FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 8 No. 3, June, (Special Issue) 2024, pp 416 - 421
DOI: https://doi.org/10.33003/fjs-2024-0803-2524

UTILIZING LOGISTIC REGRESSION TO IDENTIFY HOUSEHOLD POVERTY STATUS IN BENUE STATE OF NIGERIA

Gaav, T. A., *Akumba, B. O., Otor, S. U., and Agber, S.

Department of Mathematics/Computer Science, Benue State University Makurdi, Nigeria

*Corresponding authors' email: beatriceakumba@gmail.com

ABSTRACT
Despite numerous poverty alleviation programmes in Nigeria, household poverty continues to rise. This paper aimed to improve poverty detection by analyzing household socio-economic and environmental data using logistic regression models. Data was collected through Google forms and pre-processing and feature extractions were performed. Five logistic models were developed to represent different types of poverty. The models were evaluated for accuracy, precision, recall, and F1 score, achieving 80% for Absolute Poverty and 100% for Chronic/Structural, Conjectural/Transitory, and Locational/Spatial poverty. The analysis revealed that over half of the households were poor, with Spatial/Locational poverty being the most prevalent. Targeting this type of poverty is recommended for effective intervention.

Keywords: Detection, Poverty Status, Logistic Regression, Household, Algorithm

INTRODUCTION
Poverty, impacting both physiological and social well-being, is defined by the World Bank as unacceptable deprivation and by the United Nations (UN) as lacking the means for a minimum standard of living (WB, 2020; UNDP, 2019). Despite global efforts to eradicate poverty by 2030, extreme poverty persists in regions like South Asia and Sub-Saharan Africa, affecting about 80% of the population (IMF, 2017). In Nigeria, 63% of the population, or 133 million people, live in multidimensional poverty, worsened by the COVID-19 pandemic (NBS, 2022). Rural areas and children are especially affected. For instance in Kaduna State, about 51% of rural farmers are poor, underscoring the severe impact on agricultural households (Adamu et al., 2021) in Nigeria. Accurate poverty tracking is crucial for effective interventions, but traditional methods have limitations (Vatcheva et al., 2016). Machine learning, particularly Logistic Regression, offers a promising alternative (Blumenstock et al., 2015; GeeksforGeeks, 2023). This paper explores using reduced-variable datasets for better poverty detection, aiming to improve poverty alleviation strategies.
In the past decade, machine learning (ML) has been increasingly applied to poverty-related data, demonstrating accuracy, cost-effectiveness, and real-time assessment capabilities (Blumenstock et al., 2015; Morten, 2013; Emily, 2022). Research in Pakistan using logistic regression identified education and remittances as significant factors in reducing poverty (Majeed and Malik, 2015). Studies in Hong Kong and Jordan utilized logistic regression and other ML approaches to analyze poverty determinants, providing detailed insights (Chenhong et al., 2019; Adham et al., 2021). In Costa Rica, Ji (2021) highlighted education as the most influential factor in predicting poverty status. Alsharkawi et al. (2021) employed LightGBM and Bagged Decision Trees in Jordan, achieving accurate poverty classification. Liu et al. (2021) identified village-level factors such as access to industry, banks, and towns as key poverty determinants in rural China.
Wang & Shi (2021) introduced an RF-PCA model for predicting poverty levels among college students, achieving high accuracy. Satej et al. (2022) explored adaptive sampling strategies to enhance poverty prediction models, while Aziza (2022) reviewed advancements in AI-based poverty prediction, noting improvements in speed, accuracy, and dataset diversity. Pa et al. (2022) proposed a machine learning approach emphasizing data pre-processing and feature engineering, with the Random Forest Regressor performing best in poverty prediction.
Rekha et al. (2023) utilized Deep Neural Networks (DNNs) to predict poverty levels from satellite imagery, finding that DNNs were effective but dependent on the quality and size of the training dataset. Dedy et al. (2024) leveraged e-commerce data and support vector regression to assess socio-economic conditions in Indonesia, addressing high-dimensional data challenges with feature selection algorithms. Abdirizak et al. (2024) applied advanced ML methods to Somalia's demographic data, finding the random forest model most accurate, with key predictors including geographical region, household size, and respondent age group.
These studies collectively showcased the versatility and effectiveness of machine learning in addressing various aspects of poverty prediction and alleviation.

