STUNTING FACTOR ANALYSIS USING MACHINE LEARNING FOR UNDER AGED CHILDREN IN THE DEVELOPING COUNTRIES

BY

PRITOM SAHA
ID: 181-15-1807

MOHAMMAD SHAFIQUL ISLAM
ID: 181-15-1942

SUNIT CORNELEOUS SARKER
181-15-10591

The report has been submitted in partial fulfillment of the necessary documents for the purpose of Degree of Bachelor of Science in Computer Science and Engineering

Supervised By
Al Amin Biswas
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Co-Supervised By
Md. Mahfujur Rahman
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2022

i

APPROVAL

Project of ours is titled as "**Stunting Factor Analysis using Machine Learning for Under Aged Children in the Developing Countries**", which was submitted by Pritom Saha, Mohammad Shafiqul Islam, and Sunit Corneleous Sarker, intended to the Department of CSE, Daffodil International University, that has been considered quite satisfactory for the cause of proportional fulfillment of the required elements for the degree of B.Sc. in Computer Science & Engineering also approved by its style and contextual appropriateness. The research papers presentation has been taken place on 13th January, 2022.

<u>BOARD OF EXAMINERS</u>

Tania Khatun   **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Mahfujur Rahman   **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Swakkhar Shatabda   **External Examiner**
Associate Professor,
Department of Computer Science and Engineering (CSE)
United International University (UIU), Dhaka, Bangladesh

DECLARATION

We have studied the data multiple times over and over again by doing various extensive research and thus from our valuable findings, it can be stated that, the project has been brought to completion by us under the supervision of Al Amin Biswas, Lecturer (Senior Scale), Department of CSE, Daffodil International University. We also assure as much as the certainty of neither this project nor any portion of this project being submitted anywhere else for the intention of getting any award of any degree/diploma.
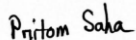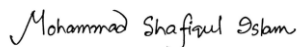
Supervised by:

_____

Al Amin Biswas
Lecturer (Senior Scale)
Department of CSE
Daffodil International University
Co-Supervised by:

_____

Md. Mahfujur Rahman
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Submitted by:

_____

Pritom Saha
ID: 181-15-1807
Department of CSE
Daffodil International University

_____

Mohammad Shafiqul Islam
ID: 181-15-1942
Department of CSE
Daffodil International University

Sunit Corneleous Sarker
ID: 181-15-1942
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

At the very beginning we would like to convey our gratitude towards our almighty God for the divine blessing he bestowed upon us and makes it plausible for us to finish the project of the final year with ease and success. It has been a complete experience.

We are also indebted to our respected course teacher **Mr. Al Amin Biswas, Lecturer (Senior Scale)**, Department of CSE, Daffodil International University, Dhaka for helping us with the aspects of working procedure and always keeping us on track. Expertise & deep knowledge of the supervisor in this particular field of "Machine Learning" to proceed on with this project. His patience, enlightening guidance, continuous encouragement and constant care with diplomatic supervisions were a such an experience to have while doing all these works and it really helped us a lot to have a better understanding of how these works are done. Reading so many inferior drafts & making corrections in them at all stages have made it possible to accomplish what we were aiming for i.e., the fruitful outcome of this paper.

We would also like to be thankful to the whole course mate team we have back in Daffodil International University, who helped us with their existing knowledge by participating in these helpful discussions while completing the course works.

And thus finally, our immense gratefulness is due in the acknowledgement of the unending and continuous help, care and protection provided through the hard work and virtue of our parents who are always patient.

ABSTRACT

Malnutrition and stunting which comes from it can be put in the list of the top problems of any developing countries. A country like Bangladesh is no different. The principal contribution of this study is to predict the Stunting status of the children and also to search for the related affecting factors which affect the nutrition status. In this research, data has been collected from Bangladesh Demographic Health Survey (BDHS)'2017-18. Nutrition status correlated with the child's age, mother's education, father's education, father's employment, family wealth index, currently breastfeeding, place of residence and division. The differential impact of some sectors like demographical and socioeconomical, environment plus health-affecting determinants on the nutritional rank within the population of under-five children in Bangladesh has been taken into account. To measure the child nutritional condition of under-five children among various methods, this Z-score method is one which we have used in this paper to find the statistics of malnutrition in Bangladesh. Methods provided by WHO (World Health Organization) was followed to find out the necessary outcomes. Here, Chi-square statistics algorithm has been applied to find out the factors which are most responsible for stunting by ranking features and then applied machine learning algorithms for a prediction model. Two algorithms have been used here and based on the performance results and parameter we find Logistic Regression gave most accuracy out of them all. It's found that based on all features the accuracy is 77.00% and based on top 5 features the accuracy is 63.00% and based on top 7 features the accuracy is 77.00% which is better than the outcomes of Random Forest Algorithm. The study is suggested to focus upon the factors which are responsible for malnutrition and it would ensure healthy nation.

