Predicting Building Energy Consumption Using Machine Learning: Analyzing Residential Heating and Cooling Loads

IYEOSE SIMON UHUMUAVBI

Advanced Decision Making: Predictive Analytics and Machine Learning

Table of Contents

Abstract

The rapid rise in global energy consumption—particularly in the residential sector's heating, ventilation, and air conditioning (HVAC) systems—necessitates accurate predictive tools for optimizing building performance and reducing carbon emissions. This study investigates the performance of four machine learning regression algorithms—Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN)—in predicting two critical continuous targets: Heating Load (HL) and Cooling Load (CL).

Using the Energy Efficiency dataset (Tsanas & Xifara, 2012) from the UCI Machine Learning Repository, a structured preprocessing pipeline was applied. This included exploratory data analysis, correlation assessments (Spearman and partial), Z-score standardization to prevent data leakage, and feature selection to enhance model stability and interpretability. DT, RF, and SVM models were trained using 10-fold cross-validation, while ANN utilized resilient backpropagation. Model performance was evaluated on a held-out test set using RMSE, MAE, and R^2 metrics.

Results show that RF and ANN achieved the highest accuracy ($R^2 = 0.997$ for HL and $= 0.969$ for CL), effectively capturing non-linear relationships, particularly those involving geometric and glazing features. Glazing Area emerged as the most influential predictor.

This research highlights the advantages of ensemble and non-linear methods in energy demand forecasting and emphasizes the value of thorough preprocessing. It lays a foundation for future work on hyperparameter tuning, the integration of alternative algorithms, and the application of explainable AI techniques to further improve model transparency.

Keywords: Residential Energy Consumption · Heating Load · Cooling Load · Machine Learning · Random Forest · Artificial Neural Network · Feature Selection · Model Evaluation

Introduction

Global energy consumption has risen significantly in recent decades, prompting the need for more efficient energy management—particularly within the building sector (Santamouris & Vasilakopoulou, 2021). Residential buildings contribute substantially to overall energy demand and greenhouse gas emissions, with heating, ventilation, and air conditioning (HVAC) systems accounting for a large share of this consumption (Mardiana & SB, 2015; Kassas, 2015). While HVAC systems are essential for maintaining thermal comfort across various climates and seasons, they are among the most energy-intensive components in residential buildings. As a result, optimizing HVAC performance is critical for promoting energy efficiency and sustainable building practices (Wei et al., 2019).

Historically, building energy consumption was modeled using physics-based simulation tools such as DOE-2, EnergyPlus, and TRNSYS. Although effective, these methods are often complex, computationally intensive, and require detailed input data (Pan et al., 2023). In recent years, there has been a paradigm shift toward data-driven approaches, with machine learning (ML) techniques gaining traction for their ability to analyze historical data, uncover hidden patterns, and generate accurate predictions without the need for explicit programming (Tri et al., 2019; IBM, 2023).

Machine learning has demonstrated superior performance in predicting building energy consumption, particularly for heating and cooling loads (Shapi et al., 2021; Seyedzadeh et al., 2018). Accurate forecasting of these loads is essential for optimizing HVAC control, minimizing energy use, and maintaining occupant comfort (Balali et al., 2023). The integration of ML algorithms—such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Decision Trees (DT)—into HVAC systems enables automated, adaptive control strategies that improve both efficiency and responsiveness (Wei et al., 2019).

This project focuses on predicting the heating and cooling loads of residential buildings using supervised machine learning algorithms. The primary objective is to model energy consumption by comparing the predictive performance and computational efficiency of four regression techniques: Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN). These models are trained and tested using consistent preprocessing protocols and a shared dataset of architectural and environmental features.

While many studies have applied ML to building energy prediction, few have systematically compared multiple algorithms under a unified framework. This research addresses that gap by evaluating both interpretable models (e.g., DT) and more complex black-box models (e.g., ANN), offering a comprehensive analysis of their strengths and limitations. The findings aim to inform best practices for energy forecasting in architectural design and retrofitting applications.

Chapter 2: Literature Review

In the global energy consumption, the building sector is one of the large contributors that accounts for a significant share of electricity use and greenhouse gas emissions. With a major aspect of the energy use tied to the heating and cooling systems (Mardiana and SB, 2015). These systems are important for the regulation of comfort indoors across the various seasons and climates. Heating and cooling loads in building energy use are important as the major energy consumers (the heating, ventilation, and air conditioning–HVAC system) help to prevent waste of energy and performance issues (Kassas, 2015, Wei et al., 2019; Tian et al., 2019). They help to improve the comfort of air quality indoors thereby preventing excessive humidity (Balali et al., 2023), allow for system efficiency and longevity, and save cost-through better insulation. This system will help to maintain occupants' thermal comfort while minimizing energy consumption.

Prior to new developments, traditional methods were about building energy performance using simulation-based tools such as DOE-2, EnergyPlus, and TRNSYS.

While they were good visualisation tools, and good for physical interpretation, they yielded complex inputs (Pan et al., 2023). This therefore led to the use of data-driven approaches–machine learning, to create improved models to predict the consumption of energy building (Tri et al., 2021). As described by International Business Machines (IBM), machine learning (ML), a subset of artificial intelligence (AI), is set to equip computers and machines with the ability to learn from experience, allowing them to perform tasks independently and enhance their accuracy and performance as they are exposed to more data. These ones automatically discover the relation between imputed features and output target–making them better equipped and useful for nonlinear and complex problems–like building energy prediction (Shapi et al., 2021). With respect to heating and cooling load estimation, ML offers some advantages. It uses different techniques or algorithms to enhance energy efficiency, reduce cost of operations, and support initiatives of green building. These algorithms developed to estimate energy usage patterns in buildings, include Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Trees and other statistical methods (Chou and Bui, 2014; Seyedzadeh et al., 2018; Pham et al., 2020).

ANN is among the most widely used algorithms for predicting building energy consumption, with various types (Ahmad et al., 2014, Amasyali and El-Gohary, 2018); and can also be combined with other methods to enhance performance (Amasyali and El-Gohary, 2018). In its operation in building each input feature of ANN, it is multiplied by its associated weight, summed with a bias term, and then passed through an activation function to produce the output. However, the result of the output (following a mathematical function), will help to predict building energy performance.

SVM is a kernel-based algorithm applicable to both regression and classification tasks. It is particularly effective in handling non-linear problems and considered one of the most accurate and robust machine learning algorithms. The SVM approach addresses non-linearity between input features and target outputs by mapping the data into a high-dimensional space, where it determines an optimal function. It then applies its kernel function to transform the complex non-linear relationships into a linear form (Chalal et al., 2016). However, to enhance its computational efficiency in energy consumption prediction, variations such as Least Squares SVM (LS-SVM) (Edward et al., 2012) and Parallel SVM (Zhao and Magoulès, 2010) have also been employed.

Decision tree algorithm uses a tree-like structure to map input instances to predict outcomes. In these models, each internal (non-leaf) node represents a specific feature, each branch corresponds to a possible value of that feature, and each leaf node signifies a prediction or class. Decision trees are highly adaptable and can expand as more training data becomes available (Domingos, 2012; Amasyali and El-Gohary, 2018).

Each of these algorithms used for developing energy consumption prediction models has their strengths and limitations. For instance, ANN and SVM deliver higher prediction accuracy compared to decision trees and statistical methods, but they require tuning of numerous parameters and can be computationally intensive. In contrast, decision trees and statistical algorithms are typically more user-friendly and less demanding in terms of computational resources, though their predictive performance tends to be moderate (Zhao and Magoulès, 2012; Amasyali and El-Gohary, 2018).

2.1 Summary of Existing Literatures

Several studies have reviewed data-driven methods for predicting building energy consumption. Zhao and Magoulès (2012) categorized models into ANN, SVM, statistical, and grey methods, comparing them on complexity, accuracy, and usability. Ahmad et al. (2014) reviewed ANN, SVM, and hybrid models, while Li et al. (2014) proposed a flowchart to guide method selection. Chalal et al. (2016) addressed both building- and urban-scale predictions, and Wang and Srinivasan (2016) compared individual and ensemble AI techniques.

While these reviews provide valuable insights, they often overlook specific prediction scopes (e.g., heating vs. cooling loads), the types of predictive features, and the nature of datasets used. This study aims to address these gaps by predicting building energy consumption—specifically heating and cooling loads—using machine learning models, with a focus on model comparison, feature relevance, and performance evaluation.Various studies have employed machine learning methods to forecast building energy consumption, with examples summarized in Table 2.1.1.