MATERIALS AND METHODS
The paper utilized a population-based cross-sectional study design, and primary data source. The required dataset are the socio-economic characteristics of households in Benue State. These characteristics include the household; income or consumption level, education attended, access to social and merit goods, physical fitness and environment factors which formed the set of explanatory variables of the model. Python was used as the programming language in a Google Colaboratory (Google Colab) Environment to develop a framework for the models. The Matplotlib and chart.js were used to plot the necessary graphs on the user interface. The study also employed different Python and Java libraries, which provided the bulk of pre-processing tools, ML algorithms and data visualization.

Data Collection
A dataset of 25 parameters were collected from 1,108 households across Makurdi Local Government Area of Benue State using google forms. This allowed the model to train and test various possible variations of the rate of types of poverty among sampled households. Noise, gaps and null values were associated with the collected raw data. In order to ensure a perfect outcome of the model, the raw data collected were subjected to preliminary processing stages of importing

Python standard libraries, checking the dimensions of the dataset, data cleaning and splitting the dataset into training and testing on a ratio of 80:20. The uploaded dataset contained binary categorical values of YES and No which were the responses of the sampled households. The loaded categorical dataset (Yes and No) were converted to binary digits of 0's and 1's, where yes was converted to 0 and No was converted to 1, in order to be accepted by the Logistic Regression model which only takes in binary digits (0s and 1s).

Logistic Regression Algorithm

According to Sonia (2022) Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of certain classes based on some dependent variables. In short, the logistic regression model computes a sum of the input features (in most cases, there is a bias term), and calculates the logistic of the result. The output of logistic regression is always between (0, and 1), which is suitable for a binary classification task. The higher the value, the higher the probability that the current sample is classified as class=1, and vice versa.

Consider that the parameters of the logistic regression model are estimated by maximum likelihood, with the likelihood function formed by assuming independence over the observations. Then the work adopted the sigmoid function: $p(Y) = \frac{1}{1+e^{-z}}$ (1)

Where:

$p(Y)$is the output (probability),

e is the base of the natural logarithm (approximately 2.71828), and

$z$is the linear combination of input features and their associated weights:

$z = \beta_0 + \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_n x_n + U_i$ (2)

If Y measures poverty status, then $p(Y)$ is the probability that **Y** might be poor (1) or non-poor (0). By taking logs and simplifying equation (1), the log likelihood simplifies to:

$LnY_i Log = {}^{P_i}/_{1 - P_i} = \beta_0 + \sum_{j-1}^{\beta} k\, X_{ki} + U_i$ (3)

Where:

$LnY_i Log$ = Natural log of Y (Household Poverty Status);

X_{ki} = A set of household's socio-economic and environmental characteristics (determinants of poverty status of individual/household);

β_k = Parameters;

U_i = A random disturbance term.

Relative Poverty Detection Model

This identifies someone who earns an income below the estimated poverty line of the population. The poverty line, in this case, is calculated as half of the median income value of the population (the gap between the rich and the poor). Based on this, the new daily per head global extreme poverty line of \$2.15, equivalent to ₦3,225 at an exchange rate of \$1 to ₦1,500, replaced the \$1.90 poverty line, which was equivalent to ₦634 at an exchange rate of \$1 to ₦333.71, based on 2017 Purchasing Power Parities (PPPs) (World Bank, 2022).

Therefore;

Median Income $(\boldsymbol{MI})$ is calculated as:

i. If the total number of observations gives odd, then the formula to calculate the **Median Income** $(\boldsymbol{MI})$ is given as:

$\boldsymbol{Median} = \left(\frac{N+1}{2}\right)^{th}$ term. So the Median Income (MI) is income at Median.

