Scientific
Research
Publishing

Leveraging Geospatial Technology for Smallholder Farmer Credit Scoring

Susan A. Okeyo, Galcano C. Mulaku*, Collins M. Mwange

Department of Geospatial and Space Technology, University of Nairobi, Nairobi, Kenya
Email: susanokeyo@students.uonbi.ac.ke, *gmulaku@uonbi.ac.ke, cmwange@uonbi.ac.ke

Abstract

According to the Food and Agriculture Organization of the United Nations (FAO), there are about 500 million smallholder farmers in the world, and in developing countries, such farmers produce about 80% of the food consumed there; their farming activities are therefore critical to the economies of their countries and to the global food security. However, these farmers face the challenges of limited access to credit, often due to the fact that many of them farm on unregistered land that cannot be offered as collateral to lending institutions; but even when they are on registered land, the fear of losing such land that they should default on loan payments often prevents them from applying for farm credit; and even if they apply, they still get disadvantaged by low credit scores (a measure of creditworthiness). The result is that they are often unable to use optimal farm inputs such as fertilizer and good seeds among others. This depresses their yields, and in turn, has negative implications for the food security in their communities, and in the world, hence making it difficult for the UN to achieve its sustainable goal no.2 (no hunger). This study aimed to demonstrate how geospatial technology can be used to leverage farm credit scoring for the benefit of smallholder farmers. A survey was conducted within the study area to identify the smallholder farms and farmers. A sample of surveyed farmers was then subjected to credit scoring by machine learning. In the first instance, the traditional financial data approach was used and the results showed that over 40% of the farmers could not qualify for credit. When non-financial geospatial data, *i.e.* Normalized Difference Vegetation Index (NDVI) was introduced into the scoring model, the number of farmers not qualifying for credit reduced significantly to 24%. It is concluded that the introduction of the NDVI variable into the traditional scoring model could improve significantly the smallholder farmers' chances of accessing credit, thus enabling such a farmer to be better evaluated for credit on the basis of the health of their crop, rather than on a traditional form of

collateral.

1. Introduction

Globally, smallholder farmers are generally understood to be those that farm on small pieces of land, often taken as 2 ha or less, and largely for subsistence; however, this size threshold varies from country to country, depending on the prevailing ecological and demographic conditions. The Food and Agriculture Organization (FAO) estimates that there are about 500 million such farmers in the world, and they produce about 80% of the food consumed in their countries [1]; their farming activities are therefore critical to the food security of their communities and the world in general, as the UN can hardly achieve Sustainable Goal no.2 (no hunger) without them. Smallholder farmers face the challenge of limited access to credit due to the fact that most of them farm on unregistered land (a common phenomenon in developing countries), which cannot be offered as collateral to lending institutions; those that farm on registered lands fear losing such land (often their only key family asset) that they should default on loan payments and hence don't apply for credit; those that still do apply are still disadvantaged by low credit scores due to poor financial history and low valued collateral. The result is financial exclusion, inability to provide optimal farm inputs, depressed yields and food insecurities [2]. An alternative approach to the evaluation of such farmers for credit, such as giving weight to the health of their crops (as an indicator of expected yields, and hence ability to service credit) would go a long way in mitigating the credit challenge for smallholder farmers.

Credit scoring is a statistical analysis which is performed by either lenders or credit reference institutions to assess a person's creditworthiness. For lenders, credit scoring is subsequently used to help decide whether to extend or deny credit to a borrower [3]. For credit reference institutions, credit scoring is described as a means of making a summary of information on a credit application so as to produce a credit score. In Kenya, specifically Migori County, most farmers are unable to access loans to finance their farming activities due to low credit scores. This is mainly attributed to the fact that they are financially excluded [2].

The need for credit scoring arises out of the fact that lending is always associated with the risk that the borrower may default on paying back their loan; for example, it has been recently reported that over 60% of small scale borrowers in Kenya defaulted on their loans in a given month [4]. The lender, therefore, desires to separate loan applicants into high risk and low risk ones, so that the lower risk ones can receive priority in getting loans, thus minimizing bad debts for lenders; a credit score is able to achieve this binary separation.

The emergence of credit scoring can be traced back to 1941 when David Du-

rand established a credit scoring system. Durand identified different variables that helped lenders distinguish between good and bad loans [5]. In 1946, E. F. Wonderlic developed a credit score guide that helped define and narrow the variables of good and bad loans [6]. The credit score guide helped to indicate the degree of risk associated with a customer. In the 1950's credit scorecards were becoming a popular instrument used in creditworthiness assessment.

The Fair Isaac Corporation was founded in 1956 and introduced its first credit scoring system in 1958 (myFICO). These scorecards were models that helped to determine if a customer would default on their loan given their current financial position. The late 1960s to early 1970s brought about technology that allowed for credit scoring models to be developed further and automated [7].

Today many lenders rely on the traditional 5Cs of credit to assess the credit risk of applicants. These are Character (e.g. reputation, credit history), Capacity (e.g. cash flow stability), Capital (e.g. net worth), Collateral (e.g. fixed asset that can be sold in case of default) and Condition (e.g. current economic condition) [8]. Smallholder farmers with no or little credit history, limited capital and collateral, inevitably fare poorly in this kind of traditional assessment.

In Kenya, credit scoring for farmers is mainly carried out by commercial banks, insurance companies, the Agricultural Finance Corporation and some civil society organizations such as the Syngenta Foundation [9]; these organizations rely on the same 5Cs approach, to the detriment of smallholder farmers.

