IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS)
e-ISSN: 2319-2380, p-ISSN: 2319-2372. Volume 13, Issue 2 Ser. II (February 2020), PP 48-57
www.iosrjournals.org

Analysis of Access to Credit from Smallholder Farmers in the Boane District, Mozambique: A Credit Rating Approach of Borrowers

Euclides Alfredo Matusse*, Cachimo Combo Assane**
**Agribusiness Research Program-Innovation in Information Technology, **Universidade Eduardo Mondlane*
Corresponding Author:Euclides Alfredo Matusse

Abstract: *This study was designed to investigate criteria and classification of credit access of small farmers in the District of Boane, Maputo Province, Mozambique using a cluster analysis approach. The objective is to group farmers according to the similarity of their characteristics (risk profiles), in relation to the determinants of access to credit that will allow the researcher to classify the study group as worthy of credit or least credited using metrics, namely accessibility to borrowing, repayment capacity, credit term and socioeconomic characteristics of borrowers. Thirty (30) farmers were selected to compose different configurations by using structured questionnaires, interviews and standard analytical test methods to determine the properties of the data applied in the selected metrics in the current literature and draw conclusions. In the classification of potential farmers benefiting from agricultural credit, the K -means clustering algorithm used identified two groups (cluster 1 and cluster 2) that are extremely heterogeneous, while, within each group, farmers tend to have homogeneous characteristics. Based on this principle, an attempt was made to build a logistic regression model for the purposes of classifying these groups according to their characteristics. The discriminatory analysis results from the Gap Statistic method, the optimal number of cluster to be defined and $K = 2$ with a high level of precision in the error analysis with the centroid group.*

The Chi-Square test, only the variable $Emp_{categoria2}$ $\left(x^2 = 4{,}72, \rho - value = 0{,}029\right)$ and statistically significant for a 5% level of significance. This means that this variable has a significant impact on the loan allocation. The value of $\exp(coef)$ = 29, 96 indicates that the chance of a farmer in the category $Emp_{categoria2}$ being in the group and 29.96 times (100 x (29.96 -1) = 2896%) greater than the chance of the farmer in the category $Emp_{categoria1}$ (category reference) belong to the same group. This means that there is a strong relationship between discriminatory functions and group variables in classifying farmers into credible or less-credited groups using theorized variables.

Keywords*: Loans and repayment, credit rating, Clustering Analysis, Logistic Regression analysis, farmers, metrics.*

Date of Submission: 24-02-2020 | Date of Acceptance: 07-03-2020

I. Introduction

Access to finance for rural families is essential for their businesses to be successful. Microfinance institutions have tried to fill the gap in accessibility to credit for farmers, despite the high level of defaults on their loans [9] [3]. This study is carried out to analyze the factors that influence criteria and classification of credit access of small farmers in the District of Boane in microfinance institutions. The findings will help microfinance institutions to arrive at appropriate measures to be applied in order to eliminate default and greater assertiveness in granting credit to low-income farmers in order to increase their agricultural income [14].

Micro finance was introduced in Mozambique to provide financing services to the poor and small medium enterprises (SMEs) to start businesses [2], however financial institutions specifically commercial banks refuse to provide micro finance to the agribusiness sector at the beginning of the value chain due the high risk arising from the interest rate, amount granted, repayment period, lack of guarantor, distance from sources of loan and critical restrictions that farmers face in the preparation of agricultural land, storage and marketing of their products [1] [6].

This study effectively responds to this aim driven by restrictions on financing credit to farmers in order to contribute to the increase in production and development of the agribusiness value chain and to reduce the deficit of agricultural products in the region's market. The financing processes are intended to be supported by metrics to guarantee the adequate flow and assertive financing decisions in order to cover the social classes with less purchasing power, however, access to formal credit in microfinance cooperatives is reduced. It is, therefore, imperative to expand and strengthen financial institutions within the cooperative society to play catalytic roles in this regard, especially in the area of supplying machines and tools, better inputs and education for farmers [4].

The main obstacle to the sustainable financing of agricultural projects in general and to rural families in particular, is related to the structure of operating costs, but also to a large extent by the needs of initial capital since the players in this sector are mainly small (family) producers with poor access to capital in the financial market [20], but also with little technical and business management knowledge, which maintains that most financial institutions only give loans to medium to large scale farmers in order to make a profit for cover the cost of production and repay the loan [18].

The cost side has to do, mainly with the necessary inputs of the agricultural production unit in the production chain, on the side of technical knowledge it has to do with the lack of systematized information and the low level of education and technique on this sector [17]. On the access to capital side, first is the fact that this sector requires large investments at the beginning of the projects, but also because the capital available on the formal market is very expensive due to the high interest rates, amortization period, amount granted and the low capacity of operators to present tangible or collateral guarantees to make credit viable [19].

Objective of the research

The general objective of the research work is to finance credit for rural families through the classification of clusters for small farmers in the district of Boane.

Specifically, this study aims to:

1. Identify, analyze variables and objects to be grouped to finance agricultural credit.
2. Study and investigate the grouping algorithm - It means grouping to group farmers for loan accessibility.
3. Determine the ideal number of clusters (K) and apply the K -means Clustering algorithm to obtain a natural group structure.
4. Estimates (predict) the probability of any farmer belonging to one of these groups and establish the importance of each characteristic for the calculation of these probabilities.

