Original Research Article

MODELING SELF HELP GROUPS' IMPACT ON LIVELIHOODS IN MURANG'A EAST SUB-COUNTY: A LOGISTIC REGRESSION APPROACH.

ABSTRACT

According to the World Bank (2022), approximately 8.9 million people, or 17% of Kenya's population, live below the poverty line of 1.9 USD on a daily basis, majority of them in the rural areas. This research aimed to analyze the impact of self-help groups on the livelihoods of rural areas of Kenya, with the goal of promoting sustainable livelihoods and reducing poverty. To achieve this, the study employed machine learning specifically thelogistic regression algorithm to model the impact of self-help groups on livelihoods in Murang'a East sub-county.The study used primary data obtained through the issuance of structured questionnaires to SHG members, on their wealth status since joining the self-help groups on areas such as ability to save, access to credit services and acquiring assets, both income generating and household. A total of 969 members of self-help groups were issued with the questionnaire. The study's findings helped identify the key predictors of members' livelihoods and provided insights into how self-help groups influence them. The results of logistic regression indicated that 91.33% of the members had seen a significant improvement on their wealth status since joining self-help groups and the significant predictor variables were income generating assets, access to basic commodities and access to loans. The model's accuracy was 88.04%.The ethical considerations in this study included ensuring no coercion or pressure to participate in the study and confidentiality and privacy of the respondents.

Key Words:*Self Help Groups; Machine Learning;Logistic Regression.*

1.Introduction

SHGs are informal groups of individuals who work together towards a common goal. These groups are formed voluntarily by individuals who have a good relationship with one another and a desire to support each other in achieving their goals. The nature of the group's objectives may vary depending on the interests of its members. SHGs are self-regulated, with members selecting their leaders from within the group.

The Kenyan economy has been identified as the most advanced in Eastern and Central Africa by a report from the World Bank in 2019. However, there is still a significant portion of the population, amounting to 16.1%, who struggle to access basic needs and are considered to be living in poverty.

This demonstrates the need to find ways and measures to assist this group of people living below the poverty lineto achieve financial independence and sustainability in their way of living.

Livelihood refers to the approach through which people acquire the essentials of life (Serrat, 2017). It encompasses the means and activities undertaken to obtain basic and supplementary needs.

As per the data provided by the Kenya Data Portal, the population of Murang'a East sub-county was 1,056,640 during the 2019 census. A significant proportion of this population, specifically 28.5%, are living below the poverty line.

This research aimed to assess self-help groups' impact on the livelihoods of their members in Murang'a East sub-county by utilizing logistic regression. The study aimed to determine whether SHGs can be effective in eliminating poverty and providing individuals with the resources necessary to obtain basic needs and beyond.

1.1 Logistic Regression Model.

According to Raghavan et al. (2016), logistic regression is utilized to identify a boundary line that can differentiate between the different categories within a dataset. The model comprises of a response variable and predictor variables, which jointly determine the outcome of the dependent variable. Logistic regression also helps identify the significant predictor variables in predictor the outcome or response variable using the p-values. The output variable in logistic regression is in binary form, yes or no.

The coefficients that are assigned are denoted by β. The first coefficient, β_0, is the intercept, while each variable from $1 \ldots \mathrm{n}$ are each assigned the coefficients, $\beta_1, \beta_2 \ldots \beta_n$. They determine the association between the independent and dependent variables. The result obtained from the analysis is in the form of odds ratio;

$$\frac{P}{1-P} = exp(\beta_0 + \sum_{i=1}^{n} \beta_i X_i)$$

The probability of the occurrence of the event is denoted by P.

1.2 Scope and Limitations of the Study.

This study aimed to analyze the impact of self-help groups on the livelihoods of their members in Murang'a East Sub-County, Kenya, using logistic regression. To achieve its objective, the study used data from self-help group members to build a predictive model.