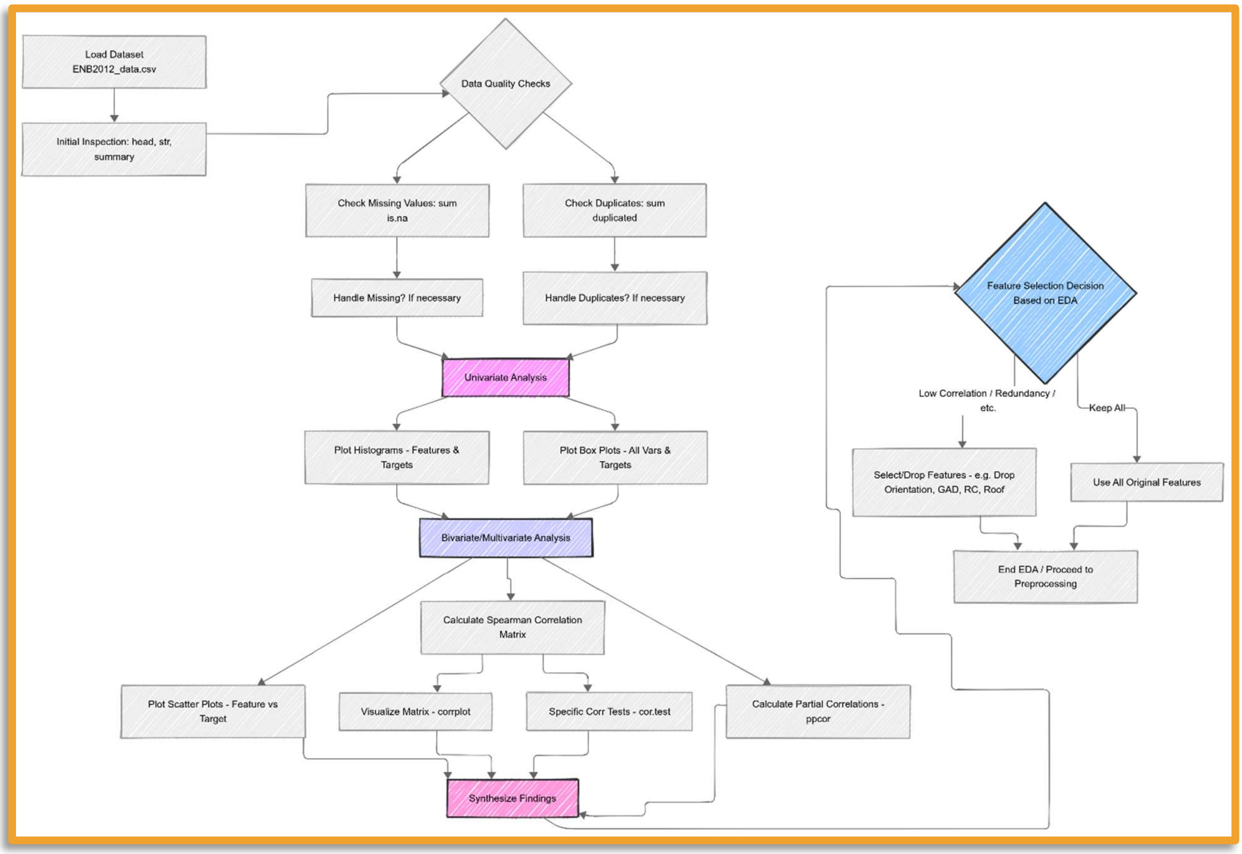
Table 2.1.1. Selected studies employing machine learning techniques to forecast building energy consumption.

References	Learning algorithm (type)	Building type	Type of energy consumption predicted	Sample size	Performance
Z. Dong, et al. 2021	Artificial Neural Network (ANN)	Non-Residential	Hourly load	507 instances	5.71 (RMSE)
Q. Li, et al. 2009	Support Vector Machine (SVM)	Non-Residential	Hourly load	1 Instance	1.17 (RMSE)
	Back Propagation Neural Network (BPNN)				2.22 (RMSE)
	the Radial Basis Function Neural Network (RBFNN)				1.43 (RMSE)
	General Regression Neural Network (GRNN)				1.19 (RMSE)
A.-D. Pham, et al. 2020	Random Forest (RF)	Non-Residential	Hourly load	5 Instances	5.53 (RMSE)
	M5 Model trees				6.09 (RMSE)
	Random tree (RT)				7.99 (RMSE)
B. Dong, C. Cao, S.E. Lee 2005	Support Vector Machine (SVM)	Non-Residential	Hourly load	4 instances	0.99 (R^2)
J.-S. Chou, D.-K. Bui 2014	Artificial Neural Network (ANN)	Residential	Hourly load	N/S	1.68 (RMSE)
	Support Vector Machine (SVM)				1.65 (RMSE)
	Decision Tree (DT)				1.84 (RMSE)
	General linear regression (GLR)				1.74 (RMSE)
R. Wang, S. Lu, W. Feng 2020	Stacking	Non-Residential	Hourly load	2 Instances	13.81 (RMSE)
	Random Forest				26.34 (RMSE)
	Decision Tree				19.20 (RMSE)
	Extreme Gradient Boosting				15.37 (RMSE)
	Support Vector Machine				16.12 (RMSE)
	K- Nearest Neighbour				17.81 (RMSE)

Chapter 3: Methodology

3.1 Introduction

This chapter details the methodology employed to predict building Heating Load (HL) and Cooling Load (CL) based on building design parameters. The objective was to develop and compare the performance of several machine learning regression models using the Energy Efficiency dataset provided by Tsanas and Xifara (2012).



3.2 Data Source and Description

The dataset used in this project is the "Energy efficiency Data Set" available from the UCI Machine Learning Repository (Dua & Graff, 2019), originally created by Tsanas and Xifara (2012). The dataset comprises 768 samples, representing various building shapes simulated using Ecotect software. It includes eight input features describing building parameters:

Variable Name	Role	Type	Description	Units	Missing Values
X1	Feature	Continuous	Relative Compactness	-	No
X2	Feature	Continuous	Surface Area	-	No
X3	Feature	Continuous	Wall Area	-	No
X4	Feature	Continuous	Roof Area	-	No
X5	Feature	Continuous	Overall Height	-	No
X6	Feature	Integer	Orientation	-	No
X7	Feature	Continuous	Glazing Area	-	No
X8	Feature	Integer	Glazing Area Distribution	-	No
Y1	Target	Continuous	Heating Load	-	No
Y2	Target	Continuous	Cooling Load	-	No

The goal is to predict two continuous target variables (considered as the dependent variable

)

1. Heating Load (HL, Y1)
2. Cooling Load (CL, Y2)

3.3 Data Preprocessing and Preparation

A systematic preprocessing pipeline was implemented to prepare the data for modeling, ensuring robustness and preventing data leakage.

3.3.1 Initial Loading and Cleaning

The dataset was loaded into the R environment from the ENB2012_data.csv file. Column names were programmatically renamed. Initial checks were performed to identify the presence of missing values (NA) and duplicate rows. In this analysis, no missing values were found, no duplicate rows were found.

3.3.2 Exploratory Data Analysis (EDA)

EDA was conducted to understand data distributions, relationships between variables, and potential outliers. Key techniques included:

Figure 3.1: First six rows of the data set

	Relative_Compactness	Surface_Area	Wall_Area	Roof_Area	Overall_Height	Orientation	Glazing_Area
1	0.98	514.5	294.0	110.25		7	2
2	0.98	514.5	294.0	110.25		7	3
3	0.98	514.5	294.0	110.25		7	4
4	0.98	514.5	294.0	110.25		7	5
5	0.90	563.5	318.5	122.50		7	2
6	0.90	563.5	318.5	122.50		7	3

	Glazing_Area_Distribution	Heating_Load	Cooling_Load
1	0	15.55	21.33
2	0	15.55	21.33
3	0	15.55	21.33
4	0	15.55	21.33
5	0	20.84	28.28
6	0	21.46	25.38

Column names were programmatically renamed to descriptive labels just as it is from the data descriptions

- *Histograms*: Plotted for target variables using ggplot2 to visualize their distributions.

Figure 3.2: Histograms showing the Distribution of Heating Load and Cooling Load

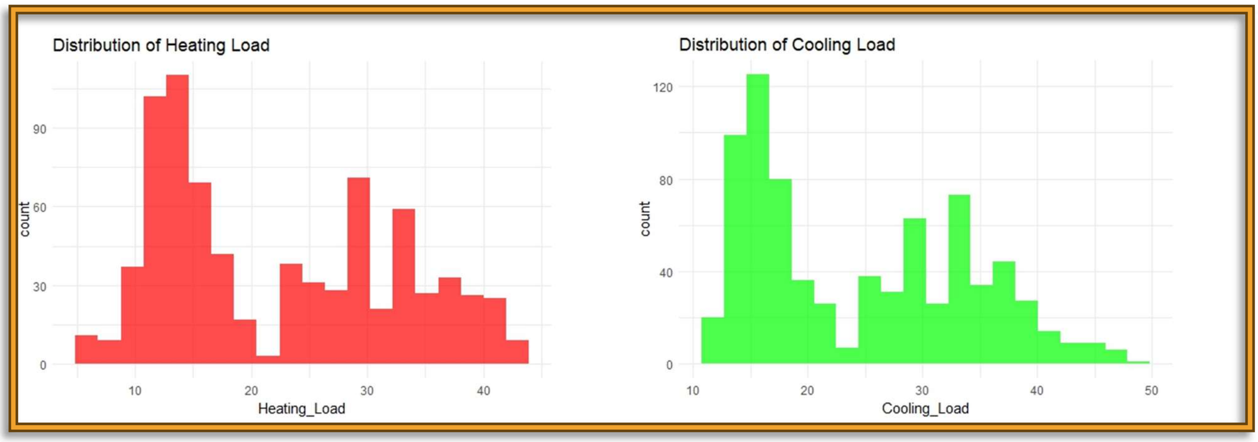


Figure 3.2 indicate that Heating Load and Cooling Load in this dataset are not uniformly distributed but instead form distinct clusters. Rather than reflecting a single, typical energy load, the data suggests multiple common performance levels influenced by variations in building design. This complexity highlights the need for non-linear machine learning models which are better equipped to capture these patterns and their intricate relationships with input features.