Therefore, Poverty line $(\boldsymbol{PL}) = \frac{\boldsymbol{MI}}{\boldsymbol{2}}$

ii. If the total number of observations gives even, then the formula to calculate the **Median Income** $(\boldsymbol{MI})$ is given as:

$$\boldsymbol{Median} = \frac{\left(\frac{N}{2}\right)^{th} \boldsymbol{term} + \left(\frac{N}{2}+1\right)^{th} \boldsymbol{term}}{2}\ \boldsymbol{term}$$

So the Median Income (MI) is income at Median.

Therefore, Poverty line $(\boldsymbol{PL}) = \frac{\boldsymbol{MI}}{\boldsymbol{2}}$

Where N is the number of household sampled.

Decision Rule

If the household's income $(\boldsymbol{HI})$ is less than Poverty line $(\boldsymbol{PL})$, then the household is suffering from Relative Poverty(P_2).

RESULTS AND DISCUSSION

The Logistic Regression model from the study achieved an accuracy of 0.80, with perfect recall and precision scores of 1.00, indicating that it correctly identified all cases of poverty without false positives. This performance surpassed other models in recall and precision, demonstrating its effectiveness in poverty detection based on household socio-economic and environmental characteristics in Benue State of Nigeria.

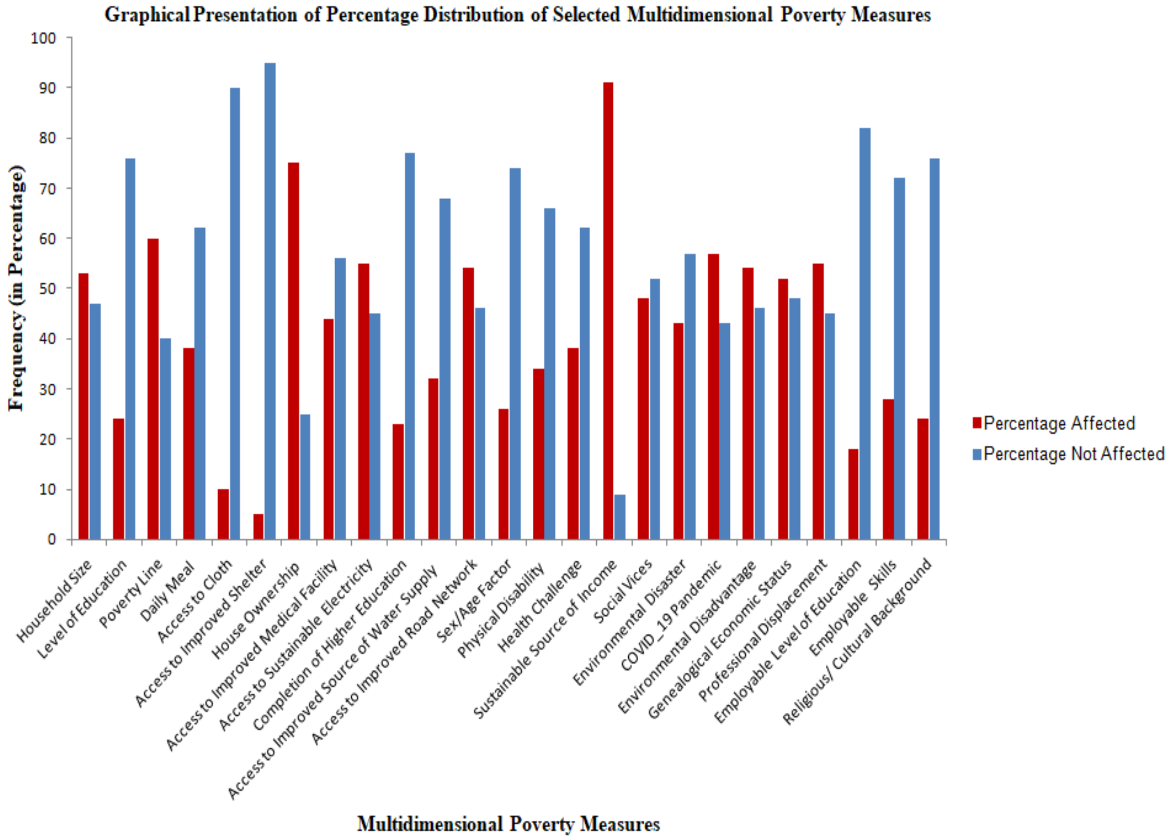

Figure 1: Percentage Distribution of Selected Multidimensional Poverty Measures

