Sachith Nimesh Yamannage\textsuperscript{1}

\textsuperscript{1}Affiliation not available

May 13, 2024

Sachith Nimesh Yamannage
https://orcid.org/0009-0002-9320-1890
Email: sachithnimesh999@gmail.com
Faculty of Science, University of Ruhuna, Matara, Sri Lanka

Abstract

This research intends to use machine learning techniques to estimate the heating load class of residential buildings, allowing for the design of more energy-efficient structures. The dataset utilized in the analysis comprises of seven hundred residential building examples, each with variables such as relative compactness, surface area, roof area, overall height, orientation, glazing area, and glazing area distribution. The target variable, heating load class, is classified as low, medium, or high.

The analysis begins with a thorough overview of the major variables, which includes summary statistics, correlation analysis, and visuals. Decision tree models are built in two ways: using the complete dataset and adopting a 10-fold cross-validation strategy. Additionally, a Neural Net Multi-Layer Perceptron (NN-MLP) model is run and compared to the decision tree models.

The results show that all models have high accuracy rates, with the NN-MLP model having the best at 99.57%. Comparative research indicates that the NN-MLP model outperforms decision tree approaches in forecasting heating load classes. However, both models perform well in accurately categorizing heating load classes.

Overall, this work emphasizes the effectiveness of machine learning algorithms in forecasting heating load classes, as well as their potential for improving the energy efficiency of residential structures.

Introduction

In today's quickly fluctuating world, the demand for sustainable and energy-efficient buildings has become critical. With climate change concerns growing and energy efficacy rules becoming stricter, precisely evaluating the energy performance of residential structures is critical. Outmoded techniques of energy estimation can be time-consuming and inflated. However, the advent of machine learning offers a promising alternative by employing data-driven methodologies to estimation heating load classes, allowing for the construction of further energy-efficient buildings.
The goal of this project is to usage machine learning procedures to forecast the heating load class for residential buildings. The dataset used in this analysis was gotten from the UCI Machine Learning Repository and comprises of seven hundred residential building samples, each of which is famed by various attributes such as relative firmness, surface area, wall area, roof area, overall height, orientation, glazing area, and varnishing area distribution. The target variable, heating load class, is categorized as low, medium, or high. This paper search for to give a complete examination of the dataset and machine learning models used to predict heating load classes. It begins by describing four important variables - Relative Compactness (X1), Surface Area (X2), Roof Area (X4), and Overall, Height (X5) - using summary statistics, correlation analysis, and graphics. It then examines the creation of decision tree models in two situations: using the complete dataset and employing a 10-fold cross-validation strategy. Furthermore, the research calculates the concert of a Neural Net Multi-Layer Perceptron (NN-MLP) model and compares it to decision tree models.

Finally, it stipulates recommendations based on a comparative checkup of the models' performance measures. The report's investigation intends to provide noteworthy insights into the use of machine learning approaches in estimating heating load classes of residential structures, foremost to the creation of more energy-efficient building designs.

**Methodology**

The methodology used for this study was a arranged strategy that included data preparation, model creation, and evaluation to estimate the heating load class of residential erections. During the initial phase, data was composed using the EnergyPerformance_Classification_SZ700 dataset from the UCI Machine Learning Repository. This dataset involved seven hundred samples of residential buildings, each with eight parameters and a target variable demonstrating the heating load class. Prior to model creation, the dataset underwent rigorous preprocessing to assure data quality and compatibility. This enters resolving missing values, storing category variables, and scaling numerical features to reduce biases and inconsistencies.

Predictive models were built using two key machine learning algorithms: Decision Tree and Neural Net Multi-Layer Perceptron (NN-MLP). For decision tree models, two situations were examined. Scenario A required training the model on the complete dataset, but Scenario B used a 10-fold cross-validation strategy to assess model performance and reduce overfitting. Similarly, the NN-MLP model was built using a 10-fold cross-validation procedure, with data normalization added to improve model convergence and performance. This meant that the neural network could efficiently learn from the data, unaffected by variations in feature sizes.

Following model development, the performance of each model was assessed using a variety of metrics, including accuracy, confusion matrix, precision, recall, and F1-score. A comparison of decision tree models and the NN-MLP model was performed using mean cross-validation accuracy and misclassification probability as metrics.

Furthermore, the ROC curve was used to show the trade-off between true positive rate and false positive rate for decision tree models in Scenario A, providing insights into the model's performance across various threshold values. The findings of each model were interpreted and analyzed to acquire insight into their efficacy in predicting heating load classes for residential structures. These findings were used to identify the best technique for classifying heating loads, considering both performance and computational economy.
Using this thorough technique, the research aims to provide useful insights into the predictive modeling process for heating load categorization in residential structures, allowing for more informed decision-making and model selection in real-world scenarios.

Results

Descriptive Analysis of Variables:
The descriptive analysis of variables X1, X2, X4, and X5 revealed valuable information about the dataset's properties. The correlation matrix indicated significant relationships between several variables, most notably between Relative Compactness (X1) and Surface Area (X2), as well as Roof Area (X4) and Overall, Height (X5). This points to potential multicollinearity issues that should be addressed during modeling. Summary statistics clarified the distribution and variability of the variables. For example, X1 had a mean of 0.763, whereas X2 ranged from 514.5 to 808.5, demonstrating significant variability in the sample.