The limitations of the study included bias from the respondents in their responses and the personal perceptions they individually have concerning self-help groups and their impact. This was however tackled using the test-retest method which involved administering the same test twice. This was achieved by conducting the pilot study and then the actual study and evaluating the accuracy of the information provided. In both the pilot study and the actual study, the questionnaire captured the information it was intended to capture consistently. In addition, the members were not left to decide whether the impact was positive or not, but their responses to various questions were analyzed and this decision arrived at using logistic regression, this played a major role in doing away with bias of the respondents.

1.3Literature Review

According to Brody et al. (2013), a SHG is a group of individuals who voluntarily unite with a shared purpose and work together to achieve that purpose. Self Help Groups (SHGs) are created to cater to the needs of underprivileged and marginalized individuals who can work together to achieve common goals. According to Deininger et al. (2009), Self-Help Groups (SHGs) have become increasingly popular in India, particularly at the grassroots level and have had a beneficial effect on the economic empowerment and access to fundamental needs of their members.

Dr. Muhammad Yunus, the founder of Grameen Bank in Bangladesh, is credited with starting the SHG movement. The Grameen Bank provided small loans to women organized in groups without requiring collateral, which led to its success and replication in over 64 countries worldwide. Women were the primary beneficiaries of this microcredit program and today the Grameen American Bank still focuses on lending to women.

A research study conducted in Bhandara District, India by Sukhdeve (2011) found that rural SHGs had a positive impact on the acquisition of household assets, with all participants reporting an increase. The study found that 76% of SHG members experienced an improvement in their homes' structures and access to electricity, resulting in greater comfort. Additionally, 66% of respondents were able to accumulate savings and 96% reported an increase in their business incomes. According to these findings, the study suggests that SHGs should consist of members of the same age group who share a common goal and receive regular training to stay informed of available services. The study employed logistic regression, which correctly classified 67% of women, indicating that the model was well-fitted to the data.

Alrefaei (2022) conducted a study on SHGs in India that aimed to reduce poverty and promote empowerment. The study identified key features of SHGs, including savings, democratic leadership and offering loans at lower interest rates than other financial institutions. The study found that SHGs contributed to economic development, created job opportunities and improved the quality of life.

The study conducted by Jose et al (2019) aimed to identify the problems encountered by women SHGs in Ernakulam District, India. The data was collected through questionnaires using the SRS technique and the challenges identified included illiteracy, inadequate training and a need for government and voluntary groups to work together to address these issues.

According to LaValley (2008), logistic regression involves analyzing binary outcomes, where the dependent variable is limited to either 0 or 1. Independent variables can be either categorical or continuous. The logistic model predicts the likelihood of an event happening, taking into account multiple predictor variables.

Sukhdeve (2011) utilized logistic regression to investigate how SHGs impacted women empowerment in Bhandara District of Maharashtra:

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

The response variable was represented by Z, while P represented the predictor variables used in the analysis.

$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

An accurate classification of 67% of rural women was made by the model, suggesting that it fit the data well.

Othman et al. (2012) employed logistic regression as one of the models in their study on Malaysian cooperatives, to identify factors that influenced cooperative membership and share increment. The LR model was used to assess independent variables including age, gender and income level on the dependent variable, which was the decision of an individual to join a cooperative or increase their shares. The study found that age and occupation were significant predictor variables in the logistic model and that the model was a suitable fit for the data. The conclusions of the study were that the cooperative performance and progress were dependent on trust and member commitment and a recommendation to reach more young people was made for more growth of the Malaysian cooperatives.

Magali (2013) employed logistic regression to examine the influence of Savings and Credit Cooperative Organizations or SACCOs and credit cooperatives on various factors in Tanzania. The study used health, education, new assets, income, business capital and renovation or building cost as independent variables. The logistic regression analysis revealed that SACCOs had a positive impact on health, education, purchasing new assets, crop yields and business capital. 73.5% of SACCO members had households' improvements, 50% experienced an improvement in health and 80% saw an improvement in physical assets. The study concluded that logistic regression fit the data well.

