TECHNICAL UNIVERSITY OF KENYA

SCHOOL OF SURVEYING AND SPATIAL SCIENCE

DEPARTMENT OF GEOINFORMATION & EARTH OBSERVATION

ESTIMATING MULTIDIMENSIONAL CHILD POVERTY IN KENYA

NAME: JOHN PAUL MBUTHIA GITATA

REGISTRATION NUMBER: ESEI/01263/2018

SUPERVISED BY: DR. SIMEON KANANI PHD

A research project submitted to the School of Surveying and spatial science in fulfillment for award of a degree in Bachelor of Technology (Geoinformation Technology)

March 2023

DECLARATION BY STUDENT

I, **John Paul Mbuthia Gitata**, declare that this is my original work and has never been submitted for the award of a degree by anyone in any university.

Name: John Paul Mbuthia

University Registration Number: ESEI/01263/2018

Signature:..

DECLARATION BY SUPERVISORS

Dr. Francis Oloo, Date …………………………….

We confirm that the work reported in this project was carried out by the candidate, under our supervision. We have read and approved this research report for examination.

Dr. Simeon Kanani

Senior Lecturer

Department

Geoinformation and Earth Observation, Technical University of Kenya

ACKNOWLEDGMENT

No amount of human effort, in my opinion, can lead to success on its own. It simply makes sense to acknowledge that our Father, who art in Heaven, is responsible for every success, regardless of its magnitude. Please accept my gratitude to you, God, for your providence in all of its manifestations, despite my good health, which together allowed me to come up with this research. I want to express my profound gratitude to my supervisor, Dr. Simeon Kanani, for his continual guidance during the course of this project. I'm thrilled and sincerely appreciative of the help. God bless you, please.

ABSTRACT

This research focuses on estimating and predicting multidimensional child poverty in Kenya. According to UNICEF, an estimated 356 million children live in poverty. The global Multidimensional Poverty Index (MPI) is an international tool that measures severe multidimensional poverty in more than 100 developing nations. Demographic Health Survey (DHS) has conducted more than 400 surveys in over 90 countries. This study will use the 2020 survey data of Kenya to estimate multidimensional child poverty in the country. After having the target variables and estimations, a machine learning algorithm, specifically a random forest regressor was used to predict multidimensional child poverty in Kenya. Highly accurate multidimensional child poverty maps are then created, which show the estimations and the predictions. These predictions were used to create geospatial visualizations in the form of poverty maps that could help county governments and organizations drafting policies for local planning. The results show the counties that have the highest scores of child poverty in Kenya such as Wajir and Mandera. The outputs from this study are crucial for governments and organizations to create policies and plans for improving the quality of life of children in Kenya.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

DHS	Demographic Health Survey
MPI	Multidimensional Poverty Index
MDCP	Multidimensional Child Poverty
GIS	Geographical Information System
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
RS	Remote Sensing
DNMP	Division of National Malaria Programme
KNBS	Kenya National Bureau of Statistics
MOH	Ministry of Health
GDP	Gross Domestic Product
PPP	Purchasing Power Parity
SDGs	Sustainable Development Goals
cf_cvg	Cloud Free Coverages
GPS	Global Positioning System
CRS	Coordinate Reference System
KNN	K-Nearest Neighbor

CHAPTER ONE: INTRODUCTION

1.1 Background

In 2023 more than 700 million people live in poverty with the largest percentage being children with half of these people living at less than $1.90 a day. According to UNICEF, an estimated 356 million children live in poverty (Ani Rudra & Solrun & Jose Cuesta & David Newhouse & David Stewart, 2019). Child poverty describes when a child is raised with limited access to or no access to the essential resources they need to survive and live well (World Vision UK, 2023). Children who have grown up in poverty often suffer from a lack of food (nutrition), sanitation, healthcare, housing, water, and education compared to other children. An important international tool for measuring severe multidimensional poverty in more than 100 developing nations is the global Multidimensional Poverty Index (MPI) (Sabina Alkire and James Foster, 2009). First introduced in 2010 by the United Nations Human Development Report Office and the Oxford Poverty and Human Development Project at the University of Oxford.

To track deprivations across 10 indicators including health, education, and standard of living, the global MPI first creates a deprivation profile for each family and everyone living in it (Alkire, Chatterjee, 2014). A household and all of its occupants are considered to be deprived, for instance, if any child is stunted or any child or adult for whom data are available is underweight; if at least one child has died in the previous five years; if any school-age child is not enrolled in school until the age at which he or she would complete class 8; or if no occupant of the household has completed six years of education; or if the household lacks access to electricity or an improved source of drinking water within a 30-minute walk round trip, an improved sanitation facility that is not shared, non-solid cooking fuel, durable housing materials, and basic assets such as a radio, animal cart, phone, television or bicycle (Alkire, Adriana, Gisela, José M, María Emma, Suman & Ana, 2016). From these multidimensional factors implemented by UNICEF, which are nutrition, sanitation, healthcare, housing, water, and education, the estimation of child poverty in the whole of Kenya is possible. Demographic and health survey programs (DHS) over the years have been able to collect, analyze, and disseminate accurate and representative data on population, health, HIV, and nutrition through more than 400 surveys in over 90 countries including Kenya (Guide to DHS Statistics DHS-7 the Demographic and Health Surveys Program, n.d.). Surveys in Kenya have been conducted in the following years, 1989, 1993, 1998, 1999, 2003, 2004, 2008-09. 2010, 2014, 2015, and 2020. In this study, the latest DHS survey carried out in 2020 is used. The Ministry of Health (MOH) implemented this survey with the organization of the Division of National Malaria Programme (DNMP) in collaboration with the Kenya National Bureau of Statistics (KNBS). The duration of the fieldwork was from November 2020 to December 2020.

1.2 Study area

A hexagonal grid map at a resolution of 7 of Kenya will be created for the study case. Each hexagon represents 5.161293360 km2. These cells will be used as the base for the analysis. The region of interest for the project in Kenya which is located in the east of Africa as shown in the map below.

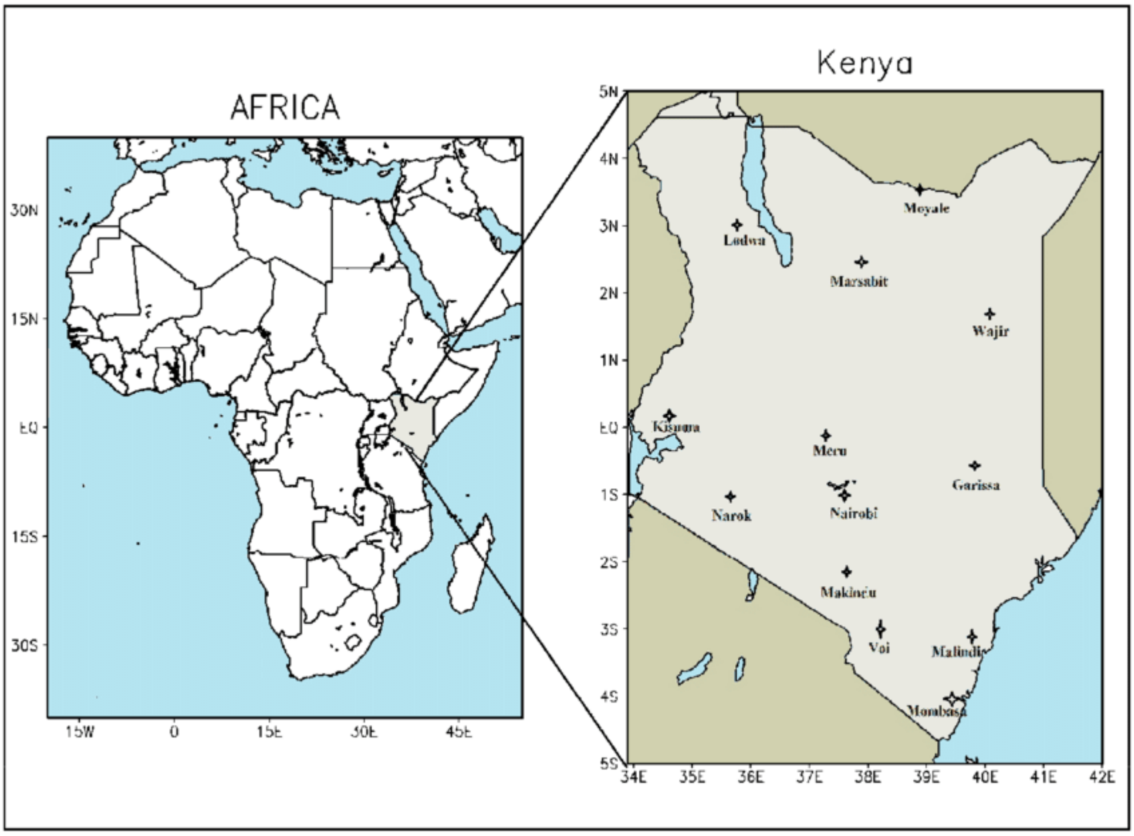


Figure 1: Study Area

1.3 Problem statement

Global poverty (World Bank, 2018) is the number of people who live on less than \$2.15 a day. One out of every ten people on the planet, which is more than 736 million people, currently lives below the poverty threshold. Children comprise one-third of the global population, they represent half of those struggling to survive on less than \$2.15 a day (Poverty Overview, 2020). However, this statement is not ideal for assessing poverty amongst children who are faced with many challenges as it focuses only on monetary poverty disintegrated by the international poverty lines of PPP. UNICEF and Save The Children have proposed a new definition of child poverty based on six factors which are nutrition, healthcare, education, water, and sanitation (nutrition, for every child, n.d.). This crucial demographic has been previously ignored while estimating global poverty. In this study, I will try to estimate and predict multidimensional child poverty in Kenya. This model will then be replicable in other countries that have DHS datasets. These predictions will be used to create geospatial visualizations in the form of poverty maps that will help county governments and organizations in their policies and local planning.

1.4 Objectives

1.4.1 Main objective

The main aim of this project is to estimate and predict multidimensional child poverty in Kenya from demographic health survey data.

1.4.2 Specific objectives

I. Estimate households that are on the extreme side of basic necessities and poverty index.

II. Create geospatial visualizations in the form of child poverty maps from the estimations.

III. Predict multidimensional child poverty in Kenya using machine learning models.

IV. Create geospatial visualizations in the form of child poverty maps from the predictions.

V. Reproduce the scripts to a python package that can be used by other countries.

1.5 Conceptual framework

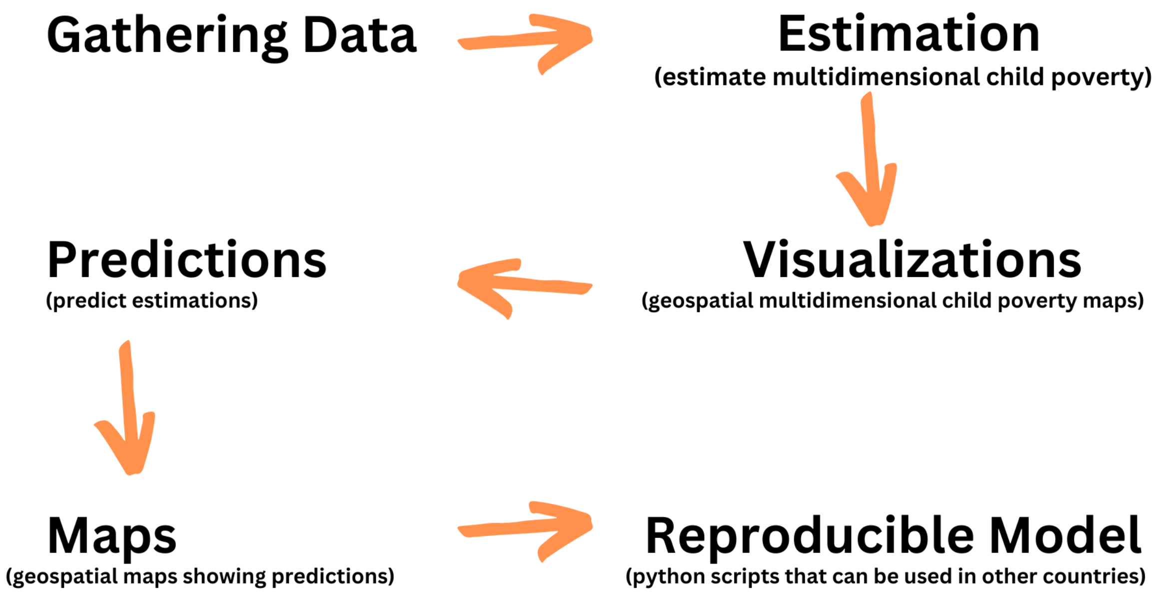


Figure 2: Conceptual framework

1.6 Scope of study

To achieve this study, ground truth data has to be available for the country to be estimated. In the study, the region of interest in Kenya. For the ground truth data estimations are created of four main factors that were collected in Kenya which are housing, sanitation, education, and water. A household sample size of 7952 was interviewed and their responses were recorded. Questions on sanitation in the household characteristics category played a very important role to decipher the sanitation of people living with children. The same also applied to all the other five factors. After cleaning the dataset and focusing on the target variables and dropping the null columns, 283 rows of households within a 5 km radius are created with the estimates and predictions. In this study, the latest datasets which are from the year 2020 are used. To support the prediction model, a lot of secondary data is used which include satellite images, critical infrastructure, connectivity datasets, economic variables of the target country, and demography.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

Child poverty is a problem that needs to be addressed on a global scale, not only because of the alarmingly large number of children who are affected (Alkire, 35–36; World Bank, 2020) but also because of the detrimental effects it has on both present-day and future human flourishing. There are various global estimates of child poverty, and these differ based on the definition of poverty. For example, child poverty can be measured based on income/consumption poverty as nationally defined (monetary poverty) and disaggregated by the child population or based on the international poverty lines of PPP $1.90/$3.20 or $5.50 disaggregated by the child population. They can also be based on the actual deprivations households face (multidimensional poverty), and the deprivations children face directly (multidimensional child poverty).

(Sen, 1985) argued that poverty should be viewed in terms of a lack of fundamental requirements or capacities. This indicates that because poverty is a multidimensional phenomenon, it should be assessed using a variety of well-being indices. The child poverty study (child poverty in Kenya National Bureau of Statistics (KnBS) and United Nations Children's Fund (UNICEF) a multidimensional approach, n.d.) indicate that overall child poverty (prevalence of deprivations) is 45 percent of all children. This translates to 9.5 million children in Kenya who are severely deprived of at least 3 or more basic needs for their well-being. This national mean conceals huge variations between the urban (19%) and rural (56%) areas, implying that 1 in 5 children and 6 in 10 children respectively in urban and rural areas are deprived in 3 or more dimensions.