- *Box Plots*: Generated using ggplot2 and pivot_longer for all and specifically for the target variables (HL and CL) to identify central tendency, spread, and potential outliers.

Figure 3.3: Separate Box Plot for All variables

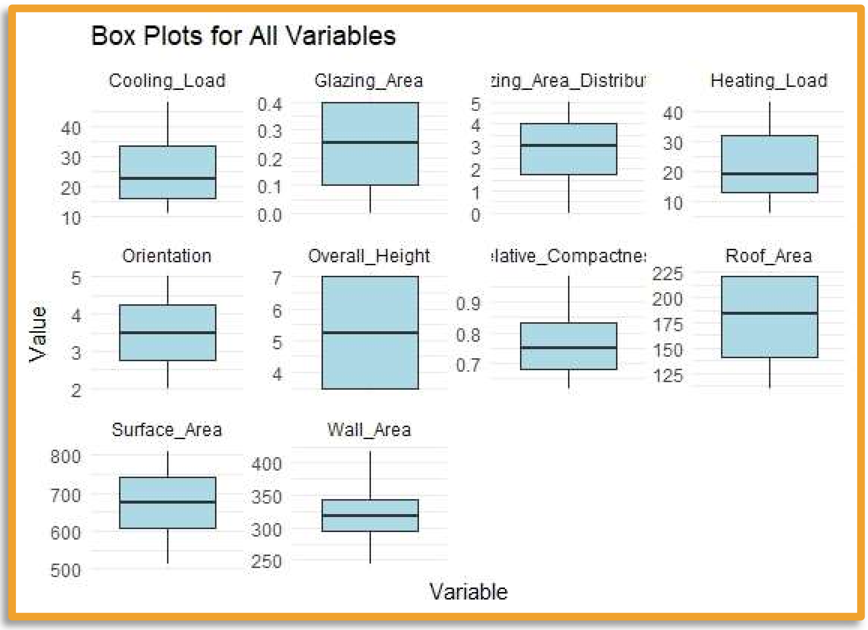


Figure 3.3 highlights the differing numerical scales across variables, emphasizing the need for feature scaling. This is particularly important for models like SVMs and ANNs, which are sensitive to variations in feature magnitudes.

Figure 3.4: Box Plots of Target Variables

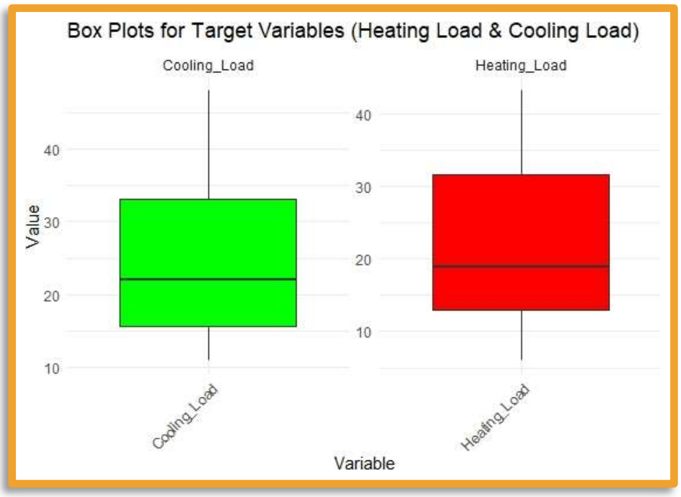


Figure 3.4 compares Cooling Load and Heating Load, revealing moderate variability in both as shown by their interquartile ranges (IQRs). This suggests that energy demands

vary with building design. The absence of outliers and the symmetry of both distributions indicate balanced data without significant skew or extreme values.

3.3.3 Correlation Analysis

To understand the monotonic relationships between the building design parameters and the target energy loads, as well as the relationships between the features themselves, a Spearman rank correlation analysis was conducted. Spearman correlation was chosen over Pearson correlation due to its robustness to potential non-linear relationships, which are common in building energy performance data, and because it evaluates monotonic relationships based on variable ranks rather than assuming linearity.

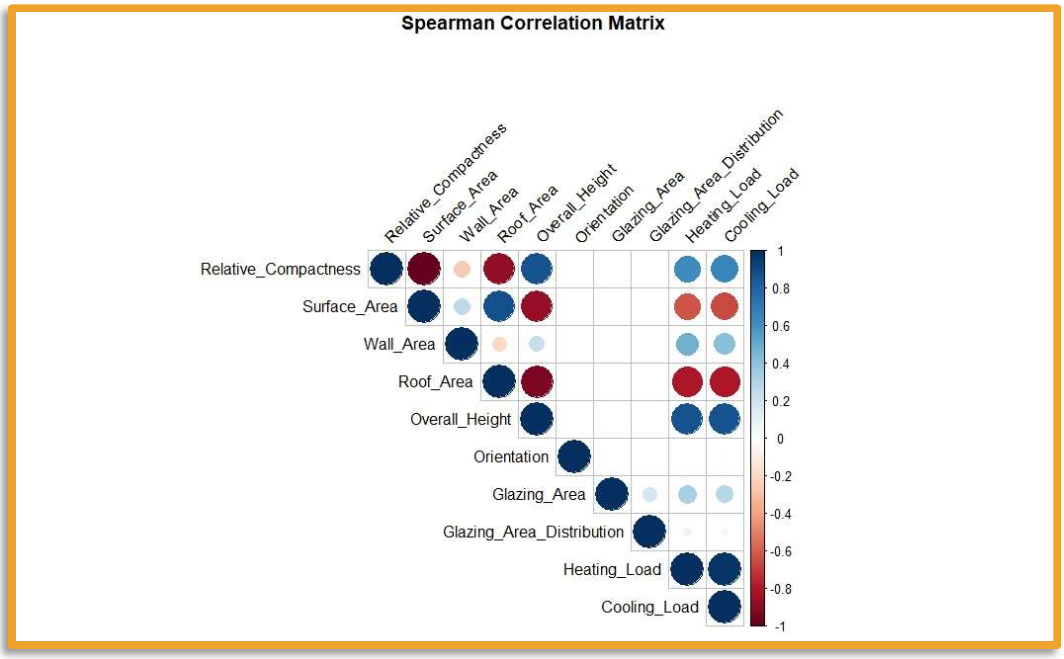
	Relative_Compactness	Surface_Area	Wall_Area	Roof_Area	Overall_Height
Relative_Compactness	1.000	-1.000	-0.256	-0.871	0.869
Surface_Area	-1.000	1.000	0.256	0.871	-0.869
Wall_Area	-0.256	0.256	1.000	-0.193	0.221
Roof_Area	-0.871	0.871	-0.193	1.000	-0.937
Overall_Height	0.869	-0.869	0.221	-0.937	1.000
Orientation	0.000	0.000	0.000	0.000	0.000
Glazing_Area	0.000	0.000	0.000	0.000	0.000
Glazing_Area_Distribution	0.000	0.000	0.000	0.000	0.000
Heating_Load	0.622	-0.622	0.471	-0.804	0.861
Cooling_Load	0.651	-0.651	0.416	-0.803	0.865

	Orientation	Glazing_Area	Glazing_Area_Distribution	Heating_Load
Relative_Compactness	0.00000	0.000	0.0000	0.62213
Surface_Area	0.00000	0.000	0.0000	-0.62213
Wall_Area	0.00000	0.000	0.0000	0.47146
Roof_Area	0.00000	0.000	0.0000	-0.80403
Overall_Height	0.00000	0.000	0.0000	0.86128
Orientation	1.00000	0.000	0.0000	-0.00416
Glazing_Area	0.00000	1.000	0.1876	0.32286
Glazing_Area_Distribution	0.00000	0.188	1.0000	0.06834
Heating_Load	-0.00416	0.323	0.0683	1.00000
Cooling_Load	0.01761	0.289	0.0465	0.97269

	Cooling_Load
Relative_Compactness	0.6510
Surface_Area	-0.6510
Wall_Area	0.4160
Roof_Area	-0.8032
Overall_Height	0.8649
Orientation	0.0176
Glazing_Area	0.2889
Glazing_Area_Distribution	0.0465
Heating_Load	0.9727
Cooling_Load	1.0000

This analysis is visualized in the correlation matrix plot presented in Figure 3.5. This plot uses circles whose color and size represent the direction (blue=positive, red=negative) and magnitude (larger=stronger) of the Spearman correlation coefficient (rho), respectively.

Figure 3.5: Correlation Matrix (Spearman) of all variables



Several key observations were derived from this analysis:

1. *Target Variable Inter-correlation*: As expected, Heating Load and Cooling Load demonstrated a very strong positive correlation with each other (rho = 0.97), indicating that buildings requiring high heating loads generally also require high cooling loads within this dataset's context.

2. *Strong Correlations with Target Loads*: Significant correlations were observed between several geometric features and the energy loads:

   - Strong positive correlations were found between both HL and CL and Overall_Height and Relative_Compactness. This suggests taller and more compact buildings in this dataset tend to have higher energy demands.

