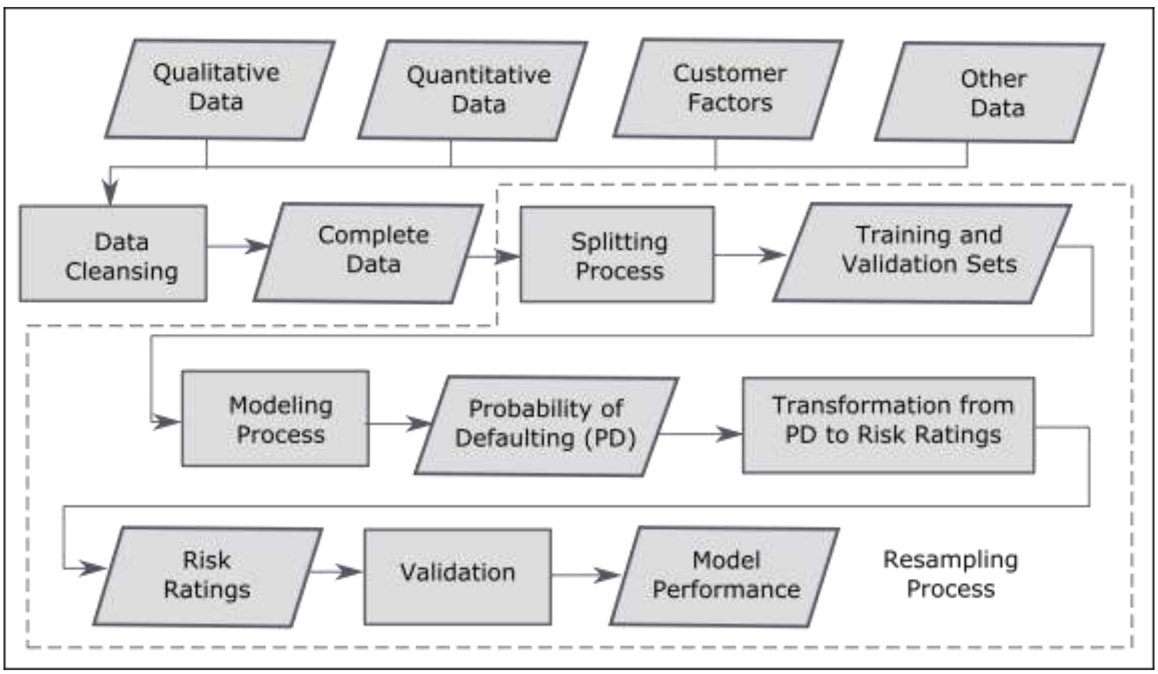

Figure 3.2: Data Processing and Modeling Procedure

A customer that has not been observed for two consecutive years is either a new customer or a retiring customer, the record was removed. Any records with missing values were also removed. Once the data was cleansed and complete, it was split into training and validation datasets. 50% was the training dataset used to fit the models, 25% the validation dataset used to estimate the prediction error for model selection, and 25% the test dataset used to assess the error of the final model chosen.

Because the dataset of randomly picked records represents a small sample, the processes of splitting, fitting, transforming and validating are performed iteratively. The training and validation datasets (the modeling datasets) are randomly selected from 2014, 2015 and 2016. These modeling datasets are recursively split by selecting a random sample with no replacement so that the training set is 2/3 of the modeling dataset and the validation dataset makes up 1/3. The test dataset is the 2017 dataset.

In order to equally assess all credit categories, the dataset had to include five main sectors: real estate, trade, production, service and transport (Lagat, Mugo and Otuya, 2013). Real estate is made up of borrowers who require money for procurement of properties such as land and buildings. Trade has individuals and companies that borrow money for the purpose of starting or expanding their business ventures. Production includes manufacturers and farmers. Service sector includes customers from non-manufacturing and non-farming areas, such as ICT services. Transport sector has customers who have the sole purpose of conducting transportation, such as purchasing motorbikes, matatus and other kinds of transport vessels.

3.4 Data Mining and Modeling

The data collected was analyzed using the R statistical tool. The data analysis involved generation of descriptive and inferential statistics. The descriptive statistics include frequency distribution tables and measures of central tendency (the mean), measures of variability (standard deviation) and measures of relative frequencies. The inferential statistics include a regression model to establish the relationship between variables. The analyzed quantitative data will be presented using tables, charts and graphs.

The regression model takes the form:

$$p(X)=\frac{1}{1+\exp(-z)}$$

Where *p(**X**)* is the probability of defaulting and *z* is defined as:

$$z=\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\beta_4 x_4+\beta_5 x_5$$

And

χ_1 = Client's age

χ_2 = Client's gender

χ_3 = Client's financial information

χ_4 = Credit terms

χ_5 = Credit risk control measures

β_0 = the intercept (value of EY when X = 0)

$\beta_{1\text{-}n}$ = the regression coefficient or change included in Y by each χ

3.5 Study Variables

The conceptual framework in Chapter 2 illustrates the interaction between independent variables and the dependent variable in the study.

In order to assess the creditworthiness of a loan applicant, several quantitative outcomes are measured. Borrowing a leaf from Moti et al. (2012), this thesis assesses credit worthiness using five key categories equivalent to the 5 Cs: Character, Capacity, Collateral, Capital and Condition.

3.5.1 Character

Character is a technique for weighting an average over several attributes of the loan applicant. The total weighted score quantifies the customer's credit worthiness (Myers and Forgy, 2005). Customer-related factors can be categorized into social, economic, personal and cultural. The social aspect of a customer is typically measured by his lifestyle, i.e. the way a customer lives, his relationship circle (reference group), consumption and entertainment trends. Credit institutions particularly seek to identify a customer's reference group as this is known to highly influence a customer's credit worthiness (Moti et al., 2012). Economic factors included information about ownership of property by the customer or business, as well as relative ownership within the reference group. Personal factors can include age, occupation, personality, economic standing and family standing. For example, families with smaller children might be more likely to default due to the obvious expenses, while more settled families with older children are likely to have stable collateral on their assets.

3.5.2 Capacity

In considering the capacity to pay, lending institutions will consider the cash flow from a customer or business, the frequency of repaying the credit and previous successes in loan repayment. According to Orlando (1990) a customer's financial ratios can assist lending institutions to know whether the borrower can pay the current expenses accruing from the credit as well as additional expenses of any new credit advancement.

3.5.3 Collateral

Collateral represents the assets that a customer will pledge as the alternative payment resource for a loan. Collateral will mostly consist of tangible capital such as land or other form of real estate, equipment from farming and manufacturing, office furniture and equipment, etc. Lending institutions will generally accept only the collateral whose lifetime value corresponds to the length of the loan period.

3.5.4 Capital

This refers to the amount of money that a borrower individual or business has invested. Capital generally quantifies the amount of risk that a borrower will incur if a business fails.

3.5.5 Condition

This quantity is an indicator of the sensitivity of a borrower to external forces such as interest rates (static or dynamic), inflation rates, business cycles and competition pressures. While most of the factors quantified here are external to the customer, condition typically indicates a customer's vulnerability that can impact on credit risk.

