BOSTON UNIVERSITY

Using Permit, Parcel Data to Investigate Residence Energy Systems

Zixin Ding, Jason Lu, Fang Qu

Collaboration with BU Spark! (Michael J. Walsh and Adam Pollack)

Abstract

Design and analyze a database of tax parcels and building permits for investigating the trends and distribution of heating and energy efficient systems.

Questions to be answered:

- Which communities are more likely to have heat pumps, based on various demographics?
- From the Boston area, which communities are benefitting or not benefitting from energy efficient systems?

Goals:

1. Make a predictive model to determine which building has a particular type of heating system
2. Visualize the distribution of heating systems

Data Sources

Boston Property Assessments: 2014-2018
Boston Building Permits:text descriptions about building renovations
Parcel Data Key by Boston Assessing Department: legend for determining the descriptions of the Boston Property Assessments attributes.

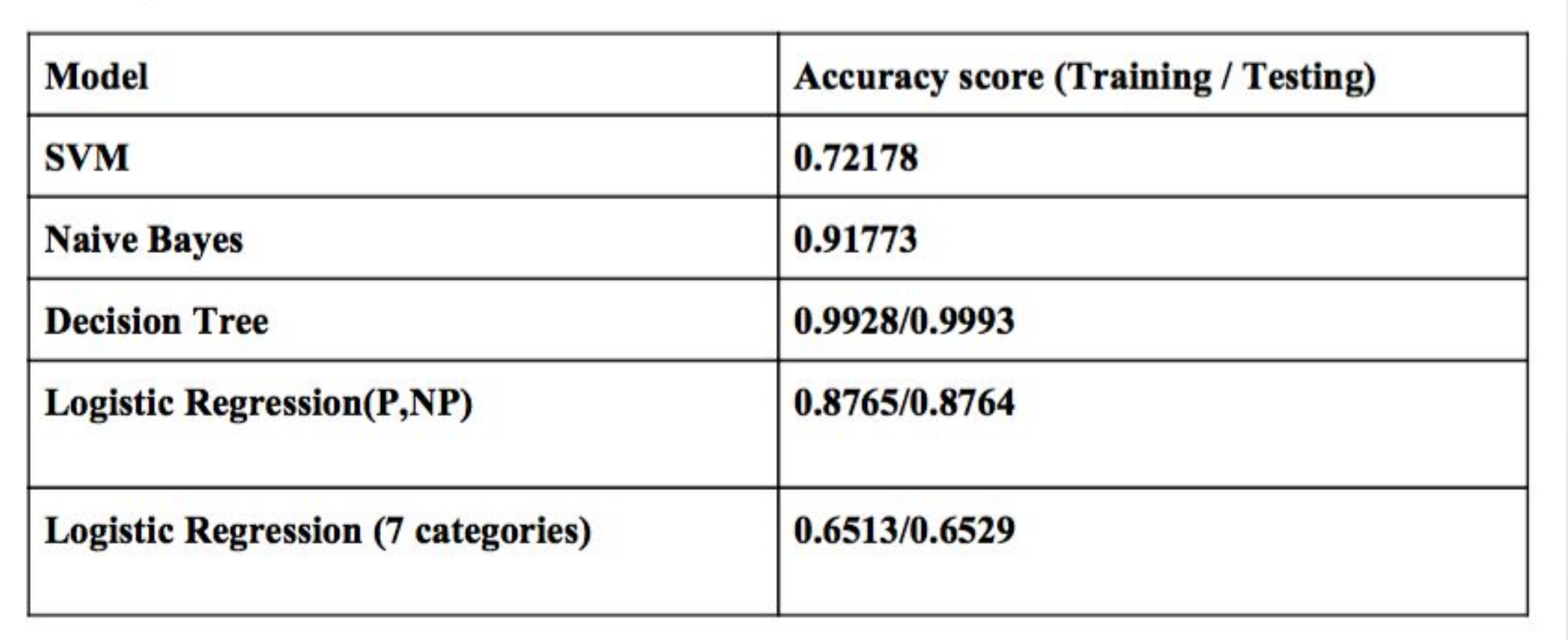
Methodology

- We applied four common statistical modeling methods (LR, GNB, SVM, DTC) to predict the heat pumps class (P).
- We figured out the rate of change of each heat type for both residential and condo buildings (shown in results table)
- We visualized heat systems distribution, average assessed value per square feet, owner occupancy type by years in several map figures
- We determined which type of communities benefit from energy efficient systems (solar panels) or from installation of heat pumps.