This study aimed to determine whether, by including a geospatial variable, specifically crop NDVI, in the credit evaluation model, a smallholder farmer's chances of accessing credit could be improved. The contribution or novelty of this research could be considered threefold: It has demonstrated how geospatial technologies can be utilized to generate non-financial data as well as how such geospatial data can complement historical information used by financial institutions to determine more farmer-friendly credit scores. The study has also contributed knowledge on the application of artificial intelligence machine learning techniques to agriculture, especially African agriculture. Finally, the research has contributed to the global debate on the financial inclusion of vulnerable smallholder farmers whose farm outputs remain critical to the achievement of food security in the world.

2. The Study Area

This study was carried out in Migori County of Kenya, illustrated in Figure 1.

Migori is a county in the former Nyanza Province located in south-western Kenya, bordering Homabay, Narok, Tanzania, Kisii and Lake Victoria. Geographically, it lies between east longitudes 33°55'42" and 34°43'50" and latitudes −1°39'06" and −1°23'21". It has eight sub-counties, namely; Awendo, Uriri, Rongo, Kuria East, Kuria West, Suna East, Suna West and Nyatike. The main livelihood activities in Migori County include agro-farming and pastoral farming. Among these livelihood activities, crop farming forms the backbone of the economy in
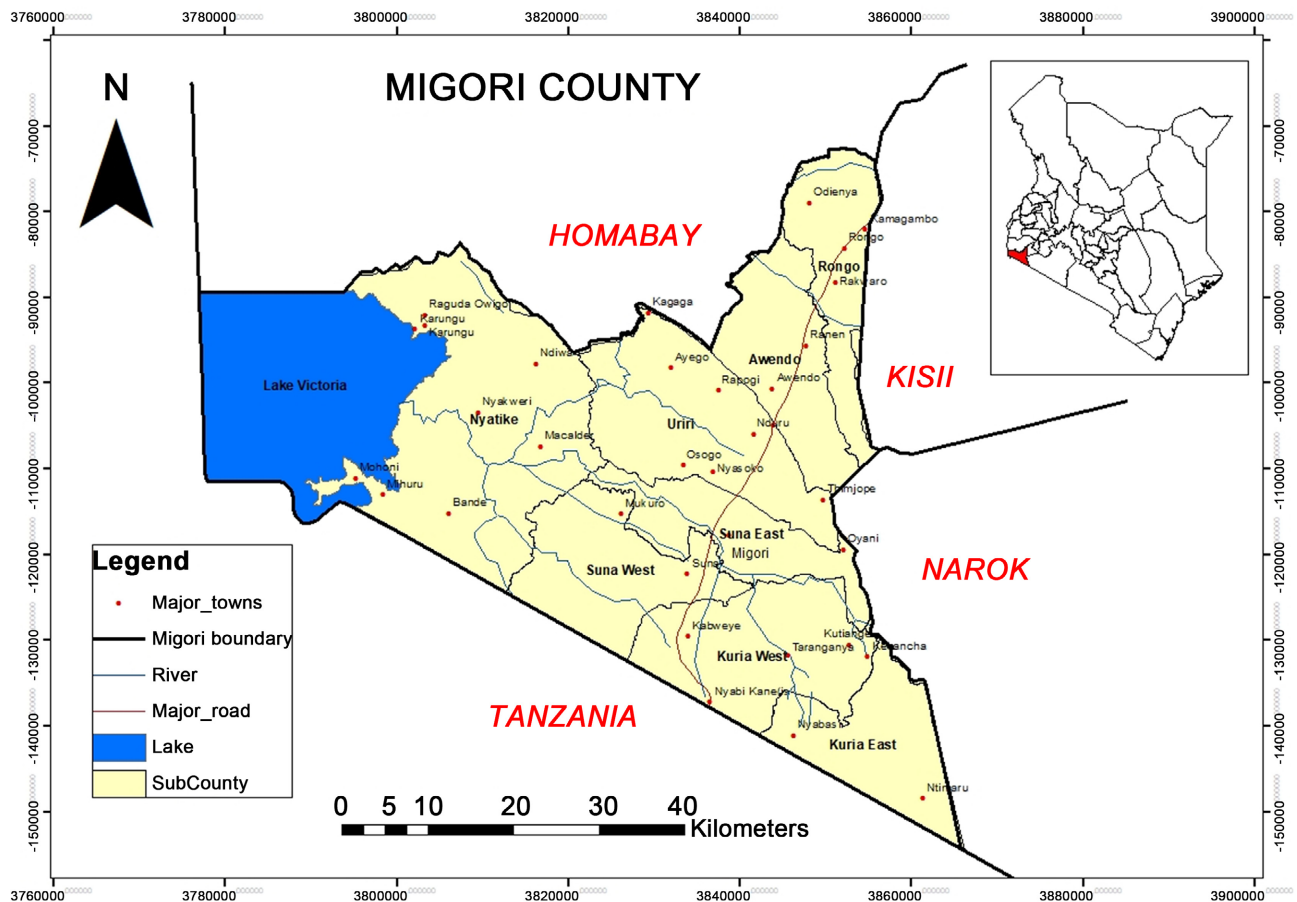
Figure 1. Map showing the study area.

the county. At least 70% of the people resident in Migori County are dependent on crop farming, and the farming is dependent on rain-fed agriculture.

Demographically, Migori County is the most diverse region of the former Nyanza province after Kisumu County. The main inhabitants are Suba people, Luo, Abakuria, Abagusii, and Abaluhya. Others are Somalis, Indians, Arabs, and Nubians. The population of Migori County according to the 2019 population census was 1.2 million people [10].

Climate-wise, Migori County has two main rainy seasons. The first rainy season starts in March and ends in May, this season constitutes the long rains. The second rainy season starts in September and ends in November. The driest months are between December and February and June and September. The average daily temperature is usually a low of 24 degrees Celsius (74 F) and a high of 31 degrees Celsius (87 F). Rainfalls come in the afternoon and the heat is often dry and thus bearable. Migori County was chosen for this study due to its ease of access for field research, and also its prevalence for smallholder farmers who do not own the land they farm on.

