Mahlon Page
CSCI 1420
Capstone Proposal

Predicting Energy Generation & Energy Use

For my project I want to use weather data to predict energy consumption and energy use in the coming hours/days. I've taken two EEPS classes at Brown including one called EEPS 0850: Weather and Climate where we talked about modeling and predicting future weather as well as the weather's impacts on various aspects of the world like energy use and consumption. In my other class EEPS 1320: Intro to GIS with Environmental Applications we've talked about modeling energy use on maps and the impact of energy on the climate. I want to bridge my interest in the two topics of machine learning and weather forecasting to predict future energy production and consumption.

For my dataset I have four years of hour by hour energy production and consumption data for the country of Spain. I also have hourly weather data for the same four years for five distinct cities in Spain. In this dataset I have weather factors such as temperature, pressure, humidity, cloud cover, and wind speed. In my energy dataset I have total energy production varied by type such as wind versus solar versus fossil fuels and also the total amount of electricity used. These will be the most useful attributes. I imagine that I can use current wind data and cloudiness data to predict solar and wind energy production as well as use temperature to predict energy consumption. One thing in particular I think might be cool is using the difference in temperatures between cities as a prediction of future production. When we learned about weather forecasting we learned that wind is caused by temperature differences between locations and air moving to try to find equilibrium between these temperature differences. I think as a whole, especially with data from cities in geographically separate regions within the same country leaves a lot of room for feature engineering. I can imagine a variety of different factors that could all be good predictors of future energy production/use such as temperature gradients, cloud cover and how those clouds might move, etc.

This is an interesting problem for a couple of reasons. First, being able to predict energy consumption in whole can help not overproduce energy. If we can know exactly how much energy we need, we won't have to burn excess fossil fuels which are bad for the environment. Secondarily, if we can predict the amount of solar or wind energy we will produce based on the weather we can also refine this same metric to reduce fossil fuels being burned. Being able to have a balanced energy understanding can also reduce the occurrence of negative energy prices. There are stories of dams using excess energy to pump water back up the dam to then refilter through the dam as a way of conserving energy for later since energy has a negative cost and batteries are expensive. Preventing this seems like a clear good for society. Negative energy prices encourage energy companies to squander produced energy which maybe shouldn't have been produced in the first place.

I envision this task will be challenging in deciding how to factor all of these various cities' data together. Thinking about data spatially can be tricky and deciding how to properly integrate the data from various locations will prove difficult. As mentioned before, this dataset has a lot of potential for feature engineering, however I think there will be difficulty in deciding what features are most important or seem relevant. Having such detailed temporal data and spatial data leaves a large space of features that can be produced and finding the right ones will be tough. The biggest difficulty though will be in finding the balance between distance in the

future of our prediction and accuracy of the prediction. A further out prediction is more useful but only if it is still accurate.

I want to use this dataset from Kaggle.