Regression/Classifier

Accuracy scores for the four common classifiers

- Logistic Regression (LR),
- Gaussian Naive Bayes (GNB),
- Support Vector Machines (SVM)
- Decision Tree Classifier (DTC).

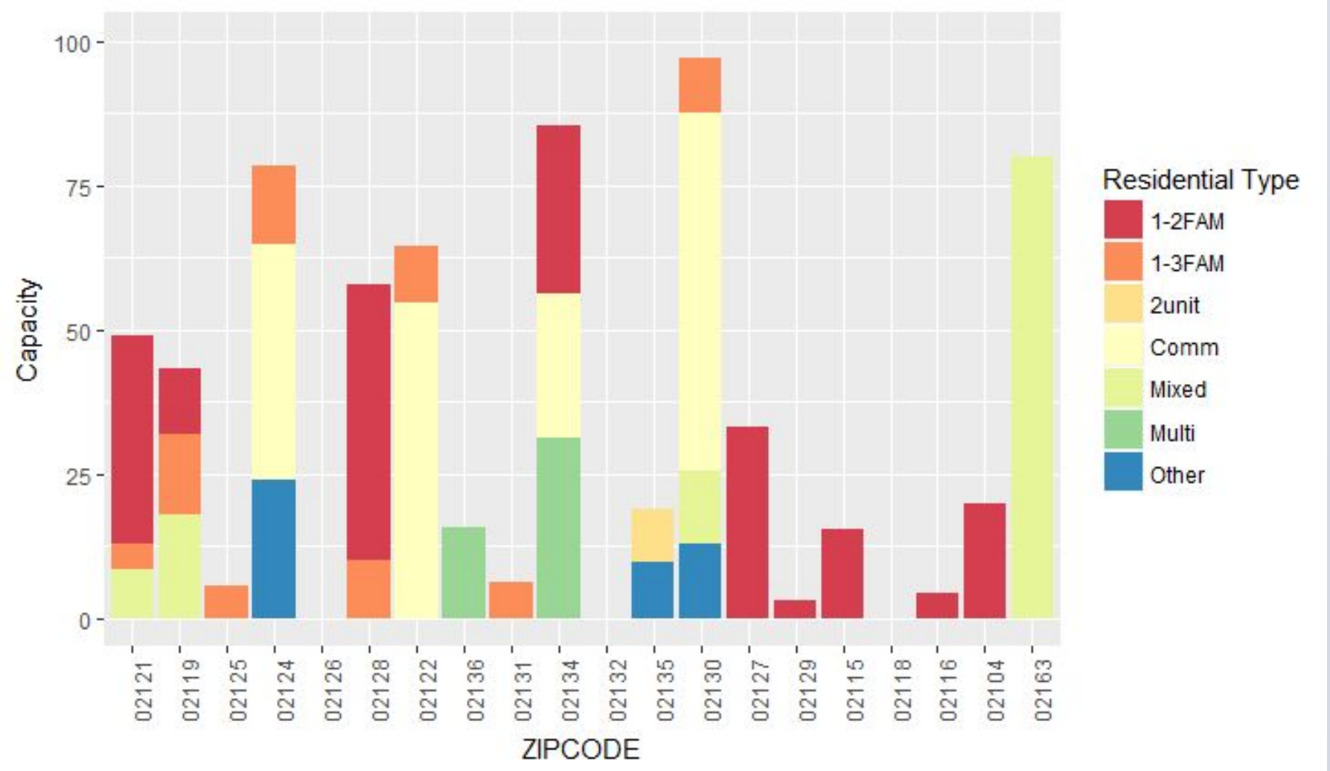


Model	**Accuracy score (Training / Testing)**
SVM	**0.72178**
Naive Bayes	**0.91773**
Decision Tree	**0.9928/0.9993**
Logistic Regression(P,NP)	**0.8765/0.8764**
Logistic Regression (7 categories)	**0.6513/0.6529**

Analysis:
Since the number of class P (heat pumps) categories is too small, we implemented:

