Smallholder Loan Performance Simulation and Monitoring

Francis Anokye (francis.anokye@aims-senegal.org / francis.anokye@manobi.com)

Mathematical Sciences for Climate Resilience Internship Program(MS4CR-AIMS)

Supervisors:
Mr. Pierre C. Sibiry Traore
ICRISAT/Manobi Africa Group Company, Senegal
Mr. Daniel Annerose
Manobi Africa Group Company, Senegal

December 2019

Contents

1. Introduction

The global food production contribution of smallholder farmers across the world is highly significant and plays a special role in the global efforts to improve food and nutrition security [12]. In this sense, the world relies heavily on the productivity of smallholders to fulfill its food needs. It is estimated that about half of the worlds' cereals, three-fourths of the world's dairy production and 60% of the world's meat are produced by smallholders [24]. In spite of their contribution in the global fight against hunger and food insecurity, they are challenged by the paradoxical fact that represents them as the world's most undernourished and food insecure population despite having agriculture and food production as their core occupations [55]. This challenge is rooted in several factors which include but not limited to the difficulties they face in the markets regarding agricultural credit acquisition, farm inputs (including improved seeds and fertilizers), insurance, and other technical assistance. Climate change and its related impacts also pose several environmental constraints on their efforts and massively impact their yield.

The World Bank through its report [30], World Development Report 2008, emphasized with the outset declaration: 'it is time to place agriculture afresh at the center of the development agenda'. Agriculture is a viable source of poverty reduction tool for every country that channels the right proportions of its resources into its sustainability. In the efforts and campaign of the united nations (UN) to achieve better and more sustainable future for all, it outlines "zero hunger" as the second [1] sustainable development goal (SDG) with the broader vision of doubling smallholders' productivity and income which holds the promise of eradicating most of the developmental deficits that exist on our planet due to its inter-connectivity with other development goals. The July 2019 report of the food and agriculture organization of the united nations (FAO) laments about the state of food security and nutrition with regards to the progress made in achieving most of the SGD goals linked to the above mentioned SDG. The report raises concerns about how off track the world has come in achieving this goal by 2030.

A decline in agricultural performance is a contributing force behind growing poverty among African smallholder farming populations, and its recovery offers the greatest prospects for rural populations to escape out of poverty. This inference can be adduced from the effect of the 'green revolution' in agriculture that swept across large parts of the developing world during the 1960s and 1970s which dramatically increased agricultural productivity and further reduced poverty [54]. The bulk of work that goes into productive agricultural activities have created such strong affinity between the best practices and usage of liquid cash which the predominant individuals who venture into farming mostly lack especially those in the rural farming communities; areas where poverty is more accentuated. They require capital investment to enable them make use of recommended and approved agricultural inputs and technological inclined methodologies to improve their crop yield to ensure long term food security.

For the above reasons and many others, have caused businesses and institutions that have the vision of breaking the spirals of poverty to venture into innovative concepts with a special focus on smallholders to improve their access to sustainable financial services by reducing the high cost associated with the evaluation of their credit creditworthiness. For example, the agricultural loan evaluation system (ALES) developed by Frankfurt School of Finance and Management functions as a comprehensive risk assessment model and loan calculator that determines the credit risk associated with a smallholder loan applicant by calculating the capital requirement and the most likely yield or income based on the applicant's agricultural activities. Similarly, in ensuring a complete cycle of value creation for smallholder farmers and agricultural investors, the agCelerant platform is a special platform deployed by Manobi Africa Group Company that serves as a digital smallholder value chain orchestration platform and implements innovative approaches to agricultural risk management to secure and transform agricultural investment by helping financial service providers to decrease their cash-out risks when lending to smallholder farmers or cooperatives (http://agcelerant.com). It combines physical asset management with digital solutions to ensure trust and scalability of service. Through the functionalities of the agricultural investment risk mapping and profiling (AIRMAP) service bundle, credit institutions and other stakeholders are able to control and reduce their financial risk cash-out and farmers are able to unlock access to credit, inputs, and market outlets.

1.1 Objective

Credit or risk scoring has not been widely applied in smallholder loans due to the substantial difference that exist in different ecological zones with their associated key level features, hence, making it more difficult to construct accurate scoring method. The goal of this project is to leverage risk scoring approaches with digital tools to track the performance of smallholder loans and to optimize their disbursement.

The specific objectives are to:

- design a pre-season risk scoring algorithm, encode it and test its robustness against data sets from rain fed sorghum systems (Nigeria and Ghana) and rain fed soybean systems (Ghana).
- develop a preliminary user guide for static (pre-season) risk scoring, with sample test data.

2. Review of Literature

In this chapter, we shall review various writing and research concerning smallholder loans and the associated risk scoring approaches employed to evaluate loan performance, and the key level variables that impact on the outcome of smallholder loans.

2.1 Concept of Smallholder Farmer

The varying viewpoints that exist in the literature on the definitions associated with 'smallholder' mostly bear very striking resemblances with the context of the general economic situation and the ecological zone under consideration. They often reflect the differences in the various stages of development across countries and the specific Continent in question. This notion is evident in several reports and studies including its influence and conception as an ambiguous term [38]. The same ideology was expressed in the definitions given in the papers [27], [33], [32], they argued that the notion of "small" varies in different contexts and that that explains why different terms such as "small-scale", "subsistence", "resource-poor" and "low-income" are frequently used interchangeably with smallholder.

Lipton [26] suggests that a smallholder is determined based on whether or not most of the labor is executed by members of the family. A similar viewpoint echoes in the definition given in the paper [42] which suggests smallholders be typically part of the informal economy and are generally understood as farmers who cannot solely manage the land they farm, and mostly rely on family labor, but at times engage other hands to supplement labor on their farms. According to Dixon et al [13], smallholders are endowed with limited resources relative to other farmers in the sector and are characterized by low returns and limited or outdated use of technology.

According to the FAO [6], The European and North American smallholder farmer is a well-identified decision-maker whose farm is clearly identified with a set of plots, possesses corresponding titles to the plots of land and supplements his/her farming activities with machinery or cattle. However, smallholders in most developing countries are associated with small farms that predominantly depend on family labor and represent a significant proportion of the rural poor whose economic importance is far from negligible since they are mostly the only source of food production.

2.2 Advancing Credit To Smallholders Through Financial Access

Prior to the establishment of the formal financial services, several hard-boiled reasons were linked to the conviction that smallholder farmers were too risky to be assisted with loans, hence, they were been charged with excessively high-interest rates by third-parties (including activities of intermediaries such as relatives and friends, traders, money lenders) who offered informal loan assistance [3], [45]. This was partly associated with the fear that their activities are under the mercy of extraneous factors which include the weather, which is very difficult to predict with certainty. The high-interest rates served as discouraging barriers for smallholders from accessing such loans in other that they do not severe their social ties with their lenders. This negatively impacted their productivity. In Nigeria for instance, the lack of adequate, accessible and affordable credit had a negative downturn effect which resulted in a systemic decline in the contribution of agriculture to the Nigerian economy [34].

The advent of the formal credit system made way for the few microfinance institutions that were established with the special mission to serve the interests of smallholders. For example, the German firm, Internationale Projekt Consult (IPC), successfully created an individual loan product for farm lending in El Salvador by using a credit-only approach to loans. The concept of the loan product was to analyze the cash flow of the farmer to evaluate the individual's repayment capacity. The success of the product saw its transfer to the Centenary Bank in Uganda where individuals who could repay their loans were graduated to larger loans. Similarly, Opportunity International maximizes the potential of smallholders by equipping them to afford high-quality seeds and fertilizer, to learn improved farming techniques, and to connect them with the best local crop buyers through its initiative across Ghana, Malawi, Mozambique, Rwanda, and Uganda. The roles played by these institutions help to bridge the gap in income and expenditure and the demand and supply of credit needed by farmers to have access to a loan in cash which they use in purchasing the most limiting production resources [29] they need. The absence of credit to poor families in emerging economies impedes the livelihoods of the people and further obstructs growth in their ventures [41]. Hence, advancing credit to these households could enable them to extend the size of their farm operations and further upgrade their level of livelihood by providing them with a consistent stream of income to sustain their livelihoods. This idea continues to stay relevant and consistent with the World Bank Group's global goal of attaining universal financial access (UFA) by 2020 by ensuring that everyone has full access to quality financial services at affordable prices in a dignified manner all over the world.

