Krishi Sharma
December 12, 2019

Capstone Project Proposal

Problem

Forecasting the rate at which the ozone layer is disappearing in different cities, and the particle matters that contribute to that decline over time, can help us understand patterns that we should avoid and issues we need to confront in order to improve air quality and living conditions in the world's largest cities.

Provided the timestamp, particle matter concentrates, nitrogen dioxide, sulfur dioxide and carbon monoxide concentrations, it is possible to predict the ozone layer concentration at any point in time in the future. Forecasting the ozone layer concentration can help us determine the rate of ozone layer depletion. Scientists and policy makers of highly affected cities can use this forecasting to prioritize air quality concerns and enact change or solutions to prevent further depletion of our ozone layer.

Data + Approach

The data for this project will come from University of California Irvine's Machine Learning Repository. Specifically, the Beijing Multi-Site Air-Quality Data Set will be used. The data set will be cleaned to remove any N/A data points (or fill with meaningful values, if relevant). Dimensionality reduction tactics will be applied to make the data set more manageable and keep only significant data points for the analysis. A Machine Learning prototype will first be created, followed by a scaled version using Spark ML. Machine learning techniques will be applied, with the feature columns being the timestamp and particle and chemical concentrates, and the target column being the ozone layer concentration. Linear Regression will be performed to forecast the ozone layer concentration on a particular day, provided the attributes aforementioned.

Product

The final product will be a web app that retrieves user inputs (mimicking the data flow as if the manual user inputs were in fact being streamed from the monitoring site itself) and provides the ozone layer prediction for a certain day or timeframe. It will be deployed using an API and will be packaged using Docker.

Requirements

This project will be completed in Domino, and any additional memory/processing capabilities will be purchased if required. The API will be produced using PaperSpace, and the web app will be hosted on the creator's portfolio website, retrofuturism.net, to allow for public viewing.

Conclusion

The purpose of this capstone is to forecast ozone layers to provide policy makers and scientists with information that can help prevent further deterioration of the ozone layer across cities. Beijing's data set, across 13 different station, will be used to develop and scale this model.