CIND820 D1H - Big Data Analytics Project - F2024

Name: Shankar Ghimire

Student #: 501177872

Program: Data Analytics, Big Data, and Predictive Analytics

Abstract

The dataset on ‘Energy and water usage of large buildings in Ontario’ provides comprehensive information on various aspects including property types, energy and water consumption, greenhouse gas emissions, and energy efficiency scores etc. The dataset contains 6,864 rows and 32 columns.

The problem being considered for the project is the classification of the properties based on their attributes and forecasting the future demand for energy and water service. For the prediction, the appropriate algorithms such as regression models will be applied on the dataset. Similarly, for the classification of the properties, algorithms such as Decision trees, random forest, K-Nearest neighbors etc. will be applied.

For the classification and prediction tasks on the given dataset, a systematic data analysis approach will be employed. Initially, the entire dataset will go a cleaning process, followed by preliminary analysis utilizing various exploratory data analysis tools. Then, experimental design and model constructions will be undertaken. Finally, the performance of the model will be evaluated with recommendations and conclusions.

References:

Dataset source

https://data.ontario.ca/dataset/0eab2faf-6186-4a5b-8de1-b15872943c24/resource/f53db89b-d5e0-49ea-aa85-70453d969453/download/energy\_large\_building\_energy\_water\_ghgs\_2022.xlsx