2.2 Definition of child poverty

Existing literature and papers have focused on these definitions of child poverty, in this review, each relevant piece of literature in the field is covered, starting with multidimensional poverty. (Yekaterina Chzhen, Lucia Ferrone, 2016) implement UNICEF's rights-based multiple overlapping deprivation analysis methodology to Bosnia in their paper titled "Multidimensional Child Deprivation and Poverty

Measurement: Case Study of Bosnia and Herzegovina." Using distinctions between rural and urban locations, the research analyzes the prevalence and severity of multidimensional child deprivation and evaluates its relationship to household-based financial poverty. Using the child rights framework, seven areas of deprivation have been identified: housing, food, clothing, educational resources, leisure, and social participation. The majority of school-age children in Bosnia and Herzegovina experience deprivation in one or more of the seven categories, and one in four experience deprivation in three or more of the seven dimensions. Children who live in consumption-poor households are more likely to experience deprivation across all examined variables and a greater number of dimensions simultaneously. Yet, there is a moderate overlap between poverty and multidimensional deprivation, indicating that child poverty cannot be eliminated by just raising household spending.

2.3 Estimating multidimensional poverty index

(Unpacking Deprivation Bundles to Reduce Multidimensional Poverty Global Multidimensional Poverty Index 2022 OPHI Oxford Poverty & Human Development Initiative, n.d.) paper that primarily focuses on the global multidimensional poverty index of all population demographics. By addressing its various components, it offers helpful insights on how to combat multidimensional poverty, which is referred to throughout as "poor." In order to estimate the values of the Multidimensional Poverty Index (MPI) for the world in 2022, household surveys were conducted in 111 different nations, including 6.1 billion people. Common deprivation profiles and bundles — combinations of deprivations experienced by 1.2 billion poor people, or roughly twice as many as live in monetary poverty — are made visible for the first time. The multidimensional poverty index facilitates cross-country analysis and presents long-term patterns, which provides significant insights into this paper. The MPI values of 72 of the 81 countries having trend data considerably decreased across at least one of the examined time periods. Of these 72 nations, 68 significantly decreased the deprivations of the poor in five or more categories throughout that time, and 46 significantly decreased deprivations in eight or more measures.

2.4 Global multidimensional poverty index

The global MPI begins by constructing a deprivation profile for each household and person in it that monitors deprivations in 10 indicators spanning health, education, and standard of living. For example, a household and all people living in it are deprived if any child is stunted or any child or adult for whom data are available is underweight; if at least one child died in the past five years; if any school-aged child is not attending school up to the age at which he or she would complete class 8 or no household member has completed six years of schooling; or if the household lacks access to electricity, an improved source of drinking water within a 30-minute walk round trip, an improved sanitation facility that is not shared, non-solid cooking fuel, durable housing materials, and basic assets such as a radio, animal cart, phone, television or bicycle. A person's deprivation score is the sum of the weighted deprivations she or he experiences.

All indicators are equally weighted within each dimension, so the health and education indicators are weighted 1/6 each, and the standard of living indicators are weighted 1/18 each. The global MPI identifies people as multidimensionally poor if their deprivation score is 1/3 or higher. MPI values are the product of the incidence of poverty (proportion of people who live in multidimensional poverty) and the intensity of poverty (average deprivation score among multidimensionally poor people). The MPI is therefore sensitive to changes in both components. The MPI ranges from 0 to 1, and higher values imply higher poverty. From this study (OPHI, 2019), child poverty is almost ignored as the report focuses on all demographics. This study (OPHI, 2019) also does not factor in the geographical distribution of global poverty by administration boundaries in each country.

2.5 Monetary child poverty

Existing literature has previously focused on monetary child poverty which includes children living in poverty based on national poverty lines, children living below \$1.90, \$3.20, and \$5.50 PPP on the international poverty line. The global estimate of children in monetary poverty (Ani Rudra & Solrun & Jose Cuesta & David Newhouse & David Stewart, 2019) paper has come out with findings that suggest that children were still disproportionately more likely to be in households living under \$1.90 PPP per day in 2017 compared to adults (17.5% vs. 7.9%). The first-ever global estimates of children living in extreme poverty, as determined by the international poverty line of $1.90 PPP per day, were published in 2016 by the World Bank and UNICEF. The study's findings show that acute child poverty has a disproportionately negative impact on youngsters. Compared to 7.9 percent of adults, kids were twice as likely (17.5 percent) to reside in homes that make less than $1.90 PPP per day in 2017.

The geographical distribution of children living in poverty is striking, with Sub-Saharan Africa having the highest rates of children living in extreme poverty at just under 45.8 percent and the largest proportion of the world's extremely poor children at 65.8 percent, it became clear during this review. Sub-Saharan Africa is now home to two out of every three children who live in extreme poverty worldwide. The technique employed to measure poverty, which is based on income or consumption per capita, contributes to higher rates of poverty among children than adults. As it connects to child poverty, the project uses income consumption per capita data to train the machine learning model. This report (Silwal et al., n.d.) focused on poverty in 175 countries worldwide while this report mainly focuses on Kenya alone. The geographical displacement and distribution of children in all counties in Kenya can also be seen. The report also finds that child poverty is more prevalent in countries prone to conflict. About 41.6 percent of children who live in fragile and conflict-affected countries affected by conflict and fragility live in extremely poor households, compared to 14.8 percent of children in other countries Because these estimates pertain to 2017, they do not factor in covid-19 pandemic and the global inflation that followed later on. This current project factors in this pandemic from the recent ground survey conducted in 2020.

2.6 Estimating child poverty rates

A method for estimating child poverty rates, projections for the short-term, and the relationship between child poverty and child care subsidy receipt at the county level (Dorabawila, V., DuMont, K., & Mitchell-Herzfeld, S., 2012) focused mainly on coming up with a method to estimate annual poverty rates at the county level. The paper developed a method conceptually similar to the regression-based modeling which was applied to estimate and project the proportion of children less than 12 years of age whose income is below the 200% poverty threshold, by county in New York State from 2008 to 2010. This paper then contributes to the literature on county-level relationships in New York between poverty rates and income-based benefits, particularly on childcare subsidies. Specifically, the association between the proportion of children receiving child care subsidies and child poverty rates varies by metropolitan, micropolitan, and non-core-based designation of counties. The relationship between children who were receiving subsidies and the elimination of children in poverty below 12 years. This paper (Dorabawila et al., 2012) mainly focused on New York which is an urban area while the paper focuses more on the whole of Kenya both rural and urban areas. In Kenya, a child is a person who is below 18 years of age, the project focuses on this demographic of people.

2.7 Economic framework for measuring poverty

The political economy of multidimensional child poverty measurement: a comparative analysis of Mexico and Uganda published by the Oxford Development Studies (Monica Pinilla-Roncancio, Amy E. Ritterbusch, Sharon Sanchez-Franco, Catalina González-Uribe, Sandra García-Jaramillo, 2021) study investigates the causes of this absence from a political economic standpoint. It creates a ground-breaking political economic framework for measuring poverty and posits that a nation will only generate and apply trustworthy multidimensional child poverty (MDCP) indicators if and only if three factors come together: consensus, capacity, and polity. With the help of two pertinent case studies—Mexico and Uganda—the concept is investigated. Only Mexico satisfies the other two conditions, however, both nations meet the capacity requirement needed to measure MDCP. Because it indicates the circumstances that must alter in various situations before the successful acceptance and usage of an MDCP measure increase, the suggested political economy paradigm is normatively significant.

2.8 Mapping child poverty rates from health surveys

(Akande, Sheerifdeen & Olaide, Akande. 2019) The goal of this study is to analyze child poverty in Minna using a multifaceted approach to identify potential solutions to the problem. The study uses a survey- and observation-based exploratory and descriptive research design. A sample size of 386 houses was determined at a 95% confidence level using the sample size calculation. The study region has 54,141 households across the 24 neighborhoods of Minna. Using a cluster and simple random sampling procedure, a multistage sampling method was used to select the respondents. The data were analyzed using descriptive statistics and a multi-dimensional poverty measurement metric after 321 completed questionnaires were received. According to the survey, there are 50.9% of children live in multidimensional poverty in Minna, compared to 28.7% who only live in unidimensional poverty. According to the study, the overall multidimensional poverty index was 0.29, and the level of poverty was 0.52. The analysis also demonstrates that while two neighborhoods have acute child poverty, thirteen other neighborhoods experience moderate child poverty. The study's findings suggest that, as can be seen at the neighborhood level, child poverty dynamics differ across spatial units. So, if child poverty is to be significantly decreased, attention must be given to the spatial imbalances that exist inside the city.

CHAPTER THREE: METHODOLOGY

3.1 Data collection

All data used in this project has already been collected and is open-sourced. 5 categories of data sources have been used in this project which are satellite images, geographical information, critical infrastructure, economic variables, and demography. The ground truth data has been collected by the Ministry of Health under DHS for the multidimensional child poverty factors which are water, sanitation, healthcare, housing, nutrition, and education. To get the required satellite image data, use the google earth engine to load and download the data on the area of interest which is the country Kenya.

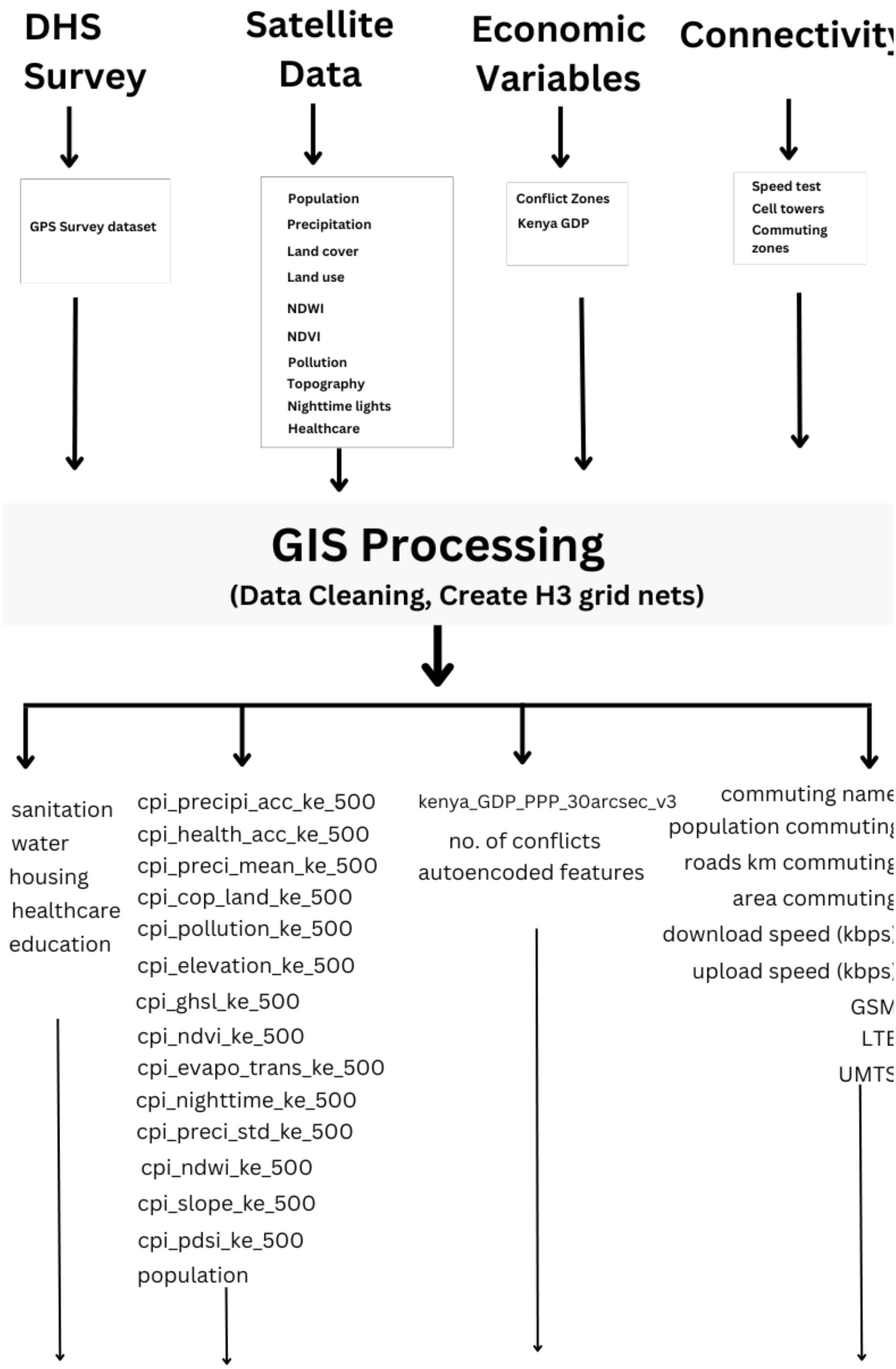
Table 1: Data, data sources, period, and spatial resolution.

Theme	Data Type	Resolution	Time	Source
Population	Raster/Imagery	100 m	2020 - 2021	World population hosted on Google Earth Engine
Precipitation	Raster/Imagery	11132 m	2012 - 2021	NASA hosted on Google Earth Engine
Land cover	Raster/Imagery	100 m	2012 - 2021	Copernicus hosted on Google Earth Engine
Land use	Raster/Imagery	38 m	2015	EC JRC hosted on Google Earth Engine
NDWI	Raster/Imagery	30 m	2012-2020	Landsat 7 hosted on Google Earth Engine
NDVI	Raster/Imagery	30 m	2012 -2020	Landsat 8 hosted on Google Earth Engine
Pollution	Raster/Imagery	1000 m	2012 - 2021	NASA LP DAAC at the USGS EROS Center

				hosted on Google Earth Engine
Topography	Raster/Imagery	90 m	2000	NASA/CGIAR hosted on Google Earth Engine
Nighttime lights	Raster/Imagery	463.83 m	2014 - 2021	Earth Observation Group, Payne Institute for Public Policy, Colorado School of Mines hosted on Google Earth Engine
Healthcare	Raster/Imagery	927.67 m	2019	Malaria Atlas Project hosted on Google Earth Engine
Conflict Zones	csv			Uppsala Conflict Data Program
GDP	Raster/Imagery	30 m	2015	Kummu, Matti; Taka, Maija; Guillaume, Joseph H. A. (2020), Data from: Gridded global datasets for Gross Domestic Product and Human Development Index over 1990-2015, Dryad, Dataset, https://doi.org/10.5061/dryad.dk1j0
Speed test	Raster/Imagery	610.8 m	2022	Ookla

Cell towers	csv		2022	OpenCelliD
Commuting zones	csv		2022	Data for Good at Meta

On the jupyter notebook linked, downloading satellite images using the package called geemap on google earth engine made it easy to access all necessary satellite images required. Using a shapefile of Kenya as the region of interest, clip the satellite tiles by filtering to the boundary of Kenya. While downloading these datasets, data quality is ensured by downloading the data from recognized institutions and organizations such as Google and respectable organizations.


