

## **CSCI 5622 Project: Air Quality**

Group Members: H. Nihar Nandan, Ksenia Lepikhina, Kristen Okorn

### Problem Statement:

To predict future levels of air quality by developing a model and testing (comparing) it against current air quality measure values.

### Motivation:

With the current state of the Earth with climate change and human activities that are increasing disruptive, air quality is more important now than ever. The World Health Organization estimates that 4.6 million people die each year from causes directly related to air pollution. Currently, there are a number of cities in developing countries with large populations with declining air quality. In the near future, this will be a problem faced by even developed countries and so it would help to develop models to improve the air quality and reduce the pollution levels in our atmosphere.

One group member researches air quality in southern California using in-house low cost air quality monitors and data from government regulatory instruments, so the data from these regulatory sites can be utilized to determine other effects of air quality. The idea originated from her PhD research.

### Data / Data Plan:

The datasets that we have - One of our group member has access to air quality data in southern California as mentioned above (datasets that relate the composition of the air to the air quality level). They can be shared upon request - there is no online link currently available. For given dates and times, the features include methane, ozone, carbon dioxide, nitrogen oxide and dioxide, volatile organic compounds, and total non-methane hydrocarbons. Depending on availability of each, these may not all be used.

Datasets we are looking for - datasets that relate the air quality levels to increasing temperatures, economic activity and population. Ideally if our project is on track, we would like to incorporate a dataset with air quality levels and sea levels since these are the 2 most important effects of global warming. These will be provided by faculty at CU. Some features that we might be able to get from these would include sea level, pollutant and quantity of pollutant.

### Planned Approach:

Depending on the data that we are able to get access to, we plan on either predicting future air quality or future sea level. There are a lot of factors which contribute and are intertwined with the air quality and sea level, and applying an ML algorithm to this will help in combating pollution. Because the future is ambiguous, an excellent option for predicting it is with machine learning models. We plan on using one or more machine learning models and methods that we have learned in class.