# Predicting Weather Dependent Energy Savings

## for Low-Income Residential Buildings for Specific Upgrades with Limited Building Data

Phillip Clayton

The University of Dayton

## Abstract

The pathway to sustainability is challenging. Multiple paths exist, but the key will be to achieve carbon reduction with the least cost. This could be achieved through large scale deployment of renewable energy; however, many studies have shown how important it is to reduce demand first.

This study employs machine learning to analyze detailed energy profiles from the National Renewable Energy Laboratory (NREL), estimating potential energy savings for natural gas heating, electric heating, and electric cooling through modifications such as insulation improvements, setpoint changes, infiltration reduction, or system efficiency enhancements. By comparing these building models with actual building data from Cincinnati, Ohio, via a nearest neighbor approach, mean savings are calculated for the ten most similar simulated houses. This process allows for the use of limited data (annual energy usage for heating and cooling and house area) to identify comparable model sets and estimate potential areas for energy-saving improvements. When savings estimates vary significantly (coefficient of variation greater than 0.2), clustering is applied to find a more consistent subgroup, enhancing the accuracy of the energy savings predictions.

This methodology proves particularly effective for high energy-consuming residences, which are often found within low-income housing sectors. By focusing on buildings with the highest potential for energy savings, this approach offers targeted insights for utilities and city planners looking to prioritize energy reduction initiatives effectively. It highlights buildings where interventions could have the most substantial impact, both in terms of energy savings and cost efficiency.

The next steps will involve validating the estimated savings against actual data. This validation process is crucial for refining the methodology and ensuring its applicability and accuracy in real-world scenarios.

Focusing on high-consumption, low-income buildings, this study aims to reduce energy demand, enhance sustainability, and help vulnerable communities achieve greater energy efficiency.

## Background

**Key to Sustainability**
Focus on reducing energy consumption in residential buildings, especially in low-income areas with older, inefficient buildings.

**Barrier to Efficiency**
Traditional methods for estimating energy savings are often resource-intensive, requiring detailed audits that are not feasible for all buildings.

**Why Prioritize High-Consumption Buildings**
Correlation of high-consumption buildings to low-income communities. Target buildings to provide substantial support to communities in need.

**NREL Datasets**
550,000 simulated buildings. About 21,000 located in Ohio used for this modeling. Simulations include data on various aspects of residential energy use and building characteristics.

**Regression Models**
Machine learning was applied using h2o Flow to create regression models from the NREL datasets. These models make predictions of a given energy consumption based on the building characteristics used to create the model.

## Methodology

**Machine learning models developed from NREL data for selected features**

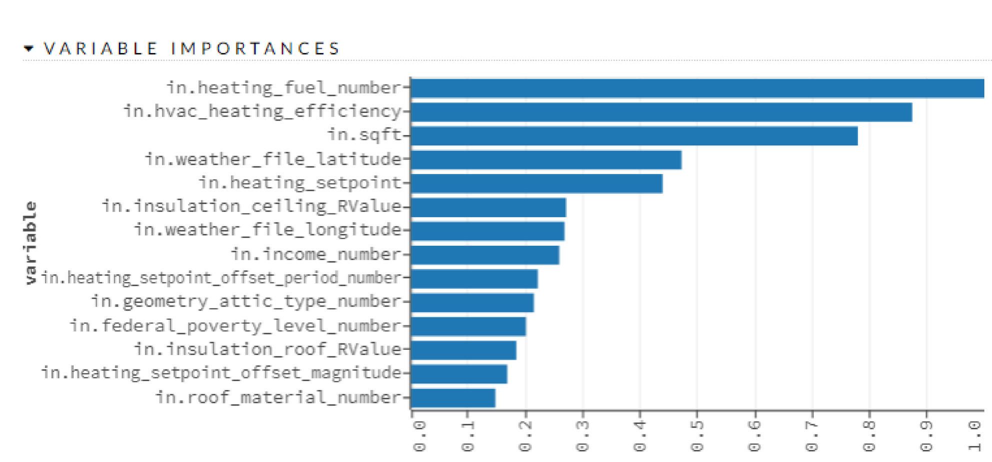
**Predicting Energy Savings**
- Natural Gas Heating
- Electric Heating
- Electric Cooling

**Based on Building Modifications**
- Attic Insulation
- Wall Insulation
- Infiltration
- HVAC Efficiency
- Heating Setpoint

**Then Applied to a Nearest Neighbor Calculation**
- 10 NREL buildings that are most like each Cincy house based on:
- Area
- Natural Gas Heating
- Electric Heating
- Electric Cooling

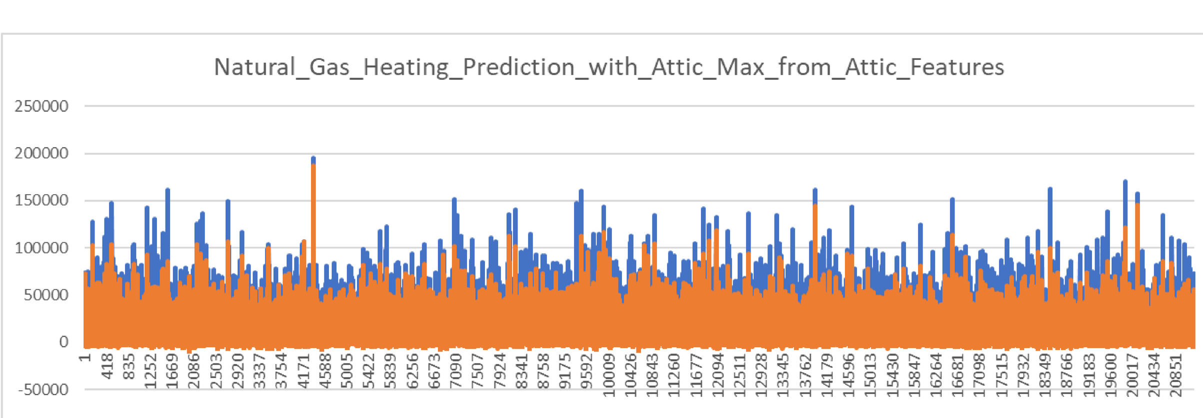
## Model of Natural Gas Heating Consumption Prediction from Attic Insulation Increase

