Target journal: Applied Energy

Targeting home energy retrofit participants with machine learning algorithms: A case study from the Midwestern US

Authors/affiliations

D. Leith Nye, Douglas Ahl, Scott Pigg, Quinn Koenig

Keywords:

Energy efficiency, machine learning, home retrofit, utility program

Recommended reviewers

Need to include 3 potential reviewers

Abstract:

We demonstrate how "out-of-the-box" machine learning algorithms can be used to analyze historical home energy retrofit data and provide valuable information for programs aiming to increase home energy retrofit uptake rates through targeted marketing strategies. Model test performance results suggest that follow-through on major insulation and air-sealing upgrades would increase approximately two-fold if identified high-potential customers were targeted by the program, and that overall participation rates could increase more than ten percent. Machine learning algorithms can uncover patterns in complex data sets that can provide new insights into factors driving energy retrofit follow-through. For example, we observed that the probability of following through on audit recommendations was highly correlated with geographic location and proximity to other homes that receive an energy audit and undertake major retrofits. The predictive modeling approaches discussed in this paper can also be used to target different demographic groups based upon various program goals. Machine learning approaches have inherent limitations as predictions are built on the assumption that patterns in the data used for building models will hold in the future. For this reason, models should be updated regularly with new program data and predictions should be followed judiciously to avoid unexpected outcomes and unintended biases in program marketing.

Introduction:

Supported by economic and environmental analyses, residential energy efficiency investments have long been touted as a cost-effective and important strategy for curtailing greenhouse gas (GHG) emissions, reducing energy burdens and managing energy demands.[1] [2] As such, government agencies, utilities and nongovernmental organizations (NGOs) have offered structured programs and incentives to encourage owners of residential buildings to install

[1] Granade et al., "Unlocking Energy Efficiency in the U.S. Economy."
[2] Onat, Egilmez, and Tatari, "Towards Greening the U.S. Residential Building Stock."

efficiency measures such as insulation, air sealing and heating and cooling system replacements . With ongoing improvements in energy modeling and advanced metering infrastructure, energy efficiency programs can more accurately target buildings with the greatest technical potential for energy savings.[3] Despite incentives, societal drivers and technical developments, market penetration of residential energy retrofits continues to fall short of expectations.

This persistently low adoption rate of cost-effective home efficiency measures has been dubbed the “efficiency gap” or “energy efficiency paradox”.[4] The efficiency gap has been attributed, among other factors, to a lack of awareness of strategic efficiency opportunities on the part of the consumer.[5] A common strategy for overcoming this information deficit, implemented since the earliest residential energy conservation programs of the 1970s, is to offer free or subsidized “home energy audits” to guide homeowners through the retrofit process.[6] In this approach, an independent consultant evaluates the building shell and equipment and provides a prioritized list of cost-effective energy-saving measures to the consumer to act on. Typical high-priority recommendations include insulation, air sealing and major equipment upgrades. Efficiency programs offer various incentives for measures installed and may require the consultant to conduct a post-retrofit inspection for quality assurance.

Despite the prevalence of home audit programs among utilities and weatherization agencies nationally, program participation remains much lower than the economic potential of home retrofits. Based upon estimates from US Energy Information Administration’s 2009 and 2015 Residential Energy Consumption Survey (RECS), only about 3 percent of households have had a home energy audit. In addition, a recent national survey of home energy audits suggests that less than half of home owners who receive the audit follow through on a significant recommendation.[7] In many cases, the follow-through rate on major building envelope and heating system improvement recommendations is likely to be significantly less than 50%. For example, only approximately10-20% of customers followed through on any major building shell, heating system or water heater measure recommended in audits in a Canadian government weatherization program from 1998 to 2005.[8]

In the interest of designing more effective policies and programs, past research has sought to uncover the decision-making process around efficiency retrofits and to identify characteristics of households that are more likely to follow-through on major audit recommendations. Information and awareness about the efficiency options is one of many factors that affect the decision-making

[3] Tsanas and Xifara, “Accurate Quantitative Estimation of Energy Performance of Residential Buildings Using Statistical Machine Learning Tools”; Marasco and Kontokosta, “Applications of Machine Learning Methods to Identifying and Predicting Building Retrofit Opportunities”; Tian et al., “Data Driven Parallel Prediction of Building Energy Consumption Using Generative Adversarial Nets”; Swan and Ugursal, “Modeling of End-Use Energy Consumption in the Residential Sector”; Zhong et al., “Vector Field-Based Support Vector Regression for Building Energy Consumption Prediction.”
[4] Gerarden, Newell, and Stavins, “Assessing the Energy-Efficiency Gap”; Hirst and Brown, “Closing the Efficiency Gap”; Jaffe and Stavins, “The Energy-Efficiency Gap.”
[5] Palmer, Walls, and O’Keeffe, “Putting Information into Action.”
[6] Walker, Rauh, and Griffin, “A Review of the Residential Conservation Service Program.”
[7] Palmer et al., “Assessing the Energy-Efficiency Information Gap.”
[8] Gamtessa, “An Explanation of Residential Energy-Efficiency Retrofit Behavior in Canada.”

process. These motivations and constraints include financial considerations, perceived non-financial benefits, perceived hassle, the technical savings potential and various contextual influences such as household composition, demographics and tenure.[9] For example, Palmer and Walls (2015) found that households that were more aware of energy use, knew others who had gotten an audit and had no immediate plans to move were more likely to get an audit.[10] In a related study, Palmer et al (2015) found that the homeowner's decision about whether to follow-through on audit recommendations was influenced by factors related to the satisfaction with the audit and costs of recommended improvements.[11] Recent research has used multivariate probit models to analyze and illustrate how these factors come into play in households as they undertake the efficiency retrofits.[12]

By mining historic program data, other studies have uncovered factors associated with households that are more likely to make significant efficiency upgrades. Gamtessa (2013) applied linear logistic regression to analyze participation data for a Canadian home efficiency retrofit program and found that those who acted on major audit recommendations were more likely to have significant savings opportunities, live in older homes and have lower incomes. A similar analysis of the 2013 Canadian Households and Environment Survey found that the probability of following through on audit recommendations was correlated with demographic characteristics such as the household size, composition and age, as well as home age and available government incentives.[13] This suggests that there are opportunities to increase program impacts through targeting marketing and incentives toward specific demographic groups with a higher propensity for participation.

Like other industries, efficiency program administrators are increasingly recognizing the potential benefits incorporating advanced analytics techniques in their program design process and marketing plans, (e.g. Sheer et al 2017; Quaid 2016; Shaban et al 2016; Dahlquist 2016).[14] Many utility programs have access to rich historic data which includes energy usage, building shell characteristics, building locations, program participation information and sometimes detailed demographic information available from third parties. When available, geolocational information for buildings can be combined with public census data to explore geospatial demographic patterns in energy usage and program participation rates. While some factors associated with increased household propensity for efficiency retrofit may generalize across various program design structures and geographic regions, other factors will be program and region-specific. For this reason, there are benefits to programs mining their own program data for

[9] Wilson, Crane, and Chryssochoidis, "Why Do Homeowners Renovate Energy Efficiently?"
[10] Palmer and Walls, "Limited Attention and the Residential Energy Efficiency Gap."
[11] Palmer, Walls, and O'Keeffe, "Putting Information into Action."
[12] Wilson, Pettifor, and Chryssochoidis, "Quantitative Modelling of Why and How Homeowners Decide to Renovate Energy Efficiently."
[13] Das, Richman, and Brown, "Demographic Determinants of Canada's Households' Adoption of Energy Efficiency Measures."
[14] Scheer, Borgeson, and Rosendo, "Customer Targeting for Residential Energy Efficiency Programs: Enhancing Electricity Savings at the Meter"; Shaban, Khan, and Meisegeier, "Mining Energy Efficiency Program Data"; "Magic Powers: Customer Targeting with Machine Learning - 0194_0286_000378.Pdf."

insights. As “off-the-shelf” advanced analytic tools become more readily accessible, this becomes increasingly feasible for organizations.

