Oh, Where Oh, Should These Heat Pumps Go, Oh Where Oh Where Can They Be?

Noel Stevens, TRC, Boston, MA
Mohammed Hossain, Ph.D., TRC, Buffalo, NY
Wendy Lu, Ph.D., TRC, Cary, NC
Jake Schefrin, TRC, Boulder, CO
Alfredo Jahn, TRC, St. Louis, MO

ABSTRACT

A key challenge many program administrators face today is identifying customers for whom conversion to heat pumps makes economic sense given costs associated with converting to heat pumps from their installed technologies and effectively marketing to those customers. Efforts to reduce greenhouse gas through electrification leads many utility programs to look to widespread adoption of heat pumps for heating, cooling, and water heating. However, the value proposition of conversion to heat pumps and electrification differs considerably depending upon the technologies each home or business uses for heating, water heating and cooling. Switching from a delivered fuel system (e.g. oil or propane) or resistance heat with room air conditioners to a heat pump is an easier case than from natural gas. Often program administrators do not have access to consumption data for one or both fuels, which complicates the challenge of identifying the heating fuel type. Public Service Gas and Electric (PSEG) of Long Island provides electric service to customers on Long Island, while gas service is provided by another utility. TRC implements PSEG LI's energy efficiency programs, including their Clean Heat Connect Program that provides rebates for residential customers installing electric heat pumps. TRC completed this study to identify customers most likely to use delivered fuel (oil/propane) or electric resistant space heating.

In this paper, TRC applied machine learning models to data often available to program managers to predict the existing fuel (electric, natural gas, or delivered fuel) used for space heating based on known characteristics. We combined past program history with third party tax assessor data, state department of environmental protection agency data, and billing analysis results to describe all single-family homes in the territory. We merged these data with thousands of home energy audits that provided evidence of the actual system types within each home. We applied the data to a simple linear statistical model to limit the number of variables that were both theoretically important and statistically relevant. We then used the identified variables in machine learning models and used k-fold evaluation to select the best fitting models to develop predict system type. We then applied the best fitting models from the k-fold evaluation to a set of data outside the modeling exercise for which we knew the actual space heating fuel type to assess the accuracy of the three best fitting models in a "real world" prediction.

Introduction

Utility programs are increasingly turning to expanding the installation of electric heat-pump for space heating and cooling and heat-pump water heaters to achieve their electrification goals and reduce green-house gas impacts. The value proposition for heat-pump conversion differs considerably depending upon a range of factors, of which the existing heating and water heating system's installed technology is one of the most crucial. In New York, and in most regions of the United States, the value proposition for heat pumps is more difficult for natural gas conversions. However, installing a heat pump or heat pump water heater makes economic sense when converting customers away from electric resistance heating and water heating or forced hot air systems powered using delivered fuel (e.g., oil or propane). The trick for program implementers is identifying customers whose existing technologies make them strong.

This paper presents TRC's approach for predicting the installed technologies for space and water heating for Public Service Electric and Gas (PSEG) Long Island's population of single-family residential customers. This study was completed by TRC's Research and Consulting team (the R&C team) in conjunction with TRC's PSEG Long Island's program implementation team (the Program Team). The Program Team sought to identify customers with existing electric resistance or delivered fuel powered heating and water heating systems. While PSEG is a dual-fuel (electric and gas) utility, their service territory on Long Island is limited to electric service only. Consequently, the program team does not have access to gas service data for their electric customers on Long Island, so the Program team cannot identify electric customers who use gas for heating and water heating from those that use delivered fuels.

Study Goal and Objectives

The goal of this study was to leverage available data to identify customers who are strong candidates for conversion to heat pumps and heat pump water heaters from their existing space heating and water heating systems. To meet this goal, this study addressed the following research objectives:

- Identify installed technology for heating, cooling, and water heating systems for a sample of customers in PSEG Long Island's territory.
- Develop a series of explanatory variables to define each customer in the population in terms of their energy consumption, building characteristics, past program participation, and other characteristics known for the entire population of customers.
- Use a sample customer with known installed technology for space heating and water heating and explanatory variables known for all customers to predict fuel and system type.
- Apply the model to known characteristics for each customer in the population to predict system fuel type for all single-family homes on Long Island.