DHS Survey
Satellite Data
Economic Variables
Connectivity
GPS Survey dataset
Population
Precipitation
Land cover
Land use
NDWI
NDVI
Pollution
Topography
Nighttime lights
Healthcare
Conflict Zones
Kenya GDP
Speed test
Cell towers
Commuting zones
GIS Processing
(Data Cleaning, Create H3 grid nets)
sanitation
water
housing
healthcare
education
cpi_precipi_acc_ke_500
cpi_health_acc_ke_500
cpi_preci_mean_ke_500
cpi_cop_land_ke_500
cpi_pollution_ke_500
cpi_elevation_ke_500
cpi_ghsl_ke_500
cpi_ndvi_ke_500
cpi_evapo_trans_ke_500
cpi_nighttime_ke_500
cpi_preci_std_ke_500
cpi_ndwi_ke_500
cpi_slope_ke_500
cpi_pdsi_ke_500
population
kenya_GDP_PPP_30arcsec_v3
no. of conflicts
autoencoded features
commuting name
population commuting
roads km commuting
area commuting
download speed (kbps)
upload speed (kbps)
GSM
LTE
UMTS

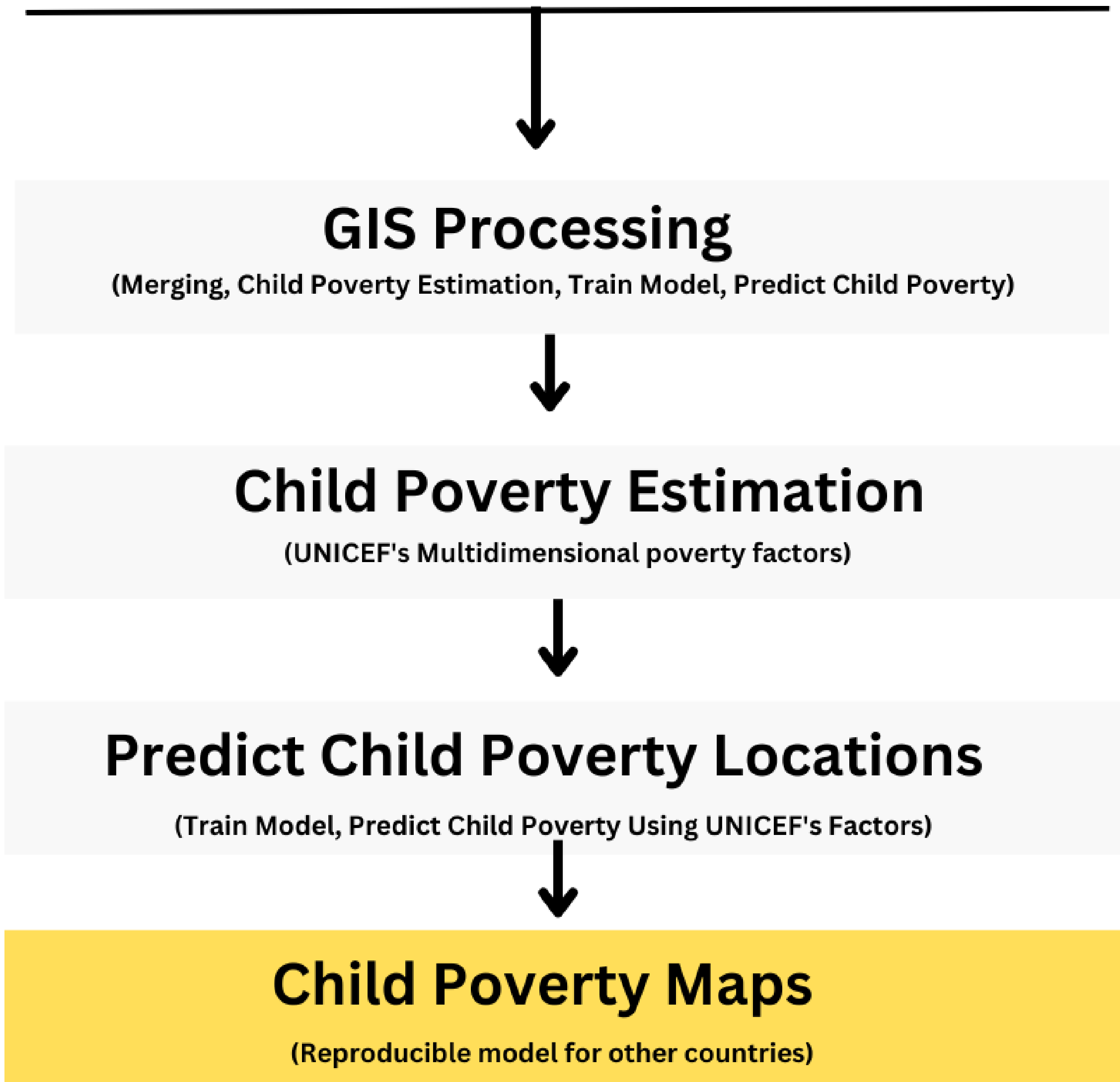


Figure 3: Methodology flow chart

3.2 Estimating multidimensional child poverty

To do the estimations, formulas provided by UNICEF help in estimating the multidimensional factors. To estimate the median years of education completed, medians are calculated from accumulated single-year percent distributions of completed years of schooling. The median is linearly interpolated between the years of completed schooling that occur before and after a cumulative 50 percent.

median = mi + (0.5 - p i) / (p i+1 - p i)

To estimate the sanitation factor, the calculation is of the percent distribution of de jure population by type of sanitation facility can be calculated using the household recode file and simply weighting the data by the sample weight (hv005/1000000) multiplied by the number of de jure household members (hv012). The estimate percent distribution of the de jure population by the source of drinking water can be calculated using the household recode file and simply weighting the data by the sample weight (hv005/1000000) multiplied by the number of de jure household members (hv012). Lastly to estimate housing characteristics the percent distribution of de jure population by housing characteristics can be calculated using the household recode file and simply weighting the data by the sample weight (hv005/1000000) multiplied by the number of de jure household members (hv012).

Households and populations with missing information are included as separate categories. The missing values are not included in the estimations and the predictions. These are the Missing data and “Don’t know” responses are grouped into a separate category in the percent distribution. Cleaning and merging the survey dataset with the GPS datasets requires joining the spatial attributes with the hexagonal resolution at 5.16 km2 to account for the survey radius. This is because of the anonymization and data protection of the households interviewed.

3.3 Survey dataset and estimations

DHS program survey dataset which is acting as the target variable has several categories from which the dataset was collected as. These categories are fieldwork questionnaire, household recode, individual recode, children’s recode and household member recode. In this study the main focus is on the household recode and children’s recode. DHS also collects the respondents locations data using a GPS, which is available on the GPS dataset. This dataset is highly restricted and should not be shared with the public until you are given permission. The dataset format is in a shapefile format.

The data format is in several formats which are Stata dataset (.dta) Flat ASCII data (.dat) SAS dataset (.sas7bdat) and SPSS dataset (.sav). The sas dataset is read by pandas, then filter it to the required columns, which are sanitation, water, housing, healthcare and education. These columns are represented by their unique IDs which are 'HV225', 'HV201', 'HV216', ‘V133’, and ‘B5’. Then drop unrequired columns and columns which have duplicates. Having the columns, the estimation of child poverty based on the calculations is done as seen in the methodology chapter. After having the target variables, merge it with the survey dataset, so as to get the spatial coordinates in the geometry column. The dataset’s crs is referenced to epsg: 4326.

This dataset is then spatial indexed using Uber’s H3 spatial index at a resolution of 7. After having a column with each unique hex code, group the dataframe by the hex codes. Later on get the hex centroids, where from the centroid extract the latitude and longitude of each hexagonal area. Lastly, to finalize cleaning the DHS dataset, get the dhs count of the hex codes and merge it with the survey dataset, in a new dataframe called survey threshold. This survey threshold will act as the target variables for the prediction model.

3.4 Population

After processing the target variables (survey), begin processing the image files. Download the dataset from google earth engine as shown in the notebooks. The population dataset is acquired from the world pop image and clipped to Kenya using the population band which is the estimated number of people residing in each grid cell. The population dataset is used as one of the variables to train the model, considering children's age is factored in.

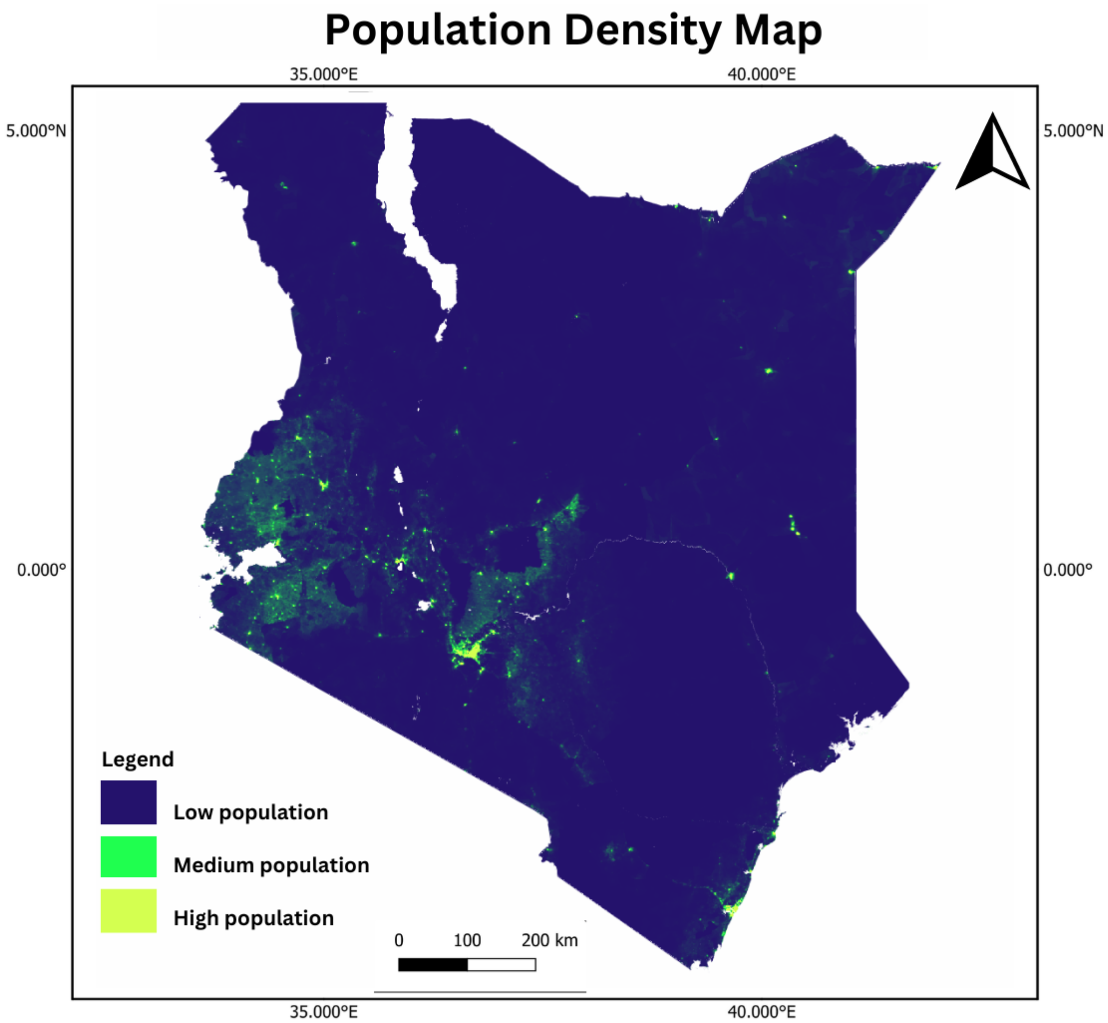


Figure 4: Kenya population density map

3.5 Precipitation

The precipitation dataset is then acquired from NASA using the precipitation band which is in mm/Hr. The filter date is for 10 years starting from 2012 to 2022 where the precipitation dataset is reduced using the ee reducer and acquire the mean of the images. Then filter by the boundaries of the country of interest. The precipitation mean and standard deviation is achieved by reducing the images and getting the mean and standard deviation respectively.

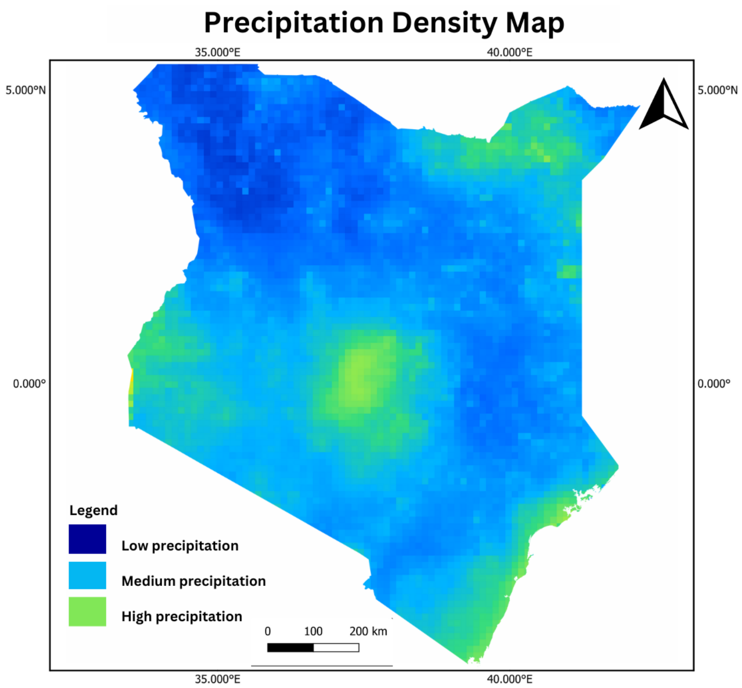


Figure 5: Kenya precipitation density map

To acquire climate precipitation, get the satellite images of TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces from the University of Idaho. This satellite image has an aet band for actual evapotranspiration, derived using a one-dimensional soil water balance model. The pr band is for precipitation accumulation. These two bands are added in the precipitation training dataset. Lastly the pdsi band is added which is responsible for Palmer Drought Severity Index. This acts as the drought dataset where the latest image is acquired and then resample it and clip it to the area of interest which is Kenya.

3.6 Land cover

For land cover, the Copernicus Global Land Cover Layers has the discrete classification band which represents the land cover classification. Areas with good vegetation cover, tend to have enough food for children which factors for nutrition. Nutrition is a major factor in multidimensional poverty. This dataset trains the machine learning model.

