Energy-Efficient Home Analysis in NYS

Abdellah Ait Elmouden

DATA698 | Analytics Master's Research Project

CUNY School of Professional Studies

Research Question

- Do energy-efficient homes save energy?
- What are the primary factors for adoption of an energy-efficient program for energy and cost saving?

The Dataset

- The model is based solely on publicly available data and comprises information available from one principle dataset: data from the New York State Energy Research and Development Authority (NY SERDA), hosted by the State of New York.

Data source : https://data.ny.gov/Energy-Environment/Residential-Existing-Homes-One-to-Four-Units-Energ/assk-vu73

Updated November 2, 2021

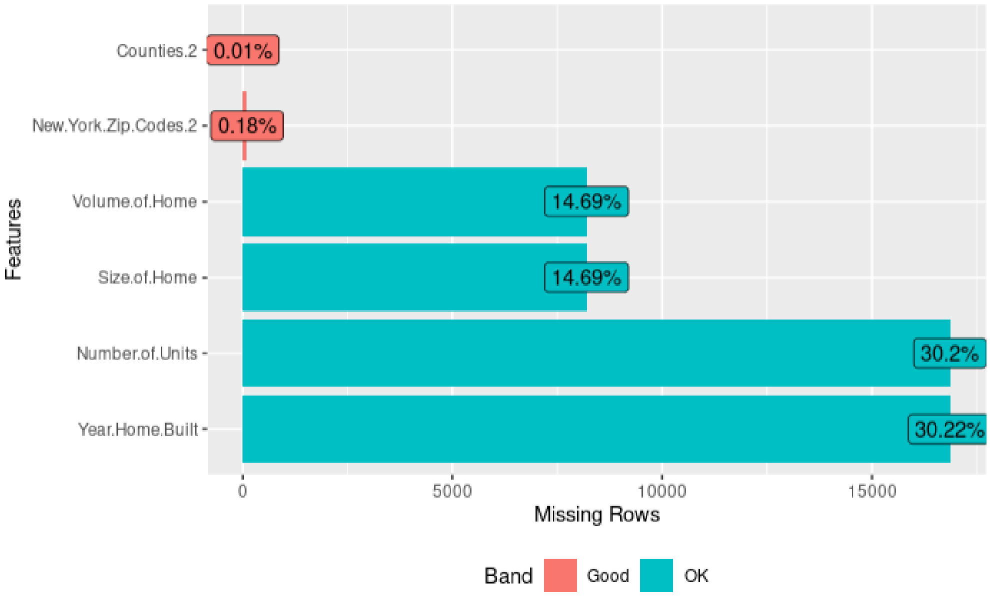
The Dataset

```
Rows: 55,830
Columns: 29
$ Reporting.Period                                                <chr> "2020-02-12T00:00:00.000", "20…
$ Home.Performance.Project.ID                                     <chr> "483911", "486389", "483374", …
$ Home.Performance.Site.ID                                        <chr> "357221", "343438", "356880", …
$ Project.County                                                  <chr> "Oneida", "Wayne", "Erie", "On…
$ Project.City                                                    <chr> "Barneveld", "Palmyra", "Buffa…
$ Project.Zip                                                     <int> 13304, 14522, 14215, 13440, 14…
$ Gas.Utility                                                     <chr> "", "", "National Fuel Gas Dis…
$ Electric.Utility                                                <chr> "National Grid", "New York Sta…
$ Project.Completion.Date                                         <chr> "2021-01-16T00:00:00.000", "20…
$ Customer.Type                                                   <chr> "Assisted", "Assisted", "Assis…
$ Low.Rise.or.Home.Performance.Indicator                          <chr> "Home Performance", "Home Perf…
$ Total.Project.Cost                                              <int> 5900, 9980, 5000, 4200, 5890, …
$ Total.Incentives                                                <int> 2950, 3250, 2500, 2088, 2945, …
$ Type.of.Program.Financing                                       <chr> "", "", "", "", "", "", "", ""…
$ Amount.Financed.Through.Program                                 <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
$ Pre.Retrofit.Home.Heating.Fuel.Type                             <chr> "", "Propane", "Natural Gas", …
$ Year.Home.Built                                                 <int> 1947, 1966, 1928, 1971, 1984, …
$ Size.of.Home                                                    <int> 2398, 1092, 1616, 980, 940, 24…
$ Volume.of.Home                                                  <int> 19184, 8517, 20516, 7840, 7520…
$ Number.of.Units                                                 <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …
$ Measure.Type                                                    <chr> "Building Shell", "Building Sh…
$ Estimated.Annual.kWh.Savings                                    <int> 0, 309, 0, 0, 0, -30, 20, 0, -…
$ Estimated.Annual.MMBtu.Savings                                  <int> 27, 3, 8, 28, 6, 13, 34, 3, 14…
$ First.Year.Energy.Savings...Estimate                            <int> 474, 111, 61, 681, 146, 15, 28…
$ Homeowner.Received.Green.Jobs.Green.NY.Free.Reduced.Cost.Audit..Y.N. <chr> "Y", "Y", "Y", "Y", "Y", "Y", …
$ New.Georeferenced.Column                                        <chr> "{'type': 'Point', 'coordinate…
$ New.York.Zip.Codes.2                                            <int> 764, 364, 71, 1366, 312, 239, …
$ Counties.2                                                      <int> 625, 631, 2041, 625, 2091, 213…
$ NYS.Municipal.Boundaries.2                                      <int> 985, 631, 300, 989, 742, 811, …
```