- L1 regularization method to prevent overfitting
- "Balanced weighting" for each class so we can get better performance when predicting class P
- Delete non-statistically important features according to the p-value for each of the dependent variables

Solar Capacity Histogram



- Extract numbers before 'KW' (kilowatt) in comments column(Solar Capacity), Residential Type and Zip Code(stacked in ascending order of Average Assessed Value per sq ft)
- Solar Capacity is highly utilized in communities, mixed, multi or other residential type of buildings as compared to 1-2 or 1-3 family occupancies.
- The percentage of 1-2 or 1-3 family usage of solar capacity is decreasing with ascending order of average assessed value per sq ft.
- Possible selection bias:
  - 1. The dataset only measures the renovation required in 2018, but not all houses report solar capacity in comments column(only 638 out of 300,000 total parcel rows).
  - 2. Zip codes might not be a good measure as area within same zip code might have varied housing prices.

Visualization

Fig 1: Map Visualization for Heating Systems Except Heat Pumps

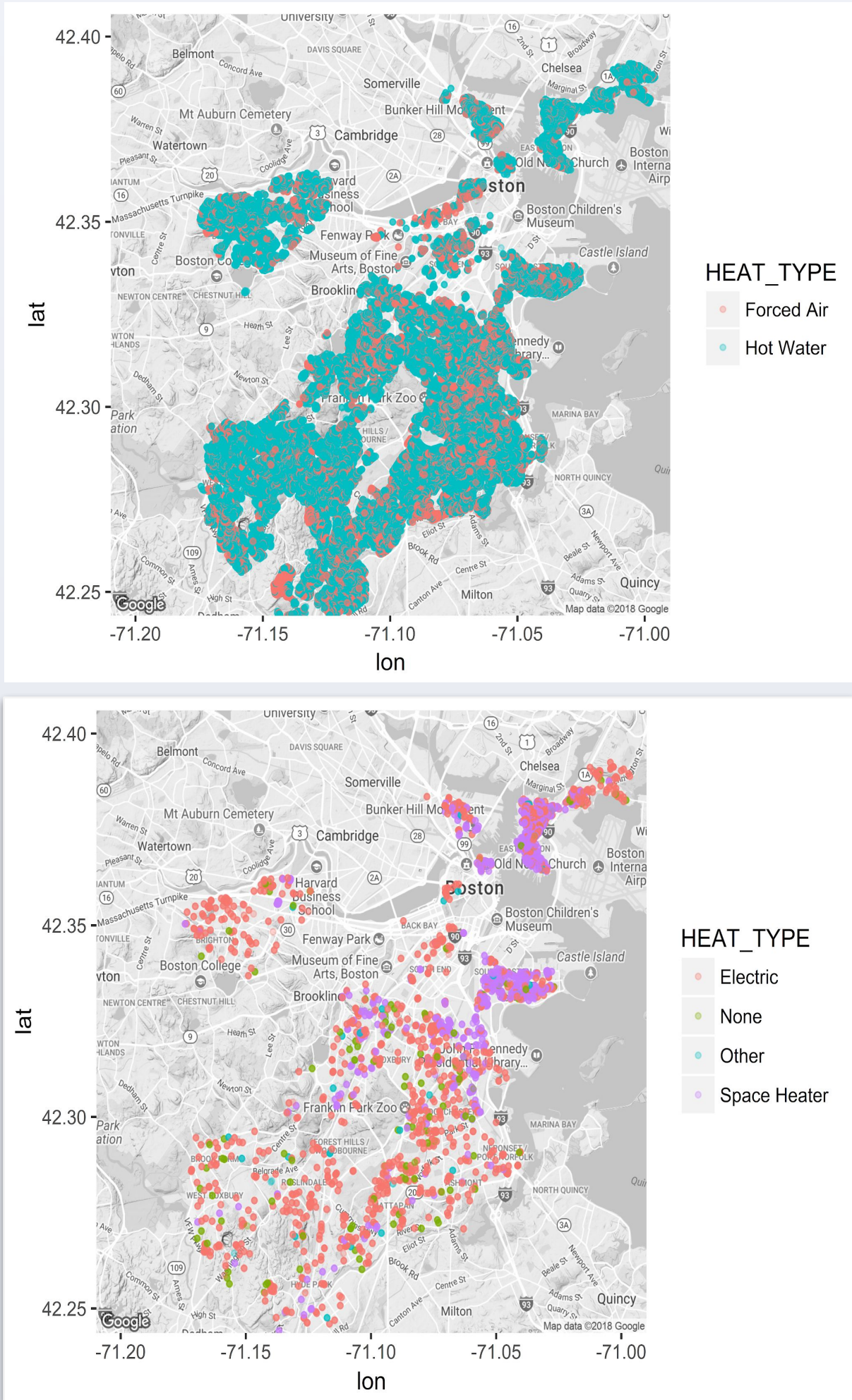


Fig 3: Facet Plot for Owner Occupancy (x: N/Y), Average Total Assessed Value per sqft(y) and Heating Systems

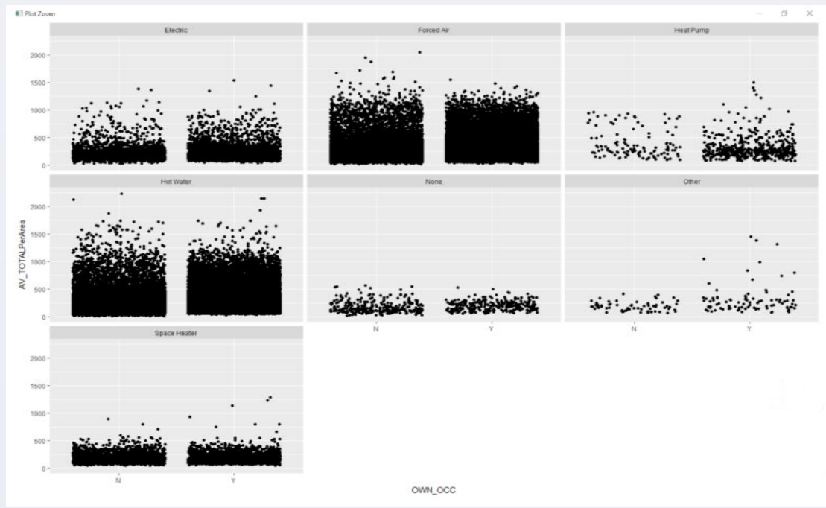


Fig 4: Residence distribution Grouped by Type of Heating Systems and Years Built in 2014 Parcel Data