3.5.6 Logistic Regression Variables

For the regression modeling, the independent variables are: client's financial information, client appraisal, gender and age, credit terms and credit risk control measures while the dependent variable is the probability of default.

3.6 Conclusion

This chapter has outlined the methodology used in the research study and described the research design that will be used, the target population, data collection instrument and lastly how the data analysis was carried out. The study expects that the research instrument used will be adequate and reliable to collect information/data that will help make inferences on the study objectives. The next chapter presents the data analysis, presentation and interpretation of the findings as analyzed from the data collected.

CHAPTER FOUR

RESULTS

4.1 Introduction

This chapter presents the main findings of the study. It begins by showing results of the summary descriptive statistics. Afterwards, results of mathematical and statistical procedures will be used to asses both the customer's creditworthiness and the probability of default. Results from different validation methods will be presented. Furthermore, a discussion of the performance of significant variables will be given.

4.2 Sample Description and Default Rates

Table 4.1 shows how the portfolio is split between the sectors identified in Chapter 3. It shows the total number of observations of the complete data and percentages for each sector, as well as the defaulting behavior per sector.

Table 4.1: Summary of Portfolio distribution across sectors and the default rate

Sector	Observations	[%]	Default Observations	[%]	Default Rate (%)
Real Estate	1163	[19.35]	71	[49.31]	6.10%
Trade	1157	[19.16]	21	[14.58]	1.82%
Production	1498	[24.81]	23	[15.97]	1.54%
Service	1244	[20.60]	10	[06.94]	0.80%
Transport	977	[16.18]	19	[13.19]	1.94%
Total	**6039**	**[100.00]**	**144**	**[100.00]**	**2.38**%

From the table, it is obvious that real estate has the highest number of defaulters. Default rate in this sector takes almost 50% of the overall default rate. However, the service and production sectors have quite low rates.

4.3 Assessing Customer Creditworthiness

Several quantitative outcomes are used to determine a customer's credit worthiness. Borrowing a leaf from Moti et al. (2012), this thesis assesses credit worthiness using five key categories that have been described in Chapter 3: Character, Capacity, Collateral, Capital and Condition.

4.3.1 Character

According to the results of the histograms, the histogram that represents all sectors appears to be normally distributed. But when different sectors are split each sector shows a different pattern. Real estate has a long tail on the lower side meaning that most of the real estate loan applicants do not score very well at the character level. Production and Trade sectors show quite a uniform distribution with most applicants concentrating in the middle. The transport sector is also skewed to the left, maybe because Real Estate and Transport applicants are self-employed and have no employers to vouch for their characters.

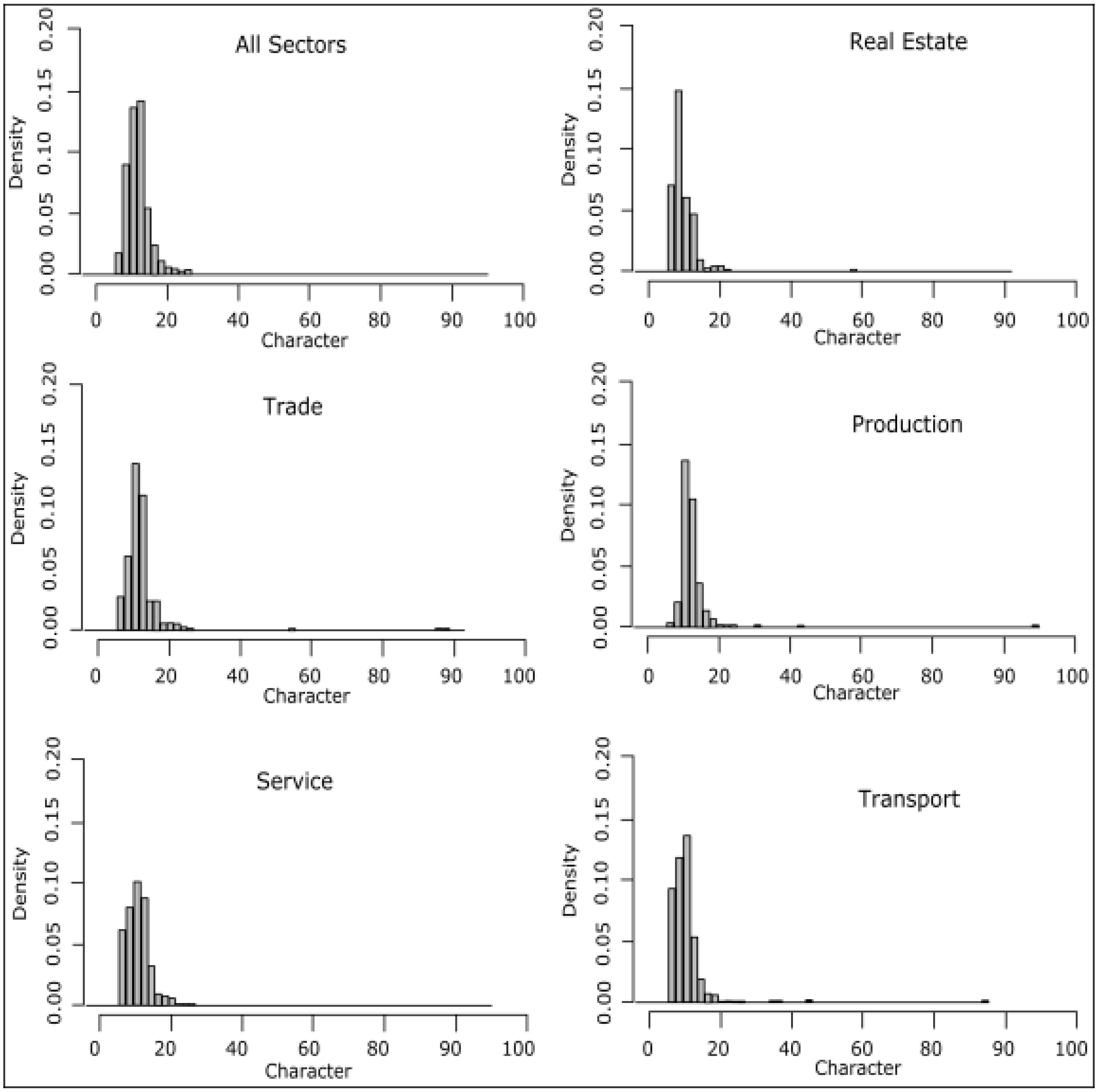

Figure 3.1: Histograms of Character Rating for all Sectors and per Sector.

4.3.2 Capacity

In terms of capacity the Real Estate, Service and Transport Sectors have the frequency of loan applicants distributed more on lower side with some applicants even having a negative value. These sectors don't seem to have enough resources to pay debts over the next 12 months. Trade and production show better distribution which are positive, and Production has the best.

