Clustering Analysis for Energy Consumption

Context: Cluster analysis or clustering is an unsupervised learning technique that involves grouping data points into different subgroups (clusters). The members within the same group share common characteristics and are considered more similar than the members who belong to various groups. The (dis)similarity between the data points is calculated according to a distance measure to divide the data points into different subgroups. Clustering is an important method that allows us to gain a better understanding of the data and discover similar observations in this problem. This thesis work is conducted as part of a collaboration project with an energy provider partner to analyze the load on energy stations. The dataset represents real-world records of the customers' energy consumption recorded at a given timestamp.

Problem: The energy providers are trying to understand their customers' behaviors and patterns to improve their services. Thus, explainable clustering methods play an essential role in uncovering hidden subgroups in the dataset to help decision-makers take action. To address this problem, four main clustering types could be used: Connectivity-based clustering, Centroid-based clustering, (Model) distribution-based clustering, and Density-based clustering.

Task: The goal of this thesis is to apply clustering techniques to the energy consumption dataset. The task is to cluster the different parts of the day based on the energy consumption used by customers. Adopting these techniques would better manage the loads on the energy substations during different parts of the day. The technical work of this thesis is intended to be implemented in the Python programming language. Python offers a wide range of built-in libraries and frameworks to visualize and process different types of datasets and implementations of various machine learning algorithms.

Contact:

Bestoun S. Ahmed

bestoun@kau.se