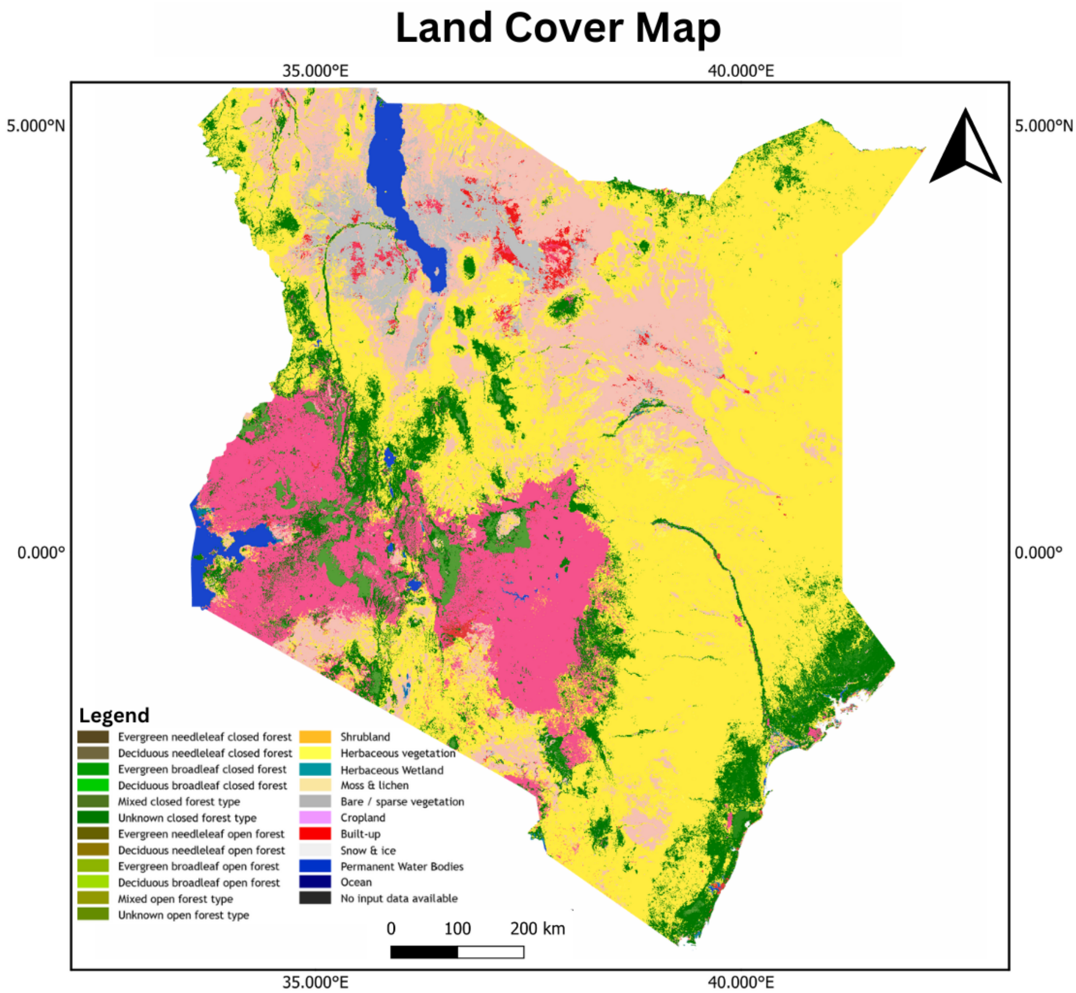


Figure 6: Kenya land cover map

3.7 Land use

In land use, the built and cnfd bands are added to the datasets which represent the multitemporal built-up presence and gaps-filled confidence grid on the built-up class aggregated for 2014. Land use especially built up areas help in training the machine learning model on the housing factor that greatly affects child poverty.

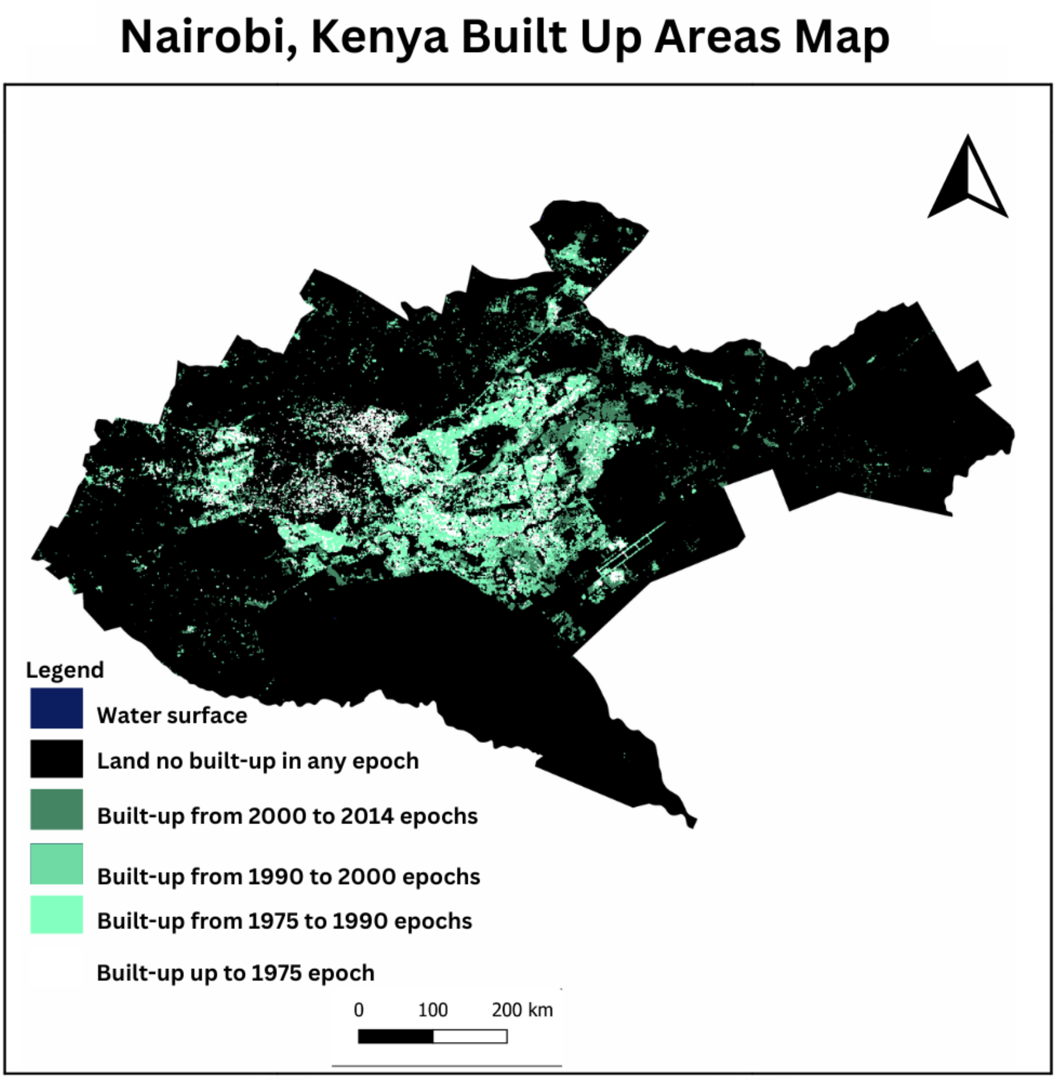


Figure 7: Nairobi county, land use built up areas map

3.8 NDWI

In NDWI, the landsat 8 Collection 1 Tier 1 Annual NDWI Composite is used and filtered to 2021 the last image available and the Normalized Difference Water Index band is selected. The satellite map below shows the NDWI of Kenya. The bright areas have high water index levels, while the dark blue areas have the lowest water index levels. Water index levels tend to influence the amount of food grown or fishing in certain areas. This factors in for nutrition and sanitation factors in multidimensional child poverty.

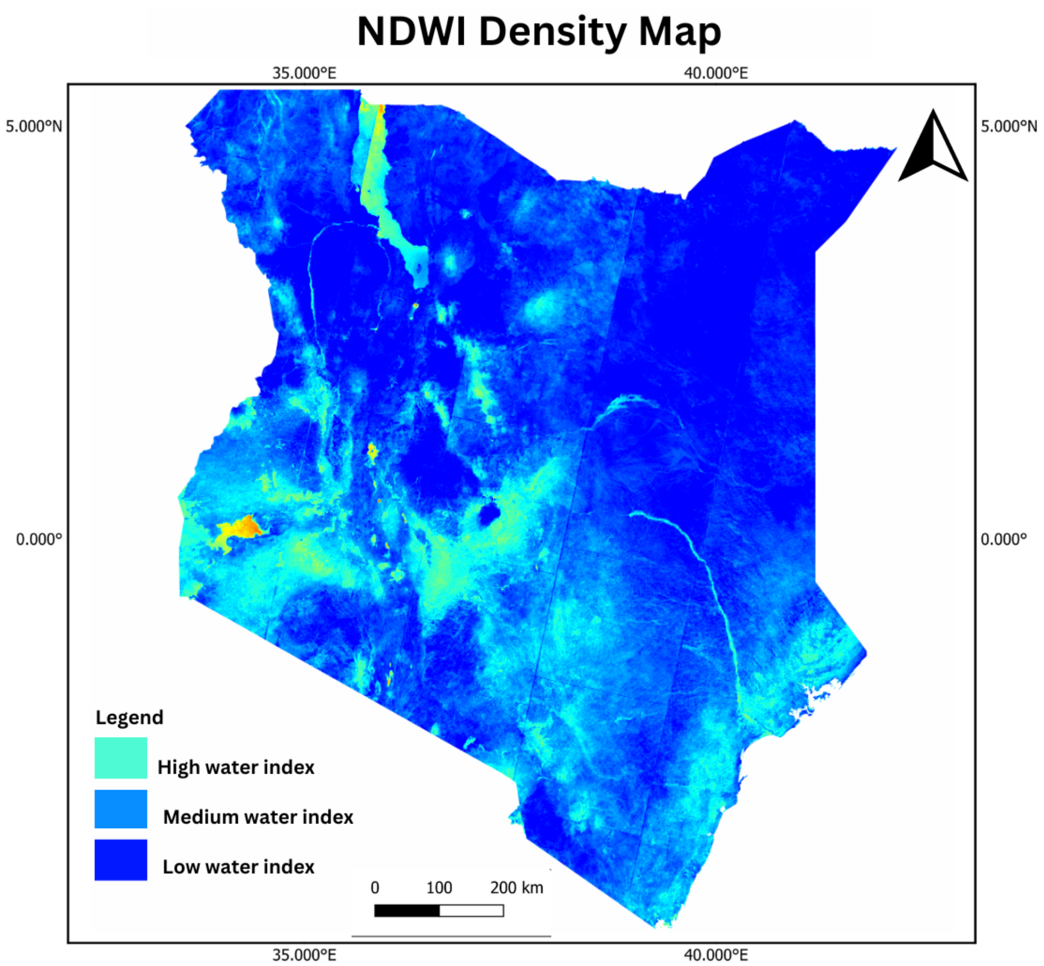


Figure 8: Kenya NDWI map

3.9 NDVI

In NDVI the landsat 7 Collection 1 Tier 1 32-Day NDVI Composite is used and selects the Normalized Difference Vegetation Index band which also the end date is in 2021. Vegetation plays a huge role in the factors of multidimensional poverty. This dataset will greatly contribute to the model.

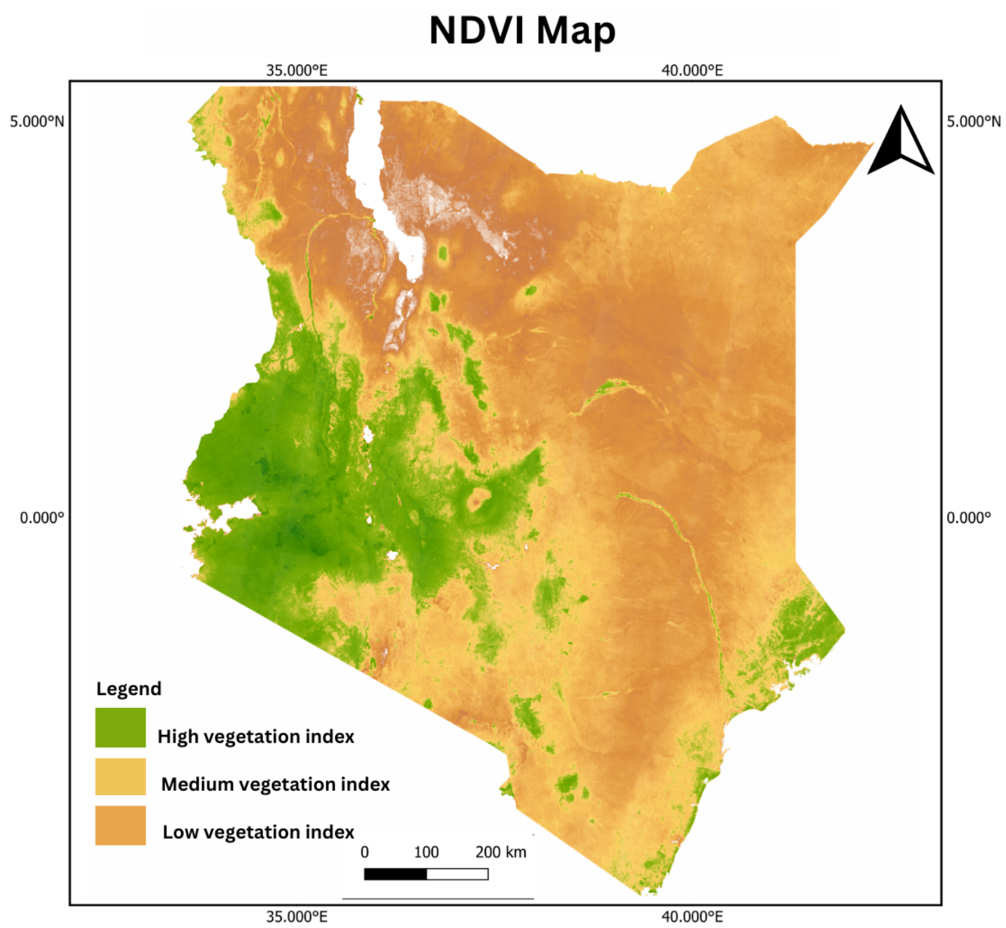


Figure 9: Kenya NDVI map

3.10 Pollution

For pollution the NASA satellite images of Terra & Aqua MAIAC Land Aerosol Optical Depth Daily 1km are used. The bands Optical_Depth_047 and Optical_Depth_055 which are the blue band (0.47 µm) aerosol optical depth over land and the green band (0.55 µm) aerosol optical depth over land. The pollution of the surface will be well represented by these two bands. Sanitation and health factors for children are greatly affected by pollution. Having this dataset will have a high correlation with the target variables.