The establishment of microfinance institutions [21] in developing countries was birthed after international development agencies came to the realization that their foreign aids such as grants and subsidies were not enough to reduce poverty in developing countries [31]. They opted for the stimulation of entrepreneurial growth as the best replacement alternative and concluded that the ability to grant more people access to financial services could serve as a competitive force in stimulating the growth they sort after, thereby, reducing poverty among the world's poor populations [5], [10]. The institutions were expected to become profitable enough to sustain [56] their existence without relying on external funding. However, the initial ones experienced a massive decline when their external funding from the government's subsidies and donor funding ceased [16]. The decline was attributed to the loan default rate which exceeded the sustainability thresholds [49] which was a threat to their existence. Adejobi and Atobatele [4] claim that loan default is the bane of agricultural financing among smallholder farmers in developing countries. It is documented as the single biggest threat to microfinance institution's profitability and sustainability [19], [23], [44].

In the early 1990s, the Development Bank of Ethiopia (DBE) and Commercial Bank of Ethiopia (CBE) had to reduce their supply of input credit (such as fertilizer, improved seeds, herbicides, and farm tools) to farmers since their existence was severely threatened by massive loan defaults [46]. Loan default discourages financial institutions from refinancing the defaulting members, which puts them once again into a vicious cycle of low productivity, dwindling household income and food insecurity. Some of the reasons associated with loan defaults sometimes extend beyond causes that fall outside the remedy of both the lender and borrower such as death, family calamity and financial instability [51], but there also exist the ones which involve the borrower's attitude towards repayment and the lender's strategy to mitigate loan defaults [40]. Derban et al. [11] affirm the above factors and further relate the suitability of a loan product to the borrower as well as the systematic risks that arise from external factors such as the economic, political and business scope within which the smallholder operate as major contributors of loan repayment capacity.

In overcoming the challenge of loan defaults to smallholders in Ethiopia for instance, most of the micro-financing schemes provide loans to organized members, who are not required to put up physical collateral but operate in a group mechanism in which risks of non-repayment are transferred to the group [46].

This lending mechanism is the brainchild of the Grameen Bank. It is a microfinance organization and community development bank founded in Bangladesh. It is noted as the first bank in the world to use group lending without collateral among poor population in Bangladesh. The Varian conceptual framework [53] was later developed to give understanding to the bank's successful implementation of the group lending approach. This concept is adopted by several financial institutions when lending credits to its customers. The framework operates based on the three (3) major principles of joint liability, collective punishment, and voluntary group formation. However, Sileshi et al [46] contend that even though the Varian conceptual framework improves loan repayment rates and helps lower transaction costs when lending to the poor, the problem of poor loan repayment performance continues to persist.

In a study by Mphaka to explore the strategies that microfinance institutions' leaders use to reduce loan default in the base of pyramid market (BOP) [31], he identifies peer selection, information symmetry and trust among group members as the most influential factors that help to prevent loan defaults in group lending since members in the group would want to avoid social rejection. He further noted that regular monitoring visits by loan officers ensure timely advice to the groups whenever they need help and also remind them to honor their repayment schedules, and the borrowers become motivated to keep a good relationship with the lending institution. He encourages lenders to combine the self-selection principle from the Varian conceptual framework with other strategies and approaches that provide incremental benefits to good borrowing groups according to their loan cycles.

2.3 Factors Influencing Smallholder Loan Performance

Since farmers in general lack control over the influence of climatic conditions on their farms, the influence of quality inputs and the use of quality soil and vegetation are not enough to guarantee them successful farming seasons. Therefore, there is a need to predetermine and monitor their loan performances. In an attempt to examine and understand the major factors that influence smallholders' loan repayment performance among farmers of the Ondo State agricultural credit corporation in Nigeria by Okorie [35], four major determinants were identified. He claims that these determinants serve as pieces of evidence to the fact that agricultural loan repayment performance in developing economies could be improved when these factors: number of supervisory visits by credit officers, the profitability of borrowers enterprise, nature and the timeliness of loan disbursement are carefully managed. Similar studies conducted in the Southeast and Ogbomoso Agricultural Zone in Nigeria by Eze [18] and Oladeebo et al [36] respectively also established some positive relationships between the loan repayment performance and the educational level of farmers. According to Sileshi et al [46], an educated farmer is able to use modern agricultural technologies to boost production by employing farming activities based on cropping calendar, and manage resources properly which improves his or her chances of loan repayment.

In Ghana, Abankwah et al [2] conducted a study in the Ejura-Sekyedumasi and Mampong Municipality, they identified farmer's age, sex, household membership, income and farming systems to be the most significant factors that influence the loan repayment performance of farmers in these areas. They further identified post disbursement monitoring, low-interest rate and moratorium and repayment schedule as institutional factors that also contribute to smallholder farmers loan repayment performance. Wongnaa and Vitor [57] examined the determinants of loan repayment among yam farmers in the Sene District of Ghana, their results showed that the educational level of farmers, their years of experience, profit, age, supervision, and off-farm income positively have an influence on their loan repayment performance. They further noted that gender and marriage have no impact on loan repayment performance, and this buttresses the findings made by Chong, Morni and Suhaimi [9] when they indicated that being a male or female borrower does not influence your loan repayment performance. Conversely, an earlier implication had been made by Roslan and Karim [43] in Malaysia in the case of microcredit repayment to Agrobank, in which they claimed that female borrowers have higher loan repayment performance due to their economic empowerment as a result of the loan and their culture of financial discipline compared to male borrowers.

In identifying the critical determinants of loan repayment performance in Malaysia, Roslan and Karim empirically quantified the extent to which some determinants affect loan repayment performance proven by their high correlations. The discovery of their work narrowed around the amount of loan which they describe as the most crucial factor that affects the loan repayment performance among borrowers. They believe that giving out reasonable loan amount to borrowers encourage them to pay hence improving their loan repayment performance. More so, borrowers with huge loan amounts are frequently visited by their lenders to monitor their activities but borrowers with small amounts do not enjoy this service this much. According to Brehanu and Fufa [7], small group lending is a viable approach that improves smallholder loan performance. They identified the total land and livestock holding size, the experience of the farmer in the use of agricultural extension services, his or her engagement with extension agents and the amount of income generated from activities outside his or her farming occupation to be the factors that positively impact on a farmer's loan repayment performance.

2.4 Risk Scoring Approaches For Smallholder Loans

Effectively rating the risk associated with a smallholder loan application in a timely manner is crucial to the control and prevention of loan defaults. This is a fundamental task undertaken by all credit facilities before credits are given out. Risk or credit scoring helps to provide information on not just what is driving loan defaults but also helps decision-makers to improve the quality of their credit products. The author of this report is in the known with regards to the evidence [39] that exist in literature concerning the negative biases associated with how some data-driven models can discriminate against certain populations or groups such as women and youth as a result of the inbuilt biases from the historical data on which such models are built. In a study of credit assessment models, 'Credit assessment models for farm borrowers: A logit analysis', Miller and LaDue [28] used weighted logit models to discriminate between good borrowers and bad borrowers. They examined the financial ratios of liquidity, capital efficiency, profitability, operating efficiency and solvency as explanatory variables. They observed the model classify loan applicants better than the naive model and the discriminant analysis approach that has been implemented in previous studies because of its objective ability to differentiate between borrower default risk. They identified the financial measure of liquidity, operating efficiency and the profitability of the farmer's activity to indicate his or her quality. Similarly, Escalante et al [15] observed the usefulness of the logit model as a significant credit assessment method to establish the repayment rating or capacity in loan approvals.

