Journal of Information and Technology

Prototyping a Credit Scoring Model for Micro Finance Institution in Kenya: A Case of Kenya Women Fund Trust (KWFT) Bank

Kibet Kipkemoi & Prof. Peter W. Wagacha

Stratford Peer Reviewed Journals and Book Publishing
Journal of Information and Technology
Volume 1||Issue 1||Page 12- 27||January||2018|
Email: stratfordjournals.org

Prototyping a Credit Scoring Model for Micro Finance Institutions in Kenya: A Case of Kenya Women Fund Trust (KWFT) Bank

[1*] Kibet Kipkemoi, [2]Prof. Peter W. Wagacha

[1]Postgraduate Student, University of Nairobi

[2] Senior Lecturer, University of Nairobi

***E-mail of corresponding author: kkibet@gmail.com**

***How to cite this article*: Kibet K. & Wagacha W., P. (2018), Prototyping a Credit Scoring Model for Micro Finance Institutions in Kenya: A Case of KWFT Bank, *Journal of Information and Technology* Vol 1(1) pp. 12-27.**

Abstract

Credit scoring models are measurable examination utilized by credit authorities to assess your value to get credit. It breaks down the attributes and characteristics of past loans to foresee the performance of future loans. The KWFT client fits the demographic profile of a woman of the age between 25 to 60 years old. Given that for the KWFT client base, much of their financial information is limited and therefore judgmental and group security approach is mainly used. The study sought to prototype a behavioral credit scoring model that predicts the probability of defaulting on loan payments. The system for building the credit scoring models included the accompanying procedure which entailed 'performers' and 'defaulters'. The target population was 800,000 clients spread across 230 branches. The study used data of 20 best credit performing clients and 20 poor performing clients in all branches in Nairobi County and environs. A total of 480 clients were therefore sampled. Modeling techniques used was logistic regression. The study involved the delivery of a software based prototype. Findings indicated that the odds of defaulting was 1.497 times greater for males as opposed to females and the relation was significant, all classes of age were not significant determinant of performance rating. The study established that individual unmarried was highly likely to default as opposed to individual married, the relation however is insignificant. Group membership was found to greatly predict the rate of default. The study concluded that defaulting was based on information, and in microfinance, this information was usually qualitative and informal and resides with group members or with loan officers. It was recommended that senior management should see the strategic value in developing, implementing, and using scorecards as an integral approach to managing risk within microfinance.

Keywords: *Credit scoring, Prototype, Defaulting, performing and Microfinance*

1.0 Introduction

1.1 Background of the Study

Credit scoring models are measurable examination utilized by credit authorities to assess your value to get credit. Scoring counts depend on installment record, recurrence of installments, measure of obligations, credit charge-offs and measure of Mastercards held. A specific weight is appointed to each calculate considered the model's equation and a FICO rating is allocated in view of the assessment (Littwin, 2007). Application credit scoring is utilized all through the world to prepare many sorts of little esteem loan exchanges. It has been connected most broadly and effectively for individual charge cards and shopper and home loan advances. Reimbursement chance for these items is firmly connected to evident elements, for example, salary, credit agency data, and statistic elements, for example, age, training, and property holder status.

There is an existing lack of reliable credit scores between the lenders. The loan lenders can help come up with reliable credit scores to help normalize the affairs of advances to the borrowers. This would also lay a credible platform where the borrowers can assess the trust levels of the lenders. Obviously, the scores are expected to vary from one lender to another which would enable the borrowers choose the most reliable. The information in credit reports varies among banks also. Whereas the CBK regulations require a minimum of monthly submission of the CRB information, banks may also be update at different leading to inconsistency in credit scoring (Diana, 2005).

KWFT is the largest microfinance institution in Kenya with a client base of over 800,000 clients. The institution has a dedicated workforce of over 2,700 staff who meet clients monthly in their areas of business operations. KWFT has mainly focused on the low income earners thus boosting their capacity to do business or invest. In addition to the branches, KWFT utilizes technology in services such as Mobile Banking, Agency Banking and ATM to help in ensuring that KWFT has an even deeper penetration into remote, rural and urban areas. The Bank utilizes its credit officers to facilitate training in the use of these technologies as well as show the customers how they can leverage on other payment solutions such as Mpesa to enhance their businesses.

The KWFT client fits the demographic profile of a woman of the age between 25 to 60 years old. She is mainly located in rural and urban areas and engages in low income economic activities that mainly involve trade, of which a majority is in retail and wholesale, as well as agriculture. The institution has over 100 deposit taking branches and over 230 field offices across the country. In relation to payment systems, KWFT has over 75 ATMs across the country, 1700 Agents and offers Mobile Banking Services. Financial record keeping is rudimentary in nature through use of manual records such as cash books and sales records. The client has in most recent years started utilizing emerging technologies in financial management and transaction processing such as Mpesa.

1.2 Statement of the Problem

Most financial movement on the planet originates from the little and medium-sized undertaking (SME) division with over half been kept running by ladies (Wendel & Harvey, 2006). In creating nations, SMEs have restricted access to formal credit. In sub-Saharan Africa, for instance, keeping money division infiltration is about one percent of the populace (Stein, 2001). In Kenya, there more than 2.2 million miniaturized scale, little and medium undertakings (Strategic Business Advisors (Africa) Ltd. SME Banking Sector Report, 2007), of which 88 percent are non-enlisted. Of this non-enlisted gathering, just 23 percent have financial balances, and just 10 percent have ever gotten credit from any formal source. KWFT has more than 80% of its customers situated in rustic regions where farming is the principle financial action.

Their size and credit request of Kenyans have outgrown the limit of microfinance foundations, which offer little, short loans for the most part through gathering loaning systems, while the limit of the SME chance profile joined with the moneylenders' absence of refined hazard evaluation strategies makes a large number of them seem undesirable as credit clients for business managing an account. These reasons in this way push up the cost of credit and furthermore, numerous Kenyans are additionally hesitant to look for credit. Credit score models rely mainly on creditor's financial information history. Given that for the KWFT client base, much of their financial information is limited and therefore judgmental and group security approach is mainly used. Given this background, this study sought to come up with a credit score model that is practical to the Kenyan microfinance banking sector and more so to address KWFT operational challenges on credit assessment.