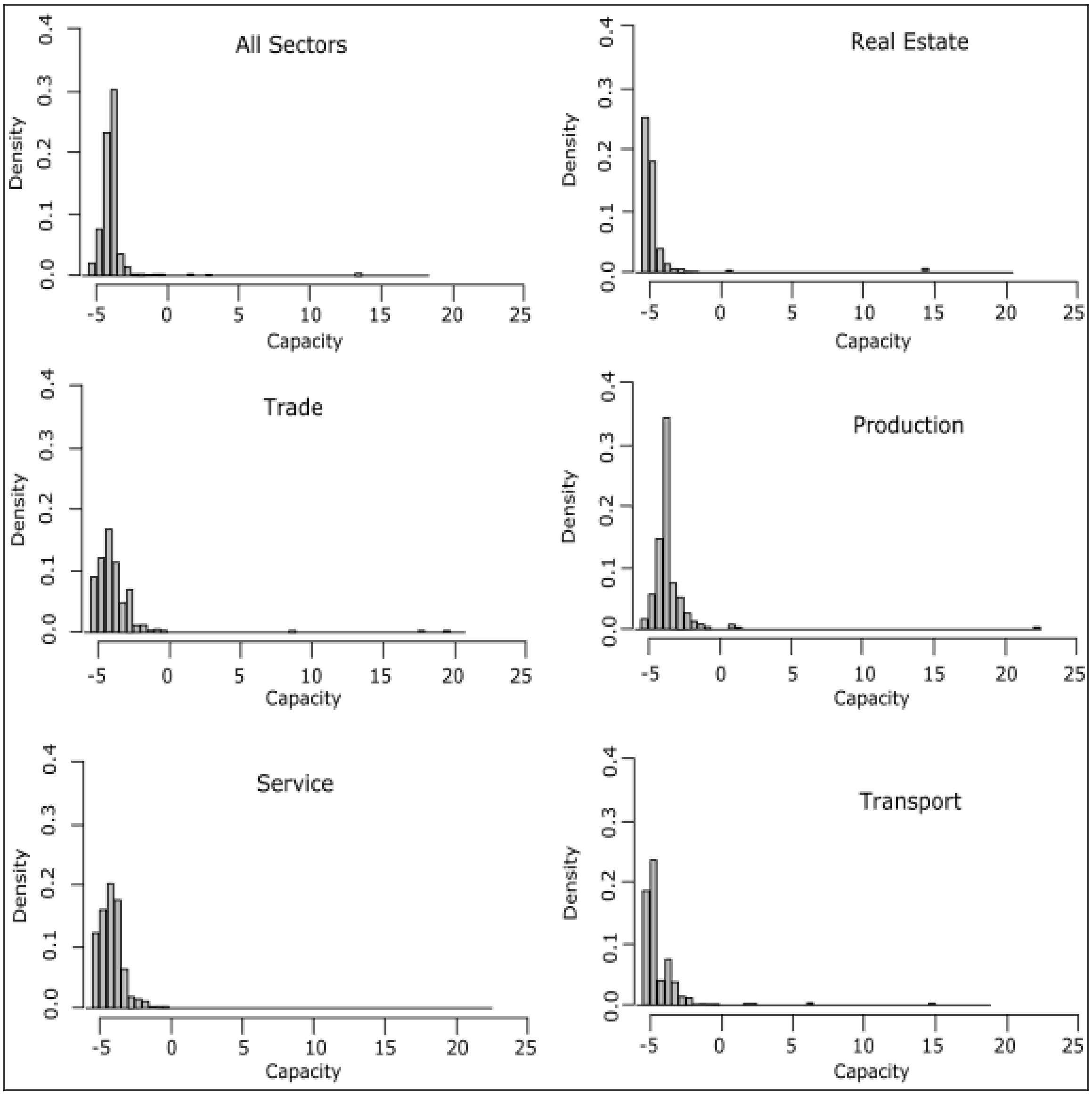

Figure 4.2: Histograms of Capacity values for all Sectors and per Sector. Values below 0.02 are not shown

4.3.3 Collateral

Table 4.2 shows summary statistics quantifying collateral over all sectors and for each individual sector. If current liabilities value is 0, then the Collateral Value (CV) is awarded 1000, otherwise it is a fraction of 1000. Since values are too close together and precision is required, results of collateral summaries are presented in a table rather than as histograms. From the results of the table, real estate has the lowest mean collateral, while production and service sectors have the largest mean collateral. Overall the collateral ratio is also highest for production sector, showing an agreement with the mean observed values.

Table 4.2: Summary of Collateral values for all sectors and across each sector

Sector	**All Sectors**	**Real Estate**	**Trade**	**Production**	**Service**	**Transport**
Min	-0.43	-0.37	-0.02	-0.03	0.00	-0.07
1st Qu.	0.75	0.15	0.96	0.81	0.44	0.47
Median	1.23	0.81	1.20	0.98	0.73	1.14
Mean	1.34	0.97	1.49	2.29	1.51	1.17
3rd Qu.	1.61	1.48	1.77	1.05	0.73	1.13
Max	20.14	19.84	52.08	60.72	26.13	11.42
CV (1000)	**0.76%**	**0.56%**	**0.34%**	**2.10%**	**0.44%**	**0.17%**

4.3.4 Capital

In the results most of the sectors are distributed towards the middle scale. But Real Estate and Production are quite heavy investments. This can also reduce the creditworthiness of a loan applicant.

