

Jefferson Fernandez

12-2020

Capstone Two: Project Proposal

### Problem Statement Formation

This project studies the problem of building regression models to estimate the energy produced by a Combined-Cycle Power Plant, as a function of given parameters”

### Context

One of the main challenges for electric power companies is balancing the load in the grid (total electricity output vs total electricity demand). If an electric power company can predict the electric output from its own power plants then it can know in advance if it could generate enough power to meet the power demand or if they would need to purchase electricity from other electric power companies. In this project we are provided a data set, as a CSV file, which contains 4 hourly average ambient variables, and the associated net hourly electrical energy output of a combined cycle power plant.

### Criteria for Success

This project will build several regression models to estimate the production of electric power as a function of the features in the provided data set. These models will be compared with respect to appropriate performance metrics, such as R-Squared, Mean Absolute Percent Error (MAPE), and upper/lower bound of worst-case residuals. Finally, the models will be used to conduct analyses to show the impact and importance of the features that might suggest important insights to stakeholders.

### Scope of solution space

The models that will be built will only be based on 4 features. If we could get more features like gas input, then we could create an even more robust model.

### Constraints within solution space

The data set does not provide time-dependent information, and therefore time series analyses will not be conducted.

### Stakeholders to provide key insight

Resource management team at the power company that is responsible for balancing electric grid load.

### Key data Sources

Excel file containing 9568 data points collected from a Combined Cycle Power Plant collected over a 6-year period while the plant was set to work with full load. There are no missing values and there are 4 attributes.