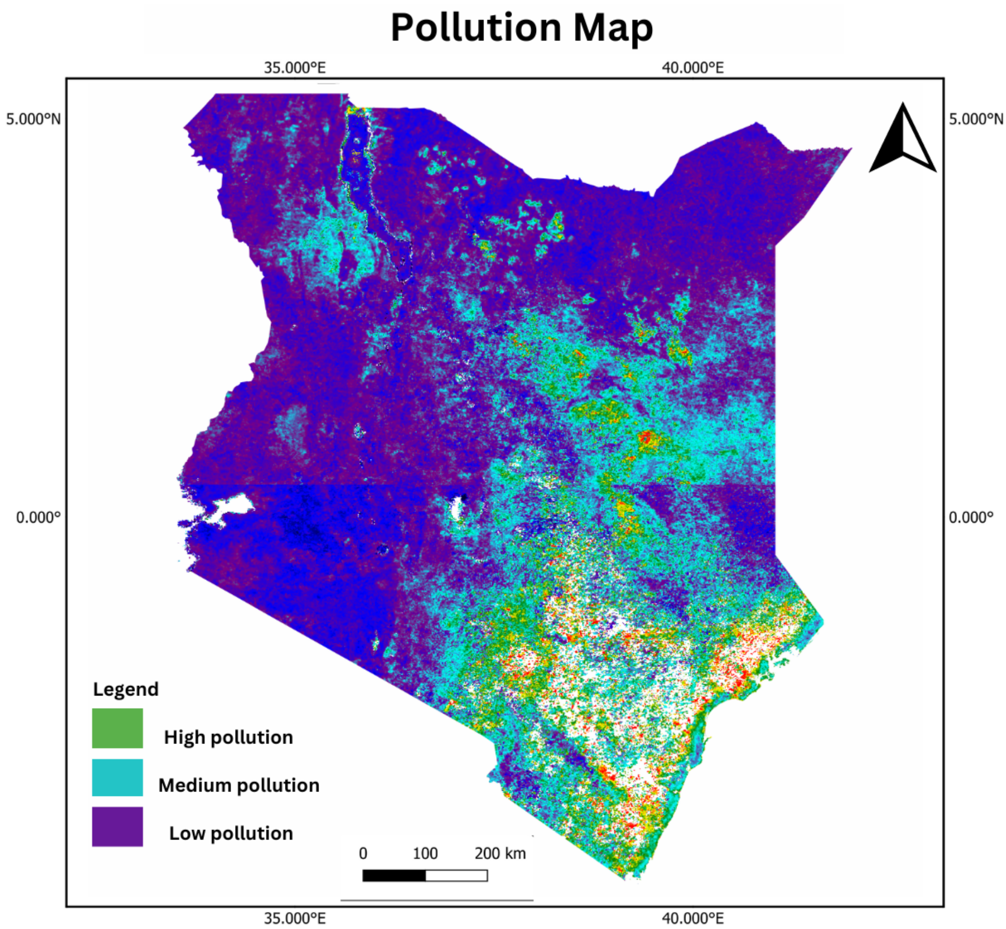


Figure 10: Kenya pollution density map

3.11 Topography

Topography has the dataset provided by NASA/CGIAR of the Shuttle Radar Topography Mission (SRTM) digital elevation which has the elevation band that represents the elevation at a 90 metre resolution. To acquire the slope the function of ee.Terrain.slope(elevation) on the elevation band is used. Slope will determine housing factor of multidimensional child poverty in the training and prediction.

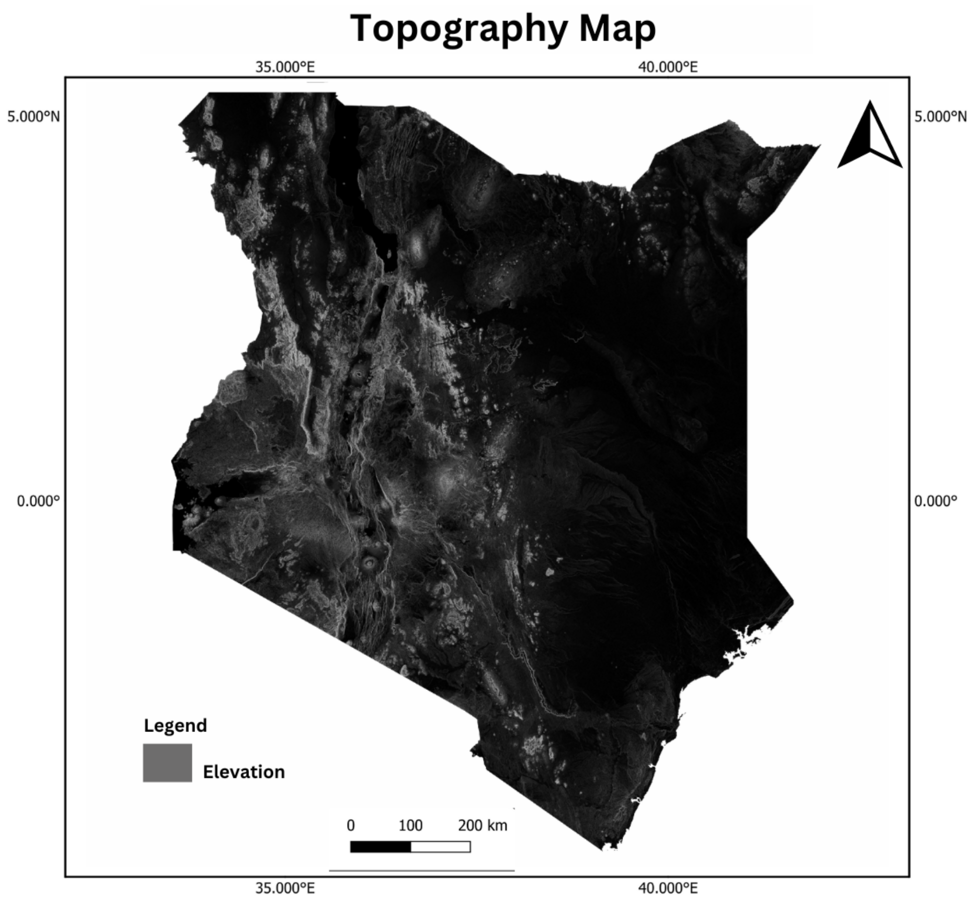


Figure 11: Kenya Topography map

3.12 Nighttime lights

Nighttime Lights provided by the Earth Observation Group, Payne Institute for Public Policy is the amount of monthly average radiance composite images using nighttime data from the Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band (DNB). These satellite images have two important bands which are the avg_rad (average DNB radiance values) and the cf_cvg band whose description is the cloud-free coverages; the total number of observations that went into each pixel. This band can be used to identify areas with low numbers of observations where the quality is reduced. Nighttime lights combined with population plus built up areas will contribute greatly to the model.

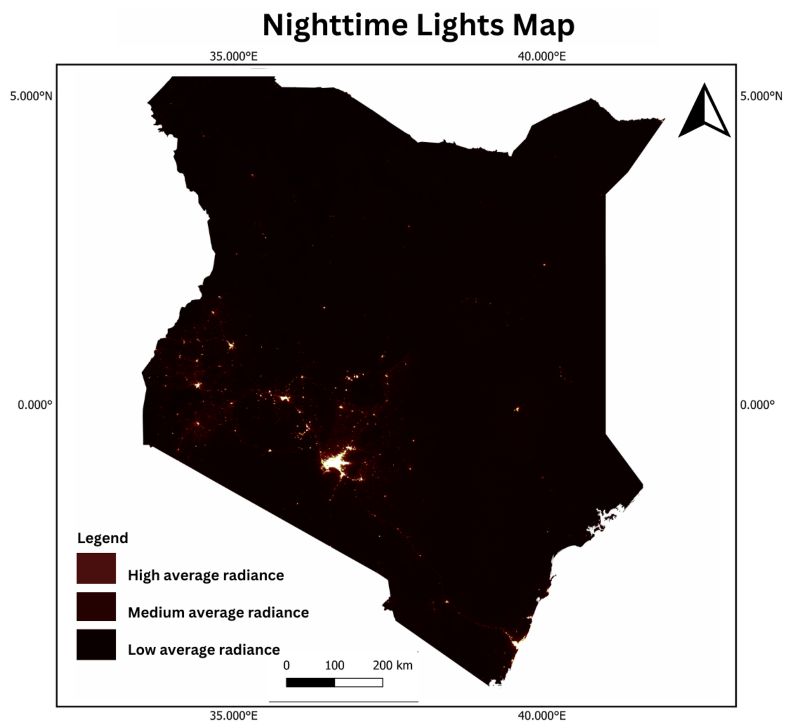


Figure 12: Nighttime lights map by average radiance

The global GDP is acquired as a satellite image where it is filtered by boundary and clipped to the region of interest, that is Kenya.

3.13 Healthcare

Lastly in the satellite images the Healthcare satellite image provided by the Malaria Atlas Project has two bands that are used which are accessibility which is the travel time to the nearest hospital or clinic. The accessibility_walking_only band is the travel time to the nearest hospital or clinic using non-motorized transport. The satellite map below shows access to healthcare via commuting. The North and eastern of Kenya have the highest time taken to commute to the nearest hospitals.

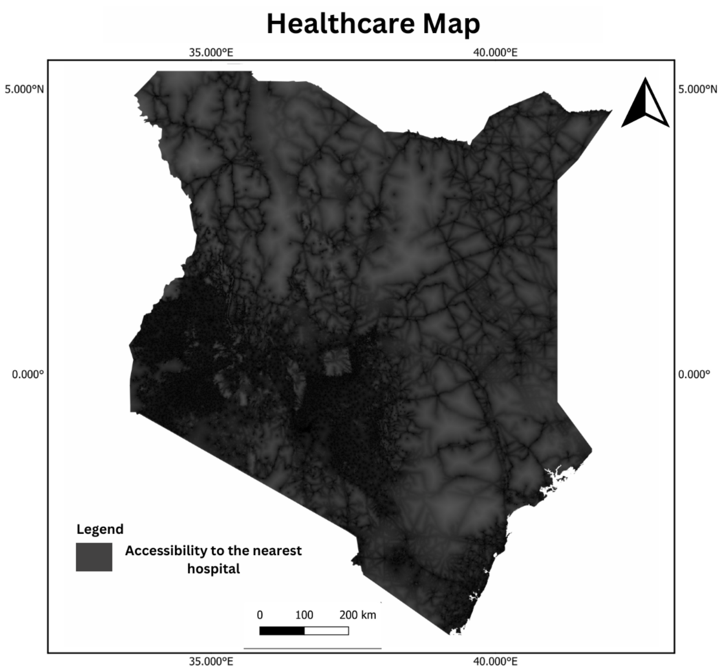


Figure 13: Kenya healthcare map

To process the satellite images, begin by converting the images which are tiff files in the format of a dataframe. Using a python package called rasterio and a python function that squeezes the data in each band at each pixel. The data for each band is applied to its own column. The (x,y) position of each data value is collected and converted to a latitude and longitude. From these new coordinates, hex codes are now created using Uber's H3 spatial index. After getting the hex codes for each image with its bands, aggregate all columns to form one dataframe. The function works very well with all the satellite images. For satellite images with more than two bands may take a while to process. To avoid this, a function called rast_to_agg_df is written which vectorizes the coordinates from the satellite image, transforms the crs and then applies the hex codes. The bands are then converted to a list and their data are read, and created as a new dataframe. Then group the data by the hex codes and drop all null values which may give us error during the GIS processing. A spatial join is then done, to join all satellite images by the hex code on the outer side. This gives us a full dataset to work with and predict child poverty based on the six indicators.

3.14 Internet speed test data

After cleaning and processing the satellite images, the connectivity dataset is cleaned and processed. To acquire speed test data, download it from ookla speedtest data which is an open data initiative. The dataset is in the form of tiles. To get Kenya internet speed test data, do a spatial join using the Kenya country boundary. Then download the dataset of the internet speed test. To process the dataset, get the centroid of each polygon in each row, and then get the coordinates. From the coordinates it is possible to get the hex codes of the speed test data using the hex code function. Group the data by the hex code and drop null and duplicate columns to have a cleaned and processed speed test data set.

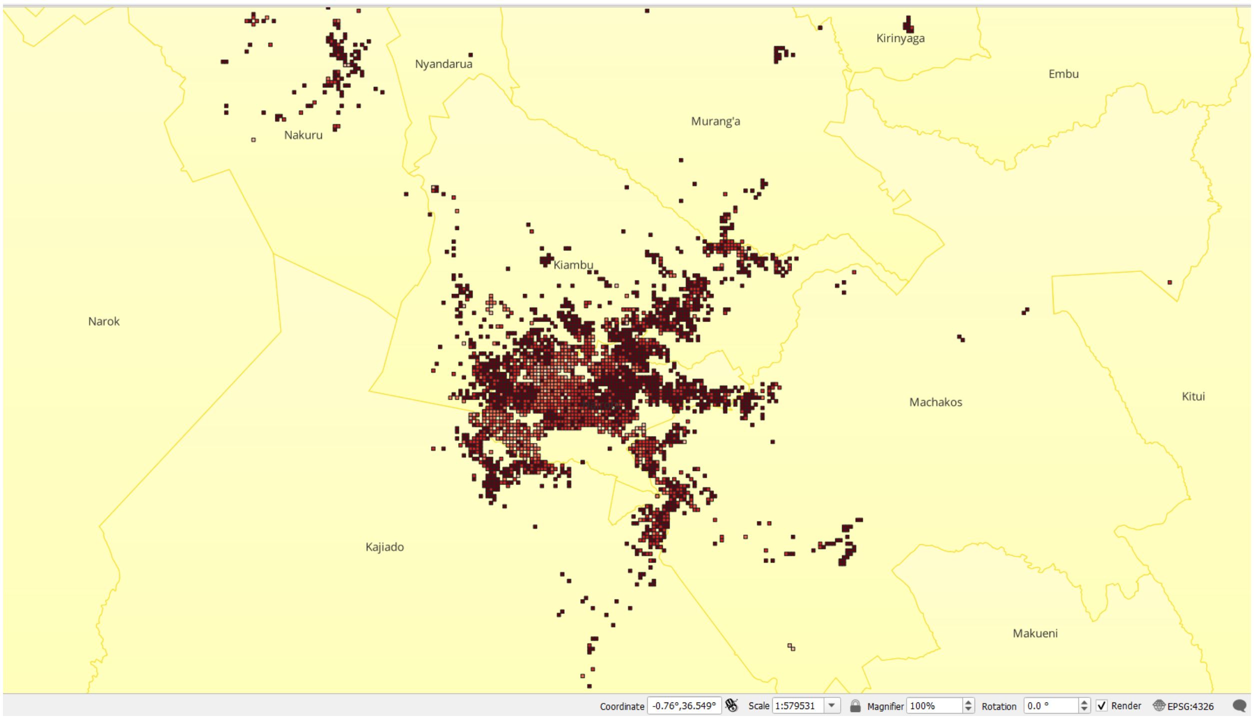


Figure 14: A zoomed in screenshot of internet speed test map of Kenya

3.15 Cell towers data

Cell towers of the target country can be acquired from opencellid.org. This organization is the world's largest collaborative community project that collects GPS positions of cell towers, used free of charge, for a multitude of commercial and private purposes. To get Kenya’s cell tower positions one has to have an api token that gives you access to download the data. The dataset is in a csv format which has several columns that are radio, mcc, net, area, cell, unit, lon, lat, range, samples, changeable, created at and updated at. For the research the radio, longitude and latitude columns are needed. From the coordinates generate the geometry. Then hex codes are created using the hex code function. Later on, group the dataframe by the hex codes and radio, to finish processing the cell towers dataset.