II. Theoretical Analytical Framework

The use of indicators provides several benefits, including the effective evaluation of the financing of agricultural loans to farmers and also support to the credit manager in making strategic decisions. A systematic review study conducted by Euclides Alfredo Matusse [5], found that 11.11% of the articles deal with the proposed solutions; evaluation survey (66.67%); research articles under analysis (16.67%); and (5.55%) experimental validation article using metrics. This implies despite the relevance of using metrics to help finance agricultural credit given the diversity of the existing business model, organizations have neglected their practice. The reason for this includes: although access to credit is a significant factor in the adoption of agricultural technologies and in increasing agricultural incomes among rural farmers, they are generally not mature enough to make use of measurements [7].

In this context, the figure below consists of evaluating the metrics to ensure that the quality of agricultural credit financing offers a perspective for the credit manager to measure assertively and monitor his experience in microcredit services. This framework is divided into (four) 4 main stakeholders, groups of non-governmental organizations, rural community (hereinafter Cooperative/Farmers Association), micro-finance and the government (figure 1). The model implements the concept of greatest need, it analyzes the variables considered or objects relevant to the study of a natural structure of groups that are the metrics of determinants of access to credit, namely accessibility to the loan, repayment capacity, credit term and characteristics socioeconomic in decision sub committees. Each sub-committee prepares an assessment of all stages of the process, including the selection of the algorithm K -means Clustering of grouping, a non-hierarchical procedure, the most used to group individuals whose initial number of cluster is pre-defined by the analyst [13]. The basic idea of the K -means Clustering algorithm is to define groups of individuals or objects based on their own characteristics, in such a way that the internal variance to the groups is minimized (homogeneous individuals within the group) and the variance between the groups maximized (groups heterogeneous) [10].

In this context, the evaluation follows a determined period, often of three to four days, within which credit analysts must complete and deliver each loan decision process, except for the credit demand that may be adopted by some credit agents and portfolio. One of the tools that will assist in the loan decision process is the logistic regression model. Considering the structure of the groups (obtained by the K -means algorithm) as a dependent variable and the determinants for credit granting as explanatory variables, through logistic regression it is possible to determine the probabilities of any individual belonging to these groups, in addition to establishing the importance of each determinant in the credit granting decision.

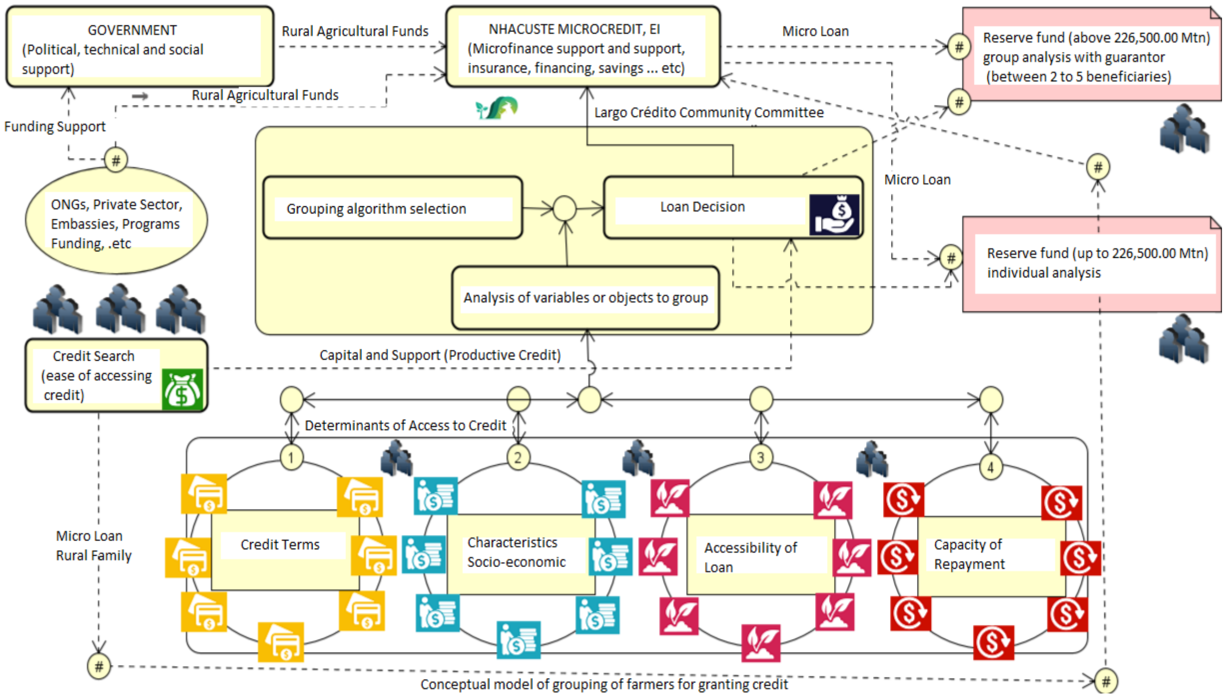

Figure 1. Conceptual model of grouping of farmers to grant agricultural credit

The committee is responsible (in consultation with its members) for asset management, operation, maintenance of microfinance services within the community. In addition, manage the finances to guarantee the full cost of credit recovery (including the loan, repayments, salaries, maintenance and fund reserve for future loans). The commission should be composed of elected representatives of the community and credit agents, for example, portfolio managers, operations managers, resource manager and microfinance services (among others, as required for specific cases). Administratively, the Largo Credit Committee must consist of two subcommittees (technical and financial) and a secretariat (secretary and treasurer).