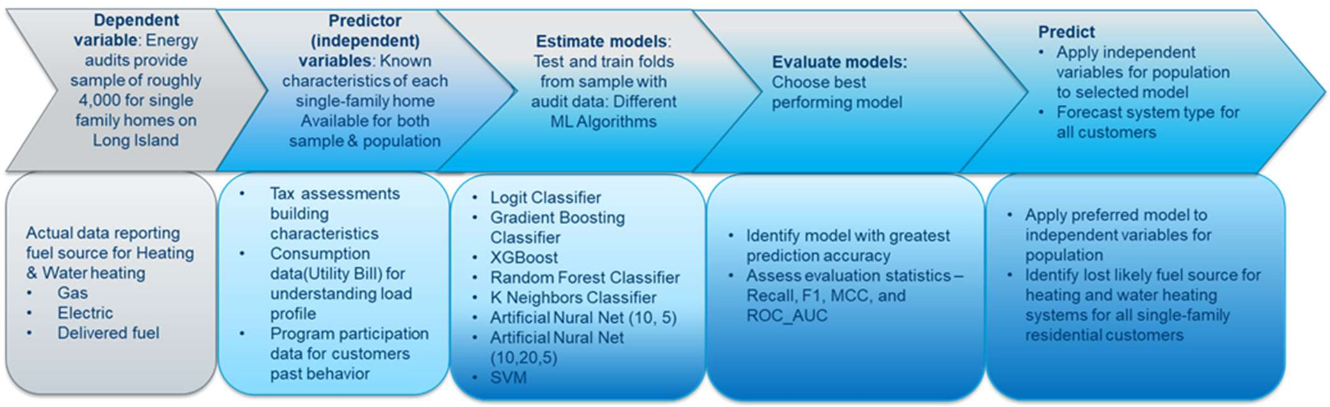
METHODOLOGY

In this section, we first provide a general overview of the modeling process to illustrate the relative simplicity and transferability of this process to other utility programs and offer a conceptual understanding of more complex data science concepts for the casual reader. Then we discuss the data used in the analysis and dive into the details of the modeling process to explain some of the more technical aspects of the data and process.

Methodology Overview

Figure 1 presents an overview of our approach to developing models we used to predict fuel type for heating of single-family residential customers on Long Island, NY. In the first step of this process, TRC used data from home energy audits to document the actual heating and water heating systems found in a sample of roughly 3,460 single family homes. We combined the audit information with data identifying each customer's electric load, building characteristics, and past program participation. We used the energy audit information to construct dependent variables for the model, and the remaining data to provide explanatory variables used to estimate the system fuel type. Where necessary, we imputed missing values for explanatory variables by building type and other key variables to provide a more robust data set. Next, we split the data set into training (70%) and testing (30%) sets for model evaluation. We then used the training data sets to estimate seven different types of machine learning models, and tested model performance on the test set of data. After selecting the model with the greatest accuracy of prediction as measured using three separate tests, we applied the explanatory variables for each account in the population of single-family homes on Long Island to the chosen model to predict the fuel type used for heating for each home.

Figure 1. Modeling process overview



Data

In this section we discuss the dependent (explained) variable and the independent (explanatory) variables we used in the modeling exercise.

Dependent Variable - Residential Audit Data

Predictive modelling requires data reporting the actual occurrence of the situation we are trying to predict, in this case the presence of fuel used for space heating in each home. Common sources for this type of data include home energy assessments (audits) or surveys. TRC maintained PDF files containing detailed information recording during residential audits for 13,833 PDF files containing residential audit information documenting actual equipment found in homes when completing audits. Because these data are contained on PDF files, they could not be readily analyzed. However, the Team was able use the software "R-Tabulizer" to read a sample of 3,460 of these PDF files into structured data sets that we also linked to information regarding the other variables used in the analysis. This provided TRC with a sample of accounts with known heating system and fuel type characteristics that had sufficient explanatory variables to construct training and test data sets for modeling.