1.3 Research Objectives

i. To develop a credit scoring model that can be used in a Microfinance institution in Kenya.
ii. To develop a prototype that uses the model to predict probability of defaulting.
iii. Test and validate the prototype.

1.4 Research Questions

i. Can model building achieve the unique Kenya microfinance credit assessment needs?
ii. Which variables determine the probability of defaulting?
iii. Can the prototype model be automated to reduce defaulting behavior?
iv. What value addition will the model contribute to credit risk cost of microfinance institutions?

2.0 Literature Review

2.1 Theoretical Review

Credit Scoring Model

Credit score model is a statistical technique that combines several financial characteristics to form a single score to represent a borrower's credit worthiness. In the view of Lahsasna *et al* (2010), credit scoring is defined as the capture of the relationship between historical information and future credit risk that is mathematically definable as:

Y, =f (X I, X2... Xm).

Where XI, X2..., Xm are each customer's attributes, while Y, is the denotation of the customer's resultant defaulting status (defaulted or not defaulted). The prediction of Y is the task of credit scoring. The target of credit scoring is to build models which rank credit customers based on their future credit risk.

There are essentially two distinct types of scoring models that can be approved measurably. These models will either utilize a factual or judgmental scoring assessment. For each situation, the end FICO rating result can shift too. A measurable scoring model uses numerous elements from one or various credit revealing organizations, corresponds them and afterward allocates weights to each element. Scoring counts depend on installment record, recurrence of installments, measure of obligations, credit charge-offs and measure of Visas held. A specific weight is allotted to each figure considered the model's equation, and a financial assessment is relegated in view of the assessment (Thomas, 2000).

2.1.1 Vantage Score Model

The Vantage Score model was presented in 2006 when the three noteworthy credit detailing authorities Experian, Equifax and TransUnion, chosen to offer FICO some opposition in the FICO assessment business. The Vantage Score demonstrate takes a gander at recognizable information, things like paying on time; keeping Mastercard adjusts low, dodging new credit commitments, ledgers and different advantages for figure its score. The real contrast in this model is that it needs just a single month of record as a consumer to build up a score, rather than the six months required for FICO and different models (Cundiff, 2004).

Different elements extraordinary to Vantage Score incorporate overlooking accumulations paid or unpaid under $250 and help for records adversely influenced by catastrophic events. The Vantage Score scoring scale is the same as FICO's 300–850; however it incorporates a letter review (A through F) to help better comprehend the score. The Vantage Score utilizes data from three credit revealing companies, Experian, Equifax and TransUnion, however measures certain elements more vigorously or less intensely than the FICO calculation. Among the elements considered are: liquidations, accumulations, missed installments and dispossessions recorded using a loan report, occupation and time at present place of employment, owning or leasing living arrangement, measure of time at current area, the quantity of investigation into credit over a timeframe, the parities of utilized credit to accessible credit, age, the length of credit report and the period of time financial record has been in the agency's database (Caire & Kossman, 2003).

2.1.2 Linear Probability and Logit models

This model uses past data such as accounting ratios into a model to explain repayment experience on old loans. This is then used to forecast probabilities on new loans (Abedi, 2000).

2.1.3 Risk Adjusted on Return on Capital (RAROC).

This model measures how much risk the lender is taking. It is calculated by evaluating the expected return against the value at risk. It helps to determine if returns are providing adequate compensation for risk and assesses if the bank is providing shareholders with an increase in value through participation in the business (Abedi, 2000).

2.1.4 Option Pricing Theory Models

This method starts with the observation that a borrower's limited liability is comparable to a put option written on the borrowers' assets. The strike price is usually equal to the value of the debt outstanding. If in some future period, the value of the borrower's assets falls below the value of the outstanding debt, the borrower is likely to default. The probability of default here is inferred from an estimate of the firms' asset-price volatility based on the observed volatility of firms' equity prices (Abedi, 2000).

2.2 Empirical Review

Giné, Jakiela, Karlan, and Morduch (2010) found, based on laboratory-style experiments in a Peruvian market, that contrary to much of the theoretical literature, joint obligation invigorates chance taking, at any rate when borrowers know the venture procedures of co-borrowers. At the point when borrowers could self-choose into gatherings there was a solid negative impact on hazard taking because of assortative coordinating. Fischer (2010) attempts comparative lab style tests and furthermore finds that under restricted data, aggregate obligation animates hazard taking as borrowers' free-ride on the protection given by co-borrowers. In any case, when co-borrowers need to give forthright endorsement for each other's undertakings, ex risk moral peril is moderated. Giné and Karlan (2010) look at the effect of joint obligation on reimbursement rates through two randomized tests in the Philippines. They find that evacuating bunch obligation, or presenting singular risk starting with no outside help, did not influence reimbursement rates over the following three years. In a related review, Carpena, Vole, Shapiro and Zia (2010) misuse a semi analyze in which an Indian MFI changed from individual to joint-obligation gets, the invert of the switch in Giné and Karlan (2010). They locate that joint risk essentially enhances credit reimbursement rates.

Morduch (1998) and Morduch and Roodman (2009) duplicate the Bangladeshi reviews and discover no confirmation of a causal effect of microcredit on utilization. Kaboski and Townsend (2005) likewise utilize non-exploratory information and archive a positive effect of joint risk microcredit on utilization however not on interests in Thailand. In light of a basic approach the creators certify this finding in Kaboski and Townsend (2011).

With advances in innovation, more smart credit scoring models are being created. Subsequently, charge card guarantors can make utilization of the data created from the models to define better gathering procedures and henceforth utilize their assets all the more viably (Cundiff, 2004). Lucas (2000), for instance, had discovered recuperation rates to normal 15.9% in 1999, up from 12.1% in the earlier year and 9.1% in 1997.

The credit scoring applied in developed countries differs from current practices in Kenya due to financial data at the individual level being limited. Most applications are done on paper at the client's location and not with real-time data entry to the bank's systems. The bank's loan officers then manually evaluate each application based on thorough knowledge of the product, region and industry, which tends to yield a good decision. However, this method is less scalable and depends on lengthy training.