**Variables used in the creation of a model ranked by their influence on the predicted output.**




**Predicted energy savings calculated from regression models**

- **Blue** is actual consumption and **orange** is predicted consumption.
- Difference between actual and predicted consumption is the predicted savings.

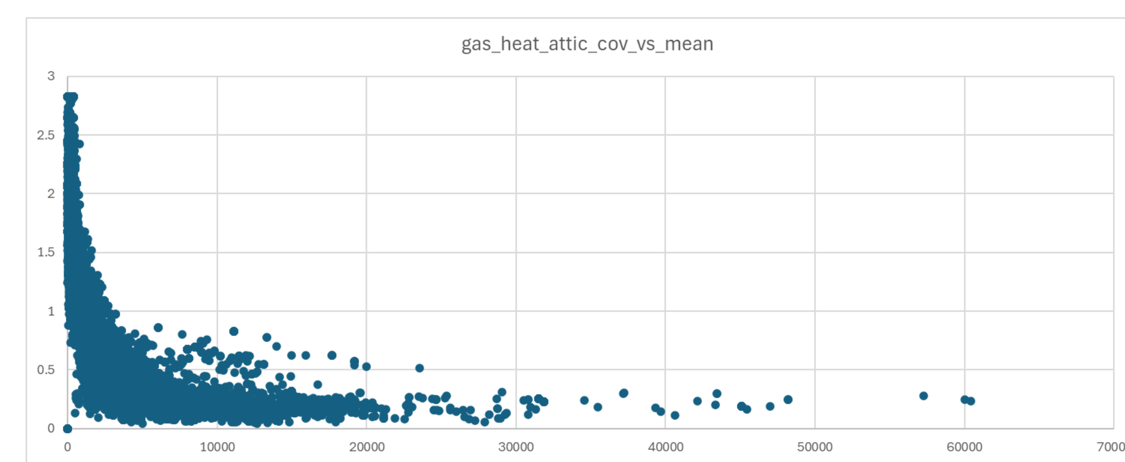



## Analysis and Insights

**Analysis Process**
- Begin with calculating mean savings from the ten nearest NREL models to estimate potential energy savings for each building.
- Use the coefficient of variation (CoV) to assess savings consistency, with a threshold of less than 0.2 indicating reliability.

**Refining Predictions**
- When CoV exceeds 0.2, implying significant variability, apply a two-step clustering to the ten nearest neighbors. Recalculate mean savings and CoV for the larger cluster to ensure focus on consistent savings estimates.

**Plot of mean savings versus coefficient of variation (CoV).**
Correlation of low CoV to high savings reveals how this approach is more accurate in low-income, high-consumption homes.




**Key Finding: Reliability in High-Consumption Buildings**
- Discovered an inverse relationship between mean savings and CoV, notably in buildings with high energy use.
- This pattern demonstrates the approach's effectiveness in pinpointing significant energy-saving opportunities in low-income, high-consumption homes, making predictions more reliable where the potential for savings is greatest.

**GUI Development**
- Created a Python script with a Graphical User Interface (GUI), enabling address-specific queries for energy savings.
- Outputs a prioritized list of potential upgrades based on estimated mean savings, arranged from highest to lowest.
- CoV displayed with each savings estimate, providing insight into the reliability of predictions.

**Future Steps**
- The next phase will focus on validating these predicted savings against actual data, further refining the methodology's accuracy and applicability.

## Acknowledgements

Thank you to Dr. Kevin Hallinan for your mentorship on this project. Your guidance on this research as well as the wealth of knowledge you have shared through various classes in the past year and a half has given me a deep understanding of renewable and clean energy, machine learning, and artificial intelligence.

## Example of Characteristics from NREL Datasets Used to Create Two of the Fourteen Regression Models

- Red text highlights characteristics specific to certain models
- Area, latitude & longitude, poverty level, and income are common in all models

| heating_energy_consumption_with_attic characteristics models: | cooling_energy_consumption_with_wall characteristics_model: |
|---|---|
| in.sqft | in.sqft |
| in.heating_setpoint | in.weather_file_latitude |
| in.heating_setpoint_offset_magnitude | in.weather_file_longitude |
| in.weather_file_latitude | in.cooling_setpoint_offset_period_number |
| in.weather_file_longitude | in.federal_poverty_level_number |
| in.federal_poverty_level_number | in.geometry_wall_exterior_finish_number |
| in.geometry_attic_type_number | in.geometry_wall_type_number |
| in.heating_fuel_number | in.hvac_cooling_EER |
| in.heating_setpoint_offset_period_number | in.hvac_cooling_space_percentage |
| in.income_number | in.income_number |
| in.insulation_ceiling_RValue | in.insulation_wall_RValue |
| in.insulation_roof_RValue | |
| in.roof_material_number | |

## Python GUI

- Input address.
- Output mean savings from ten nearest neighbors (or cluster) in order.
- CoV displayed to show reliability of results.