Limsombunchai et al [25] in Thailand, assumed that the probability of a good loan follows the logistic distribution and is a function of the borrower's characteristics, credit risk proxies, dummy variables, and the relationship related indicators. They constructed two separate models to predict farmers' creditworthiness and default risk using logistic regression and artificial neural networks (ANN). They hypothesized that the farm type, loan type, loan size and lending year have systematic effects that influence a borrower's credit risk. They claim that a borrower who has cash crops (horticulture) as the major production would require a smaller amount of credit than those into other farm types. Hence, farmers who are into such products have a higher chance of obtaining loans than others, because their activities are short-termed making them less risky to default than those of medium-term or long-term production. The total asset value, capital turnover ratio and the duration of the borrower's relationship with the bank identify as the most important factors in determining the creditworthiness of the borrower according to logistic regression. Borrowers with long relationships with the bank and also have a higher gross income to total assets tend to be in a higher position to default on their repayment. The ability of the ANN model to reduce its misclassification error on out-of-sample data in comparison with the logistic regression makes it a better algorithm. When a borrower is incorrectly considered creditworthy, when in fact, the institution should not give the borrower a loan, a type I error is committed. The inefficiency on the part of logistic regression was reflected in the higher type I error rate compared to the lower type I error rate by the ANN model. Since the cost of classifying a bad loan as a good loan (Type I error) is more expensive and significantly greater than the cost of misclassifying a good loan as a bad loan (type II error), their work supports the ANN model as the best credit scoring model with the lowest misclassification costs in screening agricultural loan applications in Thailand.

According to the study conducted under the Nigerian Agricultural Cooperative and Rural Development Bank (NACRDB) by Onyenucheya and Ukoha [37] in Abia State, there exist some positive relationships between the creditworthiness of a farmer and the years of experience, age, farm size, educational level, and the total operating expenditure-to-income ratio. Their result reflects the importance associated with discriminant analysis as a useful model in rating the creditworthiness of agricultural loan applicants. Similar results were earlier obtained by Turkey [50] when he reviewed four alternative credit scoring models for agricultural loans. The predictive accuracy of the discriminant analysis was superior in the midst of the logit, probit and linear probability models by achieving 71.5% accuracy followed by logit, probit and linear probability with accuracies 69.7%, 69.4%, and 67.1% respectively.

In overcoming the challenge of sophisticated agronomic know-how before evaluating the credit rating of agricultural loan applicants, the international advisory services (IAS) employs a very much different approach in scoring the risk associated with an agricultural loan applicant. It relies on the ALES (risk scoring system) that functions on the core elements of tech-cards to determine credit risk by using a scoring model that employs gross margin projections, agricultural production parameters and risk analysis of the farming enterprise. The credit risk is calculated based on the qualification and effectiveness of the applicant to operate an agricultural production, the potential of the applicant to produce high yields that is enough to cover all costs and further deliver acceptable gross margins and finally, the potential of the loan applicant to reduce credit risk by been able to satisfy the least safety margin levels through his or her production, costs and income taking into account the risk appetite, risk policy and the lending strategy of the financial institution within which it serves.

The scoring model of ALES incorporates parameters from the financial institutions such as the gross profit or net profit upon which lending decision is to be inferred from, the threshold percentage of capital need of applicant that the loan must cover, historic past yields of the applicant in comparison with regional averages, and applicant's minimum and maximum net income and the percentage of the net income that is expended as living expenses. Qualitative analysis is then performed making use of the applicant's background and agronomy data to determine his or her status as either being a good or bad farmer. General information covering the applicant's crop production experience with the cultivated crop, years of experience as a farmer, frequency of same crop grown on the same piece of land, ownership of the farm area compared to a rented farm area, ownership of equipment and machinery used on the farm, the irrigation status of the farm land and the possible source of water for the irrigation purpose, the total farm area used for the cultivation and the registration status of the area as a farming area.

Quantitative analysis is also performed on variables that contribute to yield changes. These variables speak to the type of seeds used, the number of times irrigation is done in a cropping cycle, the number of times fertilizer application for phosphorus, insecticides, and fungicides are applied in a single cropping cycle and the amount of nitrogen application in a single cropping cycle. The type of soil used, weed control method and how the decision for pest control is arrived at. All these are evaluated together to ascertain the productivity changes the applicant is likely to encounter. Special attention is given to the estimate of future selling prices and future income and expenses by updating data regularly to conform to the changes and expectations in the local and international markets. When ALES receives information regarding the loan amount applied for, maturity, purpose of the loan, non-farming income and expenses of the applicant, applicant's current bank indebtedness, it automatically calculates the working capital need of the applicant and the most likely yield based on the qualitative and quantitative analysis conducted to decide on the loan decision of the agricultural activity. It further projects the applicant's income-expense through a summary report with the suggested loan limits for each type of loan product applied for.

Fast forward into the 21st century, artificial intelligence(AI) and machine learning approaches are taking precedence in almost every sphere of decision-making tasks due to their abilities to learn from data through the power of automation. They are noted for showing promising results in areas in which they have been applied, and this serves us right to consolidate them with lessons on credit scoring. For example, the results from Tsai et al [48] show that 'Classification + Classification' hybrid model based on the combination of logistic regression and neural networks can provide the highest prediction accuracy for credit rating. Their results identify the logistic regression as the first classifier in combination with the neural networks (second classifier) to produce the optimal credit rating model. Similarly, Charpignon et al [8] demonstrates the superiority of ensemble models such as random forests and gradient boosting trees over multiple logistic regression when it comes to predicting loan default. The superiority of the gradient boosting algorithm lies in its ability to rely on incremental minimization of its error term, hence, improving the accuracy of its predictive function. On the other hand, random forest averages the results from multiple deep decision trees in order to reduce the variance. These algorithms hold some promise in identifying potential default customers in favor of the lending institution.

Simumba et al [47] proposed an alternative scoring approach based on risks which include fraud, interaction, and revenue for the creation of credit scoring models for financially excluded populations such as smallholder farmers about whom no financial history is available for credit evaluation. They documented three vital issues as the most important to be tackled in such instances which include the collection of appropriate data to perform credit risk evaluations, the design of alternative scoring factors, and the selection of apt data analysis methods to evaluate these scoring factors. They proposed the merger of mobile applications, credit scoring and data analysis as the multidisciplinary solution. In their design of the alternative scoring which they claim is the most crucial among the three vital issues, they examined three scoring approaches with the selected scoring factors (variables) based on non-financial data from smallholder farmers.

Multiple logistic regression was trained and tested on this data to evaluate the selected factors. A probability threshold was set and used to determine the class a borrower belongs to, where a probability above the threshold implied a "good" borrower and one below the threshold also implied a "bad" borrower. Support vector machine (SVM) models (SVM with a linear kernel & SVM with radial basis function kernel) were also trained and tested on the data to evaluate the factors using SVM's hyperplane as a separator between the data points to differentiate them into classes while maximizing the distance between the classes. Assessment of the models was done using their area under the receiver operating characteristics (ROC) curve from the training data. The support vector machine with a radial basis function kernel outperformed the support vector machine with a linear kernel and was ultimately identified to be the best performing model for the development of the fraud score. Multiple logistic regression was outperformed in all cases.

In summary, several studies have broadly explored features that influence agricultural loan performance and risk scoring approaches that have the capability to de-risk agricultural loan applications by smallhoders through the use of farmers' characteristics, financial credit history and key farm-level features. Moreover, every research study has emphazied the influence and inevitable need for financial credit history before one can determine the loan performance of an applicant. This might be that it is because too many findings and research have already skewed the fact in favor of this feature requirement, hence, causing researchers to pay less attention to other factors that could be used in the absence of historic credit information. To the best of the knowledge of the researcher, no study has been conducted especially in Africa where financial inclusion is a huge problem, and many of the credit applicants (smallholders) lack past financial credit information. Therefore, there is a need to fill in the gap of knowledge by designing risk scoring models that do not make use of historic financial credit information and monitor the loan performance thereof.

3. Mathematical Framework of Algorithms

This chapter provides a brief description of the conceptual and mathematical framework that underpins the decision-making rule of the selected algorithms, namely: logistic regression, naive Bayes, support vector classifier, decision tree, and the light gradient boosting algorithms.

3.1 Logistic Regression

The logistic regression [22] belongs to a generalized linear model (GLM) class of algorithms, and forms a special case of the linear regression when the outcome variable is binary and categorical. GLMs assume a linear relationship between our target variable (CreditWorthiness of the farmer) and the independent variables using the maximum likelihood estimation (MLE) concept.