Keywords: Malnutrition, BDHS, Z-score, Machine Learning, Random Forest, Prediction

TABLE OF CONTENTS

CONTENTS	PAGE

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

Nutrition is basically the body's consumption of food, looked from the perspective of its relation to the body's optimum dietary demands. Nutrition is needed for the healthy growth, intellectual development and well-being of the child. A lack of, excess, or imbalance in an individual's energy and/or nutrient consumption is referred to as malnutrition.[1] Bangladesh has developed quite significantly in various signs and sight of human development. Multi-level progress was also been made in the nutrition and health department. But the nutrition situation of women and children still can't be announced ideal. Stunting, wasting, and being underweight are examples of undernutrition, whereas overweight/obesity is linked to a number of non-communicable disorders (diabetes, cancer, stroke, and heart disease) [1-3] A survey which came out recently survey states that for overcoming the malnutrition in children, Bangladesh still has got a long way left to go. Malnutrition affects more than half of the population in Bangladesh.[6]. It's not necessary to state that this group of children will surely induce negative effects caused by malnutrition until the age when they turn 5 years old. The cause of real attention is, here 15% of the birth weight of the child is found to be less than optimum and albeit one-third of the child deaths are caused by fatal form malnutrition. Malnutrition implies a particular kind of nutritional deficiency or anemia. When the mother has got a unhealthy diet ridden life and got malnutrition and anemia present in her system, it has quite a terrible impact on her child. Her unborn children are also infected with this disease. So, malnutrition levels of mother are also a point for concern. The main & direct causes of malnutrition of mothers are the lack of necessary nutritious food and disease. Children are affected by malnutrition due to ill health and malnutrition of mothers. This is a big reason for the baby's weight being less than usual during their birth. Mothers suffering from malnutrition have a shortage of breast milk, so the physical and psychological growth of the newborn is interrupted. Along with the weight of the

children, stunting problems related to mother's malnutrition. On the other hand, Child marriage is responsible for one of the biggest causes of malnutrition. Child marriageable parents usually cannot take nutritious food. As a result, it has an impact on their child and the child suffers it for the rest of his life. Mothers' ignorance and indifference are also one of the reasons behind the nutritional condition of children. After birth the child is completely dependent on its mother. When the mother is not aware of what to feed her baby, how to feed when to provide supplementary food, child will be victim of malnutrition. As the family size is large, it is not possible for family members to provide good education and provide balanced food according to the needs of the body, so they are also victims of malnutrition. So, Lack of child nutrition due to lack of balanced diet for children, lack of knowledge about parental nutrition, family illiteracy and poverty, food disorders due to chronic diseases of children, diseases of diarrhea for the child long, diseases of digestion with baby lever, food processing and cooking in unhealthy food, baby worm disease, lack of adequate vitamins and minerals for food are also the reason for malnutrition.

1.2 Motivation

From World Health Organization information, a significant number of children in Bangladesh still suffer from malnutrition. Their physical body heights are found to be less than the average in age ratio. It is not quite possible for their parents to find out exactly if they will have negative and harsh side effects of malnutrition until the time, they turn 5 years old. The fact that 15 children out of every hundred have a birth weight that is less than normal, and that one-third of child deaths are attributable to deadly nutrition, is cause for alarm. The most likely cause of malnutrition is the lack of knowledge about parental nutrition and family members illiteracy and poverty. Some of the people leads a life way below the poverty line. For the inflation in food prices, a great number of populations aren't capable of affording the necessary nutritious food. The lack of proper nutrition is a major cause of malnutrition. Lacking of knowledge on nutrition, people do not eat the right food, though they can afford it. Besides, the nutritional

policies and weaknesses in the strategy, our motivation is to look for the major factors of malnutrition by finding them out and trying to solve them so that not a single child suffers from malnutrition not in Bangladesh as well as all over the world.

1.3 Scope and Challenges

Malnutrition is the major problem of any developing countries. The study utilizes nationwide data from the BDHS (Bangladesh Malnutrition is the major problem of any developing countries. The study utilizes nationwide data from the BDHS,2017-2018. The significant contributing factors of growing up of the malnutrition were found to be child's age, the mother's education, father's education, father's line of work, place of delivery and division, family size, is currently breastfeeding or not, type of residence and wealth index combined. Nutritional condition of children of workers is determined together with their social and economic background, the nutritional condition of the population is compared with respect to the data obtained from a national survey. Although those children are breastfed at a higher rate when they're compared with others but the educational status of mother and her measured body mass index are found to be lower than the Bengali citizens. The nutritional condition is worse than that of the national level. Intimate partner violence (IPV) in Bangladesh is also one of the main reasons for malnutrition. The nutritional condition of under-five children is a very alarming warning indication for this kind of countries spread over different sectors as such as health and economics. The impact of several sectors, such as demographic, socioeconomic, environmental, and health-related factors, on the nutritional status of Bangladesh's under-five children is undeniably huge. To measure the child nutritional condition of under-five children using various methods, Z-score method (Z-score is the measured value of the distribution of target reference population), Chi-square Algorithms, determining factors which are related with the responding variables for finding out the level of stunting and ML algorithms to predict the stunting of children who might have it in the future. The pervasiveness of malnutrition is assessed by three standard anthropometric indicators: underweight (children having less weight than normal), stunting (children having less

©Daffodil International University

height than normal) and wasting (weight not according to the height) following the WHO guidelines and cut-off points. Improving the general health and nutritional condition of Bangladesh's children requires getting people out of poverty and making services more affordable and accessible to the poor and masses. Particularly for this research the work will be done mainly on Stunting. These levels paly a huge role in the future buildup of a generation and it can leave severe impact on the sustainability of the population in near future. The nutritional rank of under-five children is a highly alarming sign of countries in different sectors like as health, economic.

From the previous paper, no clear ideas were found about the possible factors where the new data was taken into account. The reason why it was done lies within the outcome which was found from such a great demographic. Data collection is a major problem to develop any data mining project as they are very scarce and the more under-developed a country is, the harder it is to collect data completely and correctly. Often times it is the data that we lack. Because of applying data mining, thousands of data will be needed. But in this study data has collected from BDHS 2017-18 with their permission. So, collecting the data is no problem here at all but the challenge was to filter, cleaning, merging and finding the z-score for huge data sets and the most importantly the labeling for the huge datasets and applying the algorithm to the large datasets.

1.4 Research Questions
- What is the latest up to-date Stunting status?
- What are the associated factors with Stunting status?
- Which factors affect the children most?