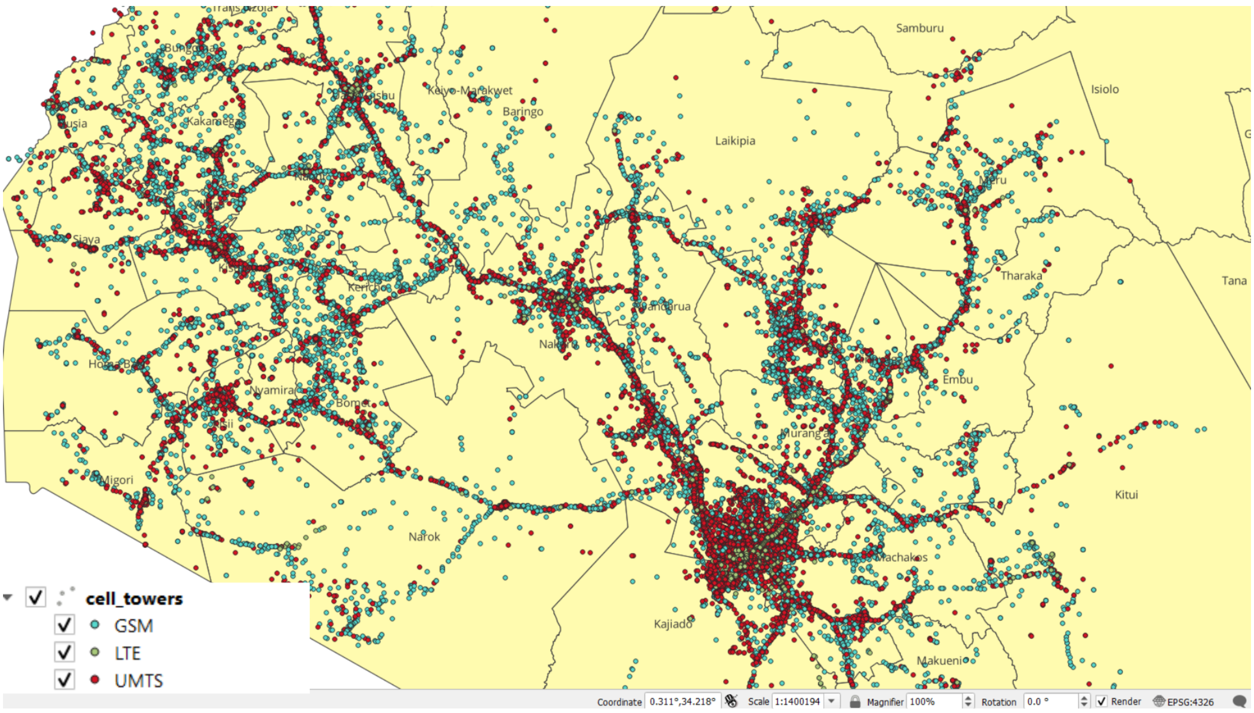


Figure 15: A zoomed in screenshot of a map showing cell towers location in Kenya

3.16 Commuting zones

Commuting zones provided by facebook, are geographic areas where people live and work and are useful for understanding local economies, as well as how they differ from traditional boundaries. To acquire a commuting zones dataset download it from data.humdata.org. Rename the geography column to be geometry and filter Kenya from the dataset. Hex codes are then created from the geometry column. The commuting zone has several important columns which are the win_population_commuting, win_roads_km_commuting, and area_commuting. This will be very crucial when training the dataset.

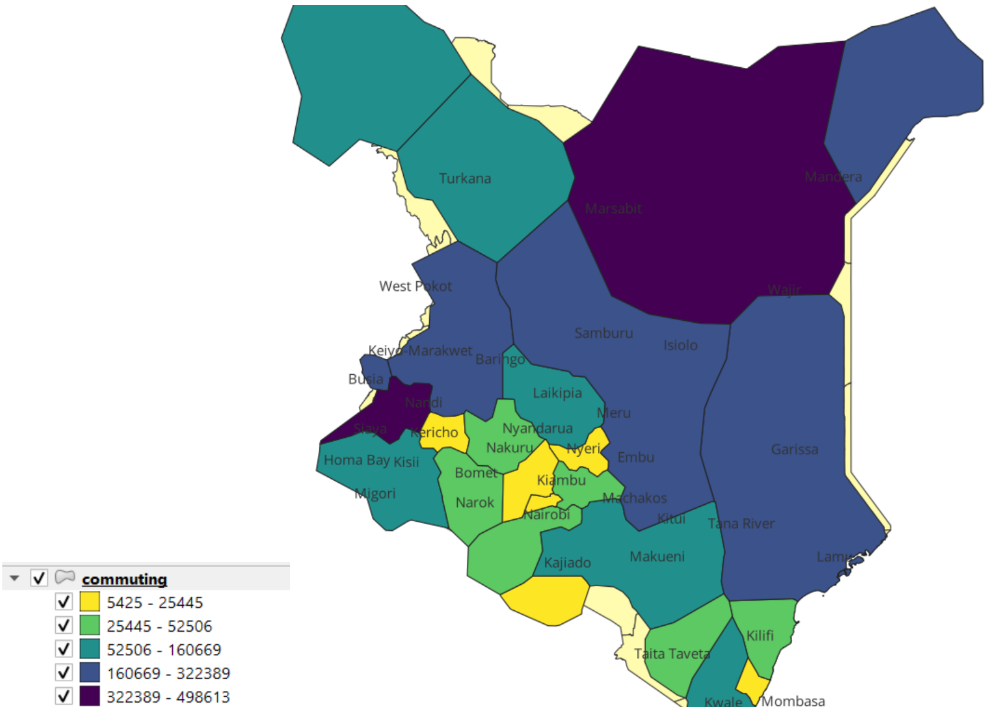


Figure 16: A screenshot of a map showing commuting distances of Kenya

3.17 Conflict zones

The conflict zones dataset acquired from UCDP covers individual events of organized violence (phenomena of lethal violence occurring at a given time and place). These events are sufficiently fine-grained to be geo-coded down to the level of individual villages, with temporal durations disaggregated to single, individual days. For areas with conflicts in Kenya, group the data by the hex code after deriving the hex code from latitude and longitudes. Then count the number of conflicts and group by that. After having all the datasets collected and processed, merge all the datasets to one dataframe using the hex codes. Also merge the satellite image data into this new dataframe.

3.18 Autoencode satellite images

Satellite images are known to have many features on them including noise and cloud cover. This noise may bring errors on the analyzed data, to extract more features and remove this noise autoencoders are used. An autoencoder is a type of artificial neural network used to learn data encodings in an unsupervised manner. The aim of an autoencoder is to learn a lower-dimensional representation (encoding) for a higher-dimensional data, typically for dimensionality reduction, by training the network to capture the most important parts of the input image. To create autoencoders for the satellite images, both encoders and decoders are made. To encode, keras from tensorflow is used with the adam optimizer algorithm as seen in the notebook.

```
input_dims = input_data.shape
model = tensorflow.keras.Sequential()
model.add(keras.Input(shape=input_dims[1:]))
model.add(Conv2D(32, (3, 3), activation="relu", padding="same", name="conv1"))
model.add(MaxPooling2D((2, 2), padding="same", name="mp1"))
model.add(Conv2D(16, (3, 3), activation="relu", padding="same", name="conv2"))
model.add(MaxPooling2D((2, 2), padding="same", name="mp2"))
model.add(Conv2D(8, (3, 3), activation="relu", padding="same", name="conv3"))
model.add(MaxPooling2D((2, 2), padding="same", name="mp3"))
model.add(Conv2D(8, (3, 3), activation="relu", padding="same", name="conv4"))
model.add(MaxPooling2D((1, 1), padding="same", name="Encoder_Output"))

model.add(Conv2D(8, (3, 3), activation="relu", padding="same", name="conv5"))
model.add(UpSampling2D((1, 1), name="us1"))
model.add(Conv2D(8, (3, 3), activation="relu", padding="same", name="conv6"))
model.add(UpSampling2D((2, 2), name="us2"))
model.add(Conv2D(16, (3, 3), activation="relu", padding="same", name="conv7"))
model.add(UpSampling2D((2, 2), name="us3"))
model.add(Conv2D(32, (3, 3), activation="relu", padding="same", name="conv8"))
model.add(UpSampling2D((2, 2), name="us4"))
model.add(
    Conv2D(
        input_dims[-1],
        (3, 3),
        activation=None,
        padding="same",
        name="Decoder_Output",
    )
)
model.compile(
    optimizer=Adam(learning_rate=learning_rate),
    loss=keras.losses.MeanSquaredError(),
)
model.fit(input_data, input_data, batch_size=batch_size, epochs=epochs)
```

Figure 17: convolution encoding model

Small images are extracted from the large Kenyan satellite image, using either the latitude, longitude or hex codes. Then convert the small images in the tiff format to a dataset that will be possible to encode. Processing the dataset to be trained by scaling it using the standard scaler and then reshaping it and transposing it and removing the nan columns. Having clean training data, the model can be trained. After encoding the model, decode the model to get the values of the images at their respective columns. This encoded image dataset is then merged with the survey threshold dataset and the hexes dataset to have the final dataset to feed to the machine learning model to predict child poverty in Kenya.

3.19 Prediction and training

To train the model to predict child poverty from dhs datasets, linear regression and random forest model were used. The final prepared dataset is split into two, the training dataset and testing dataset at a ratio of 80/20. Before splitting the dataset, null and duplicate columns are dropped. Other null columns are replaced by the KNNImputer, which uses the k-nearest neighbor to replace the missing values. 5 nearest neighbors were used to replace the missing values. The data is transformed via a pipeline with a column transformer where the numerical values go through the KNNImputer and a standard scaler. Categorical values are hot encoded where they are converted to numerical values in a binary ranging from 0 to 1. The transformed data is now ready to be trained on two regression algorithms to predict the factors that highly contribute to child poverty. The two regression algorithms used are the Linear regression algorithm and random forest regression.

The linear regression R^2 results below show that the housing target variable had a score of 0.638 while water and education had 0.664 and 0.785 respectively. A high r squared score means the model fits better. On checking the Root Mean Square Error (RMSE) of water and education variables, the model performs poorly on the training data. This leads to exploring the other regression model.

```
    rmse = np.sqrt(mean_squared_error(y_train[col], predictions))
    print(f"For target {col}:")
    print(f"R^2 = {r_squared}, Adj.R^2 = {adj_rsquared}, RMSE = {rmse}")
    scores[idx] = rmse
    models[col] = lin_reg

Loading widget...
For target housing:
R^2 = 0.6381435544461644, Adj.R^2 = -0.35251261616843466, RMSE = 0.29698476256808204
For target water:
R^2 = 0.6639359288923489, Adj.R^2 = -0.2561083313531878, RMSE = 5.69042220456803
For target education:
R^2 = 0.7848602394277783, Adj.R^2 = 0.1958710588448106, RMSE = 1.2612458641565316
```

Figure 18: Linear regression model results

Random forest model outperformed the linear regression model on the test dataset by having an R^2 score of 0.55 at a root mean square error of 1.89. The linear regression performed well on the training data but very poorly on the unseen test data. The housing prediction is the most accurate followed by water and housing. What may have contributed to these scores is the limited survey dataset after dropping the null columns, 289 rows remain from 100,000 + rows. Future implementation should be done on how to replace missing values of the target variables and add more training models such as the LGBMregressor.

```
        predictions = final_model.predict(X_test)
        rf_preds[col] = predictions
        r_squared = r2_score(Y_test[col], predictions)
        adj_rsquared = adjusted_rsquared(
            r_squared, X_test.shape[0], X_test.shape[1]
        )
        rmse = np.sqrt(mean_squared_error(Y_test[col], predictions))
        print(f"For target {col}:")
        print(f"R^2 = {r_squared}, Adj.R^2 = {adj_rsquared}, RMSE = {rmse}")
        rf_scores[idx] = rmse
        rf_models[col] = final_model
```

```
Loading widget...
 Best Parameters {'max_features': 8, 'n_estimators': 30}
 Best Estimator RandomForestRegressor(max_features=8, n_estimators=30, random_state=42)
For target housing:
R^2 = 0.013773723350546274, Adj.R^2 = 1.5110445251728988, RMSE = 0.38556646663764216
 Best Parameters {'max_features': 6, 'n_estimators': 30}
 Best Estimator RandomForestRegressor(max_features=6, n_estimators=30, random_state=42)
For target water:
R^2 = 0.20742258268227942, Adj.R^2 = 1.4106992071555462, RMSE = 8.496985573174905
 Best Parameters {'max_features': 6, 'n_estimators': 30}
 Best Estimator RandomForestRegressor(max_features=6, n_estimators=30, random_state=42)
For target education:
R^2 = 0.5575352217724105, Adj.R^2 = 1.2292772032633872, RMSE = 1.8953882550271957
```

Figure 19: Random forest regression model results

CHAPTER FOUR: RESULTS AND DISCUSSIONS

4.1 Education

The random forest regression then predicted UNICEF's data factors that contribute to multidimensional child poverty. For the education factor, that is the number of years each child in a household has spent in school. From the map below the North Eastern counties with Kajiado, Kwale will have the lowest number of children that are completing school to class 8.

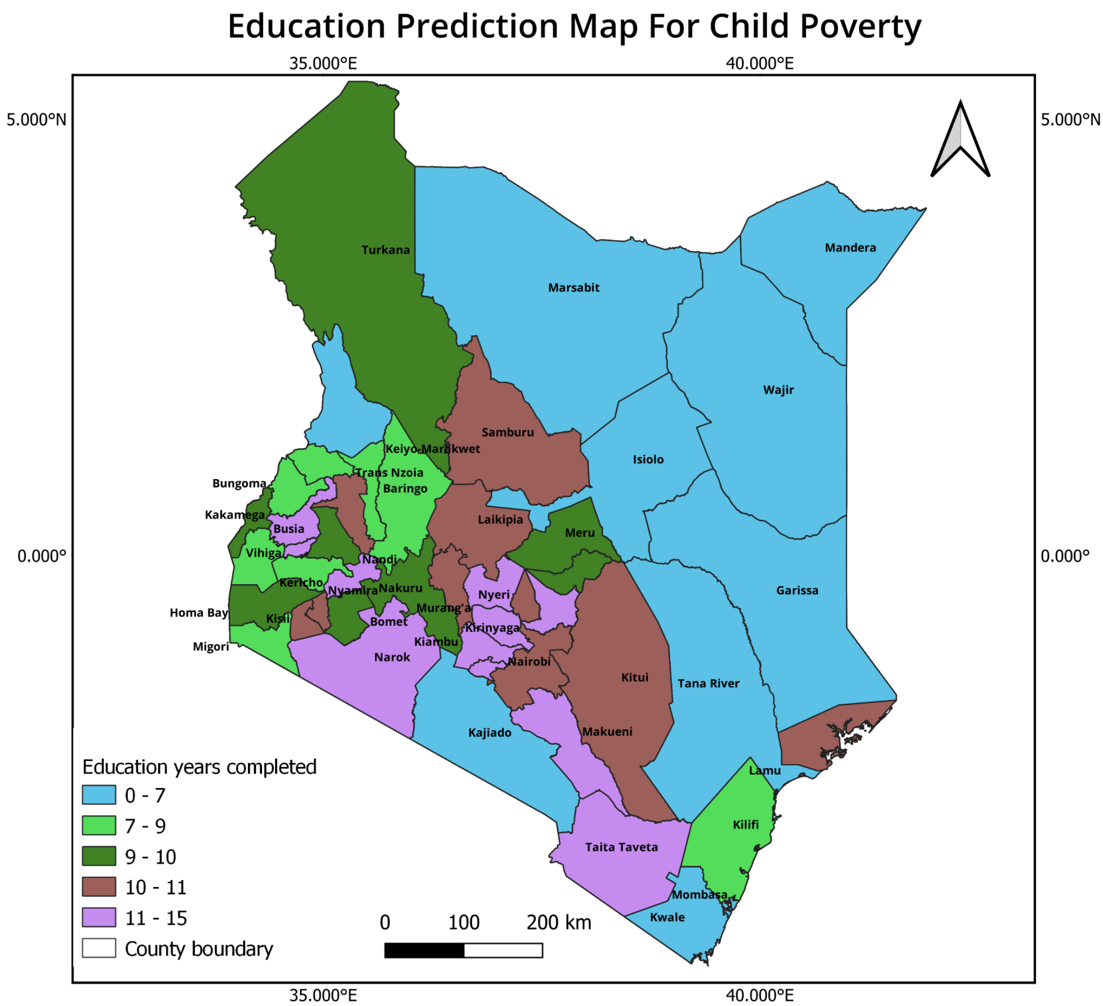


Figure 20: Education prediction map for child poverty

When the education prediction map is compared to the ground truth of the year 2020, the prediction map indicates that the number of children finishing school will increase. This is a positive sign that will contribute to eliminating child poverty.