There are very few sophisticated data models in use by banks and microfinance institutions in Kenya. Developed countries use 'instant' credit decisions that are based on systems requesting a bureau score in real-time, followed by proprietary models calculating an application score: approve / decline or red / yellow / green. The systems and models require large amounts of data, and in Kenya, organized and collated data is sparse. Sometimes, key application metrics never make it from paper into the MIS thereby limiting all down-stream modeling capabilities. There is a need to have systematic decisions that exclude inputs that might cause discrimination.

In Kenya, the microcredit's strongest risk management practice is a requirement that each borrower be part of a group. This creates a community that both spreads the risk and invokes peer pressure to help with repayments. These groups are the cornerstone of successful microfinance, and directly mitigate default risk.

2.3. Conceptual Model

Calculated model captures the procedure of the development and combination of credit scoring models, an illustrative calculation is displayed in Figure 1. The process begins with the development of individual models. This stage involves the accompanying five stages: (1)

characterize objective, (2) select factors, (3) select specimen and gather information, (4) select demonstrating apparatuses and develop models, and (5) approve and survey models. The goal is to build up a credit scoring model to anticipate the credit danger of advance candidates as terrible or great hazard. To develop the model, 20 traits containing statistic attributes and credit subtle elements are utilized.

In the event that the individual models are to be joined, the second stage is then locked in. Five stages are utilized to consolidate the models: (1) select models for mix, (2) select displaying strategies to join models, (3) join models, (4) approve and evaluate consolidated models, and (5) select best joined model for organization.

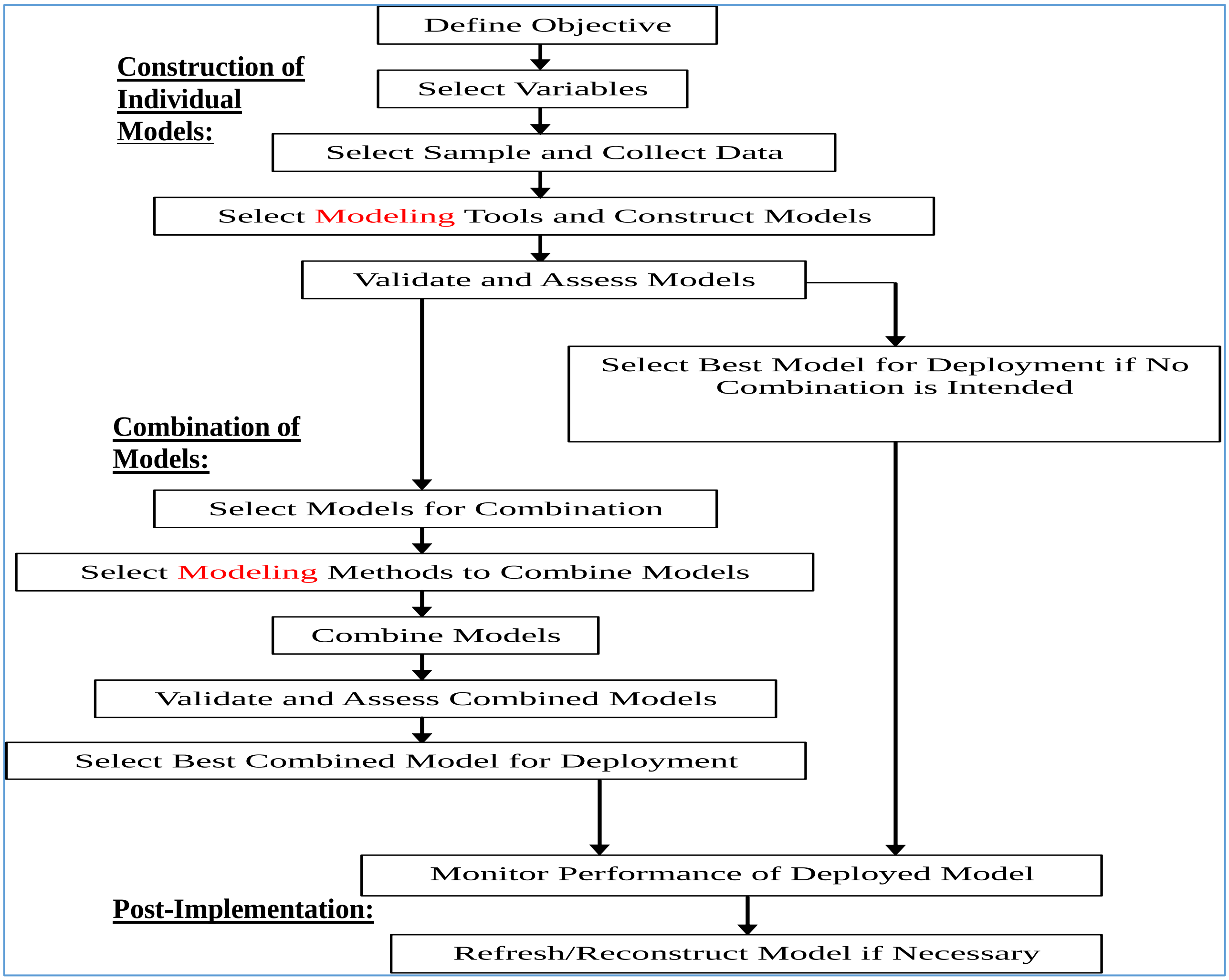

Figure 1: Conceptual Model (Adopted from Blake and Merz, 1998)

3.0 Research Methodology

The Sample, Explore, Modify, Model, Assess (SEMMA) methodology was used in the modeling process but with extensions for it to support iterations at each of the stages. The extensions of the model led to the proposed Sample, Explore, Transform, Re-categorize, Model and Assess

(SETRMA) methodology that provides an iterative modeling methodology but with finer implementation steps between two main steps. The system for building the credit scoring models included the accompanying procedure which entailed 'performers' and 'defaulters'. The target population was 800,000 clients spread across 230 branches. The study used data of 20 best credit performing clients and 20 poor performing clients in all branches in Nairobi County and environs. A total of 480 clients were therefore sampled. Modeling techniques considered included discriminant analysis, logistic regression and maximum likelihood estimation, however logistic regression was chosen. The two-tier architecture of a Staging area and a Data Mart area for the credit scoring process was adopted in the data extraction process. The Crammer V greedy algorithm, a post test to determine the strengths of significance between the variables, was used to guide the selection of variables based on their correlation to the dependent variable and inter-variable correlations. The study involved the delivery of a software based prototype.