1.5 Research Objective
- To find out the Stunting status.
- To identify the associated factors affecting malnutrition status.
- Making awareness to reduce the percentage of Child deaths in every year.
- Giving message with the effect of lacking knowledge on nutrition of peoples.

o To promote mothers' and caregivers' understanding of the relevance of complementary feeding preparation, frequency, amount, and types of feeds in maintaining their children's nutritional status.

1.6 Report Layout

- Background study explains the related works done on the same sector with data from previous years and different regions determining the stunting factors and its effects on children.
- Research Methodology describes the working procedure and necessary calculations and tabulations.
- Experimental results are the representation of the outcome that was found and discussions on the outcome.
- Conclusion holds statements on the outcome and future works that can be done on the particular area.
- Reference holds the citations made while doing the paper.

CHAPTER 2

Background Study

2.1 Related Works

Severe acute malnutrition afflicted 450,000 children, whereas mild acute malnutrition afflicted roughly 2 million.[7] According to the World Health Organization, 36% children among the whole population around here living inside Bangladesh do actually suffer for the reason malnutrition. Their calculated heights are found to be less than the average ideal in the age ratio. Over the previous few decades, children's nutritional status has gradually improved [8]. U5 stunted children accounted for 51% of all children in 2004, 43% in 2007, and 41% in 2011, whereas underweight children accounted for 43% of all children in 2004, 41% in 2007, and 36% in 2011. Similarly, in 2004, 15% of U5 children were squandered, followed by 17% in 2017, and 16% in 2011 [8]. In 2014, these percentages had dropped to 36% stunted, 33% underweight, and 14% overweight [9]. The height-for-age (stunted), weight-for-height (wasted), and weight-for-age (underweight) measurements were commonly used to determine if a kid was malnourished [9, 10]. Some of the population live below the poverty line. Malnutrition [12–14], anemia [15–17], diabetes [18], low birth weight [19–22], child mortality [23–25], and other sectors have also benefited from ML research. ML has also been used to predict underweight [12–14], stunted, and wasted [13, 14]. They did not tune the hyper-parameters of ML algorithms in earlier experiments. As a result, the precision of their machine learning system was not adequate. The goal of this study is to develop a strategy that combines a logistic regression (LR)-based risk factor identification method with machine learning (ML) classifiers to more accurately detect malnutrition. To back up this claim, we examined the accuracy and area under random forest (RF), and linear regression (LR) models for predicting malnutrition (AUC). Due to the increase in food prices, this huge number of people cannot afford the necessary nutritious food. The lack of proper nutrition of them is a major cause of malnutrition. Lacking of knowledge on

nutrition, People do not eat right food, though they can afford it. Besides, the nutritional policies and weaknesses in the strategy. More population is one of the reasons for malnutrition. The Bangladesh Demographic & Health Survey (2017-18) is the 8th DHS statement carried in BD, following all these, achieved in 1993 to 1994, 1996 to 1997, 1999 to 2000, 2003 to 04,2007 to 08, 2010 to 2011 and 2014. From looking up the earlier surveys, the final objective of the 2017-18 BDHS is to give updated latest erudition on childhood and fertility death levels; fertility choices; preparedness with knowledge, their allowance and proper use of ideal and guided family planning steps; child and maternal health; awareness and understanding towards Human Immune Virus/Acquired Immune Deficiency Syndrome and other sexually transferred infections (STI); and society-level collection of data on support and abundance of health and family plan assistance. All the married females aging from 15 to 49 those who were habitually persons from elected families and who did spend the latter half of a day before survey within the marked households were qualified to be questioned in the survey. The study was intended to give characteristic findings of this country as an integral part, for urban & rural areas separately, for each of 7 managerial groups. Alongside, the found cause of severe malnutrition in the country is the evident poverty among the people of the country who are mostly living under the ideal poverty line and the poisoning present in food caused by the huge demand of this densely populated country. There were also 47 million wasted U5 children, 14.3 million severely wasted children, 144 million stunted children, and 38.3 million overweight/obese children, according to the report. Malnutrition kills 2.6 million children worldwide each year, with undernutrition accounting for 45 percent of U5 mortality [4,5]. The association between the imbalance of wealth of the household with constant childhood under-nutrition in Bangladesh is found to be present. Social and economic Factors affecting the Nutritional rank of Under-Five children in Bangladesh has a exposition of several levels. Nutritional condition of under-5 children in Bangladesh. Nutritional rank of under-five children in Bangladesh: a multilevel investigation.

CHAPTER 3

Research Methodology

3.1 Introduction

In this research paper, relative data has been collected from the BDHS, 2017-18 which is the 8th DHS initiated in Bangladesh, followed by these attained in 1993 to 1994, 1996 to 1997, 1999 to 2000, 2003 to 04,2007 to 08, 2010 to 2011 & 2014. This survey has made by depending on both rural and urban regions of Bangladesh also keeping up for several of the seven supervisory groups and considering the country as an integral part. The 2017-18 BDHS procedure has been created to provide information that will be able to fulfill the demand of health, population, and nutrition sector development programs (HPNSDP). Experiment on malnutrition by definition implies on 2 signs Severely Stunted (z-score is less than -3) and Moderately Stunted (z-score is less than -3). But Two groups of children are differed by 'suffering from severe stunting' and 'suffering from moderate stunting', for each of the cornerstone of the provided rules in the World Health Organization and the national report of Bangladesh. Finding out the results it can used to make people of the country aware of the problematic future that can emerge upon a nation. The strength of the nation lies within its future generation and what they're capable of doing. Thus, finding out such sensitive information can leave and huge impact on a developing country.