2.Methodology

2.1 Model Fitting.

Logistic regression is a technique that is utilized when the outcomes are binary, where the anticipated value is either 0 or 1. In this method, each independent variable in the model is assigned a coefficient that gauges its effect on the output variable. These coefficients aid in comprehending the association between the independent variables and the binary outcome (Mbugua,2014).

In this study, the output variable Y had two possible values: 1 means "yes" and 0 means "no". The Y variable represents the wealth status and the study aimed at evaluating whether the wealth status had improved, which is represented by 1, or it had not improved, represented by 0. The coefficients that are assigned are denoted by β. The first coefficient, β_0, is the intercept, while each variable from $1 \ldots \mathrm{n}$ are each assigned the coefficients, $\beta_1, \beta_2 \ldots \beta_n$. They determine the association between the independent and dependent variables. The result obtained from the analysis is in the form of odds ratio;

$$\frac{P}{1-P} = exp(\beta_0 + \sum_{i=1}^{n} \beta_i X_i)$$

The probability of the occurrence of the event is denoted by P.

The equation for the multiple logistic regression function is:

$$P(x) = \frac{exp(\boldsymbol{X}'\boldsymbol{\beta})}{1 + exp(\boldsymbol{X}'\boldsymbol{\beta})}$$

In this research, multiple logistic regression was employed since the prediction of a single binary output was done using multiple variables.

In machine learning, the logistic sigmoid function is widely used to transform any input value to a probability value that ranges between 0 and 1. This is different from the hyperbolic sigmoid function which transforms data values to a range between -1 and 1 and the arctangent function which transforms data values to a range between $\frac{-\pi}{2}$ and $\frac{\pi}{2}$. Below is a graphical representation of the common sigmoid curves.

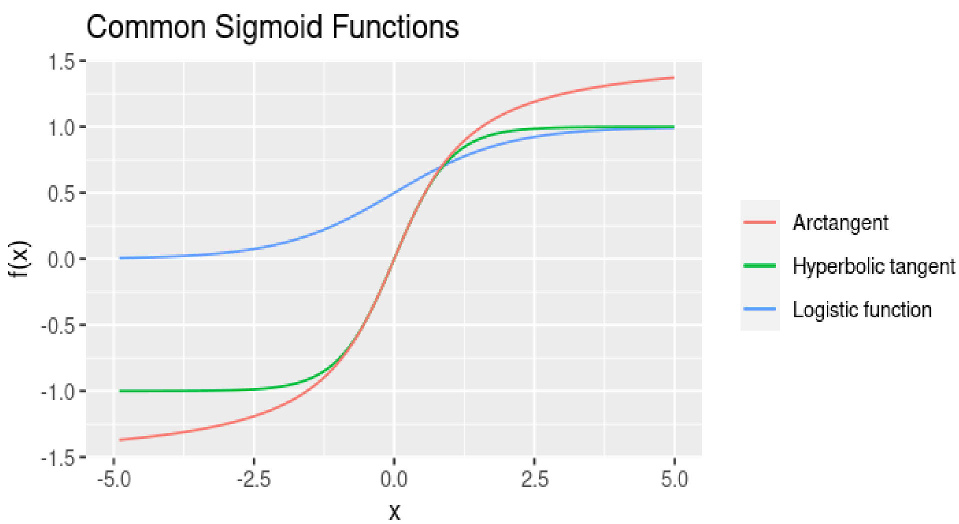

Figure 1: Sigmoid functions (Sigmoid Function Definition | DeepAI, n.d.)

The formula below represents the logistic sigmoid function:

$$P(x) = \frac{1}{1 + \exp(-X)}$$

The LR model is a powerful tool for data analysis that is simple to understand and interpret, computationally efficient and can handle datasets that are large. Additionally, updating new data is easy and it can handle outliers and data that is missing without requiring data preprocessing.

