AFRICAN CENTRE OF EXCELLENCE
IN DATA SCIENCE

Predicting Stunting Status among children under five years: The case study of Tanzania

By

Lucy Lawrence Sylivester

Registration Number:219014027

A dissertation submitted in partial fulfillment of the requirements for the Degree of Master of Science in Data Science

University of Rwanda, College of Business and Economics

The African Centre of Excellence in Data Science (ACE-DS)

Supervisor: Dr. Dieudonne N. Muhoza

September ,2020

DECLARATION

I declare that this dissertation entitled Predicting stunting status among children under five years: The case study of Tanzania is the result of my own work and has not been submitted for any other degree at the University of Rwanda or any other institution.

Student Name: Lucy Lawrence Sylivester

Signature: 

APPROVAL SHEET

This dissertation entitled "Predicting stunting status among children under five years: The case study of Tanzania" written and submitted by Lucy Lawrence Sylivester in partial fulfillment of the requirements for the degree of Master of Science in Data Science majoring in Biostatistics is hereby accepted and approved. The rate of plagiarism tested using Turnitin is 19% which is less than 20% accepted by ACE-DS.

_____

Dr. Dieudonne N. Muhoza

Supervisor

_____

Dr. Ignace Kabano

Head of Training

DEDICATION

I dedicate this work to God the Almighty, my source of inspiration, insight, understanding and knowledge. Throughout the program, he was the source of my determination.

I also dedicate this work to my husband, David Mzia, who encouraged me all the way and whose encouragement ensured that I gave everything I needed to finish what I began.

To my daughter, Gracegianna, who was affected by this quest in every possible way. To my friends who have helped and supported me in all circumstances, I appreciate your time.

ACKNOWLEDGEMENTS

Firstly, thanks to almighty GOD and St. Virgin Mary, mother of Jesus Christ for giving me the abilitytobewhereIamnowandalwayswithmewhereeverIgo.Youhavedonesomuchfor me; I'm gland thanks, Amen.

Secondly, I extend my heartfelt grateful thank to my Supervisor Dr. Dieudonne N. Muhoza for the guidance in the development of this thesis into what it is today by giving constructive comments and overall guidance. I also thank him for taking his time to advise me without a schedule at any time when I needed his hands.

I'm grateful to my family and my friends, without whom this research wouldn't be succeeded. Lastly not the least, I thank the University of Rwanda for sponsoring this thesis research.

ABSTRACT

Stunting is a public health concern for children under five years. This issue affects the physical growth and cognitive capacity of children, which affects the child's potential development as growth progresses. Stunting is caused by a nutritional deficiency where, as needed by the health guidelines, the child does not get enough nutrients. Stunting or chronic malnutrition affects children under five years in Tanzania at a rate of 34.7 percent, which is still high.

The goal of this study is to find the most risk factors for stunting in Tanzania, as well as the best classifier for predicting stunting in children under the age of five**.**

Secondary data from the Tanzania 2015 Demographic and Health Survey, Children file was used in this study. The study included a total of 8289 children under the age of five. Five algorithms, Random forest, Decision tree, K-near-neighbor, Support vector machine and Logistic regression, were used in building the model. To obtain the best classifier evaluation metrics were used to get the performance of each classifier, the metrics used are precision, recall, F1 score, accuracy and AUC.

The findings reveal that the best classifier which predicts the stunting status of children under five years was Random forest because it has performed better than the other classifier with the precision score of 89%, recall score of 84%, F1 score of 86% and accuracy of 83% and AUC score of 92%. The most risky factors of stunting were children from southern highlands, children born with mothers with primary education, male children, babies from poorest family, children between 36-47 month and children born with mothers between 15-24 years.

The study concludes that advanced technology is playing a major role in contributing to the development in health system. Machine learning as one of the technology tools is being used in health issues for predicting the diseases and identifying hidden pattern which couldn’t been revealed easily by other method. This research managed to improve a model that will be used to predict the stunting status of children under the age of five. This will aid in the reduction of child stunting.

Keywords: Stunting, Under–five years, Machine learning, Tanzania

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

WHO	:	World Health Organization
TDHS	:	Tanzania Demographic and Health Survey
NMNAP	:	National Multi-Sectoral Nutrition Action
BMI	:	Body Mass Index
TBAs	:	Traditional Birth Attendants
HAZ	:	Height-for-age Z-score
WHZ	:	Weight-for-height Z-score
WAZ	:	Weight-for-age Z-score
SD	:	Standard Deviation
ANC	:	Antenatal Care
K-NN	:	K-Nearest Neighbors
SVM	:	Support Vector Machine
RF	:	Random Forest
ML	:	Machine Learning
ROC	:	Receiver Operating Characteristics
AUC	:	Area Under the ROC Curve
MoH:	:	Ministry of Health

CHAPTER ONE

INTRODUCTION

1.1 Background

Stunting is a consequence of persistent under nutrition, which is caused by insufficient nutrition. Stunting denotes children that are short for their age, with a height for age that is -2SD below the WHO Child Growth Standards median (Nations & Unicef, n.d.). Wasting and underweight are two more side consequences of insufficient diet. Wasting is a condition in which a kid loses weight or fails to acquire weight, causing him or her to be too thin for his or her height (Motbainor & Taye, 2019). Malnutrition affects a high number of youngsters in Africa and Asia. children who have been wasted have a higher risk of sickness and mortality than their peers (Stella G. Uzogara, 2016). Despite the fact that children under the age of five rely on their mothers for nutrition, children under the age of five are disproportionately affected by these symptoms.

Globally stunting affects 154.8 millions of children below five years in the world, representing 22.9% of the world total children under five years or more than one child in five(WHO, 2018). Africa and Asia are the most regions affected by children stunting: 59.0 million and 86.5 million respectively (Unicef/ WHO/The World Bank, 2019). In these regions, children suffer from stunting compared to other regions across the world because of poor health facilities and lack of awareness among mothers on giving to their children a balanced diet (UNICEF-WHO-The World Bank, 2017).

According to WHO on global nutrition report (Fanzo, J. Hawkes, C.,Udomkesmalee, E., Afshin, A., Allemandi,L., Assery, O., Baker, P., Battersby, J., Bhutta, Z. & Chen, 2018) shows that the burden of malnutrition in the world persists to be unacceptably high and the progress of decline unacceptably slow. Malnutrition is more responsible than any other causes for diseases targeting children, with various stresses on children under the age of five (Blössner et al., 2005).

In Tanzania, 34% of children under five years are stunted. They represent about 2.7million of children. Stunting is higher in Tanzania main land with prevalence of 35% than in Zanzibar with prevalence of 24%. Across the country three regions are found to have the highest prevalence which are Rukwa (56%),

Njombe (49%) and Ruvuma (44%), while Dar es Salaam is the region with the lowest prevalence of 15%. Children who come from the families with lowest wealth quintile are the mostly affected than those who come from families with highest wealth quintile (Situation, 2017).

According to the study that was conducted on trends of stunting from 1992 to 2015 (Sunguya, Zhu, Mpembeni, & Huang, 2019), Stunting has fallen by 30% in the last 25 years, yet it still affects one out of every three children. The most affected groups are boys, children in rural regions, children from low income families, families without schooling, and female-headed households, all of whom can benefit from improved baby and young feeding habits.

Resulting from maternal malnutrition, lack of exclusive breastfeeding, poor health care and poor hygiene, stunting affects the physical development of children and their cognitive ability which results in brain problems. The consequences of stunting can lead to mortality and morbidity where the child will be susceptible to diseases since the immunity is weak. In addition to that, stunting affects the health status, educational performance and productivity of children as they grow. Unfortunately, three millions of children below five years in Tanzania are stunted (*Survey 2018*, 2019).

Although several studies have been conducted about the factors leading to stunting in Tanzania, most of them were concentrated at only district level but not at national level. In addition to this, few studies have been conducted integrating machine learning approach and to our knowledge there is no study that has been conducted for predicting the child stunting status.

Therefore, this study aims at identify the risk factors of stunting and then integrate them in machine learning algorithms to construct predictive models. The end product of this work will help to make intervention for solving under-nutritional problems among under-five children in Tanzania.

1.2 Problem Statement

Different studies employed various models and approaches to investigate the factors that contribute to stunting in children under the age of five. The most common models employed were logistic regression, survival analysis, decomposition, and multivariate analysis. The variables associated with malnutrition

for children below five years were determined using logistic regression in a study done in Northwest Ethiopia. (Belaynew W, 2014) (Mwenda & Wei, 2014).

On the other hand, the study conducted in Nigeria use multilevel analysis to examine contributing determinants of stunting (Akombi et al., 2017). However those approaches have limitations in prediction capability and failed to explain to a full extent the factors that account for the variation in stunting. The new method of machine learning presents opportunities for enhancing predictions as well as identifying the risk groups for targeting of specific interventions. It was found that this new statistical application provides more accurate estimates of statistical analyses and it uses classification algorithms to explain the outcome in terms of the predicting variables (Kraamwinkel, Ekbrand, Davia, & Daoud, 2019).