Rows: 55,830

Columns : 29

Missing Data



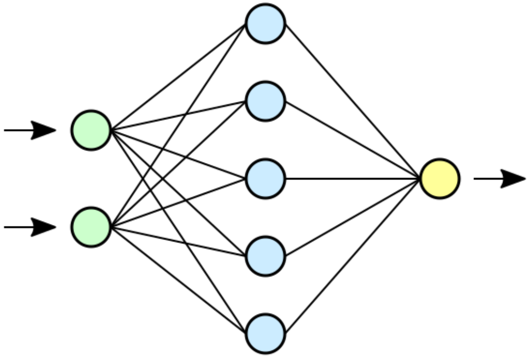
Methodology

Data Preparation:

- Data Cleaning
- Data Imputation
- Variable Classification (coding)

Data-driven prediction techniques:

- ANOVA Analysis
- Multiple linear regression
- Multilayer neural network



Methodology | Variable Categories

Gas Utility

1: NA
2: "Central Hudson Gas & Electric"
3: "Consolidated Edison"
4: "Corning Natural Gas"
5: "KeySpan Energy"
6: "Long Island Power Authority"
7: "Multiple Gas Providers"
8: "Municipal"
9: "National Fuel Gas Distribution"
10: "National Grid"
11: "New York State Electric & Gas"
12: "No Gas Provider"
13: "Orange & Rockland"
14: "Rochester Gas & Electric"
15: "Saint Lawrence Gas"

Electric Utility

1: "Central Hudson Gas & Electric"
2: "Consolidated Edison"
3: "Long Island Power Authority"
4: "Municipal (Not Qualified)"
5: "National Grid"
6: "New York State Electric & Gas"
7: "Orange & Rockland"
8: "Rochester Gas & Electric"

Fuel Type

- 1: NA
- 2: "Anthracite Coal"
- 3: "Bituminous Coal"
- 4: "Coal"
- 6: "Electricity"
- 7: "Kerosene"
- 9: "Natural Gas"
- 10: "Oil"
- 11: "Propane"
- 12: "Wood"
- 13: "Wood Pellets"

Customer Type

- 1: "Assisted"
- 2: "Market"

Measure Type

- 1: "Building Shell"
- 2: "Heating and Cooling"
- 3: "Water Heater"

HRGJ

- 2: "Y"
- 1: "N"