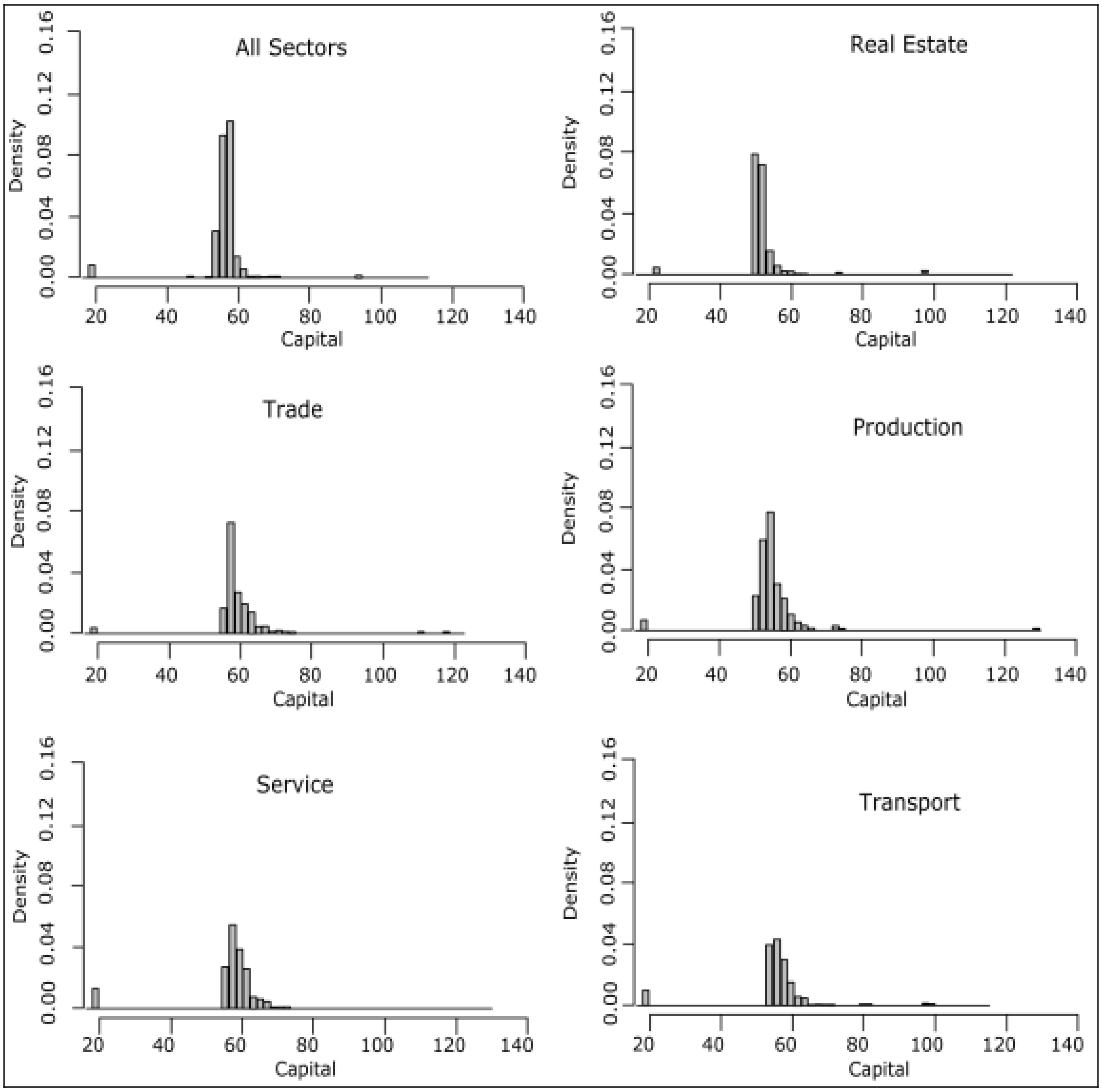

Figure 4.3: Histograms of Capital values for all Sectors and per Sector.

4.3.5 Condition

While most of the Condition factors quantified here are usually external to the customer, condition typically indicates a customer's vulnerability that can impact on credit risk. From the results, all sectors are equally affected by market prevailing conditions. Therefore issues such as inflation rates can badly affect the ability of customers from all sectors to pay back their loans.

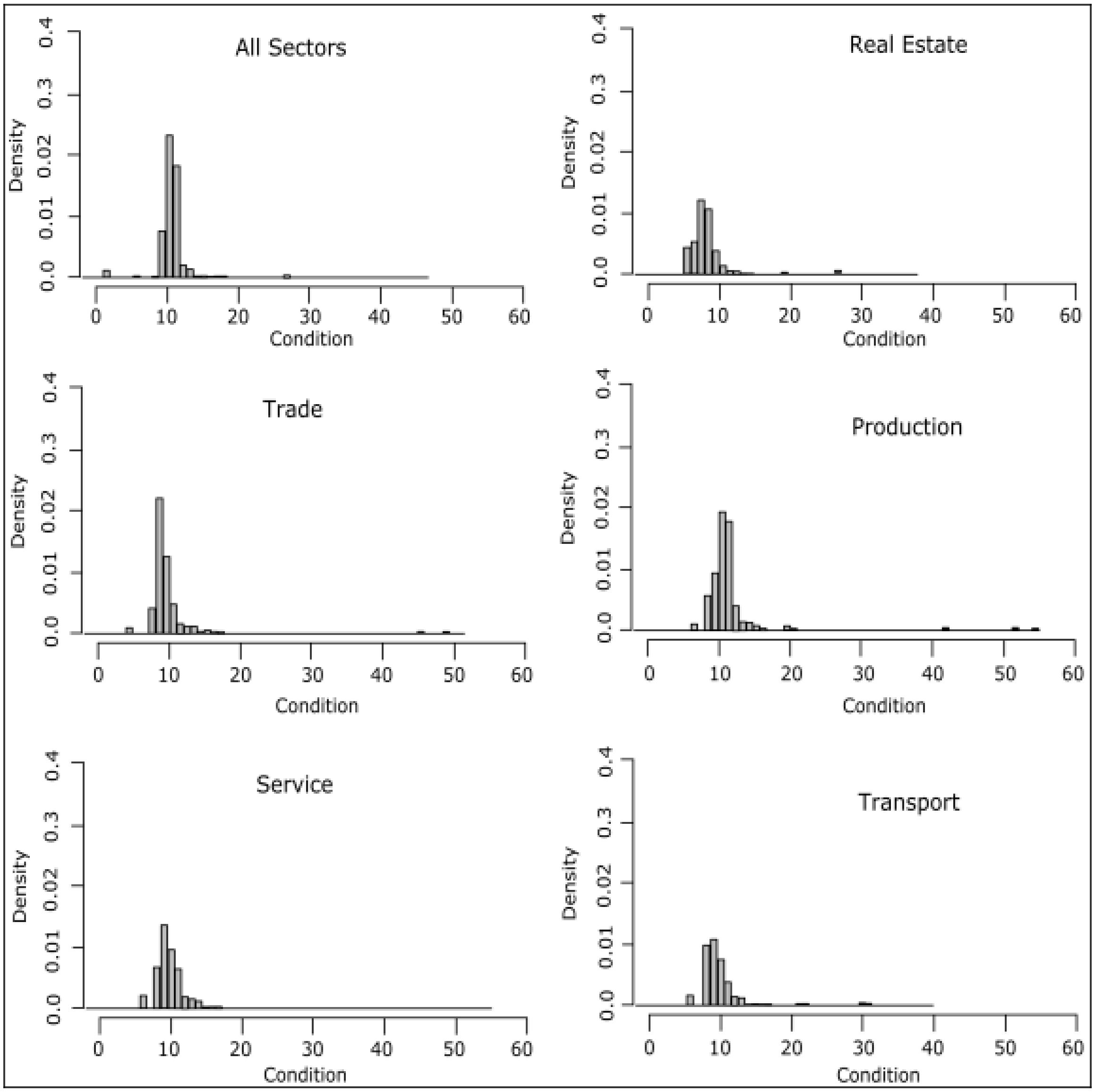

Figure 4.4: Histograms of Condition values for all Sectors and per Sector.

4.4 Predicting the Probability of Defaulting

The logistic regression model to predict the probability of defaulting used the formula outlined in Chapter 3. First, the model was tested backwards to test the ability of predicting defaults between 2014 and 2017. This test was carried out using an annual walk forward process that included all data of 223, 913 customers and 512 defaulting events. In order to make default predictions for a single year, only the information before that year was used to select model variables and calibrate the model coefficients.