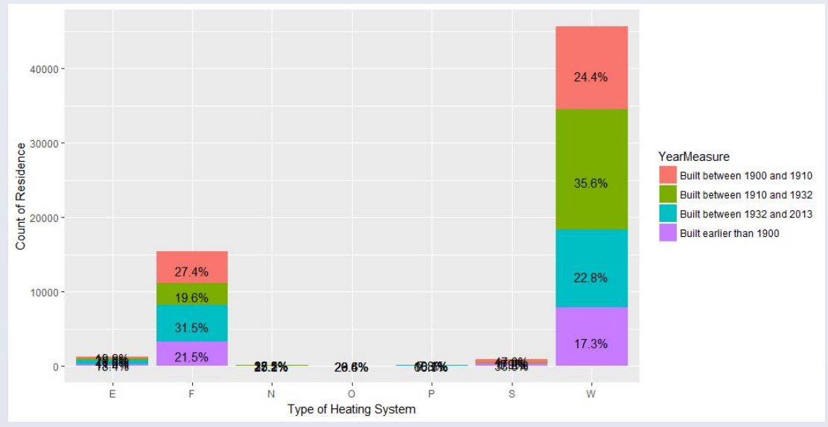


Fig 2: Map Visualization for Heat Pumps

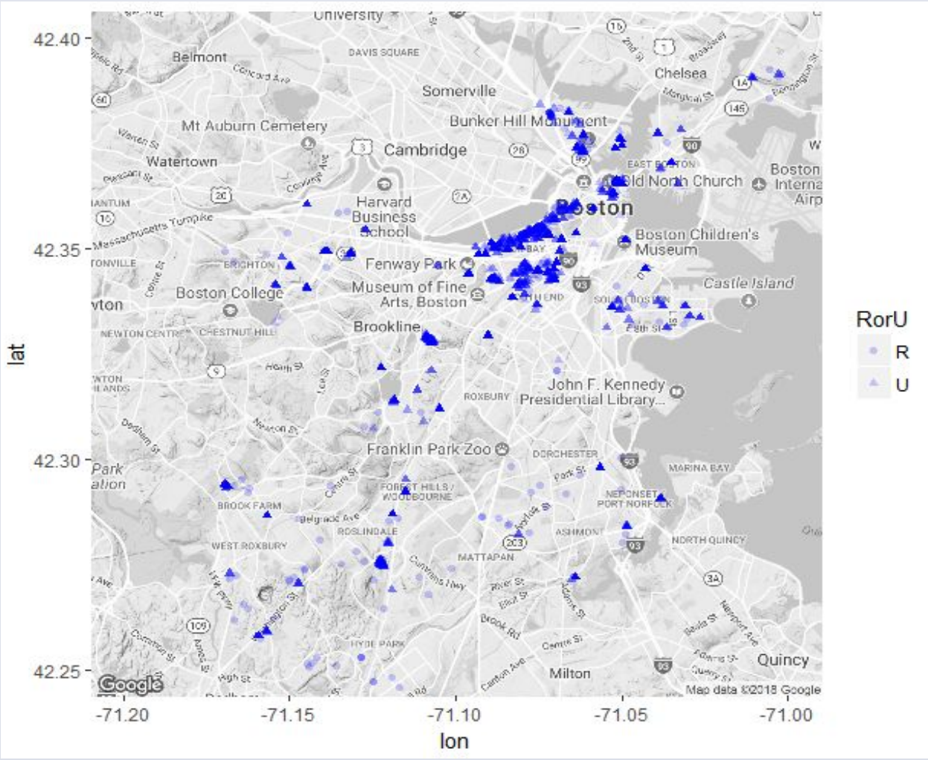
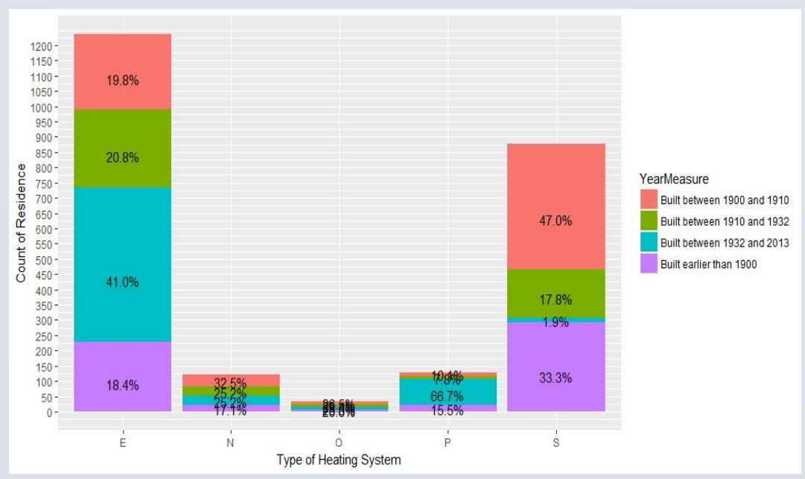


Fig 5: Residence Distribution by Type of Heating Systems and Years Built in 2014 Parcel Data Except Forced Air and Hot Water



