Thrive AAI Project A

Energy Efficiency Prediction

Student Name: Jyothi Mayuri Chatla
Email: jyothi.mayuri@gmail.com
Date: 19.09.2025

Introduction

The purpose of this project is to predict the energy efficiency of buildings using the UCI Energy Efficiency dataset. The dataset contains 768 building simulations that explore how different design choices impact heating and cooling loads. The goal is to identify which building features drive energy consumption, compare prediction models, and provide concrete business recommendations to improve energy efficiency.

Dataset Description

The dataset comes from the UCI Machine Learning Repository and Kaggle. It contains eight input features related to building design, such as surface area, wall area, roof area, orientation, and glazing area. The target variables are Heating Load (Y1) and Cooling Load (Y2). The dataset has 768 rows with no missing values.

Data Exploration & Preparation

The data was explored using summary statistics and visualization techniques such as histograms, boxplots, and heatmaps. Correlation analysis was performed to identify relationships between building features and target loads. The dataset was split into training and test sets, and features were standardized where required. Outliers and anomalies were reviewed to ensure data integrity.

Model Training & Evaluation

Model Target RMSE R^2

Linear Regression Heating Load (Y1) 3.03 0.9122
Linear Regression Cooling Load (Y2) 3.15 0.8932
Random Forest Heating Load (Y1) 0.50 0.9976
Random Forest Cooling Load (Y2) 1.73 0.9679

Results show that Linear Regression provided a reasonable baseline with R^2 above 0.89 for both targets. However, Random Forest Regressor significantly outperformed Linear Regression with much lower RMSE and higher R^2, indicating near-perfect predictive accuracy.

This demonstrates the strength of Random Forest in capturing complex, non-linear relationships in building energy data.

Feature Importance

Analysis of feature importance from Random Forest indicated that features such as glazing area, roof area, and wall area had the strongest influence on heating and cooling loads. Building orientation and relative compactness also played an important role, though to a lesser extent. These insights help to identify key design factors that directly impact energy efficiency.

Business Recommendations

Based on the analysis, the following recommendations can help improve building energy efficiency:

- *Minimize west-facing glazing areas to reduce cooling load.*
- *Optimize roof and wall area ratios to balance heating and cooling efficiency.*
- *Use energy-efficient glazing materials to reduce energy loss.*
- *Consider orientation in design to minimize energy consumption.*

Conclusion

The project successfully demonstrated the application of machine learning techniques to predict building energy efficiency. Random Forest Regressor significantly outperformed Linear Regression, achieving near-perfect accuracy for heating load prediction and high accuracy for cooling load prediction. These findings underline the importance of advanced models in capturing complex relationships in data. The insights derived provide actionable recommendations for improving building design and reducing energy consumption.