The LR model can provide interpretable results that help explain the relationship between predictor variables and outcomes.

LR has assumptions such as the requirement for a binary or dichotomous outcome variable. In addition, the predictor variables should not have high correlations, as this may affect the estimated parameters and reduce the model's accuracy and effectiveness.

2.2 Parameters Estimation.

The maximum likelihood estimator is a statistical approach that is employed to estimate the parameters of a statistical model. It strives to identify the parameter values that are most likely to account for the observed data, given a specific set of observations. When it applies to LR, the goal is to develop a model that describes the relationship between a binary dependent variable and a set of independent variables.

The form of the logistic regression is as shown below;

$$P(x) = \frac{exp(\boldsymbol{X}'\boldsymbol{\beta})}{1 + exp(\boldsymbol{X}'\boldsymbol{\beta})}$$

Below is a representation of the likelihood function;

$$l(\beta, Y = 1/X = x_i) = \pi_i \left[\frac{\exp(\boldsymbol{X}'\boldsymbol{\beta})}{1 + \exp(\boldsymbol{X}'\boldsymbol{\beta})}\right]^{\sum_i^n y_i} \left(1 - \left[\frac{\exp(\boldsymbol{X}'\boldsymbol{\beta})}{1 + exp(\boldsymbol{X}'\boldsymbol{\beta}))}\right]^{n - \sum_i^n y_i}\right)$$

2.3 Ethical Considerations.

Ethical considerations are an essential aspect of any research project that involves human subjects. This research ensured that the welfare and privacy of the participants was protected. The ethical considerations surrounding this research included consent from respondents, confidentiality, privacy and avoiding harm to participants. Informed consent from the participants was obtained before conducting the research. This means that participants were informed of the research's purpose and benefits before agreeing to participate.

Confidentiality and privacy were also crucial ethical considerations in this research and the steps to protect the participants' identities and personal information were taken ensuring that this information was not disclosed. Additionally, this research ensured that it did not cause any harm to participants, whether it be physical, psychological, or emotional.

The study also ensured there was no coercion or pressure to participate in the study but that it was absolutely by their own choice.

3. Results and Discussions

The data used in the analysis was primary data, obtained by issuance of questionnaires to members of self-help groups. The population (N) was composed of 2250 individuals, from 150 self-help groups, with at least 15 members per group.

Yamane's formula (1967) allowed for the calculation of a representative sample size.

$$n = \frac{N}{1 + N(e)^2} \qquad)$$

Where:

$n =$ size of the sample

$N =$Target population

$e =$ Margin of error. To calculate the size of the sample for this study, an error margin of 0.025 was applied. Using a smaller margin of error resulted in a larger sample size.

$$n = \frac{2250}{1 + 2250(0.025)^2} = 935$$

By this, the size of the sample for this study was determined to be 935 individuals.

The data was split into 80% for training and 20% for testing. A higher percentage of the data was used for training to enable the model to identify patterns within the data and the other percentage was used to test the model's performance.

Logistic regression model fits data with predictor variables and a dependent variable which is in dichotomous form. This section gives the results of the predictor variables when fit in the logistic regression model. Predictor variables such as access to household assets, access to income generating assets, access to loans, access to basic commodities and ability to save were fit to the model to investigate their relationship with the dependent variable, wealth status. The R software was used in the analysis. Standard errors, z-values and p-values which determine the significance of the predictor variables are shown below. Logistic regression in R was fit using the glm function.

The table below shows the results when the predictor variables in this study which included access to household assets, access to income generating assets, access to basic commodities, access to loans and ability to save were fit in the logistic regression model using the R environment. The table shows the estimated coefficients, the std error, the z-value, the p-value and the odds ratio and the 95% CI.