3. Methodology

The methodology followed in this study as illustrated in **Figure 2**.
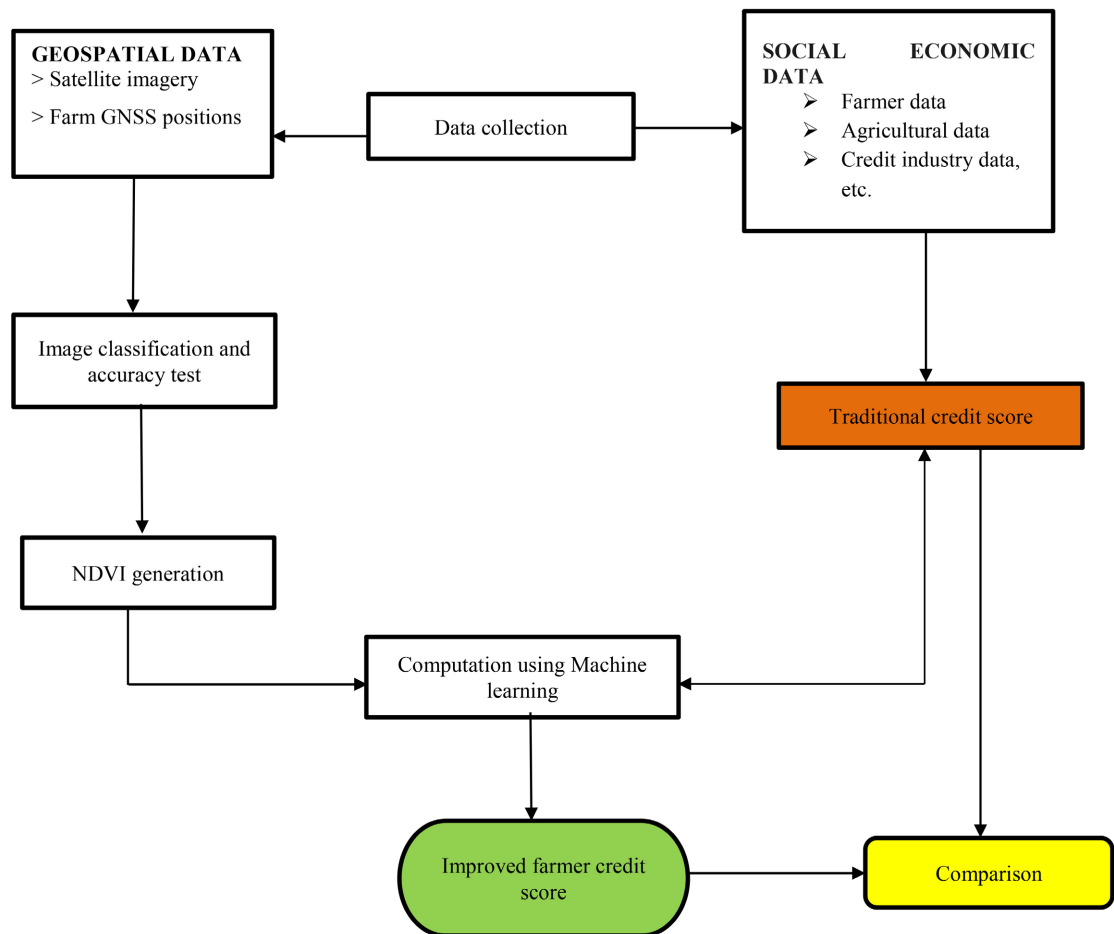
Figure 2. Methodology flow chart.

3.1. Data Collection

Data was collected in two phases. In phase 1, small scale farmers were identified through stratified random sampling and relevant data was collected from them, through a field questionnaire survey conducted in all the eight sub-counties of Migori County. The data collected from each respondent included personal information (such as name, identification, gender, presence or absence of bank account, etc.), occupation, any assets, types of crops grown and access to credit. 337 questionnaires were administered with 320 being completed and returned, giving a very good return rate of 95%. Interviews with key informants in the financial institutions and Credit reference bureaus in Kenya were also conducted, in order to get a feel of their awareness of and use of credit scoring.

In phase 2, Sentinel 2 imagery covering the study area was collected and LULC was classified using Google Earth Engine. The date of the imagery was November 2021. In addition, reference points for later use in image analysis were selected in a well-distributed pattern over the study area and positioned using hand-held GNSS. Phase 2 data collection started in Kuria sub-County, due to the rains and the poor road infrastructure there. 158 reference points were collected from seven sub-counties; Nyatike was excluded due to minimal farming there. To ensure even

distribution of the points, it was decided to pick 5 points in each sub-County, representing the west, east, north, south and central areas of the sub-County; for example, in Suna East Sub-County, points were picked from Rabuor in the east, Nyarongi in the west, Godjope in the north, Witharaga in the south and Ngege in central. Two GNSS instruments were used during the reference data collection, a Trimble TDC100 mobile mapper and a Garmin Trex 10 hand-held. A research assistant was trained to operate the second instrument for the purpose of validation and backup of the data from the first instrument which was operated by the researcher herself. The image composite plus the collected reference points are shown in Figure 3.

3.2. Data Analysis

Data analysis was carried out in three steps, as indicated below.