Explanatory Variables

Once we identified data for the dependent variable, the next challenge was to construct possible explanatory variables available for all sample observations as well as for the entire population of accounts. We used data for the sampled accounts to construct and test the different machine learning models; we needed data for the rest of the population to apply the estimated model to all accounts so we could predict heating and water heating fuel for each single-family home in the population. We discuss these variables in the following sections below: building characteristics, variables providing evidence of gas or delivered fuel, evidence of electric heating, cooling, and baseloads, and past program participation variables.

Building Characteristics

The Team obtained tax assessor data from Lightbox for all buildings on Long Island. From Lightbox, a 3rd party data vender. These data report building and owner characteristics of all buildings contained in town and/or county Tax Assessor records. Lightbox also provides estimated or modeled data for key fields such as building type, square footage, or number of rooms for instances in which the tax assessment records are incomplete. The Lightbox data provided TRC with a number of variables that were crucial for predicting system fuel type, including building type, whether the building was owner occupied, square footage, number of rooms, and number of baths in the home. Where available, Lightbox also reports very high-level information regarding the heating, water heating, and cooling systems in homes, which we discuss in more detail below.

Evidence of Gas and Delivered Fuel

TRC analyzed over 20 years of PSEG Long Island's program history to identify customers whose program tracking records provide evidence of gas or delivered fuel systems being replaced through previous program engagements. In addition, TRC used data maintained by the county level offices of the Department of Environmental protection for Queens, Suffolk, and Nassau counties to identify homes with above-ground storage tanks greater than 1,000 gallons that content delivered fuel (e.g., #2, #4, #6 Fuel Oil, Petroleum Products, Chemical Bulk Storage). We used these data to provide evidence that delivered fuel was present on site at some time in the past.

Identify Accounts with Electric Heating, Cooling, and Substantial Base Loads

TRC used statistical regression analysis commonly used for account-level billing analysis to develop individual account-level models that identify customers who have temperature-sensitive electric (heating and/or cooling) and base loads. We combined the monthly account-level consumption data the Program Team provided with historical temperature data to estimate the relationship between heating and cooling degree days and consumption as well as base loads. We used linear regression analysis to estimate four separate models for each account by regressing monthly consumption on the corresponding heating degree days (HDD) and cooling degree days (CDD) series as specified in the models expressed in the equations below.

- Heating and cooling model: $E = \alpha + \beta_1 * CDD_{ij} + \beta 2 * HDD_{ij} + \mu_{ij}$
- Heating only model: $E = \alpha + \beta_2 * HDD_{ij} + \mu_{ij}$
- Cooling only model: $E = \alpha + \beta_1 * CDD_{ij} + \beta 2 * HDD_{ij} + \mu_{ij}$
- Baseload only model: $E = \alpha + \mu_{ij}$

Where:

E = *Monthly consumption (electric or gas)*

α = Intercept term measures size of baseload when heating and cooling degree days = 0

β_1 = Measure of relationship between monthly consumption and CDD. The parameter estimates the impact of a 1-degree change in CDD on monthly consumption.

CDD_{ij} = Measure of cooling degree days for month i and location j

β_2 = Measure of relationship between monthly consumption and HDD. The parameter estimates the impact of a 1-degree change in HDD on monthly consumption.

HDD_{ij} = Measure of cooling degree days for month i and location j

μ_{ij} = error term

The Team used results from these models to identify customers with statistically significant temperature sensitive electric cooling and heating loads (i.e., customers with statistically significant terms for β_1 and/or $\beta 2$, respectively). Once we identified the best models for each home, we applied Typical Meteorological Year 3 (TMY3) temperatures to each account's respective model to estimate the magnitude of heating, cooling, and base loads under typical weather conditions. The Team used these models to construct the following explanatory variables describing the heating, cooling, and baseloads of each electric account:

- Presence of electric heating, cooling load
- Relative magnitude of electric heating, cooling load

Imputing Missing Values

TRC used the following imputation methods to fill missing values for the key fields used in the analysis.

