Wildfire Risk Factors (dataset)

Certain environmental/human factors can determine the probability of a wildfire starting. Based on our research we looked for datasets that contain the following factors:

- **Oxygen level**: combustion needs oxygen to occur
    - Possible correlation between CO level and oxygen level (MOPITT)
- **Temperature**: heat helps breaking chemical bonds between carbon chains, hence initiate the combustion process
    - Surface temperature can be found in MOPITT
- **Wind:** Moving moist air away from fuels, causing them to dry out faster. Moving moist air over fuels causing an increase in moisture content. Carrying burning embers that have been lifted aloft by convection air and starting spot fires ahead of the perimeter.
    - Wind influences the rate of spread and intensity of the fire.
    - Wind increases the supply of oxygen, which results in the fire burning more rapidly. It also removes the surface fuel moisture, which increases the drying of the fuel.
- **Precipitation level**
- **Humidity**: more likely to catch fire if region is dry
- **Weather condition**: natural phenomena such as lighting can act initiate a wildfire
- **Current fire condition**: if there's already a fire in a region, we need to update the map
- **Human activity**: a large percentage of wildfire is actually man-made. If we can trace activity level/type in regions that's tat risk of a wildfire, then we can send notifications/reminders
    - Artificial activities can change our analytic results drastically (ex: factory releasing chemicals as byproducts, causing an abnormally high level of CO)
    - This one is a lot more dynamic/variable, so we might not have enough time to fully study/gather data on the impact of these activities. We should aim for a simulation instead

Input Data (Before Analysis)

Each country will be separated into small regions in our data representation. We can treat each unit of region as an object that has the following attributes:

- **Coordinate** [Longitude, Latitude] (normalized to represent a regional area)
- **Risk factors** (listed above)

Output Data (After Analysis)

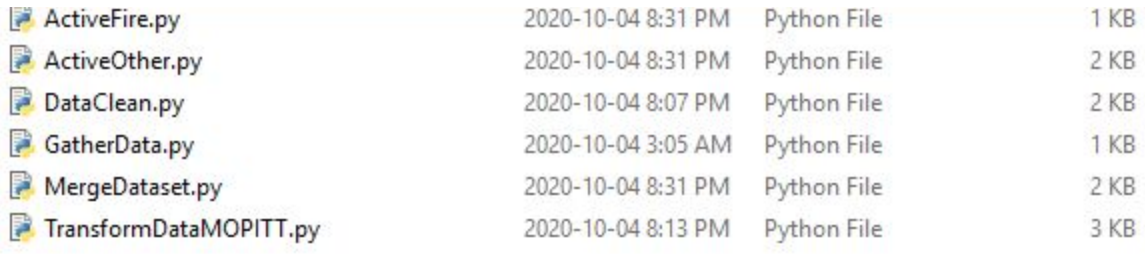
The map will be separated into individual regions in our data representation. Each unit of region contains the following fields:

- **ID** (set unique ID to each region so front-end can update map visualization efficiently)
- **Coordinate** [Longitude, Latitude] (normalized to represent a regional area)
- **RiskLevel** (1-6)

    - 1 = very low, 2 = low, 3 = slightly low, 4 = slightly high, 5 = high, 6 = very high
- **IsCurrentOnFire** (bool)
    - Confirmed that there's a fire
    - Mark surrounding regions to be high risk & predict fire track
    - send evacuation notification
- **ReportOnFire** (bool)
    - Reported that there's a fire, not yet confirmed
    - Send warning notification

Data Gathering/Filtering/Transformation:

Some datasets given for this challenge are easy to download manually, but for a large number of data files (such as from MOPITT), scripts were created to help automate the data gathering process. Since we had data from multiple sources, additional files were needed to process, and transform these data so they are all in a matching format that can be fed into our model later. These files are stored in the "CSA_Hackathon\Analysis\ProcessData" folder:



Model Type:

Since we have previous years data, we can split the data into training data & test data. With these data sets, we planned to use supervised learning to train our model.

- Reference site for examples of modeling technique that we can use
    - https://towardsdatascience.com/selecting-the-correct-predictive-modeling-technique-ba459c370d59
    - https://towardsdatascience.com/weather-forecasting-with-data-science-approaches-cb8f2afd3f38
- *NOTE:* Unfortunately we didn't have enough time to work on the model, but the idea is to map/find correlation between factors we gathered and the risk level of a wildfire occurrence.
    - With a confirmed correlation, we can then calculate to which degree these factors can affect the likelihood of a fire starting
    - Once a model has been trained/developed, we'd then feed a different set of historical data to test out the accuracy of its predictions
    - Refine model with better data, more fitting techniques until its predictive accuracy increases to an acceptable level

Future Consideration:

- Look outside of known factors and find out if there are any new environmental/human factors that can affect the likelihood of a fire starting. Also, how can we capture that relationship and convert it into useful data?