3.2.1. Step 1: Identification of Smallholder Farms and Farmers

The smallholder farms and farmers that would later feature in the credit scoring analysis were identified according to the following criteria:

- Farms size was to be less or equal to 2 hectares.
- The farmers had to either have title to their farms or be farming on family land.
- For even distribution across Migori, 15 farms per sub-county were picked except for Nyatike sub-county where farming is minimal. In total, 101 farms were selected.
- The farms was to be in a cloud-free area in the imagery (as indicated by the corresponding reference data), to enable later generation of NDVI.
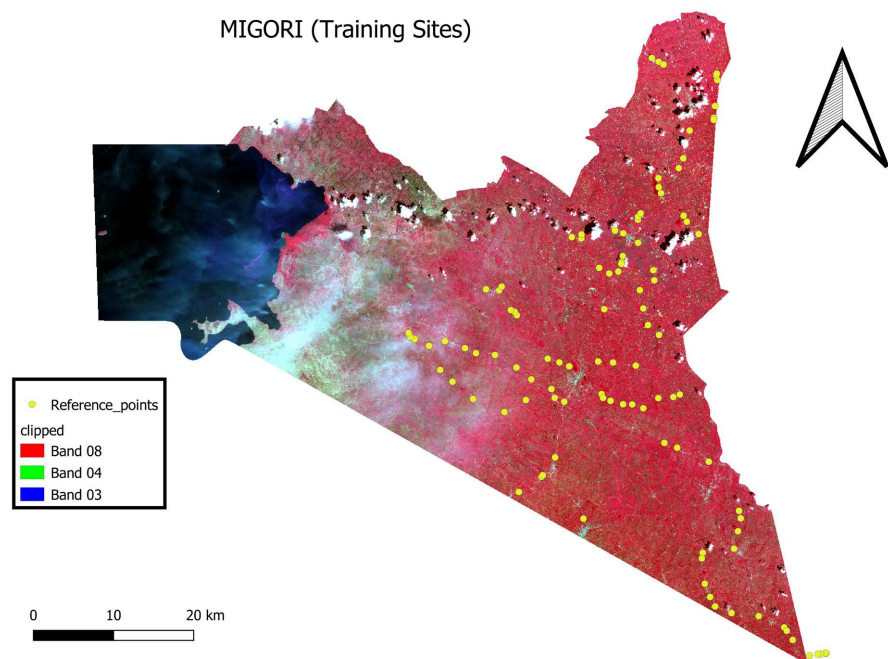
Figure 3. The imagery plus reference points.

3.2.2. Step 2: Image Classification and NDVI Generation

The image was classified using the supervised classification approach and the maximum likelihood classifier. The classification was based on the five classes shown in Table 1, and an overall classification accuracy of 0.861 was obtained. The classified image is shown in Figure 4.

Following the image classification, the classified image was masked to retain only Pixels covered by maize, and the masked image was then used to generate NDVI per pixel using the formula: NDVI = NIR − R/NIR + R.

The resulting NDVI image is shown in Figure 5.

On the assumption that any smallholder farm (≤2 ha) could fit within a 1 km × 1 km grid square, such a grid was overlaid onto the NDVI image and the average NDVI per grid square was computed; this would be the representative NDVI for any farm falling within that grid square. Table 2 shows an abstraction of the

Table 1. Image classification results (Classification accuracy = 0.861).

class	Description	Confusion matrix
2	Bare land	[0, 0, 30, 0, 0, 0, 1]
3	Other crops	[0, 0, 0, 25, 0, 0, 2]
4	Built up	[0, 0, 0, 0, 5, 0, 3]
5	Water	[0, 0, 0, 0, 0, 3, 0]
6	Maize	[0, 0, 2, 3, 0, 0, 5]
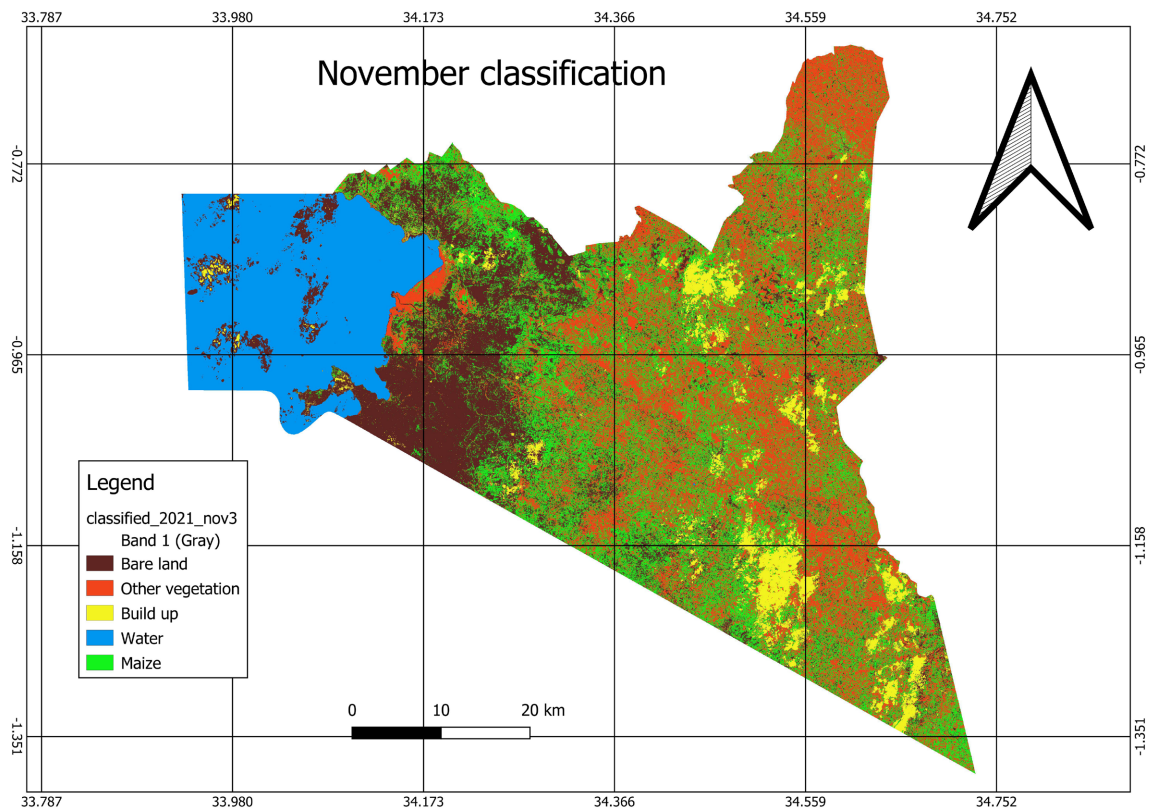
Figure 4. The classified image.