Measurement of stunting included having all the data portraying perfect values otherwise the outcomes accuracy might not be that much high. Here, level of stunting weather it is severely stunted or moderately stunted were practiced to examine the magnitude of stunting among under-five children in Bangladesh using the most recent BDHS data which is of 2017-18. A double staged laminated sample of those households has been used for doing the survey. The method of anthropometric for estimating the nutritional condition includes three publicly used pointers to judge the germination of children: Severely Stunted (z-score is less than -3) and Moderately Stunted (z-score is less than -3)

. Children can be inspected of having stunted if (s) he is, in terms of height-for-age, more than two standard deviations below thus the number of children who has got the height for age z-score below -3.0 standard deviations (SD) below the mean on World Health Organizations Child Growth Standards (height/age standard deviation (new who) < -300) & serial of children who has got height for age z-score below -2.0 standard deviations (SD) below the mean on the WHO Child Growth Standards (height/age standard deviation (new who)< -200). In this paper, the Stunting status of under-five (5) children on the basis of Z-scores were found out for every attribute and thus the final outputs were compared with the previous year's scores. According to WHO given process, Mean Z-score for height-for-age: Z-scores Sum of children with a non-flagged height with relation to age score (Σ height/age standard deviation (new who)/100). Z-score lesser ($-$2) in any of these 2 criteria makes it fall under the category of Stunting. The data were examined by using Stata on PC platform.
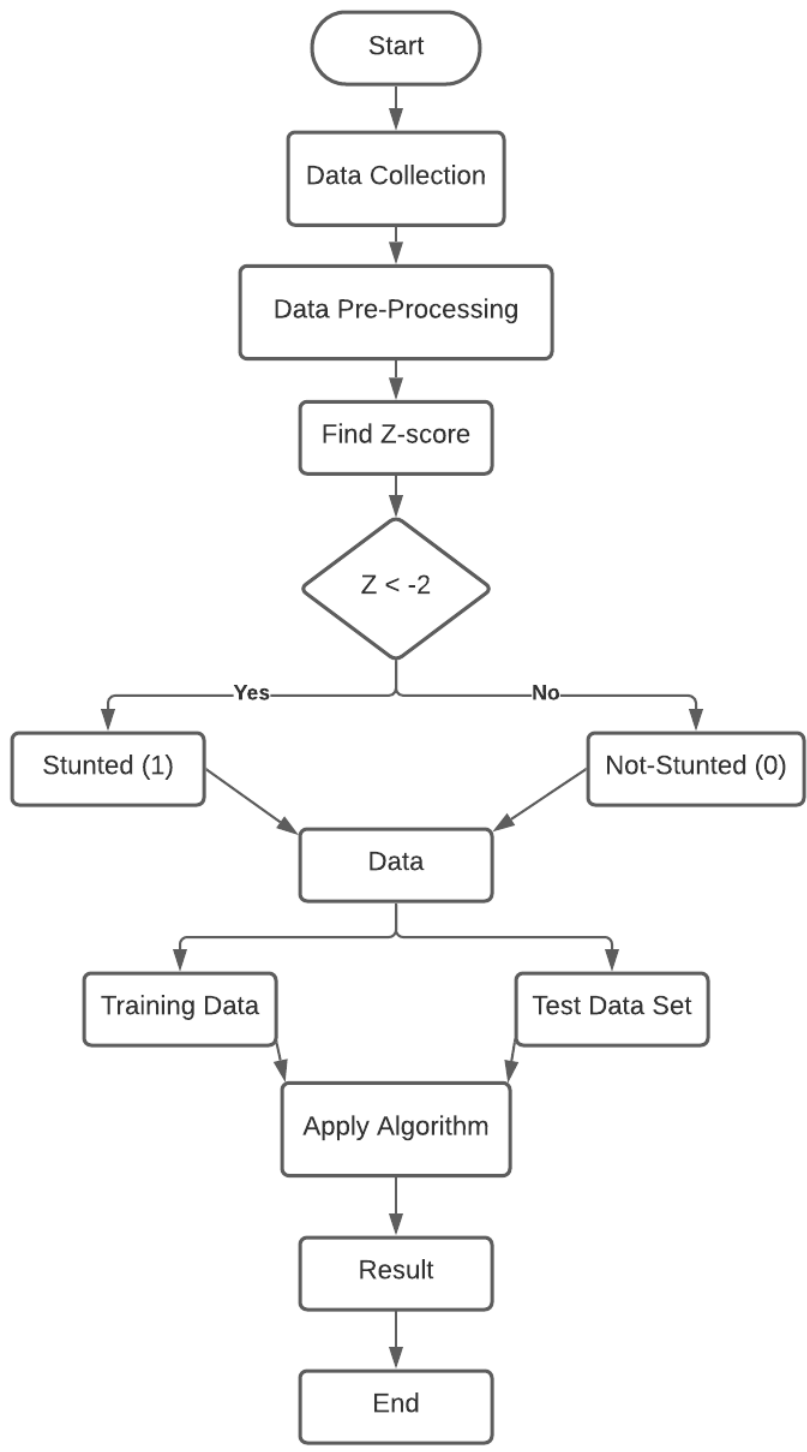
Figure 1: Working Procedure for Stunting Prediction

3.2 Research Subject and Instrumentation

The rising size of data in advanced technology demands for even more complicated and refined instruments. Although progressions in the technology of Data mining have invented large data acquisition much cautiously is all the time obliging and a consistent need is there for new techniques and instruments that might be able to help in transforming this data into valuable data points and understanding. Concept & procedures continues the tradition of providing users added the realization & real-life application of various theories & practices of finding out patterns hidden in very large data sets and making new discoveries. Our research topic is "Stunting Factor Analysis using machine learning among children under-five in Bangladesh" and in this paper, we are using Stata, Google Colab, Jupyter Notebook and Python.