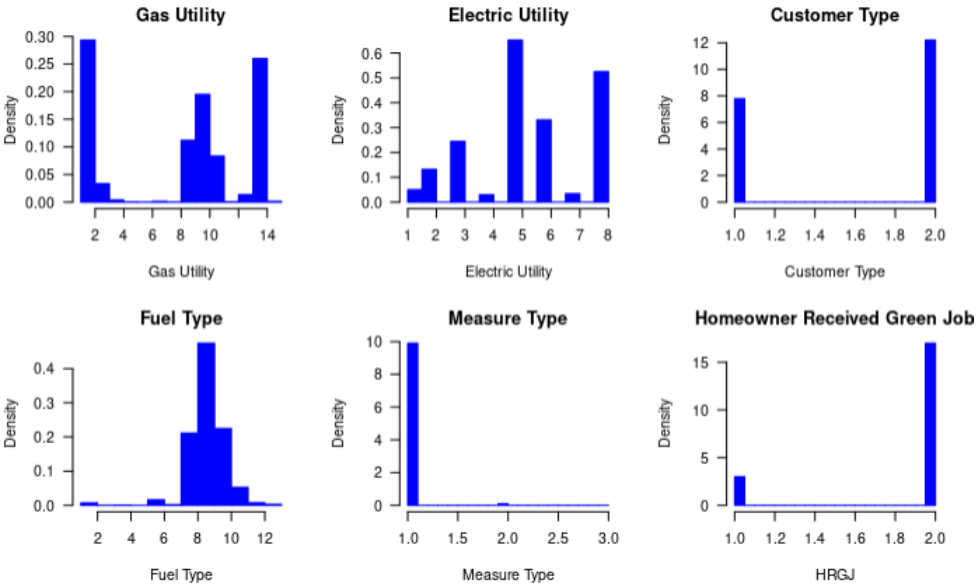
Descriptive statistics

Table 1: Variables Summary Statistics (N = 55830)

	Mean		Variance	Skewness		Kurtosis	
	Statistic	Std. Error	Statistic	Statistic	Std. Error	Statistic	Std. Error
Gasutility	8.16	0.022	26.602	-0.399	0.01	-1.433	0.021
Elecutility	5.43	0.008	3.915	-0.273	0.01	-0.732	0.021
customertype	1.61	0.002	0.238	-0.453	0.01	-1.795	0.021
totalprojectcost	8572.8	23.628	31169565.735	2.063	0.01	8.209	0.021
totalincentive	1687.31	7.219	2909413.163	0.961	0.01	0.262	0.021
fueltype	9.04	0.005	1.393	-2.159	0.01	14.648	0.021
yearhomebuilt	1945.78	0.16	1436.601	-0.932	0.01	1.004	0.021
sizeofhome	1909.51	2.889	466074.787	0.772	0.01	0.12	0.021
masuretype	1.01	0	0.01	10.382	0.01	110.132	0.021
FYES	596.21	2.893	467243.075	2.395	0.01	7.425	0.021
HRGJ	1.85	0.002	0.127	-1.962	0.01	1.85	0.021