The goal of this study was to explore how readily-available, open-source machine learning tools could be used to predict different characteristics associated with household propensity to participate in a utility home energy retrofit program. One of the benefits of machine learning algorithms is that they can more readily model complex, non-linear patterns in data in ways that are more difficult with traditional approaches like linear regression.

Working with a cohort of over one million residential customers for natural gas customers of a Midwestern U.S. utility, we compared the performance of several standard machine-learning approaches for predicting probabilities that customers would both have significant retrofit opportunities and act on audit recommendations as a function of monthly gas usage, building characteristics, household demographics and spatial proximity to other program participants. We demonstrate here how the best models could be used to strategically target customers for program participation, identify some of the key predictors of participation and discuss some of the potential limitations of using machine learning in this context. Although prior studies have used machine learning algorithms to predict energy usage in buildings and the technical savings potential, this is the first study that we are aware of to use these methods to predict both technical potential as well as household propensity for efficiency retrofits.

Methodology:

The problem

It is common in the U.S. for natural gas residential retrofit programs to provide subsidized diagnostic energy audits and offer rebates for specific energy efficiency improvements such as attic and wall insulation, air sealing, and high efficiency heating systems. In addition, income-qualified customers may receive higher incentive levels for installed measures. Owner-occupied, single-family homes typically participate the most in these programs because they both receive the benefits of weatherization and have an incentive to invest in the home. The natural gas retrofit program used in this study conducts an audit on 3,000 to 6,000 customers’ homes annually. Of these, an average of 35% follow through on major recommended improvements to the building shell or heating system.

To increase the energy savings per dollar invested in the program, the utility programs generally would like to target customers that: a) live in homes with significant energy savings potential; b) are likely to seek an energy audit; and, c) are likely to follow-through on major recommended efficiency upgrades. In addition, the utility programs that focus on serving lower-income customers are interested to know the customers whose income level would qualify them for additional incentives.

Approach

Our objective was to quantify how well readily available machine learning tools could predict key targeted attributes and outcomes of past program participants using relevant data that is currently available. Because our goal was to predict the potential of households for program participation *prior* to receiving an audit, we were constrained to using data sources that could be obtained without doing an on-site visit to the premise. This includes data on past natural gas consumption, building characteristics, household and neighborhood demographics and geospatial patterns.

Using these potential predictors, we trained logistic regression, k-nearest neighbors, support vector machines, random forests, and gradient boosting machines models to classify binary outcomes observed in audit recommendations and past program participant actions for 2012 through 2018. We compared model performance on test datasets and selected the best models to make predictions on the full customer database and identify a subset of customers predicted to have a higher propensity for program participation and a higher propensity of needing rebated measures. We then compared actual program outcomes for 2019 to model predictions. Table 1 summarizes the approach and training data used to build classification models to identify each of the four target groups. For example, to identify homes that are likely to need insulation, air sealing or a new heating system we trained models to predict the inclusion of these recommendations in past audit reports based on existing information about the building and household characteristics

Table 1: Summary of approach and historical data used to predict inclusion in target group. The sample size represents the number of homes used for the model training process prior to splitting between training and test datasets. For each dataset, 80% of the records were used in the training and the remaining 20% were reserved for the test data.

Target group	*Approach*	*Binary variable(s)*	*Training data description*	*Sample size*	*Prevalence of positive class*
Homes likely to have significant savings opportunities	Predict the inclusion of insulation, air sealing or heating system upgrades in prior audit reports	Recommend: -Heating system? -Air sealing? -Insulation?	Audit recommendations for sample of prior program participants	11,534	*27% Heating rec.* *61% Insulation rec.* *61% Air sealing rec.*
Households likely to get audits	Predict that a customer will get an audit through the program.	Will get audit?	All customers in service territory with complete data	665,437	*5% get audits*
Households likely to make significant upgrades after audit	Predict that a program participant will act on air sealing, insulation and heating-system recommendations after an audit.	Will install major measure?	Outcomes for a sample of prior program participants	15,858	*37% install at least one major measure*
Households likely to be income-qualified	Predict that program participants will be income-qualified.	Is income-qualified?	Sample of past program participants with income classification	23,442	*34% income-qualified*

Predictor variables

The available data sources used for training models fell into five categories. These categories and their associated variables associated are described in detail below and summarized in Table 2.

Natural gas consumption

For each home and each program year (2012-2018), we calculated weather-normalized annual consumption—as well as statistically disaggregated estimates of consumption for space-heating, and non-heating end uses—using a variable-base heating degree day regression model.[15] Keeping only data for years to prior to getting an audit where there was a good fit for the consumption (model $r^2 > 0.7$), we calculated long-term and recent normalized gas consumption. When building area was available, we calculated the energy use intensity (mean normalized annual gas consumption per square meter).

Building attributes

Home audit data from each program year was combined with purchased third-party data which included actual or estimated building attributes for many homes that might be predictive of household retrofit potential and propensity. These included building area, building age, estimated home value, property type, and exterior cover type (siding).

Household-level demographics

Also available with the third-party marketing data were variables that estimate the demographic characteristics of the household. These included the number of adults and children, educational level, length of residence, home ownership status, household composition, marital status and age of household members.

Neighborhood-level demographics

U.S. Census Bureau tract-level data was matched to home addresses and included information on the estimated number of people who identified in different racial groups, median income, overall median age and median age by gender.

Geospatial attributes

We estimated the geographic coordinates for most addresses in the customer database using a free online geocoding service provided by the U.S. Census Bureau. Geographic coordinates, latitude and longitude, were then transformed to the Universal Transverse Mercator (UTM) projected coordinate system zone designation for the region using a geographic information system. These UTM coordinates were included as predictors for training models. We also generated variables that indicated the proximity of a customer to other customers who have participated in the retrofit program either by getting an audit or by both getting an audit and acting on major efficiency upgrades. Because we were not sure at what scale potential neighborhood influences might be occurring, we calculated this predictor in different ways, including distance to closest one, five or ten program participants or the number of participants within expanding radii from 500 to 40,000 meters (Table 2).

---

[15] Fels, "PRISM."

Data Preprocessing

We took some initial steps to separate the customer list into those who would likely qualify for the home retrofit program and those with very low incomes that would qualify for a separate program. The separate program receives commitments to install major measures before the audit and is completely free to the customers. We wanted to remove them to reduce the potential confounding effects of free services to predicting whether or not a customer will participate in the home retrofit program. Therefore, we removed households that had either previously participated in the low-income program or that would likely qualify based upon the fact that they had recently gotten billing assistance from the utility. Similarly, we also removed multifamily dwellings and past participants in a new homes program because they likely would not qualify for this specific home retrofit program. The utility also runs a marketing campaign in selected geographic areas involving working with local leaders to drive participation in the program using special promotions for that area. These customers were also removed as they would create a stronger correlation with the geospatial attributes than would be observed in the general population.

Most of the modeling approaches that we used could not handle missing values. One way to deal with this is to impute missing data. Since imputation adds and additional layer of uncertainty to model predictions and since we had a relatively large dataset that was representative of the full customer population, we decided not to impute data. Instead we selectively removed predictors if they had a high proportion of missing values, i.e. >15%, and then removed records that had missing values for any of the remaining predictors.

Prior to model training, we also screened predictor variables with either near-zero variance or that were highly collinear (correlation > 0.9). The final set of 34 predictor variables were then centered, scaled and Box-Cox transformed. We used R (version 3.6.1), a free software environment for statistical computing and graphics, for all data processing and analysis steps.[16] In particular we relied heavily on the tools and methods included in the "caret" (Classification and Regression Training) package version 6.0-84.[17] The final analytic dataset used for model training and testing included 45 predictor variables between 11,534 and up to 665,437 records depending on the data available for the response variable. For example, we had measure recommendations from audits and predictor variables for 11,534 homes but income qualification status and complete predictors for 23,442 homes (Table 1).

