Predicting Residential Energy Cost and Consumption Through Machine Learning

(Electricity)

Description: This study will present the development, testing, and evaluation of a supervised machine learning model in order to predict the electricity consumption for residential homes. With the rapid evolution and adoption of sustainable energy technologies due to efforts to limit our ecological footprint, our dependence on fossil fuel energy has convinced homeowners to transition to solar energy. Increasing government regulation and legislation surrounding energy consumption in an effort towards achieving greater sustainability necessitates the accurate forecasting of energy cost, especially for residential homeowners who may want physical evidence to support their transition to solar energy or discover ways to reduce their monthly energy bill.

Energy consumption can be influenced by a variety of factors, such as the number of appliances, the size of the house, the number of residents, etc. These are features that are related to the household characteristics. Even some environmental factors like the season, economical factors like the current price of fuel and electricity determined by energy companies, etc.. all have significant influence over how much electricity is consumed and the price of this consumption throughout the year.

Dataset:
https://www.kaggle.com/datasets/claytonmiller/2015-residential-energy-consumption-survey/data

Learning Goal:

- **Problem that we want to solve**:

    We want our model to grant homeowners the ability to make decisions on how to reduce their energy consumption based on their predicted consumption levels. They could pinpoint what factors contribute the most to the overall consumption and devise a strategy to reduce the consumption of those particular features.

- **Is it possible with the data available and how will we evaluate the success of this project?**

    This dataset provides us with over 700 labeled and very well-described features to use in our model. We find this to be more than sufficient data to make educated predictions with. Given the surplus of features, it may be difficult to identify which ones play the biggest role in increasing energy consumption. We are highly likely to select features that we assume to be more influential. The data is very clean and includes many integer encoded features, as well as little to

no missing values. We will have to normalize some of the data as there are highly varying values in different features.

- Questions:
    - What characteristics of building have an impact on energy consumption? (temperature, location(rural, urban,...etc), appliances(AC, Water heater, oven, dishwasher, ,solar system, fridge, pool, microwave, heat pump,fireplace, ceiling fan), # of bedrooms and bathrooms, size of the building, # of residents, material of the building)
    - Pattern in energy consumption?
    - Owned / Rent
    - How does the year of the house built affect types of energy consumption?
    - Ethnicity, Age, income, marital status?
    - Price of each energy source
- Goal - Predicting resident's monthly energy (electricity) usage and cost.
- Dataset - The dataset includes Expenditures, Structure of residential buildings, locations, demography of household, appliances. We will need to preprocess features as some of the data are missing or not numerical (one hot encoding).
- Evaluation - Mean Squared Error, Coefficient of determination

Group Members: Cameron Cobb, Cole Simpson, Ryo Taono