Results of the probability of defaulting from the model showed a high level of accuracy in prediction of default risk (Figure 4.5). The model was effective in separating potential defaulting customers from non-defaulting ones over 1-year periods.

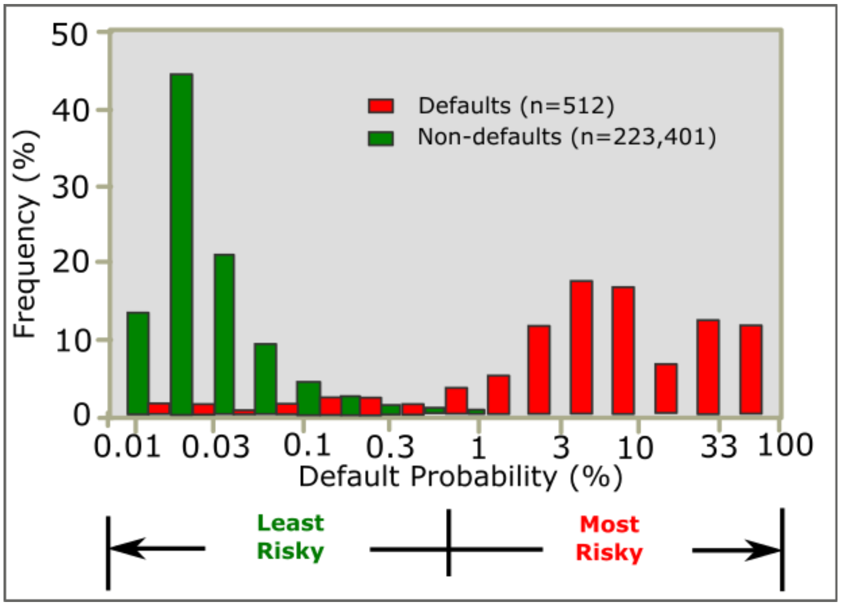

Figure 4.5: Distribution of 1-year probabilities of Defaults for defaulting and non-defaulting customers

As shown in Figure 4.6, loan applicants in the highest percentile included 97% of the defaults within the next year. The prediction power of the model decreased as the prediction window was extended over time, but still the model performance was good enough over the five years. The number of defaulters within 10% of the population declined from 95% to 81%, then to 68%, 54% and finally 40% at 2, 3 4 and 5 year horizons respectively.

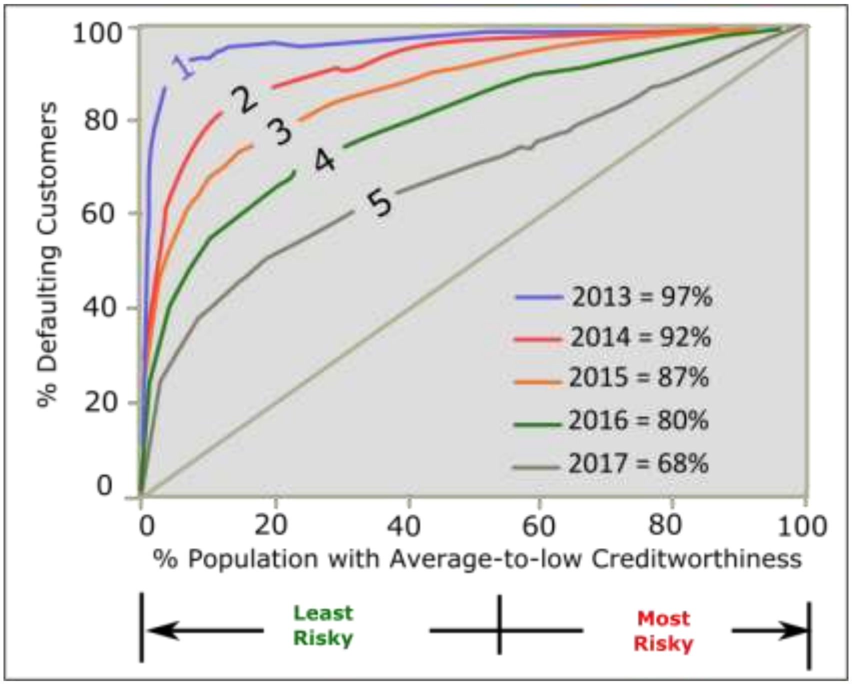

Figure 4.6: Cumulative accuracy profiles (CAP) curves for building the probability of Default model at 1-year horizon till 5-year horizon.

4.5 Logistic Regression Modeling

Table 4.1 shows results of the logistic regression with the *p*-values of significant variable shown in bold. The first most significant variable is Female with a *p*-value of 0.003. The odds ratio for the variable is 0.005 showing a very low probability of defaulting. The second most significant variable to the probability of defaulting was loan applicants of age 21 – 40 years. The odds ratio of this group is very high, showing that there is high likelihood. Risk control measure put in place by the bank were also significant, but not as significant as the personal characteristics of the loan applicants. Variables like financial information and credit terms were not significant at all in predicting the defaulting probability.

Table 4.3: Results of Logistic Regression model for Predicting the probability of Defaulting

Variables	Odds ratio	Confidence Interval (95%)	*p*-value
Age			
21 – 40	1.921	1.42 – 2.03	**0.021**
41 – 60	1.235	0.95 – 1.57	0.452
61 – 80	0.572	0.29 – 1.17	0.133
Gender			
Male	2.081	1.44 – 2.39	**0.047**
Female	0.005	0.29 – 0.51	**0.003**
Financial information	1.267	1.09 – 1.74	0.064
Credit terms	0.772	0.48 – 1.31	0.322
Risk control measures	1.654	1.06 – 2.08	**0.049**
-2 Log Likelihood		10.68	
Chi-square		4.32 ($p < 0.07$)	
Degrees of freedom		2,364	
Akaike Information Criteria (AIC)		16511	

Predictions from the regression model were compared to the actual defaults observed and plotted in Figure 4.7. The regression line showed in red shows a lot of accuracy in the prediction. This proves that the model is adequate for the prediction.

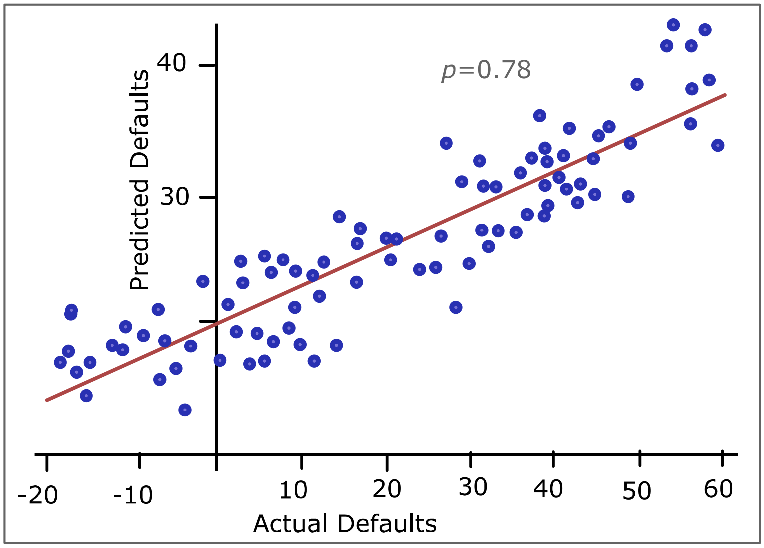

Figure 4.7: Scatterplot of observed defaults and predicted defaults from the logistic regression model

It was also necessary to validate the logistic regression model using only the validation dataset described in Chapter 3. The output of this validation is shown in Figure 4.8.