4.0 Results and Findings

4.1 Description of the Basic Dataset

The Basic Dataset in this study refers to the raw dataset that was sourced from KWFT for this study. It therefore excludes any such variables as are product of any transformations.

4.1.1 Gender

Gender findings were analyzed from the data set. The results were as shown in figure 2. Majority from the analyzed dataset were female with a total of 76%. Men represented 24%.

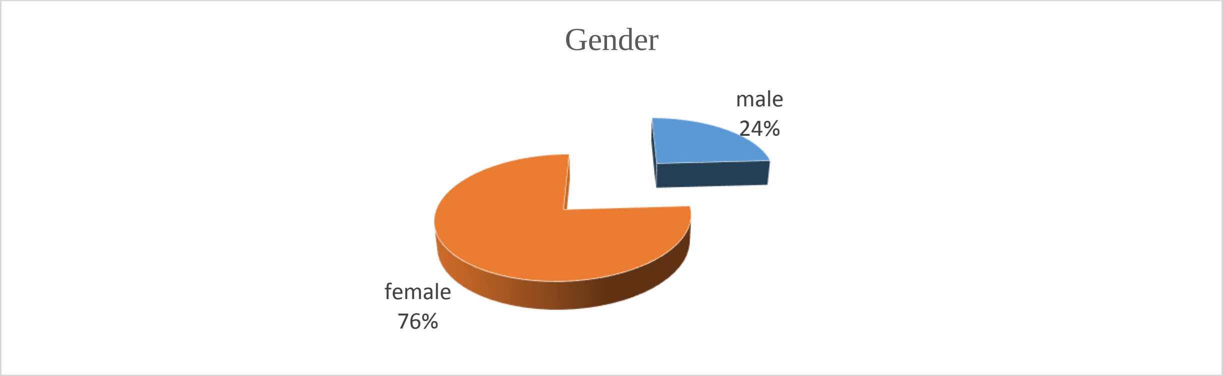

Figure 2: Gender

4.1.2 Annual Level of Income

From the dataset, data on annual income was captured and presented in table 1. Majority, represented by 43% earned and annual income of between 500,000 and 1 million, 28% earned annual income of between 200,000 and 500,000, 21% earned and income of less than 200,000, the least, represented by 8% earned annual income of over 1 million.

Table 1: Annual Level of Income

Income level	Percentage
less than 200,000	21%
200,000-500,000	28%
500,001-1 Million	43%
Over 1 million	8%

Stratford Peer Reviewed Journals and Book Publishing
Journal of Information and Technology
Volume 1||Issue 1||Page 12- 27||January||2018|
Email: stratfordjournals.org

4.1.3 Marital Status

Information on marital status was captured from the dataset in WFT databank. The findings were analyzed as shown in figure 3. 55% were married while 45% were unmarried.

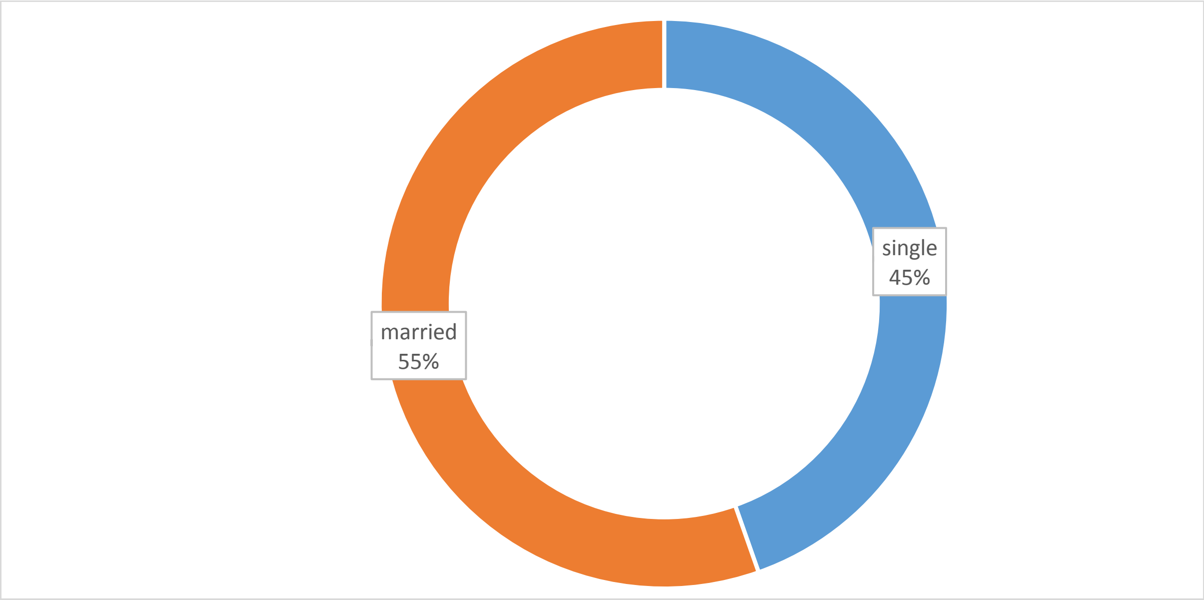

Figure 3: Marital Status

4.1.4 Group Membership Status

The observation membership to a grouped was analyzed and presented in the figure 4. Majority of the observation were group members, 260 belonged to a group while 130 were no associated with any membership group.