- **Building square footage** – We used two methods to impute square footage. The first method used the building type and assessed square footage value were available in Lightbox data. We used these fields to calculate the mean square footage per dollar of assessed value by building type and zip code. We then multiplied this value by the assessed value for records that had missing square footage data but assessed value data available. The second method was used for records where the assessed value was not available. In these cases, we computed the average square footage by zip code and consumption size bin (small, medium, and large), and used this value to fill the remaining missing square footage values.
- **Building age** – We calculated the average building age by building type and census block in the Lightbox data, and assigned these values to missing values for building age.
- **Number of stories** – We used non-missing data from Lightbox to estimate a linear model that predicted the number of stories by census block as a function of the total square footage. We used the linear model results to fill missing values for the number of stories.
- **Number of baths** – We assumed one of four values for the number of baths, based on the square footage of the home: Homes less than 1,000 sqft were assigned one bath; Homes from 1,001 sqft to 1,600 sqft were assigned 1.5 baths; 1,601 sqft to 2,200 sqft were assigned 2 baths; and homes 2,201 sqft or greater were assigned 3 baths.

Data Limitations

TRC notes that the relevance of estimated models from this study to the population of accounts on Long Island was constrained by the extent to which the sample homes for which we had audit data were representative of the population. We identify the following differences that will impact the ability to extrapolate results to the population:

1. **Geographic distribution** – 72% of the accounts with audit data were located in Suffolk County compared to 52% for the population, while 28% of sample accounts were located in Nassau County compared to 47% for the population, and 0.4% were from Queens, compared to 0.6% for the population.
2. **Total consumption sampling** – The median annual consumption for the sample was somewhat lower 10,595 kWh /year than the population with a median of 11,495 kWh / year for the population, or 92%.

3. **Heating load sampling** – The median heating load of the sampled accounts 5,554 kWh /year, which was 97% of the population's heating estimate of 5,723.

While these data do not show substantial differences between the two data sets, the study would benefit from expanding the sample of accounts beyond the relatively small sample of accounts with energy audits.

Model Specification

In this section we discuss our process of defining and selecting various models to predict the fuel and system types used for space heating in each single-family residential property in PSEG Long Island's territory.

The generalized approach included the following steps:

1. Compiled the necessary data
2. Estimated a simple linear model to narrow down the set of variables
3. Ran successive training and test machine models through multiple folds
4. Compared accuracy to predictions for the training (modeling) and test (control) data sets to select the most accurate models
5. Applied selected model to population data to predict fuel type for space heating
6. Compare predicted fuel types to actual fuel types from program tracking data set

Primary Linear Model – Identifying Significant Variables

In general, TRC attempted to estimate the likelihood or probability of a customer having a particular fuel for space heating as a function of their electric heating, cooling, and baseload profile, building / ownership characteristics, and past evidence of oil or delivered fuel in the home. While this seems straightforward, the initial modeling dataset contained an extensive number of possible explanatory variables, many of which were likely to be highly correlated with other variables or not logically relevant to the choice of fuel used for space heating. Machine learning models are not bound by engineering principles or economic theory, so it is common for them to find patterns in the data that may not be logically tied to the event we are trying to model. To avoid model misspecification, we first estimated a linear logistic model that allowed us to pare down the number of variables considered in the model to those that were most relevant from an engineering and economic perspective. While the precise specification of this model was not the objective, this process narrowed the set of relevant variables to the following:

- Location characteristics: Zip code
- Building characteristics: Building use, square footage, number of rooms, number of baths, number of stories, number of rooms, number of units, building age, assessed value,
- Occupant characteristics: Owner occupied
- Consumption characteristics: Total annual consumption, Annual heating consumption, Annual cooling consumption, Annual baseload.

TRC used the linear model to attempt various combinations of variables that are most predictive of gas space heating fuel type.

