

# Forecast Power against Weather

## Stage-1: Explanatory Data Analysis

Basic steps for raw data analysis:

- Load individual CSV files into dataframes
- Preliminary examination
  - Check for datatypes for individual columns in the dataframe
  - Check for NULL values across the dataframes
  - Find max and min values for individual columns/attributes
  - Plot data against specific time periods (smaller time intervals) to understand the kind of data we are looking at.

## Stage-2: Prepare Data

Based on preliminary examination, following steps are performed to prepare the data for analysis:

- Clean up the data
  - Replace NULL values
  - Convert columns containing Datetime from object to datetime format.
  - Reindex the data based on datetime.
  - Rollup the actual data into hourly and daily units.
    - In case of Power data, we sum up the power output to get the daily power generation.
    - In case of Weather data, we have calculated a mean of the values over the day as these don't vary much over a given day.
  - Remove un-necessary/redundant columns
- Merge Datasets
  - Once the data is indexed and rolled up to common units/time interval, the data sets are merged together for further analysis.
- Analyse impact of weather attributes on power generation from the merged dataset. This is done by plotting graphs of power generation against various attributes of weather . This shows the correlation between weather and power.
- Reindex the data to prepare training sets

- Create X\_train and Y\_train data from the merged dataset based on the correlation identified in the previous steps.

## Stage-3: Load Weather Forecast Data

Since we are provided with the weather forecast data, we load this data from CSV to a dataframe. All the same cleanup and preparation steps are performed on this forecast dataset to prepare it for being passed as test data to the trained model.

## Stage-4: Train The Model

After we have the train dataset available, we can use the sklearn library to create a **linear regression** training model. We pass the X\_train and Y\_train data to LinearRegression object's **fit** method to train the model.

## Stage-5: Predict

Once the model is trained, we pass the weather forecast data to the model's **predict** method as our test data. The output of this is the power generation forecast against the test dataset, which in our case is available in weather\_forecast.csv.

The output from this model is finally merged with the weather forecast data and written back to a CSV file as output.