Final Project for Group 2 - Energy Efficiency

Energy Efficiency Dataset by Athanasios Tsanas, Angeliki Xifara from the UC Irvine Machine Learning Repository

David Liu, Saim Zafar

STAT444, Spring 2025

Research Question and Data

Research Question

Can we create a Machine Learning model to evaluate the heating and cooling demands of a residential building using only its engineering design variables?

Variable	Mean	Std	Min	Max
Relative_Compactness (X1)	0.7642	0.1058	0.62	0.98
Surface_Area (X2)	671.71	88.09	514.50	808.50
Wall_Area (X3)	318.50	43.63	245.00	416.50
Roof_Area (X4)	176.60	45.17	110.25	220.50
Overall_Height (X5)	5.25	1.75	3.50	7.00
Orientation (X6)	3.50	1.12	2.00	5.00
Glazing_Area (X7)	0.2344	0.1332	0.00	0.40
Glazing_Area_Distribution (X8)	2.8125	1.5510	0.00	5.00
Heating_Load (Y1)	22.31	10.09	6.01	43.10
Cooling_Load (Y2)	24.59	9.51	10.90	48.03

Data collected from UC Irvine Machine Learning Repository.

Methods

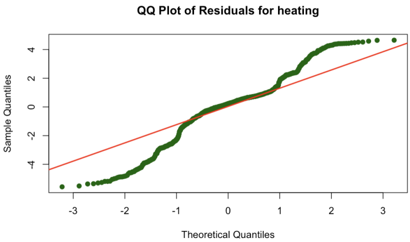
Machine Learning Models used to evaluate Heating and Cooling demands:

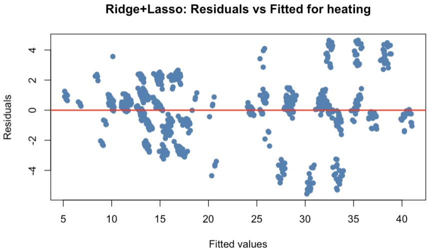
- Ridge Regression
  - Serves as a baseline to capture linear relationships
  - Performed 10-fold CV over $\lambda \in [0, 50]$; recorded optimal $\lambda(0.25)$ and corresponding to the minimum MSE
- Lasso Regression (Ridge + Lasso)
  - Select interaction terms using Lasso regularization (combined with Ridge)
  - Performed 10-fold CV over $\alpha \in [0, 1]$ and $\lambda \in [0, 50]$; recorded the optimal (α, λ) and choosing $(\alpha = 1, \lambda = 0.002)$ which has minimum MSE.
- Spline Regression with Lasso Regularization:
  - Fitting a spline regression with the mgcv library.
  - Regularize using a L1 regularization.
- KNN:
  - Motivation - "Similar" residential structures may have similar energy load.
  - Run 10-fold CV for $k \in [1, 50]$, using the best k(5) with the lowest MSE.
- Random Forest: As proposed by the Initial Researchers.

Results

Models with 10-fold Cross Validation error for Heating and Cooling load.

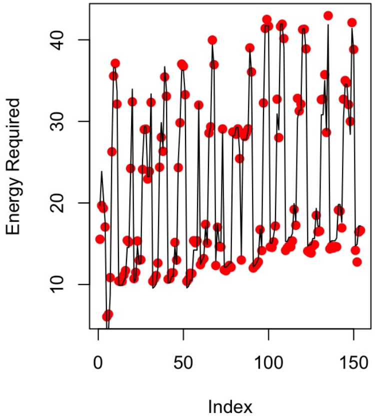
Model	**Cooling load**	**Heating load**
Random Forest	1.03	6.59
Spline & Lasso	1.09	3.07
KNN	6.05	10.66
Ridge & Lasso	7.3	4.85
Ridge Regression	10.26	8.64





Conclusion

Actual vs. Predicted Heating Load (Y1)



Actual vs. Predicted Cooling Load (Y2)