There is a gap in applying machine leaning methods to analyze stunting in Tanzania and there is no study conducted in Tanzania applying of machine learning methods in analysis of stunting status. Therefore, the purpose of this study is to examine the stunting status of Tanzanian children under the age of five using machine learning approaches (Baumgartner & Graber, 2007).

1.3 Objectives of the study

1.3.1 General objective

The general purpose of this research is to identify a classifier that predicts better stunting among children under the age of five in Tanzania.

1.3.2 Specific objectives

a) To identify the most risk factors which contribute to stunting among children under five years.
b) To evaluate the performance of evaluation metrics based on machine learning algorithms.
c) To identify the best classifier in predicting stunting for children under five years.

1.4 Significance of the study

This research will develop an effective model for predicting the occurrence of stunting in Tanzanian children under the age of five. The model will have numerous benefits to Ministry of Health and Social

Welfare, regions and stakeholders for intervention of child under-nutrition problems. The Government has been trying to improve child nutritional by introducing different programs. In addition to this, the findings of this study can be used by The National Multi-Sectoral Nutrition Action Plan (NMNAP, 2016-2021) where among the targets aimed to reduce the percentage of stunted children in Tanzania from 34.5% to 28% by 2021 for making better decision regarding early detection and prevention of under-nutrition as well as supporting equitable nutritional logistic distribution throughout the country.

The model will help health planners to understand the nature and patterns of the occurrence of under-nutrition in Tanzania. Additionally, the end product of this study can be used as a baseline that will helps as a reference for conducting further research in the future. Generally, the products of this study add a clue for intervention of under nutrition, and then improve quality life of individual as well as economic status of the community.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Other researchers' perspectives on the consequences of stunting, factors linked with stunting status, results from other scholars on the part of machine learning, and the trajectory of stunting in Tanzania are discussed in this chapter.

2.2 Consequences of Stunting

Stunting is a condition in which linear growth failure is a symptom of numerous pathological conditions linked to increased morbidity and mortality, decreased physical growth capacity, impaired neurodevelopment and cognitive function, and a higher risk of adult chronic disease (Prendergast & Humphrey, 2014). Infection-related morbidity and death, notably pneumonia and diarrhea, are linked to stunting (Olofin et al., 2013), but also sepsis, meningitis, TB, and hepatitis, implying a systemic immunological dysfunction in severely stunted children.

2.3 Risk factors for stunting in children under age of five

The finding from the study that was conducted in Nigeria (Akombi et al., 2017) revealed that child's sex (male), Pre-mature babies, children born in poorest and middle wealth index, breastfeeding period more than one year, geopolitical zone (North East and North West) found to be factors associated to stunting.

Moreover, the study conducted in Tanzania on the determinants related to stunting among children below five years. The findings revealed that an increase in child age is direct proportion to stunting for children between 0-23 months; boys were expected to be stunted than girls. Pre-mature babies were expected to be stunted than full term babies. Children born by mothers below twenty years were at risk of being stunted than those born by mothers between 20-29 years.

Children who lived in the household with no clean drinking water were at high risk of being stunted than the ones who can get safe drinking water. The fathers who had no education and worked in farms

their children were expected to be stunted than the one who had education. Children who were delivered at home and mothers who did not attend antenatal clinic were at higher risk of being stunted (Chirande et al., 2015).

Similarly, the research conducted in Zambia (Mzumara, Bwembya, Halwiindi, Mugode, & Banda, 2018)disclosed the risk factors that contribute stunting. Boys found to be stunted than girls, Children whose mother drinking water intake was low were expected to be stunted than the ones drinking water intake was high, children from poor households and children with mothers without education were susceptible to be stunted.

Several studies have helped to recognize the risk factors that can cause stunting among children below five years, with various findings showing that children at risk of stunting are those from poor backgrounds, mothers without an education history, male children, poor water facilities, and poor health facilities (Akombi et al., 2017).

Furthermore, in Bangladesh the study was conducted on determining factors leads to stunting children between 0-23 months. The findings indicated that the gender of the baby (male) was more likely to be stunted, birth weight. Children of mothers without education were susceptible to stunting than the ones with at least primary education. Toilet hygiene status and food safety status were significantly correlated with stunting (Mistry et al., 2018).

2.4 Related study on Machine learning and malnutrition prediction

A variety of data mining and machine learning methods have recently been used in the healthcare industry to derive secret information from nutritional data. The goal of these strategies is to provide valuable for healthcare decision-makers and policy-makers to make interventions for eliminating stunting. Machine learning is assisting the prediction and interpretation of the effects of malnourished patients.

Machine learning algorithms were expected to predict the nutritional status of children under the age of five in a study done in Afghanistan. Afghanistan Nutrition SMART Survey data was used to apply

Random forest classifier data. The results were compared to those obtained using the Logistic Regression statistical technique. Random forest performed well for malnutrition with the highest accuracy. The study described how classification techniques for data mining can identify malnutrition status for children below the age of five (Momand, Mongkolnam, Kositpanthavong, & Chan, 2020).

The investigator used J48, Naïve Bayes and PART rule induction algorithms to construct the model for the prediction in the study conducted in Ethiopia for predicting under five nutritional status. He found that J48 was the best algorithm with precision, WTPR and WROC of 65.3 percent, 63.6 percent and 82.2 percent respectively, compared to the other algorithms (Khare, Kavyashree, Gupta, & Jyotishi, 2017)

2.5 Trend in stunting status among children under the age of five in Tanzania between 1991 and 2016

For the past 25 years, stunting has decreased from 50 percent to 35 percent in Tanzania since 1991 to 2016. The Tanzania demographic Health Survey was used to analyze the pattern of stunting to show the decrease. However, the government concentrated to diminish the prevalence of stunting, but the problem is persistently greater and varies between regions. The variation of stunting between regions is due to children living in female-headed households, children aged 24-35 months and with inconsistent or non-breastfeeding families, mothers with low education, children who come from poor families were the associated risk factor for stunting as per the report (Sunguya et al., 2019).

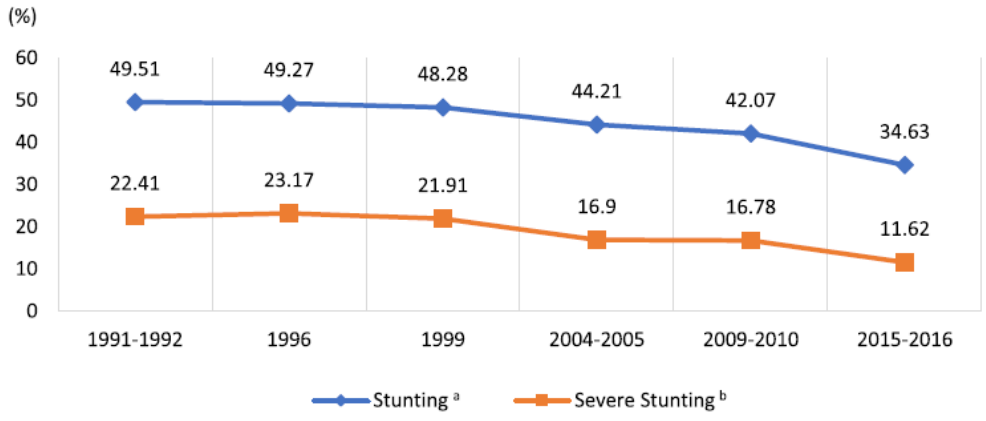

Figure 1: Trend in prevalence of stunting in Tanzania from 1991 to 2016 TDHS Surveys

CHAPTER THREE

DATA AND METHODS

3.1 Introduction

The data source and methods utilized in the data classification process by employing classifiers to predict stunting in children under the age of five are described in this chapter. The analysis uses five classifiers to achieve the goal of prediction: logistic regression, K-nearest-neighbors, support vector machine, decision tree, and random forest are the classifiers.

3.2 Data source

The research uses data from the Demographic and Health Survey of Tanzania (2015-2016 TDHS).The dataset is cross-sectional data and is geographically representative. For this study, the child's file is used for anthropometry estimation concerning children between 0-59 months. The missing values in the TDHS dataset from 2015-2016 were omitted from the study.

3.3 Sample design and sample size

The sample design for the Tanzania Demographic and Health Survey in 2015-2016 was intended to provide estimates for the entire country, as well as mainland Tanzania and Zanzibar. In the first stage, sample sites (clusters) were chosen from enumeration areas (EAs) drawn up for the 2012 Tanzania Population and Housing Census. There were a total of 608 clusters chosen. A systematic survey of households was conducted. Prior to the fieldwork, all 608 clusters were given a comprehensive list of households. The list was then randomly selected 22 households from each cluster, providing a 13,376-household representative probability sample for the 2015-16 TDHS. The 2015-16 TDHS included all women between the ages of 15 and 49, whether they were regular residents or guests in their homes, and they were all eligible for an interview.

