Predicting Household Energy Consumption

Dhananjay Mandalkar,
ReDi School NRW

October 1, 2025

The objective of this study is to develop a predictive model for estimating house hold electricity consumption based on key household characteristics. The dataset under consideration consists of three variables.

Please download the file electricityConsumption.csv

- Home Size (SqFt) Representing the total area of the house in square feet.
- Heating Type Indicating the primary heating system used in the household (e.g., Electric, Gas, Heat Pump, Solar Hybrid).
- Monthly kWh Target variable representing the monthly household energy con sumption.

The dataset includes missing values that must be properly addressed through data pre processing techniques. The primary research task is to apply Multiple Linear Regres sion (MLR) to model the relationship between home size, heating type, and monthly energy consumption.

The goals of this work are as follows

1. To analyze the influence of home size and heating type on monthly electricity usage.
2. To construct a regression equation that predicts energy consumption ($\hat{y}$) based on the explanatory variables.
3. To evaluate the model's predictive performance using appropriate regression metrics such as R^2, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

The general form of the multiple linear regression model is

$$\hat{y} = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \cdots + \beta_p \cdot x_p$$

where

- $\hat{y}$ is the predicted monthly energy consumption,

- β_0 is the intercept,
- $\beta_1, \beta_2, \ldots, \beta_p$ are the regression coefficients,
- $x_1, x_2, \ldots, x_p$ represent the input variables.

The outcome of this analysis will provide both a statistical understanding of the determinants of household electricity consumption and a practical prediction tool for future estimation.