Confusion Matrices

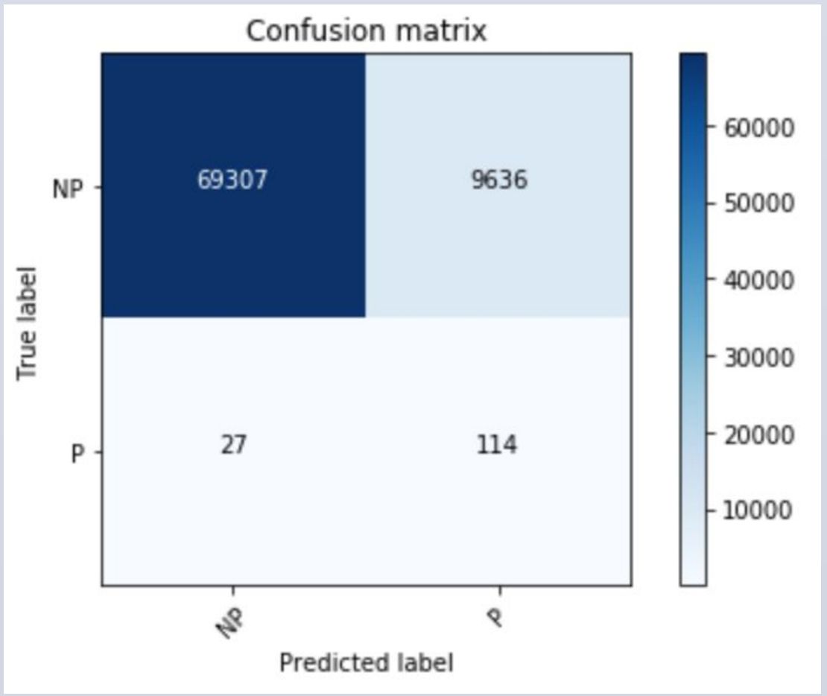


Fig. 1: Logistic Regression with L1 regularization and balanced class weights.

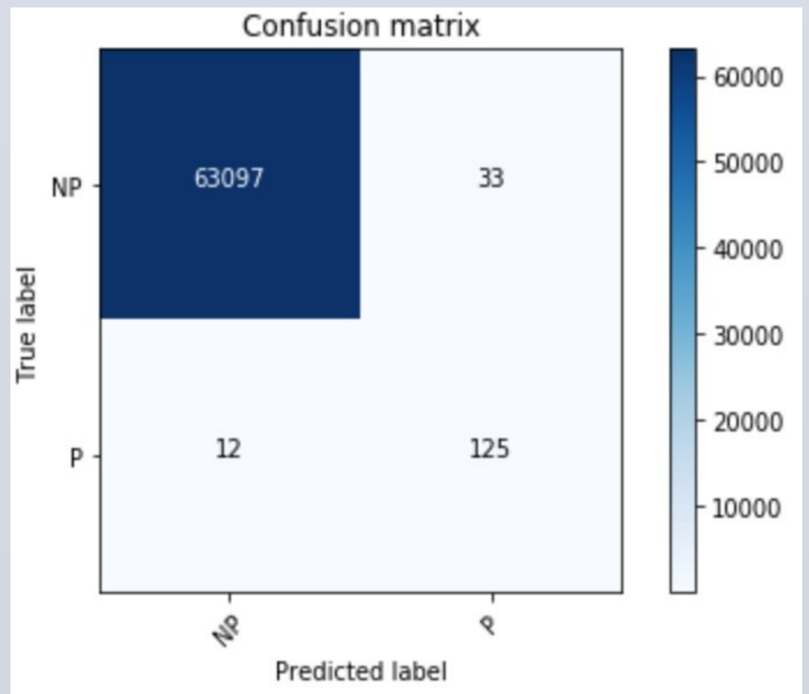
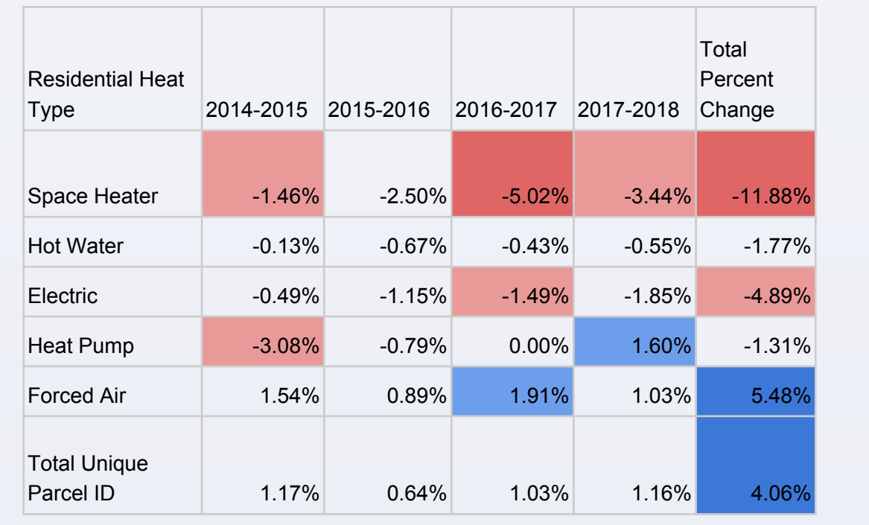