Figure 1 illustrates the percentage distribution of selected multi-dimensional poverty measures, highlighting the proportion of the population affected and not affected by various poverty indicators. The key areas of concern include sustainable income, house ownership, and living below the poverty line, with over 50% of the population impacted in these categories. In contrast, access to improved shelter and clothing shows minimal impact, affecting less than 10% of the population. This visual representation underscored the diverse challenges faced and areas needing targeted intervention.

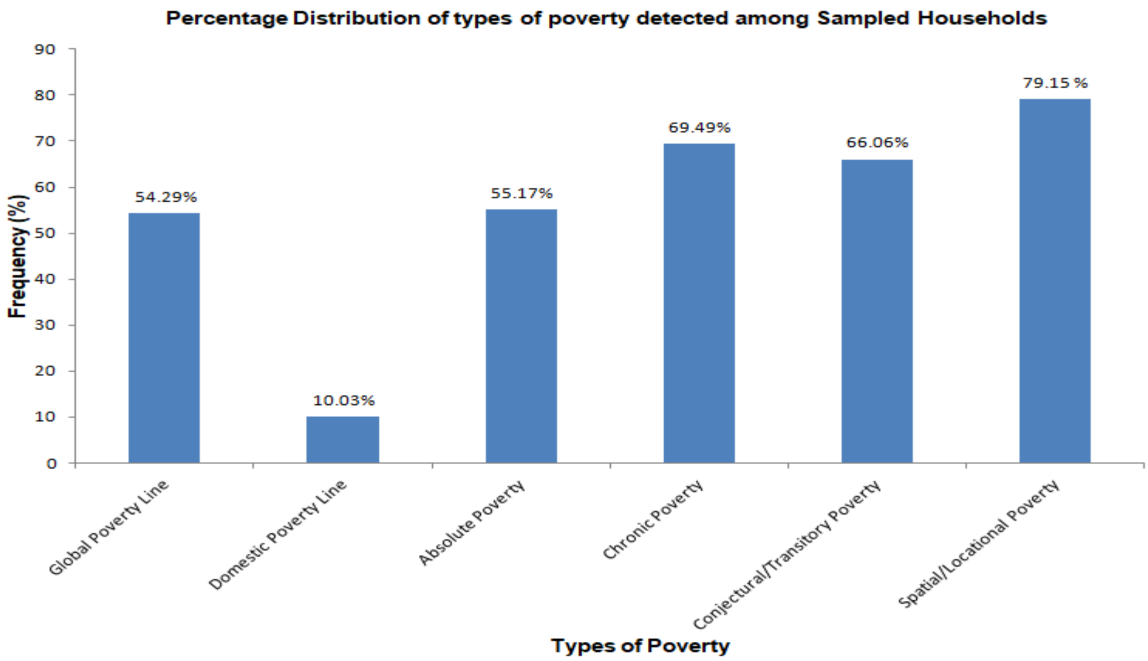

Figure 2: Graph of rate of types of poverty detected among Sampled Households

Figure 2 illustrates the distribution of the household poverty types detected. Spatial/Locational Poverty was the most prevalent, affecting 79.15% of households, followed by Chronic Poverty at 69.49% and Conjectural/Transitory Poverty at 66.06%. Absolute Poverty impacts 55.17%, while Relative Poverty (Global Poverty Line) affects 54.29%. Relative Poverty (Domestic Poverty Line) is the least common, impacting only 10.03% of households, using the estimated Domestic Poverty Line (DPL) of $0.90 equivalent of ₦1,350 at an exchange rate of $1 to ₦1,500 per day. This analysis highlighted the varying prevalence of poverty types, aiding in targeted intervention planning.