Table 1 Logistic Regression Model Results

Variable	**Estimate**	**Std Error**	**z-value**	**p-value**	**OR**	**95% CI**	
						Upper	**Lower**
Intercept	-5.9973	1.8858	-3.170	0.001526 **	0.0025	-9.72	-2.31
Household Asset	-0.0607	0.2196	-0.276	0.782463	0.9412	-0.497	0.366
Income Generating Assets	0.6633	0.1853	3.580	0.000344 ***	1.9411	0.299	1.027
Basic commodities	0.5154	0.1921	2.684	0.007284 **	1.6743	0.1321	0.888
Loans	0.8611	0.2522	3.415	0.000638 ***	2.3658	0.369	1.361
Savings	0.2494	0.2631	0.948	0.343176	1.2832	-0.267	0.766

Out of the fit predictor variables, access to income generating assets, access to basic commodities and access to loans were found to have significant impact on the dependent variable, wealth status. Savings and access to household assets did not have a significant impact on the dependent variable.

The intercept is -5.9773 $(OR = 0.0025, 95\%\, CI: -9.72 - 2.31)$. It represents the log-odds of the response variable (wealth status) when all other predictor variables are set to 0. The coefficient for access to household assets is -0.0607 ($OR = 0.9412, 95\%\ CI: -0.497\ 0.366$). A one-unit increase in the variable access to household assets is associated with a decrease of 0.0607 in the log-odds (probability of the event occurring to the probability of the event not occurring) of the response variable. However, it is not statistically significant (p-value = 0.782463) the 95% CI also includes zero indicating that the coefficient is not statistically significant at 0.05 level of significance and hence was removed from this study. The coefficient for access to income generating assets is 0.6633 ($OR = 1.9411, 95\%\ CI: 0.299\ 1.027$) .The 95% CI also does not include zero indicating that it is statistically significant and the fact that it is positive shows that the relationship between this variable and the outcome variable is positive, an increase in income generating assets leads to an increase in wealth status. A one-unit increase in the variable access to income generating assets is associated with an increase of 0.6633 in the log-odds of the response variable. It is statistically significant (p-value = 0.000344) and thus has a significant impact on the response variable. The coefficient for access to basic commodities is 0.5154 ($OR = 1.6743, 95\%\ CI: 0.1321\ 0.888$). A one-unit increase in the variable access to basic commodities is associated with an increase of 0.5154 in the log-odds of the response variable. It is statistically significant (p-value = 0.007284). The coefficient for access to loans is 0.8611 ($OR = 2.3658, 95\%\ CI: 0.369\ 1.361$). A one-unit increase in the variable access to loans is associated with an increase of 0.8611 in the log-odds of the response variable. It is statistically significant since its p-value = 0.000638. The coefficient for ability to save is 0.2494 ($OR = 1.2832, 95\%\ CI: -0.267\ 0.766$). A one-unit increase in the variable ability to save is associated with an increase of 0.2494 in the log-odds of the response variable. However, it is not statistically significant (p-value = 0.343176) meaning it does not affect the response variable and hence it was removed from this study.

The table below shows the results when the demographic factors were fit in the logistic regression model. The factors included age, marital status, education level, gender and source of income. The table shows the estimated coefficients, the std error, the z-value, the p-value and the odds ratio and the 95% CI.

Table 2Logistic Regression Results for social demographic variables

Variable	**Estimate**	**Std Error**	**z-value**	**p-value**	**OR**	**95% CI**	
						Lower	**Upper**
Intercept	2.8699	0.6112	4.695	2.66e-06 ***	17.636	-9.7188	-2.3092
Age (years)							
18-30 (ref)	-	-	-	-	-	-	-
31-45	-0.1484	0.3729	-0.398	0.6907	0.8621	-0.9118	0.5618
46-60	-0.2600	0.3899	-0.667	0.5048	0.7710	-1.0512	0.4894
Over 60	-0.4649	0.4154	-1.119	0.2631	0.6282	-1.2986	0.3434