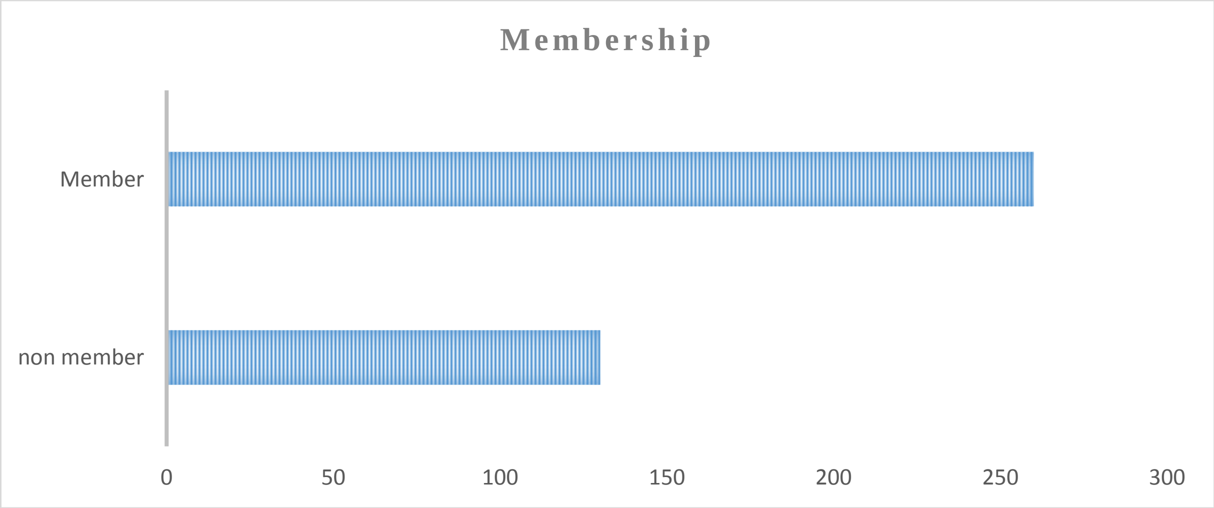

Figure 4: Group Membership Status

4.1.5 Group Membership Duration

For observation who associated with group membership, the duration of being members was checked for analysis. The results were as presented in figure 5. The study found out that 46% had been group member for between 2-5 years, 30% had been group members for over 5 years, 24% had been group members less than two years.

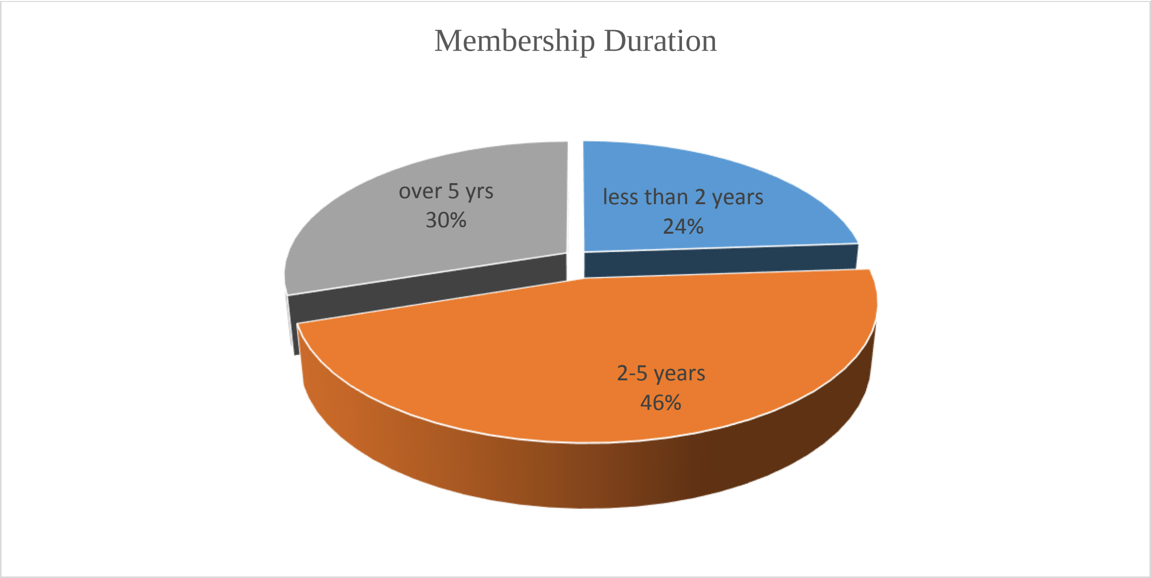

Figure 5: Group Membership Duration

4.1.6 Number of Dependents'

Number of dependents were a key factor of concern in determining the defaulting likelihood, the results were as presented in table 2. The findings show that majority, who were 45.1% had more than 5 dependents, 28.3% had between 3-5 dependents, and 19.2% had 1-3 dependents' while only 7% had no dependents.

Table 2: Number of Dependents

	Percent
No Dependents	7.3%
1-3 Dependents	19.2%
3-5 Dependents	28.3%
More than 5	45.1%
Total	**100.0%**

4.1.7 Amount of Loan

Amount of loan borrowed were also investigated to see whether it affected the rate of defaulting. The results findings were as shown in figure 6. The study found out that majority on the observation list borrowed of a loan of between 101,000-500,000 as shown by 46.4%, this was followed by 25.1% who borrowed a loan of less than 100,000, 20.8% borrowed a loan of between 500,001-1 million, while 7.7% who were the list borrowed a loan of over 1 million.