   - Strong negative correlations were observed between both HL and CL and Surface_Area and Roof_Area

3. *Weak Correlations with Target Loads:* Of particular relevance to subsequent feature selection steps, certain variables exhibited negligible relationships with the energy loads:

   - Orientation: Showed a very weak, near-zero correlation with both Heating Load and Cooling Load. The corresponding cor.test results confirmed this lack of significant monotonic relationship (p-value > 0.05).

```
                Spearman's rank correlation rho

data:  energy_data$Heating_Load and energy_data$Orientation
S = 75811645, p-value = 0.9083
alternative hypothesis: true rho is not equal to 0
sample estimates:
         rho
-0.004163071
```

- Glazing_Area_Distribution: Similarly displayed almost no correlation with either HL or CL, also confirmed by cor.test results (p-value > 0.05).

```
            Spearman's rank correlation rho

data:  energy_data$Heating_Load and
energy_data$Glazing_Area_Distribution
S = 70337594, p-value = 0.05834
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
0.06834346
```

- Glazing_Area: Showed only a weak positive correlation with both HL and CL. While present, this relationship was considerably weaker than those observed for height, compactness, or surface/roof area.

```
                Spearman's rank correlation rho


data:  energy_data$Heating_Load and energy_data$Glazing_Area
S = 51122247, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.3228603
```

3.3.4 Partial Correlation Analysis:

Significant relationships between predictor variables were also noted, indicating potential multicollinearity. The most prominent was the extremely strong negative correlation between Relative_Compactness and Surface_Area, which is expected given their definitions. Strong negative correlations also existed between Roof_Area and features like Overall_Height and Relative_Compactness. The ppcor package was used to investigate the correlation between two variables while controlling for the effect of a third variable.

Relative Compactness vs. HL (Controlling for Surface Area):

```
estimate          p.value   statistic    n  gp   Method
```

```
-0.3191372    1.280388e-19   -9.313939  768    1  pearson
```

Surface Area vs. HL (Controlling for Relative Compactness):

```
  estimate          p.value     statistic     n    gp       Method
-0.4112467    1.163483e-32    -12.47857    768     1      pearson
```

With a partial correlation estimate of -0.4112, Surface Area has a stronger and more significant unique linear relationship with Heating Load than Relative Compactness, when controlling for the other variable. This suggests that, although both features are related to Heating Load, Surface Area contributes more unique variance after accounting for the effect of Relative Compactness.

Roof Area vs. HL (Controlling for Overall Height):

```
   estimate       p.value     statistic     n     gp    Method
0.02963426    0.4124697     0.8200032   768      1    pearson
```

When controlling for Overall_Height, the partial correlation between Heating_Load and Roof_Area became statistically insignificant (r = 0.03, p = 0.41). This suggests that the strong negative relationship observed between Roof Area and Heating Load in the initial Spearman matrix might be largely explained or mediated by their mutual association with Overall Height.

Overall Height vs. HL (Controlling for Roof Area):

```
   estimate          p.value   statistic      n    gp    Method
 0.4343021    1.254328e-36   13.33552    768     1    pearson
```

Conversely, the partial correlation between Heating_Load and Overall_Height, after controlling for Roof_Area, remained strongly positive and highly significant.

This implies Overall Height is a much more important predictor of Heating Load than Roof Area. Roof Area contributes almost no unique information about Heating Load once Overall Height is considered.

3.3.5 Feature Selection (FS)

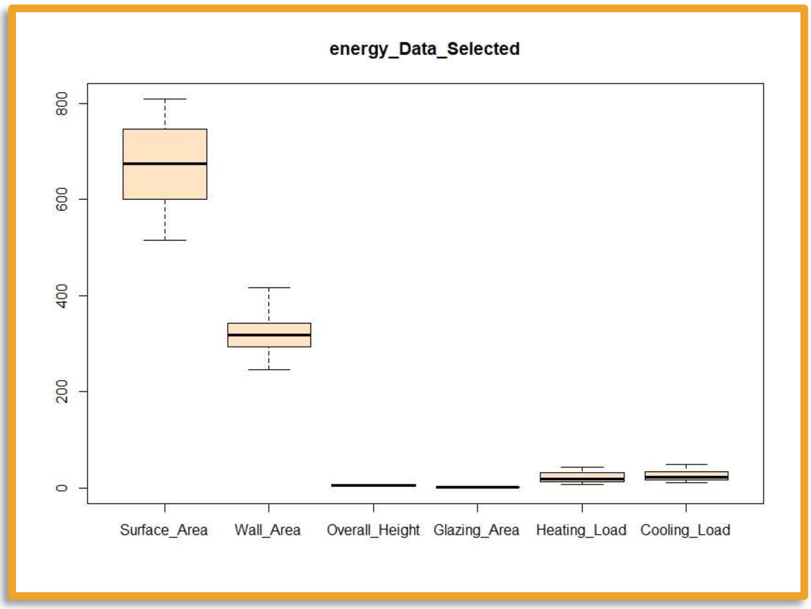
This plays a critical role in balancing the accuracy and complexity of predictive models. It is a key step in the development of machine learning models, as not all features contribute meaningfully to the model's performance. FS helps identify and eliminate irrelevant or insignificant features, enhancing model efficiency (Zhao, H. and Magoulès, F., 2012.). Applying feature selection is essential to achieving optimal model performance (H.A. Alaka, *et al.* 2018).

A feature selection step was undertaken to simplify the models, potentially improve performance by removing irrelevant or redundant predictors, and reduce multicollinearity issues.

The following features were selected for removal from the dataset prior to model training:

1. *Orientation:* This feature exhibited a near-zero Spearman correlation coefficient with both Heating Load (HL) and Cooling Load (CL). Statistical tests (cor.test) confirmed the lack of a significant monotonic relationship ($p > 0.05$).
2. *Glazing_Area_Distribution:* Similar to Orientation, this feature showed almost no Spearman correlation with either HL or CL, and the relationship was statistically insignificant ($p > 0.05$).
3. *Relative_Compactness:* initially showed strong correlations with Heating and Cooling Loads but had an extremely high negative correlation with Surface_Area. This level of multicollinearity can disrupt model stability and interpretability. Therefore, Relative_Compactness was removed, with Surface_Area retained as the more reliable geometric predictor.
4. *Roof_Area:* This feature displayed a significant negative Spearman correlation with both HL and CL. However, the partial correlation analysis provided critical context. This suggests that the predictive information contained in Roof_Area regarding energy load is largely captured by Overall_Height. Therefore, Roof_Area was removed to reduce redundancy and model complexity.

Figure 3.6: Box plot of the selected Variables Dataset



3.3.6 Data Splitting

To ensure an unbiased evaluation of model performance on unseen data, the preprocessed dataset was split into training (80%) and testing (20%) sets *before* any scaling was applied.
The caret::createDataPartition function was utilized, this resulted in 614 samples for training and 154 samples for testing.

3.3.7 Feature and Target Scaling

Machine learning models such as Support Vector Machines and Neural Networks are sensitive to feature scales, so Z-score standardization was applied. To avoid data leakage, the mean and standard deviation were calculated only from the training set and then used to scale both training and test data.

Figure 3.7: Box plot of Scaled Data Set

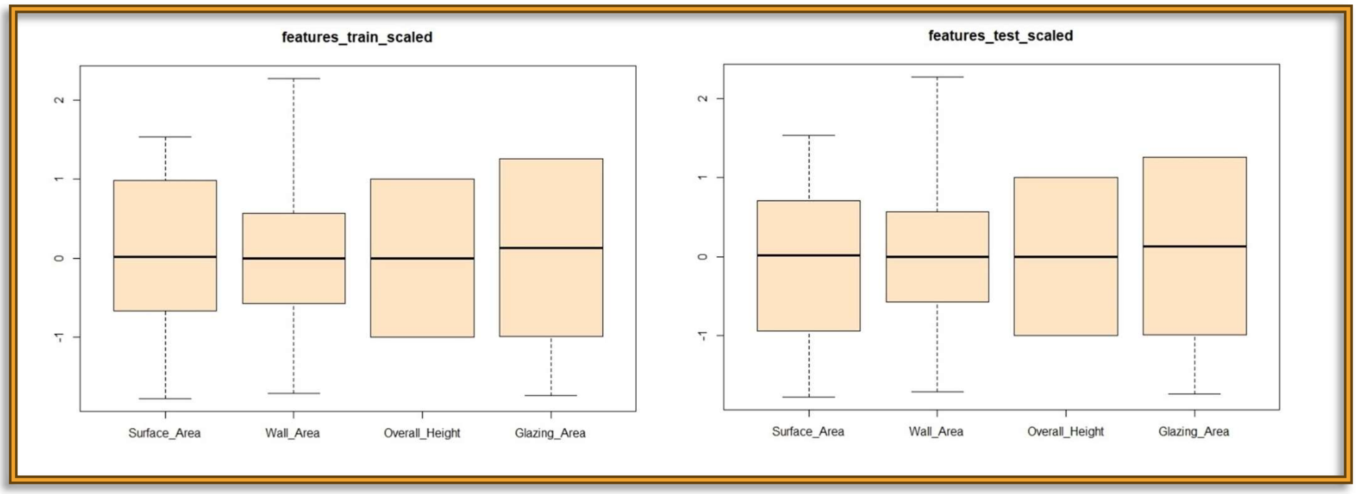


Figure 3.8: Box plot of Scaled Features

The effect of this scaling is visualized in Figure 3.7 and Figure 3.8.