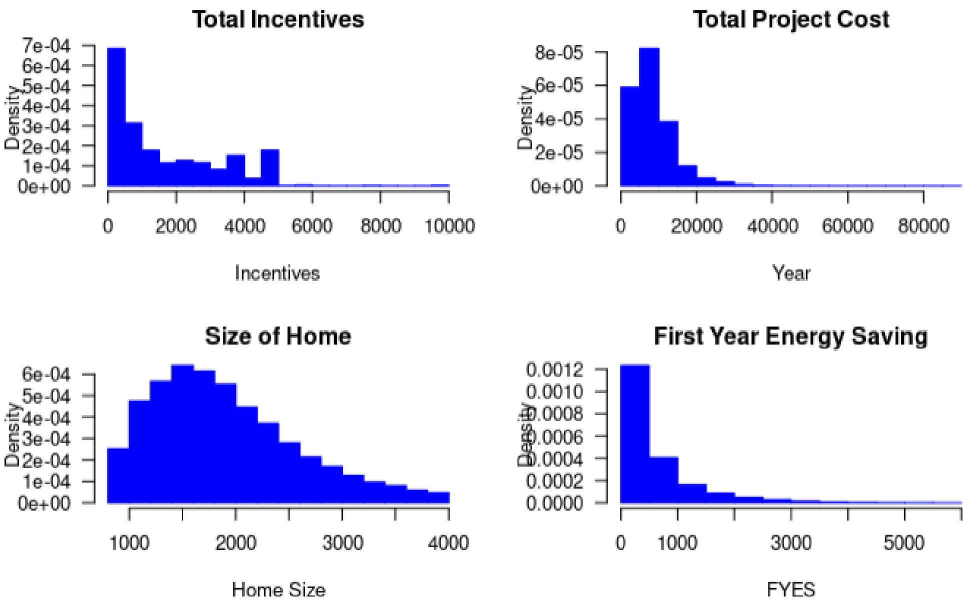
Descriptive statistics

Categorical Variables

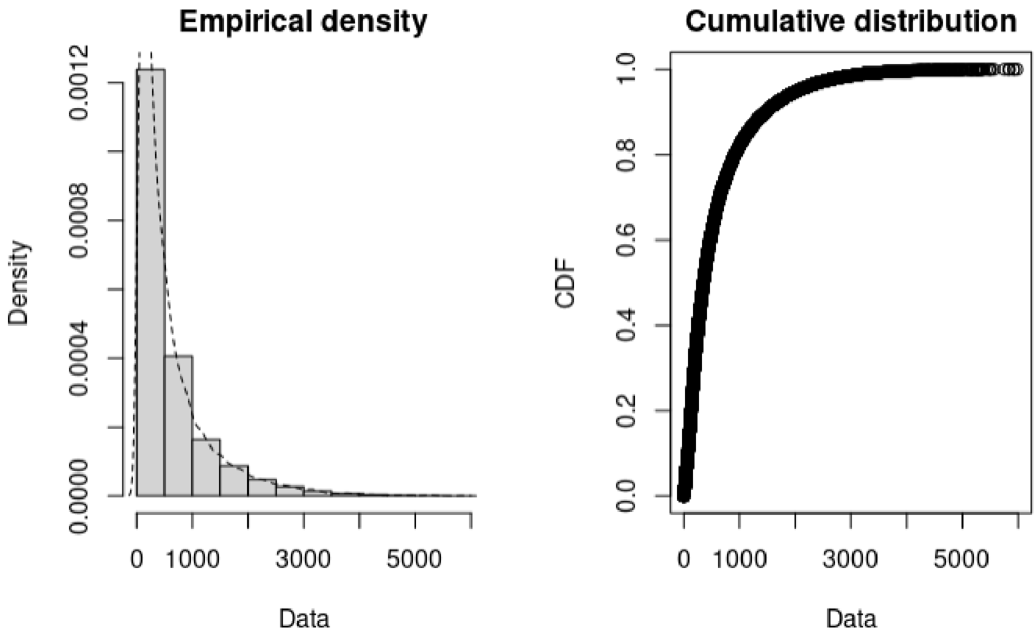


Descriptive statistics

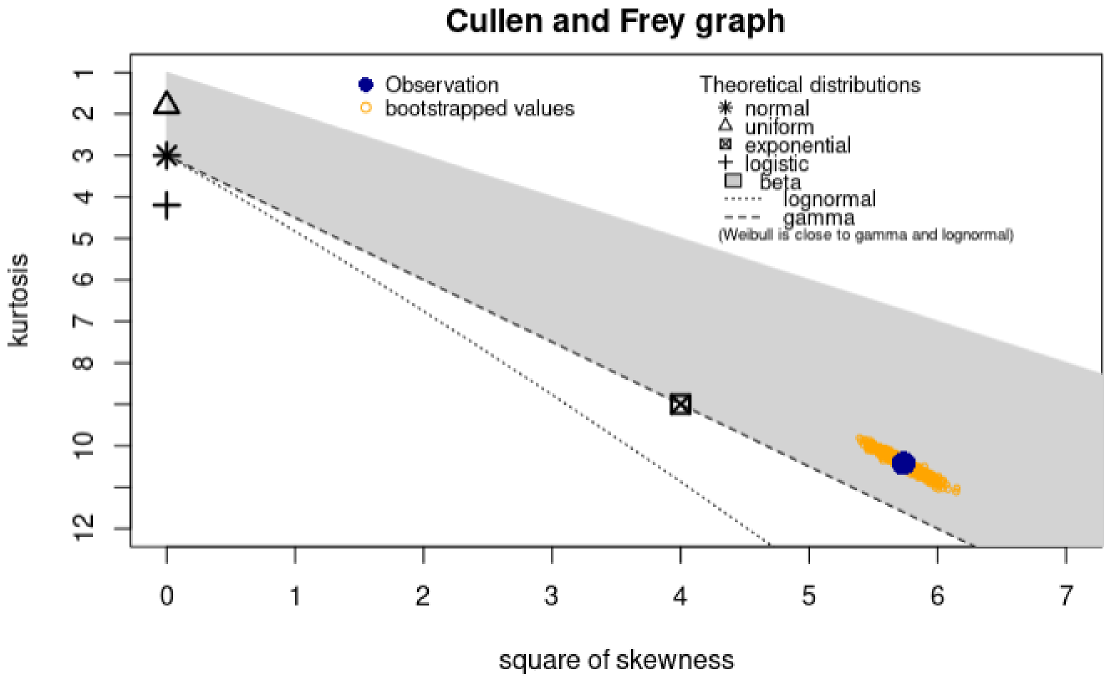
Numerical Variables



Descriptive statistics



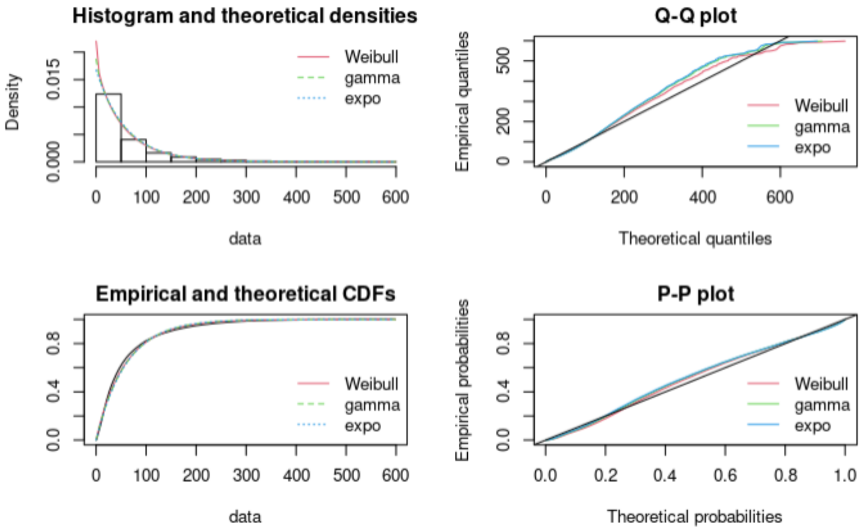
Descriptive statistics



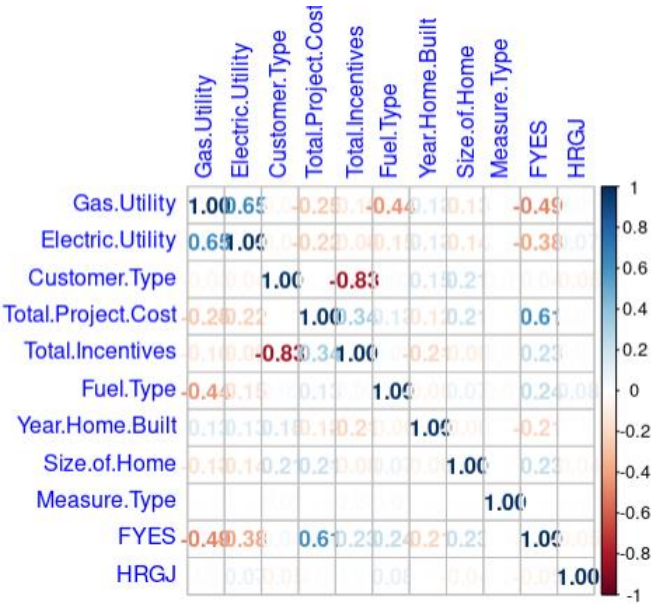
Correlation

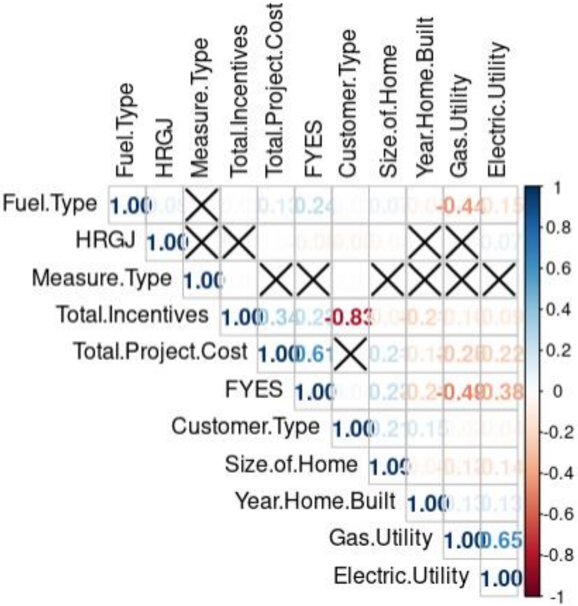
Table 2: Goodness-of-fit-statistics

Goodness-of-fit-statistics			
	1-mle-weibull	2-mle-gamma	3-mle-exp
Kolmogorov-Smirnov-statistic	0.042	0.051	0.055
Cramer-von-Mises-statistic	36.676	53.192	60.093