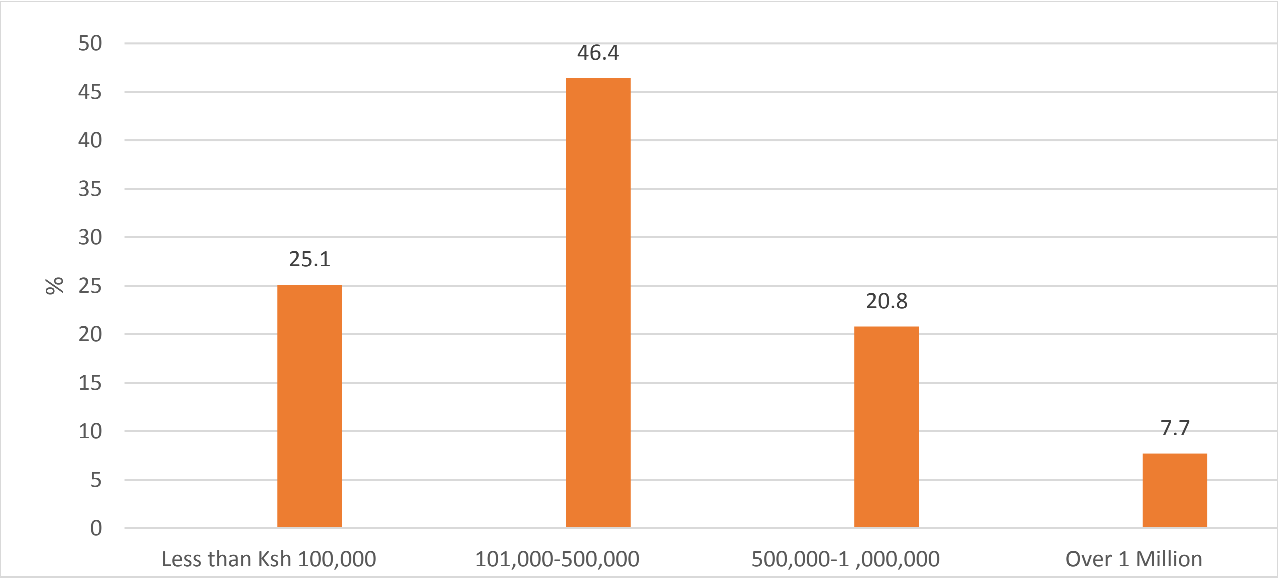

Figure 6: Amount of Loan Borrowed

4.1.8 Loan Repayment Period

From loam amount borrowed, observation was also made to determine the duration of the loan they took to be repaid, out of an observation of 390 in figure 7, 113 took 6-12 months to repay the loan, 105 took a period of less than 6 months to repay the loan, 87 took a loan of over 24 months to be repaid, lastly 85 took a loan that was repaid in between 12 and 24 months.

Figure 7: Loan Repayment Period

4.2 Performance verses Defaulting Ratings

From the observation, dependent variable which is rating whether one was a defaulter or a performer was analyzed and shown in figure 8. From the observation, 48.5% were in defaulting category while 51.5% were in a performing category.

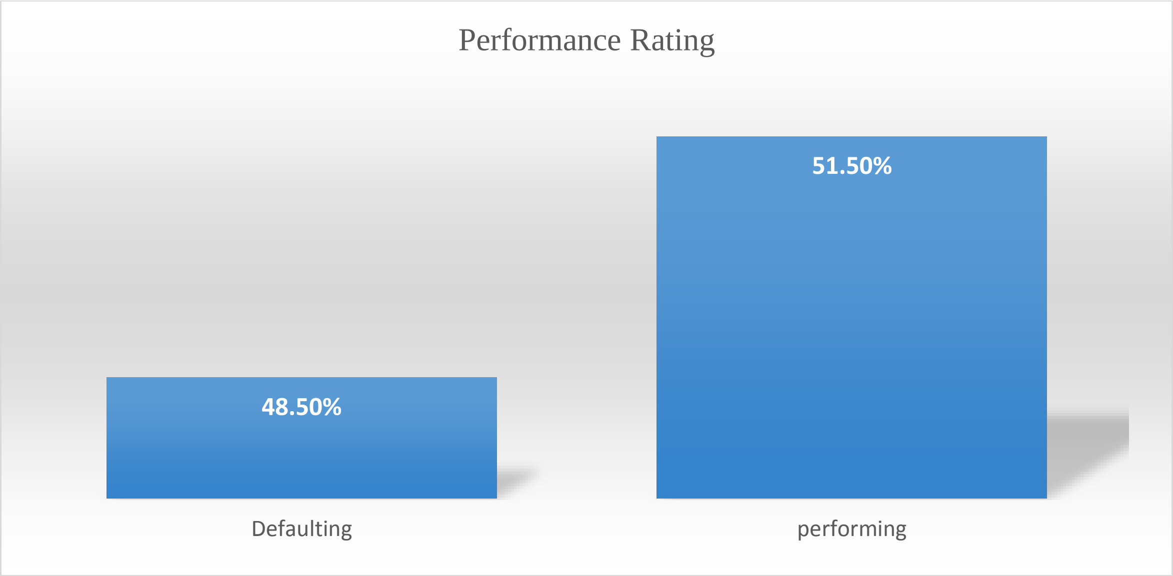

Figure 8: Performance Rating

4.3 Credit Scoring Using Logistic Regression

Logistic regression was the appropriate regression analysis to conduct the study since the dependent variable, Ranking, was dichotomous (binary). Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. It predicts the probability that an observation falls into one of two categories of a dichotomous dependent variable based on one or more independent variables that can be either continuous or categorical.

The assumptions of Logistic regression were met for this study to be carried out. The assumption includes; that the outcome must be discrete, otherwise explained as, the dependent variable should be dichotomous in nature, for this study the outcome was Defaulting vs. Performing. The second assumption is that there should be no outliers in the data, this was assessed by converting the continuous age predictors to standardized, or z scores. The third assumption is that there should be no high inter-correlations (multicollinearity) among the predictors. This was assessed by carrying out a correlation matrix among the predictors. Tabachnick and Fidell (2012) suggest that as long correlation coefficients among independent variables are less than 0.90 the assumption is met. The results in table 3 confirmed that there was no high inter-correlations (multicollinearity) among the predictors.

