

Project proposal: Electricity consumption segmentation and prediction using daily data

The recent evolution of the electricity business regulation has given new possibilities to the electricity providers for formulating dedicated tariff offers. A key aspect for building specific tariff structures is the identification of the consumption patterns of the customers, in order to form specific customer classes containing customers exhibiting similar patterns.

In this project, you will be given a dataset which provides daily electricity usage of 6445 of houses and companies between 2009 and 2010. The purpose of the project is to group customers with similar electricity consumption behavior, analyze their behavior and predict load consumption of each cluster.

The contributions to this work are listed as bellow:

1. Clean the data (Handle missing values).
2. For each customer, add different features to the current dataset. These features should provide necessary and meaningful information for segmentation of load profiles. They can be statistical features, such as average daily consumption, percentage of consumption on weekdays, etc. They can also be calendar features like day of the year, type of the day (public holiday or not), etc. you should think of any other kind of features which improve the segmentation process.
3. Consider a clustering technique from Scikit-learn to do the segmentation (clustering).
4. Justify your choice regarding the clustering algorithm and validate your choice regarding number of the clusters, using a suitable technique.
5. Illustrate the differences among clusters based on the distinguishing features (Use plotting and visualization techniques).
6. In order to find which group of customers has more predictable consumption behavior, use a machine learning algorithm to predict the total daily consumption of each cluster in the last 80 days of the dataset.
7. Evaluate results of the prediction model using different measures such as RMSE and MAPE.

To have access to the dataset, refer to this Kaggle link:
<https://www.kaggle.com/zonnalobo/daily-electricity-usage>