Anderson-Darling-statistic	224.160	281.895	307.472
Goodness-of-fit-criteria			
	1-mle-weibull	2-mle-gamma	3-mle-exp
Akaike's-Information-Criterion	567029.00	567244.30	567262.60
Bayesian-Information-Criterion	567046.90	567262.10	567271.50



Model development | Correlation





Model development | ANOVA Analysis

```
                    Df    Sum Sq   Mean Sq  F value    Pr(>F)    
Electric.Utility      1 3.778e+09 3.778e+09 16375.45  < 2e-16 ***
Gas.Utility           1 2.675e+09 2.675e+09 11595.86  < 2e-16 ***
Total.Project.Cost    1 6.609e+09 6.609e+09 28644.67  < 2e-16 ***
Total.Incentives      1 8.369e+06 8.369e+06    36.28 1.72e-09 ***
Size.of.Home          1 1.356e+08 1.356e+08   587.62  < 2e-16 ***
Residuals         55824 1.288e+10 2.307e+05                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Model development | Multiple Regression

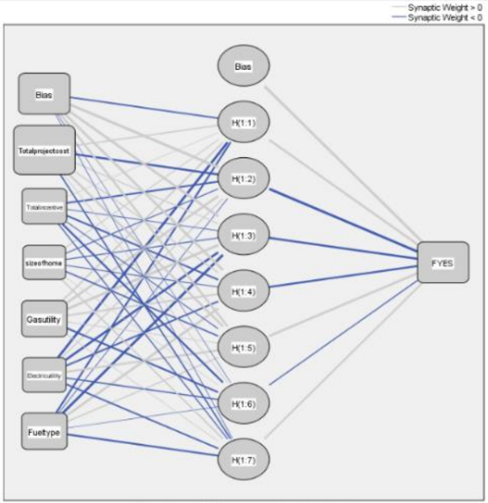
```
Call:
lm(formula = eng_imputed$FYES ~ eng_imputed$Size.of.Home + eng_imputed$Total.Incentives +
    Gas.Utility, data = eng_imputed, subset = Electric.Utility +
    Fuel.Type, weights = Total.Project.Cost)

Weighted Residuals:
    Min      1Q  Median      3Q     Max
-105513  -51072   -3101   17456   84658