Table 3: Correlation Matrix of Predictor Variables

	Gender	Occupation	Marital status	membership	membership duration	Number of dependents	Amount of Loan	Repayment period
Gender	1.000							
Occupation	0.039	1.000						
Marital status	-0.018	0.066	1.000					
membership	0.006	-0.019	-0.07	1.000				
membership duration	-0.005	0.049	-0.031	0.215	1.000			
Number dependents	0.021	-0.068	-0.097	-0.224	-0.129	1.000		
Amount of Loan	0.033	-0.015	0.003	0.045	-0.007	0.069	1.000	
Repayment period	0.032	-0.058	0.044	0.114	0.149	-0.044	-0.024	1.000

Having met all the assumption. The study proceeded to predict the probability in which the observation falls into between defaulting and performing. The finding was as shown in table 2 and 3. Table 3 presented the model summary that explain how much variation in the dependent variable can be explained by the model. The table contains the Cox & Snell R Square and Nagelkerke R Square values, which are both methods of calculating the explained variation. These values are referred to as pseudo R^2 values. The explained variation in the dependent variable based on the model ranges from 33.4% to 44.5%, depending on the reference of the Cox & Snell R^2 or Nagelkerke R^2 methods, respectively.

Table 4: Model Summary

-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
342.809	0.334	0.445

The study models analyzed the characteristics of applicants such as income, marital status, amount of loan and loan duration were used to classify new borrowers into defaulting or performers (Chen and Huang, 2003). The Variables in the Equation table (table 3) shows the contribution of each independent variable to the model. The Wald test ("Wald" column) is used to determine statistical significance for each of the independent variables. The statistical significance of the test is found in the "Sig." column. "Variables in the Equation" table predicts the probability of an event occurring based on a one unit change in an independent variable when all other independent variables are kept constant.

The table shows that the odds of defaulting is 1.497 times greater for males as opposed to females and the relation is significant, all classes of age are not significant determinant of performance rating. Annual income was analyzed to determine its prediction in terms of performance rating in loan repayment. The study used annual income of less than 200,000 as the reference point in the analysis. The study found that odds of defaulting is 0.039 times less likely for individual earning an annual income of between 200,000-500,000 as opposed to individual earning income of less than 200,000, the relation is significant. Also, the odds of defaulting is 0.147 times less likely for an individual earning between 500,001-1 million as opposed to and individual earning income of less than 200,000, the relation is significant. Further, the findings show that the odds of defaulting is 0.263 times less likely for an individual earning annual income of over one million as opposed to individual earning income of less than 200,000, the relation is significant.

The study sought to know whether marital status could predict the rate of performance ranking. Married individual was used as the reference point. The study established that the odds of defaulting is 0.403 times less likely for an individual unmarried as opposed to individual married, the relation however is insignificant.

Group membership is a common factor that exist in KWFT and several microfinance banks clients and one of the basis of loan approval due to group and social guarantee. This was a key predictor which the study was keen to establish the prediction impact. The study found out that the odds of defaulting is 4.069 times greater for a non-member individual as opposed to individual who is a member to a group, the relation is significant. For membership duration the study found out that the odds of defaulting is 0.393 times less likely for an individual who has been a group member for between 2-5 years as opposed to individual who has been a member for a duration of less than 2 years, the relation is significant. The study also established that the odds of defaulting is 0.632 times less likely for an individual who has been a group member for between over 5 years as opposed to individual who has been a member for a duration of less than 2 years, the relation is significant.

Number of dependents supported by the loan borrower was analyzed to establish the impact on loan repayment performance and ranking performance. The study established that the odds of defaulting is 0.243 times greater for a loan borrower with between 1-3 dependents as opposed to one with no dependents, the relation is significant, additionally the study found out that the odds of defaulting is 0.741 times greater for a loan borrower with between 3-5 dependents as opposed to one with no dependents, the relation is significant. Equally importantly the study established that the odds of defaulting is 1.941 times greater for a loan borrower with between more than 5 dependents as opposed to one with no dependents, the relation is significant.

The study further sought to determine the rate at which amount of Loan borrowed predicted the ranking performance. The study found that the odds of defaulting is 1.705 times greater for a loan borrower with a loan of between 101,000-500,000 as opposed to one with a loan amount of less than 100,000, the relation is significant. The study equally established that the odds of defaulting is 1.784 times greater for a loan borrower with a loan of between 500,001-1 million as opposed to one with a loan amount of less than 100,000, the relation is significant. The study further established that the odds of defaulting is 1.954 times greater for a loan borrower with a loan of over 1 million as opposed to one with a loan amount of less than 100,000, the relation is significant. This study finding agree with (Schreiner, 2000) that the risk of loans with monthly installments increases by about 3 percentage points for each additional installment, he found out that given the number of installments, a loan repaid monthly was about 0.6 percentage points riskier than a loan repaid weekly.

The study sought to know whether loan repayment period could predict the rate of performance ranking. Repayment period of over 24 months was used as the reference point. The study established that the odds of defaulting is 0.577 greater for a loan borrower with the repayment period of between 12-24 months as opposed to individual with the repayment period of over 24 months, the relation is significant. The study further established that the odds of defaulting is 0.819 greater for a loan borrower with the repayment period of between 6-12 months as opposed to individual with the repayment period of over 24 months, the relation is significant. Finally, the study found out that the odds of defaulting is 1.109 greater for a loan borrower with the repayment period of less than 6 months as opposed to individual with the repayment period of over 24 months, the relation is significant.