Marital status							
Divorced	-	-	-	-	-	-	-
Married	0.0906	0.3388	-2.316	0.0206 *	1.0949	0.1305	1.4660
Single	-0.7847	0.3778	0.240	0.8105	0.4562	-0.6516	0.8413
Widowed	-0.0009	0.3631	-0.002	0.9981	0.9991	-0.7194	0.7140
Education							
College	-	-	-	-	-	-	-
No Education	-0.4367	0.3824	-1.142	0.2534	0.6462	-1.2026	0.3075
Primary	-0.4542	0.3658	1.242	0.2143	0.6349	-1.1921	0.2522
Secondary	-0.601	0.3730	-1.611	0.1072	0.5484	-1.3521	0.1201
Gender							
Female (ref)	-	-	-	-	-		
Male	0.5479	0.2577	2.126	0.0335 *	1.7295	0.0498	1.0633
Income							
Retired	1.1503	1.0981	1.048	0.2949	3.1590	-0.6549	4.1083
Self Employed	-0.3620	0.425	-0.852	0.3943	0.6963	-1.2793	0.4105
Unemployed	-1.1465	0.5805	-1.975	0.0483 *	0.3177	-2.3125	-0.0044
Employed (ref)	-	-	-	-	-	-	-

Demographic factors such as gender, source of income and marital status were found to be significant predictors of wealth status. Factors such as education level and age were not found to be significant predictors to wealth status.

The intercept, which is the log-odds when all other predictor variables are zero had an estimated coefficient of 2.87 and a p-value 2.66e-06 indicating that the intercept is significantly different from zero $(OR = 17.636, 95\% \, CI: -9.7188 - 2.3092)$.

Age was categorized into four groups i.e., 18-30, 31-45, 46-60 and over 60 years. The category 18-30years was used as the reference category or the baseline category against which the other age categories are compared. The estimated log-odds for this category are assumed to be zero. The estimated coefficient for the age category 31-45 is -0.1484. This represents the change in the log-odds of the response variable for individuals aged 31-45 compared to those aged 18-30. The standard error for this estimate is 0.3729. The z-value is -0.398 and the associated p-value is 0.6907, indicating that the coefficient is not statistically significant. The odds ratio (OR) for this category is 0.8612 while the 95% confidence interval for the odds ratio is between 0.7710 (lower) and 0.9582 (upper). The estimated coefficient for the age bracket 46-60 years is -0.2600, representing the change in the log-odds of the response variable for individuals aged 46-60 compared to those in the reference category. The associated p-value is 0.5048, indicating that the coefficient is not statistically significant. The odds ratio (OR) for this category is 0.7708 while the 95% confidence interval for the odds ratio is between 0.6619 (lower) and 0.8970 (upper).

The estimated coefficient for Over 60 is -0.4649, representing the change in the log-odds of the response variable for individuals aged over 60 compared to those aged 18-30.

The p-value is 0.2631, indicating that the coefficient is not statistically significant. The odds ratio (OR) for this category is 0.6282 and the 95% confidence interval for the odds ratio is between 0.5618 (lower) and 0.7033 (upper). None of the age categories (31-45, 46-60, Over 60) have coefficients that are statistically significant at the 0.05 significance level.

Gender of the respondents was categorized as either male or female. The female category was used as the reference category and hence its coefficient was assumed to be zero. The coefficient for the male gender was 0.5479. This shows that compared to the reference category, being male is associated with an estimated increase in the log-odds (OR=1.7295) of the response variable by 0.5479. The p-value is 0.0335 which is less than 0.05, indicating that this effect is statistically significant at the 0.05 significance level. The 95% CI was 0.0498 (lower) and 1.0633(upper) which does not include zero indicating the variable is statistically significant and the relationship between this predictor variable and the response variable is positive.

