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

IEPEC Denver, CO: October 7, 2025

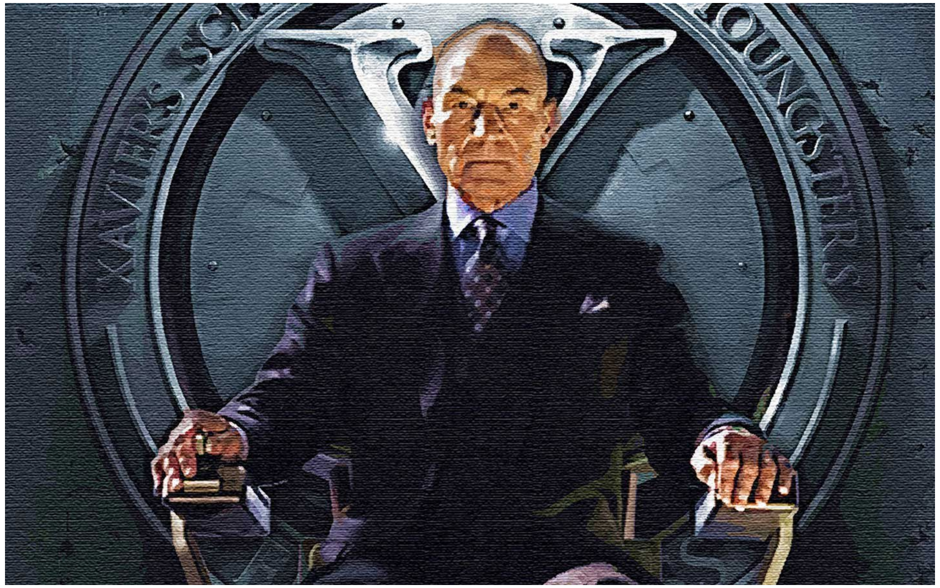
Presenter: Noel Stevens, TRC

Co-authors: Mohammed Hossain, Ph.D., Wendy Lu, Ph.D., Jake Schefrin, and Alfredo Jahn

Identify fuel used for space heating
Natural gas, electric resistance, delivered fuel

Objectives:

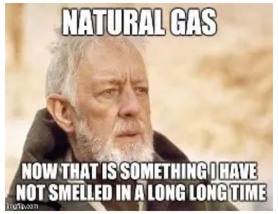
- Sample of known systems: Identify installed technology for heating systems in PSEG Long Island's territory.
- Describe all customers: Construct explanatory variables known for the entire population of customers.
- Model: Machine learning on a sample of customers with known heating systems to predict heating system fuel.
- Extrapolate: Apply the model to the population to find all single-family homes on Long Island with gas, electric, or delivered fuel for space heating.
- Evaluate: Compared predicted and actual fuel type for a separate group to test model performance.



Where are they?

Overview of approach
Machine learning models: Random forest, logit, k-neighbor, GB, XGB

Audit data

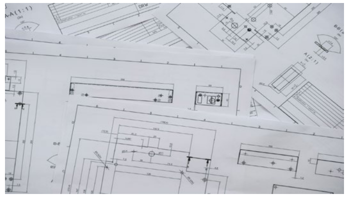


f

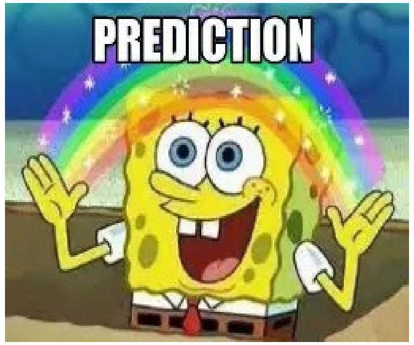
Consumption Data

Property tax data

Participation data



=



Apply to population

Test on independent dataset

Space heating Fuel is function of….

Electric heating, cooling, baseloads

Building and owner characteristics:

- SQFT
- Building age
- Assessed value,
- location

What we know about them:

- e.g., Pulled out gas water heater, or stove in 2013

Predict space heating fuel – all buildings:

- Apply the model to data each building in the state

Use separate sample

- Compare predicted results to known equipment

Selecting variables for ML models

Simple linear model – First pass

Model		Predicted Non-Gas	Predicted Gas	accuracy	recall	f1	MCC
Logit Classifier	Actual Non-Gas	5	503	69.4%	50.4%	81.9%	6.9%
	Actual Gas	1	1,140				
Gradient Boosting Classifier	Actual Non-Gas	221	287	81.3%	70.8%	87.9%	54.2%
	Actual Gas	22	1,119				
XGBoost	Actual Non-Gas	503	5	99.7%	99.5%	99.8%	99.3%
	Actual Gas	-	1,141				
Random Forest Classifier	**Actual Non-Gas**	**505**	**3**	**99.8%**	**99.7%**	**99.9%**	**99.6%**
	Actual Gas	**-**	**1,141**				
K Neighbors Classifier	Actual Non-Gas	218	290	75.9%	66.8%	83.9%	38.9%
	Actual Gas	107	1,034				

Upon review, the model excluded theoretically important variables.

- Heating load only – no cooling, base load, or total load.
- Did not differentiate properties by location (e.g., couldn’t tell who was on the gas network).
- Essentially built on building attributes and little else.

Data science needs a sanity check.

First pass – Looks good, huh?

K-fold evaluation – Models are highly predictive

- Random forest MCC - 99.6%
- XG Boost MCC - 99.3%

Very good at finding the pattern within the finite sample, but was it predictive of the broader population?