NDVI IMAGE

Legend

NDVI Value

High : 0.803727

Low : -0.277657

Miles
0   3.5   7   14   21   28
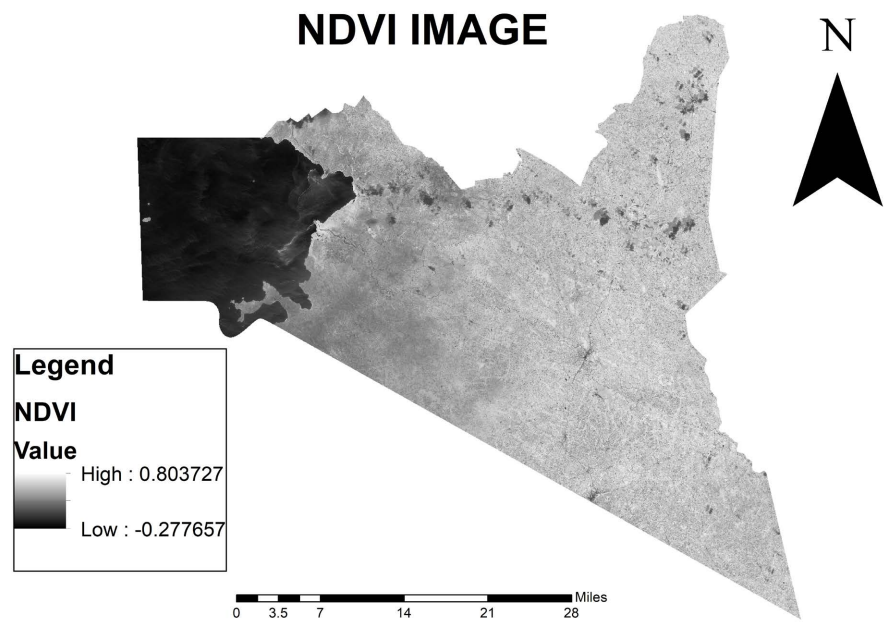
Figure 5. Farm NDVI image.

Table 2. Abstract of NDVI averages.

Farm points (Easting, Northing)	Mean NDVI
(679440.57, 9860357.50)	0.53
(679322.61, 9859664.19)	0.51
(679692.43, 9856491.25)	0.54
(680359.45, 9854840.07)	0.54
(680988.03, 9853659.18)	0.54
(683031.34, 9852836.49)	0.52
(686422.71, 9852453.27)	0.52
(689603.28, 9851039.78)	0.55
(689916.28, 9850623.30)	0.54
(690661.04, 9849483.76)	0.55

NDVI averages for maize farms within the study area as identified by the reference points that had been positioned within them.

3.2.3. Step 3: Development of a Score Model

This step involved the development of a score model that computes both the traditional score and the new score for any farmer, with the latter score computation including the geospatial variable NDVI. Machine learning using logistic regression was used. This is a predominant method in credit analysis and has become the benchmark method against the other methods for such credit analysis [11].

This process involved modeling farmers' data to come up with a credit scoring tool to assess the eligibility of a farmer accessing a loan, with additional empha-

sis on crop geospatial data/information. This was based on sample farmers from Migori County who were interviewed with regard to access to credit. Weight of Evidence (WOE), a statistical technique commonly used in credit scoring to evaluate the predictive power of various features or variables was employed. This method enabled the transformation of raw data into meaningful and informative predictors, providing a solid foundation for accurate credit risk assessment. Using the R programming language, the step-by-step processes which were undertaken in implementing credit scoring using WOE and the Generalized Linear Model are now outlined:

1) Start Variables

The start variables which were obtained from the questionnaire are shown in Table 3.

2) Correlation Testing

All the start variables were tested for correlation. This is because the predictor variables used in logistic regression should, ideally not be correlated. Correlation among such variables can cause model problems such as multi collinearity, leading to unstable and unreliable estimates of the regression coefficients. In such cases, the regression coefficient may change dramatically with minor changes to the data.

3) Final Variable Selection

Following the dropping of correlated variables, six final predictor variables were selected based on their information value (IV) and domain knowledge (i.e. knowledge of the credit scoring industry). IV is a tool in machine learning used to assess the predictive power of any variable for a given feature in a dataset and is often used in credit scoring. It quantifies the extent to which a variable can differentiate between different outcomes, such as default and non-default, and provides insights into the variable's contribution to the predictive model's performance. IV depends on a variable's weight of evidence (WOE).

Statistically,

$$\text{WOE} = \ln\left(\frac{\%\ of\ non\text{-}events}{\%\ of\ events}\right)$$

and IV = (% of non-events − % of events)

In the case of separating loan defaulters from non-defaulters, default represents an event while non-default represents a non-event.

Generally the higher the IV, the better the variable for the intended prediction; however the best IV values range between 0.3 and 0.5; although values between 0.1 and 0.3 may also be accepted for model development as medium predictors [11].