Table 2. Final list of predictors used in model training

Category	Variable
Building attributes (5)	Conditioned area
	Home value
	Siding type
	Year built
	Residence type

[16] R Core Team, *R: A Language and Environment for Statistical Computing*.
[17] Wing et al., *Caret: Classification and Regression Training*.

Neighborhood demographics (8)	Median age (female)
	Median age (male)
	Asian population
	African American population
	Other race population
	White population
	Mixed race population
Gas consumption (4)	Normalized annual consumption (past year)
	Normalized annual base consumption (historic mean)
	Normalized annual base consumption (past year)
	Heating energy use intensity (mean annual heating consumption per square foot in past year)
Household demographics (8)	Number of adults
	Number of children
	Highest education level
	Household composition category
	Home ownership status
	Length of residence
	Marital status
	Birth date
Geospatial attributes (9)	Longitude (UTM)
	Latitude (UTM)
	Distance to nearest participant
	Mean distance to closest 5 participants
	Number of participants within 500 meters
	Number of participants within 2,000 meters
	Number of participants within 5,000 meters
	Number of participants within 10,000 meters
	Number of participants within 20,000 meters
	Number of participants within 40,000 meters

Model Training, Evaluation and Interpretation

For each of the target binary response variables, we initially compared the performance of five predictive modeling approaches implemented in R packages: logistic regression (base R), k-nearest neighbors (“class” package v7.3-15)[18], support vector machines with radial basis function kernel (“kernlab” package v0.9-27)[19], random forests (“randomForest” package v4.6-14)[20], and gradient boosting machines (“gbm” package v2.1.5)[21]. For each target response variable, we allocated 80% of the data for model training and the remaining 20% was set aside to

[18] Venables and Ripley, *Modern Applied Statistics with S*.
[19] Karatzoglou et al., “Kernlab – An S4 Package for Kernel Methods in R.”
[20] Liaw and Wiener, “Classification and Regression by RandomForest.”
[21] Greenwell et al., *Gbm: Generalized Boosted Regression Models*.

test final model performance (Table 1). When predicting unbalanced classes, such as audit uptake where only 5% of customers have gotten audits, we compared the effect of different up- and down-sampling techniques.

All models were trained to optimize overall accuracy using 10-fold cross validation. Final model performance was evaluated based upon predictions on the test data. We compared their overall accuracy (ratio of correctly predicted observation to the total observations) compared to the no information rate (accuracy attributed largest proportion of observed outcomes), precision (ratio of true positives over sum of true and false positives) and sensitivity (ratio of true positives over sum of true positives and false negatives). Our objective was to identify a cohort of customers that should have a higher rate of program qualification and participation than the historic baseline. For this reason, our primary criteria for model selection was a balance of precision and sensitivity. For all target response categories, we found that iterative tree-based classification methods, especially random forest models, consistently performed better than other methods. Table 3 shows the test data metrics for the five modeling approaches when predicting whether a home will install a major measure recommended in an audit. Although not shown here, the random forest and gradient boosting machine algorithms performed better than the other methods for all customer participation categories.

Table 3. Test data metrics for models used to predict probability that customer would install major measure recommended in an audit. Random forest and gradient boosting machine algorithms performed better than the other methods for all program participation categories.

MODEL	*Accuracy*	*Precision*	*Sensitivity*	*Neg. Pred. Value*
Random Forest	0.79	0.77	0.59	0.79
Gradient Boosting Machines	0.77	0.72	0.59	0.79
Support Vector Machines	0.70	0.66	0.36	0.71
K-Nearest Neighbors	0.69	0.67	0.31	0.70
Logistic Regression	0.69	0.63	0.38	0.71

We then used the selected random forest models to make predictions on the larger customer data set and select a cohort of customers with a greater than 50% probability of a positive response for each target category. Using the model performance measures from the test data, i.e. precision, we estimated the counts of expected true positive and false positives in each customer cohort. These numbers were then used to evaluate the theoretical potential for future program participation and consider applications for targeted marketing campaigns. To further assess the expected accuracy predictions in future years, we compared the model predictions generated using the 2013-2018 data to the outcome observed in 2019. We compared the model accuracy metrics for predicting outcomes in 2019, including participation detection rate, i.e. sensitivity, to that calculated using the historic test data. We also compared the mean predicted probability participation among 2019 participants to the mean probabilities for all potential participants.

Untangling the detailed mechanics for how random forest machine learning algorithms are making predictions can be difficult or impossible. However, by comparing the Gini index for model training iterations when a variable was included versus excluded a relative “variable importance” score can be estimated for each predictor. By comparing the variable importance metrics for predictors and exploring patterns in predictor values among customers expected to fall into each participation category, we gleaned insights into some of the important determinants of program participation. For each predicted participation category, we compared the mean importance of predictor variables falling in each of the five categories shown in Table 2.

To further explore potential patterns driving model predictions, we compared how predictor values varied between new customers (potential future program participants) assigned to different participation categories by the random forest model. For example, we examined how building age correlated with probability of predicting insulation recommendations and examined how mean distance to past program participants correlated with probability of predicting a new household’s participation. Given that spatial variables were important for many of models, we also used contour plots to explore spatial patterns in the distribution of future program participants predicted to fall in different categories.

Results

Model performance and predictions

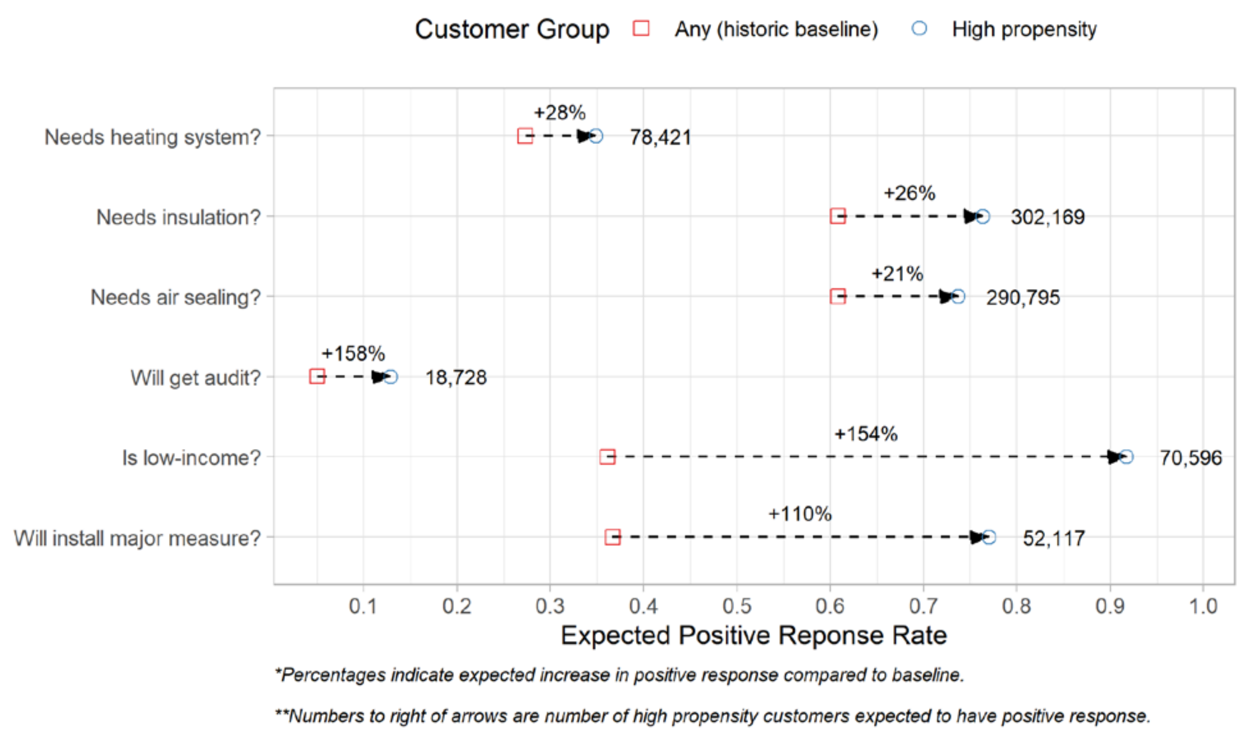
The performance metrics for the random forest model predictions on the test data for each of the six target program participation categories are summarized in Table 3. Only the models for predicting the recommendation of a heating system and that a customer would get an audit had accuracy below the no information rate, i.e. the level of accuracy that could be achieved by just choosing the category of highest proportion participation. The test precision for all models exceeded the baseline positive prediction rate, i.e. the proportion of the positive class in the test and training data. This indicates that the models are able to identify a subset of households that have a higher rate of program qualification and propensity for participation. For example, although on average 37% of program participants act on major retrofit recommendations after an audit, this rate of 67% among the group the model predicts will follow through on recommendations. For all response categories, the cutoff for assigning a positive outcome was 0.5 or greater predicted probability.