Government refers to all ministries that play a role in financing rural services, ensuring quality and sustainable access. After reforms and decentralization in the area of agriculture, activities by different actors must, however, be managed and monitored by the Ministry of Agriculture and Rural Development - MADER (or the regulation of services provided by the microcredit agency).

ONGs, a group of non-governmental organizations, are responsible for financing micro-credits through specialized funds for rural development, including productive credit funds. This could be through budget support from the Ministry of Economy and Finance - MEF.

These funds can be used to subsidize micro loans:

- Subscribe to microfinance loans used exclusively for provisioning, including deferred repayments to close the construction period for micro credit facilities, during which there is no generation of income;
- Supervise the activities of micro credit as provided by the legislation in force in microfinance or banking.

So Nhacutse microcredit EI, must:

- Design different micro loans / credit products in collaboration with the government and groups of non-governmental organizations;
- Disbursing and monitoring / auditing micro loans in collaboration with the community credit sub-committee;
- Provide savings opportunities for groups of farmers, among other services (insurance, etc.);
- Identify and hire private actors (if necessary) in collaboration with the project; and
- Discuss and ensure the viability and sustainability of proposed business plans.

This framework confirms that rural communities can acquire agricultural credit financing through the implementation of a new hybrid financing model that uses productive credit and micro loans to rural families through analysis of loan accessibility, repayment capacity, credit term and socioeconomic characteristics for production and expanding services. However, the model cannot be sustained in the long term without the active participation of members of the cooperative society. In addition, the framework depends on support from external stakeholders including a strong political will to prioritize and support the financing of the family agricultural sector, for example, in specialized farmer training funds. So the proposed framework can be used to channel targeted support (financial, technical) to poor and low-income families.

III. Methodology

A. The Study Area

The study was conducted in the District of Boane. The district is located to the southwest of Maputo province, being limited to the north by the district of Moamba, to the south and east by the district of Namaacha, and to the west by the city of Matola and district of Mautuine, located 30 km from the city of Maputo and it lies between longitude 32 ° 23 '20 "east and latitudes 26 ° 1' 44" south. With an area of 815 Km^2 and a population density of 101 habitants / Km^2, the population is 42% young under 15 years old, mainly female (masculinity rate 47%) and urban and semi-urban matrix (urbanization rate 68%).

The waterways in the district of Boane belong to the hydrographic basins of the Umbeluzi, Tembe and Matola rivers. The Umbeluzi valley has soils with good agricultural and livestock potential, which should be explored by a vast fabric of private and family farming. Agriculture is the basis of the district economy, with vegetables, corn, manioc, beans, bananas and citrus fruits as the main crops. Agricultural development depends mainly on the uses of loans and their facilities for micro credit institutions to help borrowers use modern technologies and advanced practices in production fields.

In this view, micro credit plays a significant role in providing loans for farmers to leverage the production and productivity of their agricultural income [16]. The primary data were collected directly from 30 farmers to compose different configurations using structured questionnaires, interviews and standard analytical test methods were used to determine the properties of the data applied in the selected metrics in the current literature. Secondary data were also collected from published and unpublished data from DPIC[1], Boane.

In relation to the statistical treatment of the data, a cluster analysis (grouping) was carried out, with the objective of grouping farmers according to the similarity of their characteristics (risk profiles), in relation to the determinants of access to credit. Additionally, using the structure of the groups obtained in the cluster analysis, a logistic regression model was built to assist in the decision-making process regarding the granting of credit.

B. Analysis of the variables and objects to be grouped

The variables considered relevant to the study of a natural group structure are the metrics of determinants of access to credit, namely accessibility to the loan, repayment capacity, credit term and socioeconomic characteristics of the farmers. It is important to note that this technique is highly sensitive to the inclusion of variables with outliers, that is, observations that deviate from the expected pattern in each variable. Furthermore, if there are variables with missing data (missing values), they must be estimated or the corresponding individuals must be removed [13].

The verification of the presence of outliers was done using boxplot graphics. With the exception of the variable “reimbursement”, there is no indication of the presence of outliers in the data regarding the determinants of access to credit. Individually evaluating the boxplot of the variable “repayment”, despite the observation of the borrower 13, appearing to be outliers as shown in figure 2, does not represent a sufficient reason for the elimination of this individual from the database.

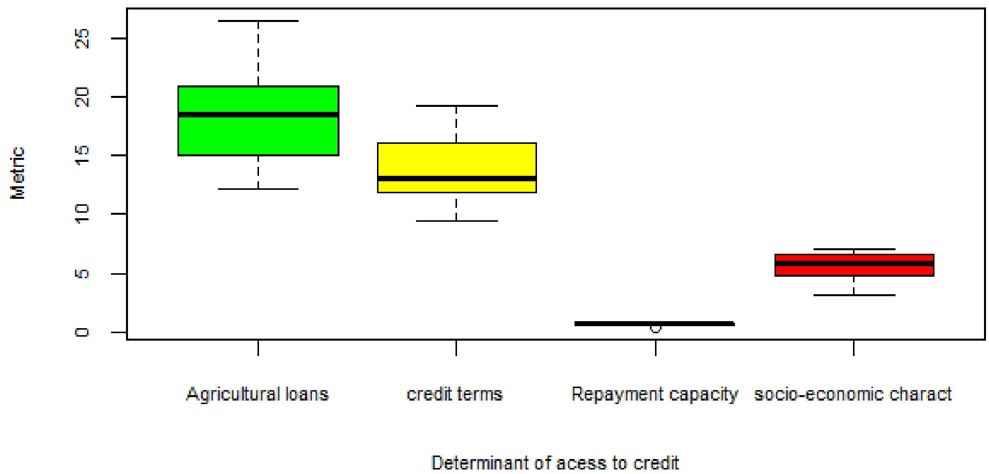

Figure 2.Boxplots for checking the presence of outliers

[1] Direcção Provincial da Industria e Comércio (https://www.pmaputo.gov.mz/)