3.3 Data source

The rising size of data in advanced technology demands for even more complicated and refined instruments. The data was collected by individuals through extensive field work assigned by the DGHS. They were thoroughly checked and taken with caution and consent from the person they were collecting the data from. Although progressions in the technology of Data mining have invented large data acquisition much cautiously is all the time obliging and a consistent need is there for new techniques and instruments that might be able to help in transforming this data into valuable data points and understanding. Concept & procedures continues the tradition of providing users added the realization & real-life application of various theories & practices of finding out patterns hidden in very large data sets and making new discoveries. Our research topic is "Stunting Factor Analysis using machine learning among children under-five in Bangladesh" and in this paper, we are using Stata, Google Colab, Jupyter Notebook and Python.

3.4 Data Collection Procedure

Data has been collected from the Demographic Health Survey (BDHS), 2017-18 archive. We are working on Malnutrition of children who are under 5, our main key variables are Height and age, and there are thousands of data variables present in the dataset. From the collective data set of thousands of people there were a lot of arrangements. There is a huge amount of data in the dataset with the division of IR, KR, BR, and PR. The ".dta" format data was used in data analysis in Stata format to find out the statistics in percentage to get how many children are suffering from stunting. That's why it was needed for the Height/Age standard deviation (new WHO) to calculate the Z-Score. Then the data set is prepared for further work. The data was collected with permission and caution.

3.5 Data Preparation

Collection of data was easier than the preparation for applying algorithm. Data was readable in Stata tool. IR refers to Women, KR refers to Children, BR refers to birth-record and PR refers to Parents. We used KR file to analyze our data. A huge number of data was missing in the data set, that's why missing data was cleaned and duplicate data was removed by using commands in the Stata app. After deleting the null and duplicate values 4583 data in total were found. Class level is unbalanced with 0(3175) and 1 (1408). Under sampling has been used to balance the class data. In this way the most accurate output was found. After balancing the class attribute both 0(1408) and 1(1408) was the output.

3.6 Data Splitting

Sklearn library was used to split the train and test data. 30% data was used for testing and 70% data for training.

3.7 Machine Learning Algorithms

Logistic regression

LR is a widely used statistical model for classification problems. It applies the maximum likelihood estimation procedure to estimate the parameter of interest.

If there are n numbers of features denoted by $A_i = \left(A_{1,...,A_n}\right)'$ and $\beta = \left(\beta_{1,...,\beta_n}\right)'$ express parameters of the model, then logistic regression model is like: $log\left[\frac{\pi}{1-\pi}\right] = \alpha_0 + \beta'A_i$, $i = 1,2,...n$; where π represents the probability of occurrence and event $(1 - \pi)$ represents the probability of not occurrence an event.

Random forest

RF is a classification strategy that relies on a group of tree-structured classifiers "developing." To categorize another person, attributes of the person are used to classify them, and every classification tree in the forest is employed. The grown trees are arranged in a random order, with each tree assigning a classification (or "vote") to a class label. The result is based on a majority of votes cast over the total number of trees in the forest. [26]
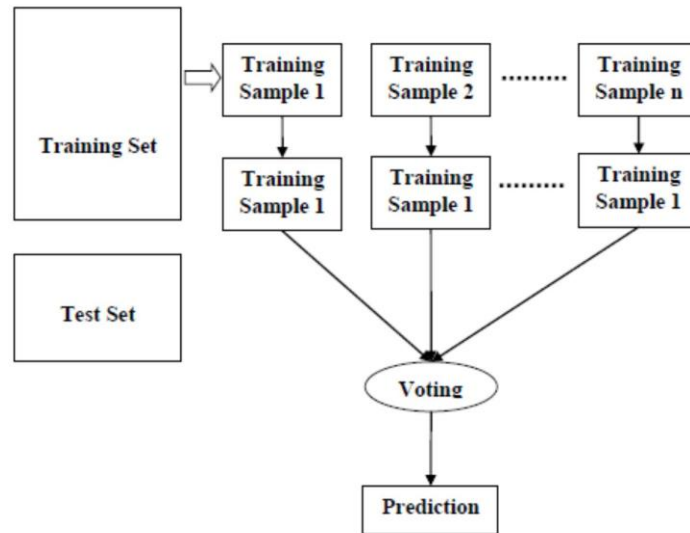
Figure 2: Working Procedure of Random Forest Classifier

CHAPTER 4

Experimental Results & Discussion

4.1 Experiment Results

Applied algorithms and processes:

- **Feature Ranking:**
    - Both our input and output variable are categorical. That's why we used Chi-Square algorithm.
    - After applying chi-square, we got the feature ranking (by top scores) below:

Table 1: All Features by Ranking

	Factors	Scores
1	ChildAge	865.155817
2	breastfeeding	105.577488
3	PartnerEdu	63.418384
4	drinkingwater	50.331352
5	bmiCAT	21.105837

6	edu	19.705338
7	division	14.008973
8	childistwin	10.411596
9	currentlyworking	6.322889
10	toilet	3.290326
11	residence	3.050905

- **Logistic Regression:**

Table 2: Logistic Regression performance evaluation metrics

Features	Accuracy	Precision	Recall	F1-score
All Features	77.00%	83.00%	67.00%	74.00%
Top 5	63.00%	65.00%	60.00%	62.00%
Top 7	77.00%	84.00%	68.00%	75.00%

Table 3: Confusion Matrix of Logistic Regression based on All features

Logistic Regression		Predicted Values	
		Positive	**Negative**
Actual Value	**Positive**	288	141
	Negative	57	359

Table 4 : Confusion Matrix of Logistic Regression based on Top 5 Features

Logistic Regression		Predicted Values	
		Positive	Negative
Actual Value	Positive	258	171
	Negative	139	277