Table 3. *Test data performance metrics for the best random forest models trained to predict propensity for of households to meet program participation criteria. Models that more effectively classify customers in each category will have an overall accuracy greater than the “no information rate” and have a precision greater than the baseline positive prediction rate. *Indicates test accuracy greater than the no information rate.*

CATEGORY	*Accuracy*	*Precision*	*Sensitivity*	*Baseline Pos. Pred. Rate*	*No info. rate*
Needs air sealing?	0.74*	0.74	0.88	0.61	0.61
Needs insulation?	0.76*	0.76	0.88	0.61	0.61
Needs heating system?	0.58	0.35	0.63	0.27	0.73
Is income-qualified?	0.84*	0.92	0.61	0.36	0.64
Will get audit?	0.76	0.13	0.68	0.05	0.95
Will install major measure?	0.78*	0.77	0.59	0.37	0.63

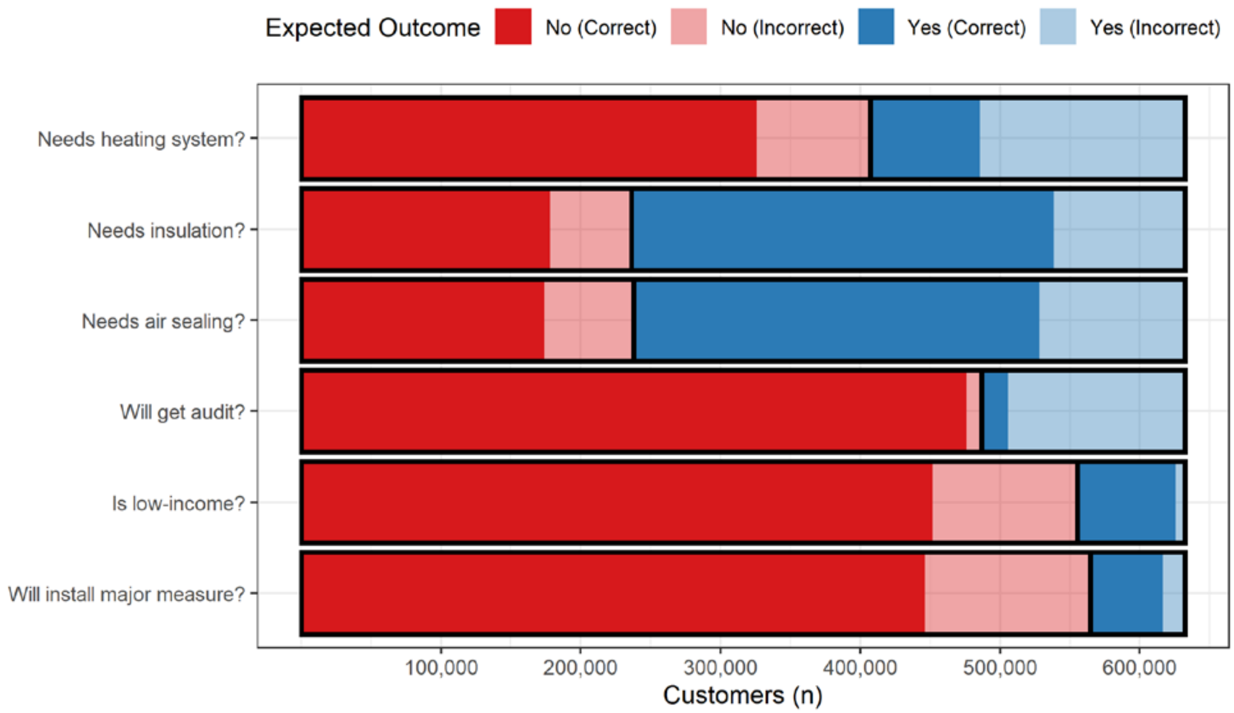
Figure 1 provides a visual comparison of the rate of positive outcomes for all customers in the test data compared to the subset predicted by the model to have a positive outcome, i.e. to need insulation, get an audit, etc. For predicting measures that would be recommended on an audit report, the increase in rate of the positive class among those predicted to need a measure is modest (23-40%). For other categories, the relative change is more substantial (110-172%). However, this conversion rate change is only useful if models can identify a sufficiently large cohort of customers. The numbers to the right of each line represent the expected number of correct predictions among 630 thousand customers. For example, the model predicts that approximately 145 thousand customers would get audits but, based upon the measured precision for the test data, we would only expect about 19 thousand (13%) of those prediction to be correct. On the other, the model was able to accurately identify a subset of customers expected to be income-qualified for additional incentives. Of the 77 thousand predicted to be income-qualified, we would expect 70 thousand (92%) to be classified correctly.

Figure 1. Comparison between baseline rate of positive outcomes to true postive rate of customers identifed by the model. The numbers to the right of each line represent the number of correct predictions out of 630 thousand potential customers.



Similarly, by applying the observed error rates from the test data to the predictions on the larger set of 630 thousand potential participants, we could estimate the number of positive outcomes that the models likely will not detect. Figure 2 shows the estimated counts of false positive and false negative errors for these predictions. For example, the model for predicting income qualification would be expected to have a low false positive rate (8%), but will likely misclassify a larger fraction of customers predicted not to be low-income (19% false negative rate) and therefore miss many customers who are truly income-qualified.

Figure 2. Estimated outcomes for the 630,000 potential customers predicted to fall in each category. Misclassifcation counts rates were estimated by applying the error rates from the measured on the test data to the predictions on the new customer data.



We found that the performance metrics for predicting the outcomes for 2019 participants was similar to the historical test data metrics for predicting air sealing or insulation recommendations and income status (Table 4) but performance metrics were generally lower than the test data for the other categories. For example, 81% of the customers with insulation recommendations in an audit were among the cohort predicted to need insulation by the random forest model.

The detection rate was lower for predicting audits and whether customers would install a major measure. Only 16% of the 2019 households who installed major measures were among those predicted by the models to do so and 46% of those predicted to install a major measure were among those who did. This number is still higher than the overall installation rate for the year 32% which suggests that the model was able to discriminate among customers with a higher probability of following through on recommendations. We thought that the measure installation detection rate might be low in part because customers with audits later in 2019 have not had time to follow through on recommendations. But we saw similar results when analyzing only households that received audits in the first six months of 2019. Audits and heating system recommendations were rare which makes predicting them inherently challenging. Less than 1% customers get audits on a given year and only 5% of the customers who got audits in 2019 had heating system replacement recommendations both of which are much lower than the overall fractions of customers in these groups, 5% and 27% respectively (Table 3).

