Predicting Weather-Dependent Energy Savings for Low-Income Residential Buildings for Specific Upgrades with Limited Data

Phillip Clayton

https://doi.org/10.52202/077496-0010

Abstract

Energy efficiency in low-income residential buildings offers significant potential for reducing both carbon emissions and energy costs. This study develops a machine learning-based methodology to predict weather-dependent energy savings in low-income homes, utilizing limited building data. Leveraging detailed simulations from the National Renewable Energy Laboratory (NREL) and applying a nearest-neighbor approach, the model estimates potential energy savings for natural gas heating, electric heating, and electric cooling based on specific building modifications. A two-step clustering process refines predictions by removing outliers and improving accuracy. Results highlight that this approach is particularly effective for high-consumption buildings, which are often found in low-income areas. Additionally, a Python-based graphical user interface (GUI) enables both professionals and the public to easily access the results to make informed decisions about energy-saving improvements. This methodology provides targeted insights for utilities and city planners to prioritize energy-reduction initiatives, with significant implications for enhancing sustainability and supporting vulnerable communities.

Keywords: energy savings, machine learning, low-income households, energy efficiency, building upgrades

Introduction

The pathway to sustainability is multifaceted, with one of the most effective strategies being the reduction of energy demand. While large-scale deployment of renewable energy is crucial, reducing energy consumption in existing buildings, particularly in low-income residential areas, offers significant potential for achieving carbon reduction at a lower cost. This study employs machine learning techniques to predict weather-dependent energy savings for low-income residential buildings based on specific upgrades, using limited building data.

The primary objective of this study is to develop a methodology that uses machine learning models to estimate potential energy savings for natural gas heating, electric heating, and electric cooling through various building modifications. By leveraging detailed energy profiles from the National Renewable Energy Laboratory (NREL) and comparing them with actual building data from Cincinnati, Ohio, the study aims to provide reliable savings estimates with minimal data input.

Literature Review

Despite the widespread availability of energy modeling tools, there is a gap in the literature regarding tools that integrate machine learning with publicly available data to provide specific upgrade recommendations for low-income households. This study aims to address this gap by using a nearest-neighbor approach to estimate energy savings with minimal data requirements.

https://doi.org/10.52202/077496-0010

Amasyali and El-Gohary (2018) reviewed the application of data-driven techniques in energy-consumption prediction, highlighting the strengths and limitations of machine learning models in building energy modeling. Their work underscores the potential of machine learning in capturing complex patterns in energy use, which this study leverages by focusing on the data constraints of low-income households (Amasyali & El-Gohary, 2018).

Ensemble machine learning models combine multiple learning algorithms to improve predictive performance compared to using a single model. This approach allows for more accurate energy savings predictions by aggregating the strengths of different algorithms. In the context of energy efficiency, Doukas (2023) demonstrates that ensemble models capture diverse patterns of energy consumption across buildings, which can be particularly beneficial when applied to data with varying characteristics. The lessons from Doukas' work are reflected in this study's focus on low-income households, where minimal data inputs are required, making the models accessible and practical for implementation at scale (Doukas, 2023).

Hu et al. (2022) discussed the use of k-nearest neighbor estimation in functional nonparametric regression models. K-nearest neighbor estimation is a nonparametric method used for classification and regression. The method identifies the k closest training examples in the data space and uses their values to predict the target variable for a new instance. Although the research of Hu et al. (2022) is centered on functional data analysis, it supports the theoretical foundation of this study's approach to energy savings estimation in residential buildings (Hu, Wang, Liu, & Yu, 2022).

Hallinan et al. (2011) conducted a multivariate analysis of energy consumption that provides a basis for understanding complex energy-use patterns in residential buildings. This study builds on their findings by applying advanced machine learning techniques to predict energy savings with limited data inputs (Hallinan et al., 2011).

Building on these foundational studies, this research employs a nearest-neighbor approach to provide specific and actionable upgrade recommendations for low-income households. By leveraging publicly available data and advanced machine learning techniques, this approach aims to fill a significant gap in the existing literature on energy efficiency and building energy modeling.