Tanzania was divided into nine geographic zones to estimate geographic differentials for some demographic indicators, MoHCDGEC's Reproductive and Child Health Section also uses this classification system.

3.4 Measurement of child nutritional status

Children's nutritional status is an important indicator of their overall health. Nutritional status is determined by anthropometric height, weight, and age. These parameters were used to measure the nutritional status of Tanzanian children in the 2015-16 Tanzania Demographic and Health Survey. These data are critical in identifying malnourished children under the age of five, who are at an elevated risk of death (Survey & Survey, 2015).

The anthropometric details of children's weight, height, and age were used to create nutritional status indexes. Formulated indexes include height-for-height, weight-for-height, and weight-for-age. According to WHO standards, such indices are converted into z-scores and categorized as height-for-age score (HAZ), weight-for-height score (WHZ), and weight-for-age score (WAS) (WAZ). According to WHO, the height-for - age score (HAZ) known as stunting is expressed as standard deviation (SD).

Stunting (Height-for-Age Index): Anthropometric measures of both height and age are used for the HAZ estimation. This index helps to generate data on the chronically stunted group of under-nutrition, according to the WHO. If HAZ falls below -2SD from WHO standards in children below five years, the children are recorded as stunted (WHO Child Growth Standards, n.d.).

3.5 Variables included in the study

The TDHS 2015-2016dataset, child’s file contains a total of 10,233 cases of children information and 1240 variables, with all births of interviewed women age 15-49 years in the 5 years preceding the survey.

3.5.1 Dependent Variable

Stunting status, which is divided into two categories: stunted and unstunted, is the dependent variable in this study. This variable is derived from the dataset's variable named height-for-age. The WHO (2006) child growth standards are used to establish this indicator, and height-for-age in the dataset is recorded in terms of standard deviation. According to WHO (2006), a child is termed stunted if his or her height-for-age Z-score is less than 2 SD, If the height-for-age Z-score is between -2 and 2 SD, it is considered normal. These growth criteria were created using data from the WHO Multicenter Growth

Reference Study (WHO Child Growth Standards, n.d.) and expressed in standard deviation units based on the Multicenter Growth Reference Study median.

3.5.2 Independent Variables

Based on literature reviewed the study considered the following as independent variables in the table 1 below.

Table 1: Independent Variables included in the Study

Variable Group	Variable Name	Measurement (categories)
Socioeconomic factors	Mother education	No education, Primary, Secondary, Higher
	Wealth Index	Poorest, Poorer, Middle, Richer, Richest
	Residence	Urban, Rural
Geographic	Geographic Zones	Western, Northern, Central, Southern highlands, Southern, Southern west highlands, Lake, Eastern, Zanzibar
Bio demographic factors	Mother age in years	15-24, 25-34, 35 and above
	Child age in month	0-10, 11-21, 22-32, 33 and above
	Sex of the child	Male, Female
	Sex of the house hold head	Male, Female
Environmental factors	Source of Drinking water	Protected, Unprotected
	Type of toilet	Flushed, Not-flushed
Behavioral factors	ANC	Once, Twice, Trice and above
	Previous birth interval	below 24 months, 24 months and above

3.6 Methods of data pre-processing

Data pre-processing is the machine learning techniques used to clean data since the data is collected in different format of which data pre-processing is required to fit the purpose of the study. Different stages are been conducted during data pre-processing where data cleaning can be done according to the target of the study (Gaurav & Patel, 2020). The following are the process conducted during data pre-processing.

3.6.1 Data cleaning

Data cleaning is an important aspect to be done before building the model since it will affect the analysis if the data is not clean. Errors can happen during the data collection, data entering, data transformation and extraction. All these are possible to make the data not good for analysis, hence data cleaning is needed (*DATA*, 2016). The study includes the children who were under five years, the one who were above five years where excluded from the study together with children who did not met the anthropometric measurement of the stunting status. In addition, individuals with high missing values were deleted from the dataset.

3.6.2 Missing values handling

In the process of handling missing value was crucial because some of the variable contained missing values. To solve this problem imputation was done where the missing values were replaced with the means.

3.7 Methods of Analysis

Descriptive analysis was done to check the distribution of independent variable with dependent variable for the understanding of the data. The other part was machine learning where five (5) machine learning classifiers Logistic Regression, Random Forest, Decision Tree, Support Vector Machine and K-near-neighbor were used to classify the data and make the prediction of stunting status of the children under five years.

3.7.1 Descriptive Analysis

The distribution of independent variables to the dependent variable, stunting status, was checked in this section by cross tabulating the dependent and independent variables. Every independent variable showed the distribution.

3.7.2 Algorithms for building the predictive model

Machine learning algorithms are model-free methods for solving categorization problems efficiently.. Hence, the performances of these ML algorithms will be compared with the statistical classifier. The following are the methods which will be considered in this study.

a) Decision trees (DT)

A decision tree is a non-parametric technique that classifies a data set based on the problem's predictive structure. If the dependent variable is continuous, decision tree creates regression trees, and produces classification trees for categorical variable. It will be used to identify a binary outcome variable for this research which is the status of stunting for children below five years (Diego, 1984).

b) K-nearest neighbors (K-NN)

K-near-neighbors is a robust and adaptable classifier that belongs to the supervised learning algorithm family. Since the distribution of data does not make any explicit assumptions, K-NN is a non-parametric algorithm. This approach keeps track of all existing cases and categorizes new ones using a similarity score. A case is classified using a distance function and the majority vote of its neighbors, with the case being allocated to the most common class among its nearest k neighbors (Cover & Hart, 1967).

c) Support vector machines (SVM)

SVM is a kernel-based supervised machine learning technique that is often used in classification problems.. By optimizing the margin between the classes, the SVM algorithm creates a hyperplane that precisely divides the training observations according to their class labels. SVM allocates a test observation to a class based on which side of the hyperplane it is on (Le Thi, Nguyen, & Ouchani, 2008).

d) Random forest (RF)

RF is a technique of classification focused on the "rising" of an ensemble of ordered classifiers of trees. Features of this identity are often used for classification using each classification tree in the forest in order to identify a new entity. The grown trees are constructed randomly, and for a class name, each tree provides a classification (or 'voting'). The decision is based on the votes of the majority over all the forest trees (Breiman, 2001).

3.8 Methods of Performance Evaluation for Predictive Models

Evaluating a model is important because it allows you to evaluate how well a classifier performs as a general model. This means that the input-output associations developed from the training dataset should work in the test and validation datasets as well.

To check the accuracy of a ML classifier, the most novel way is to extensively test the classifier on a set of independent samples in a way to incorporate all possible sources of variability to be experienced. To estimate the predictive accuracy of the classifier from the training data, one way is to develop a k-fold cross-validation. This approach offers "internal estimates" of the classification models' predictive accuracy. In k-fold cross-validation, splitting the data into k subsets of almost equal size, the model is trained k times. One of the subsets is randomly set aside in each of the k cycles of model training, which in turn is used to determine the efficiency of the classifier. In this way, all possible cases of the entire dataset are trained and evaluated, leading to a lower variance in the set estimator and less bias in the true rate estimator, which is the main benefit of using this process.

However, this approach ensures a more precise prediction, while being both computer-intensive and time-consuming. We have focused on the 10-fold Cross Validation approach in our analysis, which has been used in many studies related to health care and medical studies (Kang, Cho, & Zhao, 2010), (Wiener, 2003).Using the criteria uncertainty matrices and receiver operating characteristics (ROC), the efficiency of the algorithms is generally assessed. There are four possible prediction results, which are TR= true positives, TN= true negatives, FP= false positives, and FN= false negatives, in an uncertainty matrix for a two-class case with classes '0' and '1'. Using these four possible results, various metrics

such as accuracy, sensitivity, specificity are commonly measured to evaluate the classifier, as defined by,

a) Accuracy: Is the number of correctly predicted cases of interest out of all cases

   Accuracy= $\frac{TP+TN}{TP+TN+FP+FN}$.

b) Sensitivity: The true positive rate is often referred to as the recall. It is the proportion of positive cases that were identified as positive. It measures how well the classifier classified positive correctly.

   Sensitivity = $\frac{TP}{TP+FN}$

c) Specificity is also called the true negative rate. It is the proportion of correct negative cases that were classified as negative. It measures how well the classifier classified the negative cases.

   Specificity = $\frac{TN}{TN+FP}$

d) Precision: This implies the percentage of cases classified as positive by the classifier that are actually positive.

   Precision = $\frac{TP}{TP+FP}$

The receiver operating characteristic (ROC) curves were created using the expected and true outcomes as inputs. The AUC of the ROC was averaged for the test data sets to compare the discriminating powers of the algorithms (Liu, Fang, Liu, Wang, & Chou, 2016).In theory, the AUC is between 0 and 1, where the maximum value of 1 can be taken by a perfect classifier. However, for random classification, the realistic lower bound is 0.5, and classifiers with an AUC substantially greater than 0.5 have at least some potential to differentiate between cases and non-cases.

