Machine Learning – Final Project

Team: Kashifa Ahmed, Mohammed Irshad, Venkata Ratnam Mantripragada, Joan E Stone-Mays and Carol Pietro

Topic: Healthcare and Machine Learning

Goal: To determine if machine learning can predict the likelihood of infant mortality based on location and / or other factors such as race, mother's education, urban versus rural, access to healthcare, health insurance.   We are planning to use one of these factors first to test our theory and then build upon the model to see how well the model predicts based on the fewest factors.

Data sources: We are using data from the CDC Wonder Database on Infant Deaths (linked birth / infant death records) for US data, Ohio public health information warehouse, Ohio Department of health, US Vital Statistics records, and other related sources for validation, such as march of dimes perinatal data, medical journal reports on infant mortality and others.

Project Plan: 1. The plan is to gather as much data as possible, to determine which will be the best at defining and testing our model. Additionally, to use any data gathered for visualizations, especially, if not useful for the model. 2. Clean and evaluate the data using pandas. 3. Create a web page for display and visualization of the following: Introduction, to the basic topic of infant mortality and the rates within Ohio as well as other trends, introduction to the team and a break down of each of our roles; A page for all the visualizations of the topic on infant mortality and the variables that impact infant mortality, such as race, mothers education, access to healthcare, and the health of the mother; A page for interaction with the model, a drop down form for input of the variables we use and receive the information from the model; A page for a summary of all learnings from the project – HOW difficult IS it to find clean data? What were we hoping for and what did we finalize on; and finally a page for all the resources and technologies used in the production of this project. 4. We plan to build a machine learning model to find out if we can predict outcomes of infant birth, by county first in the State, then possibly the country, and adding variables until we have a good model at predicting outcomes. 5. We plan to use flask routes and a python application to control our web page. 6. We plan to use postgres to load and store our data for interaction with the model and forms.