A critical issue to be taken into account in the cluster analysis is multicollinearity between variables [10]. According to James, et al. [12], the best way to measure multicollinearity and computing the variance inflation factor (VIF). The VIF for each variable can be calculated using the individual formula in the database.

$$VIF(\hat{\beta}_j) = \frac{1}{1 - R^2_{\chi_j|\chi_{-j}}}$$

Onde $R^2_{\chi_j|\chi_{-j}}$ is amount of the variable in question that is explained by all other independent variables in the regression model. In practice, a VIF value above 5 or 10 indicates a multicollinearity problem [12].

Table 1 shows the VIF values (below 5) for the variables under study. Based on the aforementioned criterion, there is no evidence of multicollinearity among the determinants of access to credit.

TABLE I. VIF VALUE FOR EACH DETERMINANT

Variable	Loan	Repayment	Term	Characteristics
VIF	2,82	1,49	4,14	1,83

The use of variables with different measures / scales can distort the structure of the grouping. This problem is overcome by standardizing the variables. The most used form for standardization is to transform each variable into a standard score (Z Scores), in order to present zero mean and standard deviation [8]. The Z score is calculated from the following formula:

$$Z = \frac{x - media}{\text{standard deviation}}$$

```
            Emp        Rpc        Term       Cart
Mut1 -0.24722061 -0.5944401 -1.2728259 -0.5903928
Mut2 -0.12261519  0.0424600 -0.3754742 -0.6296129
Mut3  1.65301204  1.3162601  1.6577658  1.1431343
Mut4 -0.02656518 -0.5944401 -0.3754742  0.2724488
Mut5 -0.46528010  0.6793601  0.9383699  0.8529058
Mut6 -1.28559911 -1.2313401 -0.6935229  0.2724488
```

Figure 3. Part of the standardized data

Where Z is the amount of the variable in question that is explained by all the others, x -related variables in the regression model, Emp-loan metrics, Rpc-repayment metrics, $Term$- credit term metrics and $Cart$- borrower characteristics metrics.

C. Similarity or distance measures (dissimilarity)

The identification of groupings of subjects is only possible with the adoption of some measure of similarity that allows for an objective comparison between the subjects. Similarity represents the degree of correspondence between objects across all characteristics used in the analysis [10]. There are several ways to measure similarity or distance between objects, which results in a matrix of distances [13]. In this work, the Euclidean distance measure is adopted, recommended when the variables are not multicollinear.

In the Euclidean distance, the distance between two observations (i e j)corresponds to the square root of the sum of the squares of the differences between the pairs of observations (i ; j) for all p variables, that is,

$$d_{ij} = \sqrt{\sum_{k=1}^{p} \left(\chi_{ik} - \chi_{jk}\right)^2}$$

Where χ_{ik} is the amount of variable k for observation i and χ_{jk} represents variable k for observation j. Figures 4 and 5 show the values of the distance matrix. The intensity of the color is proportional to the value of similarity (or dissimilarity) between observations: pure red indicates $d_{jk}(x_i; x_j) = 0$, greater similarity; and pure blue indicates $d_{jk}(x_i; x_j) = 1$less similarity.

```
     Mut1 Mut2 Mut3
Mut1  0.0  1.1  4.3
Mut2  1.1  0.0  3.5
Mut3  4.3  3.5  0.0
```