CHAPTER FOUR

PRESENTATION OF RESULTS

4.1 Introduction

This chapter describes the results of the descriptive analysis, multivariate logistic regression, and five classifiers used to predict stunting in children under the age of five. Logistic regression, random forest (RF), support vector machines (SVM), decision tree, and K-nearest neighbors are the five techniques employed (K-NN) (Conference, 2017), (Agarwal, 2014).

We applied five alternative algorithms in the training dataset based on the numerous relevant criteria acquired from the univariate feature selection approach in table 3.7.3. Based on the strongly related risk factors, these algorithms were used to classify research participants as stunted or not stunted..

4.2 Summary of Descriptive Statistics

A total of 8289 children were involved in the study, with 5309 (64%) of them not being stunted and 2980 (36%) being stunted. The majority of stunted children came from the South West Highlands, where 46.4 percent of children were stunted, compared to a national average of 34 percent. In comparison to children in urban areas, 38.4 percent of children in rural areas were stunted. When compared to mothers who had at least a primary education, their children were stunted 41.1 percent of the time. Children in poorer households were 40.9 percent stunted compared to other households, and mothers aged 15-24 years show that their children are 37.3 percent stunted compared to other age groups. Male children are stunted by 37.8% more than female children, children aged 36-47 months are stunted by 41.5 percent, and children who are not breastfed are stunted by 40.1 percent more than children who are breastfed. Children in un-piped water households are 39.4 percent stunted when compared to children in piped water households. Table 2 below summarizes the findings.

Table 2: Distribution of the sample according to selected variables and stunting status

Variable	Category	Stunted (%)	Not stunted (%)	Total %(N)
Geographic zones	Western	32.7	67.3	100(798)
	Northern	39.5	60.5	100(623)
	Central	33.9	66.1	100(831)
	Southern highlands	46.0	54.0	100(604)
	Southern	37.2	62.8	100(323)
	South west highlands	46.4	53.6	100(946)
	Lake	36.8	63.2	100(2304)
	Eastern	26.8	73.2	100(693)
	Zanzibar	27.4	72.6	100(1167)
Place of residence	Urban	27.6	72.4	100(1854)
	Rural	38.4	61.6	100(6435)
Maternal Education	No education	41.1	58.9	100(1824)
	Primary	36.8	63.2	100(5017)
	Secondary	27.1	72.9	100(1379)
	Higher	14.5	85.5	100(69)
Wealth index	Poorest	40.6	59.4	100(1960)
	Poorer	40.9	59.1	100(1748)
	Middle	40.7	59.3	100(1603)
	Richer	30.2	69.8	100(1691)
	Richest	23.8	76.2	100(1287)
Mother Age in years	15-24	37.3	62.7	100(2399)
	24-34	34.3	65.7	100(3659)
	35 and above	37.2	62.8	100(2231)
Child's sex	Male	37.8	62.2	100(4152)
	Female	34.1	65.9	100(4137)

Variable	Category	Stunted (%)	Not stunted (%)	Total %(N)
Child's age in month	0-5	20.9	79.1	100(554)
	6-11	21.9	78.1	100(935)
	12-35	37.5	62.5	100(2003)
	36-47	41.5	58.5	100(1569)
	48-59	33.4	66.6	100(1523)
Birth interval in month	Below 24	35.1	64.9	100(3279)
	Above 24	36.5	63.5	100(5010)
Breast feeding	No	40.1	59.9	100(3247)
	Yes	33.3	66.7	100(5024)
Sex of household head	Male	35.5	64.5	100(7036)
	Female	38.5	61.5	100(1253)
Source of Drinking Water	Piped	33.4	66.6	100(4750)
	Un-piped	39.4	60.6	100(3539)
Type of Toilet Facility	Improved	26.4	73.6	100(2402)
	Not improved	39.9	60.1	100(5887)
Number of Antenatal care visits (ANC)	Once	36.0	64.0	100(771)
	Twice	36.3	63.7	100(471)
	More than two visits	32.4	67.6	100(105)

4.3 Multivariate Logistic Regression Results

The risk factors associated with stunting status among children under the age of five are classified using multivariate logistic regression in this study. Table 3 summarizes the results of the complete model and the reduced model derived from the study. Risk factors associated with stunting status have been found to be geographical areas, type of toilet facility, age of mother, sex of the child, wealth index, child's age, and mother education. These risk factors were statistically significance as shown in table 5.

Table 5 shows that children from the Southern Highlands were 1.96 times more likely than children from the Western Zone to be stunted (OR = 1.96, 95 percent CI [1.57, 2.44]). Children from the central zone were 0.90 times (OR = 0.90, 95 percent CI [0.73, 1.12]) less likely to be stunted than children from the western zone, while children from the eastern zone were 0.96 times (OR = 0.90, 95 percent CI [0.73, 1.12]) less likely to be stunted. Children from the western zone are less likely to be stunted (OR = 0.96, 95 percent CI [0.77, 1.12]). Children born to mothers with a primary education were 0.85 times less likely to be stunted than children born to mothers with no education (OR = 0.85, 95 percent CI [0.76, 0.96]). Children from poorer homes were 1.01 times more likely to be stunted than children from the poorest households (OR = 1.01, 95 percent CI [0.88, 1.16]). Children born into households without improved toilets were 1.76 times more likely to be stunted than children born into households with renovated toilets (OR = 1.76, 95 percent CI [1.57, 1.98]). Female children were 0.84 times more likely than male children to be stunted (OR = 0.84, 95 percent CI [0.77, 0.92]). Children of female household heads were 1.15 times more likely to be stunted than children of male household heads (OR = 1.15, 95 percent CI [1.01, 1.31]). Children aged 36-47 months were 3.26 times more likely than children aged 0-5 months to be stunted (OR = 3.26, 95 percent CI [2.56, 4.10]). Children born to mothers aged 35 and over were 0.90 times more likely (OR = 0.90, 95 percent CI [1.08, 1.36]) to be stunted than children born to mothers aged 15 to 24.

Table 3: Multivariate logistic regression model of factors associated with stunting status

Variable name	**Full model**				**Reduced model**			
	Odds Ratio	**95% CI**		**P-value**	**Odds Ratio**	**95% CI**		**P-value**
Geographic Zones								
Western	1.00				1.00			
Northern	1.65	1.31	2.08	0.000	0.56	1.25	1.95	0.000
Central	1.11	0.90	1.38	0.327	1.06	0.86	1.31	0.570
Southern highlands	2.18	1.73	2.74	0.000	1.96	1.57	2.44	0.000
Southern	1.19	0.90	1.57	0.223	1.22	0.93	1.60	0.152
South west highlands	1.93	1.58	2.36	0.000	1.81	1.49	2.20	0.000
Lake	1.28	1.07	1.52	0.006	1.21	1.02	1.43	0.032

Variable name	Full model				Reduced model			
	Odds Ratio	95% CI		P-value	Odds Ratio	95% CI		P-value
Eastern	1.07	0.85	1.36	0.553	0.96	0.77	1.21	0.754
Zanzibar	1.13	0.90	1.42	0.284	1.07	0.87	1.31	0.538
Place of Residence								
Urban	1.00							
Rural	1.16	1.0	1.36	0.057				
Maternal Education								
No education	1.00				1.00			
Primary	0.85	0.76	0.96	0.007	0.85	0.76	0.96	0.008
Secondary	0.75	0.63	0.91	0.003	0.75	0.63	0.91	0.003
Higher	0.43	0.21	0.87	0.019	0.43	0.21	0.87	0.019
Wealth Index								
Poorest	1.00				1.00			
Poorer	1.00	0.87	1.15	0.999	1.01	0.88	1.16	0.853
Middle	0.98	0.84	1.13	0.774	0.99	0.85	1.13	0.842
Richer	0.72	0.60	0.87	0.000	0.70	0.59	0.83	0.000
Richest	0.65	0.51	0.86	0.002	0.61	0.48	0.78	0.000
Sex of child								
Male	1.00				1.00			
Female	0.84	0.77	0.92	0.000	0.84	0.77	0.92	0.000
Source of Drinking Water								
Piped	1.00							
Un-piped	0.99	0.89	1.10	0.844				
Type of Toilet Facility								
Improved	1.00				1.00			
Not improved	1.20	1.00	1.44	0.045	1.76	1.57	1.98	0.000
Breastfeeding								

Variable name	Full model				Reduced model			
	Odds Ratio	95% CI		P-value	Odds Ratio	95% CI		P-value
Yes	1.00							
No	0.91	0.81	1.02	0.093				
Birth interval								
Below 24	1.00							
Above 24	0.97	0.84	1.11	0.636				
Number of Antenatal care visits (ANC)								
Once	1.00							
Twice	1.10	0.90	1.34	0.347				
More than two visits	1.00	0.65	1.54	0.983				
Sex of household head								
Male	1.00				1.00			
Female	1.14	1.01	1.31	0.034	1.15	1.01	1.31	0.029
Child's age in month								
0-5	1.00				1.00			
6-11	1.14	0.88	1.48	0.335	1.14	0.88	1.48	0.332
12-35	2.54	2.02	3.19	0.000	2.57	2.04	3.22	0.000
36-47	3.07	2.41	3.90	0.000	3.26	2.56	4.10	0.000
48-59	2.83	2.22	3.60	0.000	2.97	2.36	3.76	0.000
Mother Age in years								
15-24	1.00				1.00			
24-34	0.84	0.73	0.97	0.015	0.83	0.74	0.93	0.001
35 and above	0.91	0.77	1.08	0.286	0.90	0.79	1.02	0.099
constant	0.24	0.16	0.35	0.000	0.25	0.18	0.35	0.000

4.4 Performance of the Classifiers

Building machine learning models requires classifiers where the performance is evaluated with the metrics to obtain the best model as the goal of the study. We use five classifiers for building the predictive models of stunting status of the child under five years and the results are presented in the tables below for each classifier.