The fundamental equation of generalized linear models is given as:

$$g(f(x)) = \alpha X + \beta + \epsilon$$

- g: the link function that 'links' the expectation of the target varaible (creditworthiness of the farmer) to the linear independent variables.
- X: represents the independent set of farmers' characteristics, their farming activities and farm-level features.
- α, β & ϵ: represent the weights or coefficients and error terms respectively used in the linear combination of the independent variables.

$$f(x) \quad = \quad \alpha X + \beta = z \tag{3.1.1}$$

The link function, g, assumes that:

- the probability of a smallholder farmer who is not creditworthy (probability of success), p, is always positive and lies between 0 and 1.
- the probability of a creditworthy farmer (failure in this case) is also given as $1 - p$.

Since probability must always be positive, We denote $g(z)$ with p and then put the linear equation (4.1.1) in exponential form

$$p \quad = \quad \exp\left(\alpha X + \beta\right) \tag{3.1.2}$$

and then divide p by a number greater than p in order to satisfy the first assumption of the link function.

$$p \quad = \quad \frac{\exp\left(\alpha X + \beta\right)}{\exp\left(\alpha X + \beta\right) + 1} \tag{3.1.3}$$

Hence, the probability of failure, $q = 1 - p$ is given as:

$$q \quad = \quad 1 - \frac{\exp\left(\alpha X + \beta\right)}{\exp\left(\alpha X + \beta\right) + 1} \tag{3.1.4}$$

We divide p by q to obtain the odd ratio:

$$\frac{p}{q} = \frac{p}{1-p} = \frac{\frac{\exp{(\alpha X + \beta x)}}{\exp{(\alpha X + \beta)} + 1}}{1 - \frac{\exp{(\alpha X + \beta)}}{\exp{(\alpha X + \beta)} + 1}} = \exp{(z)} \tag{3.1.5}$$

We take logarithm of both sides. This transformation allows us to model a non-linear association in a linear way, and this yields:

$$\log{(\frac{p}{1-p})} = \log{(\exp{(z)})} \tag{3.1.6}$$

$$\log{[\frac{p}{1-p}]} = z \tag{3.1.7}$$

$$\phi(z) = \frac{1}{1 + \exp{(-z)}} \tag{3.1.8}$$

This is the equation used in Logistic Regression. Here $(p/1-p)$ is the odd ratio. Whenever the log of the odd ratio is found to be positive, the probability of success (a smallholder identified as not creditworthy for a loan) is always more than 50%. Figure 4.1 is an image of the logit function.

The Logistic Function

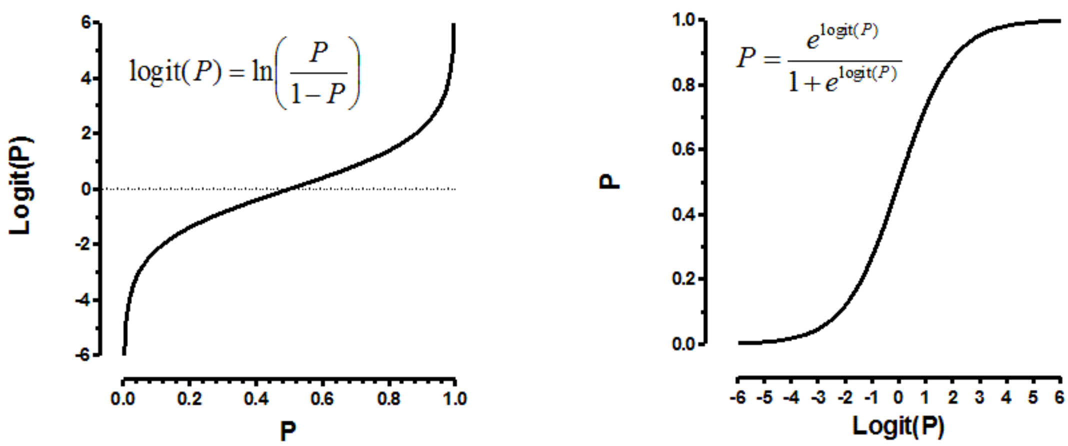


Figure 3.1: shape of the logit function

3.2 Naive Bayes

The naive Bayes algorithm operates with an assumption of independence [58] among all the independent variables based on the principle of the Bayes' Theorem. Bayes theorem is built on top of conditional probabilities and comes into effect when several events denoted $\mathcal{H}_i$ form an exhaustive set with another event, also dented $\mathcal{F}$. The event $\mathcal{F}$ can be defined as:

$$\mathcal{F} = \sum_{i=1}^{n} \mathcal{F} \cap \mathcal{H}_i \tag{3.2.1}$$

Hence, probability of $\mathcal{F}$ can be written as:

$$P(\mathcal{F}) = \sum_{i=1}^{n} P(\mathcal{F} \cap \mathcal{H}_i) \tag{3.2.2}$$

But since;

$$P(\mathcal{F} \cap \mathcal{H}_i) = P(\mathcal{F}|\mathcal{H}_i) \times P(\mathcal{H}_i) \tag{3.2.3}$$

Hence, replacing $P(\mathcal{F})$ in the above equation of conditional probability produces the equation of the Bayes Theorem represented as:

$$P(\mathcal{H}_i|\mathcal{F}) = \frac{P(\mathcal{F}|\mathcal{H}_i) \times P(\mathcal{H}_i)}{(\sum_{i=1}^{n}(P(\mathcal{F}|\mathcal{H}_i) \times P(\mathcal{H}_i)))} \tag{3.2.4}$$

The Bayes Theorem enables us to calculate:

1. the posterior probability of $\mathcal{H}_i$ denoted by $P(\mathcal{H}_i|\mathcal{F})$ given
2. the predictor prior probabilities denoted by $(\sum_{i=1}^{n}(P(\mathcal{F}|\mathcal{H}_i) \times P(\mathcal{H}_i)))$
3. $P(\mathcal{H}_i)$ represent the class prior probability, and
4. $P(\mathcal{F}|\mathcal{H}_i)$ is the probability of likelihood of the predictor given the class.

3.3 Support Vector Machine

The support vector machine (SVM) [52] executes its classification tasks by selecting an optimal hyperplane that best segregates a given data set into the number of distinct target classes prior to the maximization of its margin. An SVM model represents data points in space so that the points are mapped into high-dimensional feature spaces such that the different individual classes of the data are clearly separated by a decision boundary that is as wide as possible. Aside its ability to perform linear classification tasks, it also relies on kernel trick to perform efficient non-linear classifications.

Given a training data set composed of feature space $\mathcal{S}_i$ and label y_i paired together as $(\mathcal{S}_i, y_i)$, where $i \in \mathcal{R}_n$ for the i^{th} observation and $y \in \{0,1\}^l$

Mathematically, SVM requires the solution of an optimization problem that:

$$\begin{aligned} \min_{\omega,\mu,\sigma} \quad & \frac{1}{2}\omega^T\omega + K\sum_{i=1}^{l}\sigma_i \\ \text{subject to} \quad & y_i(\omega^T\rho(\mathcal{S}_i) + \beta) \geq 1 - \sigma_i, \\ & \sigma_i \geq 0. \end{aligned} \tag{3.3.1}$$

The ρ function maps the training vectors $\mathcal{S}_i$ into a high-dimensional space and finds a linear or non-linear hyperplane (kernel eg. radial basis function) that best separates the training vectors into their respective classes with the maximal margin using K as the penalty for the error term σ_i.