Table 5: Confusion Matrix of Logistic Regression based on Top 7 Features

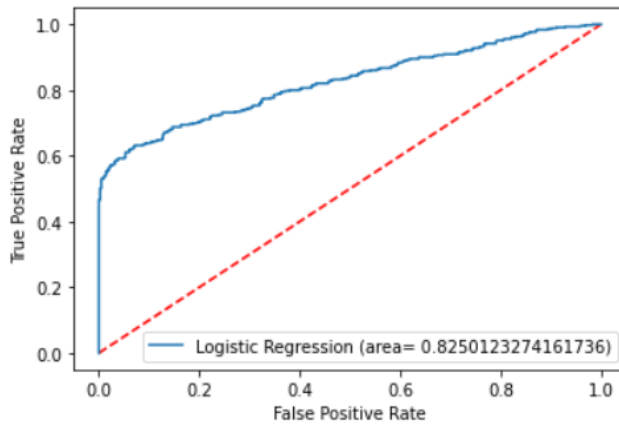| Logistic Regression		Predicted Values	
		Positive	Negative
Actual Value	Positive	293	136
	Negative	55	361

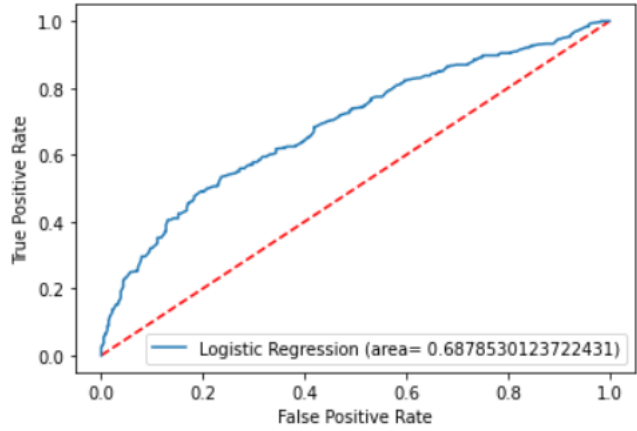Figure 3: ROC Curve of Random Forest based on All Features

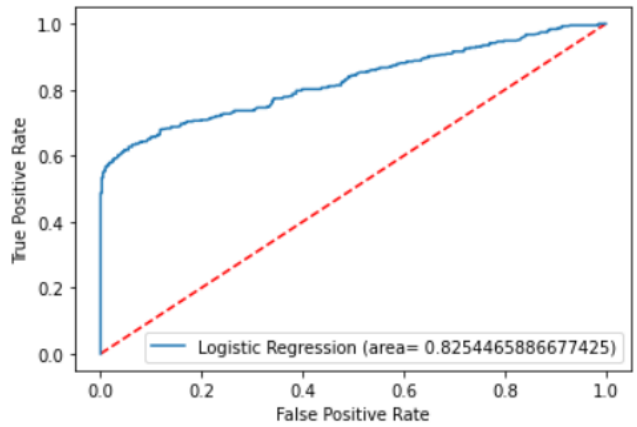Figure 4: ROC Curve of Random Forest based on Top 5 Features

Figure 5: ROC Curve of Random Forest based on Top 7 Features

- **Random Forest:**

Table 6: Random Forest performance evaluation metrics

Features	Accuracy	Precision	Recall	F1-score
All Features	75.00%	79.00%	69.00%	74.00%
Top 5	59.00%	60.00%	59.00%	59.00%
Top 7	78.00%	83.00%	71.00%	77.00%

Table 7: Confusion Matrix of Random Forest based on All features

Random Forest		Predicted Values	
		Positive	Negative
Actual	Positive	295	134
	Negative	77	339

Table 8: Confusion Matrix of Random Forest based on Top 5 Features

Random Forest		Predicted Values	
		Positive	Negative
Actual	Positive	252	177
	Negative	166	250

Table 9: Confusion Matrix of Random Forest based on Top 7 Features

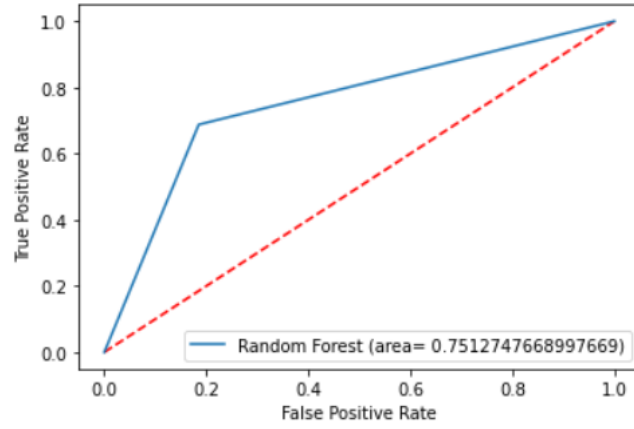| Random Forest		Predicted Values	
		Positive	Negative
Actual	Positive	308	121
	Negative	63	353

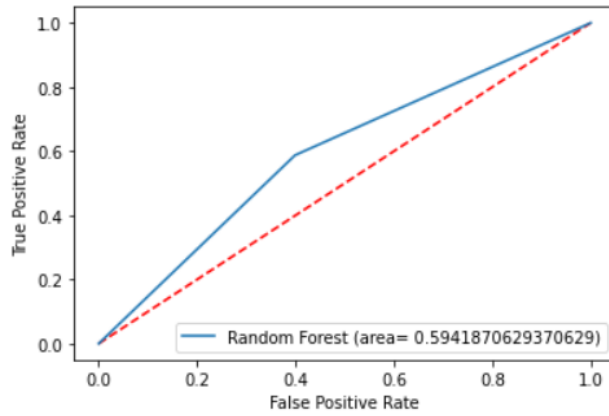Figure 6: ROC Curve of Random Forest based on top of All Features