Figure 3.8 provides a direct comparison of the distributions of the scaled features between the training set and the test set. The plots demonstrate that for each feature, the distributions are highly comparable between the two sets after scaling using the training set parameters.

3.3.8 Target Variable Scaling (for Neural Network)

The neuralnet package, used for the ANN implementation, often benefits from having the target variables scaled, particularly for stable gradient descent during training. Therefore, a similar Z-score standardization process was applied to Dependent variables, again fitting the scalers *only* on the training set target values.

The resulting scaled training targets were used for training the neuralnet models.

```
HL_scaled          CL_scaled
 Min.   :-1.6164   Min.   :-1.4400
 1st Qu.:-0.9224   1st Qu.:-0.9411
 Median :-0.3303   Median :-0.2602
 Mean   : 0.0000   Mean   : 0.0000
 3rd Qu.: 0.9337   3rd Qu.: 0.8846
 Max.   : 2.0700   Max.   : 2.4784
```

The summary statistics confirmed that the scaled training targets have a mean of approximately zero and are scaled appropriately. The scaling parameters (mean and standard deviation) for both target variables were stored. These parameters are essential for the inverse transformation (unscaling) of the neuralnet predictions back to their original physical units during the model evaluation phase.

3.4 Modeling Approaches

Four different regression algorithms were trained and evaluated for predicting both HL and CL independently. The caret package was used to streamline the training and tuning process for the first three models, employing 10-fold cross-validation for robust performance estimation during training and basic hyperparameter tuning. The neuralnet package was used for the ANN.

3.4.1 Decision Tree (DT)

A single Decision Tree model was trained using the rpart method within caret::train. caret automatically tuned the complexity parameter (cp) using tuneLength = 10.

Figure 3.9: Decision Tree Model for Heating Load (HL)

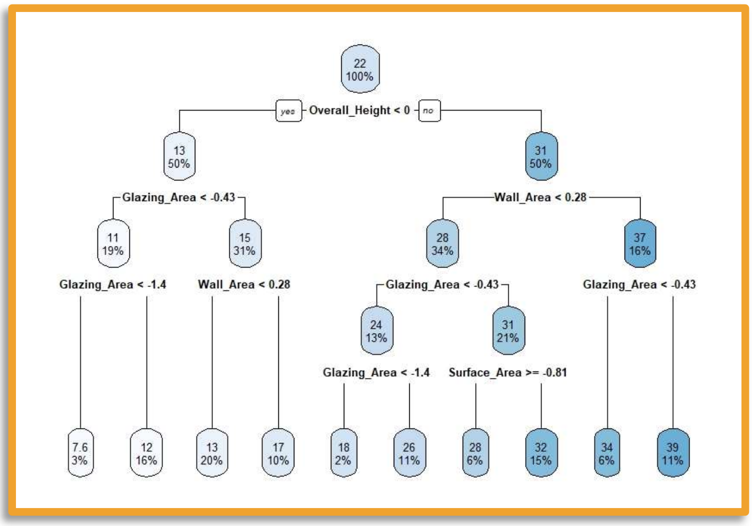
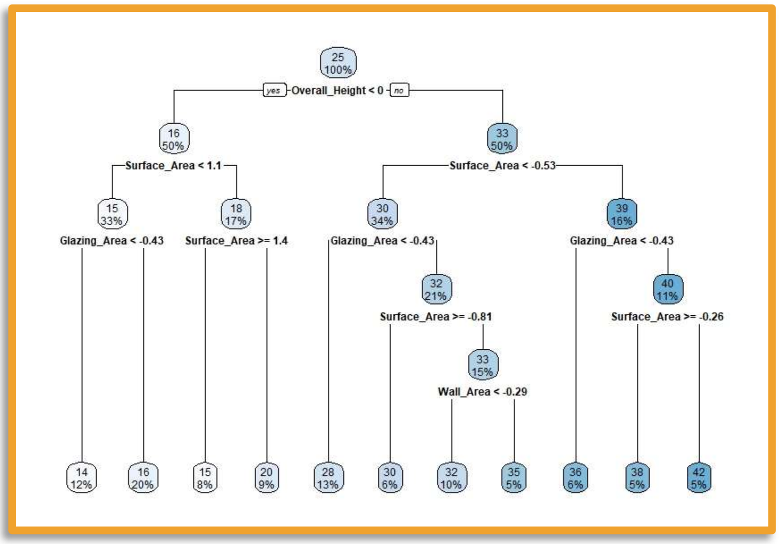


Figure 3.9.1: Decision Tree Model for Cooling Load (CL)



Figures 3.9 and 3.9.1 show pruned decision trees for predicting Heating Load and Cooling Load using the rpart algorithm. Overall_Height is the primary splitting variable, highlighting its major role in energy load prediction. Additional splits on Surface_Area, Glazing_Area, and Wall_Area show their secondary importance. The tree structures capture interactions among these features, and the terminal nodes represent groups of buildings with varying energy performance.

3.4.2 Random Forest (RF)

A Random Forest model, an ensemble method based on multiple decision trees, was trained using the rf method within caret::train. The number of variables randomly sampled at each split was tuned. Variable importance scores were calculated.

Figure 3.2.2: Random Forest Model for Heating Load (HL)

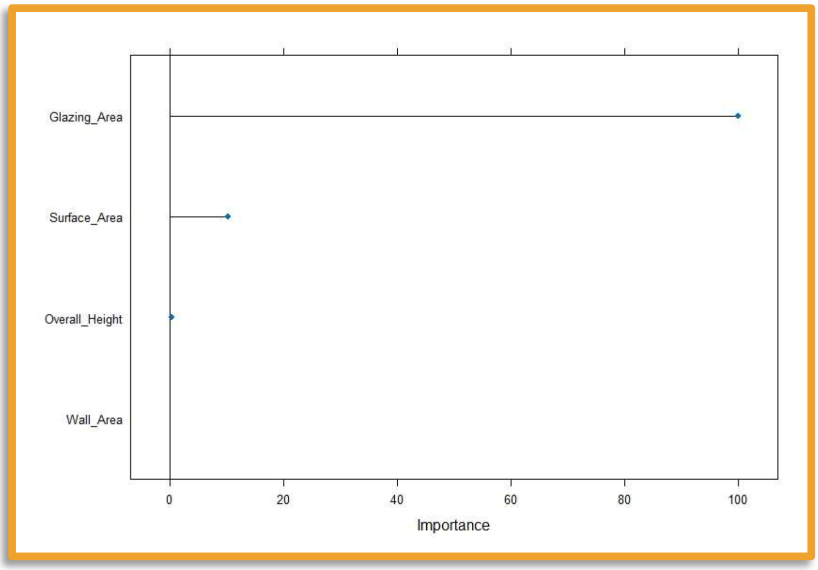
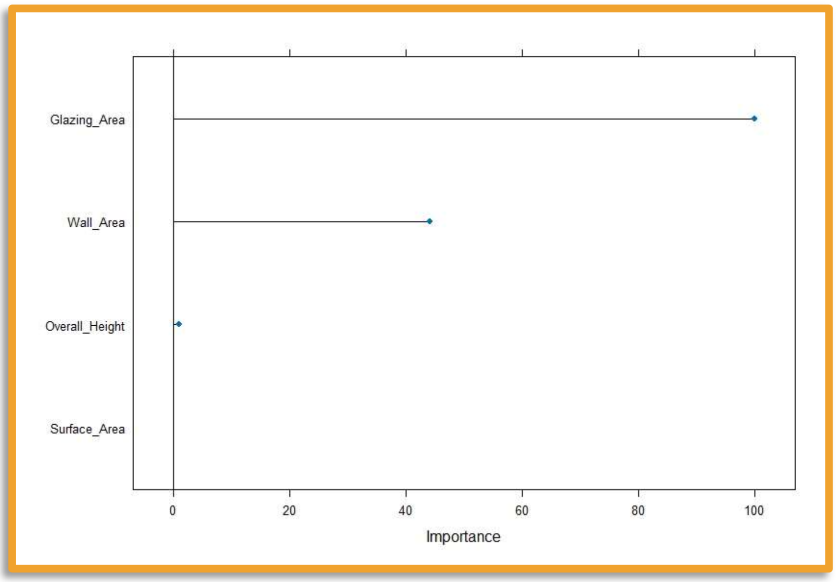


Figure 3.2.3: Random Forest Model for Cooling Load (CL)



Figures 3.2.2 and 3.2.3 show that in both Random Forest models for Heating Load and Cooling Load, **Glazing_Area** is the most important predictor, with an importance score close to 100. For Heating Load, **Surface_Area** is the second most important feature, while for Cooling Load, **Wall_Area** ranks second. Other features contribute significantly less to the models.

3.4.3 Support Vector Machine (SVM)

A Support Vector Machine model with a non-linear Radial Basis Function (RBF) kernel was trained using the svmRadial method within caret::train. Key hyperparameters, the cost parameter (C) and the RBF kernel parameter (sigma), were tuned via tuneLength = 5.

3.4.4 Artificial Neural Network (ANN)

An Artificial Neural Network (ANN) model was developed using R's native neuralnet package. The model used Z-score scaled input and target data, with a feed-forward architecture comprising two hidden layers of 6 and 4 neurons. A linear activation function was applied in the output layer to suit the regression task. Training employed either the resilient backpropagation (rprop+) or standard backpropagation algorithm, with key parameters including a maximum of 1 million iterations and a convergence threshold of 0.05.