Table 5: Variables in the Equation table

Variables	B	S.E.	Wald	df	Sig.	Exp(B)
Gender(Male)	0.404	0.282	2.048	1.000	0.152	1.497
Annual Income(Less than 200,000)			18.450	1.000	0.006	
Annual Income(Between 200,000-500,000)	-0.302	0.892	10.668	1.000	0.001	0.039
Annual Income(between 500,001-1 Million)	-0.127	0.342	6.283	1.000	0.020	0.147
Annual Income(Over 1 million)	-0.525	0.364	1.995	1.000	0.080	0.263
Marital status(unmarried)	-0.908	0.284	10.668	1.000	0.061	0.403
membership(member)	0.757	0.302	6.283	1.000	0.012	4.069
Membership duration(less than 2 years)			1.995	1.000	0.018	
Membership duration(2-5 years)	-1.645	0.377	18.998	1.000	0.000	0.393
Membership duration(over 5 years)	-0.459	0.325	1.995	1.000	0.018	0.632
Number of dependents (No dependents)			30.777	3.000	0.000	
Number of dependents (1-3 Dependents)	0.663	0.401	2.742	1.000	0.009	0.243
Number of dependents(3-5 Dependents)	0.300	0.455	0.435	1.000	0.041	0.741
Number of dependents(More than 5)	1.413	0.416	11.517	1.000	0.001	1.941
Amount of Loan(Less than Ksh 100,000)			3.497	3.000	0.321	
Amount of Loan(101,000-500,000)	0.534	0.377	2.009	1.000	0.016	1.705
Amount of Loan(500,001-1 ,000,000)	0.579	0.387	2.241	1.000	0.013	1.784
Amount of Loan(Over 1 Million)	0.670	0.413	2.627	1.000	0.011	1.954
Repayment period(Over 24 months)			33.333	3.000	0.000	
Repayment period(12-24 months)	1.284	0.391	10.775	1.000	0.001	0.577
Repayment period(6-12 months)	1.516	0.395	14.766	1.000	0.000	0.819
Repayment period(Less than 6 months)	2.219	0.393	31.938	1.000	0.000	1.109
Constant	22.525	40192.991	0.000	1.000	1.000	6.060
Reference category for Dependent Variable: Defaulting						

4.4 Non Discriminant Interpretation of the Scoring Model

The non-discriminant analysis represents an effective method for multivariate data analysis, often being use to extract relevant information from large and heterogeneous amounts of data. As a technique for classifying a set of observations into predefined classes, the discriminant analysis highlights their similarities and differences between them, creating an important advantage in describing the variability of a data set.

In this study, the dependent variable is a dichotomy of either performers or defaulters whereby performer was given a score of 1 while defaulter was given a score of zero. Therefore, using non discriminant analysis, the binary categories were further broken down as follows;

0.00-0.30=High rank defaulters

0.31-0.49=Moderate rank defaulters

0.50-0.69=Performers

0.70-0.100=High rank Performers

The general credit scoring model was as shown below

$$Prob\left(\frac{P/D}{1.00}\right) = \alpha + \beta 1GENDER + \beta 4INCOME + \beta 5MARITAL\ STATUS + \beta 6MEMBERSHIP + \beta 7MEMBERSHIP\ DURATION + \beta 8N0.DEPENDANTS + \beta 8AMT\ LOAN + \beta 9REPAYMENT\ PERIOD + \varepsilon$$

Therefore, the specific credit scoring model as per the results in table in table 4.4.3 above is as shown below;

$$\begin{aligned} Prob\left(\frac{P/D}{1.00}\right) = {} & \alpha + 0.40\ GENDER(male) - 0.302INCOME(200{,}000 - 500{,}000) \\ & - 0.127INCOME(500{,}001 - 1\ Million) - 0.525\ INCOME(Over\ 1\ Million) \\ & - 0.908\ MARITAL\ STATUS(Unmarried) \\ & + 0.757GROUP\ MEMBERSHIP(Member) \\ & - 1.645\ MEMBERSHIP\ DURATION(2 - 5Yrs) \\ & - 0.459\ MEMBERSHIP\ DURATION(Over\ 5Yrs) \\ & + 0.663N0.DEPENDANTS(1 - 3) + 0.300\ NO.DEPENDANTS(3 - 5) \\ & + 1.413\ NO.DEPENDANTS(Over\ 5) \\ & + 0.534AMT\ LOAN(101{,}000 - 500{,}000) \\ & + 0.574AMT\ LOAN(500{,}001 - 1\ Million) \\ & + 0.670AMT\ LOAN(Over\ 1\ Million) \\ & + 1.284REPAYMENT\ PERIOD(12 - 24Months) \\ & + 1.516REPAYMENT\ PERIOD(6 - 12Months) \\ & + 2.219REPAYMENT\ PERIOD(Less\ than\ 6\ Months) + \varepsilon \end{aligned}$$

5.0 Conclusions

The study concluded that the essence of credit scoring is the prediction of the risk of whether borrowers can repay the loan advanced to them. Defaulting is based on information, and in microfinance, this information is usually qualitative and informal and resides with group members or with loan officers. This model takes a different tack. It predicts the rate of default based on the quantitative information that resides in the management-information system of the lender. Up to now, microfinance lenders have depended almost exclusively on informal, qualitative information.

6.0 Recommendations

The study recommended that KWFT and other microfinance to adopt the findings of this study in credit decision-making. The loan application process should allow for branch managers to override the scorecard decision. Small numbers of overrides be acceptable since scorecard becomes invalid if this occurs to frequently due to lack of faith in the scorecards ability to rank order risk or avoiding confrontation with a long-term client. It is important for staff to trust and understand the abilities of the scorecard. The study also recommended that senior management should see the strategic value in developing, implementing, and using scorecards as an integral approach to managing risk within microfinance.

7.0 References

Blake, C.L., & C.J. Merz. (1998).UCI repository of machine learning databases [http://www.ics.uci.edu/~mlearn/MLRepository.html], University of California, Department of Information and Computer Science.

Business Banking Board (2000). *Credit Scoring for High-end Small Business Customers*.

Caire, D., & Kossman, R. S. (2003). Credit Scoring: Is It Right for Your Bank? In *Bannock Consulting, Credit Risk Conference*.

Cundiff, K. (2004). Closing the loop: How credit scoring drives performance improvements along the financial value chain, *Business Credit* 106(3) 38-42.

Diana, T. (2005). Credit risk analysis and credit scoring – now and in the future, *Business Credit* 107(3) 12-16.

Lahsasna, A. Ainon, R.N. and Wah, T.Y., 2010.Credit Scoring Models Using Soft Computing Methods: A Survey. *The International Arab Journal of Information Technology*, 7(2) pp.116-122.

Littwin, A. (2007). Beyond usury: A study of credit-card use and preference among low-income consumers. *Tex. L. Rev.*, *86*, 451.

Stein (2001). Peer. *Making Small Business Finance Profitable*. Proceedings from the Global Conference on Credit Scoring, Washington, D.C.

Thomas, L.C. (2000). A survey of credit and behavioral scoring – forecasting financial risk of lending to consumers, *International Journal of Forecasting* 16(2) 163-167