Following the WOE/IV analysis, the variables (Land parcel, Collateral, Reasons and How long) were selected for traditional score model development. The NDVI would be deliberately introduced for computing the new score. NDVI (Normalized Difference Vegetation Index) is an important index used in remote sensing to assess and monitor vegetation health and vigour, and it is the geospatial variable

Table 3. Model start variables.

Variable	Meaning
Gender	Gender of respondent
Age	Age of respondent
Marital	Marital status
Occupation	What farming activities does the respondent conduct?
Year of occupation	How long has the respondent practiced farming?
How much land	How much land holding does the respondent have?
Land ownership	What ownership do you have on the land you are farming on?
Location	Where is the land located?
Land parcel	Do you have more than one parcel of land?
How many	If yes how many parcels do you have?
Land hold	Is the land freehold or leasehold?
Market	How and where do you sell your maize product?
Collateral	What do financial institutions require as collateral before granting loan?
Credit access	What are challenges faced when applying for loan?
Credit repayment	If you have previously applied for loan, did you pay back on time?
Reasons	If no specify the reasons
Acquire	Where do you acquire your loan from?
Capacity	What is your credit capacity?
How long	For how long have you been taking loans over the years?
Outstanding	Do you have outstanding loans?
Loan duration	If yes when is your loan due?
Set	How much have you offset so far?
Employment	Are you currently employed?
Employer	If yes, indicate your employer?
Income	If no, indicate your source of income?
Cycle	Indicate the number of crop cycle per year in your farm?
Pest	What type of pest attacks do you encounter in your farming activities?
Organic	Do you use any organic pest control technique?
Specify	If yes, specify
Irrigation	Do you use any type of irrigation techniques ?
Specify irrigation	If yes, specify
Ploughing	Which ploughing methods do you use in farming?

that was introduced into the model to indicate whether the farmer had healthy crops or not.

4) Data Modelling

After converting the original data points to WOE, a Generalized Linear Model was fitted based on the following logistic regression formula [12].

$$\ln\left(\frac{P(X)}{1-P(X)}\right) = \text{intercept} + B_1 x + B_2 x + \cdots + B_n x$$

The R link function used to generate results is the logit value.

$$\begin{aligned}\text{model\_fit} = \text{glm}\big(\text{formula}\\ = \text{Credit}\sim\text{Landparcel}_{\text{woe}} + \text{Collateral}_{\text{woe}} + \text{Reasons}_{\text{woe}} + \text{Howlong}_{\text{woe}}\\ +\text{NDVI}_{\text{woe}}, \text{family} = \text{binomial}\big(\text{link} = \text{logit}, \text{data} = \text{train\_woe}\big)\end{aligned}$$

Script 1: Data Modelling

The output of this model, which was in logit was converted to odds and then probability as follows:

$$\text{Odds} = e^{\text{logit}}$$

$$\text{Probability} \quad P(X) = \frac{\text{odds}}{\text{odds}+1}$$

After running the model, the coefficients achieved were in terms of log odds; which were converted to probabilities and then final scores (scaling). These scores were originally in the scale of 0 to 1000. The probabilities were converted to scores using the Scorecard package in R using the following approach (script 2).

$$\text{score} = \text{offset} - \text{factor} \ln(\text{odds})$$

$$\text{factor} = \frac{\text{pdo}}{\ln(2)}$$

$$\text{offset} = \text{Ts} - \text{factor} \ln(\text{To})$$

Replace odds with logit.

$$\text{odds} = e^{\text{logit}}$$

$$\text{score} = \text{offset} - \text{factor logit}$$

Ts—target score.
To—arget odds.
Pdo—slopes.

Script 2: Score Computation

The original scores were finally scaled to the FICO score, which is more universal and frequently used [13].

A FICO score is a three-digit number, typically in a 300-850 range, that tells lenders how likely a consumer is to repay borrowed money based on their credit history. Only FICO Scores are created by the Fair Isaac Corporation and are used by over 90% of top lenders when making lending decisions.

The FICO score range is explained in Table 4.

Out of the 101 farmers and farms identified, 67 (about 2/3) were chosen for training the model, with the remaining 34 reserved for later testing of the model.

Table 4. FICO score range.

<580	Poor	Your score is well below the average score of the consumers and demonstrates to lenders that you are a risky borrower.
580 - 669	Fair	Your score is below the average score of the consumers, though many lenders will approve loans with this score.
670 - 739	Good	Your score is near or slightly above the average of the consumers and most lenders consider this a good score.
740 - 799	Very good	Your score is above the average of the consumers and demonstrates to lenders that you are a very dependable borrower.
800+	Exceptional	Your score is well above the average score of the consumers and clearly demonstrates to lenders that you are an exceptional borrower.

5) Model Evaluation

The performance of a binary classification model, such as the regression model used in this study, is often evaluated by plotting the Receiver Operating Characteristic (ROC) curve and determining the Area under the Curve (AUC). The curve checks how well the model is able to distinguish and separate events from non-events; it is a plot of the rate of true positives (events correctly predicted to be events) on the y-axis against the rate of false positives (non-events wrongly predicted to the events) on the x-axis. Generally the higher the AUC curve *i.e.* the bigger the area under the curve, the better the model, and 75% is the recommended acceptable minimum. Script 3 was used to determine this AUC. A related evaluation measure is the Gini coefficient defined as: Gini = (2 * AUC − 1).