Table 1: Result of Household's Poverty Status Detection Model

Poverty Status	Number of Sampled Households	Percentage (%)
Poor	630	56.86
Not Poor	478	43.14
Total	1,108	100.00

The models performance evaluation metrics used in this paper are; Accuracy, F1_Score, Recall, Mean Square Error (MSE), Relative Mean Square Error (RMSE), R-Squared (R-Squared) and Precision and their respective values are as presented on Table 2. The performance metrics for the logistic regression model showed perfect scores (accuracy, F1, recall, precision, R Square) of 1.0 for Chronic, Conjectural/Transitory, and Spatial/Locational Poverty, indicating excellent model performance for these categories. Absolute Poverty metrics, with scores of 0.8, MSE of 0.2, and RMSE of 0.4, are still satisfactory, indicating a good predictive/detective performance with slightly more variability in predictions compared to the other categories.

Table 2: Result of Performance Metrices of the Estimated Logistic Regression

Performance Metrics	Absolute Poverty	Chronic Poverty	Conjectural/Transitory Poverty	Spatial/Locational Poverty
Accuracy	0.8	1.0	1.0	1.0
F1 Score	0.8	1.0	1.0	1.0
Recall Score	0.8	1.0	1.0	1.0
Precision	0.8	1.0	1.0	1.0
MSE	0.0	0.2	0.0	0.0
RMSE	0.0	0.4	0.0	0.0
R Square	0.98	0.98	0.98	0.98

The results from the estimated model revealed that 56.86 % of the sampled households in Benue State were poor. This result is slightly higher that 32.9% report of statista, 2022 (Doris, 2022). It was also detected that, 54.29% of the sampled households where living below poverty line of $2.15 per day while 10.03% were domestically relatively poor based on the estimated poverty threshold of $0.90 per day. It was also revealed that, 55.17% of the sampled households were absolutely poor while the per cent of the households affected by chronic/structural poverty was 69.49%. Also, 66.06% of the sampled households were detected of conjectural/transitory poverty while spatial /locational poverty top with 79.15%.

Based on the outcomes of the performance metrics in Table 2, the model was seen to have performed well with high accuracy, F1 score, recall, and precision values respectively. The MSE and RMSE values indicated low average deviations between the predicted and actual values, which suggested a good predictive/detective performance for logistic regression tasks. The R-Square value of 0.98 indicated a perfect fit to the data, implying that the model explained all the variability observed in the dependent variable.

CONCLUSION

In this paper, we employed Logistic Regression algorithm to develop a model that was able to detect household's poverty status and the rate at which each type of poverty affected some households in Benue State. The model indicated that more than half of the sampled households were classified as poor. It also detected that an average household in Benue State was living below global poverty line of $2.15 equivalent to ₦3,225 at an exchange rate of $1 to ₦1,500, per day. Using the estimated Domestic Poverty Line (DPL) of $0.90, equivalent of ₦1,350 at an exchange rate of $1 to ₦1,500 per day, only10.03% of sampled households were classified as domestically relatively poor. The work further identified Spatial/Locational Poverty as the major types of poverty predominantly affected by most households in Benue State, followed by chronic poverty. Moreover, the study detected more than half of the sampled households affected by absolute and chronic/structural poverty respectively. Similarly, more than half of the households were affected by conjectural/transitory. The performance metrics presented in Table 2 indicated that the model performed well, however showed that the system had a good ability to predict or detect poverty status. The MSE and RMSE values were low, indicating a small average deviation between predicted and actual values. This implies that the model had a good predictive/detection performance for logistic regression tasks. Furthermore, the R-square value of 98 % indicated a good fit to the data, suggesting that the model explained all the observed variability in the dependent variable.

REFERENCES

Adamu, B. D., Tanko, F., Barnabas, T. M. a. & Adejoh, E. U. (2021). Determinants of Household's Poverty Among Crop Farmers In Kaduna State, Nigeria. FUDMA Journal of Sciences (FJS), 5(1):529-530.