Table *1* shows the key variables for predicting gas used in space heating. The variables we found to be most predictive included the customer's total consumption, electric heating load, number of baths and zip code. After isolating the key driver variables, our next step was to use these variables in different machine learning models to define various models for predicting space heating fuel type.

Table 1. Primary linear model parameters and significance – Space heating

Variable	Estimate	Std Error	t Value	Pr (>| t |	Stat Sign.
(Intercept)	0.0562	0.1793	0.314	0.7538	
log(baseload_est + 1)	0.0350	0.0156	2.238	0.0254	*
log(heating_est + 1)	0.0084	0.0036	2.311	0.0209	*
heating_per_sqft	0.2740	0.1907	1.437	0.1510	
log(cooling_est + 1)	-0.0112	0.0063	-1.773	0.0764	.
kwSizemedium	-0.0676	0.0368	-1.838	0.0663	.
kwSizesmall	-0.0578	0.0872	-0.662	0.5078	
building_age	-0.0007	0.0008	-0.893	0.3720	
baths	-0.0537	0.0172	-3.12	0.0018	**
zip3110:log(imputedSqFoot2)	0.0112	0.0160	0.7	0.4843	
zip3115:log(imputedSqFoot2)	0.0352	0.0142	2.481	0.0132	*
zip3116:log(imputedSqFoot2)	0.0507	0.0264	1.917	0.0554	.
zip3117:log(imputedSqFoot2)	0.0709	0.0134	5.276	0.0000	***
zip3118:log(imputedSqFoot2)	0.0623	0.0177	3.518	0.0004	***
zip3119:log(imputedSqFoot2)	0.0994	0.0137	7.253	0.0000	***

Machine Learning Models Considered in the Analysis

In addition to the linear regression model, TRC estimated the following 7 machine learning model classifiers to identify the most likely fuel source for space heating in single family homes on Long Island:

- Logit classifier
- SGD Boost
- Gradient boosting classifier;
- X-Gradient boosting classifier
- Random forest classifier;
- K neighbor classifier; and
- Support Vector Classifier.

Model Evaluation to Select Specific Machine Learning Model

Machine learning models look to find patterns formed by combinations of different variables and values of those variables that together depict the most likely group of observations (i.e., customers) to have a certain outcome (i.e., specific fuel type for space heating). Each model run produces a predicted fuel type which we compared to the actual fuel type for each customer. The difference between the predicted and actual fuel type helped us identify the percentage of customers with different sources of error.

Figure 2 shows that the training and test sets can each be separated into four sources of error:

- **True positives (TP)** – The percentage of observation predicted to have a certain fuel type and actually did have that fuel type.
- **True negatives (TN)** – The percentage of observations predicted to NOT have a certain fuel type and actually did not have that fuel type.
- **False positives (FP)** – The percentage of observations predicted to have a certain fuel type but actually did not have that fuel type.
- **False negatives (FN)** – The percentage of observations predicted to not have a certain fuel type and did have that fuel type.

Figure 2. Identification of true and false outcomes from machine learning models

		Predicted	
		Negative (N)	Positive (P)
Actual	Negative	True Negative (TN)	False Positive (TP) Type I Error
	Positive	False Negative (FN) Type II Error	True Positive (TP)

Identifying TP, TN, FP, and FN for each model allowed us to construct two separate criteria for assessing the accuracy of each model: Accuracy Score, and Matthews Correlation Coefficient (MCC) Score. TRC examined each of these metrics to assess the performance of each model and selected the preferred machine learning technique for predicting space heating within each single-family home, which we define as follows.

1. **Accuracy score** – The ratio between the number of correctly classified samples and the overall number of samples.[1] This approach works well when the number of samples varies substantially.