4.4.1 Logistic Regression

The results from table 4 below show that the logistic regression classifier has a precision of 73%. It implies that 73% of the cases that the classifier predicted as stunting were correct. A recall score of 56% on the other hand, is the proportion that the classifier picked from the children with stunting status. The F1 score of 63% is the harmonic mean of recall and precision score. F1 score approaching to 1 the better the performance of the model. The accuracy of 59%tells us the model accurately predicted 83% of the cases as either stunting or not stunting. Finally, from figure 2, we observe that the Area under the Receiver Operating Curve (AUCROC) is at a score of 64%. This indicates that the performance of Logistic regression is at 64%.

Table 4: Performance of Logistic Regression classifier

Model	Precision	Recall score	F1 score	Accuracy	AUC score
Logistic Regression	0.7274	0.5575	0.6312	0.5862	0.6368

Figure 2 Logistic Regression ROC Curve

4.4.2 Support Vector Machine

The precision of the support vector machine classifier is 78 percent, according to the results in table 5. It means that the classifier was correct in 78 percent of the cases it predicted as stunting. A recall score of 63 percent, on the other hand, represents the proportion of children who were classified as stunted by the classifier. The harmonic mean of recall and precision is 70 percent for the F1 score. The closer the F1 score gets to 1, the better the model's performance. With a 65 percent accuracy rate, the algorithm correctly predicted 56 percent of the cases as stunting or not stunting. Finally, we can see from Figure 2 that the (AUCROC) is 72 percent. This means that the Support vector machine's performance is at 72 percent.

Table 5: Performance of Support Vector Machine classifier

Model	Precision	Recall score	F1 score	Accuracy	AUC score
Support Vector Machine	0.7825	0.6289	0.6972	0.6532	0.7212

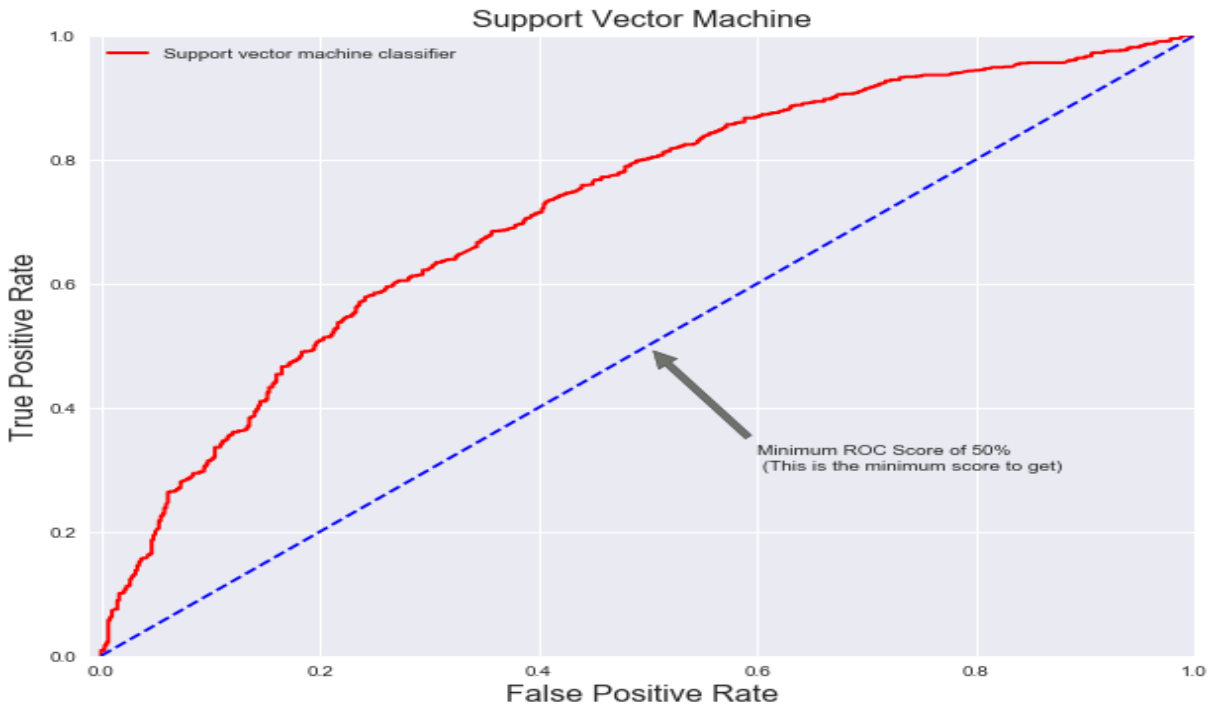

Figure 3: Support Vector Machine ROC Curve

4.4.3 K-Nearest Neighbor

The precision of the k-nearest neighbor classifier is 77 percent, according to the data in table 6. It means that the classifier was correct in 77 percent of the cases it predicted as stunting. A recall score of 82 percent, on the other hand, represents the proportion of children who were classified as stunted by the classifier. The harmonic mean of recall and precision is 79 percent for the F1 score. The closer the F1 score gets to 1, the better the model's performance. The 73 percent accuracy means that the model correctly predicted 73 percent of the cases as stunting or not stunting. Finally, we can see in figure 2 that the (AUCROC) is 77 percent. This means that K-nearest neighbors have a 77 percent success rate.

Table 6: Performance of K-Nearest Neighbors classifier

Model	Precision	Recall score	F1 score	Accuracy	AUC score
KNN	0.7684	0.8224	0.7945	0.7298	0.7670

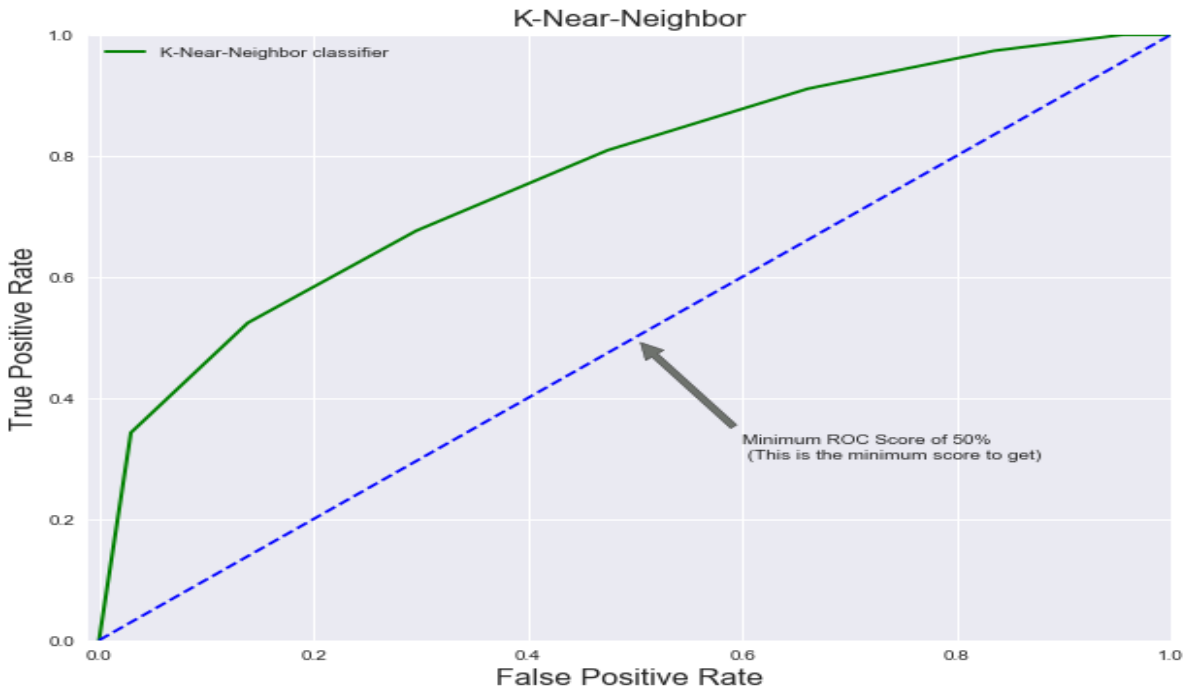

Figure 4: K-Nearest Neighbor ROC Curve

4.4.4 Random Forest

The Random forest classifier has an 89 percent precision, according to the data in table 7. It means that the classifier was correct in 89 percent of the cases it predicted as stunting. A recall score of 84 percent, on the other hand, represents the proportion of children who were classified as stunted by the classifier. The harmonic mean of recall and precision score is 86 percent for the F1 score. The model's performance improves when the F1 score approaches 1. The classifier properly predicted 83 percent of the occurrences as stunting or not stunting with an accuracy of 83 percent. Finally, Figure 2 shows that the AUCROC (Area Under the Receiver Operating Curve) is 92 percent. This demonstrates that Random Forest performs at a 92 percent rate.