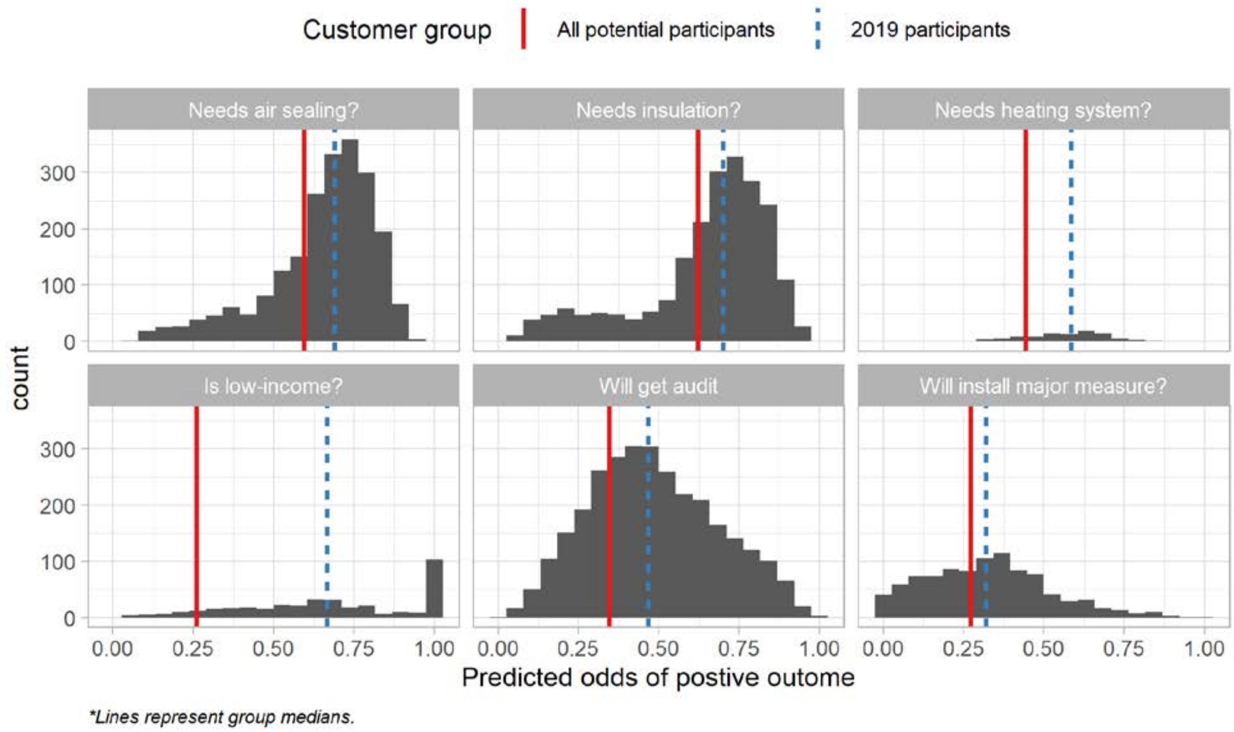
Table 4. Model prediction metrics for the out-of-sample 2019 participation data. "" Indicates accuracy greater than they no information rate for 2019.*

Category	*Participants in 2019*	*No Info. Rate (2019)*	*Accuracy*	*Precision*	*Sensitivity*	*Neg. Pred. Val.*
Needs air sealing?	2,139	0.72	0.81*	0.89	0.84	0.74
Needs insulation?	2,119	0.71	0.78*	0.87	0.81	0.70
Needs heating system?	93	0.03	0.58	0.05	0.73	0.58
Is low-income?	382	0.87	0.94*	0.80	0.72	0.97
Will get audit	2,974	<0.99	0.77	0.01	0.44	0.77
Will install major measure?	951	0.68	0.67	0.47	0.16	0.91

.

We observed that for all categories the median predicted odds of participation was higher for those households in participating in 2019 compared to all households in the customer base (Figure 3). This suggests that future program participation is correlated with higher predicted odds of participation. Median predicted odds that a household would be income-qualified was 26% for all customers but 67% among the 2019 participants who were income qualified, a 41% absolute difference. The difference between groups was more modest for other categories absolute increases of 5-14% compared to the median predicted odds for all customers.

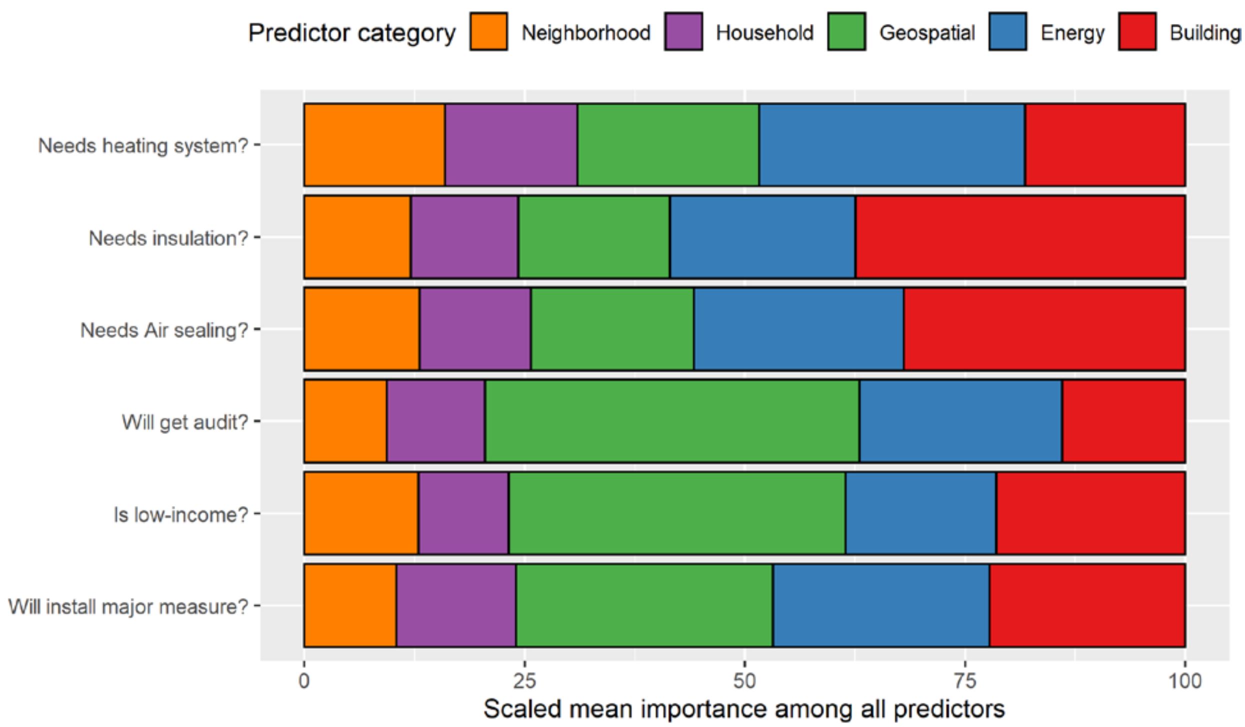
Figure 3. Distribution of predicted odds of household falling into categories for 2019 participants. Vertical lines represent the median odds of all households compared to the households participating in 2019.



Explanatory variables

We compared of the relative importance of predictor variables in different categories for improving model accuracy. Figure 3 shows the fractional importance of variables in each category among the twenty variables with the highest importance values. We noted that models predicting some program attributes like insulation and air sealing opportunities, rely more about building attributes, while model predictions of program participation outcomes (audits, measure installations) are driven more by geospatial patterns. Overall neighborhood demographic predictors were the least important for model performance.

Figure 4. Approximate relative contribution of variables in each category associated with a target outcome based upon the scaled mean importance value.



Comparing the relative contribution of different predictor categories, however, can mask cases where one or several has outsized importance compared to others. For example, home age is five times more important than the next most important variables (energy consumption) for predicting air sealing and insulation opportunities. Similarly, home age is twice as important as the next variable (distance to other participants) for predicting whether a customer will act on major efficiency improvements (Table 5). Among the household demographic predictors, the household composition is the one that appears most frequently.