Equation 1. Accuracy Score Formula

$$Accuracy = \frac{TP + TN}{TP+TN+FP+FN} = \frac{Correct\ Predictions}{Total\ Predictions}$$

2. **MCC Score** – The differences between actual values and predicted values. MCC Score is equivalent to the chi-square statistic for a 2 x 2 contingency table. Chicca and Jurman (2020) argue that the MCC score is more desirable than the F1 score, therefore we default to the MCC score in this paper.[2]

Equation 2. MCC Score Formula

$$\mathrm{MCC} = \frac{\mathrm{TP\ x\ TN - FP\ x\ FN}}{\sqrt{\mathrm{(TP + FP)x(TP + FN)x(TN + FP)x(TN + FN)}}}$$

RESULTS

In this section we first present our results from the machine learning modeling runs used to evaluate and choose the most effective machine learning model based on the evaluation criteria outlined above that rely on evaluation of successive model runs (or folds). Each model run compares predicted results to actual results from treatment and control groups and uses the MCC Score and Accuracy score to assess the model performance. This process, called k-fold evaluation, allowed us to narrow down the number of possible models to those that perform best.

Evaluation of Space Heating Model Performance

In **Error! Reference source not found.**, we present the results and evaluation of the Space Heating model performance for the five machine learning model types. The table shows the logistic regression, k-nearest neighbors (KNN), and random forest produced comparable accuracy scores during the k-fold cross-validation. This suggested that, based solely on this evaluation metric, none of the models had a clear advantage over the others. The Logistic classification model achieved an accuracy of 57%. However, a closer look reveals that it performs better in predicting gas instances, with an accuracy of 64%, compared to just 53% for non-gas. The k-nearest neighbors model slightly outperformed logistic regression in the k-fold cross-validation. Notably, k-nearest neighbors excelled at predicting Gas cases, but it struggled significantly with non-gas, offering little predictive value for that category. The random forest model showed comparable performance to the logit models during k-fold cross-validation. However, in actual predictions, its performance was highly imbalanced, achieving only 16% accuracy for gas predictions, while performing significantly better with 89% accuracy for non-gas.

Table 2. Evaluation of space heating system models

Space heating	**Model evaluation test**	**Fold1**	**Fold2**	**Fold3**	**Fold4**	**Fold5**	**Average**
Logit Classifier	**MCC**	**39%**	**39%**	**20%**	**19%**	**17%**	**27%**
	Accuracy	**76%**	**77%**	**70%**	**67%**	**73%**	**73%**
SGD Classifier_test_MCC	MCC	26%	35%	20%	11%	20%	23%
	Accuracy	73%	76%	71%	66%	43%	66%
Gradient Boosting Classifier	MCC	32%	30%	33%	31%	17%	29%
	Accuracy	69%	68%	71%	70%	64%	68%
XGBoos	MCC	32%	25%	29%	25%	14%	25%
	Accuracy	70%	68%	70%	68%	64%	68%
Random Forest Classifier	**MCC**	**27%**	**28%**	**26%**	**26%**	**27%**	**27%**
	Accuracy	**72%**	**72%**	**72%**	**70%**	**74%**	**72%**
K Neighbors	**MCC**	**39%**	**39%**	**42%**	**26%**	**30%**	**35%**
	Accuracy	**75%**	**73%**	**77%**	**69%**	**74%**	**74%**
Support Vector	MCC	25%	29%	28%	26%	24%	26%
	Accuracy	0.6104	0.5974	0.6384	0.6678	0.5961	0.6220

Evaluation of Space Heating Model Predictions

While the k-fold evaluation allowed us to identify those models that performed well within the treatment and control groups, they do not necessarily show how well the models predict in the real world, i.e., outside the modeling process. Therefore, we obtained a sample of 2,775 that were separate from those in the audit data. We used these data to compare the predicted space heating fuel to the actual space heating fuel reported in the program tracking records.

TRC linked the predicted space heating fuel type from each model to program tracking records from the 2,775 single-family residential heat pump installations to check the accuracy of predictions from each of the models discussed above. This allowed us to check the model performance against real world data not used to construct the models. We then linked the predicted values to program tracking records and compared the predicted and actual space heating fuel type.