Results show a fairly accurate prediction power where the actual numbers shown in red are not very different from the predicted numbers shown in black.

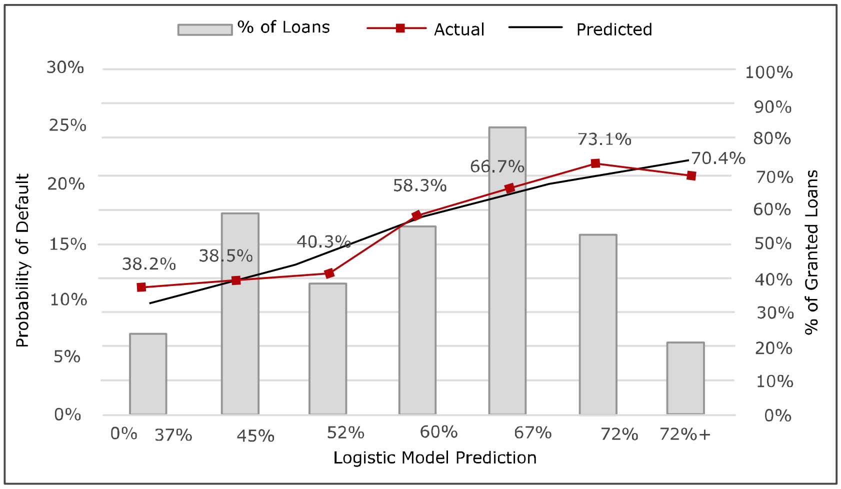

Figure 4.8: Validation of the logistic regression model using the validation dataset samples

4.6 Conclusion

Based on the results reported in this chapter, it can be concluded that the assessment of credit risk has been successful as enough variables have been analyzed. The model is very effective for identifying the probability of defaulting to advise the cooperative bank more accurately, and this model can be applied to other financial institutions to assess risk and determine the probability of defaulting. Additionally, a unique set of models has been constructed consisting of descriptive and probability ratings. These models can also be applied by both commercial bank and other financial institutions to rate customer creditworthiness.

CHAPTER FIVE

DISCUSSION OF RESULTS AND CONCLUSION

5.1 Introduction

This chapter provides a discussion of the results which have been presented in the previous chapter (Chapter 4) and compares these results with results on credit risk assessment that have been reported by other research studies. It provides a critical assessment of how the research objectives have been achieved. It also identifies how the existing research gaps have been filled and presents the value addition of this research project. Finally, the limitations and challenges of the project are highlighted, and recommendations are made for areas of future research.

5.2 Discussion of Results

In general, the results presented have shown that examining credit risk and the factors that play roles in reducing credit risk has great importance to banks and other lending institutions. Here the results agree with the empirical literature relating to credit scoring modeling. First, descriptive analysis determined that the Real Estate had the highest default rate, which has also been observed by other studies in Kenya e.g. Githua (2015) and Loyford and Moronge (2014). When assessing customers' creditworthiness using the 5Cs (character, capacity, collateral, capital and condition), the Real Estate and Transport sectors scored very low. On assessing Capital, Real Estate and Production had quite heavy investments, meaning if risk is not measured very well it can lead to a collapse of the economy. Results from logistic regression model to predict the probability of defaulting showed a high level of accuracy in prediction of default risk (Figure 4.5). The model was effective in separating potential defaulting customers from non-defaulting ones over 1-year periods. The logistic model was also examined using the Cumulative accuracy profiles (CAP) curves in order to evaluate the model's performance. The curve showed that the prediction power of the model decreased as the prediction window was extended over time, but still the model performance was good enough over the five years.

Secondly, the logistic regression model determined four independent variables which contributed to identifying the probability of default of Cooperative Bank personal loans. These variables were represented as the age group between 21 and 40 years, the gender (male, female) of the borrower, and risk control measures. The scores and *p*-value for measuring significance were used as a measure for selecting those variables. Furthermore, two measures were used for testing overall model fit. The first is the likelihood ratio for the change in the -2 likelihood value from the baseline model. The results indicated that the existence of a relationship between the independent variables and dependent variable is supported. The second measure is the Akaike Information Criteria (AIC). This measure showed the model fits the data well because the chi-square was not significant at the 0.05 level. Therefore, the measure showed that the null hypothesis was not rejected, and that the four variables which were significant can be used by financial institutions to determine credit risk. This also means that the model fits the data well and the logistic model has a high sensitivity to the variables selected.

Moreover, the logistic model showed high classification accuracy for the analysis sample data. Measures were used for evaluating achieved classification accuracy and showed that the levels of classification accuracy were acceptable. In addition, AIC, likelihood ratio, and change in log likelihood measures were used for testing coefficients. All measures showed that the entered variables are significant. Therefore, it can be concluded that the objectives have been met and the developed model is fit for the definition of good and bad loans.

5.3 Study Contributions and policy implications

This project work has made significant contributions to the analysis of credit risk in banks and the credit literature. First, many studies have focused on corporate loans and ignored the risk in personal loans, where according to CBK (2017) default rates are largely contributing to the deteriorating economy. Secondly, no studies have done a proper assessment that applies both descriptive analysis and credit risk assessment modeling for personal loans in the Kenyan context. Some studies have applied the 5Cs, such as Moti et al. (2012), but this assessment is not sufficient for determining good vs. bad loans. This study has however combined assessment with 5Cs with regression modeling that included both personal and socioeconomic variables in better assessment of risk. Furthermore, because of the lack of extensive research on credit risk modelling in Kenya, this project work attempts to fill this gap by discovering a neglected field in previous studies. Due to difficulties found by Kenyan banks and other lending institutions in identifying credit risk of their clients resulting in the scarcity of the implement such these models, this thesis helps banks and other financial institutions to more clearly define their credit risks by applying new credit risk models. Therefore, this thesis contributes to the literature of credit scoring models in general and for Kenya particular.

Finally, this thesis constructed an advanced assessment model in the credit risk context and this model resulted in variables which help to identify the credit risk in Kenyan lending institutions. Furthermore, the model was tested in terms of its assumptions. This ensures that variables selected in the model represent the best selection of this thesis and support the results of this thesis. In addition, this thesis provided comparison between the constructed credit assessment model for Cooperative Bank in order to select the best technique depending on their performance. Therefore, it can be said that this research makes a new contribution to credit assessment models in respect of their performance in Kenyan banks and new independent variables that have not been examined in the literature of credit assessment models.