3.5 Model Evaluation

The performance of the trained models (DT, RF, SVM, ANN) was evaluated on the held-out, unseen test set (features_test_scaled and the original target_hl_test, target_cl_test).

3.5.1 Prediction Generation

Predictions were generated for both HL and CL using each model's respective predict function (for caret models) or neuralnet::compute function (for the ANN model) applied to the scaled test features (features_test_scaled).

3.5.2 Inverse Transformation (Unscaling)

Since the ANN model was trained on scaled targets and produced scaled predictions, these predictions (nn_pred_scaled_hl, nn_pred_scaled_cl) were transformed back to the original scale of Heating Load and Cooling Load using the stored means and standard deviations (scale_params_hl, scale_params_cl) before evaluation. This ensures comparability with the predictions from other models and the original target values.

3.5.3 Performance Metrics

The following standard regression metrics were calculated to quantify model performance by comparing the (unscaled) predictions against the actual test set target values (target_hl_test, target_cl_test):

- *Root Mean Squared Error (RMSE):* Measures the square root of the average squared difference between predicted and actual values, sensitive to large errors (calculated via Metrics::rmse). Lower is better.
- *Mean Absolute Error (MAE):* Measures the average absolute difference between predicted and actual values, providing an error magnitude in the original units (calculated via Metrics::mae). Lower is better.
- *R-squared (R^2):* Represents the proportion of the variance in the dependent variable that is predictable from the independent variables. Calculated manually for robustness as 1 - SSE/SST. Higher (closer to 1) is better.

These metrics were calculated separately for Heating Load and Cooling Load predictions for each of the four models.

Chapter 4: Results and Discussion

4.2 Model Performance Evaluation

The core quantitative results comparing the predictive performance of the four models on the held-out test set are summarized below. Lower RMSE and MAE values indicate better accuracy (smaller prediction errors), while higher R^2 values indicate a better fit (more variance explained).

4.2.1 Heating Load (HL) Prediction Performance

Table 4.1 presents the evaluation metrics for predicting Heating Load.

Table 4.1: Performance Metrics for Heating Load (HL) Prediction on Test Set

Model	RMSE	MAE	R^2
Decision Tree	2.148	1.621	0.956
Random Forest	0.598	0.399	0.997
Support Vector Machine	0.786	0.638	0.994
NeuralNet	0.583	0.416	0.997

4.2.2 Cooling Load (CL) Prediction Performance

Table 4.2 presents the evaluation metrics for predicting Cooling Load.

Table 4.2: Performance Metrics for Cooling Load (CL) Prediction on Test Set

Model	RMSE	MAE	R^2
Decision Tree	2.376	1.808	0.940
Random Forest	1.734	1.071	0.968
Support Vector Machine	2.123	1.373	0.952
NeuralNet	1.703	1.127	0.969

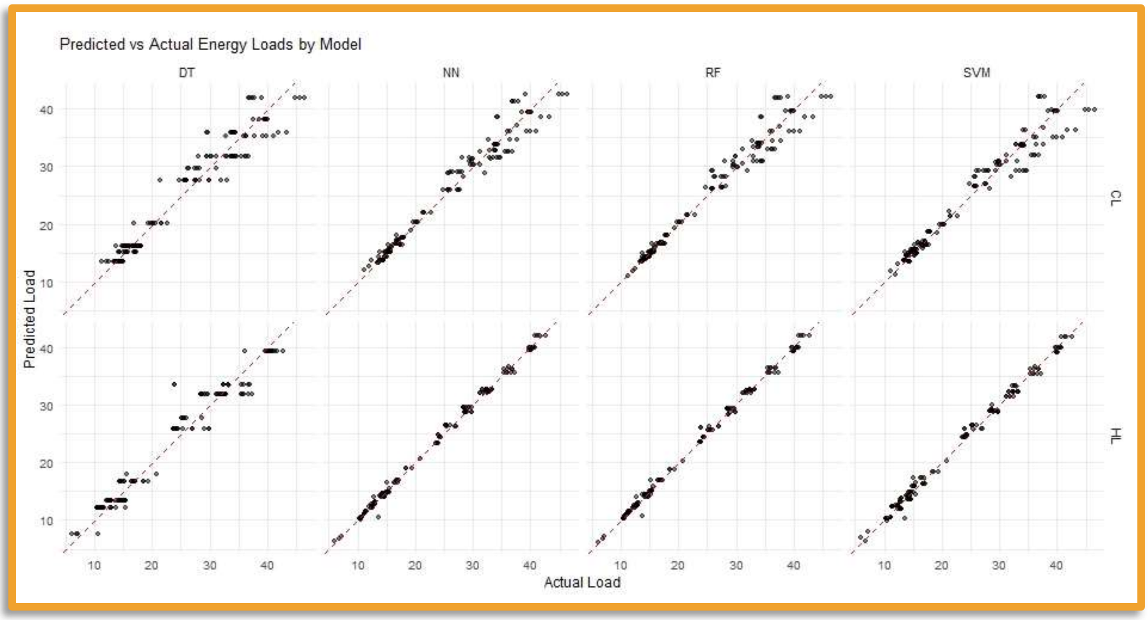
4.3 Visual Analysis of Model Predictions

To complement the quantitative metrics, visualizations were generated to assess the models' predictive behaviour.

4.3.1 Predicted vs. Actual Values

Figure 4.1 displays scatter plots comparing the predicted HL and CL values against the actual values from the test set for each of the four models. The red dashed line represents the ideal scenario where predicted equals actual (y=x). Points clustering tightly around this diagonal line indicate higher accuracy.

Figure 4.1: Predicted vs. Actual Heating Load (HL) and Cooling Load (CL) for All Models on the Test Set.



4.3.2 Variable Importance (Random Forest)

The Random Forest models provide a measure of variable importance, indicating the relative contribution of each input feature to the model's predictive accuracy. Figures 4.2 and 4.3 [Adjust figure numbers] illustrate the feature importance scores for predicting HL and CL, respectively.

Figure 4.2: Random Forest Variable Importance for Heating Load (HL) Prediction.

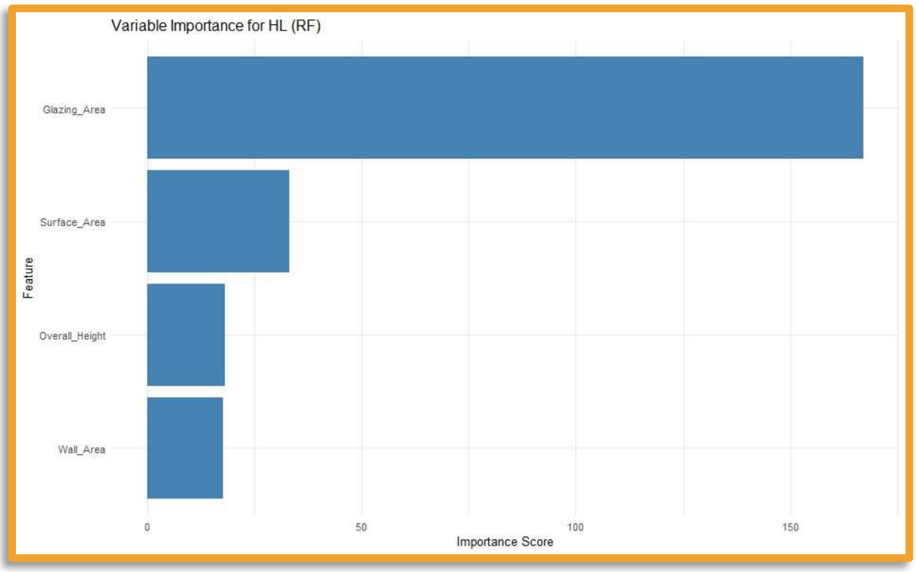
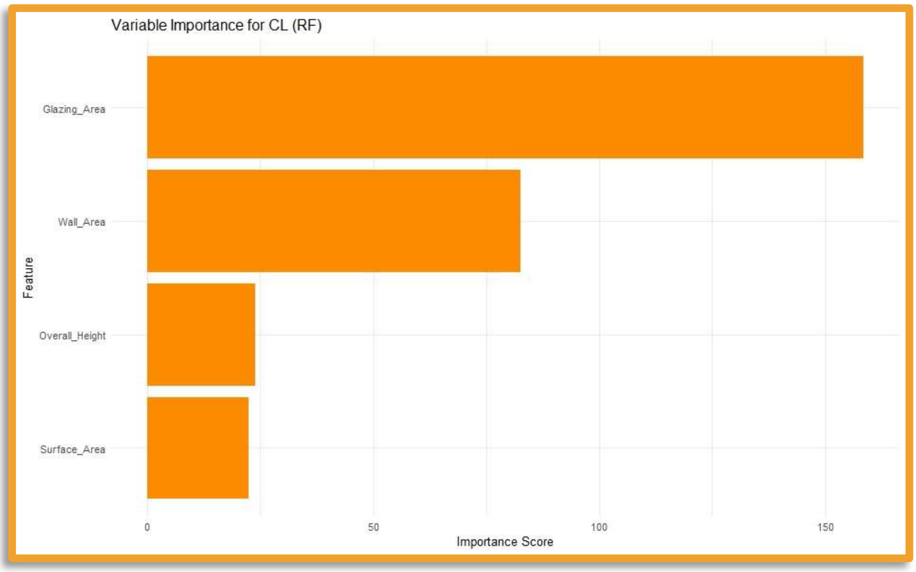


Figure 4.3: Random Forest Variable Importance for Cooling Load (CL) Prediction.