Fig. 2: Decision Tree classifier with balanced class weights

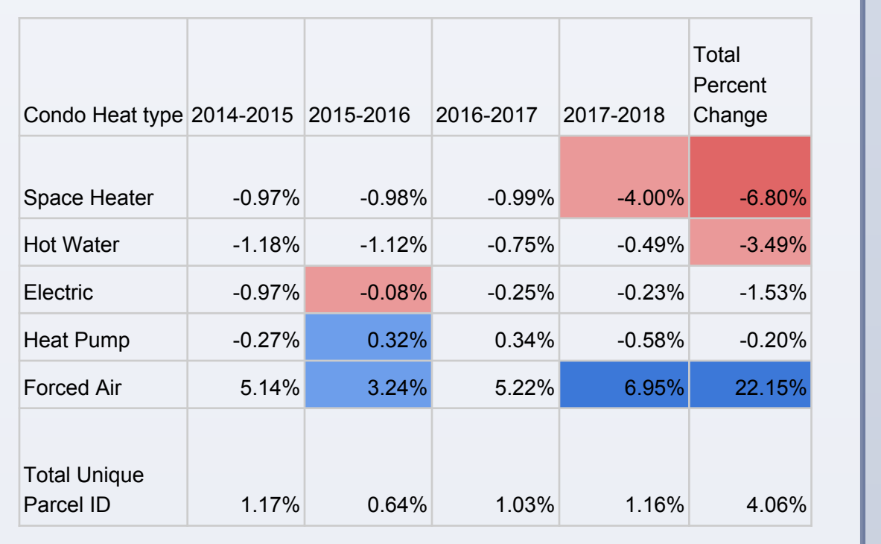
Heating Type Rate of Change

Residential Heating Type



Residential Heat Type	2014-2015	2015-2016	2016-2017	2017-2018	Total Percent Change
Space Heater	-1.46%	-2.50%	-5.02%	-3.44%	-11.88%
Hot Water	-0.13%	-0.67%	-0.43%	-0.55%	-1.77%
Electric	-0.49%	-1.15%	-1.49%	-1.85%	-4.89%
Heat Pump	-3.08%	-0.79%	0.00%	1.60%	-1.31%
Forced Air	1.54%	0.89%	1.91%	1.03%	5.48%
Total Unique Parcel ID	1.17%	0.64%	1.03%	1.16%	4.06%

Condominium Heating Type



Condo Heat type	2014-2015	2015-2016	2016-2017	2017-2018	Total Percent Change
Space Heater	-0.97%	-0.98%	-0.99%	-4.00%	-6.80%
Hot Water	-1.18%	-1.12%	-0.75%	-0.49%	-3.49%
Electric	-0.97%	-0.08%	-0.25%	-0.23%	-1.53%
Heat Pump	-0.27%	0.32%	0.34%	-0.58%	-0.20%
Forced Air	5.14%	3.24%	5.22%	6.95%	22.15%
Total Unique Parcel ID	1.17%	0.64%	1.03%	1.16%	4.06%

Color key: for significant values in the table (from 2014-2018)
Red: percentage decrease
Blue: percentage increase

Light red: 1 to 5% decrease
Light blue: 1 to 5% increase
Dark red: 5% + decrease
Dark blue: 5% + increase

Noticable changes:

- Significant decrease of space heater usage
- The condominium heating type has more percentage change in forced air
- Less people are using electric heating for residential
- Greater variation of change for residential housing

Conclusions

- We found that our models (Logistic Regression, Naive Gaussian Bayes, SVM, Decision Tree) were not enough to get information for our interested class P (heat pumps).
- Over the span of the past five fiscal years, we noticed that the usage rate of the heat pumps has been steady for condominium housing, but varies greatly for residential housing
- We can also calculate the solar capacity usage of each household with a stacked bar histogram, which shows both the averaged assessed value per zip code and land prices increasing.
- Homeowners can subscribe to "community solar gardens"
- Homeowners can generate solar electricity without having solar panels on their rooftops.

Future Work

- Put more descriptions (type of housing, parcel characteristics, etc.) for each individual permit
- Use the American Census Housing Survey to get a detailed inventory of each building parcel in Boston.
- Incorporate a larger time interval (defined as the past decade) for determining the solar panel capacity and the heat pump rate of change.

GitHub Repo

https://github.com/cxhugh/CS506_Project_Spark