Project 2: Energy-Efficiency-Statistical-Analysis, Draft 1

Christina Mourad, Victor Um, Joe De Leon, Martin Ha

2024-12-02

Abstract

Utilizing UC Irvine's "Energy Efficiency" dataset, we developed multiple statistical machine learning models in order to create predictive models that will depict the most efficient heating and cooling loads based on the building characteristics of residential buildings, aiming to provide insights that can inform sustainable design practices. We analyzed 8 input variables (relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, glazing area) and its relationship with two output variables, heating load (HL) and cooling load (CL), to optimize model performance by experimenting with different regression techniques and feature selection methods. We initially explored the dataset discovering the data distribution, correlation strength of each feature/target variable with each other, and the feature's significance with each target variable. Afterwards, we constructed multiple linear regression models to discover any additional associations, building upon our initial discoveries and fine-tuning our model, as well as comparing it against a more robust machine learning model, in this case decision trees, when predicting HL and CL. Lastly, to provide more actionable, real-world insights, we transformed the original regression tasks into a classification problem by implementing a multi-class logistic regression model. This approach, alongside the use of multiple linear regression models and decision trees for discovering associations, allowed us to not only predict outcomes but also offer more concise, clear, practical recommendations.

Introduction

Methodology

Data Analysis

Conclusion