Figure 4. Part of the Euclidean distance matriz

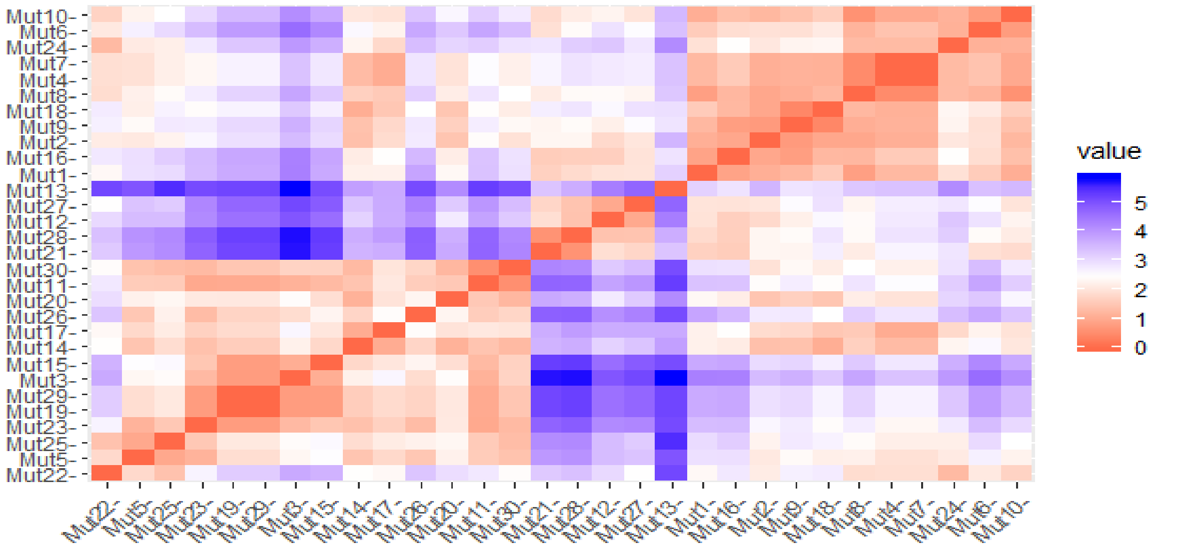

Figure 5. Red: high similarity— Blue: low similarity

D. Selection of the grouping algorithm

This work uses the K -means Clustering algorithm. It is a non-hierarchical procedure, the most used to group individuals whose initial cluster number is pre-defined by the analyst [13]. The basic idea of the K -means Clustering algorithm is to define groups in such a way that the internal variance to the groups is minimized and the variance between the groups maximized. The distance criterion for the formation of groups in K -means clustering and the Euclidean distance.

The standard algorithm -means Clustering and "Hartigan-Wong algorithm", which defines the internal variation to groups as the sum of the quadratic Euclidean distances between individuals and the corresponding centroid:

$$W(C_K) = \sum_{\chi_i \varepsilon C_k} (\chi_i - \mu_k)^2$$

Where χ_i is an observation that belongs to the cluster C_K; k is the average value (the centroid) of the observations allocated to the cluster C_K. Each observation χ_i is allocated to a given cluster C_K in such a way that the sum of the quadratic distances of the observation in relation to the respective centroid k is minimal. The total variation within groups is defined as:

$$\sum_{k=1}^{k} W(C_K) = \sum_{K=1}^{K} \sum_{\chi_i \varepsilon C_k} (\chi_i - \mu_k)^2$$

The K -means Clustering algorithm can be summarized as follows:

- Specifies the number of clusters to be created (done by the analyst);
- Randomly selects individuals from the data set as starting centroides;
- Allocates each individual to the cluster whose centroid is closest, based on the Euclidean distance between the individual and the centroid;
- For each of the clusters, the centroid is updated, calculating new average values of all observations within the cluster;
- Iteratively and minimizing the total internal variation to the clusters.

E. Determination of the optimal number of clusters

To determine the optimal number of clusters (K) to be defined, the Gap Statistic method, proposed by Tibshirani et al. [23]. The method compares the total internal variation of the clusters for different values of with their expected values under the null hypothesis of the reference distribution of the data.

Based on the results of the Gap Statistic method, the optimal number of clusters to be defined e K = 2. Applying the K -means Clustering algorithm, the results below are obtained:

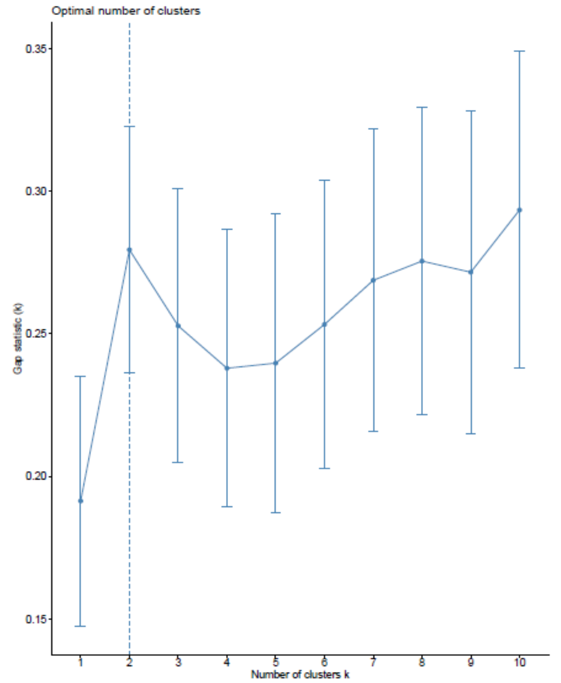

Figure 6. Determination of the optimal cluster number by the Gap

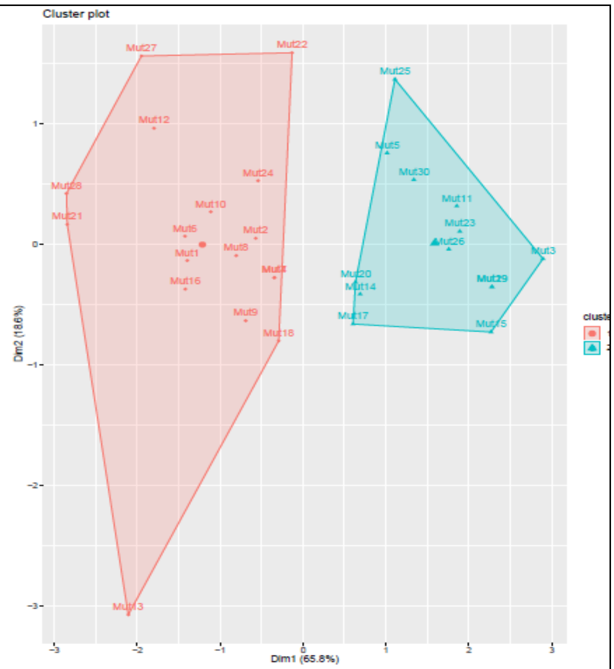

Figure 7. Grouping of borrowers according to K-means Clustering

IV. Empirical and Result Discussion

To arrive at the final profile of farmers based on the similarity of their characteristics in relation to the determinants of access to credit, the algorithm -means Clustering was used, considering the Euclidean distance between the individual and the centroid. The decision on the optimal number of groups (centroids) to consider in the analysis was based on the Gap Statistic method. The structure of two groups found through the K -means clustering algorithm indicates that the characteristics referring to determinants of credit access (Emp , $Term$, $Cart$ and Rpc) for farmers belonging to the two groups (cluster 1 and cluster 2) are extremely heterogeneous, whereas, within of each group, farmers tend to have homogeneous characteristics. Based on this principle, an attempt was made to build a logistic regression model for the purpose of classifying farmers to these groups according to their characteristics. This test aims to judge the relative importance of discriminatory variables. Thus, the result shows a good estimate, since the value defined by $k = 2$ and the optimal number of clusters to be formed by a group of 30 farmers.

