Energy Efficiency of Residential Building using Decision Tree Classification

Fatmah Mohame Alkaabi
Business Analytics Programme
Abu Dhabi School of Management
Abu Dhabi, United Arab Emirates:
adsm-215057@adsm.ac.ae

https://doi.org/10.63962/BWZY2529

***Abstract*— Building cooling demand is rising due to urbanization and climate change, especially in hot regions like Abu Dhabi[1]. Traditional HVAC systems are energy-intensive[2], increasing the need for innovative, efficient solutions. This study presents a decision tree–based model utilizing C4.5, CART, and Decision Stump algorithms to predict residential cooling load based on thermal and environmental data. Among these, CART achieved the highest predictive accuracy. The proposed model is transparent, interpretable, and cost-effective, supporting energy efficiency efforts. It can help reduce operational costs and guide decision-making for sustainable residential energy planning in rapidly urbanizing and climate-vulnerable regions [3] [4]**

Keywords*— *Artificial Intelligence; CART; Decision Tree; Decision Stump; Cooling Load

I. Background

Energy consumption in buildings constitutes a significant portion of global energy use, particularly in hot climates where cooling demands are intensified by climate change and urbanization. Traditional estimation methods often face challenges in capturing dynamic factors such as weather variability, building geometry, and occupant behaviour.

II. Motivation

With increasing urbanization and climate change, there is a pressing need for accurate and efficient cooling load predictions in residential settings. Prior research demonstrates that machine learning can significantly improve energy efficiency and operational cost savings, especially in commercial sectors[5]. This study aims to extend those benefits to residential applications by leveraging data-driven approaches

III. Problem Statement

Despite advances in predictive modeling, current cooling load estimation techniques often lack transparency and adaptability. Furthermore, explainable artificial intelligence (XAI) remains underutilized in this domain, making it difficult for stakeholders to trust and implement AI-based solutions.

IV. Research Questions

How accurately can decision tree algorithms (C4.5, CART, and Decision Stump) predict cooling loads in residential buildings?
Can explainable models improve stakeholder trust and usability in energy forecasting?

V. Objectives

To develop a predictive, explainable decision tree–based model for cooling load estimation using thermal and environmental data.
To evaluate and compare the performance of C4.5, CART, and Decision Stump algorithms.
To provide interpretable results that support both residential users and policy decision-makers.

VI. Contributions

This research contributes an interpretable and efficient modeling approach that enables accurate cooling load forecasting while enhancing transparency through XAI techniques. It supports energy efficiency strategies, reduces operational costs, and aligns with sustainability goals.

VII. Scope and Value

The study focuses solely on decision tree (DT) algorithms and their ability to forecast cooling loads in residential buildings. Other machine learning models and broader energy systems are outside the scope. The proposed framework provides actionable insights for homeowners, energy providers, and policymakers to establish more innovative, AI-integrated energy strategies. Future research will address the implementation and evaluation of the developed system.

VIII. Methodology and Results

This study compares three decision tree (DT) algorithms—C4.5, CART, and Decision Stump—to predict cooling loads in residential buildings. While existing literature highlights the effectiveness of decision tree models in energy forecasting, most studies refer to general DT approaches or ensemble methods (such as Random Forest) without directly comparing these specific standalone classifiers. By focusing on interpretable DT algorithms, this study aims to evaluate which model offers the best trade-off between prediction accuracy, computational efficiency, and explainability.

Prior research supports this direction: Moon et al. (2024) demonstrated that ensemble-based DT models like Random Forest and Gradient Boosting yield high predictive accuracy for residential electricity forecasting, supported by SHAP for interpretability[6]. Similarly, decision tree–based models achieved strong performance ($R^2 > 0.95$) in predicting heating and cooling loads [4] [7]. However, comparative benchmarking of standalone DT algorithms such as C4.5, CART, and Decision Stump remains limited in energy-related applications,

highlighting the need for this study.

TABLE I. Key Aspects of Literature Review Summary

Author/Year	Purpose/Objective	Methodology	Findings
Moon et al. (2024)	Residential building electricity consumption forecasting using explainable AI.	Random Forest, Gradient Boosting, Decision Tree Bagging, SHAP for interpretability.	Random Forest and Gradient Boosting outperformed regression models, particularly in cooling load prediction.
Khorrami et al. (2024)	Comparative study on heating and cooling loads forecasting.	Decision Tree, Linear Regression, Neural Networks.	The decision Tree model achieved 98.96% accuracy for the heating load and 93.24% for the cooling load.
Moradzadeh et al. (2020)	Performance evaluation of machine learning for heating and cooling loads.	Decision Trees, Multilayer Perceptron (MLP), Support Vector Regression (SVR).	MLP achieved $R^2 > 0.95$, outperforming SVR in heating and cooling load predictions.

The methodology is guided by the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework[8]. which is widely adopted in practical data science workflows for its structured and iterative development process[9].