4.3.3 Neural Network Structure (NeuralNet)

Figures 4.4 and 4.5 show the structure of the trained neuralnet models for HL and CL, visualizing the input nodes, hidden layers/neurons, output node, and the connection weights.

Figure 4.4: Neural Network Structure (neuralnet) for Heating Load (HL) Prediction.

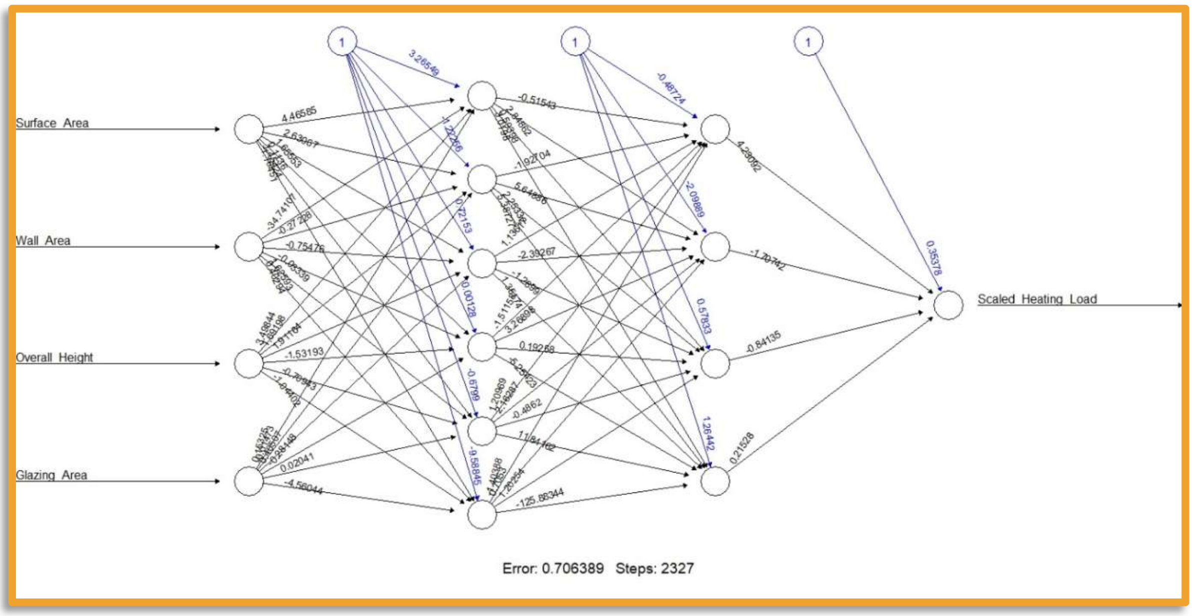
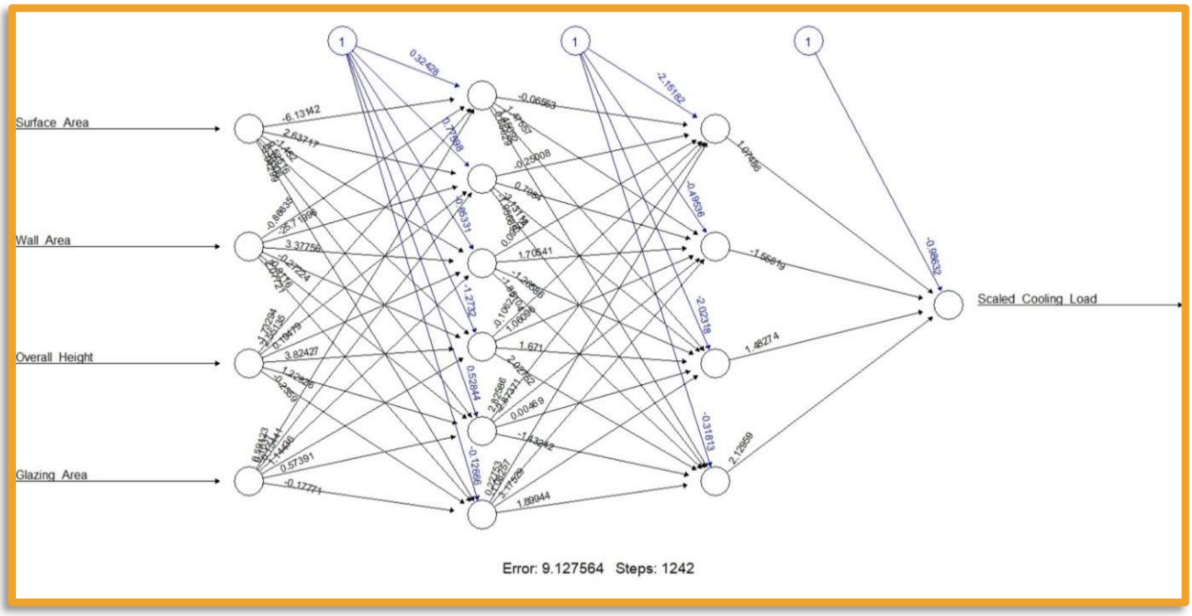


Figure 4.5: Neural Network Structure (neuralnet) for Cooling Load (CL) Prediction.



4.4 Discussion

4.4.1 Comparative Model Performance

Model evaluation revealed clear differences in predictive performance. For **Heating Load**, the Neural Network (RMSE = 0.583, R^2 = 0.997) and Random Forest (RMSE = 0.598, R^2 = 0.997) achieved the highest accuracy, closely followed by SVM (RMSE = 0.786, R^2 = 0.994). The Decision Tree showed the weakest performance (RMSE = 2.148, R^2 = 0.956).

A similar pattern emerged for **Cooling Load**, though overall accuracy was slightly lower. The Neural Network again led (RMSE = 1.703, R^2 = 0.969), with Random Forest (RMSE = 1.734, R^2 = 0.968) and SVM (RMSE = 2.123, R^2 = 0.952) performing well. The Decision Tree model again had the highest error (RMSE = 2.376, R^2 = 0.940).

These results highlight the strength of ensemble and non-linear methods in capturing complex relationships in building data. While ANN showed the best accuracy, Random Forest offered a strong balance of performance and efficiency, training faster with minimal tuning. Future work should consider computational cost and model interpretability for real-world deployment.

4.4.2 Interpretation of Predicted vs. Actual Plots

The Predicted vs. Actual scatter plots Figure 4.1 visually corroborate the metrics. The points for the Random Forest (RF), SVM, and NeuralNet (NN) models cluster very tightly around the y=x diagonal for both HL and CL, indicating excellent prediction accuracy across the range of observed loads. In contrast, the Decision Tree (DT) plots show visibly more scatter and potentially some step-like patterns, consistent with its higher error metrics.

4.4.3 Key Predictive Features

The variable importance plots from the Random Forest models (Figures 4.2 and 4.3) indicate that Glazing_Area was the most influential predictor for both HL and CL within the RF framework. For Heating Load, Surface_Area ranked second, followed by Wall_Area and Overall_Height having minimal importance according to RF. For Cooling Load, Wall_Area ranked second, followed by Surface_Area and Overall_Height. This suggests that after the dominant effect of windows, overall surface area is slightly more critical for HL prediction (heat loss), while wall area is slightly more critical for CL prediction (heat gain) in the RF models' decision process.

4.4.4 Neural Network (neuralnet) Performance and Considerations

The neuralnet package successfully implemented an ANN that achieved top-tier predictive performance, yielding the lowest RMSE for both HL and CL in this evaluation. The training times were relatively short (5 seconds for HL, 4 seconds for CL) on this dataset with the specified architecture (6-4 hidden neurons) and convergence parameters. The structure plots (Figures 4.4, 4.5) confirm the network topology but offer limited direct interpretability regarding feature importance compared to RF. The necessity of scaling target variables adds a step compared to the caret models but was handled effectively.

4.5 Limitations

This study has several limitations:

1. Dataset Scope: The analysis is based on a specific dataset generated via simulation, which may not fully capture the variability and complexities of real-world building energy consumption.
2. Model Selection: Only four specific algorithms were compared. Other potentially powerful methods like Gradient Boosting Machines (XGBoost, LightGBM) were not included.
3. neuralnet Implementation: The chosen neuralnet package, while R-native, lacks the performance optimizations (e.g., GPU support) and advanced features (e.g., diverse layer types, optimizers, callbacks) available in frameworks like Keras/TensorFlow or Torch. This may limit its performance ceiling on complex tasks.
4. No External Validation: The models were evaluated only on a single train-test split of the original dataset. Validation on a completely independent dataset would provide stronger evidence of generalizability.