What you feed into machine learning matters!
Models can provide misleading results – Overfitting

Revised model added:

- Electric heating, cooling, and baseloads.
- Specific zip code indicators interacted with square footage.

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	***

- Logistic regression model is now the preferred model.

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

All we did was change the variables fed into the k-fold evaluation process.

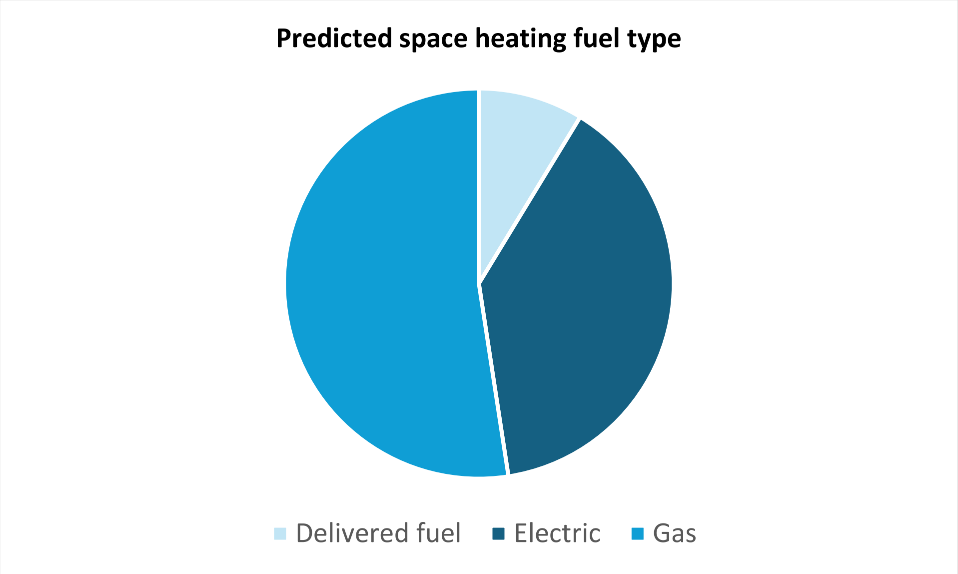
It is critical to have a theoretical basis for the variables you feed into the ML process.

Results

Estimated number of homes with natural gas, electric, and delivered fuel for space heating

- Applied explanatory variables from the logistic model to each customer in the population of single-family homes on Long Island.
- Predicted which homes had gas and non-gas space heating systems.
- Split out non-gas into electric and delivered fuel based on which homes had electric heating loads sufficient to account for electric heat.
- Provided predicted heating fuel for 604,000 single family homes on Long Island.

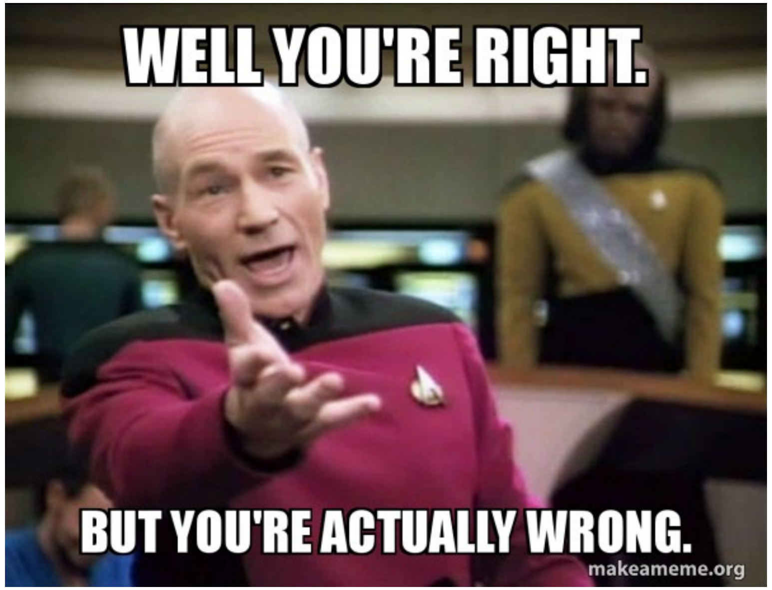
Distribution of homes by predicted heating fuel type on Long Island, NY.



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

Verification against an independent data set show a difference between the model's k-fold accuracy and "real world" accuracy in predicting

- Used a sample of 2,775 accounts from tracking data separate from those in the audit data.
- Compared the predicted space heating fuel to the actual space heating fuel reported in the program tracking records.



Comparison of predicted space heating system fuel and actual fuel from separate sample

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%	

- K-nearest neighbor model correctly predicted all gas accounts but only 5 non-gas accounts' gas systems.
- XGBoost and GBoost models predicted all systems to be gas, resulting in zero non-gas systems.

Conclusions and limitations

Conclusions

- Repurposing energy audit data can save budget on expensive primary research to track installed technologies for use in predictive models.
- Machine learning tools should build upon a solid theoretical model to ensure model results reflect known parameters essential to the system type.
- Models can appear to be highly predictive but are actually overfitting the data rather than finding strong predictors that are present outside the sample included in the k-fold evaluation.
- Identifying a separate data set to test predictions against actual data that were not included in the model estimation is necessary to gauge the accuracy of model results.

Limitations

- Our analysis was constrained by a limited sample of accounts with complete audit data. A larger sample of audits will likely improve model accuracy.
- Models only assign one fuel type for each customer and cannot be used to identify the probability of customers' different fuel types.

Questions?

Noel Stevens, Principal Consultant – TRC

Nstevens@trccompanies.com

Phone: 603-533-3091