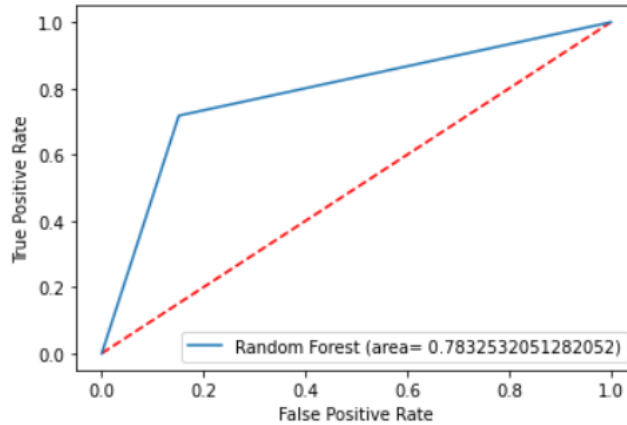Figure 7: ROC Curve of Random Forest based on Top 5 Features

Figure 8: ROC Curve of Random Forest based on Top 7 Features

For Logistic Regression, in all features an accuracy of 0.77 and AUC score 0.83 was found. For top 5 features, an accuracy of 0.63 and AUC score 0.69 was found. For top 7 features, an accuracy of 0.77 and AUC score 0.83 was found. For Random Forest, in all features an accuracy of 0.75 and AUC score 0.75 was found. For top 5 features, an accuracy of 0.59 and AUC score 0.59 was found. For top 7 features, an accuracy of 0.78 and AUC score 0.78 was found. Here it's clear that, while using Logistic Regression, the model where all attributes has been taken into consideration and where 7 attributes has been taken into consideration shows most accuracy. On the other hand, while using Random Forest, the model where 7 attributes has been taken into consideration shows most accuracy. From Features Ranking, it's been found out that Childs Age, Breast Feeding, Partner(parents) Education, Drinking Water, BmiCAT, Education and Division affect the occurrence of stunting the most. So, child's age and how long it's breastfed and how educated its parents are leaves the most impact on the child's health. Also, quality of the water they drink, parents' weight and health, their education level and the division they live in also impacts a child's health.

CHAPTER 5

Conclusions & Future work

5.1 Conclusions

The Stunting of under-five children is a very alarming indication in the country on various sectors like health and economics. This study reveals that several demographics, social, environmental, and health-related factors have a differential impact on nutritional status among Bangladeshi children under the age of five whose parents are behind in the education sector. The illiteracy plays a huge role for the unawareness of the parents thus causing bad maintenance of the health of the child. It concerns the physical mass and legitimate improvement of children. Malnutrition is measured using three traditional anthropometric indicators: underweight, stunting, and wasting, all of which are based on WHO criteria and cut-off points. For all pointers from the source population, the unions of Z-scores for the research youngsters revealed significantly far downwards arrangements. The paper discussed the status of malnutrition of Bangladesh. It investigated all of the major factors which are responsible for child malnutrition like father's occupation, child's age, father's education, mother's education, currently breastfeeding, place of residence and division. Mother's health history is also in concern here. It's a common knowledge that this Malnutrition is unbounded by a family limitation rather it is a problem which also affects people outside the family as well. It is also a major hindrance in the case of nationwide multi-skilled human resources who are the driving force of a nation. This paper provides us with a complete knowledge which expresses all the major factors for causing Stunting. The lack of proper nutrition of them is a major cause of malnutrition. Lacking knowledge of nutrition, People do not eat the right food, though they can afford it. Besides, the nutritional policies and weaknesses in the strategy.so our motivation is pinpointing the major factors of stunting and thus try to solve them so that not a single child suffer from malnutrition not in Bangladesh as well as all over the world. The result from this study ensures that the nutrition percentage of children is raising day by day. It also confirmed all of the major factors which play the

vital role for child malnutrition. To reduce malnutrition problem, the government should take steps to work with nongovernment organizations. Community people can work together.

5.2 Recommendations

To overcome malnutrition, the government will have to do more work in different fields including education, counseling, and training, awareness building, and increasing income. For this reason, the role of hospitals and media is also significant. The government needs to take proper steps to improve Water and sanitation system in unhealthy and rural areas. Many of us have a misconception about nutrition. Because many people think of eating stomach as the main. But there are different nutrients in different foods. Along with rice, the number of pulses, fish, meat vegetables, eggs, milk, and fruits are very important for the body. Nutritional needs are needed for all. But many people the necessary nutritious food. People need extensive awareness about the issue of poverty.

5.3 Implication for Further Study

Though, in recent year Bangladesh is more concerned to reduce rate of prevailing malnutrition. In past study in malnutrition, parents smoking habit was not in concern. Children form smoker parents either father or mother can be affected by the health condition. Parents living environment and their food habit may affect their child health. On the other hand, parent's physical exercise is good for their unborn child health which my contribution to improving childhood health. All of these factors are important to know for data analysis to find the main reason for malnutrition, which may also suggest the better way of good child health. This side may take in the study. Future Work: There are several adaptations, experiments, research, exploration, and tests have been done and some workers have been left for the future due to lack of time. Some algorithm we will use in the future. Future work anxiety deeper analysis of special mechanisms, a new scheme to try different tools, processes, methods, or simply curiosity. This thesis has

been mainly focused on the use of Bangladesh Demographic and Health Survey (BDHS), 2017-2018 which is the eighth DHS offered in Bangladesh. This examination has created by depending on both rural and urban regions of the country as well as for each of the seven organizational constituencies and for the country as a whole. But in future, we will try to find out the malnutrition status under 5 in the whole world. For that, we have to go on the different process, and we have to use different module also. It could be interesting to consider the regions in the different model, tools, and algorithms with different importance, depending on the data or their special meaning with respect to the sustaining process.