Thus, the selection of potential farmers receiving agricultural credit should have a function of the magnitude and sign of the metrics for including variables in the model, if all other considerations are made constant. In the analysis of the socioeconomic characteristics of farmers for cluster 1, they consisted of sex, age, education level, agricultural experience, size of cultivation area, number of households and agricultural income. Of the male respondents (55.2%) while (47.5%) of them were female. This shows that there was a significant difference in lending between men and women. The reimbursement capacity metric variable (Rpc) almost remained constant at an insignificant average of 0.6573 and standard deviation 0.0628 which shows that female farmers have a better performance reimbursement than men.

The probability of a loan default rate for male farmers was higher than that of women, this can be explained by the behavior and concern of women not to lose their good reputation and identity in society. When age was analyzed, it was found that 8.1% of respondents aged over 50 years, 21.1% aged between 41 and 50 years old, 30.3% aged between 31 and 40 deserving a loan. This indicates that adults were more likely to lend than young people because they are rational in making decisions and applying for loans in agricultural operations.

The result of the size of the cultivated area varies between 0.10-4.99 hectares; this result means that most farmers are small-scale and have difficulties in accessing agricultural loans from formal sources. This result was consistent with Kuye findings [15], who observed that most financial institutions only provide loans to medium to large scale farmers in order to obtain a profit to cover the cost of production and repay the loan. When the number of households was analyzed, it was found that among the respondents 1-2 people had dependents 50.0% received loans, among those who had dependents 3-5 people correspond 32.3% and among those who fell in the range 6-10 people 17.7%.This indicates that respondents who have fewer dependents are likely to receive more loans than those who have more dependents, because they tend to divert the application to family expenses, such as medical care, tuition fees, food, which negatively affects their search for future loans.

The majority (53.3%) received agricultural income less than 60,000 MTN, with an average agricultural income of 36,722 MTN. This clearly shows why farmers tend to increase their income with other assets that can be used as collateral for the loan, such as being employed in the civil service, losing focus on the use of credit granted for production, which hinders formal agricultural lending.

Regarding the level of education analyze the granting of loans, 21.76% of people have primary education, 76.93% of people with secondary education, 1.31% of people with university education. This shows that the cases of lending were low among farmers with primary education compared to those who had completed secondary education, government institutions should encourage farmers to improve their educational level and the membership of the cooperative society to obtain better classification in the granting of credit.

These results were similar to those obtained by Abiodun, et al. [1], who examined the factors that determine farmers' access to formal and informal agricultural loans among farmers in the Obubra agricultural zone in Cross River State, Nigeria. Their results from the multiple regression analysis showed that amount of loan obtained by farmers; years of agricultural experience with credit use and level of education were the main factors that positively influenced access to loans from formal sources.

Descriptive statistics of the dependent and independent variables of the sample used show that the average of 64.5 loans were credit worthy of cluster 2, the sexual distribution showed almost 68% of the farmers were male and the remaining 32% of them were female. Borrowers' ages ranged between twenty-one and fifty years of age, with an average of thirty-five years for all respondents belonging to the group. In addition, the above results showed that the majority (86%) of borrowers were married and, on average, they have 4 members of their family. On average, they were disbursed for the purpose of growing vegetables (47%) and other crops (53%). The education level distribution showed that 61% of borrowers had education, while 37% had secondary education. Only 2% of them reached higher education in the study area. It suggests that middle-aged borrowers are economically active workers with energy and, therefore, they should adopt new techniques and innovations to increase agricultural production and productivity.

Traditional and rudimentary agriculture still predominates in the agricultural production system through the use of human force, this means that old farmers will face severe restrictions and will be less productive than younger and more energetic farmers.

A. Logistic regression analysis

After defining the structure of groups of farmers, based on the similarities of their characteristics, logistic regression was used to estimates (predict) the probability of any individual belonging to one of these two groups and to establish the importance of each characteristic for the calculation of these probabilities. The results of this model are important in the decision-making process for granting credit to farmers.

Due to the small sample size, it was decided to estimate a regression model using the maximum likelihood estimator penalized, as described by Rainey and McCaskey [22], whose name of the statistical package in the R and logistf software.

A model is defined as logistic if the function follows the following equation:

$$f(z) = \frac{1}{1+e^{-(Z)}}, sendo\ Z = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1\chi_1 * \beta_2\chi_2 + \ldots + \beta_k\chi_k$$

Where p it indicates the probability of the occurrence of a certain event of interest (ie, probability of belonging to a specific group), $X_1, X_2, X_3 ... X_k$ are the explanatory variables (determinants of access to credit), α and β_i ($i = 1; 2 ... ; k$) are the model parameters. The term $\ln\left(\frac{p}{1-p}\right)$ is called logit and the term $\frac{p}{1-p}$ represents the chance ($odds$) of occurrence of the event of interest. Simply put, the function $f(z)$ can be understood as the probability that the dependent variable is equal to 1, given the behavior of the explanatory variables.