Table 5. Top 5 variables by relative importance value

Response variable	*Predictor (relative importance value)*
Needs Air sealing?	1. Year built (100) 2. Household composition category (26) 3. Birth date (24) 4. Heating energy use intensity (23) 5. Normalized annual consumption (past year) (23)
Will get audit?	1. Mean distance to closest 5 participants (100) 2. Number of participants within 1,000 meters (69) 3. Home value (39) 4. Household composition category (34) 5. Longitude (34)

Will install major measure?	1. Year built (95) 2. Number of participants within 5,000 meters (61) 3. Longitude (56) 4. Household composition category (52) 5. Normalized annual consumption (past year) (45)
Needs heating system?	1. Heating energy use intensity (past year) (100) 2. Year built (100) 3. Household composition category (99) 4. Normalized annual consumption (past year) (80) 5. Home value (82)
Is low-income?	1. Home value (100) 2. Number of participants within 5,000 meters (93) 3. Longitude (64) 4. Number of participants within 20,000 meters (57) 5. Latitude (52)
Needs insulation?	1. Year built (100) 2. Home value (22) 3. Household composition category (22) 4. Heating energy use intensity (past year) (18) 5. Birth date (17)

Random forest models can generate predictions based on non-linear patterns and multiple correlations between predictors. As such it can be difficult to uncover simple mechanistic explanations for why one customer would be assigned a probability to be in a class by the model. However, by plotting the model generated probabilities of positive outcomes in the test data versus continuous variables with high importance, we can see some strong linear patterns.

For example, whether an auditor will recommend insulation is mostly dependent on the age of the home and a customer must receive a recommendation to be able to receive rebates on that measure. The random forest model detects this pattern and assigns a much higher probability of getting the insulation recommendation homes built before 1975 (Figure 4). Similar to the prediction of insulation needs, the model is predicting whether a household will be income qualified primarily based on the estimated home value. The median home value for past income qualified customers in the training data is 117 thousand dollars compared to 215 thousand dollars for customers not income-qualified. However, whether a participant installs major upgrades, is correlated with household proximity to past program participants, i.e. homes who have made major upgrades through the program. As shown in Figure 4, we see that the models are detecting a negative relationship between distance to past participants and likelihood of following through on heating, insulation or insulation improvements. In other words, there appears to be a significant “neighborhood effect” or spatial clustering associated with program participation.

Figure 4. Predicted probability of receiving recommendation for insulation improvements on audit report. For clarity, homes built before 1900 were excluded. The actual outcome is indicated by point color.

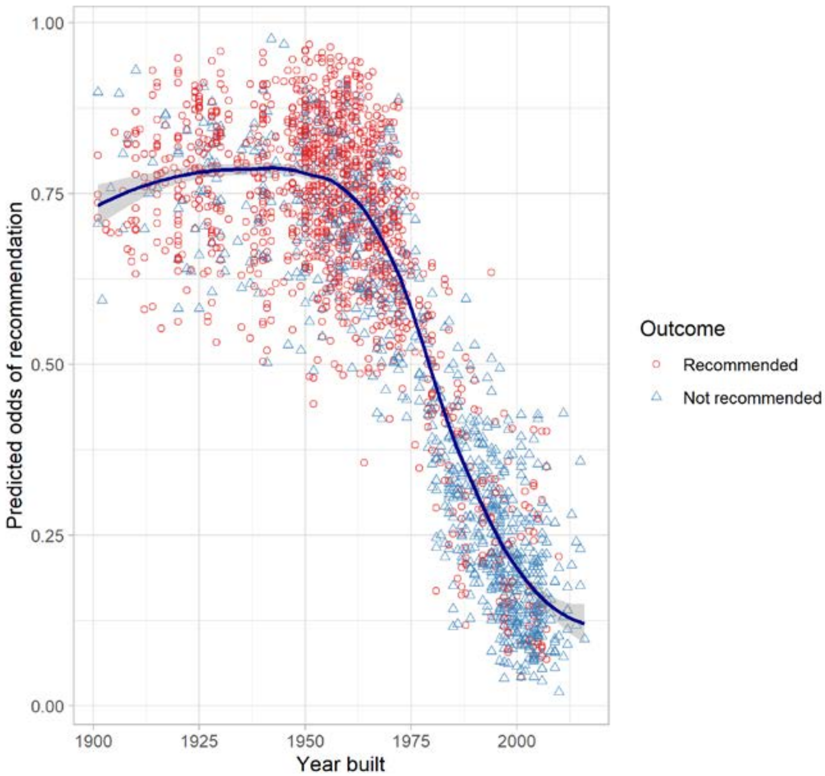
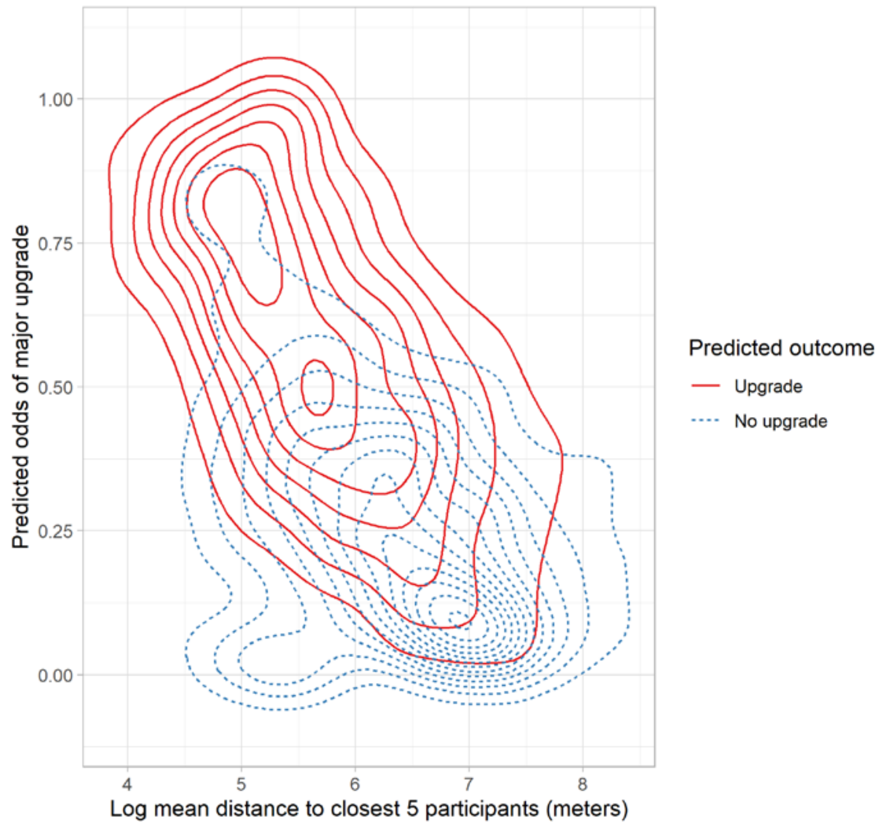


