Machine Learning Final Project Proposal

Spring 2021

Sidhikha Tripathee, Alexandra Baratta, Muhammmad Ghazi Randhawa

The Central Hypothesis:

Environmental health variables like particulate matter concentration determine the rate of asthma disease in a given area. Given the environmental injustices that have marked the development of the United States, certain communities are more at risk of suffering asthma. Placement of heavy polluting industries like waste facilities might increase air pollution and exacerbate health conditions. This project will aim to compare models and determine which factors correlate the most with asthma rates in communities across the USA. We also aim to create a map displaying our results across the map of the US.

Problem Description:

Environmental justice can be described as a movement advocating for active and equal involvement of communities of all peoples as equal chaperons and beneficiaries of environmental conservation and rejuvenation. In the context of the USA, it has been noticed that communities of color and minorities have been subjected to more than their right share of environmental damages. Historically, zoning laws have enabled certain communities to effectively outsource their waste disposals to less affluent communities, specifically Black and poor Americans. An example portraying this injustice can be seen that Chester's trash incinerator is the second largest incinerator in the country and that trash from New York, New Jersey, Maryland, and Philadelphia is sent there. Moreover, this pattern is repeated for distribution of sites of heavy polluting industries like fossil fuels power plants and waste facilities. This pattern of heavy polluting industries translates to higher incidence of air pollution and higher rates of health diseases in these communities.

USA's Clean Air Act of 1963 along with its revisions in subsequent decades make it the responsibility of the Environmental Protection Agency to maintain air quality in pursuit of public health. Another relevant law is the Presidential Executive Order # 12898 which implores executive agencies to consider environmental justice concerns when making decisions, but it does not mandate the agencies to base their decisions on these concerns. Furthermore, the agencies have an unwillingness to remediate these environmental justice grievances because legally environmental injustice can be difficult to be proved. The only permanent solution is to increase public awareness in an attempt to generate pressure on the regulatory agencies to enforce their mandate, provide borderline irrefutable evidence on the problem and to generate pressure onto corporations involved to follow these regulations. As such, a model that is able

to accurately classify whether a community is at heightened risk of air pollution related health problems and demonstrate which factors most affect at-risk communities can help highlight this problem. The primary beneficiaries of this project would be environmental justice movement organizations and the EPA who can use this model to pinpoint which communities are in the need for increased air quality the most. It can also be used to disseminate the information about the air quality problems to the public.

Algorithms:

We will be using several regression models for this project. We will use logistic regression, random forests, SVMs, and decision trees and compare the weights for each model to analyze the results. We will use the Scikit-learn libraries to implement all of these models.

Data:

The data that we will be using comes from a variety of sources. Because there is no one clean dataset that meets our needs, we will be creating one. We will be getting our initial datasets from the CDC, EPA/Kaggle, Data.org, and Homeland Infrastructure Foundation-Level Data (HIFLD) They are:

U.S. Chronic Disease Indicators: Asthma (2016)

EPA Historical Air Quality

Daily Census Tract-Level PM2.5 Concentrations, 2016

Solid Waste Landfill Facilities

Experiments:

After applying various regression models to our data, we will compare how each algorithm weighs the features: pm2.5 concentrations, age, gender, race, waste facilities, population density. We will then look for patterns in our results to determine which features likely contribute most significantly to the prevalence of asthma among US adults. For each feature, we will find the average assigned weight in order to present a probable ranking of contributing features. We predict that high pm2.5 concentrations will correspond to higher rates of asthma, and close proximity to waste facilities (which might correlate with race) will also correspond to higher rates of asthma.

Impacts:

Our project will likely provide specific evidence of environmental injustice related to the prevalence of asthma within the United States. By studying the correlation between harsh environmental conditions and

asthma rates, we hope to raise awareness about the harmful health-related effects caused by the placement of waste facilities, particularly in marginalized communities. Our results will likely highlight specific and measurable evidence of injustice, which will hopefully promote action and accountability. With a greater understanding of how man-made environmental conditions lead to higher rates of asthma in Black and unprivileged communities, we hope to bring attention and change to a pressing instance of environmental injustice in the United States.

References:

Air Pollution and Asthma | AAFA.org

Environmental Justice in Chester | Chester Digital

Summary of the Clean Air Act

Summary of Executive Order 12898 - Federal Actions to Address Environmental Justice in Minority Populations and Low-Income Populations

The Distributive Effects of Risk Prediction in Environmental Compliance: Algorithmic Design, Environmental Justice, and Public Policy