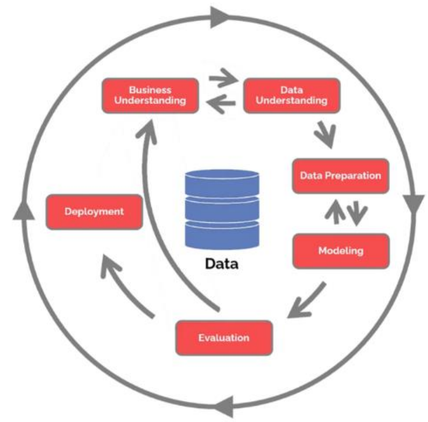

Fig. 1. CRISP-DM Process Model (Adapted from Satyasai, 2024).

- Data Source: This study utilizes the energy efficiency dataset retrieved from Kaggle, which includes 768 instances and eight building-related input features such as wall area, roof area, glazing area, and orientation[10].
- Tools: Model development was performed using standard machine-learning environments that support classification, feature selection, and evaluation.
- Preprocessing: Data cleaning, normalization, and feature selection were applied to improve model robustness.
- Modeling: The three DT algorithms—C4.5, CART, and Decision Stump—were trained and tested using a 70/30 split. Performance was assessed using accuracy, precision, recall, and F1-score.

This methodology answers:

- How do C4.5, CART, and Decision Stump compare in predicting residential cooling loads?
- Which model best balances performance and interpretability for real-world energy applications?

This short paper proposed a machine learning–based framework for predicting cooling load in residential buildings using three interpretable decision tree (DT) algorithms: C4.5, CART, and Decision Stump. The approach is grounded in the CRISP-DM methodology and leverages publicly available thermal and environmental building data. Although prior literature supports the utility of DT-based forecasting in energy modeling, this study contributes by comparatively analyzing specific DT variants that have not been widely benchmarked together. The proposed methodology supports energy efficiency efforts by enabling transparent and replicable decision-making that can inform residential energy planning and regulatory strategies.

To extend this research, future studies may consider:

- **Validation**: Validate predictive reliability by applying and testing the trained models on real-world residential energy datasets across various climate zones and building types.
- **Model Enhancement:** Explore hybrid models by integrating DT algorithms with ensemble learning or neural networks to improve performance and scalability.
- **Broader Applications**: Extend the methodology beyond residential buildings to other sectors—such as commercial real estate and infrastructure—where accurate energy forecasting supports sustainability, cost reduction, and operational planning.

This work serves as a foundation for developing practical tools to support long-term energy sustainability strategies.

To extend this research, future studies may consider:

- **Validation**: Validate predictive reliability by applying and testing the trained models on real-world residential energy datasets across various climate zones and building types.
- **Model Enhancement:** Explore hybrid models by integrating DT algorithms with ensemble learning or neural networks to improve performance and scalability.
- **Broader Applications**: Extend the methodology beyond residential buildings to other sectors—such as commercial real estate and infrastructure—where accurate energy forecasting supports sustainability, cost reduction, and operational planning.

This work serves as a foundation for developing practical tools to support long-term energy sustainability strategies.

REFERENCES

[1] Department of Energy Abu Dhabi. (2023). Sustainability Report 2023. Abu Dhabi: DOE.

[2] Ionescu, C., Baracu, T., Vlad, G. E., Necula, H., & Badea, A. (2015). The historical evolution of energy-efficient buildings. Renewable and Sustainable Energy Reviews, 49, 243–253. https://doi.org/10.1016/j.rser.2015.04.062

[3] Bekdaş, G., Aydın, Y., Işıkdağ, Ü., Sadeghifam, A. N., Kim, S., & Geem, Z. W. (2023). Prediction of the cooling load of tropical buildings with machine learning. *Sustainability*, 15(9061).

[4] Khorrami, B. M., Soleimani, A., Pinnarelli, A., Brusco, G., & Vizza, P. (2024). Forecasting heating and cooling loads in residential buildings using machine learning: A comparative study of techniques and influential indicators. *Asian Journal of Civil Engineering*, 25, 1163–1177.

[5] Zingre, K. T., Wan, M. P., & Ng, B. F. (2021). Prediction of the cooling load of tropical buildings with machine learning. *Energy and Buildings*, 252, 111403.

[6] Moon, J., Maqsood, M., So, D., Baik, S. W., Rho, S., & Nam, Y. (2024). Advancing ensemble learning techniques for residential building electricity consumption forecasting: Insight from explainable artificial intelligence. *PLOS ONE*, 19(11), e0307654.

[7] Moradzadeh, A., Mansour-Saatloo, A., Mohammadi-Ivatloo, B., & Anvari-Moghaddam, A. (2020). Performance evaluation of two machine learning techniques in heating and cooling loads forecasting of residential buildings. *Applied Sciences*, 10(11), 3829.

[8] Satyasai, U. S. (2024). What is CRISPR-DM? Retrieved April 28, 2025, from https://medium.com/@udaysrisatyasai/what-is-crisp-dm-cbdae4dfcc84

[9] Li, X., Zhang, C., & Zhuang, J. (2021). A systematic literature review on applying CRISP-DM in data mining. *Proceedings of the 2021 International Conference on Intelligent Computing (ICIC)*, 245–251.

[10] Chowdhury, U. (2022). *Energy Efficiency Data Set* [Data set]. Kaggle. https://www.kaggle.com/datasets/ujjwalchowdhury/energy-efficiency-data-set