The marital status category was divided into single, married, widowed and divorced. Divorced was used as the reference category. The coefficient for marital status married was 0.0906, being married was associated with an estimated increase in the log-odds of the response variable by 0.0906 (OR 1.0949). The p-value of 0.0206 is less than 0.05, indicating that this effect is statistically significant at the 0.05 significance level. The 95% CI was 0.1305 (lower) and 1.4660 (upper). The coefficient for marital status single was estimated to be -0.7847. The negative coefficient suggests a decrease in the log-odds for individuals who are single. The high p-value (0.8105) indicates that this category is not statistically significant, meaning we fail to reject the null hypothesis that the coefficient is zero. The OR were 0.4562 and the 95% CI was -0.6516 (lower) and 0.8413 (upper). The coefficient for marital status widowed was -0.0009, the p-value was (0.9981) suggesting that this category was not statistically significant $(OR = 0.9991, 95\%\ CI{:}\ -0.7194\ \ 0.7140)$.

The level of education was categorized into no formal education, primary education, secondary education and college or tertiary level. The college category was used as the reference category. The coefficient for no formal education was -0.4367 implying a negative impact on the log-odds for individuals with no education. The p-value was 0.2534 $(OR = 0.6462, 95\%\ CI{:}\ -1.2026\ \ 0.3075)$. This category was not found to be statistically significant. Primary education had a coefficient -0.4542 having only primary education appears to have a negative impact on the log-odds compared to the reference category, but the effect is not statistically significant (p-value = 0.2143) $(OR = 0.6349, 95\%\ CI{:}\ -1.1921\ 0.2522)$. The coefficient for secondary education (-0.601) suggested a negative impact of having secondary education on the log-odds, but the effect was not statistically significant (p-value = 0.1072) $(OR = 0.5484\ CI{:} -1.3521\ 0.1201)$.

The last category was source of income which was divided into unemployed, self-employed, employed and retired. The employed category was used as the reference category. Individuals with a source of income as retired had a positive impact on the log-odds, coefficient 1.1503, but the effect was not statistically significant (p-value = 0.2949) $(OR = 3.1590\ 95\%\ CI{:} -0.6549\ 4.1083)$.The coefficient (-0.3620) for the self-employed category suggested a negative impact on the log-odds for individuals who are self-employed, but again, the effect was not statistically significant (p-value = 0.3943)$(OR = 0.6963\ 95\%\ CI{:} -1.2793\ 0.4105)$. Individuals with a source of income as unemployed had a negative impact on the log-odds, coefficient -1.1465 and the effect was statistically significant at a 5% significance level (p-value = 0.0483) $(OR = 0.3177\ 95\% CI{:} -2.3125 - 0.0044)$.

The logistic regression model was then fit using only the significant predictor variables and the coefficients of this model were extracted. The logistic regression equation with these coefficients is expressed as shown below.

$$ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = -4.4359 + 0.5915\ x_1 + 0.4994x_2 + 0.7142x_3$$

Where;

x_1 = Income generating assts

$x_{2=}$Basic Commodities

$x_{2=}$ Access to loans

The model classified 84 members under class 0 which represented a total percentage of 8.67% and 885 under class 1 which was a percentage of 91.33% of the entire dataset. The accuracy of the logistic regression model in classifying members of SHG in this study as either having an improved wealth status or not since joining a SHG was 88.04%.

The model's performance was evaluated by construction of a confusion matrix. The metrics of a confusion matrix such as sensitivity, precision, F1 score and specificity were used and the results are as shown below. The metrics of the confusion matrix are calculated using the test set or the validation set. The test set tests the model's performance on new, unseen data that it has not been exposed to during training

Table 3 Confusion Matrix for Logistic Regression Model

	Actual Positives	Actual Negatives
Predicted Positives	179 (TP)	11 (FP)
Predicted Negatives	14 (FN)	5 (TN)

$$Recall\ or\ sensitivity = \frac{TP}{FN + TP} = \frac{179}{14 + 179} = 0.9275 = 92.75\%$$

$$Precision = \frac{TP}{FP + TP} = \frac{179}{11 + 179} = 0.9421 = 94.21\%$$

$$Accuracy = \frac{TP + TN}{TN + TP + FN + FP} = \frac{179 + 5}{5 + 179 + 14 + 11} = 0.8804 = 88.04\%$$

$$F1\ Score = \frac{2 * (sensitivity * precision)}{sensitivity + precision} = \frac{2 * (0.9275 * 0.9421)}{0.9275 + 0.9421} = 0.9347 = 93.47\%$$