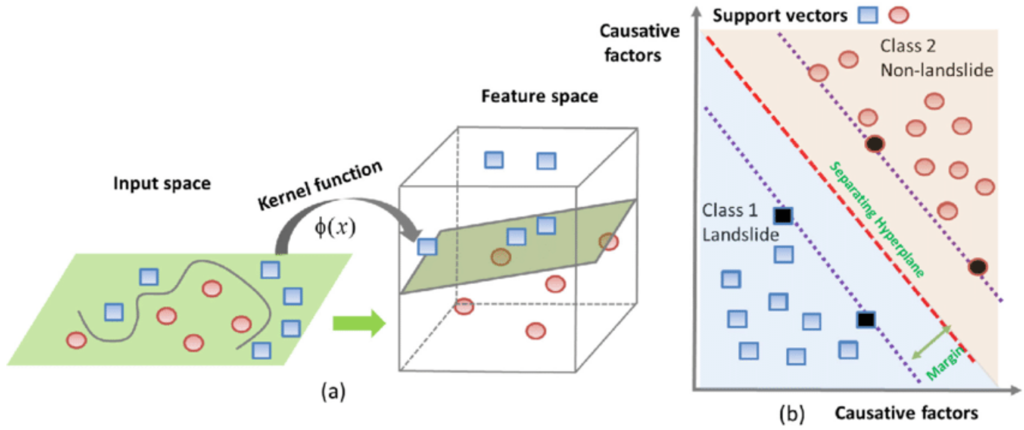


Figure 3.2: Illustration of support vector machine (SVM) principle: (a) Input space is mapped to the feature space with the help of a kernel function; (b) Separating hyperplane and margin for classification [14]

3.4 Decision Tree

The decision tree algorithm differs from other machine learning techniques in the sense that, it splits the entire sample data (root node) into two or more homogeneous sets based on the most important splitters in the input variables. The two homogeneous sub-nodes become decision nodes when they split into further sub-nodes. When these sub-nodes from the decision nodes refuse to split further they become leaves or terminal nodes. This technique works for both categorical and continuous input and output variables. Decision trees use different algorithms to decide on which node to split into two or more sub-nodes. The splitting of the sub-nodes into further child-nodes helps to increase the homogeneity in the resultant sub-nodes thereby increasing the purity of the node with respect to the target variable. Splitting is performed on the nodes of all available variables and then the selection is done in favor of the split that produces the homogeneous sub-nodes.

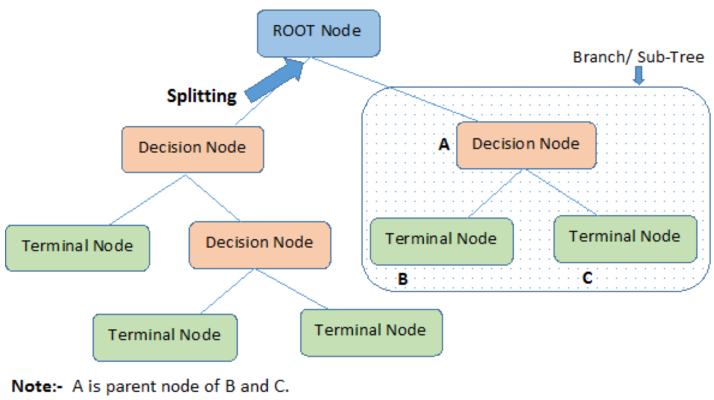


Figure 3.3: Decision trees

Information gain measures the quantity of "information" a feature holds about the target class. During the selection of nodes on which splitting could be done to produce homogeneity of the sub-node, the entropy of the node is measured to define the degree or level of impurity, disorganization or uncertainty associated with the node (data). An entropy of zero suggests a completely homogeneous sample, and if the disorganization in the node is evenly divided then its entropy is 1. The information gain algorithm relies on the assumption that the lesser the entropy, the better it is for that split to be done on the associated node. Hence, it chooses the split which has the lowest entropy compared to the parent node and other splits.

The entropy of a sample data, $\mathcal{S}$, is defined as:

$$\mathcal{H}(\mathcal{S}) = \sum_{i \in X} -p_i \log_2 p_i \tag{3.4.1}$$

where, p_i is the frequentist probability of an element/class 'i' in sample data.

Mathematically, information gain can be simplified as:

$$InformationGain(IG) = \mathcal{H}(S_{parent}) - [weightedaverage] * \mathcal{H}(S_{children}) \tag{3.4.2}$$

3.5 Light Gradient Boosting Model (Light GBM)

Light GBM is an ensemble model of decision trees that are trained in sequence [20] and falls under the class of boosting algorithms. Boosting algorithms convert weak learners (base learners) into strong learners by combining the prediction of each base learner using the weighted average after considering the predictions that have a higher vote. Weak learners are identified by their weak prediction rules after each iterative process of applying base learning algorithms with a different distribution. These weak prediction rules are further combined into a single strong prediction rule using the boosting algorithm. The light GBM derived its name from the above process. It is 'light' due to its super-fast nature. Its high-performance gradient boosting framework makes it more useful for training on a limited amount of data. It is ideal when the training model is constrained by time and also requires less expertise in tuning its parameters.

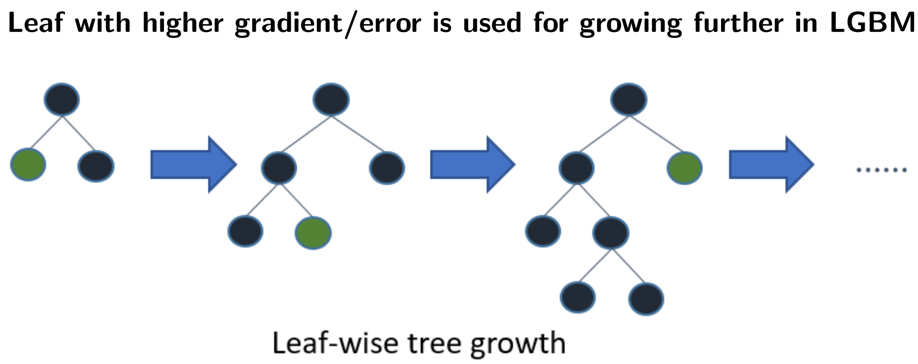


Figure 3.4

This algorithm employs a leaf-wise split of trees which enables it to reduce more loss in comparison to those that make use of the level-wise split. The gradient-based one-side sampling (GOSS) technique is a novel approach that is used by the algorithm to filter out the data samples for finding a split value. The technique maintains all the data samples with large gradients (the slope of the tangent of the loss function) and randomly samples on the data samples with smaller gradients. Splits are performed at each node with the most informative feature (with the largest information gain) which is measured by the variance after splitting. There is always an increase in the model complexity which may result in overfitting, hence, the algorithm overcomes this by using the max-depth parameter to control the depth of splitting.

4. Data and Experiments

In this chapter, we provide a description of the data we used in this report taking into account the various characteristics associated with the different features that accompanied our data set as obtained from the agCelerant platform. Experiments were performed using the agCelerant data to design five (5) machine learning algorithms which were compared based on their recall score as the performance metric. Models identified with higher recall in comparison to others are deemed to hold the best promise of reducing false positives (incorrectly identifying a farmer as creditworthy, when in fact, he or she is not).

4.1 Data Description and Feature Characteristics

The relevance and accuracy of the data obtained for this report is in agreement with the formal approval of Manobi Africa PLC and its partners. Data was obtained from the agCelerant platform and covers rain-fed information about smallholder farmers' activities on sorghum and soybean production in the northern regions of Ghana. All data regulations that pertain to the confidentiality of the individuals from whom data were obtained are strictly adhered to in conformity with standard data practices and policies. Data from 980 smallholder farmers are used in this report and spans the sowing period between April and November 2017 and harvesting period between June 2017 and April 2018. Originally, the data were exported into comma-separated values (CSV) format from the agCelerant platform and existed in separate categories and covered information on the village, farmer, plot and agricultural season as shown in Figure 4.1.

Directory in agCelerant

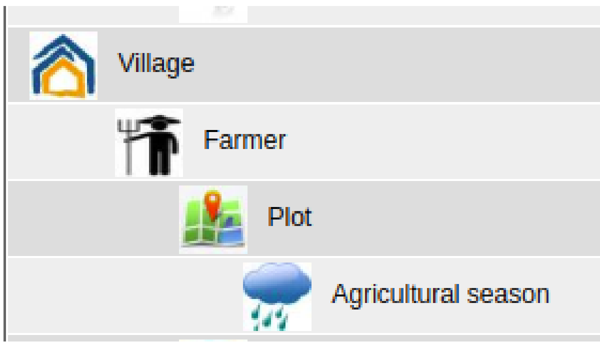


Figure 4.1: Directory structure in agCelerant

Data is composed of 451 individual sorghum farmers and 529 soybean farmers. The variables considered are similar to those identified by ALES and Henning et al [17] except for the exclusion of past financial credit history which was absent in the data by default, and are characterized by 162 original feature columns after merging data from farmer, plot and the agricultural season using id and parent id as keys. This section briefly classifies key data features into personal information, production activities, plot information, post harvest records and commitments and expected income.