Summary Statistics:

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>X4</th>
<th>X5</th>
</tr>
</thead>
<tbody>
<tr>
<td>count</td>
<td>700.000000</td>
<td>700.000000</td>
<td>700.000000</td>
<td>700.000000</td>
</tr>
<tr>
<td>mean</td>
<td>0.763057</td>
<td>672.500000</td>
<td>177.240000</td>
<td>5.2250000</td>
</tr>
<tr>
<td>std</td>
<td>0.107454</td>
<td>88.501219</td>
<td>45.170798</td>
<td>1.751073</td>
</tr>
<tr>
<td>min</td>
<td>0.628000</td>
<td>514.500000</td>
<td>110.250000</td>
<td>3.500000</td>
</tr>
<tr>
<td>25%</td>
<td>0.600000</td>
<td>612.500000</td>
<td>147.000000</td>
<td>3.500000</td>
</tr>
<tr>
<td>50%</td>
<td>0.740000</td>
<td>686.000000</td>
<td>228.500000</td>
<td>3.500000</td>
</tr>
<tr>
<td>75%</td>
<td>0.820000</td>
<td>759.500000</td>
<td>220.500000</td>
<td>7.000000</td>
</tr>
<tr>
<td>max</td>
<td>0.900000</td>
<td>880.500000</td>
<td>220.500000</td>
<td>7.000000</td>
</tr>
</tbody>
</table>

Decision Tree Model:
In Scenario A, the decision tree model had an amazing accuracy of 98.57%. The confusion matrix produced low misclassification rates across all classes while maintaining high precision and recall values. Furthermore, the model achieved a mean cross-validation accuracy of 95.29%, demonstrating its impressive performance.
In Scenario B, where a 10-fold cross-validation strategy was used, the decision tree model obtained 97.57% accuracy. Precision, recall, and F1-score were all high, as in Scenario A, showing consistent performance over multiple validation folds.

### Neural Net Multi-Layer Perceptron (NN-MLP) Model:

The NN-MLP model performed better than the others, with an accuracy of 99.57%. In terms of mean cross-validation accuracy (92.99%) and prediction accuracy, this fared better than the decision tree models. All classes exhibited consistently good precision, recall, and F1-score, with very few instances of misclassification.

**Predicted Heating Load Class:**

Using assumed X variable values, the estimated heating load class for the first residential building was judged to be "Low." This result demonstrates the model's ability to accurately classify heating load classes using input features.
Comparative Analysis:

In the scenario A and scenario B, mean cross-validation accuracy of the decision tree models was 95.29% and 95.29%, respectively, while the NN-MLP model produced a better mean accuracy of 92.99%. An insignificant risk of misclassifying buildings with low heating loads as high and a near-zero possibility of misclassifying buildings with high heating loads as medium were shown using probability calculations. Given the circumstances, the NN-MLP model emerges as the recommended option because of its greater accuracy and robustness in heating load classification tasks, even though decision tree models offer simplicity and interpretability.

Discussion and Conclusion

The results of our investigation provide insight into the effectiveness of several machine learning models in forecasting the residential building heating load class. Strong correlations between some variables were found through descriptive analysis, suggesting possible multicollinearity problems that might affect the performance of the model. Despite this, decision tree models functioned remarkably well, attaining high accuracy, and proving to be durable in scenario-based and cross-validation methods alike.

In contrast, the NN-MLP model fared better than the decision tree models in terms of F1-score, accuracy, precision, and recall. Its exceptional performance can be ascribed to its interpretability-sacrificing capacity to capture intricate nonlinear relationships within the data. Furthermore, minimal misclassification rates were shown by probability calculations, highlighting the model's dependability in forecasting heating load classes.

Our research concludes by demonstrating how well machine learning algorithms can anticipate the heating load classes of residential structures. The NN-MLP model excels in accuracy and resilience, whereas decision tree models offer simplicity and interpretability. Therefore, the NN-MLP model becomes the recommended option for real-world applications where precision is crucial, such energy efficiency evaluations in building design. Future studies, however, might investigate hybrid or ensemble approaches to improve interpretability and predictive performance even more in this area.

References


Appendix


- First five rows in dataset

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
<th>X7</th>
<th>X8</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.74</td>
<td>680.0</td>
<td>245.0</td>
<td>220.5</td>
<td>3.5</td>
<td>4</td>
<td>0.0</td>
<td>0</td>
<td>Low</td>
</tr>
<tr>
<td>1</td>
<td>0.74</td>
<td>686.0</td>
<td>245.0</td>
<td>220.5</td>
<td>3.5</td>
<td>3</td>
<td>0.0</td>
<td>0</td>
<td>Low</td>
</tr>
<tr>
<td>2</td>
<td>0.74</td>
<td>686.0</td>
<td>245.0</td>
<td>220.5</td>
<td>3.5</td>
<td>3</td>
<td>0.0</td>
<td>0</td>
<td>Low</td>
</tr>
<tr>
<td>3</td>
<td>0.74</td>
<td>686.0</td>
<td>245.0</td>
<td>220.5</td>
<td>3.5</td>
<td>2</td>
<td>0.0</td>
<td>0</td>
<td>Low</td>
</tr>
<tr>
<td>4</td>
<td>0.71</td>
<td>710.5</td>
<td>269.5</td>
<td>220.5</td>
<td>3.5</td>
<td>2</td>
<td>0.0</td>
<td>0</td>
<td>Low</td>
</tr>
</tbody>
</table>