Table 7: Performance of Random Forest classifier

Model	Precision	Recall score	F1 score	Accuracy	AUC score
Random Forest	0.8858	0.8395	0.8620	0.8293	0.9187

Figure 5: Random Forest ROC Curve

4.4.5 Decision Tree

The precision of the Decision tree classifier is 92 percent, as shown in table 8 below. It means that the classifier correctly predicted 92 percent of the cases of stunting. A recall score of 81 percent, on the other hand, represents the proportion of children who were classified as stunted by the classifier. The harmonic mean of recall and precision score is 86 percent for the F1 score. The 83 percent accuracy means that the model correctly predicted 83 percent of the instances as stunting or not stunting. Finally, we can see from Figure 2 that the Area Under the Receiver Operating Curve (AUCROC) is 91 percent. This means that the Decision Tree's performance is 91 percent.

Table 8: Performance of Decision tree classifier

Model	Precision	Recall score	F1 score	Accuracy	ROC score
Decision Tree	0.9162	0.8101	0.8599	0.8323	0.9076

Figure 6: Decision Tree ROC Curve

4.4.6 General ROC curves

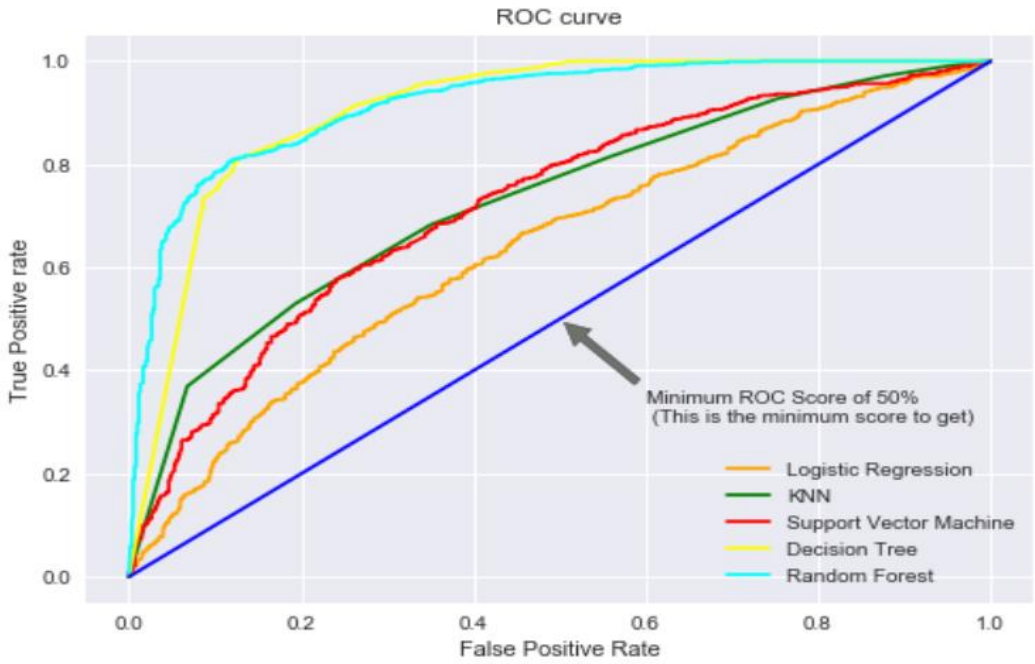

Figure 7: General ROC Curves

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

After finding the most risk indicators, the goal of this study was to find the best classifier for predicting stunting status among Tanzanian children under the age of five. The research focused on three particular aims to achieve this general goal: identify the most significant risk factors for stunting in children under the age of five, assess the effectiveness of assessment metrics based on machine learning algorithms, and choose the best classifier for predicting stunting in children under the age of five.

Researchers used five machine learning methods to develop prediction models of stunting status in children under the age of five: logistic regression, random forest, decision tree, K-Nearest Neighbor, and super vector machine classifiers. A multivariate logistic regression model was used to investigate risk factors for stunting.

In compared to multivariate logistic regression, the findings of this study revealed that machine learning methods were the most effective for developing high-model-accuracy stunting prediction models. In order to choose the best classifier, the F1 score, which is the harmonic mean of precision and recall, the accuracy, and the AUC score were used to evaluate the classifier's performance. Among the five classifiers employed in this study, the random forest classifier and decision tree performed nearly identically, with similar F1 scores and accuracy but a small difference in AUC score. The AUC score was used as the criterion for classifier evaluation. With an AUC value of 0.9, Random forest was found to be the most effective predictor of stunting status as shown in table 9 below.

Table 9: Summary of evaluation metrics

Model	Precision score	Recall score	F1 score	Accuracy	AUC score
Logistic regression	73%	56%	63%	59%	64%
K-NN	77%	82%	79%	73%	77%
SVM	78%	63%	70%	65%	72%
Decision tree	92%	81%	86%	83%	91%
Random forest	89%	84%	86%	83%	92%

Machine learning plays an important role in improving the health sector in various ways; the prediction of diseases has help in reducing death rate in diseases such as cancer and diabetes. Using machine learning helps in identifying some hidden information which traditional methods could not identify such patterns.

The goal of this study was to find the most risk factors that could be identified using multivariate logistic regression, as well as to find a classifier that could be used to predict the stunting status of children under the age of five.

The risk variables for stunting in children under the age of five were identified using multivariate analysis. Children from the eastern and central zones, children born to moms who do not have a high school diploma, male children who are more likely to be stunted, children from low-income families, and children who do not have access to proper bathroom facilities are among these traits.

The second goal was accomplished by the development of five machine learning models: logistic regression, K-NN, support vector machine, decision tree, and random forest. The final goal was to choose the best model by evaluating its performance using the accuracy and AUC scores as evaluation measures. Random forest was shown to have the best AUC score among the five models when compared to the other classifiers.

5.2 Future Work and Recommendation

Random Forest as the best classifier identified in this research should be applied in the field and monitored to validate the findings. For the purpose of reducing stunting under five years. This will help to come up with different interventions which will facilitate to reduce the problem of stunting for children under five years.

The data set used in this study was five years old (2010-2016). It is Demographic and Health Survey which is nationally representative including child's information. In the future, for prediction and by applying time series machine learning algorithms data from hospital where the prevalence of stunting is high and data of more than five years could be the best to predict stunting.

This study focused on only five algorithms: Logistic Regression, Support Vector Machine, K-Nearest Neighbor, Random Forest and Decision Tree but there are other classifiers for machine learning such as XGboost and GradientBoosting which have higher predicting power.These algorithms are favored when the data has a mixture of categorical and numerical feature.

Policy makers and health service providers should use and implement the findings to real field practices and provision of care to the children who are affected by stunting. As policy makers should make polices which will help the mother to get the necessary education about nutrition on how to feed the child with balance diet without forgetting some of programs which encourage women to practice exclusive breastfeeding.

REFERENCES

Agarwal, S. (2014). Data mining: Data mining concepts and techniques. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. https://doi.org/10.1109/ICMIRA.2013.45

Akombi, B. J., Agho, K. E., Hall, J. J., Merom, D., Astell-burt, T., & Renzaho, A. M. N. (2017). Stunting and severe stunting among children under-5 years in Nigeria : A multilevel analysis. *BMC Pediatrics*, 1–16. https://doi.org/10.1186/s12887-016-0770-z

Baumgartner, C., & Graber, A. (2007). Data mining and knowledge discovery in metabolomics. *Successes and New Directions in Data Mining*, *39*(11), 141–166. https://doi.org/10.4018/978-1-59904-645-7.ch007

Belaynew W, B. G. (2014). Assessment of Factors Associated with Malnutrition among Under Five Years Age Children at Machakel Woreda, Northwest Ethiopia: A Case Control Study. *Journal of Nutrition & Food Sciences*, *04*(01), 1–7. https://doi.org/10.4172/2155-9600.1000256

Blössner, M., Onis, M. De, Prüss-üstün, A., Campbell-lendrum, D., Corvalán, C., & Woodward, A. (2005). *Malnutrition Quantifying the health impact at national and local levels*. (12).