4.6 Future Work

Building upon this study, future work could explore:

1. Advanced Hyperparameter Optimization: Employ techniques like Grid Search, Random Search, or Bayesian Optimization using caret or specialized packages (mlr3, tidymodels, specific model packages) for potentially improving all models.
2. Alternative Algorithms: Include Gradient Boosting models (XGBoost, LightGBM) known for high performance on tabular data. Explore other ANN implementations like h2o.
3. Feature Engineering: Create interaction terms or polynomial features to potentially capture more complex relationships.
4. Ensemble Methods: Combine predictions from multiple models (e.g., stacking) to potentially achieve better performance than any single model.
5. Interpretability Techniques: Apply model-agnostic methods (e.g., SHAP, LIME) to better understand predictions, especially for the less interpretable models like SVM and ANN.
6. External Validation: Test the final selected model(s) on different building energy datasets.

4.7 Conclusion

This project successfully developed and compared DT, RF, SVM, and ANN models for predicting Heating and Cooling Loads in buildings. NeuralNet and Random Forest models achieved the highest accuracy, with SVM also performing well. The Decision Tree model lagged in accuracy. Feature importance analysis in RF identified *Glazing Area* as the most influential predictor, confirming the value of geometric design features in load forecasting.

While the models demonstrated strong predictive power, a key limitation was the use of default or basic hyperparameter settings, particularly for SVM and ANN. Future work should explore optimization methods such as grid search or Bayesian tuning to enhance performance. Additionally, the analysis was based on a static dataset; incorporating climate variability and real-world usage patterns could improve generalizability.

Model interpretability is another area for improvement. While Random Forests offered insight into feature importance, black-box models like ANNs would benefit from post-hoc techniques such as SHAP or LIME to clarify feature influence and increase transparency. Integrating these approaches will support more informed decision-making in energy-efficient building design.

5 References

HaiXiang Zhao, F. Magoulès

Feature selection for predicting building energy consumption based on statistical learning method

J. Algorithm Comput. Technol., 6 (1) (2012), pp. 59-77, 10.1260/1748-3018.6.1.59

H.A. Alaka, *et al.*

Systematic review of bankruptcy prediction models: towards a framework for tool selection

Expert Syst. Appl., 94 (2018), pp. 164-184, 10.1016/j.eswa.2017.10.040

Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F., Abdullah, H. Saidur, R. (2014) *A review on applications of ANN and SVM for building electrical energy consumption forecasting.* Renewable and Sustainable Energy Reviews, 33, pp.102–109. http://dx.doi.org/10.1016/j.rser.2014.01.069

Amasyali, K. & El-Gohary, N.M. (2018) A *review of data-driven building energy consumption prediction studies.* Renewable and Sustainable Energy Reviews, 81, pp.1192–1205. http://dx.doi.org/10.1016/j.rser.2017.04.095

Balali, Y., Chong, A., Busch, A. & O'Keefe, S. (2023) *Energy modelling and control of building heating and cooling systems with data-driven and hybrid models—A review.* Renewable and Sustainable Energy Reviews, 183, 113496. https://doi.org/10.1016/j.rser.2023.113496

Chalal, M.L., Benachir, M., White, M. & Shrahily, R. (2016) *Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: a review.* Renewable and Sustainable Energy Reviews, 64, pp.761–776. http://dx.doi.org/10.1016/j.rser.2016.06.040

Chou, J.-S. & Bui, D.-K. (2014) *Modeling heating and cooling loads by artificial intelligence for energy-efficient building design.* Energy and Buildings, 82, pp.437–446.

Domingos, P. (2012) A *few useful things to know about machine learning.* Communications of the ACM, 55, pp.78–87. http://dx.doi.org/10.1145/2347736.2347755

Edwards, R.E., New, J. & Parker, L.E., 2012. Predicting future hourly residential electrical consumption: a machine learning case study. Energy and Buildings, 49, pp.591–603. http://dx.doi.org/10.1016/j.enbuild.2012.03.010

IBM (2023) *Machine learning*. [online] IBM. Available at: https://www.ibm.com/think/topics/machine-learning [Accessed: 16 April 2025].

Kassas, M. (2015) *Modeling and simulation of residential HVAC systems energy consumption.* Procedia Computer Science, 52, pp.754–763. https://doi.org/10.1016/j.procs.2015.05.123

Kaur, K. & Gupta, O.P. (2017) A *machine learning approach to determine maturity stages of tomatoes*. Oriental Journal of Computer Science and Technology, 10(3), pp.683–690.

Li, W., Zhang, J. & Zhao, T. (2019) *Indoor thermal environment optimal control for thermal comfort and energy saving based on online monitoring of thermal sensation*. Energy and Buildings, 197, pp.57–67. https://doi.org/10.1016/j.enbuild.2019.05.050

Li, Z., Han, Y. & Xu, P. (2014) *Methods for benchmarking building energy consumption against its past or intended performance: an overview.* Applied Energy, 124, pp.325–334. http://dx.doi.org/10.1016/j.apenergy.2014.03.020

Mardiana, A., SB, R. (2015) *Building energy consumption and carbon dioxide emissions: Threat to climate change.* Journal of Earth Science & Climatic Change, S3, pp.1-5. https://doi.org/10.4172/2157-7617.S3-001

Ngo, N.-T., Pham, A.-D., Truong, T., Truong, N.-S., Huynh, N.-T. & Pham, T. (2021) *An ensemble machine learning model for enhancing the prediction accuracy of energy consumption in buildings.* Arabian Journal for Science and Engineering, 47, pp.1–13. https://doi.org/10.1007/s13369-021-05927-7

Pan, Y., Zhu, M., Lv, Y., Yang, Y., Liang, Y., Yin, R., Yang, Y., Jia, X., Wang, X., Zeng, F., Huang, S., Hou, D., Xu, L., Yin, R. & Yuan, X. (2023) *Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies. Advances in Applied Energy,* 10, 100135. https://doi.org/10.1016/j.adapen.2023.100135

Pham, A.-D., Ngo, N.-T., Truong, T.T.H., Huynh, N.-T. & Truong, N.-S. (2020) *Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability.* Journal of Cleaner Production, 260, 121082. https://doi.org/10.1016/j.jclepro.2020.121082

Santamouris, M. & Vasilakopoulou, K. (2021) *Present and future energy consumption of buildings: Challenges and opportunities towards decarbonisation.* e-Prime - Advances in Electrical Engineering, Electronics and Energy, 1, 100002. https://doi.org/10.1016/j.prime.2021.100002

Seyedzadeh, S., Pour Rahimian, F., Glesk, I. & Roper, M. (2018) *Machine learning for estimation of building energy consumption and performance: a review.* Visualization in Engineering, 6, Article 5. https://doi.org/10.1186/s40327-018-0064-7

Shapi, M.K.M., Ramli, N.A. & Awalin, L.J. (2021) *Energy consumption prediction by using machine learning for smart building: Case study in Malaysia.* Developments in the Built Environment, 5, 100037. https://doi.org/10.1016/j.dibe.2020.100037

Wang, Z. & Srinivasan, R.S. (2015) *A review of artificial intelligence based building energy prediction with a focus on ensemble prediction models.* In: 2015 Winter Simulation Conference (WSC), pp.3438–3448. http://dx.doi.org/10.1109/WSC.2015.7408504

Wang, Z. & Srinivasan, R.S. (2016) *A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models.* Renewable and Sustainable Energy Reviews. http://dx.doi.org/10.1016/j.rser.2016.10.079

Wen, J.X. (2014) *Review of building energy modeling for control and operation.* Renewable and Sustainable Energy Reviews, 37, pp.517–537. http://dx.doi.org/10.1016/j.rser.2014.05.056

Zhao, H. & Magoulès, F. (2010) *Parallel support vector machines applied to the prediction of* *multiple buildings energy consumption.* Journal of Algorithms and Computational Technology, 4, pp.231–249.

Z. Dong, *et al.*

Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification

Energy Build., 241 (2021), p. 110929, 10.1016/j.enbuild.2021.110929

C. Fan, F. Xiao, Y. Zhao

A short-term building cooling load prediction method using deep learning algorithms

Appl. Energy, 195 (2017), pp. 222-233, 10.1016/j.apenergy.2017.03.064

Q. Li, *et al.*

Applying support vector machine to predict hourly cooling load in the building

Appl. Energy, 86 (10) (2009), pp. 2249-2256, 10.1016/j.apenergy.2008.11.035

View PDFView articleView in ScopusGoogle Scholar

Q. Li, *et al.*

Predicting hourly cooling load in the building: a comparison of support vector machine and different artificial neural networks

Energy Convers. Manag., 50 (1) (2009), pp. 90-96, 10.1016/j.enconman.2008.08.033

A.-D. Pham, *et al.*

Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability

J. Clean. Prod., 260 (2020), p. 121082, 10.1016/j.jclepro.2020.121082

B. Dong, C. Cao, S.E. Lee

Applying support vector machines to predict building energy consumption in tropical region

Energy Build., 37 (5) (2005), pp. 545-553, 10.1016/j.enbuild.2004.09.009

J.-S. Chou, D.-K. Bui

Modeling heating and cooling loads by artificial intelligence for energy-efficient building design

Energy Build., 82 (2014), pp. 437-446, 10.1016/j.enbuild.2014.07.036

R. Wang, S. Lu, W. Feng

A novel improved model for building energy consumption prediction based on model integration

Appl. Energy, 262 (2020), p. 114561, 10.1016/j.apenergy.2020.114561

6 Appendix

Attached to this document below is the raw CSV. File downloaded from UCI Machine Learning repository and the R script used for this project.

ENB2012_data.csv ML R-script.txt

ENB2012_data.csv ML R-script.txt