4.1.1 Personal (Farmer) Information. A total of 980 smallholder farmers were observed using key variables that provide information on their personal and family lives. Features considered to be personal include the age of the farmer, gender, years of farming experience and educational qualification. Others include the number of spouses of the farmer together with their individual activities engaged in, the number of children of the farmer with their respective age categories and activities engaged in. A farmer's means of transportation, access to water, health care and electricity as well as the building material used for his or her housing facility were also considered.

Out of the total sample data, 70% were males with an average age of 45 years and 22 years of working experience on average. The remaining 30% were females with their age averaged at 61 years and 14 years of working experience. It was observed that 36% of the farmers were without spouses, but remaining 64% were married with exactly 1 spouse on average. However, only 6% were observed to have been engaged with more than 2 spouses. Amongst the farmers identified as literates, 52.49% of them had obtained educational training up to the primary level, 37.57% to the secondary level and less than a percentage educated to the tertiary level. Their predominant means of transportation were identified as bicycle, foot and motorcycle with their respective percentages ordered as 52.49%, 30.61% and 16.94%. It was revealing to know also that only 2.55% of the farmers had access to very good health care, 36.43% and 36.80% could access good and average health facilities. The remaining 26.22% either had no access to health care at all or considered their access as a poor one. 'Water is life', just as the saying goes. Our exploratory analysis further revealed that 8.16% of the farmers were drinking from lake, 11.53% and 17.04% were accessing well water and rivers respectively. Only 19.18% had access to tap water and the remaining 44.08% had access to hydraulic pumped water sources. The above revelation somehow feeds into the farmers obvious accessibility to electricity. It was noted that 56.12% of the farmers are connected to the national grid and just a little above a percentage were accessing solar panels. 42.76% had no access to electricity at all.

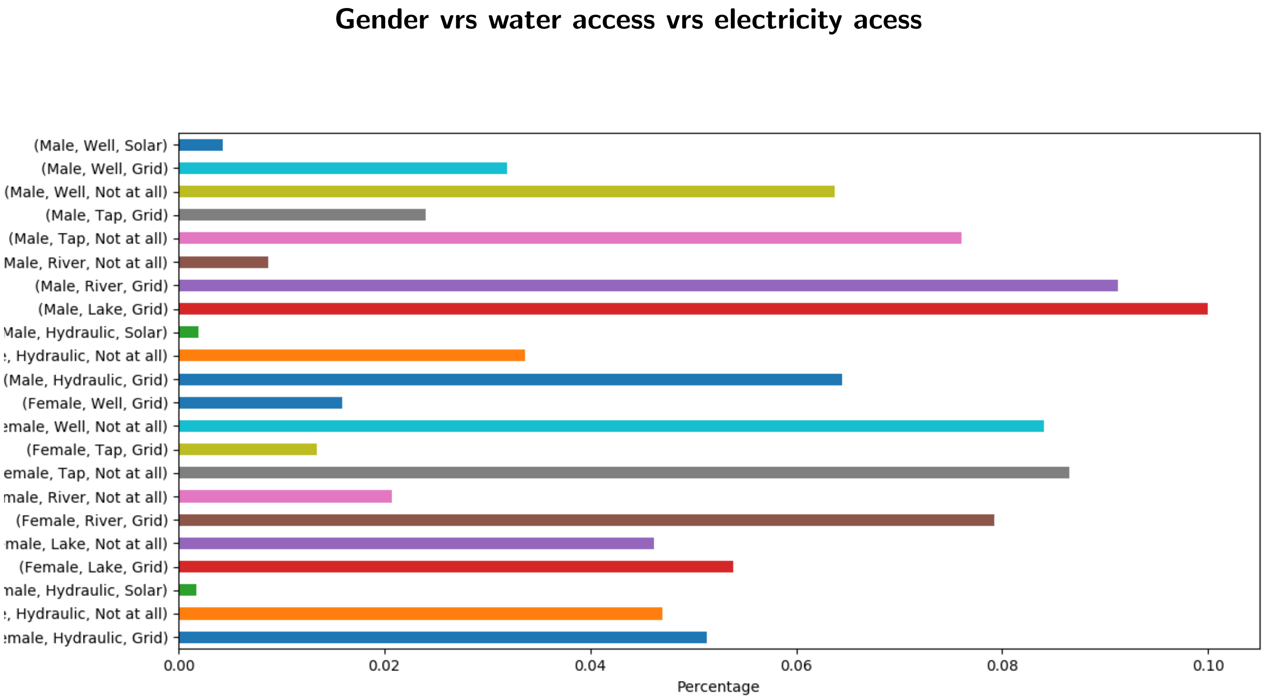


Figure 4.2: Male Applicants who drink from lakes, rivers and hydraulics have a higher chance of being connected to the grid on average than those who drink from tap and well water.

4.1.2 Plot Information. Land is rated by far as one of the most important resources in agricultural production in comparison to other productive resources. The status of its fertility based on the topography and location, soil type, soil color and soil quality, in association with its size and other attributes make it an inevitable resource in every agriculture-related activity. We observed the mean total farm plot declared by farmers to be 3.56 ha, but farmers on average cultivated 1.21 ha of the plots they had declared in a single agricultural season. On average, 99.39% of the soil used by the farmers were of good quality, only 0.61% were considered bad. 71.12% of the soil used were identified as brown, 23.67% were black. The remaining were attributed as belonging to other soil colors which were difficult to describe. Almost half of the soil cultivated by the farmers were of the loamy type. Sandy soil was about 24.90% and clay a little above 2%. The remaining soli types constituted a blend of sandy and loamy and clay and loamy.

Soil type vrs Soil quality

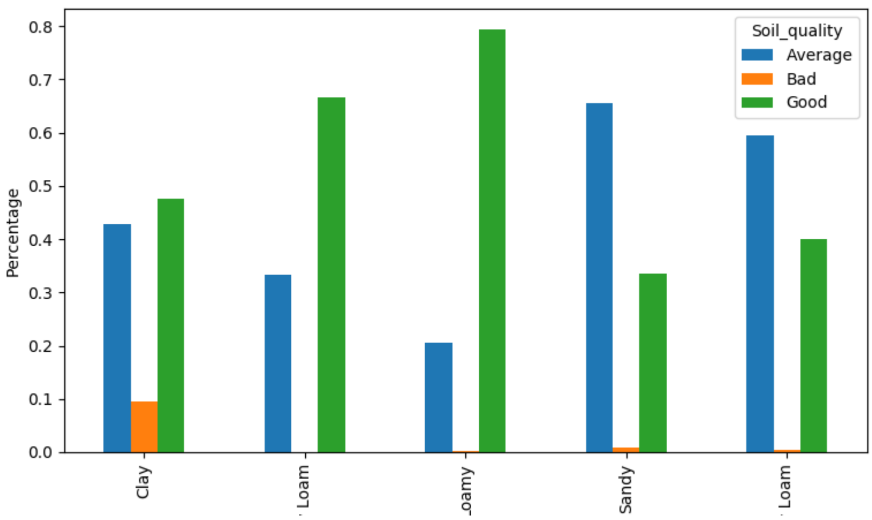


Figure 4.3: Almost all the soil types were of good and average types on average. Only clay has quite a number of bad qualities.

4.1.3 Agricultural & Production Activities. Since climate change and its related factors present inherent uncertainties and risks that influence crop yield for which farmers naturally lack absolute control over, therefore, there is the need for them to take good charge of what they have influence over in order to benefit from the outcome of their sweat in the fields, by way of engaging in result-oriented farming practices and activities. Farming practices have been identified as a strong contributor to the potency of a smallholder to either repay or default on a loan application. For this reason, the specific types of crops cultivated by individual farmers both as either primary or secondary crops were considered. The varieties of their chosen crops in addition to their motivation for the their selection were also included. As obvious and shockingly as it may be, majority of the smallholders were motivated by the income that they generated out of their produce at the end of the agricultural season.