Starting from the group structure, obtained in the cluster analysis, a dichotomous dependent variable of the logistic regression model was created, having as categories 1 (for the first group) and 0 (second group), forming a dependent variable (dichotomous) of the logistic regression model. In addition, for reasons of flexibility of the model in question, the original observations of each of the determining variables of access to credit were also categorized into three categories. Thus, the logistic regression model, as proposed by Rainey and McCaskey [22], was implemented in software R, using the logistf package (Heinze et al, 2018) [11], with the following results being obtained:

	coef	SE(coef)	exp(coef)	Chisq	p-value	IC (95%) para exp(coef)	
						L_{inf}	L_{sup}
(Intercept)	-4,23	2,15	0,015	11,65	0,000	0,000	0,284
$Emp_{categoria2}$	3,40	1,81	29,96	4,72	0,029*	1,36	11849,01
$Emp_{categoria3}$	3,62	3.10	37,34	1,38	0,239	0,045	202804,96
$Rpc_{categoria2}$	0,56	1,57	1,75	0,112	0,738	0,018	252,14
$Rpc_{categoria3}$	-0,07	2,28	0,93	0,001	0,974	0,001	788,39
$Term_{categoria2}$	2,02	2,61	7,54	0,531	0,466	0,037	38177,44
$Term_{categoria3}$	2,27	2,41	9,68	1,11	0,291	0,177	20743,74
$Cart_{categoria2}$	0,40	1,74	1,49	0,042	0,239	0,01	595,86
$Cart_{categoria3}$	-0,03	2,23	0,97	0,002	0,738	0,002	214,86

Figure 8. Result of the logistic regression model

B. Interpretation of logistic regression model parameters

The correlation analysis tells us the degree and direction of association between three categories, but it does not give the magnitude of the change in the dependent variable due to a change in an independent variable. In this study, the logistic regression model is used to estimate the probability of a farmer belonging to one of the two groups, obtained by the K -means Clustering algorithm, in addition to establishing the degree of importance of each determinant in the concession decision process credit. As previously mentioned, the variable dependent on the regression model is related to the structure of the groups, while the independent variables are all the determining characteristics of access to credit, namely loan accessibility (Emp), repayment capacity (Rpc), credit term ($Term$) and socioeconomic characteristics ($Cart$).

The figure above (figure 8) shows the results of the estimated parameters. The first column shows the test categories for each variable. Note that for all variables, category 1 is omitted, which means that it was considered as a reference category (which is the basis for comparison with the other categories). The second column represents the values of the coefficients of the model parameters; in the third, it is the standard error of the estimate; the fourth column indicates the exponential coefficients of the model, which estimate the ratio of the chances of the dependent variable when moving from a reference category to the test category of the independent variable. Values $\exp(coef)$ greater than 1 $(\beta > 0)$ indicate an increase in chances, while values less than 1 $(\beta < 0)$ indicate a decrease in chances when moving from the reference category to the category under test. In the fourth and fifth columns, the statistics of the Chi-Square test are presented (similar to the multiple regression t test); and finally, the lower and upper limits of the confidence interval $\exp(coef)$.

According to the Chi-Square test, only the variable $Emp_{categoria2}$ $\left(x^2 = 4{,}72, \rho - value = 0{,}029\right)$ is statistically significant at a 5% level of significance. This means that this variable has a significant impact on the loan allocation. The value of $\exp(coef)$= 29,96 indicates that the chance of a farmer in the category $Emp_{categoria2}$ being in the group is 29,96 times (100 x (29,96 -1) =2896%) greater than the chance of the farmer in the category $Emp_{categoria1}$ (category reference) belong to the same group. To calculate the probability that a new farmer with specific characteristics (converted into categories) belongs to group 1, simply apply the following formula:

$$P(Y=1) = \frac{1}{1+e^{-(-4{,}23+3{,}4*Emp_{cat2}+3{,}6*Emp_{cat3}+...+0{,}40*Cart_{cat2}-0{,}03*Cart_{cat3})}}$$

V. Conclusions

The study analyzed results of discriminatory analyzes between independent variables as determinants of credit for small farmers and members who used agricultural credit. The discovery led to the conclusion that promoting fragmentation of farmers 'loans would require the conscious use of policies designed to increase the size of the loan farmers' and farms or reduce the size of the family in a small cooperative.

Regarding the statistical analysis of the data, using machine learning techniques, it was possible to create a strategy that will assist credit analysts in making decisions about granting credit to farmers. Through the K -means Clustering algorithm, two heterogeneous groups of farmers were defined based on their determining characteristics of access to credit and, based on this grouping, a logistic regression model was constructed to classify new farmers from the same population in terms of pre-defined, in addition to identifying characteristics with significant impact for this classification. This means that, with this strategy, it is possible to predict which group a new farmer will belong to, considering their specific characteristics.

The results of the regression model indicate a significant influence of the loan accessibility variable on the classification of farmers.

As a future work, we intend to develop a platform, which will use the strategy created in this study, to classify farmers and, mainly, to estimate the “ideal amount” of credit to be granted, according to their characteristics or risk factors.