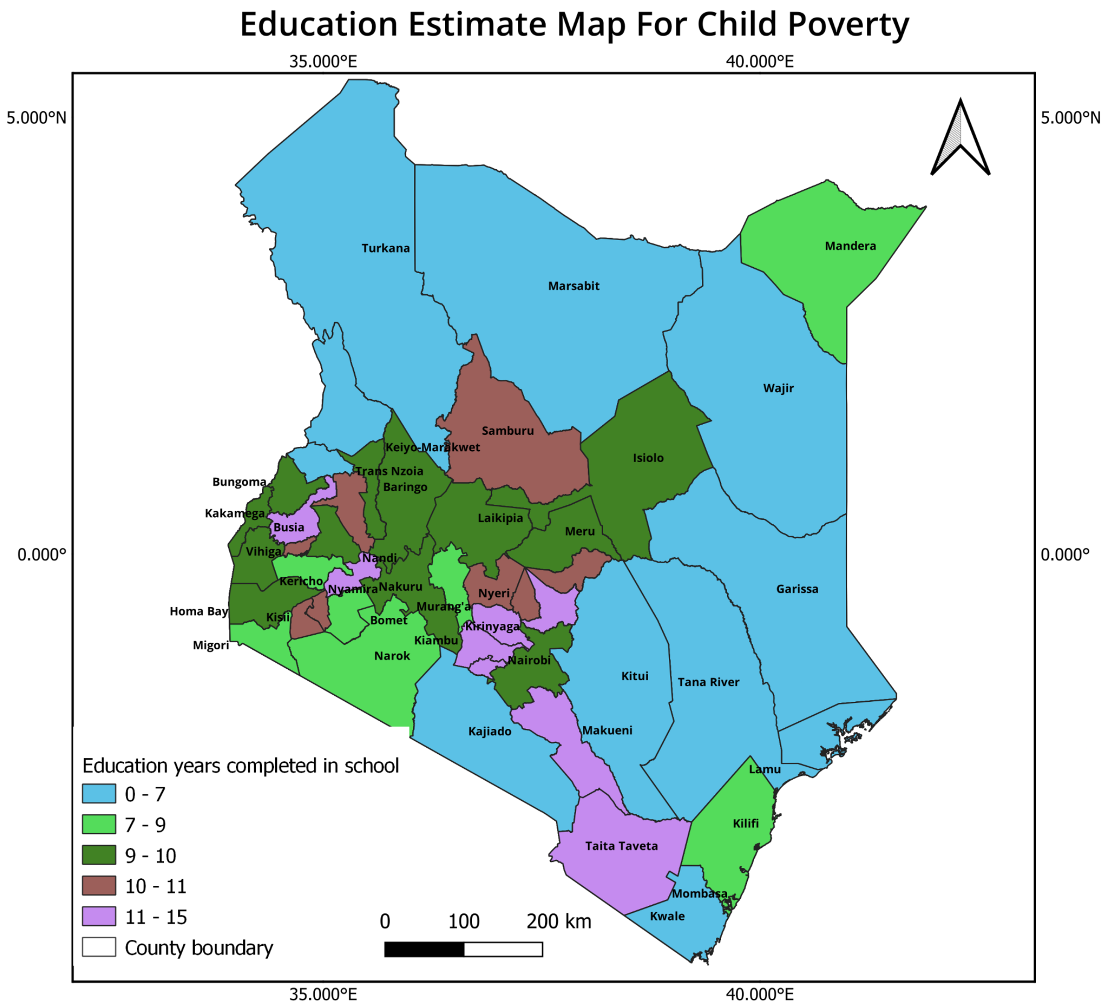


Figure 21: Education estimate map for child poverty

4.2 Water

The model had to predict the number of households by improved or unimproved drinking water source such that improved houses values were from 11:14,21,31,41,51 while unimproved drinking water source ranged from 32,42,43,61,62. The model predicts that most counties will have improved drinking water sources in the near future. Counties in yellow such as Isiolo, Laikipia, Garissa, Baringo have the lowest score of improved drinking water source. This means that children will face a greater challenge of accessing clean drinking water, which will highly contribute to child poverty.

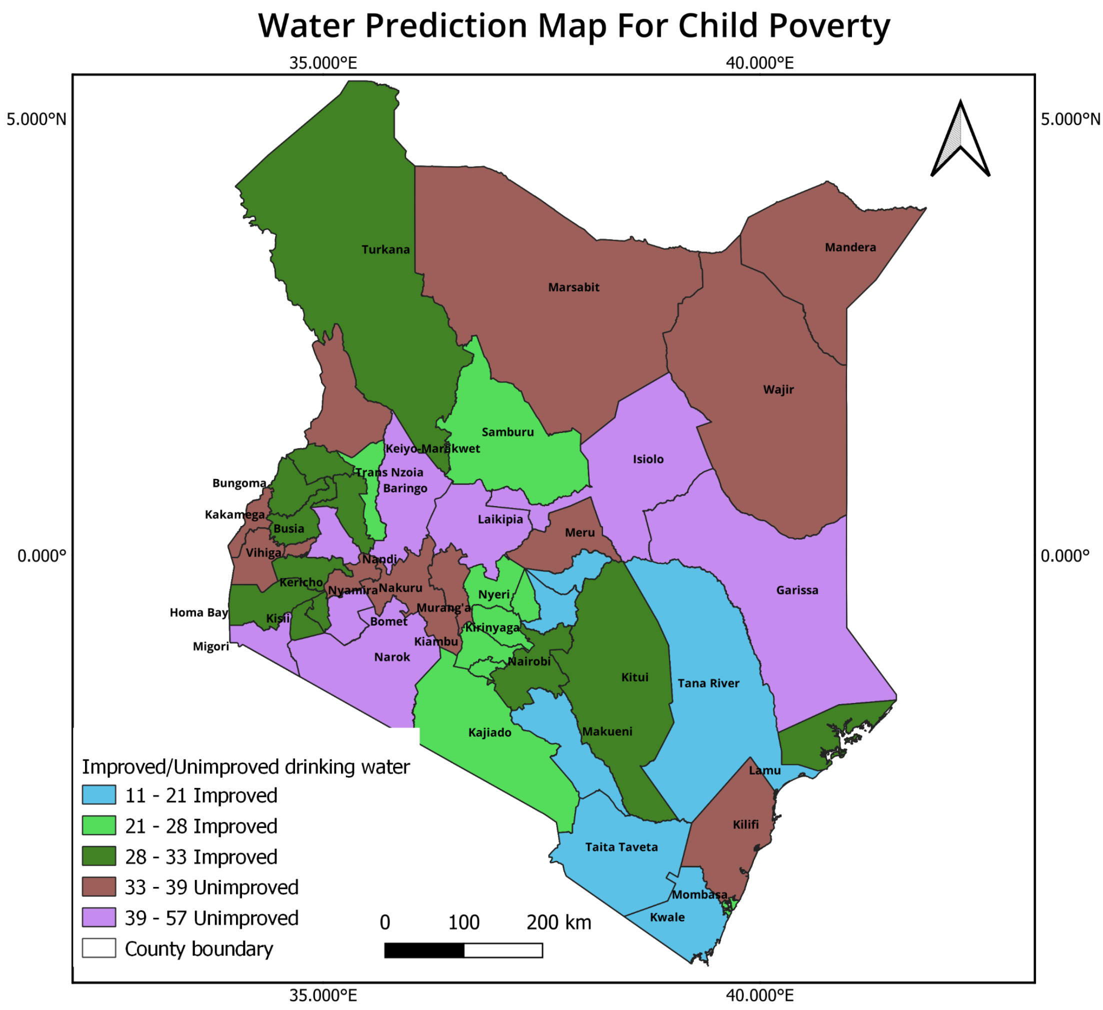


Figure 22: Water prediction map for child poverty

When compared to the actual ground truth on the maps below, counties such as Turkana and Marsabit will have improved drinking water sources if they follow the current trajectory they are on. This may be attributed to the high number of boreholes and wells being dug in the areas.

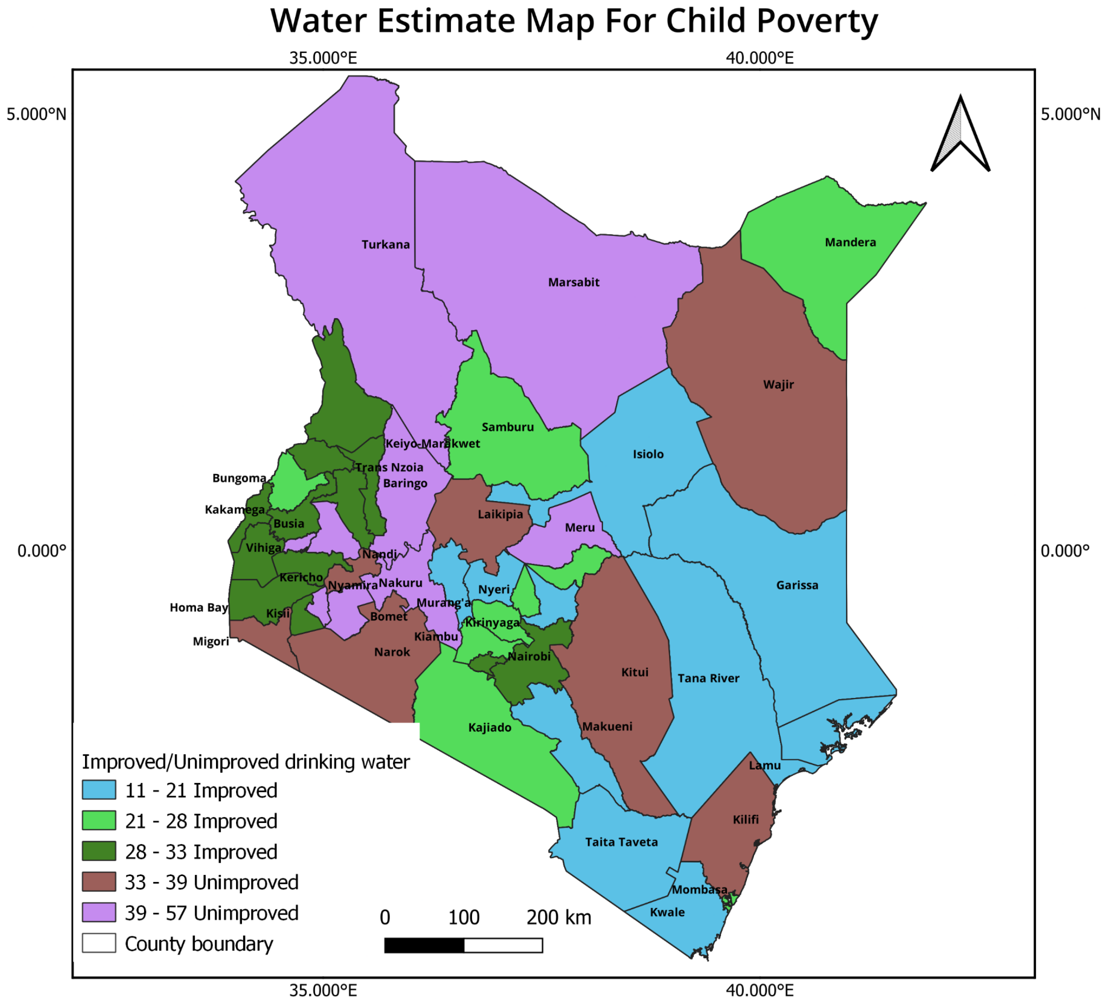


Figure 23: Water estimate map for child poverty

4.3 Housing

The last factor that the model predicted is housing. That is the number of bedrooms in each house that children are sleeping in. The model predicts that counties in the North of Kenya have the lowest number of bedrooms in each household. Most houses have less than 2 bedrooms for children to sleep. This indicates that the Northern counties have high child poverty when it comes to housing. Kwale, Kirinyaga, Kilifi and Vihiga have the highest number of households with a high number of bedrooms ranging at 2.4 bedrooms and above.

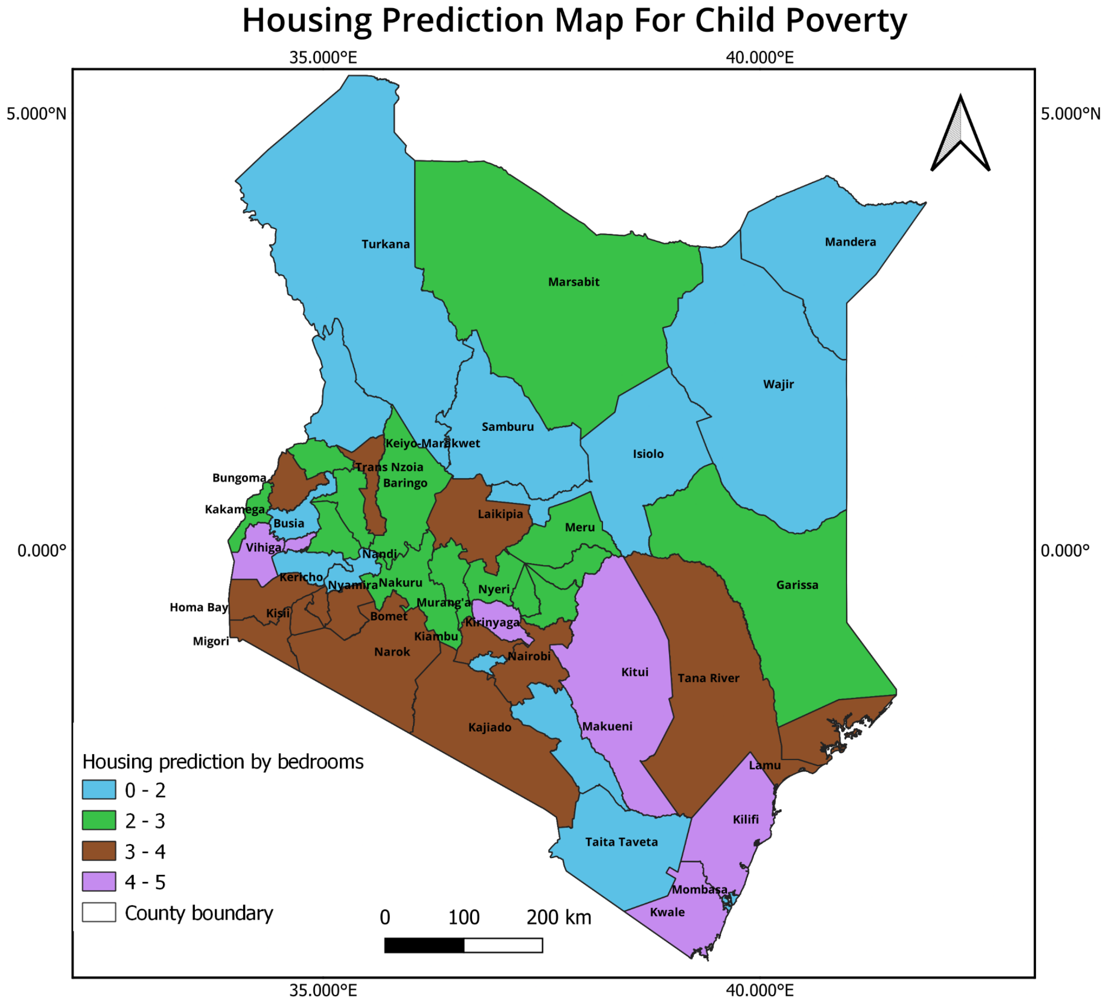


Figure 24: Housing prediction map for child poverty

When compared to the ground truth, the model is not far off from the actual survey. The actual survey still shows that most counties have under 2 bedrooms, with the worst affected being Wajir, Narok, Kericho. While the counties with the highest number of bedrooms remain in the model which include Kwale, Mombasa and Homabay.