Methods

Data and Model Development

Data from the National Renewable Energy Laboratory (NREL), encompassing 550,000 simulated buildings, approximately 21,000 of which were located in Ohio, were used for model development. These simulations included detailed information on residential energy use and building characteristics. Machine learning models were developed using H2O Flow, an open-source machine learning platform for building predictive models. These models focused on predicting energy consumption based on features such as

https://doi.org/10.52202/077496-0010

attic insulation, wall insulation, infiltration, HVAC efficiency, and heating setpoint adjustments.

Prediction Methodology

The methodology involves comparing actual building data from Cincinnati with NREL's simulated data through a nearest-neighbor approach. For each building, the 10 most similar simulated buildings are identified based on criteria such as area, natural gas heating, electric heating, and electric cooling. Mean savings are calculated for these 10 nearest neighbors.

Variability and Clustering

To address variability in savings estimates, the coefficient of variation (CoV) is used as a measure of reliability. When CoV exceeds 0.2, indicating significant variability, clustering is applied to refine the predictions. Clustering groups the nearest neighbors into subgroups with similar values. In this research, the most common result was to have one cluster of zero values and one cluster of non-zero values. Clustering can be a good way of removing outliers. Often in this analysis, it would determine if the prevalent value is zero or not. If the larger cluster is all zeros, then the mean savings is concluded to be zero. If not, then the smaller cluster is removed to prevent them from skewing the data. This two-step clustering process enhances the accuracy of energy savings predictions.

Results

The analysis demonstrated a notable inverse relationship between mean savings and CoV, particularly in buildings with high energy consumption. This indicates that the approach is particularly effective in identifying significant energy-saving opportunities in low-income, high-consumption homes. A Python-based graphical user interface (GUI) was developed to enable address-specific energy savings queries, providing a prioritized list of potential upgrades based on estimated mean savings and their associated CoV.

Discussion

Low-income households often reside in older, less-efficient buildings, leading to disproportionately high energy costs relative to their income. Targeting these households for energy efficiency improvements can provide substantial benefits, both in terms of cost savings and quality of life. While financial constraints may hinder their ability to implement these upgrades without assistance, this study's findings can guide policymakers and utility companies in designing targeted support programs. By prioritizing high-consumption buildings, the methodology ensures that interventions are both impactful and cost-effective, helping to bridge the gap between energy efficiency and affordability. The methodology developed in this study offers a scalable and cost-effective approach to identify and prioritize energy-saving interventions. By using limited data inputs, the approach overcomes the barrier of resource-intensive traditional audits, making it accessible and practical for broader use. Focusing on high-consumption

https://doi.org/10.52202/077496-0010

buildings in low-income areas is essential, as these buildings often present the greatest opportunities for energy savings and carbon reduction.

Future Work

The next steps involve validating the estimated savings against actual energy consumption data to refine the methodology further. This validation process is crucial to ensure the accuracy and applicability of the predictions in real-world scenarios. Additionally, expanding the model to include more diverse building types and regions could enhance its generalizability.

Conclusion

This study presents a novel approach to predicting energy savings in low-income residential buildings using machine learning and limited building data. By focusing on high-consumption buildings, the methodology provides targeted insights that can help utilities and city planners prioritize energy reduction initiatives effectively. This approach not only enhances sustainability but also supports vulnerable communities in achieving greater energy efficiency.

Conflict of Interest

The author declares no conflict of interest.

https://doi.org/10.52202/077496-0010

References

Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, *81*, 1192-1205.

Doukas, H. (2023). Estimating the energy savings of energy efficiency actions with ensemble machine learning models. *Applied Sciences, 13*(4), 2749.

Hu, X., Wang, J., Liu, L., & Yu, K. (2022). K-nearest neighbor estimation of functional nonparametric regression model under NA samples. *Axioms*, *11*(3), 102.

Hallinan, K. P., et al. (2011). Multivariate analysis of energy consumption. *Energy and Buildings, 43*(10), 2822-2831.

https://doi.org/10.52202/077496-0010