March 18, 2021
Springboard: Data Science
Capstone 3: Project Proposal & Description
Spencer Rubin

Machine Learning Forecasting of Energy (Electricity) Consumption

Problem:

Energy sourced from utilities varies by region, by population, and throughout different periods of the year. Utilities require accurate planning for future demand and subsequent sourcing and transmission of the power based on fluctuations with weather, population changes, and seasonality. By forecasting energy demand based upon historic energy consumption within a specified area, and utilizing machine learning modelling, we can improve the accuracy of future needs, demands, and changes, thus providing increased efficiency within utility and energy distribution processes, and subsequently reduce potential power outages and issues with customer energy consumption. This can be most critical during high-demand seasonal periods such as winter and summer where extreme weather conditions tend to present challenges for forecasting energy consumption. Also, it may provide an avenue for reducing spikes in energy cost to customers by smoothing and averaging annual demand-related fluctuations in consumer costs

- Can machine learning modeling of hourly energy consumption from various utility companies' data provide a more accurate prediction to future energy demands compared to a baseline persistence model?

Context and Path for Success:

PJM Interconnection LLC is a multi-state, regional utility transmission company that provides electricity to a large multi-state region in the eastern U.S. Utilizing hourly energy consumption provided from historic PJM datasets will provide an accurate historical dataset to build a forecasting model of energy consumption.

Wrangling the data across the time series, identifying trends using moving averages and smoothing of the time series may improve the accuracy of a time series model. This may improve on identification of the time series statistics and significance (i.e stationarity within the historic data of energy consumption, seasonality, and trends over the time period of analysis). Also, the model should improve customer satisfaction with PJM's service of reliable energy, and thus have a reduced error compared to the persistence model.

Scope and Constraints:

A multi-year historical hourly data consumption data (electricity quantified in megawatts) for a multi-state region served by PJM will be utilized to build the forecasting model. Pandas, Numpy, SciKitLearn, and other data analysis packages will be utilized.

Constraints may include:

- Improving on future predictions relative to only using historical data.
- Identifying important data features and appropriate scaling to improve model accuracy.
- Unplanned for events, such as a hurricane or extreme weather situation, that is unforeseen in the dataset.

Stakeholders:

- PJM Interconnection LLC
- Local governments and customers
- Regional utility companies

Data:

Data will be sourced from Kaggle's "Hourly Energy Consumption" dataset, accessible at the following url: https://www.kaggle.com/robikscube/hourly-energy-consumption. The data is available as in .csv. Multiple .csv files will be downloaded, reviewed, and merged appropriately into a pandas dataframe. The documentation will be uploaded to GitHub using Jupyter Notebooks and Python language, and relevant figures, features, graphs, and statistical analyst will be provided within the documentation.