One of the limitations of this work was the small sample size (only 30 farmers) used for the statistical analysis of the data. This question may impact the representativeness of the relative sizes of the groups (found in the cluster analysis) in the real population, for the purpose of building the logistic regression model. The solution found to minimize this fact was to estimate a regression model using the maximum likelihood estimator penalized. Nevertheless, it is recommended, for the next studies, to increase the sample size.

References

[1]. Abiodun S. O., Kuye O. O., Edet, E. U. Determinants of Access to Formal and Informal Agricultural Loan amoung Farmers in Obubra Local Government Area, Cross River State, Nigeria. International Journal in Management and Social Science, Volume 07 Issue 03, March 2019 ISSN: 2321-1784 Impact Factor: 6.178, pages 1-10, 2009.

[2]. Banco de Mocambique. Relatório anual. Volume 24. Relatório Anual nº 26 – Maputo, 2017.

[3]. Eze C., Ibekwe U. Determinants of loan repayment under indigenous financial system in Southeast, Nigeria. Medwell Journal, 2(2), 116-120, 2007.

[4]. Ejefobihi U., Imoagwu C., Ezeanyeji, Clement. I. *Microfinance - One of the Key Drivers of Financial Inclusion in Nigeria*. Journal of Policy and Development Studies. Journal of Pol Dev Studies DOI: 10.12816/0047552.An Open Access Journal ISSN: 0189-5958: 6.178, pages 1-19.

[5]. Euclides Alfredo Matusse. Metrics and Indicators to Aid in the Financing of Agricultural Loans to Rural Families: A Systematic Review. Global Journal of Science Frontier Research Volume XIX Issue IV Version, PP 50-58, September 2019.

[6]. Euclides Alfredo Matusse. Experimental Validation of Factors Related to Agricultural Credit Based on Metrics from the Perspective of Borrowers. IOSR Journal of Agriculture and Veterinary Science. Vol 12, Issue 9 Ser. II, PP 1-10, September 2019.

[7]. Euclides Alfredo Matusse. Experimental and Theoretical Expansion of Access to Credit among Rural Farmers: Case Studies in Boane District, Mozambique. Global Journal of Science Frontier Research Volume XIX Issue IV Version I, Estados Unidos, PP. 1 - 11, Dezembro. 2019.

[8]. Fávero, L.P., Belfiore, P., Da Silva, F. L., Chan, B.L. Análise de Dados: Modelagem Multivariada para Tomada de Decisões. Elseiver, 2009.

[9]. Fred Nimoh, Enoch Kwame Tham-Agyekum2,Samuel Ayisu. Factors Influencing Credit Default: ACase Study of Maize Farmers in the Asante AkimNorth District of Ashanti Region. International Journal of Agriculture and Forestry, Vol. 2(2),PP. 24-28, 2012.

[10]. Hair, Jr., J. F., Black, W.C., Babin, B. J. Anderson, R. E., Tatham, R. Multivariate Data Analysis, 7th Ed. Pearson, 2014.

[11]. Heinze, G., Ploner, M., Dunkler, D. Southworth, H. Logistf: Firth’s Bias-Reduced Logistic Regression. 2018

[12]. James, G., Witten, D., Hastie, T., Tibshirani, R. An Introduction to Statistical Learning with Applications in R. Springer, 2013.

[13]. Kassambara, A. Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning.1st Ed. Sthda, 2017.

[14]. Khalid Mohamed. Access to Formal and Quasi-Formal Credit by Smallholder Farmers and Artisanal Fishermen: A Case of Zanzibar. Research on Poverty Alleviation.2003.

[15]. Kuye O. O., Abiodun S. O., Edet, E. U. Determinants Of Access To Formal And Informal Agricultural Loan Amoung Farmers In Obubra Local Government Area, Cross River State, Nigeria. International Journal in Management and Social Science, Volume 07 Issue 03, March 2019 ISSN: 2321-1784 Impact Factor: 6.178, pages 1-10, 2009.

[16]. Mosca, J., Dadá, Y. A. Contributo para o estudo dos determinantes da produção agrícola. Observador Rural Nº 5. Observatório do Meio Rural. Abril, Maputo. 2013.

[17]. Mosca, J., Dadá, Y. A. Investimento no Sector Agrário. Observador Rural Nº 14. Observatório do Meio Rural. Fevereiro, Maputo. 2014.

[18]. Nawai N., Shariff M.N.M. Determinants of Repayment Performance in Microcredit Programs: A Review of Literature, International Journal of Business and Social Science Vol. 1 No. 2; November 2010, pp 1-10.

[19]. Njoku, Maria-Stella E. Analysis of Factors Affecting Agribusiness Cooperators Access to Credit from Formal Sources in Abia State, Nigeria, Agro-Science Journal of Tropical Agriculture, Food, Environment and Extension Volume 15 Number 2,pp. 23 - 28,2016.

[20]. Ndiege B. O., Mataba L., Msonganzila M., Nzilano K. L., The link between financial performance and loan repayment management in Tanzanian SACCOS, African Journal of Business Management, 28 February, 2016, Vol. 10(4), pp. 89-97.

[21]. Rainey, C., McCaskey, K. Estimating Logit Models With Small Samples. Inproceedings. 2017.

[22]. Tibshirani, R., Walther, G., Hastie, T. Estimating the number of clusters in a data via the gap statistic. J.R. Statist. Soc. B. Vol. 63, Part 2, pp 411-423.2001.

Euclides Alfredo Matusse. “Analysis of Access to Credit from Smallholder Farmers in the Boane District, Mozambique: A Credit Rating Approach of Borrowers.” *IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS)*, 13(2), 2020, pp. 48-57.