It was observed that 70.92% of the farmers did not fertilize their cultivated farm plots. Only 29.08% did apply fertilizer ranging from NPK, TSP, NPK+Urea and DAP+Urea. In the control of pests and insects on the farms, it was surprisingly noted that only 0.92% of the farmers applied insecticides on their farm, the remaining 99.08% did not. Fungicide treatment also received another shocking welcome by the farmers since it was noted to be practiced by only 0.41% of the farmers throughout the agricultural season. 28.47% of the farmers used herbicides to control unwanted vegetation such as weeds. The remaining resorted to manual weeding and other forms which were not specified except for the use of weeder by 0.31% of the farmers.

Prior to the start of every farming season, farmers are expected to clear their farmlands which requires a lot of resources which may include the use of self-propelled and heavy equipment in achieving this objective. They may fall on tractors, plow, harrowers to even off their plowed plots, ridges for creating sand walls and a host of other implements. In the case of this data, we noticed that 40.92% of the smallholders had access to tractors to supplement their activities. Only 0.41% had acces to harrowers, 2.76% used threshing machine on their farms, 28.98% plowed their lands and 29.90% used sprayers as well. It was also observed that 23.47% relied on animal power as their source of traction on the farm. For those that engage the use of animals, 27.24% rely on bovines, 9.29% on equine but the remaining failed to disclose the animals they use for their traction.

Percentage of means of traction vrs variety sown

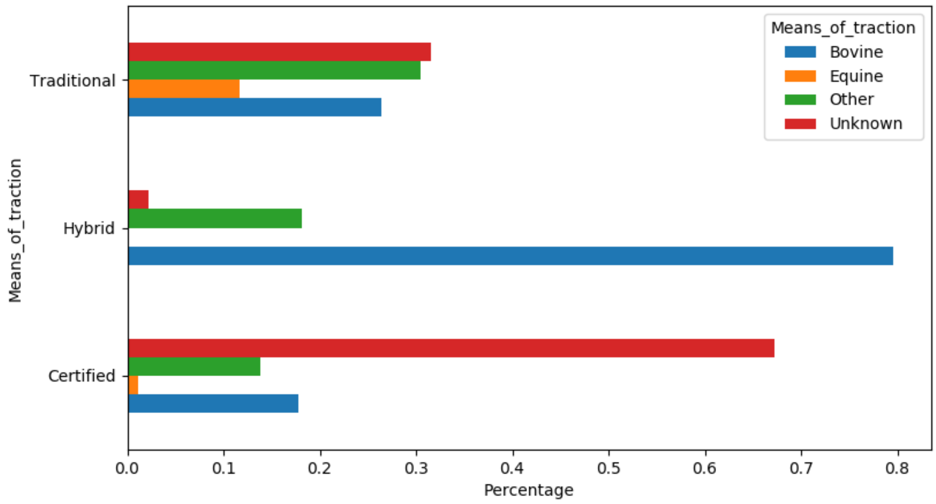
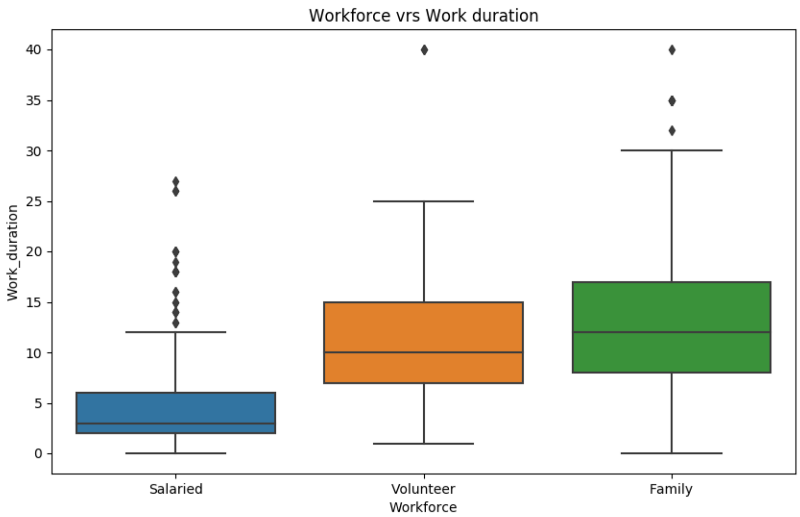


Figure 4.4: Farmers who sowed the hybrid variety of seeds mostly relied on bovine. Those who used certified seeds mostly do not disclose their means of traction. Traditional varieties have a fair share between the various means of traction.

It was identified that the majority (77.76%) of the farmers relied heavily on traditional seeds as their choice of seed varieties. 17.76% embraced the used of certified seeds and only 4.49% also used hybrid seeds. The remaining failed to disclose their choice of seed variety sown. On the average, 5.88kg of seeds were sown by each individual with 3 seeds per hole during the time of sowing while ensuring a 21.77 mean spacing between rows. 61.94% of the workforce used by the farmers are family members with 21.63% hired as salaried workers and the remaining 16.43% been volunteers. The family workers were observed to have the highest number of work duration followed by volunteers and then salaried workers spending the least time. This makes sense in the sense that salaried workers are constrained by their agreed time duration on the far while family and volunteers can work as long as they wish to.

On average, all farmers who made use of inputs, applied the recommended dose having in mind the usefluness of its benefits. It was again noted that about 79.39% of farmers expressed no interest in insurance, only 17.86% had expressed their interest and the remaining 0.03% been undecisive. Majority (46.33%) of the farmers who had subscribed to agricultural services engaged the services of the Ministry of Food and Agriculture (MoFA) and the remaining engaging the services of other parties. These services were generally rated as been good by the farmers with an average visits of 1 to 2 per year. 59.08% had no interest in personalizing their advisory services, only 26.63% of them had expressed interest and the



remaining not decisive about their interest.

4.1.4 Post Harvest & Forecasting Records.

4.1.5 Engineered Features.

4.1.6 Committments & Expected Income.

4.1.7 Credit Rating.

References

[1] Un sustainable development goals. https://www.un.org/sustainabledevelopment/sustainable-development-goals/.

[2] V. Abankwah, D. Awunyo-Vitor, and Y. Seini. Economic and institutional elements of loan repayment capacity of smallholder farmers in the transitional zone of ghana. *Asian Journal of Agricultural Extension, Economics & Sociology*, 9(3):1–13, 2016.

[3] D. W. Adams, D. H. Graham, J. D. Von Pischke, et al. Undermining rural development with cheap credit. *Undermining rural development with cheap credit.*, 1984.

[4] O. Adejobi and J. Atobatele. An analysis of loan delinquency among small scale farmers in south-western nigeria: application of logit and loan performance indices. *East African Agricultural and Forestry Journal*, 74(3):149–157, 2008.

[5] B. Armendáriz, B. D'Espallier, M. Hudon, and A. Szafarz. Subsidy uncertainty and microfinance mission drift. *Available at SSRN 1731265*, 2013.

[6] J.-M. Boussard. *The impact of structural adjustment on smallholders*. Number 103. Food & Agriculture Org., 1992.

[7] A. Brehanu and B. Fufa. Repayment rate of loans from semi-formal financial institutions among small-scale farmers in ethiopia: Two-limit tobit analysis. *The Journal of Socio-Economics*, 37(6):2221–2230, 2008.

[8] M.-L. Charpignon, E. Horel, and F. Tixier. Prediction of consumer credit risk. *Standford University*.

[9] F. Chong, F. Morni, and R. Suhaimi. Demographic factors and repayment performance of nbfi customers in kuching. In *2010 International Conference on Science and Social Research (CSSR 2010)*, pages 1340–1345. IEEE, 2010.

[10] S. K. Dary and H. Issahaku. Technological innovations in microfinance institutions in the three northern regions of ghana. 2013.

[11] W. K. Derban, J. M. Binner, and A. Mullineux. Loan repayment performance in community development finance institutions in the uk. *Small Business Economics*, 25(4):319–332, 2005.

[12] B. M. Dioula, H. Deret, J. Morel, E. Vachat, and V. Kiaya. Enhancing the role of smallholder farmers in achieving sustainable food and nutrition security. In *International Conference on Nutrition-better nutrition better lives, Paris*, 2013.