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -1.309e+03  1.229e+01  -106.5   <2e-16 ***
eng_imputed$Size.of.Home     9.723e-01  4.655e-03   208.9   <2e-16 ***
eng_imputed$Total.Incentives -2.826e-02  2.125e-03   -13.3   <2e-16 ***
Gas.Utility                  2.842e+01  1.128e+00    25.2   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 40940 on 55826 degrees of freedom
Multiple R-squared:  0.6168,    Adjusted R-squared:  0.6168
F-statistic: 2.996e+04 on 3 and 55826 DF,  p-value: < 2.2e-16
```

Model development | Multilayer Neural Network



Hidden layer activation function: Hyperbolic tangent

Output layer activation function: Identity

Parameter Estimates

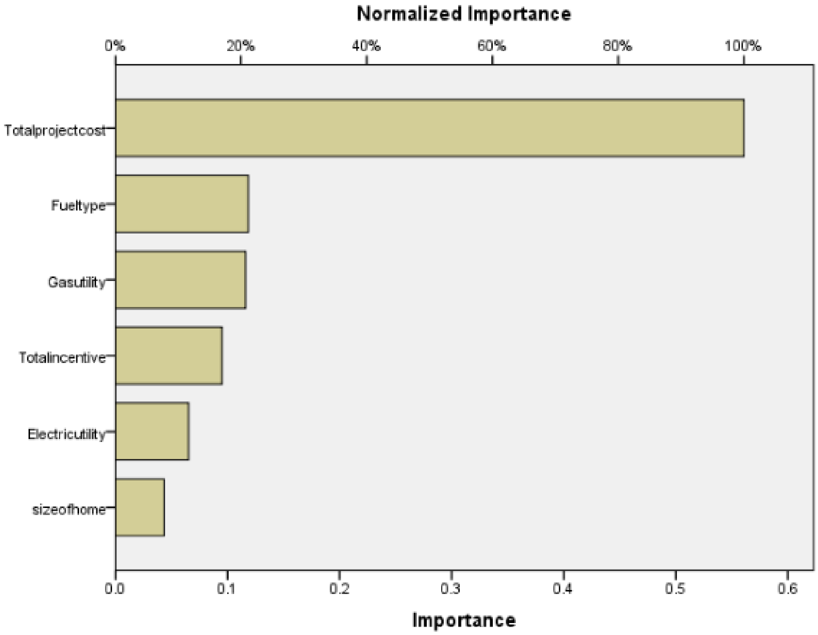
Predictor		Predicted							
		Hidden Layer 1							Output Layer
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	FYES
Input Layer	(Bias)	-.260	3.200	1.799	1.049	.566	-.005	-.049	
	Totalprojectcost	.254	-.542	.193	.208	.714	.109	-.235	
	Totalincentive	.068	-.321	-.074	-.319	-.149	-.392	-.168	
	sizeofhome	.166	-.149	-.162	-.143	-.228	.030	.281	
	Gasutility	.298	1.019	2.070	.321	.112	-.331	.072	
	Electricutility	-1.301	.199	-.824	-.313	.534	-.209	-.286	
	Fueltype	-.337	-.073	-1.323	.965	.220	-.009	-.346	
Hidden Layer 1	(Bias)								2.239
	H(1:1)								.643
	H(1:2)								-2.000
	H(1:3)								-.607
	H(1:4)								-.405
	H(1:5)								.821
	H(1:6)								-.189
	H(1:7)								.347

Figure 10: Neural Network Parameter Estimates

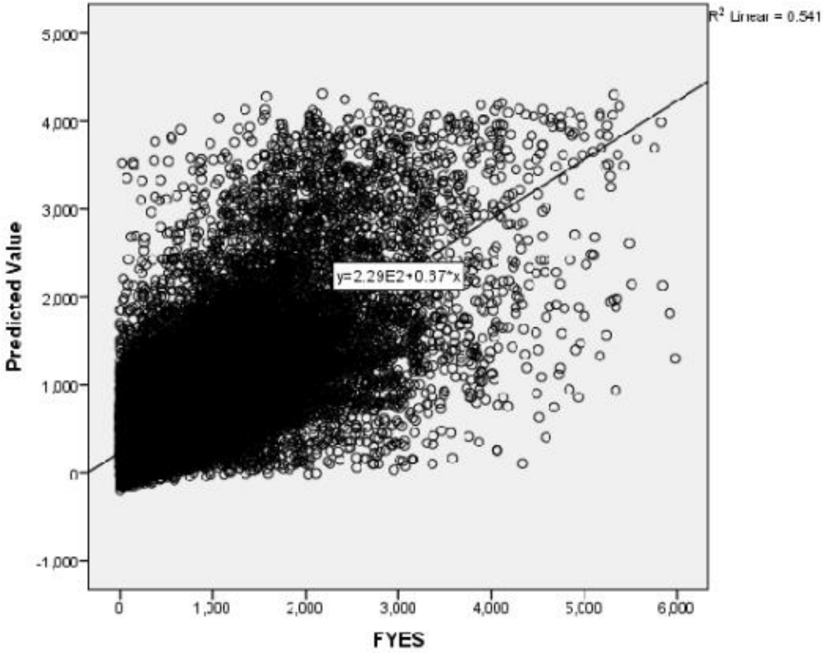
Model development | Multilayer Neural Network

Table 4: Independent Variable Importance

	Importance	Normalized Importance
Total Proje Cost	0.561	100.00%
Total Incentive	0.095	17.00%
Size of Home	0.044	7.80%
Gas Utility	0.116	20.70%
Electric Utility	0.065	11.70%
Fuel Type	0.119	21.10%



Model development | Multilayer Neural Network



Model development | Multilayer Neural Network

$$V_{0bias(7x1)} = \begin{bmatrix} -.26 \\ 3.2 \\ 1.8 \\ 1 \\ 0.57 \\ -0.005 \\ -0.049 \end{bmatrix} \text{ and } W_{bias(1x1)} = \begin{bmatrix} 2.24 \end{bmatrix} \text{ and } W_{6x1} = \begin{bmatrix} .643 \\ -2 \\ -.607 \\ -.405 \\ .821 \\ -.189 \end{bmatrix}$$

$$V_{(6x7)} = \begin{bmatrix} .254 & -.542 & .193 & .208 & .714 & .109 & -.235 \\ .068 & -.321 & -.074 & -.319 & -.149 & -.392 & -.168 \\ .166 & -.149 & -.162 & -.143 & -.228 & .030 & .281 \\ .298 & 1.019 & 2.070 & .321 & .112 & -.331 & .072 \\ -1.3 & 0.2 & -.824 & -.313 & .534 & -.209 & -.286 \\ -.337 & -.073 & -1.323 & .965 & .220 & -.009 & -.346 \end{bmatrix}$$