Table 3 shows how the predicted space heating fuel type for each of the single-family homes that were also in the program tracking data compared to the actual systems fund in the tracking records. These data clearly showed there was a considerable difference between the model k-fold accuracy and "real world" accuracy in predicting. The table shows that only the logit model and random forest models successfully predicted both gas and non-gas systems. The k-nearest neighbor model correctly predicted all gas accounts but only 5 non-gas accounts gas systems, while the XGBoost and GBoost models predicted all systems to be gas, resulting in zero non-gas systems. Given these results, TRC selected the logit classifier for predicting space heating system type as it was one of the top performing models from the k-fold evaluation and also performed best in the test against actual system types from the tracking data.

Table 3. Evaluation of space heating system models

Model	**Fuel type**	**Actual fuel from program data**	**Predicted to have gas**	**Predicted to be non-gas**	**Percent of gas predictions correct**	**Percent of non-gas predictions correct**	**Overall accuracy (green box/all box)**
Random Forrest	Gas	1,231	**200**	1,031	16%		56.9%
	Non-gas	1,544	165	**1,379**		89%	
Logit	Gas	1,231	**789**	442	**64%**		**57.8%**
	Non-gas	1,544	728	**816**		**53%**	
K Neighbor	Gas	1,231	**1,231**	-	100%		44.5%
	Non-gas	1,544	1,539	**5**		0.3%	
Extreme Gradient Boost	Gas	1,231	**1,231**	-	100%		44.4%
	Non-gas	1,544	1,544	**-**		0.0%	
Gradient Boost	Gas	1,231	**1,231**	-	100%		44.4%
	Non-gas	1,544	1,544	**-**		0.0%	

A challenge with machine learning approaches is that they lack a theoretical framework for the underlying decision process of model specification. The approaches are flexible enough to fit specific data sets extremely well but can fail to capture theoretical structure that governs patterns seen in the observed, i.e. the broader population data generating process. K-fold cross validation can mitigate, but not fully avoid that problem because even splitting a specific data source into multiple subsets can only provide a limited amount of variation.

What we saw in this example was that, when moving to a different set of customers who were experiencing the same heating fuel choice, having a solid theoretical framework helped guide the modeling to a more robust outcome. Throughout the model specification process, we used the linear model to assess the set of variables that defined the underlying theoretical structure and fed those into the machine learning process. Some linear models we considered lacked theoretically logical variables, such as electric consumption (heating, cooling and baseloads) or location characteristics, but presented as highly predictive within the k-fold evaluation process (i.e. they reported 90+% accuracy). However, these models were far less effective at predicting when applied to data outside the treatment and test set (i.e., the broader population). This finding highlights the importance of employing a theoretical framework to the underlying variable selection. Moreover, it underscores a key limitation of using machine learning alone, absent a theoretical framework for model specification, can result in a false sense of accuracy.

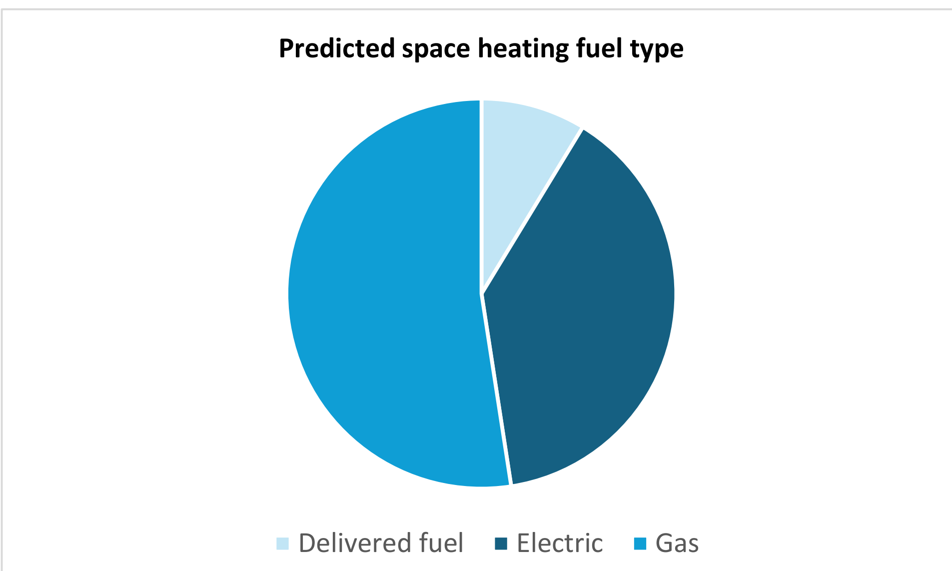
Predicted Space Heating System Fuel Types

