Enhancing Short Term AQI Predictions with Location + Meteorological Data

Data Set
https://www.epa.gov/outdoor-air-quality-data
We will process the meteorology data like temperature, relative humidity, pressure, wind direction, and wind speed in addition to the available daily AQI data.

Project Idea
Wind currents are likely a large factor in the shifting air quality measurements for specific locations. This is especially prominent during times of wildfires, in which wind currents can move the pollutants to different locations over a certain time period. Many studies have shown that meteorological data have a lot to do with air quality. We're interested in exploring the correlation of air quality in different locations around Santa Clara county that is created by changes in meteorological factors like wind currents. For example, we want to see if including wind factors can help improve time series forecasts to use information from locations from where the wind was blowing from. This proposal would likely become quite complicated, so we plan to split up our explorations into the following components/milestones.

Components

- Split our dataset into times with wildfire vs normal times. Evaluate these sections separately to see if location data is more useful during one or the other. This can take the form of a multi-modal model.
- Fit basic arma, arima and sarima models as baseline performance.
- Develop cross validation method for assessing performance
- Build basic LSTM model so that we can add features/covariates in a time series setting
- Try different approaches to fit meteorological data into features
- Build the meteorological data into LSTM
- Explore whether we can tune parameter transformations or hyper-parameters so that different AQI estimates from other locations can be conditionally used in our model if the wind current data happens to cause one location to blow to the other

Software and Packages
Python, Keras, Pytorch, Statsmodels, sktime, R, ardl, forecast

Papers to read

1. Spatiotemporal prediction of air quality based on LSTM neural network
2. An LSTM-based aggregated model for air pollution forecasting
3. Regional Spatiotemporal Collaborative Prediction Model for Air Quality
4. A Spatial-Temporal Approach for Air Quality Forecast in Urban Areas

Team

1. Kushagra Gupta (kushgpt@stanford.edu)
2. Quentin Hsu (qhsu@stanford.edu)

Milestone Results

1. Baseline ARIMA models developed with evaluation metric
2. Meteorological data explored and processed into main data set
3. Basic LSTM explored