4.Conclusion and Recommendation

As earlier indicated, approximately 8.9 million people, or 17% of Kenya's population, live below the poverty line of 1.9 USD on a daily basis a majority of them in the rural areas. Several tools have been used to reduce and even completely eliminate poverty in the best-case scenarios. One of those tools is SHGs. This is what prompted this research to establish what impact these groups have had on lives and livelihoods in terms of wealth status of the group members.

These establishments were arrived at using logistic regression.The model achieved a high level of accuracy in the analysis which can be attributed to the high amount of data used for training which enabled the model to identify patterns within it.

In conclusion, SHGs had had a positive impact on majority of the members and hence it is a tool that can be recommended for use in the fight against poverty in our society.

A recommendation on further studies to help understand poverty and how it can be eliminated or reduced in Kenya and even in Africa should be done and other tools that can be used to achieve this discussed. This will help place Africa in the global stage and improve lives and livelihoods within the continent. The goal to make people live decent lives within the continent should be a topic that should be widely researched on and explored to ensure that lives are made better.

A future study on groups that were started but collapsed should be made to further understand what members of those groups have to say on factors affecting group performance.

Also, since most of the members responded that interest rates on loans were higher for some SHGs than other financial institutions, research should be done on this to establish what should be done to make the loans more affordable for members.

References

[1] Alrefaei, N., Aquinas, P. G., & Al-Maamari, O. A. (2022). *Self-help group (SHG) in India*: a path toward empowerment and poverty reduction. Social Work with Groups, 1-15.

[2] Brody, C., Dworkin, S., Dunbar, M., Murthy, P., & Pascoe, L. (2013). PROTOCOL: *The Effects of Economic Self-Help Group Programs on Women's Empowerment*: A Systematic Review. Campbell Systematic Reviews, *9*(1), 1-43.

[3] Deininger, K., & Liu, Y. (2009). *Economic and social impacts of SHGs in India*. WorldBank Policy Research Working Paper, (4884).

[4] Hoffmann, V., Rao, V., Surendra, V., & Datta, U. (2021). *Relief from usury: Impact of a self-help group lending program in rural India*. Journal of Development Economics, *148*, 102567.

[5] Jose, S., Chockalingam, S. M., & Velmurugan, R. (2019). *Problems of Women Self Help Group Members in Ernakulam District*. Journal of Critical Reviews, *7*(1), 141-143.

[6] Magali, J. (2013). *Impacts of rural savings and credits cooperative societies (SACCOS') loans on borrowers in Tanzania*. International Journal of Management Sciences and Business Research, 2(12).

[7] Mbugua, M. D. (2014). *Application of logistic regression in identifying key determinants of Domestic Violence in Kenya* (Doctoral dissertation, University of Nairobi).

[8] Mwai, J. (2017). *Impact of Saccos in Poverty Eradication among Farmers in Murang'a County, Kenya.*

[9] Othman, A., Kari, F., Jani, R., & Hamdan, R. (2012). *Factors influencing cooperative membership and share increment: an application of the logistic regression analysis in the Malaysian cooperatives.* World Review of Business Research, *2*(5), 24-35.

.

[10] Serrat, O. (2017). *The sustainable livelihoods approach*. In Knowledge solutions (pp. 21-26). Springer, Singapore.

[11] Sukhdeve, M. I. (2011). *Factors Influencing Empowerment of Rural SHG Women in Bhandara District of Maharashtra: Logistic Regression Model*. International Journal of Microfinance, 1(1), 102-108.

[12] Wright, R. E. (1995). Logistic *regression*. *Reading and understanding multivariate statistics* (pp. 217–244).