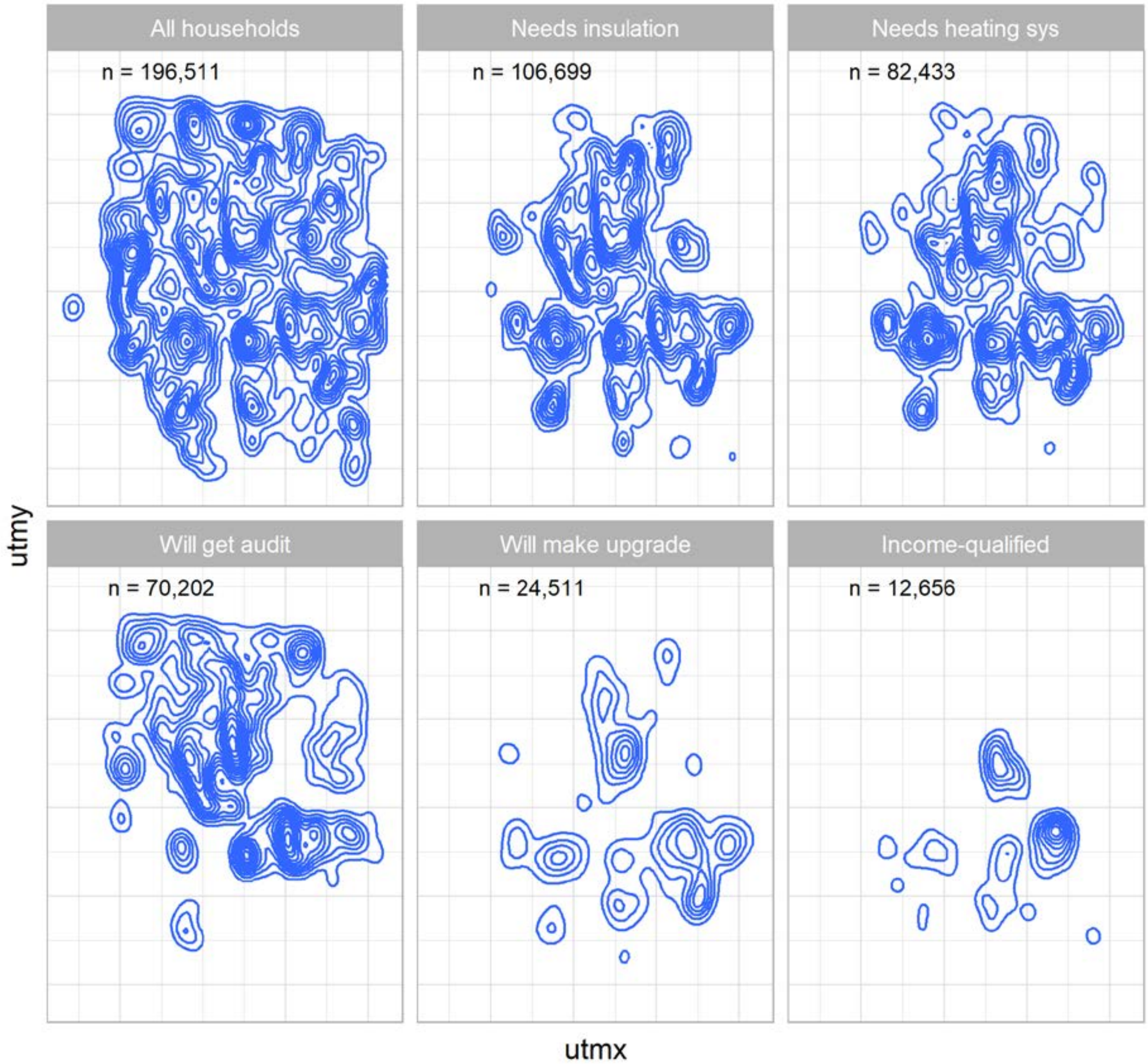
Figure 5.Contour plot of test data predicted probability of installing a major measure after an audit as a function of mean distance to the closest five households which previously installed measures through the program. Although there is significant noise in the relationship, the random forest model assigns lower probability of installing measures with increasing distance of a household from past program participants.



This apparent "neighborhood effect" is visible when comparing the the spatial distribution of new customers expected to fall in different program participation categories. When examining the geographic distribution of potential new customers predicted to meet program participation criteria, we observe spatial clustering of the households expected to fall in different categories compared to the overall distribution of households in a region. For the same county in the utility territory, for example, we see that predicted spatial density for households predicted to need insulation or a new heating system are very similar but more centrally clustered compared to the overall distribution of homes in the county (Figure 6). The distribution of customers predicted to get an audit is overlapping but somewhat shifted compared to the insulation and heating categories. Similarly, the distribution of customers predicted to install a major measure or be income-qualified show a similar distribution in specific regions but with significant overlap with the other groups.

Figure 6. Contour plot showing the spatial density of potential new program participants by predicted program participation category for a sample county in the utility territory. Categories include all potential households with sufficiency data for predictions (top left), households predicted to need insulation (top center), those predicted to need a heating system upgrade (top right), those predicted to get an audit (bottom left), those predicted to install major upgrade (bottom center) and those predicted

to be income-qualified (bottom right). UTM coordinates have been centered and scaled. Numbers represent the number of households in each category.



Discussion and conclusion:

Readily available machine learning algorithms can have a useful role in helping to explain household efficiency retrofit participation. To yield significant energy savings, a participating household must have both significant technical potential *and* high propensity to make major upgrades. The technical savings potential is determined primarily by the building characteristics. The household propensity to act on efficiency improvements is determined by a variety of contextual and demographic factors that are difficult to measure and model. In predicting which past program participants would make significant air sealing, insulation and heating system improvements after an audit, random forest models can uncover patterns in data that indicate a higher probability that *both* technical potential and household propensity exists.

We saw that technical potential for energy savings through building shell improvements appeared to be largely a function of building attributes such as home age, but also energy usage and, to a lesser extent location. This observation aligns with our unpublished analyses of home energy audit data for the same utility as well as the results of national occupant surveys about building insulation levels.[22] Building codes, especially after 1980 typically have more stringent requirements for insulation. Because insulation predictions are based upon observations and recommendations made during a sample of home energy audit, it is also possible that there is some bias on the part of auditors to recommend additional insulation in older homes.

Compared to air sealing and insulation, the models were less accurate in identifying homes that would need a new heating system. The two most important variables for predicting heating system replacement opportunities were home age and heating energy use intensity (therms/m^2.). Even if a building has unusually high heating intensity, without additional on-site measurements, it is difficult to determine from the available data whether this would indicate an inefficient heating system, inefficient building shell, or both.

In addition to building and energy usage, technical opportunities were also related to spatial variables but to a lesser extent. The observed spatial patterns are likely due in part to correlations between building characteristics and location. For example, the models could be targeting neighborhoods with a higher fraction of older homes. Similar correlations likely exist between estimated home value and location as lower and higher-income neighborhoods are often spatially segregated.[23] As a customer's propensity to be income-qualified was primarily determined by the home value this may explain the some of the spatial patterns in homes predicted to fall in this category.

Although building attributes were more important variables in prediction building technical potential building efficiency upgrades and income status, spatial, household and neighborhood characteristics were relatively more important for models predicting *propensity* for program participation, including getting audits and installing major upgrades. Indeed, the top five variables for predicting whether a household would get an audit related to proximity to other customers who participated in the program. However, it is worth noting that the model was not very effective at identifying customers who would get audits with a test precision of only 13% and a sensitivity of 70%. Assuming a customer got an audit, however, the random forest model was more effective at identifying which customers would follow through on major measures with a test precision 77% and sensitivity of 59%. Customers that install major upgrades must have both a technical opportunity for insulation, air sealing or heating system replacement and propensity to do so. This is reflected in combination of important variables which include home age, energy usage, location and proximity to past participants. Although we do observe some evidence of an apparent neighborhood effect where potential customers who are near past participants are predicted to have a higher odds of participating in a program it is not clear to what extent this reflects local diffusion of information between households, unmeasured

[22] Hojjati, *Householder's Perceptions of Insulation Adequacy and Drafts in the Home in 2001*.
[23] Owens, "Building Inequality."

variables that correlate spatially, or a reinforcement of historic program marketing efforts which might target specific areas in the utility territory.

With further research and analyses it might be possible to improve understanding of how combinations of spatial patterns, neighborhood-level diffusion of information, building attributes and demographic characteristics determine the household decision-making process around making energy efficiency retrofits. One of the benefits of using machine learning methods is that they can allow us to leverage these existing patterns in the data to make useful predictions without having to understand the nuances of how the model works. As such, we see machine learning methods as tool that has potential to be used judiciously by home retrofit program implementers to mine existing program data to develop predictions about new customers who have a greater likelihood for successful participation. Using the results of the models presented here, for example, a program implementer might develop a tiered marketing and outreach approach where broader "light touch" campaigns target the larger cohort of customers predicted to need major building shell or heating system measures. More intensive outreach efforts could target the smaller cohort of customers that are predicted to install a major upgrade, since we would feel more confident that this group had both technical savings opportunities and propensity to participate. Based on the error rates observed in the test data, participation rates in these targeted groups should be significantly higher than the historical baseline.