Breiman, L. (2001). ST4_Method_Random_Forest. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1017/CBO9781107415324.004

Chirande, L., Charwe, D., Mbwana, H., Victor, R., Kimboka, S., Issaka, A. I., … Agho, K. E. (2015). Determinants of stunting and severe stunting among under-fives in Tanzania: Evidence from the 2010 cross-sectional household survey. *BMC Pediatrics*, *15*(1), 1–13. https://doi.org/10.1186/s12887-015-0482-9

Conference, A. I. (2017). *Yusuf Maitama Sule University , Kano Faculty of Science*. (November), 16–20.

Cover, T. M., & Hart, P. E. (1967). Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

DATA. (2016). (April).

Fanzo, J. Hawkes, C.,Udomkesmalee, E., Afshin, A., Allemandi,L., Assery, O., Baker, P., Battersby, J., Bhutta, Z. & Chen, K. (2018). Global Nutrition Report: Shining a light to spur action on nutrition. In *Global Nutrition Report*. https://doi.org/http://dx.doi.org/10.2499/9780896295643

Gaurav, K. A., & Patel, L. (2020). *Machine Learning With R*. https://doi.org/10.4018/978-1-7998-2718-4.ch015

Khare, S., Kavyashree, S., Gupta, D., & Jyotishi, A. (2017). Investigation of Nutritional Status of Children based on Machine Learning Techniques using Indian Demographic and Health Survey Data. *Procedia Computer Science*, *115*, 338–349. https://doi.org/10.1016/j.procs.2017.09.087

Kraamwinkel, N., Ekbrand, H., Davia, S., & Daoud, A. (2019). The influence of maternal agency on severe child undernutrition in conflict-ridden Nigeria: Modeling heterogeneous treatment effects with machine learning. *PLoS ONE*, *14*(1), 1–16. https://doi.org/10.1371/journal.pone.0208937

Le Thi, H. A., Nguyen, V. V., & Ouchani, S. (2008). Gene selection for cancer classification using DCA. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5139 LNAI*, 62–72. https://doi.org/10.1007/978-3-540-88192-6-8

Liu, B., Fang, L., Liu, F., Wang, X., & Chou, K. C. (2016). IMiRNA-PseDPC: MicroRNA precursor identification with a pseudo distance-pair composition approach. *Journal of Biomolecular Structure and Dynamics*, *34*(1), 220–232. https://doi.org/10.1080/07391102.2015.1014422

Mistry, S. K., Hossain, B., Khanam, F., Akter, F., Parvez, M., Yunus, F., … Rahman, M. (2018). *Individual- , maternal- and household-level factors associated with stunting among children aged 0 – 23 months in Bangladesh*. (11). https://doi.org/10.1017/S1368980018002926

Momand, Z., Mongkolnam, P., Kositpanthavong, P., & Chan, J. H. (2020). Data Mining Based Prediction of Malnutrition in Afghan Children. *2020 12th International Conference on Knowledge and Smart Technology (KST)*, 12–17.

Motbainor, A., & Taye, A. (2019). Wasting in under five children is significantly varied between rice producing and non-producing households of Libokemkem district, Amhara region, Ethiopia. *BMC Pediatrics*, *19*(1), 1–11. https://doi.org/10.1186/s12887-019-1677-2

Mwenda, K., & Wei, G. (2014). *Assessing Indicators of Chronic Child Undernutrition using Machine Learning Techniques*. 11–13. Retrieved from http://www.issn.org

Mzumara, B., Bwembya, P., Halwiindi, H., Mugode, R., & Banda, J. (2018). *Factors associated with stunting among children below five years of age in Zambia : evidence from the 2014 Zambia demographic and health survey*. 1–8.

Nations, U., & Unicef, F. (n.d.). *IMPROVING CHILD NUTRITION The achievable imperative for global progress*.

Olofin, I., McDonald, C. M., Ezzati, M., Flaxman, S., Black, R. E., Fawzi, W. W., … Penny, M. E. (2013). Associations of Suboptimal Growth with All-Cause and Cause-Specific Mortality in Children under Five Years: A Pooled Analysis of Ten Prospective Studies. *PLoS ONE*, *8*(5). https://doi.org/10.1371/journal.pone.0064636

Prendergast, A. J., & Humphrey, J. H. (2014). The stunting syndrome in developing countries. *Paediatrics and International Child Health*, *34*(4), 250–265. https://doi.org/10.1179/2046905514Y.0000000158

Situation, F. S. (2017). *Tanzania : Nutrition Profile*. *2010*.

Stella G. Uzogara. (2016). Underweight, the Less Discussed Type of Unhealthy Weight and Its Implications: A Review. *American Journal of Food Science and Nutrition Research*, *3*(5), 126–142. Retrieved from https://www.researchgate.net/publication/306275675_Underweight_the_Less_Discussed_Type_of_Unhealthy_Weight_and_Its_Implications_A_Review

Sunguya, B. F., Zhu, S., Mpembeni, R., & Huang, J. (2019). Trends in prevalence and determinants of stunting in Tanzania: An analysis of Tanzania demographic health surveys (1991-2016). *Nutrition Journal*, *18*(1), 1–13. https://doi.org/10.1186/s12937-019-0505-8

Survey 2018. (2019).

UNICEF-WHO-The World Bank. (2017). Levels and trends in child malnutrition, Joint Child Malnutrition Estimates, Key findings of the 2017 edition. *UNICEF-WHO-The World Bank*, 1–16. Retrieved from https://data.unicef.org/wp-content/uploads/2017/06/JME-2017_brochure_June-25.pdf

Unicef/ WHO/The World Bank. (2019). Levels and Trends in Child malnutrition - Unicef WHO The World Bank Joint Child Malnutrition Estimates, key findings pf the 2019 edition. *Unicef*, 4. Retrieved from http://www.unicef.org/media/files/JME_2015_edition_Sept_2015.pdf%0Ahttps://pubmed.ncbi.nlm.nih.gov/30430613/

WHO. (2018). Reducing Stunting In Children. In *Equity considerations for achieving the Global Nutrition Targets 2025*. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/260202/9789241513647-eng.pdf?sequence=1

WHO Child Growth Standards. (n.d.).

Predicting stunting status among children under five years: The case study of Tanzania

ORIGINALITY REPORT

19%	15%	13%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	bmcpediatr.biomedcentral.com Internet Source	1%
2	Nidhi Wali, Kingsley E. Agho, Andre Renzaho. "Wasting and Associated Factors among Children under 5 Years in Five South Asian Countries (2014–2018): Analysis of Demographic Health Surveys", International Journal of Environmental Research and Public Health, 2021 Publication	1%
3	onlinelibrary.wiley.com Internet Source	1%
4	ir-library.ku.ac.ke Internet Source	1%
5	doctorpenguin.com Internet Source	1%
6	www.mdpi.com Internet Source	1%
	link.springer.com	

7	Internet Source	<1 %
8	Ha-Neul Yeom, Myunggwon Hwang, Mi-Nyeong Hwang, Hanmin Jung. "Study of Machine-Learning Classifier and Feature Set Selection for Intent Classification of Korean Tweets about Food Safety", Journal of Information Science Theory and Practice, 2014 Publication	<1 %
9	Submitted to University of Cape Town Student Paper	<1 %
10	academic.oup.com Internet Source	<1 %
11	arno.uvt.nl Internet Source	<1 %
12	www.researchsquare.com Internet Source	<1 %
13	repository.sustech.edu Internet Source	<1 %
14	www.tandfonline.com Internet Source	<1 %
15	Damitie Kebede, Yidnekachew Merkeb, Eyerusalem Worku, Hayat Aragaw. "Prevalence of undernutrition and potential risk factors among children under 5 years of	<1 %

	age in Amhara Region, Ethiopia: evidence from 2016 Ethiopian Demographic and Health Survey", Journal of Nutritional Science, 2021 Publication	
16	storage.googleapis.com Internet Source	<1 %
17	Submitted to Mount Kenya University Student Paper	<1 %
18	dokumen.pub Internet Source	<1 %
19	wiredspace.wits.ac.za Internet Source	<1 %
20	Submitted to University of Science and Technology, Yemen Student Paper	<1 %
21	Submitted to University of Technology, Sydney Student Paper	<1 %
22	Ziaullah Momand, Pornchai Mongkolnam, Pichai Kositpanthavong, Jonathan H. Chan. "Data Mining Based Prediction of Malnutrition in Afghan Children", 2020 12th International Conference on Knowledge and Smart Technology (KST), 2020 Publication	<1 %
23	intjem.biomedcentral.com Internet Source	<1 %

