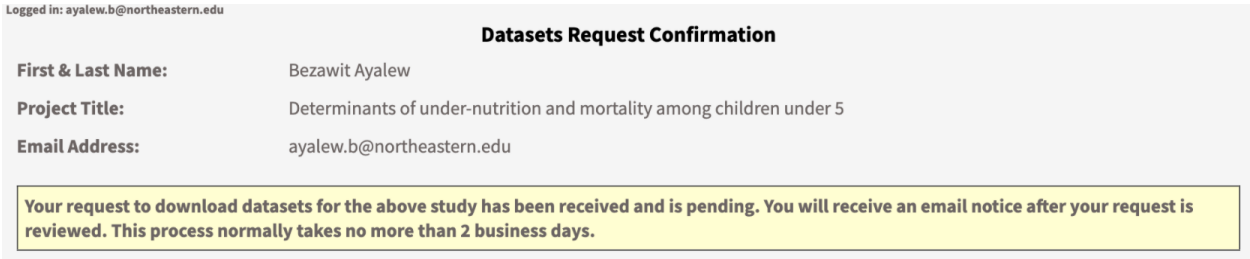
Machine learning for predicting undernutrition among under-five children

Bezawit Ayalew
Hwijong Im
Anish Rao

Introduction

Our project goal is to build Machine learning algorithms to predict undernutrition among under-five children. The study draws on data from the Ethiopian Demographic and Health Survey of 2016. However, based on data protection, we have to request to access the data sets. We are still working on the discussion of selecting models, but roughly 5 models will be selected and compared with each accuracy and performance by Confusion matrix. Also, to get better performance, we will use the Ensemble method(Bagging & Boosting) to reinforce. Each member will participate in the whole process of the entire project and will not be divided into particular parts of the project.



Logged in: ayalew.b@northeastern.edu

Datasets Request Confirmation

First & Last Name:	Bezawit Ayalew
Project Title:	Determinants of under-nutrition and mortality among children under 5
Email Address:	ayalew.b@northeastern.edu

Your request to download datasets for the above study has been received and is pending. You will receive an email notice after your request is reviewed. This process normally takes no more than 2 business days.

Data Source

https://dhsprogram.com/data/dataset/Ethiopia_Standard-DHS_2016.cfm

The dataset contains data from the Demographic and Health survey of Ethiopia in 2016.

Project Direction

We are planning on using machine learning algorithms discussed in class to predict the determinants of undernutrition among children under the age of 5. For example, for predicting the likelihood of a child being malnutritioned, Logistic regression could be used because it allows us to model the probability of a binary outcome in relation to a predictor variable (in this case the binary outcome would be whether they are malnutritioned or not). And the predictors would be certain markers from the dataset that would affect the likelihood of someone being malnutrioned). There are many more techniques such as Decision trees, Bayesian Learning etc. that we can use to further refine the data and extract more useful information.

Libraries and Tools

- Pandas, Numpy (Data manipulation)
- Matplotlib, Seaborn (graphing)
- Scikit-learn (for algorithms)

Result

The ideal outcome of the project was classify the under-five years old children who have the undernutrition circumstance and get a high performance in predicting undernutrition under-five children. We expect to see possible variations in determinants based on region, location, climate and accessibility to resources and other possible unexpected factors that directly impact undernutrition. This type of model classifier can possibly help make policy makers identify non obvious determinants of undernutrition that need to be addressed and highlighted amongst different regions.