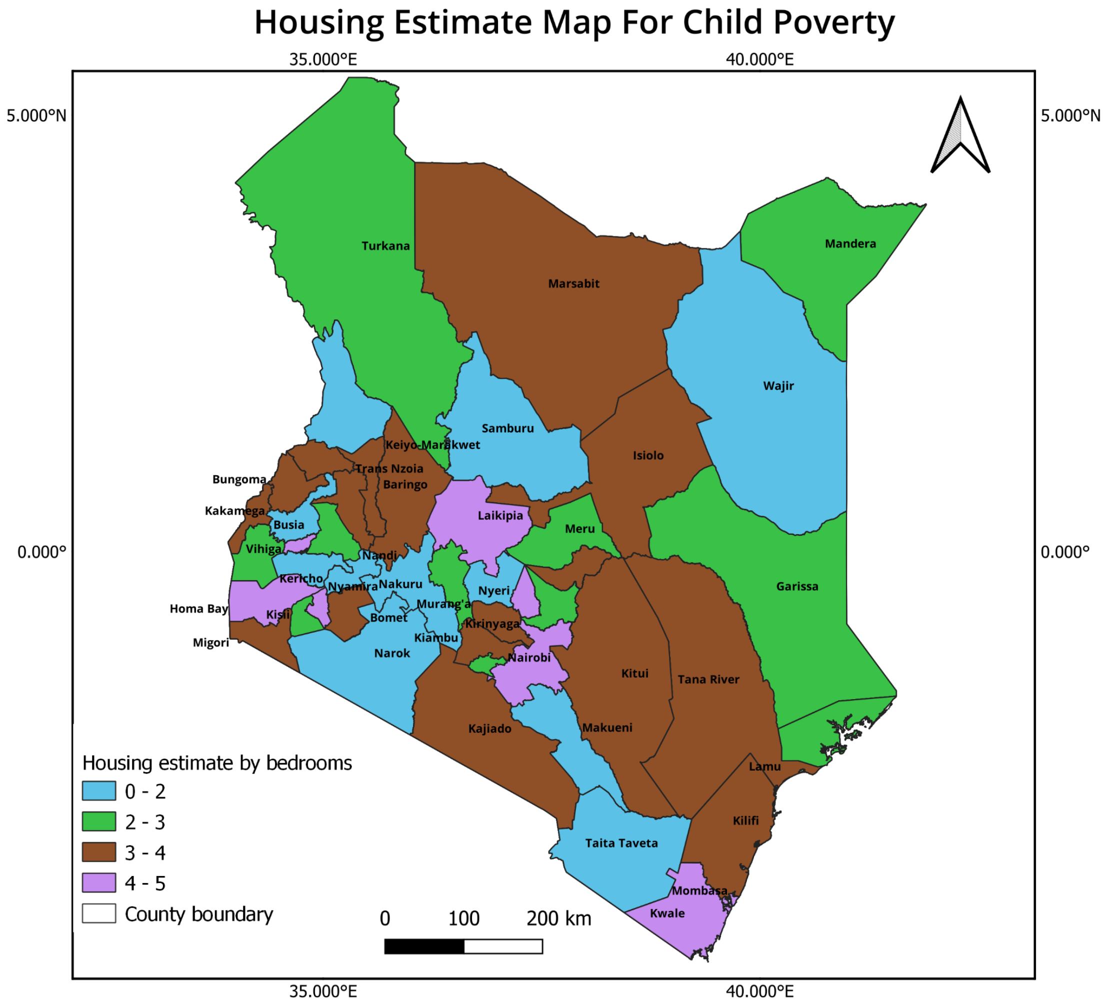


Figure 25: Housing estimate map for child poverty

4.4 Sanitation

Sanitation estimates houses with improved sanitation have a weight score of 1, houses with unimproved sanitation or open defecation have a weight score of less than 1. To estimate and calculate the percent distribution of de jure population by type of sanitation facility can be calculated using the HR file and simply weighting the data by the sample weight (hv005/1000000) multiplied by the number of de jure household members (hv012). From the estimate map below counties in the north and eastern of Kenya, Kajiado, Nyeri, Nyandarua, Muranga, and Nandi have unimproved sanitation. Nairobi county does not perform poorly on sanitation having better sanitation at a range of 0.294 to 0.0.389.

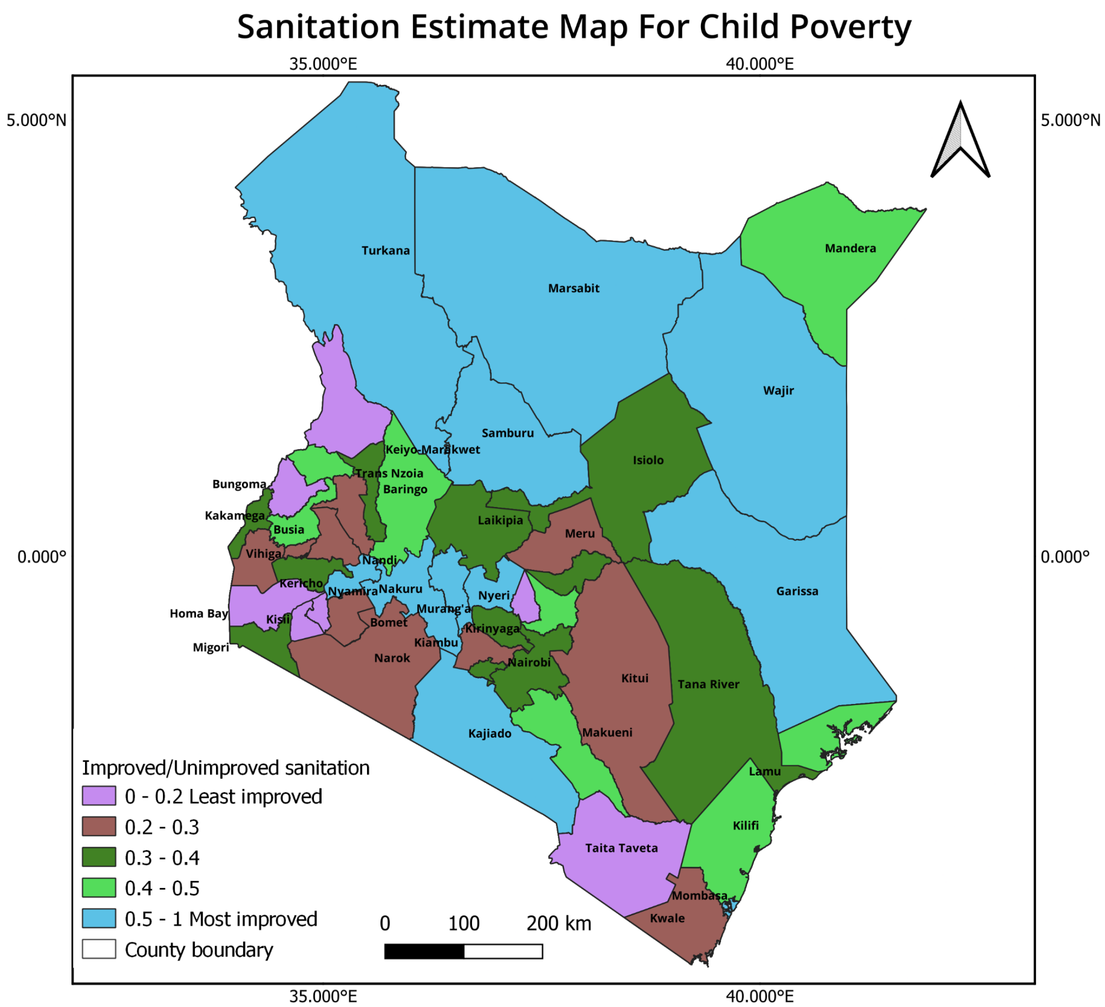


Figure 26: Sanitation estimate map for child poverty

This study analyzed multidimensional deprivation for children in Kenya, with the three main UNICEF rights-based factors being housing, water, and education. The study has also applied UNICEF's methodology to estimate multidimensional child poverty. This study finds that the majority of counties are deprived in one or more dimensions. About one in three counties is deprived in two or more dimensions. There is a stark rural/urban divide, as rural children are more likely to be deprived of many numbers of factors. Compared to previous studies such as that by (Yekaterina Chzhen, and Lucia Ferrone, 2016) of school-age children in Bosnia and Herzegovina experience deprivation in one or more of the seven categories, and one in four experience deprivation in three or more of the seven dimensions. This is almost similar to what is happening in Kenyan counties. Children in the northern and eastern counties are more deprived of these factors, tending to have the highest amount of multidimensional child poverty in Kenya. For the government to eradicate multidimensional child poverty, these counties should be given the highest priority.

Counties in the north and eastern of Kenya are currently on a good trajectory as seen in the predictions. Despite being the most affected by climatic conditions in Kenya and having the most pastoralist communities, with the current variables in the model, child poverty is on the decline. The findings of the study by (Milu Muyanga, and Phillip Musyoka, 2014) underscore the increasing importance of post-secondary education in the welfare of rural households. The success of any education policy in reducing poverty hinges on participants excelling beyond secondary schools and acquiring skills that are in demand in the job market. This study predicts an increase in the number of years spent on schooling helps in reducing multidimensional child poverty which supports that Kenya is on a good trajectory for eradicating multidimensional child poverty. As the model has factored the gdp ppp of each hexagon, the monetary poverty has been factored in the prediction maps. Counties with low gdp ppp are facing the highest multidimensional child poverty. Implementing more resources such as schools, and clean water in these counties may help in reducing these multidimensional factors that greatly contribute to child poverty.

CHAPTER FIVE: CONCLUSION AND RECOMMENDATION

This study has argued and estimated factors that contribute to multidimensional child poverty. Factors that the government and local communities should focus more on to eliminate child poverty have been identified. The model has been able to estimate and predict multidimensional child poverty. Highly visualized maps have also been made indicating areas facing high child poverty from the factors. In the next update, more focus would be on trying more regression models with the DHS survey data. Overfitting of the regression models should also be factored in when the target dataset variable is not large. After the next survey has been done, probably next year (2024) by DHS in conjunction with the Ministry of Health, the results of the model would be compared with that new data. More data on households should be corrected and complete data should be added to the survey dataset such as nutrition. Children who are malnourished are a great indicator of multidimensional child poverty.

REFERENCES

Alkire, S., & Foster, J. E. (2009, December). Counting and multidimensional poverty measurement. (OPHI Working Paper 32). Oxford: Oxford Poverty and Human Development Initiative. *https://www.ophi.org.uk/wp-content/uploads/OPHI-wp32.pdf*

Guide to DHS Statistics DHS-7 The Demographic and Health Surveys Program. (n.d.).
https://www.dhsprogram.com/pubs/pdf/DHSG1/Guide_to_DHS_Statistics_DHS-7.pdf
World Bank. (2018). WDI - Poverty and Inequality. Worldbank.org. *https://datatopics.worldbank.org/world-development-indicators/themes/poverty-and-inequality.html*

Unicef. (2014). Child Poverty. Unicef.org; Unicef. *https://www.unicef.org/social-policy/child-poverty*

Chzhen, Y., Ferrone, L. Multidimensional Child Deprivation and Poverty Measurement: Case Study of Bosnia and Herzegovina. Soc Indic Res 131, 999–1014 (2017). *https://doi.org/10.1007/s11205-016-1291-8*

Jose Cuesta, Mario Biggeri, Gonzalo Hernandez-Licona, Ricardo Aparicio
& Yedith Guillén-Fernández (2020): The political economy of multidimensional child poverty
measurement: a comparative analysis of Mexico and Uganda, Oxford Development Studies, DOI: 10.1080/13600818.2020.1739261
https://doi.org/10.1080/13600818.2020.1739261

Pradhan J, Ray S, Nielsen MO, Himanshu (2022) Prevalence and correlates of multidimensional child poverty in India during 2015–2021: A multilevel analysis. PLoS ONE 17(12): e0279241. *https://doi.org/10.1371/journal.pone.0279241*

Chen WH, Corak M. Child poverty and changes in child poverty. Demography. 2008 Aug;45(3):537-53. doi: 10.1353/dem.0.0024. PMID: 18939660; PMCID: PMC2831388.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2831388/

Unpacking deprivation bundles to reduce multidimensional poverty Global Multidimensional Poverty Index 2022 OPHI Oxford Poverty & Human Development Initiative. (n.d.).
https://hdr.undp.org/system/files/documents/hdp-document/2022mpireportenpdf.pdf

Silwal, A., Engilbertsdóttir, S., Cuesta, J., Newhouse, D., & Stewart, D. (n.d.). Global Estimate of Children in Monetary Poverty: An Update.
https://documents1.worldbank.org/curated/en/966791603123453576/pdf/Global-Estimate-of-Children-in-Monetary-Poverty-An-Update.pdf

(Milu Muyanga, Phillip Musyoka, 2014) Households' Incomes and Poverty Dynamics in Rural Kenya: A Panel Data Analysis .
https://www.africaportal.org/documents/17799/RP287.pdf

Akande, Sheerifdeen & Olaide, Akande. (2019). child poverty mapping: towards effective child poverty reduction poverty mapping as a tool for effective child poverty reduction.
https://www.researchgate.net/publication/332371396_child_poverty_mapping_towards_effective_child_poverty_reduction_poverty_mapping_as_a_tool_for_effective_child_poverty_reduction/