24	www.strokegenetics.com Internet Source	<1 %
25	Papia Sultana, Md. Mahfuzur Rahman, Jahanara Akter. "Correlates of stunting among under-five children in Bangladesh: a multilevel approach", BMC Nutrition, 2019 Publication	<1 %
26	Baru Ruth Sharon Apio, Ratib Mawa, Stephen Lawoko, Krishna Nand Sharma. "Socio-economic Inequality in Stunting among Children Aged 6-59 Months in a Ugandan Population Based Cross-sectional Study", American Journal of Pediatrics, 2019 Publication	<1 %
27	distinctionessays.com Internet Source	<1 %
28	lib.ugent.be Internet Source	<1 %
29	Submitted to UNIVERSITY OF LUSAKA Student Paper	<1 %
30	erepository.uonbi.ac.ke Internet Source	<1 %
31	krishikosh.egranth.ac.in Internet Source	<1 %
32	scitepress.org Internet Source	<1 %

33	"Advances in Cyber Security", Springer Science and Business Media LLC, 2021 Publication	<1 %
34	Submitted to University of Queensland Student Paper	<1 %
35	diklatbpom.files.wordpress.com Internet Source	<1 %
36	zaguan.unizar.es Internet Source	<1 %
37	Collins Adu, Edward Kwabena Ameyaw, Pascal Agbadi, Ebenezer Agbaglo, Justice Kanor Tetteh, Jerry John Nutor. "Inequalities in Stunting Prevalence Among Children Under Age 5 in Ghana Between 1998 and 2014", Research Square, 2021 Publication	<1 %
38	Submitted to King Mongkut's University of Technology Thonburi Student Paper	<1 %
39	dspace.sliit.lk Internet Source	<1 %
40	www.cambridge.org Internet Source	<1 %
41	Andrew Banda, Elizabeth T. Nyirenda, Chabila C. Mapoma, Bwalya B. Bwalya, Nkuye Moyo. "Influence of Infant and Young Child Feeding	<1 %

	Practices on Stunting in Children Aged 6 – 23 Months in Zambia ", Research Square, 2021 Publication	
42	Festo Kasmir Shayo, Keiko Nakamura, Saber Al-Sobaihi, Kaoruko Seino. "Is the source of domestic water associated with the risk of malaria infection? Spatial variability and a mixed-effects multilevel analysis", International Journal of Infectious Diseases, 2021 Publication	<1 %
43	Submitted to University of Venda Student Paper	<1 %
44	www.iiste.org Internet Source	<1 %
45	www.oak.go.kr Internet Source	<1 %
46	"Childhood stunting: a global perspective : Childhood stunting: a global perspective", Maternal and Child Nutrition, 2016. Publication	<1 %
47	Submitted to Chester College of Higher Education Student Paper	<1 %
48	Submitted to South Bank University Student Paper	<1 %

49	Submitted to University Der Es Salaam Student Paper	<1 %
50	Submitted to University of Strathclyde Student Paper	<1 %
51	j.mecs-press.net Internet Source	<1 %
52	peerj.com Internet Source	<1 %
53	rehab.jmir.org Internet Source	<1 %
54	Jennifer Batamuliza, Damien Hanyurwimfura. "Identity based encryption with equality test", Information Security Journal: A Global Perspective, 2020 Publication	<1 %
55	Md. Ashfikur Rahman, Md. Sazedur Rahman, Shakif Mohammad Shakur, Md. Hasan Howlader et al. "Risk factors of chronic childhood malnutrition: an analysis of the Bangladesh demographic and health survey 2014 data", Journal of Public Health, 2020 Publication	<1 %
56	Submitted to University of Cambridge Student Paper	<1 %
57	ro.uow.edu.au Internet Source	<1 %

58	scholarworks.waldenu.edu Internet Source	<1 %
59	www.hackster.io Internet Source	<1 %
60	www.scribd.com Internet Source	<1 %
61	ihi.eprints.org Internet Source	<1 %
62	ir.uiowa.edu Internet Source	<1 %
63	nutritionj.biomedcentral.com Internet Source	<1 %
64	www.aessweb.com Internet Source	<1 %
65	www.i-scholar.in Internet Source	<1 %
66	Betregiorgis Zegeye, Gebretsadik Shibre, Dina Idriss-Wheeler, Sanni Yaya. "Trends in inequalities in childhood stunting in Ethiopia from 2000 to 2016: a cross sectional study", Journal of Public Health, 2020 Publication	<1 %
67	Lecture Notes in Computer Science, 2015. Publication	<1 %

68	Sandra Nkurunziza, Bruno Meessen, Jean-Pierre Van geertruyden, Catherine Korachais. "Determinants of stunting and severe stunting among Burundian children aged 6-23 months: evidence from a national cross-sectional household survey, 2014", BMC Pediatrics, 2017 Publication	<1 %
69	Yunhee Kang, Jihye Kim. "Risk factors for undernutrition among children 0–59 months of age in Myanmar", Maternal & Child Nutrition, 2019 Publication	<1 %
70	bmcpregnancychildbirth.biomedcentral.com Internet Source	<1 %
71	journalofethics.ama-assn.org Internet Source	<1 %
72	www.jstage.jst.go.jp Internet Source	<1 %
73	www.scielo.br Internet Source	<1 %
74	"Machine Learning, Image Processing, Network Security and Data Sciences", Springer Science and Business Media LLC, 2020 Publication	<1 %

75	Dinabandhu Mondal, Pintu Paul. "Association between Intimate Partner Violence and Child Nutrition in India: Findings from Recent National Family Health Survey", Children and Youth Services Review, 2020 Publication	<1 %
76	Submitted to International Health Sciences University Student Paper	<1 %
77	Lauriane Nyiraneza, Rex Wong, Olushayo Olu, Marie-Rosette Nahimana, Eliud Birachi, August Musoni, Chantal Gegout, Jenae Logan. "Risk Factors Associated With Childhood Stunting in Rwanda: A Secondary Analysis of the 2014 Nutrition, Markets and Gender (NMG) Survey", Journal of Management and Strategy, 2019 Publication	<1 %
78	Vivian C. Pun, Russell Dowling, Sumi Mehta. "Ambient and household air pollution on early-life determinants of stunting—a systematic review and meta-analysis", Environmental Science and Pollution Research, 2021 Publication	<1 %
79	freeproject.com.ng Internet Source	<1 %

80	hdl.handle.net Internet Source	<1 %
81	journals.plos.org Internet Source	<1 %
82	mchandaids.org Internet Source	<1 %
83	researchonline.lshtm.ac.uk Internet Source	<1 %
84	sersc.org Internet Source	<1 %
85	tampub.uta.fi Internet Source	<1 %
86	www.fao.org Internet Source	<1 %
87	www.research.manchester.ac.uk Internet Source	<1 %
88	www.unicef.org Internet Source	<1 %
89	Blessing J. Akombi, Kingsley E. Agho, Andre M. Renzaho, John J. Hall, Dafna R. Merom. "Trends in socioeconomic inequalities in child undernutrition: Evidence from Nigeria Demographic and Health Survey (2003 – 2013)", PLOS ONE, 2019 Publication	<1 %

90	Fabiola V. Moshi, Christopher H. Mbotwa. "Determinants for choice of home birth over health facility birth among women of reproductive age in Tanzania: an analysis of data from the 2015-16 Tanzania demographic and health survey and malaria indicator survey", BMC Pregnancy and Childbirth, 2020 Publication	<1 %
91	Gitanjali Sathiadas, A Annieston, V Arunath, T Kasthuri, V P Wickramasinghe. "Nutritional status of school children living in Northern part of Sri Lanka", Research Square, 2020 Publication	<1 %
92	www.repository.cam.ac.uk Internet Source	<1 %
93	"Advanced Computing", Springer Science and Business Media LLC, 2021 Publication	<1 %
94	"Agriculture, Food and Nutrition Security", Springer Science and Business Media LLC, 2021 Publication	<1 %
95	Abebe Ayelign, Taddese Zerfu. "Household, dietary and healthcare factors predicting childhood stunting in Ethiopia", Heliyon, 2021 Publication	<1 %

96	Mohammad Redwanul Islam, Md. Moinuddin, Ayeda Ahmed, Syed Moshfiqur Rahman. "Association of sickle cell disease with nutritional status among under-five children and mediating role of hemoglobin level: secondary analysis of data from 2018 Nigeria Demographic and Health Survey", Research Square, 2020 Publication	<1 %
97	Mulat Mossie Menalu, Alebachew Demelash Bayleyegn, Michael Amera Tizazu, Nakachew Sewnet Amare. "Assessment of Prevalence and Factors Associated with Malnutrition Among Under-Five Children in Debre Berhan Town, Ethiopia", International Journal of General Medicine, 2021 Publication	<1 %
98	Tri D. B. Le, Man M. Ngo, Long K. Tran, Vu N. Duong. "Chapter 41 Applying LSTM to Predict Firm Performance Based on Annual Reports: An Empirical Study from the Vietnam Stock Market", Springer Science and Business Media LLC, 2021 Publication	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		