$$FYES = [W_{bias(1x1)} + (W_{6x1})^T + \tanh[V_{0bias(7x1)} + (V_{(6x7)})^T * W_{6x1})] * Std_{FYES} + \overline{X}_{FYES}]$$

Conclusion

This work indicates that MNN models have significant potential for use in predicting efficient-home Energy Saving. the model proposed that Energy efficiency is known to be a cost-effective investment when projects cost are higher, but the actual amount of savings differs based on the upgrade in question, also Fuel switching type, and Gas Utility. Although MNN can help to show trends in total energy saving patterns, they cannot explain the various components that contribute to energy saving at the household level. For this type of analysis it is necessary to conduct investigations at the micro-level.

Bottom-up methods take a disaggregated approach and estimate energy demand and emissions using high resolution data using a combination of physical, social, behavioural and demographic properties for a household.

References

[1] Wokje Abrahamse and Linda Steg. How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? *Journal of Economic Psychology*, 30(5):711–720, October 2009.

[2] Abdo Abdullah Ahmed Gassar, Geun Young Yun, and Sumin Kim. Data-driven approach to prediction of residential energy consumption at urban scales in London. *Energy*, 187:115973, November 2019.

[3] Fiona Burlig, Christopher Knittel, David Rapson, UC Davis, Mar Reguant, and Catherine Wolfram. Machine Learning from Schools about Energy Efficiency. page 45.

[4] Birgitta Gatersleben, Linda Steg, and Charles Vlek. Measurement and Determinants of Environmentally Significant Consumer Behavior. *Environment and Behavior*, 34(3):335–362, May 2002. Publisher: SAGE Publications Inc.

[5] Andrea L. Hicks and Thomas L. Theis. Residential energy-efficient lighting adoption survey. *Energy Efficiency*, 7(2):323–333, April 2014.

[6] Eunsil Lee, Nam-Kyu Park, and Ju Hyoung Han. Factors Affecting Environmentally Responsible Behaviors in the Use of Energy-efficient Lighting in the Home. *Family and Consumer Sciences Research Journal*, 41(4):413–425, June 2013. Publisher: John Wiley & Sons, Ltd.

[7] Joe O'Doherty, Sean Lyons, and Richard S.J. Tol. Energy-using appliances and energy-saving features: Determinants of ownership in Ireland. *Applied Energy*, 85(7):650–662, July 2008.

[8] Geoffrey K. F. Tso and Kelvin K. W. Yau. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32(9):1761–1768, September 2007.

[9] Sanna Tuomela, Mauricio de Castro Tomé, Netta Iivari, and Rauli Svento. Impacts of home energy management systems on electricity consumption. *Applied Energy*, 299:117310, October 2021.

Thank you!