THE ML PROJECT - ENERGY FORECASTING

We are interested in forecasting energy consumption within the US. To do this, our initial approach will be to model the consumption at the state and potentially sectoral level using electricity and weather time series data. While we are still in the process of refining the policy question, but the aim is to identify situations that require additional energy infrastructure, are at risk of energy shortages, and are at the risk of energy spikes.

US Energy Information Administration

The US Energy Information Administration (EIA) publishes US power and electricity data by power source (e.g. electricity broken down by generation source, primary use of natural gas, primary use of petroleum oil) by month and state going back over 2 decades. Sparser data is available going back as far as 1970. Furthermore, powerplant-specific data is data in many cases going back as far as 2001. While we have not settled on a specific set of time series to use to construct our dataset, we are tentatively considering using the following:

Form EIA-923 collects detailed monthly and annual electric power data on electricity generation, fuel consumption, fossil fuel stocks, and receipts at the power plant level. The data series goes back to 2001

Forms EIA-861M and EIA-826 collect monthly electricity sales data at the state-sector level. A summary of the more recent data is available here.

The EIA offers several other datasets, including renewable energy production, net metering, which is related to EIA-861M, and others.

Weather (via NOAA)

The National Oceanic and Atmospheric Administration (NOAA) offers several datasets that can be used to construct a weather time series.

The files are available for bulk download here.

Team Members: Manjiri Girish Satam, Miriam Runde, Jackson Luckey, and Aditya Narayan Rai