Electricity Consumption Prediction

By
Vishal Wadagavi
26th Nov 2021

Electricity Consumption Prediction

Problem Statement:

Context
Company of Electrolysia supplies electricity to the city. It is looking to optimise its electricity production based on the historical electricity consumption of the people of Electrovania.

The company has hired you as a Data Scientist to investigate the past consumption and the weather information to come up with a model that catches the trend as accurate as possible. You have to bear in mind that there are many factors that affect electricity consumption and not all can be measured. Electrolysia has provided you this data on hourly data spanning five years.

For this competition, the training set is comprised of the first 23 days of each month and the test set is the 24th to the end of the month, where the public leader board is based on the first two days of test, whereas the private leader board considers the rest of the days. Your task is to predict the electricity consumption on hourly basis.

Note that you cannot use future information to model past consumption. For example, you cannot use February 2017 data to predict last week of January 2017 information.

Content
It represents a fictitious time period wherein we are to predict future electricity consumption.

Electricity Consumption Prediction

Approach:

1. Data Gathering:
    - Dataset is fetched from a data source called Kaggle.
    - Data is already in the form of csv format, so it can be directly read using pandas.
2. Data Cleaning:
    - Since there are no missing values there is no need of replacement.
    - There are no duplicates in data set so there is no need of dropping data points.
    - Feature dropping if there is high correlation coefficients between features.
    - Scaling numerical features with Standardscaler()
    - Outlier treatment if there are any outliers.
    - Encoding Categorical features using labelencoder or onehot encoder based on type of feature.
    - Extracting meaningful feature out of datetime feature as data is recorded at every hour over the period of time.

Electricity Consumption Prediction

3. Data Exploration:
    - Exploring 5 points summary of all the features to know how data is varying from regular interval of percentiles.
    - Relationship among the features are explored using pairplots.
    - Heat map of correlation between features are plotted using which dropping of features is decided.
    - Distribution plots and boxplots are plotted in order to explore the outliers.
    - Bi-variate analysis is done to observe how feature values are influencing to the target variable.
4. Model Building:
    - Pre processed numerical and categorical features are concatenated to form final data frame which will be used for model building.
    - Final data frame is split into independent and target features.
    - And also data is split into train and test data.
    - Train data set is used for the training the model and test data is used for validating it.
    - From the pairplot, there is no linear relationship found and data is randomly scattered. So its better to build some ensemble or boosting or other models in which kernel could be controlled using parameters.

Electricity Consumption Prediction

5. Evaluation:
    - Built models are evaluated against test data.
    - Individual model’s parameters are tuned to have good R2_score.
    - Models are compared with Mean Absolute Percentage Error (MAPE).