References:

1. Rahman, A. and Biswas, S.C., 2009. Nutritional status of under-5 children in Bangladesh. South Asian Journal of Population and Health, 2(1), pp.1-11.

2. Rahman, M.S., Rahman, M.A., Maniruzzaman, M. and Howlader, M.H., 2020. Prevalence of undernutrition in Bangladeshi children. Journal of biosocial science, 52(4), pp.596-609.

3. Dat, T.Q., Le Nguyen Huong Giang, N.T., Loan, T. and Van Toan, V., 2018. The prevalence of malnutrition based on anthropometry among primary schoolchildren in Binh Dinh province, Vietnam in 2016. AIMS Public Health, 5(3), p.203.

4. Black, R.E., Allen, L.H. and Bhutta, Z.A., 2008. Caulfi eld LE, De Onis M, Ezzati M, Mathers C, Rivera J, Maternal and child undernutrition study group. Maternal and child undernutrition: global and regional exposures and health consequences. The lancet, 371(9608), pp.243-60.

5. Black, A., Luna, P. and Lund, O., 2005. Wa lker. Sue (coord.).

6. Zarocostas, J., 2006. Over 300 million children chronically malnourished. Bmj, 333(7560), p.166.

7. Fiorentino, M., 2015. Malnutrition in school-aged children and adolescents in Senegal and Cambodia: public health issues and interventions (Doctoral dissertation, Université Montpellier).

8. Bangladesh Demography and Health Survey Report, 2014.

9. World Health Organization, 2006. WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development. World Health Organization.

10. Akombi, B.J., Agho, K.E., Merom, D., Hall, J.J. and Renzaho, A.M., Multilevel analysis of factors associated with wasting and underweight among children under-five years in Nigeria. Nutrients 2017; 9: 44.

11. Paul, G.K., Nesa, M.K., Mondal, S.K., Salan, S.A. and Mim, F.N., 2018. Application of proportional odds model in identifying contributing factors of under-five child malnutrition in Bangladesh: A case study in Tangail district. *Journal of Health Research and Reviews*, *5*(3), p.128

12. Thangamani, D. and Sudha, P., 2014. Identification of malnutrition with use of supervised datamining techniques–decision trees and artificial neural networks. Int J Eng Comput Sci, 3(09).

13. Shahriar, M.M., Iqubal, M.S., Mitra, S. and Das, A.K., 2019, July. A Deep Learning Approach to Predict Malnutrition Status of 0-59 Month's Older Children in Bangladesh. In 2019 IEEE

International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT) (pp. 145-149). IEEE.

14. Talukder, A. and Ahammed, B., 2020. Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. Nutrition, 78, p.110861.

15. Sanap, S.A., Nagori, M. and Kshirsagar, V., 2011, December. Classification of anemia using data mining techniques. In International Conference on Swarm, Evolutionary, and Memetic Computing (pp. 113-121). Springer, Berlin, Heidelberg.

16. Jaiswal, M., Srivastava, A. and Siddiqui, T.J., 2019. Machine learning algorithms for anemia disease prediction. In Recent Trends in Communication, Computing, and Electronics (pp. 463-469). Springer, Singapore.

17. Islam, M.M., Rahman, M.J., Roy, D.C. and Maniruzzaman, M., 2020. Automated detection and classification of diabetes disease based on Bangladesh demographic and health survey data, 2011 using machine learning approach. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 14(3), pp.217-219.

18. lam MM, Rahman MJ, Roy DC, Islam MM, Tawabunnahar M, Ahmed NAMF, et al. Risk factors identification and prediction of anemia among women in Bangladesh using machine learning techniques. Curr Women 's Health Rev. 2021; 17: 1.

19. Eliyati, N., Faruk, A., Kresnawati, E.S. and Arifieni, I., 2019, July. Support vector machines for classification of low birth weight in Indonesia. In Journal of Physics: Conference Series (Vol. 1282, No. 1, p. 012010). IOP Publishing.

20. Senthilkumar D, Paulraj S. Prediction of low-birth-weight infants and its risk factors using data mining techniques. Int Conf on Indus Eng Oper Manag. 2015; 186–194.

21. Hange, U., Selvaraj, R., Galani, M. and Letsholo, K., 2017, May. A data-mining model for predicting low birth weight with a high AUC. In International Conference on Computer and Information Science (pp. 109-121). Springer, Cham.

22. Borson, N.S., Kabir, M.R., Zamal, Z. and Rahman, R.M., 2020, July. Correlation analysis of demographic factors on low birth weight and prediction modeling using machine learning techniques. In 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4) (pp. 169-173). IEEE.

23. Alves LC, Beluzo CE, Arruda NM, Bressan R, Carvalho T. Assessing the Performance of Machine Learning Models to Predict Neonatal Mortality Risk in Brazil, 2000-2016. medRxiv. 2020 Jan 1.

24. Jaskari J, Mylla¨rinen J, Leskinen M, Rad AB, Hollme´n J, Andersson S, et al. Machine learning methods for neonatal mortality and morbidity classification. IEEE Access. 2020; 8:123347–58.

25. Mboya, I.B., Mahande, M.J., Mohammed, M., Obure, J. and Mwambi, H.G., 2020. Prediction of perinatal death using machine learning models: a birth registry-based cohort study in northern Tanzania. *BMJ open*, *10*(10), p.e040132.

26. Liaw, A. and Wiener, M., 2002. Classification and regression by randomForest. R news, 2(3), pp.18-22.