The spread of credit scoring, particularly its growing use in consumer loans, should lead to increased competition among banks and increased availability of credit for consumers. Traditionally, banks usually lend to consumers from a branch in the borrower’s area. This branch provides the lender good information of the area, which is thought to be useful in the credit decision. Consumers are likely to have current accounts at the bank, and the information the bank can obtain by examining the financial ability of borrowers can provide the bank an advantage in lending to these consumers. However, credit scoring is changing the way banks make consumers loans by processing applications using automated and centralized systems. These banks can provide large volumes of consumer loans even in areas where they do not have extensive branch networks. Applications are accepted over several communications including the phone and mail, as credit card lenders do.

Credit scoring may also support more lending because it provides banks a tool for more accurately pricing risk. The price can be adjusted according to the risk. However, the relationship may be different according the size of the bank. The typical bank-borrower relationship, which is built up over years of lending, allows for considerable flexibility in loan terms. A long-term relationship allows the bank to offer special prices to a borrower facing temporary credit problems, which the bank can later make up for when the borrower overcomes the credit problems.

Credit assessment models can be used to improve credit practices for banks. Banks usually focus on corporate loans, but consumer loans are rapidly growing. These loans are usually represented in small amount with large borrowers. Therefore, banks cannot control the quality across their branches and avoid making mistakes. With credit assessment, the bank applies an equation in order to obtain a numeric quantification of the risk. This can change the required information and credit assessment. For example, the bank will use several Credit Risk Modelling in Kenyan practices for credit scoring models resulted from the information available about the applicants. Some banks may use cut-off point to approve or disapprove.

5.4 Limitations and Recommendations for Further Research

This section examines limitations of the current research and considers the impact they have on the research conclusion. It also provides some guidance for future research that could be perceived as possible extensions to the present research.

There are several intrinsic limitations related to research of this nature. Firstly, this research selected Cooperative banks in Kenya. Therefore, research results are limited to Kenyan banks. This also means the results may differ if the research is conducted in other country in another region. Therefore, the findings of this research are limited generalizations. Thus, by applying credit scoring models to different Kenyan banks or elsewhere and comparing the results would provide better scoring models for personal loans. Moreover, comparison studies of Kenyan credit scoring models and those in other developing counties and models in developed countries could contribute to an improvement in these models.

Secondly, this research used cross-sectional data in the credit risk models. Cross-sectional research design may be attractive for their advantages of saving time and cost, however the design strictly limits the researcher's capacity to address changing or development issues or to recommend fundamental interpretations. In addition, the research used loans granted only over 4 years to build credit scoring models. It may include some characteristics that differ from other years particularly as Kenya features in a transitional economy. Consequently, it can be suggested for further study that time series data could provide better results. For example, the selection of a sample across 10 or 20 years instead of four years or using cross-sectional time series data could consider whether time plays an important role in building models. Moreover, the research would be to expand the model to include other variables. The variables included in this research model were those found in the literature and in those that resulted for collecting data.

Finally, this research focused only on three credit assessment techniques; those from descriptive analysis, assessment of the 5Cs and logistic analysis. Other credit assessment models were not applied because of the limitation of the research period and the requirements of other models. Accordingly, it is suggested that other credit scoring models, including non-parametric models, could provide another contribution to the Kenyan context.

REFERENCES

Wu, X. (2008). *Credit Scoring Model Validation*. Master's thesis, University of Amsterdam, Korteweg-de Vries Institute for Mathematics.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989), "User acceptance of computer technology: A comparison of two theoretical models", *Management Science, 35*, 982–1003.

Githua, A. N. (2015). The Role of Development Financial Institutions in Real Estate in Kenya: A Case of Shelter Afrique (Doctoral dissertation, United States International University-Africa).

Loyford, M. M., & Moronge, M. (2014). Effects of economic factors on performance of real estate in Kenya. European Journal of Business Management, 1(11), 181-200.

Thun, C. (2011). Credit Risk Models Buy vs. Build. *Moody's Analytics*.

West, D. (2010). Neural Network Credit Scoring Models. *Computers & Operations Research, 27* (11-12): 1131-1152.

Chuang, C., & Lin, R. (2009). Constructing a reassigning credit scoring model. *Expert Systems with Applications 36* (2/1), 1685-1694.

Abdou, H. & Pointon, J. (2011). 'Credit scoring, statistical techniques and evaluation criteria: a review of the literature', *Intelligent Systems in Accounting, Finance & Management*, 18 (2-3), 59-88.

Auronen, L. (2003). Asymmetric Information: Theory and Applications. *Paper presented in the Seminar of Strategy and International Business as Helsinki University of Technology*, May 21st 2003.

Bernard, R. (2013). *Research methods in anthropology: Qualitative and quantitative approaches (4th edition)*. Altamira Press, Toronto Canada.

Bofondi, M., & Gobbi, G. (2003). *Bad Loans and Entry in Local Credit Markets*. Bank of Italy Research Department, Rome.

Central Bank of Kenya (2014). Bank Supervision Annual Report.

Chen, M., & Huang, S. (2013). Credit Scoring and Rejected Instances Reassigning Through Evolutionary Computation Techniques. *Expert Systems with Applications, 24*, 433-441.

Central Bank of Kenya (CBK), (2016). *The Kenya Financial Sector Stability Report, 2015*. August 2016, Issue No. 7

Central Bank of Kenya (CBK), (2017). *Credit Survey January-March 2017. Annual Report & financial statements 2017.* Available from https://www.centralbank.go.ke/uploads/banking_sector_reports/623284779_Credit%20Survey%20Report%20for%20the%20Quarter%20ended%20March%202017.pdf.

Chijoriga, M. M. (2011). Application of multiple discriminant analysis (MDA) as a credit scoring and risk assessment model. *International Journal of Emerging Markets, 6* (2), 132-147.

Cocheo, S. (2009). Latest Fair-Lending Case Contains Lessons For All Banks, and a $3 Million Bill For One, A.B.A. BANKING J., 7.

Cooper, D. R., & Schindler, P. S. (2011). *Business Research Methods*. (11th ed). New York: McGraw Hill International Edition.

Creswell, J. W. (2014). Research Design: Qualitative, Quantitative and Mixed Methods Approaches (4th ed.). London: Sage Publications Ltd.

Dastoori, M., & Mansouri, S. (2013). Credit Scoring Model for Iranian Banking Customers and Forecasting Creditworthiness of Borrowers. *International Business Research, 6* (10), 25-39

Dobbie, W., & Skiba, P. (2012). Information asymmetries in consumer credit markets: Evidence from payday lending. *Working paper, National Bureau of Economic Research*.

Einav, L., Jenkins, M. & Levin J. (2013). The impact of credit scoring on consumer lending. *RAND J. Econom. 44*(2), 249–274.

Hanna, N. K., and Boyson, S. (2013). *Information Technology in World Bank Lending: Increasing the Developmental Impact.* Washington, DC: World Bank.

Kombo, D. K., & Tromp, D. L. A. (2009). *Proposal and Thesis Writing: An Introduction.* Paulines Publications Africa, Don Bosco Printing Press, Nairobi Kenya.