Again, a higher Gini represents a better predictive model.

```
## A performance instance
## 'Area under the ROC curve'
####Make prediction
```

Testing of the predictive power of the model on test data that had been earlier isolated is done.

```
## [1] "K-S Statistic = 0.846718005133847"
## [1] "Area under the curve = 0.961037770443711"
## [1] "Gini Coefficient = 0.922075540887423"
```

##(Intercept)	Landparcels	Landhold	Collateral	Creditrepayment
##−5.2637155148	−1.5550051782	−0.5318544217	0.4488074659	−0.6338934531
##Reasons	Acquire	Howlong	Outstanding	Set
##−2.6601538908	0.6818583697	0.3192265714	3.1808766951	−0.0006021895
##Pest	Organic			
##0.0635936312	1.5129961303			

Script 3: Evaluating the model by AUC

4. Results

The final model developed is shown in Table 5, with the indicated predictor variables.

The model was trained on 67 farms and farmers; an abstract of the results is shown in Table 6.

In respect of Table 6, it should be noted that the model was first run **without** NDVI, to generate the "traditional" score column. NDVI was then introduced, to generate the "with 1.0 NDVI" column.

For experimental purposes and in view of the central role that NDVI was to play in this whole arrangement, the weight of NDVI was deliberately biased to 1.2, 1.3, 1.4 and 1.5 the original weight and this generated the correspondingly labeled columns.

The performance of the model in this training was evaluated by plotting the ROC curve and determining the AUC; the resultant curve is shown in Figure 6.

Table 5. Credit scoring model statistics.

	Estimates	Std. Error	z value	Pr(>|z|)	Significance
Intercept	−1.2873	0.3705	−3.474	0.000512	***
Landparcels_woe	1.0331	0.5107	2.023	0.043081	*
Collateral_woe	0.8495	0.3020	2.813	0.004903	**
Reasons_woe	1.4431	0.4171	3.460	0.000540	***
Howlong_woe	1.0761	0.1874	5.742	9.38e−09	***
NDVI_woe	2.0734	0.8762	2.366	0.017967	*

Table 6. Abstract of model training results.

Farm serial	Farm ID	Farmer ID	Land parcels > 1?	Collateral	Reasons	How long	NDVI	Previous credit	Traditional score	With 1.0 NDVI	With 1.2 NDVI	With 1.3 NDVI	With 1.4 NDVI	With 1.5 NDVI
1	1	20184113	2	1	Null	Null	0.64	0	685	722	806	848	850	850
2	2	27186633	2	5	Null	Null	0.62	0	599	641	709	743	777	811
3	3	8144389	1	1	Null	Null	0.63	0	639	663	735	771	807	844
4	4	221586	2	1	Null	Null	0.62	0	685	722	806	848	850	850
5	5	32261560	2	Null	Null	Null	0.56	0	726	661	733	769	805	841
6	7	2803912	2	4	Null	20	0.52	1	367	408	430	441	452	463
7	8	37703733	2	2	Null	1	0.30	0	510	499	538	558	578	598
8	10	362547	2	Null	Null	Null	0.60	0	726	661	733	769	805	841
9	11	30142286	2	Null	Null	5	0.33	0	565	549	598	623	648	673
10	13	22211547	2	Null	Null	Null	0.58	0	726	661	733	769	805	841
11	15	13386264	1	4	1	10	0.65	0	462	474	509	526	543	561
12	17	112548	2	1	Null	2	0.58	0	521	447	476	491	506	520
13	20	10233261	2	7	Null	Null	0.61	0	637	676	751	789	827	850
14	41	2771059	1	Null	Null	Null	0.49	0	681	717	800	842	850	850
15	43	23251161	1	2	Null	5	0.68	0	449	463	495	512	528	544

From the curve, the AUC was found to be 95%, with a Gini Coefficient of 0.9.
Following these encouraging results from model training, the model was now
tested on the remaining 34 farms and farmers, and the results are shown in Ta-
ble 7.
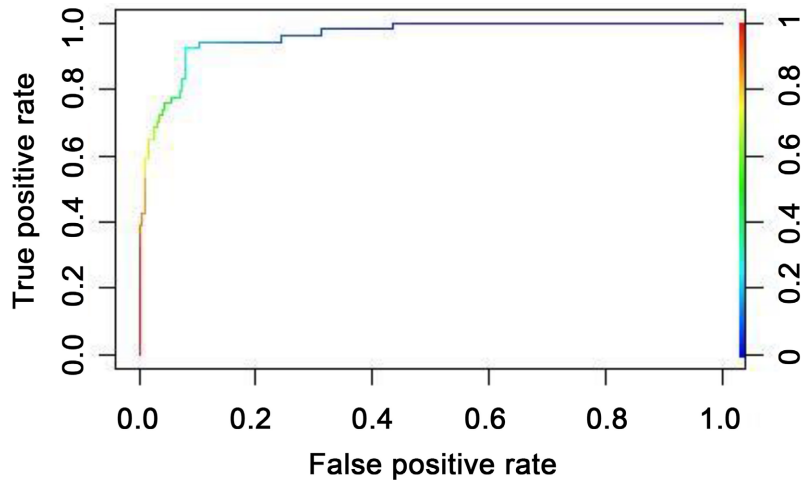
Figure 6. The model AUC curve.

Table 7. Model testing results.