[13] J. Dixon, K. Taniguchi, and H. Wattenbach. Approaches to assessing the impact of globalization on african smallholders: Household and village economy modeling. *Proceedings of a working session on Globalization and the African Smallholder Study. FAO (Agricultural Support Systems Division [AGS] and Agricultural and Development Economics Division [ESA]) and the World Bank. Rome: Food and Agriculture Organization of the United Nations*, 2003.

[14] J. Dou, A. P. Yunus, D. Tien Bui, M. Sahana, C.-W. Chen, Z. Zhu, W. Wang, and B. T. Pham. Evaluating gis-based multiple statistical models and data mining for earthquake and rainfall-induced landslide susceptibility using the lidar dem. *Remote Sensing*, 11(6):638, 2019.

[15] C. L. Escalante, R. L. Brooks, J. E. Epperson, and F. E. Stegelin. Credit risk assessment and racial minority lending at the farm service agency. *Journal of Agricultural and Applied Economics*, 38(1):61–75, 2006.

[16] A. for Financial Inclusion. (2014). Rwanda's financial inclusion success story: Umurenge saccos (afi case study).

[17] J. Henning. *Credit scoring model: Incorporating entrepreneurial characteristics*. PhD thesis, University of the Free State, 2016.

[18] C. Ibekwe. Determinants of loan repayment under the indigenous financial system in southeast, nigeria. *The social sciences*, 2(2):116–120, 2007.

[19] B. Ibtissem and A. Bouri. Credit risk management in microfinance: The conceptual framework. *ACRN Journal of Finance and Risk Perspectives*, 2(1):9–24, 2013.

[20] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, pages 3146–3154, 2017.

[21] D. Kent and M. T. Dacin. Bankers at the gate: Microfinance and the high cost of borrowed logics. *Journal of Business Venturing*, 28(6):759–773, 2013.

[22] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein. *Logistic regression*. Springer, 2002.

[23] O. Kodongo and L. G. Kendi. Individual lending versus group lending: An evaluation with kenya's microfinance data. *Review of Development Finance*, 3(2):99–108, 2013.

[24] C. Kremen, A. Iles, and C. Bacon. Diversified farming systems: an agroecological, systems-based alternative to modern industrial agriculture. *Ecology and Society*, 17(4), 2012.

[25] V. Limsombunchai, C. Gan, and M. Lee. An analysis of credit scoring for agricultural loans in thailand. 2005.

[26] M. Lipton. *The family farm in a globalizing world: The role of crop science in alleviating poverty*, volume 40. Intl Food Policy Res Inst, 2005.

[27] C. Machingura. Factors that can be used to identify potentially successful smallholder farmers: a case study of mbashe and ngqushwa.

[28] L. H. Miller and E. L. LaDue. Credit assessment models for farm borrowers: A logit analysis. Technical report, 1988.

[29] S. Million, R. Nyikal, S. Wangia, et al. Factors affecting loan repayment performance of smallholder farmers in east hararghe, ethiopia. *Developing Country Studies*, 2(11):205–213, 2012.

[30] B. Mondiale. World development report 2008: Agriculture for development. 2008.

[31] P. Mphaka. Strategies for reducing microfinance loan default in low-income markets. 2017.

[32] O. Nagayets. Small farms: current status and key trends. *The future of small farms*, page 355, 2005.

[33] S. Narayanan and A. Gulati. Globalization and the smallholders: A review of issues, approaches, and implications. Technical report, 2002.

[34] I. A. Ojiako and B. C. Ogbukwa. Economic analysis of loan repayment capacity of smallholder cooperative farmers in yewa north local government area of ogun state, nigeria. *African Journal of Agricultural Research*, 7(13):2051–2062, 2012.

[35] A. Okorie. Major determinants of agricultural smallholder loan repayment in a developing economy: Empirical evidence from ondo state, nigeria. *Agricultural Administration*, 21(4):223–234, 1986.

[36] J. Oladeebo and O. Oladeebo. Determinants of loan repayment among smallholder farmers in ogbomoso agricultural zone of oyo state, nigeria. *Journal of Social Sciences*, 17(1):59–62, 2008.

[37] F. Onyenucheya and O. Ukoha. Loan repayment and credit worthiness of farmers under the nigerian agricultural cooperative and rural development bank (nacrdb). *Agricultural Journal*, 2(2):265–270, 2007.

[38] G. F. Ortmann and R. P. King. Small-scale farmers in south africa: Can agricultural cooperatives facilitate access to input and product markets? Technical report, 2006.

[39] Oxfam. Understanding structural barriers & hidden bias in access to credit for women-led businesses. https://www.oxfamamerica.org/static/media/files/REPORT_OXFAM_GUATEMALA_3-web_gMpm7nD.pdf,Retrieved from Oxfam America website,2019, Aug 23.

[40] Ö. Özdemir and L. Boran. An empirical investigation on consumer credit default risk. Technical report, Discussion Paper, 2004.

[41] A. Park. Ren and s. wang, 2003. micro-finance, poverty alleviation reform in china. In *Workshop on Rural Finance and Credit Infrastructure in China, Paris, France*.

[42] L. C. Roberts. *An analysis of determinants of access to and use of credit by smallholder farmers in Suakoko District, Liberia*. PhD thesis, University of Nairobi, 2017.

[43] A. H. Roslan and M. A. Karim. Determinants of microcredit repayment in malaysia: The case of agrobank. *Humanity & Social Sciences Journal*, 4(1):45–52, 2009.

[44] L. M. Sama and R. Mitch Casselman. Profiting from poverty: ethics of microfinance in bop. *South Asian Journal of Global Business Research*, 2(1):82–103, 2013.

[45] A. Siamwalla, C. Pinthong, N. Poapongsakorn, P. Satsanguan, P. Nettayarak, W. Mingmaneenakin, and Y. Tubpun. The thai rural credit system: public subsidies, private information, and segmented markets. *The world bank economic review*, 4(3):271–295, 1990.

[46] M. Sileshi, R. Nyikal, and S. Wangia. Factors affecting loan repayment performance of smallholder farmers in east hararghe, ethiopia. *Developing country studies*, 2(11):205–213, 2012.

[47] N. Simumba, S. Okami, A. Kodaka, and N. Kohtake. Alternative scoring factors using non-financial data for credit decisions in agricultural microfinance. In *2018 IEEE International Systems Engineering Symposium (ISSE)*, pages 1–8. IEEE, 2018.

[48] C.-F. Tsai and M.-L. Chen. Credit rating by hybrid machine learning techniques. *Applied soft computing*, 10(2):374–380, 2010.

[49] F. Tumwine. *Savings and credit cooperatives (SACCOs) services terms and members economic development in Rwanda: a case study of Zigama SACCO LTD*. PhD thesis, PhD Thesis, Jomo Kenyatta University of Agriculture and Technology, Nairobi, 2015.

[50] C. G. Turvey. Credit scoring for agricultural loans: a review with applications. *Agricultural finance review (USA)*, 1991.

[51] M. van den Berg, R. Lensink, and R. Servin. Loan officers' gender and microfinance repayment rates. *The Journal of Development Studies*, 51(9):1241–1254, 2015.

[52] V. Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

[53] H. R. Varian. Monitoring agents with other agents. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift Für Die Gesamte Staatswissenschaft*, pages 153–174, 1990.

[54] M. Walpole, J. Smith, A. Rosser, C. Brown, B. Schulte-Herbruggen, H. Booth, M. Sassen, A. Mapendembe, M. Fancourt, M. Bieri, S. Glaser, C. Corrigan, U. Narloch, L. Runsten, M. Jenkins, M. Gomera, and J. Hutton. Smallholders, food security, and the environment, 03 2013.

[55] S. Wiggins, S. Keats, et al. Smallholder agriculture's contribution to better nutrition. *ODI, London*, 2013.

[56] M. Wijesiri, L. Viganò, and M. Meoli. Efficiency of microfinance institutions in sri lanka: a two-stage double bootstrap dea approach. *Economic Modelling*, 47:74–83, 2015.

[57] C. Wongnaa and D. Awunyo-Vitor. Factors affecting loan repayment performance among yam farmers in the sene district, ghana. *Agris on-line Papers in Economics and Informatics*, 5(665-2016-44943):111–122, 2013.

[58] H. Zhang. The optimality of naive bayes. *AA*, 1(2):3, 2004.