Lagat, F. K., Mugo, R., & Otuya, R. (2013). *Effect of credit risk management practices on lending portfolio among savings and credit cooperatives in Kenya*.

Moti, H. O., Masinde, J. S., Mugenda, N. G. & Sindani, M. N. (2012). Effectiveness of Credit Management System on Loan Performance: Empirical Evidence from Micro Finance Sector in Kenya. *International Journal of Business, Humanities and Technology*. 2 (6), 99-108.

Mugenda, A.G. (2008). *Social Science Research: Theory and Principles*. Acts Press, Nairobi.

Ong, M. K. (Ed.). (2007). *The Basel Handbook: A guide for financial practitioners*. Risk books.

Pagano, M. & Jappelli, T. (1993). Information Sharing in Credit Markets. *The Journal of Finance 43*(5), 1693-1718.

Richardson, C. J. (2009). Credit Scoring of the Future, COLLECnONS & CREDrr RIsK, April, at 19.

Samreen, A., & Zaidi, F. B. (2012). Design and Development of Credit Scoring Model for the Commercial banks of Pakistan: Forecasting Creditworthiness of Individual Borrowers. *International Journal of Business and Social Science, 3*(17), 2219– 1933.

Schreiner, M. (2010). Credit Scoring for Microfinance: Can It Work? *Journal of Microfinance Risk Management, 2* (2), 105-118.

Einarsson, A. I. (2008). Credit risk modeling (Master's thesis, Technical University of Denmark, DTU, DK-2800 Kgs. Lyngby, Denmark).

Anolli, M., Beccalli, E., & Giordani, T. (Eds.). (2013). Retail credit risk management. Palgrave Macmillan.

Frydman, H., Edward I. Altman, & Duen-Li Kao (1985) Introducing recursive partitioning for financial classification: The case of financial distress. The Journal of Finance, 40(1):269–291, ISSN 00221082. URL http://www.jstor.org/stable/2328060.

Yu, L, Wang, S. & Lai, K.K. (2007) Credit risk assessment with a multistage neural network ensemble learning approach. Expert Systemswith Applications

Hastie, T., Tibshirani, R. & Friedman, J. (2009) The Elements of Statistical Learning: Data mining, Inference and Prediction. Berlin, Springer

Altman, E.I., Marco, G. & Varetto, F. (2004) Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the italian experience). Journal of Banking and Finance, pages 505–529.

Huang, C. L., Chen, M. C., & Wang, C. J. (2007). Credit scoring with a data mining approach based on support vector machines. Expert systems with applications, 33(4), 847-856.

Yap, B. W., Ong, S. H., & Husain, N. H. M. (2011). Using data mining to improve assessment of credit worthiness via credit scoring models. Expert Systems with Applications, 38(10), 13274-13283.

Bravo, E. R., Talavera, A. G., & Serra, M. R. (2017). Prediction and Explanation in Credit Scoring Problems: A Comparison between Artificial Neural Networks and the Logit Model. In Data Analytics Applications in Latin America and Emerging Economies (pp. 153-170). Auerbach Publications

APPENDICES

Appendix I: Budget

Budget Items	**Cost (Kenya Shillings)**
Proposal development- Printing, stationery, internet costs	12,000
Data collection a) Research assistant fees	15,000
b) Stationery and printing	5,000
Data analysis and report a) Printing and stationery	5,000
b) Hard cover Binding	6,000
Transport Fuel to Campus, airtime	8, 000
Miscellaneous	5,000
TOTAL BUDGET	**KES 56,000**

Appendix II: Work Plan

Timelines / Activity	Jan - April 2017	April- May 2018	June- July 2018	July 2018	August – Sept. 2018	Sept- Oct
Proposal Development (Done)						
Development Research Instruments (Done)						
Addressing Feedback/ Supervisor Comments						
Proposal Defense						
Addressing Panel Feedback/ Corrections						
Pilot Testing						
Field Data Collection						
Data analysis and Report Writing						
Addressing supervisor Feedback						
Project Defense						
Addressing Feedback from panel						
Final Report Submission						

Appendix III: Sample Source Code for Regression Analysis

```
library(ggplot2)
library(RColorBrewer)
library(gridExtra)
library(png)
library(reshape2)
setwd("C:/Areba")
df <- read.csv("LoansData.csv", header=TRUE)
dir.create(file.path("output"), showWarnings = FALSE)
png(filename="output/LoansData.png", height=750, width=1000,
    bg="white", res=300)
dat <- melt(df, id="DefaultingYear")
ggplot(dat, aes(x=time, y=value, fill=value),stat = "identity",
position = "stack",alpha=.9)+

scale_fill_brewer(palette="Paired",breaks = sort(levels(dat$variable)))
+ geom_density(stat="identity")
dev.off()

dir.create(file.path("output"), showWarnings = FALSE)
png(filename="output/DefaultProbability.png", height=1000, width=1000,
res=300)
dev.off()

ggplot(dat, aes(x=Time, y=value)) +
  geom_area(aes(fill=variable, colour=variable),position='stack')
dir.create(file.path("output"), showWarnings = FALSE)
png(filename="output/DefaultProbability1.png", height=1000, width=1000,
res=300)
dev.off()

dat <- melt(df, id="DefaultingYear")
ggplot(dat, aes(x=Time))+ geom_density(aes(y=value, fill=variable,
colour=variable))
dir.create(file.path("output"), showWarnings = FALSE)
png(filename="output/DefaultProbability2.png", height=1000, width=1000,
res=300)
dev.off()

# Creating the Cumulative accuracy profiles (CAP) curves
defaultFile <- data.frame("Score"=runif(100,1,100),
                 "hasDefaulted"=round(runif(100,0,1),0))

# Ordering the dataset
defaultFile <- defaultFile[order(defaultFile$Score),]

# Creating the cumulative density
defaultFile$cumden <-
cumsum(defaultFile$hasDefaulted)/sum(defaultFile$hasDefaulted)

# Creating the % of defaulting customers
defaultFile$perpop <- (seq(nrow(defaultFile))/nrow(defaultFile))*100
```

```
# Ploting
plot(defaultFile$perpop,defaultFile$cumden,type="l",xlab="% Population
with Average-to-low Creditworthiness",ylab="% of Defaulting Customers")

# Creating the regression model

dfdata <- read.csv("Loans.csv", header=TRUE)
summary(dfdata)
par(mfrow=c(1,8))
for(i in 1:8) {
    hist(dfdata[,i], main=names(dfdata)[i])
}
glm.fit <- glm(Default ~ Age1 + Age2 + Age3 + Male + Female +
FinancialInfo+ CreditTerms+ RiskMeasures, data = dfdata, family =
binomial)
summary(glm.fit)

xweight <- seq(-20, 60, 10)
yweight <- predict(model_weight, list(wt = xweight),type="response")
plot(dfdata$actual, dfdata$predict, pch = 16, color = "blue", xlab =
"Actual Defaults (g)", ylab = "Predicted Defaults")
lines(xweight, yweight)
```