Farm serial	Farm ID	Farmer ID	Land parcels > 1?	Collateral	Reasons	How long	Previous credit	NDVI	Traditional score	With 1.0 NDVI	With 1.2 NDVI	With 1.3 NDVI	With 1.4 NDVI	With 1.5 NDVI
1	6	13388005	2	Null	Null	Null	0	0.74	726	760	850	850	850	850
2	9	6912420	2	1	2	Null	0	0.44	825	823	850	850	850	850
3	12	13112933	2	Null	Null	2	0	0.67	562	585	642	670	699	727
4	14	21186481	2	Null	Null	Null	0	0.56	726	661	733	850	805	841
5	16	21086953	2	Null	Null	3	0	0.57	515	434	461	580	488	501
6	18	24244325	2	2	2	5	0	0.58	635	564	616	735	669	695
7	19	3369874	2	Null	2	Null	0	0.61	850	850	850	850	850	850
8	39	3654821	2	1	Null	4	1	0.46	474	511	553	526	596	617
9	40	8159461	1	Null	Null	Null	0	0.51	681	717	800	795	850	850
10	42	10130257	2	Null	Null	Null	0	0.69	726	760	850	850	850	850
11	50	24031606	1	5	Null	4	1	0.60	342	354	365	354	376	382
12	51	9771763	2	Null	Null	Null	0	0.61	726	760	850	850	850	850
13	52	33724440	2	Null	2	2	0	0.62	703	726	811	823	850	850
14	56	5643871	2	Null	Null	Null	0	0.45	726	721	805	850	850	850
15	57	27993343	2	Null	Null	Null	0	0.38	726	721	805	850	850	850
16	60	24922360	2	1	Null	Null	0	0.49	685	738	825	801	850	850
17	63	23890276	1	8	Null	10	1	0.64	360	369	383	378	396	403
18	70	22407658	2	4	2	10	1	0.40	507	494	532	570	571	590

Continued

19	71	22304385	1	1	Null	16	1	0.63	408	414	437	440	460	472
20	72	24503941	2	1	Null	Null	0	0.59	685	622	687	801	751	784
21	75	37752062	1	1	Null	3	1	0.45	428	396	416	467	435	444
22	125	11215846	2	6	Null	4	1	0.57	388	314	317	414	320	322
23	126	30002630	2	1	Null	5	0	0.65	524	550	600	591	650	675
24	127	6912045	2	1	Null	7	0	0.48	524	566	619	591	672	699
25	129	3972355	1	10	Null	2	1	0.54	428	342	350	466	359	363
26	132	24222361	1	1	Null	5	0	0.26	478	451	481	532	511	526
27	134	26811913	2	1	Null	3	1	0.56	474	396	415	526	434	444
28	136	32241069	2	1	Null	Null	0	0.58	685	622	687	801	751	784
29	145	29547139	2	1	Null	Null	0	0.52	685	738	825	801	850	850
30	204	11224046	2	2	Null	Null	0	0.62	656	694	773	763	850	850
31	210	7.21E+08	1	1	Null	Null	0	0.54	639	563	616	741	668	694
32	217	25696211	2	7	2	Null	0	0.50	777	833	850	850	850	850
33	221	11412683	2	1	2	Null	0	0.62	825	850	850	850	850	850
34	222	13244463	2	Null	8	Null	0	0.31	850	850	850	850	850	850

5. Discussion

The credibility of the scoring model developed is evidenced by the ROC curve in Figure 6, whose high AUC indicates a good choice of predictors. The results in Table 7 show that out of the 34 farmers scored, 14 were poorly scored (below 580) and hence would not be recommended for credit. On introduction of the NDVI variable, nearly half of all farmers had their scores improved, and one farmer was able to move to a score of over 580 and hence become eligible for credit. On enhancing the weight of NDVI up to 1.5 NDVI, 5 more farmers were able to transition to eligible status. Therefore by the end of the NDVI experiment, 6 farmers out of the originally ineligible 14 had changed status and could now get credit; the level of non-eligibility had therefore changed from 41% to 24%. This is a promising result that can be built upon if the lending industry was to warm up to it.

However, since the model is developed using machine learning techniques, it needs fine tuning using a larger sample of farmers, and also testing in more geographic locations and agro climatic zones.

6. Conclusion

The results achieved from this study have demonstrated that NDVI can be a useful tool in improving smallholder farmers' credit scoring. Lending institutions can adopt the model in order to improve the chances of smallholder farmers' access to credit for the sustenance of their farming activities. However further research is needed in order to fine-tune the model using a larger data set.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] FAO (2014) The State of Food and Agriculture. Flagship Report, UN Food and Agriculture Organization.

[2] Okeyo, S.A., Mulaku, G.C. and Mwange, C.M. (2022) Statistical Analysis of Smallholder Farmer Financial Exclusion: Case Study of Migori County, Kenya. *Open Journal of Statistics*, **12**, 733-742. https://doi.org/10.4236/ojs.2022.125043

[3] Costa, A., Deb, A. and Kubzansky, M. (2016) Big Data, Small Credit: The Digital Revolution and Its Impact on Emerging Market Customers. *Innovations: Technology, Governance, Globalization* 2015, **10**, 49-80. https://doi.org/10.1162/inov_a_00240

[4] Standard Nation-Kenya, 9th September 2023.

[5] Gutiérrez-Nieto, B., Serrano-Cinca, C. and Camón-Cala, J. (2016) A Credit Score System for Socially Responsible Lending. *Journal of Business Ethics*, **133**, 691-701. https://doi.org/10.1007/s10551-014-2448-5

[6] Weston, L.P. (2012) Your Credit Score: How to Improve the 3-Digit Number That Shapes Your Financial Future. 4th Edition, Pearson Education, London.

[7] Thomas, L.C., Ebelman, D.B. and Crook, J.N. (2004) Readings in Credit Scoring: Foundations, Developments, and Aims. Oxford University Press, Oxford.

[8] High Radius.com. https://www.highradius.com/

[9] Ngare, P., Kweyu, M. and Huka, C. (2015) Modeling Risk of Financing Agribusiness in Kenya. Kenya Bankers Association.

[10] KNBS (2019) The Kenya National Bureau of Statistics: Kenya Population and Housing Census. https://www.knbs.or.ke/

[11] https://www.listendata.com

[12] Sirak, M. and Rice, J.C. (1994) Logistic Regression: An introduction. JAI Press 3, 191-245.

[13] https://www.myfico.com