Abdirizak A. H., Abdisalam H. M. & Christophe C. (2024). Machine learning study using 2020 SDHS data to determine poverty determinants in Somalia. https://www.nature.com/articles/s41598-024-56466-8

Adham, A., Mohammad, A., Maha, D., Heba, S. & Musa, A. (2021). Poverty Classification Using Machine Learning: The Case of Jordan. Multidisciplinary Digital Publishing Institute. https://www.researchgate.net/publication/348898452_Poverty_Classification_Using_Machine_Learning_The_Case_of_Jordan .

Alsharkawi, A., Al-Fetyani, M., Dawas, et al. (2021). Poverty Classification Using Machine Learning: The Case of Jordan Sustainability. https://www.mdpi.com/2071-1050/13/3/1412

Aziza Usmanova (2022). Utilities of Artificial Intelligence in Poverty Prediction: A Review. Sustainability. Econpapers. https://econpapers.repec.org/article/gamjsusta/v_3a14_3ay_3a2022_3ai_3a21_3ap_3a14238-_3ad_3a959346.htm

Blumenstock, J. E., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. https://www.unhcr.org/innovation/wp-content/uploads/2016/11/blumenstock-science-2015.pdf

Chenhong P., Lue F., Julia S. W. & Paul S. F. Y. (2019). Determinants of Poverty and Their Variation Across the Poverty Spectrum: Evidence from Hong Kong, a High-Income Society with a High Poverty Level. Social Indicators Research, 144(1), 219–250.DOI: 10.1007/s11205-018-2038-5

Dedy R., Raden I., Fadhilah I., Elis H. & Wawa W. (2024). Poverty prediction using E-commerce dataset and filter-based feature selection approach. Scientific Report. https://www.nature.com/articles/s41598-024-52752-7

Doris D. S. (2022). Poverty headcount rate in Nigeria 2019, by state. https://www.statista.com/statistics/1121438/poverty-headcount-rate-in-nigeria-by-state/#statisticContainer

Emily, A., Suzanne, B., Dean, K., Chris, U. & Joshua, E. B. (2022). Machine Learning and Phone Data can Improve Targeting of Humanitarian Aid. https://www.nature.com/articles/s41586-022-04484-9.

GeeksforGeeks (2023). Logistic Regression in Machine Learning. Retrieved from https://www.geeksforgeeks.org/understanding-logistic-regression/ .

International Monetary Fund. (2017). Malawi Economic Development Document.

Ji , Y. K. (2021). Using Machine Learning to Predict Poverty Status in Costa Rican Households. Johns Hopkins University, Carey Business SchoolWashington, DC. https://www.semanticscholar.org/reader/fe12f04fdd4d38a94909ce8a008d7ddb06c3c0de.

Liu, M., Hu, S., Ge, Y., Heuvelink, G. B., Ren, Z., & Huang, X. (2021). Using multiple linear regression and random forests to identify spatial poverty determinants in rural China. Spatial Statistics, 42, 100461. https://research.wur.nl/en/publications/using-multiple-linear-regression-and-random-forests-to-identify-s.

Majeed, M. T., & Malik, M. N. (2015). Determinants of Household Poverty: Empirical evidence from Pakistan. The Pakistan Development Review, 54(4I-II), 701–718. https://www.jstor.org/stable/43831356#:~:text=determinants,.-6.&text=of%20household%20poverty%20in%20Pakistan,location%2C%20household%20size%20and%20remittances .

Morten, J. (2013). Poor numbers. In Poor Numbers. Cornell University Press.

National Bureau of Statistics. (2022). Nigeria Launches its Most Extensive National Measure of Multidimensional Poverty. Press Released in Abuja. https://nigerianstat.gov.ng/news/78#:~:text=Highlights%20of%20the%202022%20Multidimensional,quarter%20of%20all%20possible%20deprivations .