As noted, the results of these models and the predictions should be used judiciously to guide program design and outreach strategy. The predictive models generate predictions based the assumption that patterns driving participation, including incentive levels and program structure, that existed in the historic data used for model training will continue in the future. They are also built on the assumption that the sample of households used in the model training process are representative of the characteristics of the larger potential customer based on which new predictions are generated.[24] For this reason, care should be taken both in training models on a representative data set, and in interpreting predictions made on a larger population of customers. We advocate the use of machine learning approaches as one additional tool at program managers disposal. The information generated by these models should evaluated along with existing sources of information and evolving strategies but could provide additional value for ratepayer funded utility programs.

Acknowledgements

This work was funded jointly by a Midwestern U.S. utility and Slipstream Inc.

References:

Das, Runa, Russell Richman, and Craig Brown. "Demographic Determinants of Canada's Households' Adoption of Energy Efficiency Measures: Observations from the Households and Environment Survey, 2013." *Energy Efficiency* 11, no. 2 (February 1, 2018): 465–82. https://doi.org/10.1007/s12053-017-9578-4.

[24] Riley, "Three Pitfalls to Avoid in Machine Learning."

Fels, Margaret F. "PRISM: An Introduction." *Energy and Buildings* 9, no. 1–2 (February 1986): 5–18. https://doi.org/10.1016/0378-7788(86)90003-4.

Gamtessa, Samuel Faye. "An Explanation of Residential Energy-Efficiency Retrofit Behavior in Canada." *Energy and Buildings* 57 (February 1, 2013): 155–64. https://doi.org/10.1016/j.enbuild.2012.11.006.

Gerarden, Todd D., Richard G. Newell, and Robert N. Stavins. "Assessing the Energy-Efficiency Gap." *Journal of Economic Literature* 55, no. 4 (December 2017): 1486–1525. https://doi.org/10.1257/jel.20161360.

Granade, Hannah Choi, Jon Creyts, Anton Derkach, Philip Farese, Scott Nyquist, and Ken Ostrowski. "Unlocking Energy Efficiency in the U.S. Economy," n.d., 165.

Greenwell, Brandon, Bradley Boehmke, Jay Cunningham, and G. B. M. Developers. *Gbm: Generalized Boosted Regression Models*, 2019. https://CRAN.R-project.org/package=gbm.

Hirst, Eric, and Marilyn Brown. "Closing the Efficiency Gap: Barriers to the Efficient Use of Energy." *Resources, Conservation and Recycling* 3, no. 4 (June 1, 1990): 267–81. https://doi.org/10.1016/0921-3449(90)90023-W.

Hojjati, Behjat. *Householder's Perceptions of Insulation Adequacy and Drafts in the Home in 2001*, 2004.

Jaffe, Adam B, and Robert N Stavins. "The Energy-Efficiency Gap." *Energy Policy* 22, no. 10 (1994): 7.

Karatzoglou, Alexandros, Alex Smola, Kurt Hornik, and Achim Zeileis. "Kernlab – An S4 Package for Kernel Methods in R." *Journal of Statistical Software* 11, no. 9 (2004): 1–20.

Liaw, Andy, and Matthew Wiener. "Classification and Regression by RandomForest." *R News* 2, no. 3 (2002): 18–22.

"Magic Powers: Customer Targeting with Machine Learning - 0194_0286_000378.Pdf." Accessed May 1, 2019. https://aceee.org/files/proceedings/2018/node_modules/pdfjs-dist-viewer-min/build/minified/web/viewer.html?file=../../../../../assets/attachments/0194_0286_000378.pdf#search=%22dahlquist%22.

Marasco, Daniel E., and Constantine E. Kontokosta. "Applications of Machine Learning Methods to Identifying and Predicting Building Retrofit Opportunities." *Energy and Buildings* 128 (September 15, 2016): 431–41. https://doi.org/10.1016/j.enbuild.2016.06.092.

Onat, Nuri Cihat, Gokhan Egilmez, and Omer Tatari. "Towards Greening the U.S. Residential Building Stock: A System Dynamics Approach." *Building and Environment* 78 (August 1, 2014): 68–80. https://doi.org/10.1016/j.buildenv.2014.03.030.

Owens, Ann. "Building Inequality: Housing Segregation and Income Segregation." *Sociological Science* 6 (2019): 497–525. https://doi.org/10.15195/v6.a19.

Palmer, Karen L., Margaret Walls, and Lucy O'Keeffe. "Putting Information into Action: What Explains Follow-Up on Home Energy Audits?" *SSRN Electronic Journal*, 2015. https://doi.org/10.2139/ssrn.2630120.

Palmer, Karen, and Margaret Walls. "Limited Attention and the Residential Energy Efficiency Gap." *American Economic Review* 105, no. 5 (May 2015): 192–95. https://doi.org/10.1257/aer.p20151009.

Palmer, Karen, Margaret Walls, Hal Gordon, and Todd Gerarden. “Assessing the Energy-Efficiency Information Gap: Results from a Survey of Home Energy Auditors.” *Energy Efficiency* 6, no. 2 (May 1, 2013): 271–92. https://doi.org/10.1007/s12053-012-9178-2.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2019. https://www.R-project.org/.

Riley, Patrick. “Three Pitfalls to Avoid in Machine Learning.” *Nature* 572, no. 7767 (August 2019): 27–29. https://doi.org/10.1038/d41586-019-02307-y.

Scheer, Adam M, Sam Borgeson, and Kali Rosendo. “Customer Targeting for Residential Energy Efficiency Programs: Enhancing Electricity Savings at the Meter,” n.d., 41.

Shaban, Hassan, Haider Khan, and David Meisegeier. “Mining Energy Efficiency Program Data,” n.d., 11.

Swan, Lukas G., and V. Ismet Ugursal. “Modeling of End-Use Energy Consumption in the Residential Sector: A Review of Modeling Techniques.” *Renewable and Sustainable Energy Reviews* 13, no. 8 (October 1, 2009): 1819–35. https://doi.org/10.1016/j.rser.2008.09.033.

Tian, Chenlu, Chengdong Li, Guiqing Zhang, and Yisheng Lv. “Data Driven Parallel Prediction of Building Energy Consumption Using Generative Adversarial Nets.” *Energy and Buildings* 186 (March 1, 2019): 230–43. https://doi.org/10.1016/j.enbuild.2019.01.034.

Tsanas, Athanasios, and Angeliki Xifara. “Accurate Quantitative Estimation of Energy Performance of Residential Buildings Using Statistical Machine Learning Tools.” *Energy and Buildings* 49 (June 1, 2012): 560–67. https://doi.org/10.1016/j.enbuild.2012.03.003.

Venables, W. N., and B. D. Ripley. *Modern Applied Statistics with S*. Fourth. New York: Springer, 2002. http://www.stats.ox.ac.uk/pub/MASS4.

Walker, J A, T N Rauh, and K Griffin. “A Review of the Residential Conservation Service Program,” n.d., 31.

Wilson, C., L. Crane, and G. Chryssochoidis. “Why Do Homeowners Renovate Energy Efficiently? Contrasting Perspectives and Implications for Policy.” *Energy Research & Social Science* 7 (May 1, 2015): 12–22. https://doi.org/10.1016/j.erss.2015.03.002.

Wilson, C., H. Pettifor, and G. Chryssochoidis. “Quantitative Modelling of Why and How Homeowners Decide to Renovate Energy Efficiently.” *Applied Energy* 212 (February 15, 2018): 1333–44. https://doi.org/10.1016/j.apenergy.2017.11.099.

Wing, Max Kuhn Contributions from Jed, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, et al. *Caret: Classification and Regression Training*, 2019. https://CRAN.R-project.org/package=caret.

Zhong, Hai, Jiajun Wang, Hongjie Jia, Yunfei Mu, and Shilei Lv. “Vector Field-Based Support Vector Regression for Building Energy Consumption Prediction.” *Applied Energy* 242 (May 15, 2019): 403–14. https://doi.org/10.1016/j.apenergy.2019.03.078.