ORIE 4741 Project Proposal

Vrushali Samant(vms59), Yash Mundra(ym267)

Hydraulic fracturing--more known widely as fracking--is a highly controversial activity at the root of various climate change policies that shepherd the movement towards clean and natural energy sources. While the process of injecting high volumes of chemicals, sand and water at high pressure into rock formations does release trapped gases and oil, allowing them to rise to the surface, the side effects of this activity are numerous and often condemned.

Of these various effects, we can highlight the following: water depletion, water contamination, air pollution, and earthquakes. For the scope of this project, however, we aim to predict water contamination[1] of a public water source given what we know about the fracking site[2]. Our two main sources of information are EPA, the United States Environmental Protection Agency, and FracFocus, a public chemical disclosure registry. The EPA dataset contains information such as the various locations of water contamination, the pollutants involved and the size of the water sources. The FracFocus dataset on the other hand gives us information detailing the fracking locations throughout the US, how long the sites have been running and the water volume involved. Given that the FracFocus dataset has information throughout the US, we hope to focus our analysis on a subset of states in order to control the dataset size, as well as increase our effectiveness on data analysis.

Having visually examined our data, we have identified several features that will likely factor into the relationship between a fracking event and resulting water contamination: duration of the activity, proximity to the water well, and fracking fluid ingredients, to name a few. Since the dataset contains mostly ordinal and nominal values, it makes sense that we do a lot of feature engineering. After that we hope to use various methods such as linear regressions and decision trees in order to predict whether there is a correlation between fracking and water contamination sites. Other potential techniques and methods would also be explored in this project.

We believe that fracking and its potential effect on water contamination is an issue that everyone should be aware of. By achieving high levels of predictability in water contamination spatial and temporal proximity (not to mention the various other features that are likely to factor in to this relationship) of the fracking site and the public water source will hopefully allow us to issue in advance warnings to the general public when their drinking water is likely to become toxic.

---

[1] https://www.epa.gov/dwucmr/occurrence-data-unregulated-contaminant-monitoring-rule#3
[2] https://www.fracfocus.org/index.php?p=data-download