Pa, P. M., Yen, W. G., Siti, N. B. H., Thian, S. O., & Shohel, S. (2022). Poverty Prediction Using Machine Learning Approach. Journal of Southwest Jiaotong University. Faculty of Information Science and Technology, Multimedia University Melaka, Malaysia. http://jsju.org/index.php/journal/article/view/1162

Rekha G S, Shivanshu P, Samarth C., Shweta P. & Rohan S. (2023). Poverty Detection Using Deep Learning and Image Processing. IJraset Journal for Research in Applied Science and Engineering Technology. DOI Link: https://doi.org/10.22214/ijraset.2023.50518

Satej, S., Emily, A., Esther, R. & Joshua, B. (2022). Can Strategic Data Collection Improve the Performance of Poverty Prediction Models? 36th Conference on Neural . Information Processing Systems (NeurIPS 2022). https://www.researchgate.net/publication/365448864_Can_Strategic_Data_Collection_Improve_the_Performance_of_Poverty_Prediction_Models .

Sonia, J. (2022). How Does Logistic Regression Work. https://www.kdnuggets.com/author/sonia-jessica.

United Nations Development Programme. (2019). Multi-Dimensional Poverty ReportReveals Wide Inequalities Among the Poor United Nations. The Sustainable Development Goal Report 2019. New York: United Nations. https://sdg.iisd.org/news/undp-multi-dimensional-poverty-report-reveals-wide-inequalities-among-the-poor/.

Vatcheva, K. P., Lee, M., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. Epidemiology (Sunnyvale, Calif.), 6(2), 227.

Wang S. and Shi, Y. (2021). Prediction Poverty Levels of College Students Using a Machine Learning Model. Chuzhou University. Posted Date: https://doi.org/10.21203/rs.3.rs-919541/v1 .

World Bank (2020). DataBank. https://databank.worldbank.org/source/poverty-and-equity-database. https://econpapers.repec.org/article/gamjsusta/v_3a14_3ay_3a2022_3ai_3a21_3ap_3a14238-_3ad_3a959346.htm

World Bank (2022).Understanding Poverty 2022 https://www.worldbank.org/en/topic/poverty/overview.

APPENDIX
Questionnaire

1. Do you have up to three (3) dependents? (YES/NO)
2. Have you completed any higher level of education (post-secondary, vocational training, etc.)? (YES/NO)
3. Is your monthly income up to ₦30,000? (YES/NO)
4. Do you and your household members feed at least three times a day? (YES/NO)
5. Do you and your household members buy at least one new cloth in a year? (YES/NO)
6. Are you staying in a zinc roof and cemented walls/floors house? (YES/NO)
7. Do you own a house? (YES/NO)
8. Do you and your household members afford medical bills of standard private, specialized and general hospital? (YES/NO)
9. Do you and your household members have electricity for at least twelve (12) hours per day? (YES/NO)
10. Did any of your household member complete any higher (Advanced) level of Education (college degree, professional certification, etc.)? (YES/NO)
11. Do you have access to very good drinking water? (YES/NO)
12. Is there a good road network in your place of residence that can facilitate business? (YES/NO)
13. Your sex or age does not limit your chances of generating income? (YES/NO)
14. You do not have any physical disability? (YES/NO)
15. You do not have any sickness that affects your job? (YES/NO)
16. Are you satisfied with your current source of income? (YES/NO)
17. You or any of your household member have never been displaced by war, crises, attack or violence? (YES/NO)
18. You or your household members have never been affected by flood or any other environmental disaster? (YES/NO)
19. COVID_19 did not affect your level of income generation? (YES/NO)
20. Your current place of residence is not affecting your chances of employability? (YES/NO)
21. Your parents were/are not poor? (YES/NO)
22. You have never lost a job because of change in organisational or governmental policy? (YES/NO)
23. Do you have any certificate that can secure a good job for you? (YES/NO)
24. Do you have any advanced skills that can get you a good job? (YES/NO)
25. Your religion or cultural background does not affect the level of your income generating? (YES/NO)