TRC applied explanatory variables for each customer in the population of single-family accounts to the estimated model provided estimates of gas and non-gas space heating and water heating systems. We discuss these results in the sections that follow. We used the estimated Space Heating model to separate customers into gas and non-gas space heating. We then split the non-gas space heating customers into electric and delivered fuels using estimated electric heating lead from the set-point modeling used to disaggregate monthly consumption into heating, cooling, and baseloads. **Error! Reference source not found.** and **Error! Reference source not found.** present the estimated number of accounts in PSEG Long Island's territory that have electric, gas, and delivered fuel for space heating. The analysis identified roughly 96,000 customers likely to have delivered fuel for space heating, accounting for 16% of the population of accounts. In addition, we identified 242,000 customers who have electric space heating. PSEG Long Island reports installing roughly 19,000 heat pumps, leaving 232,000 accounts with electric resistance heating, who are also good candidates for heat pump replacement.

Table 4 Number of customers in PSEG Long Island's territory by predicted heating system fuel

Predicted heating fuel	Predicted Space heating fuel type
Delivered fuel	53,522
Electric	238,811
Gas	322,189
Grand Total	**604,331**

Figure 3 Proportion of customers in PSEG Long Island's territory by predicted space heating system fuel



SUMMARY, CONCLUSIONS, AND LIMITATIONS

As utility programs look to electrification as a means of reducing green-house gas emissions, a key challenge is isolating customers who have installed systems that are more economically viable to convert to heat pump space and water heating. Many utility programs operate with incomplete information regarding fuels used for systems they wish to electrify because they only provide electricity within a territory. For these utility programs, the trick is to identify customers who have delivered fuel or electric resistance heat for installed technologies.

There are vast data resources and data science tools now available to utility programs, program administrators and implementers. This paper showed how to overcome a key challenge to identifying customers who represent high valued prospects for electrification. Repurposing energy audit data enabled PSEG Long Island to conserve research dollars by using already available data reporting actual installed technologies rather than administering surveys to collect such information. We demonstrated how we could combine audit information with load-disaggregation analysis, building characteristic data, and past program history to construct a robust data set used to define a series of models to predict space and water heating system fuel type.

The results of this analysis provide the PSEG Long Island team with a list of high-value customers to target those who have either delivered fuel or less efficient electric resistance systems. In conclusion, TRC's PSEG Long Island residential program can use results from this analysis to target customers with existing delivered fuel or electric resistance heating systems. PSEG LI and TRC will be using the results of this analysis in their 2026 marketing and outreach strategy to help target high valued customers, i.e. those with delivered fuel or electric heating. TRC also incorporates the existing heating system type into propensity models we developed to identify customers who are most likely to participate in the next 12 months.

TRC identifies several limitations to this study. First, the analysis was limited to properties with existing energy audit data, which included single-family homes only. Second, use of energy audit data for the dependent variable does impose specification bias by inferring that the installed technology and characteristics of audit customers were like non-audit customers. Finally, the nature of machine learning models does not provide for variation in the likelihood of predicted values. The models only assign one fuel type for each customer and cannot be used to identify the probability of customers' different fuel types.

REFERENCES

1. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation | BMC Genomics | Full Text
2. Kaden, M. et al., (2014). Optimization of General Statistical Accuracy Measures for Classification Based on Learning Vector Quantization. ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence