The Concrete Jungle

MACHINE LEARNING PROJECT

AUTHOR: KYUNGHWAN WILLIAM KIM

BUSINESS PROBLEM

- Since 2009, the NYC government has carried out an ongoing effort to gather relevant data on how energy is being consumed by NYC buildings. An attempt to audit every single property would be unfeasible in terms of both cost and time

- Local Law 84 requires annual benchmarking data to be submitted by NYC building owners.

- The main focus will be using the LL84 dataset, build a model that can predict a building's ENERGY STAR Score, then interpret the results to find the features that are most related to the prediction of the target variable

THE DATA

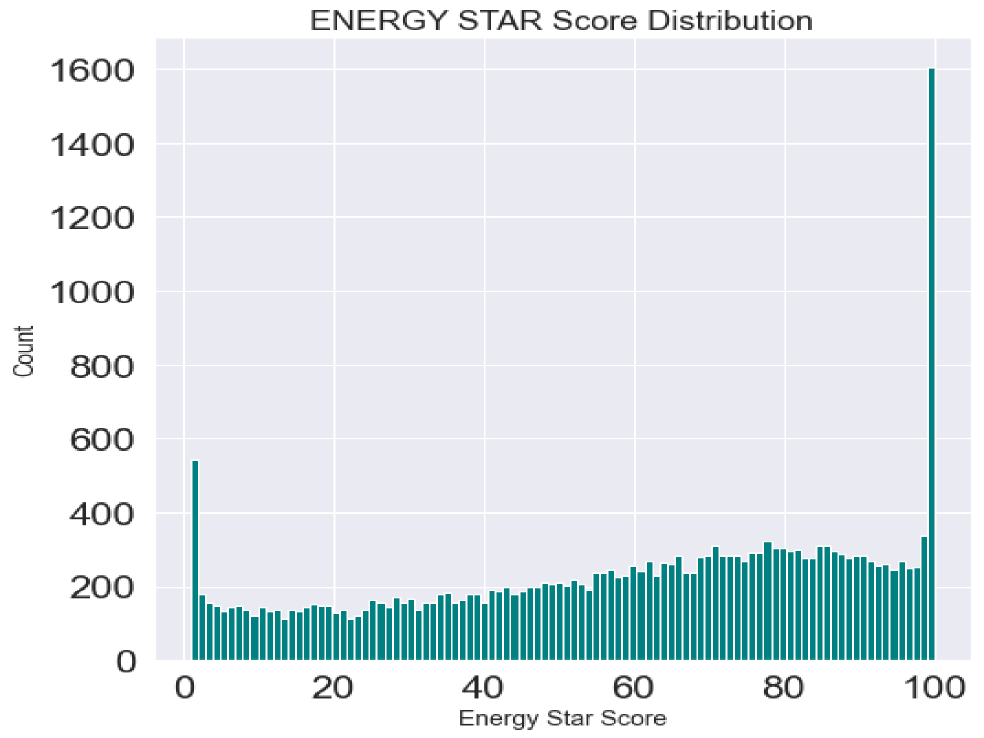
- The NYC Benchmarking Law requires owners of large buildings to annually measure their energy and water consumption in a process called benchmarking. The law standardizes this process by requiring building owners to enter their annual energy and water use in the U.S. Environmental Protection Agency's (EPA) online tool, ENERGY STAR Portfolio Manager and use the tool to submit data to the City.

- This data gives building owners information about a building's energy and water consumption compared to similar buildings, and tracks progress year over year to help in energy efficiency planning.

- The data comes from the year 2021, with 29,842 buildings and 249 various features if energy usage, emissions, and other information.

THE TARGET

- ENERGY STAR Score is the target variable, which is a measure of building performance relative to similar properties.
- Looking at the distribution, there are two nodes in the data, one at 1 and the other at 100.
- ENERGY STAR Score might not be as objective.

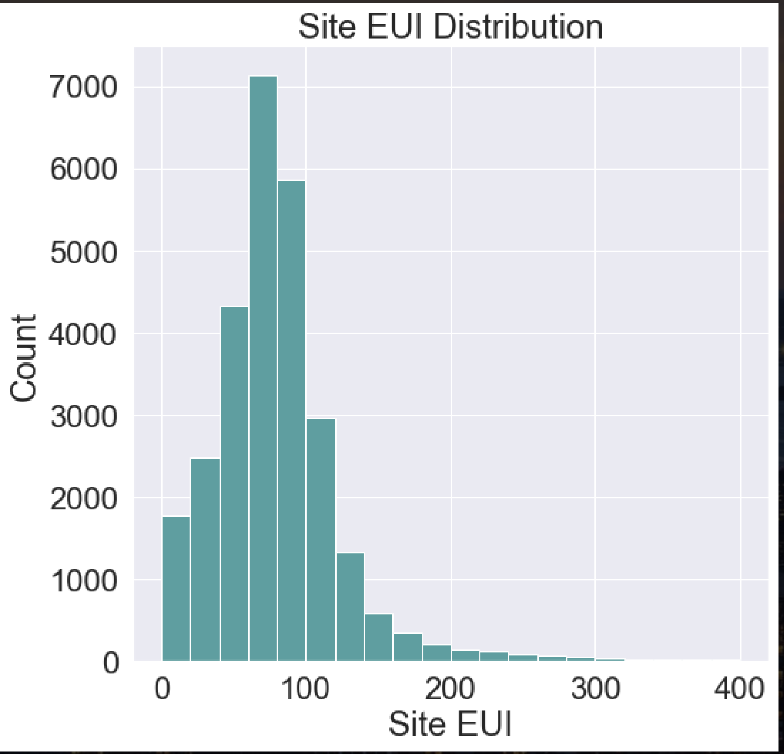


EUI (ENERGY USAGE INTENSITY)

- In contrast to the Energy Star Score, the Energy Use Intensity (EUI) is based on actual energy use as determined by the utility.

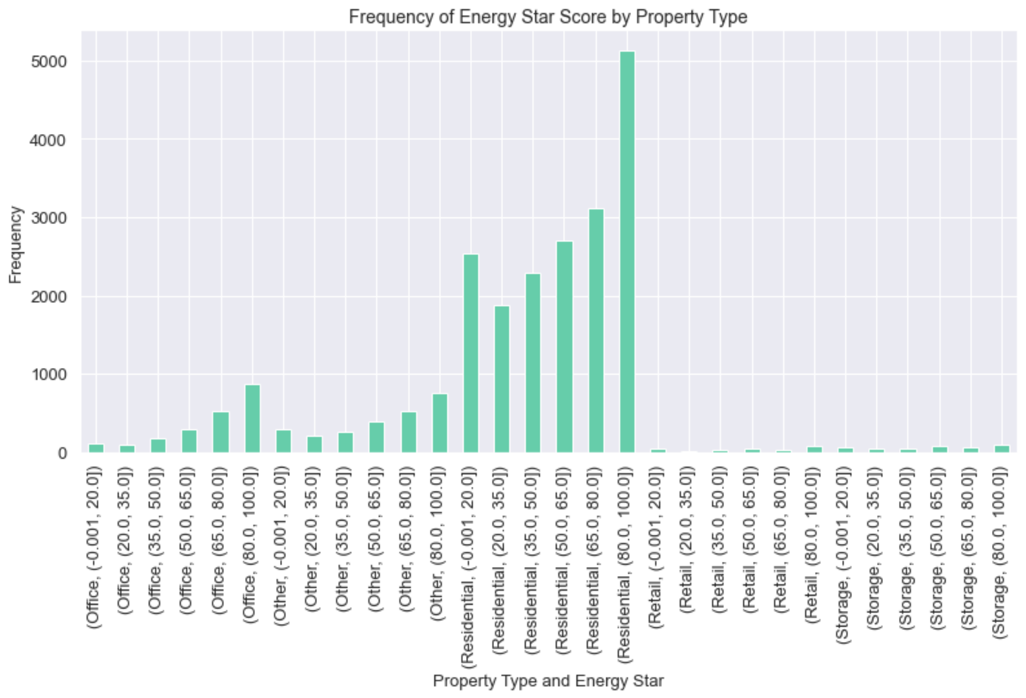
- Plot is more normally distributed and is a better measure of building performance.

- Buildings with Energy Star Scores of 100 have lower Site EUI than buildings with Energy Star Scores of 1.

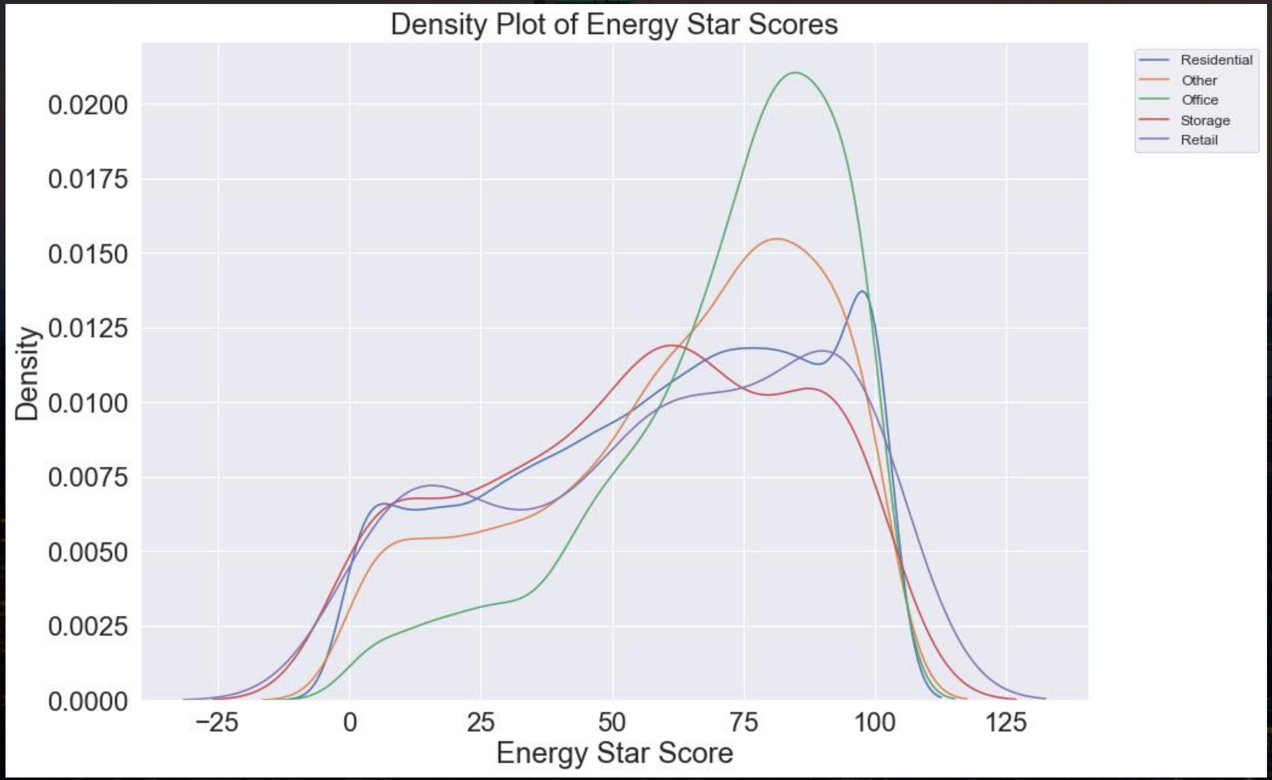


PROPERY TYPE

- There were 74 different types of property types.
- Most common property types in NYC are Multifamily Housing, then Office.
- Categorized the 74 property types into 5 building types:
  - Residential
  - Office
  - Retail
  - Storage
  - Other



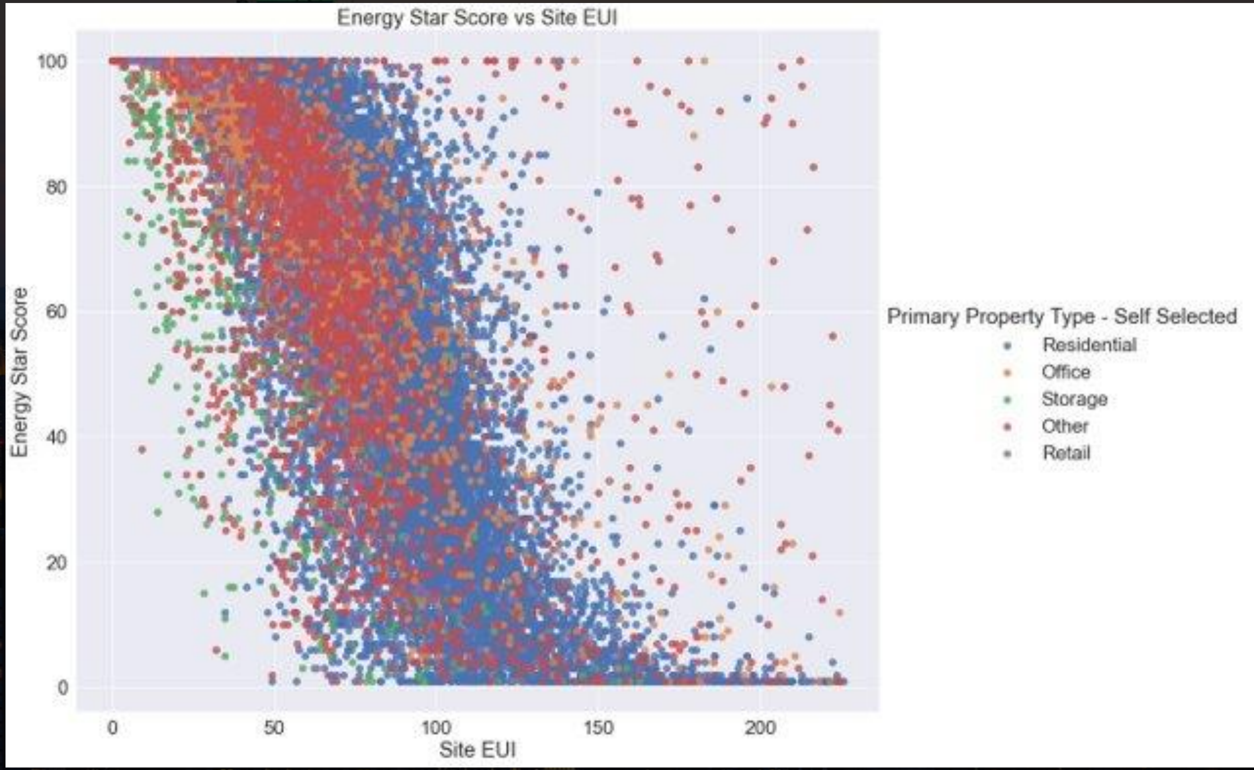
ENERGY STAR SCORE BY PROPETY TYPE



TWO VARIABLE PLOT

- The plot shows the expected negative relationship between Energy Star Score and Site EUI

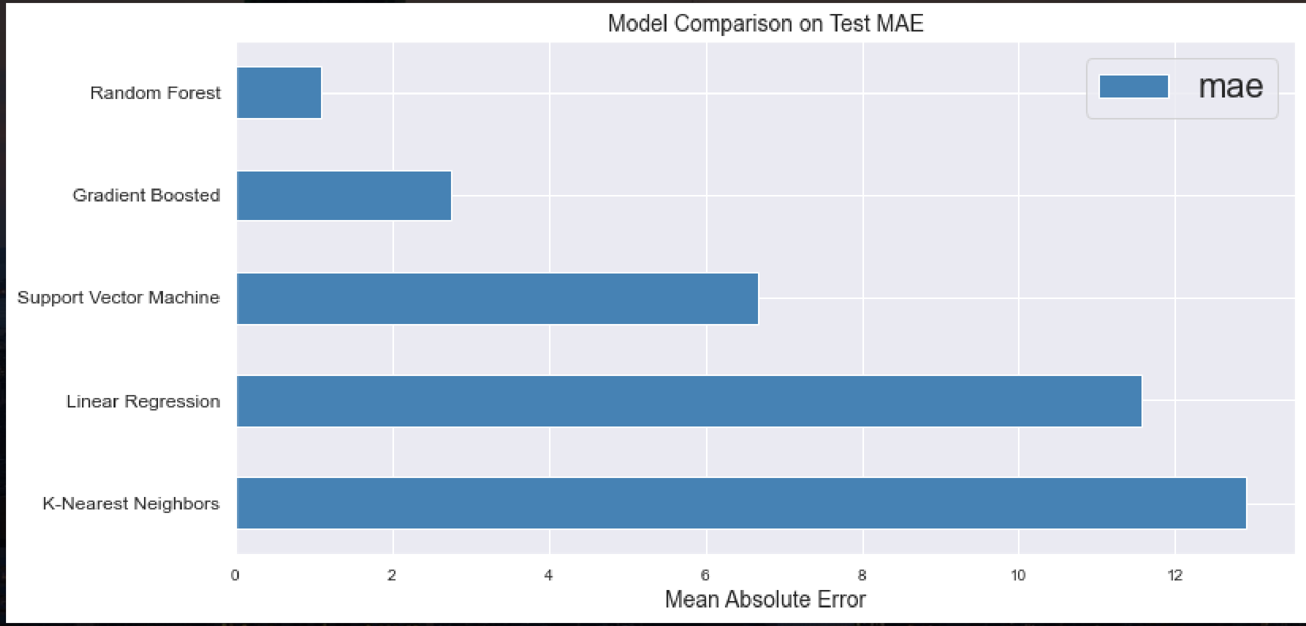
- This relationship holds across all property types.



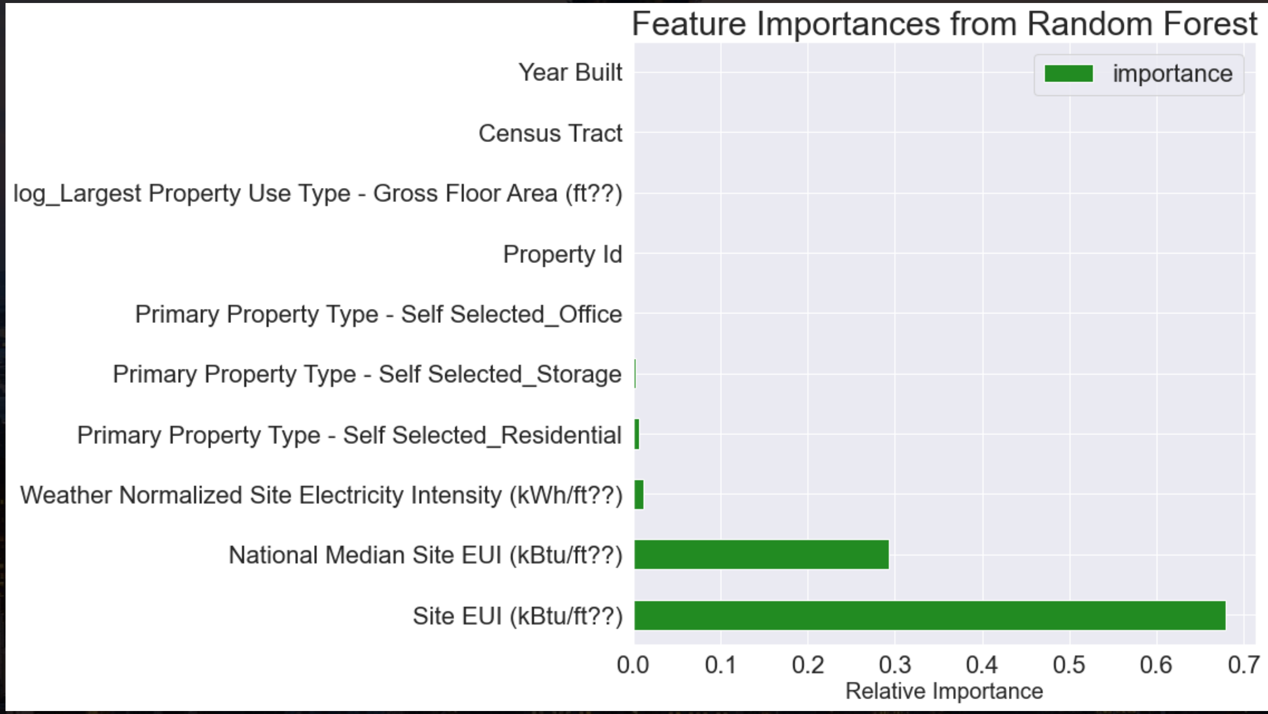
MACHINE LEARNING MODELS

- Metric: Mean Absolute Error, represents the average amount our estimate is off by in the same units as target value.

MODEL	MAE
Baseline	24.62
LR	11.55
KNN	12.89
SVM	6.68
GB	2.74
RF	1.29



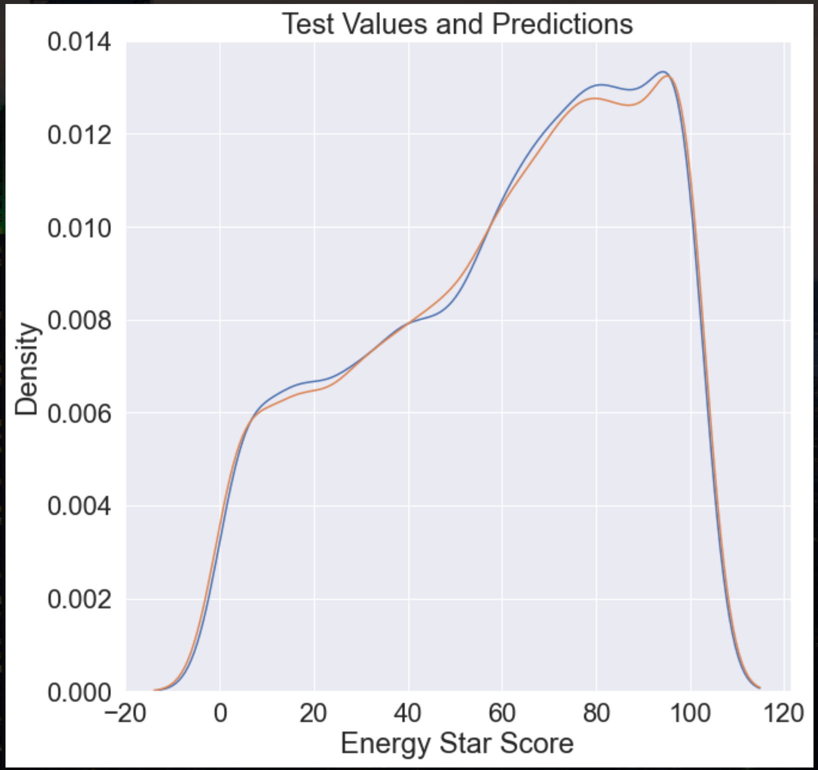
FEATURE IMPORTANCE



BEST ML MODEL

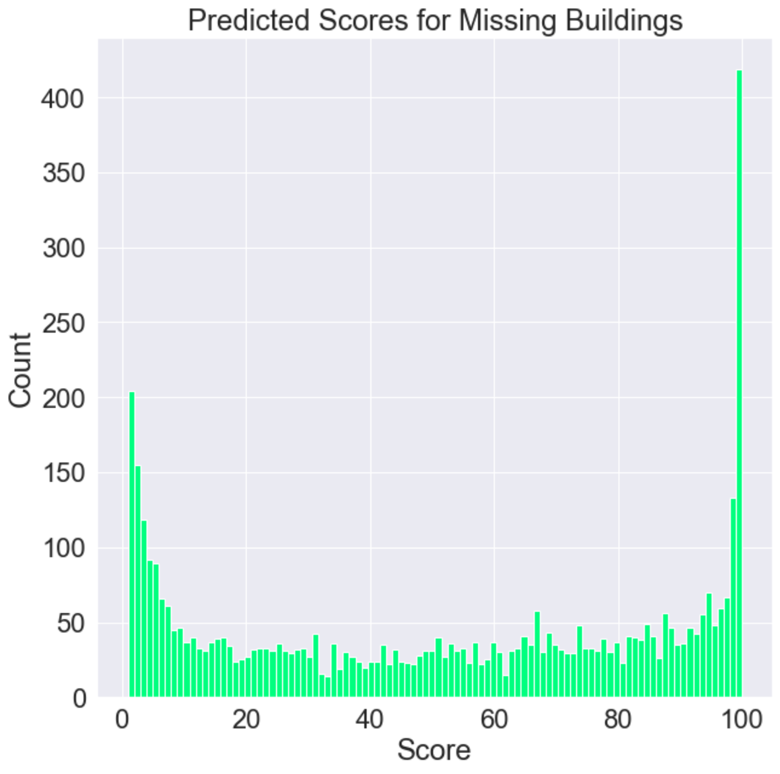
- The distribution looks to be nearly the same although the density of the predicted values is closer to the median of the test values rather than to the actual peak at 10

MODEL	MAE
RF-Base	1.29
RF-CrossVal	1.07
RF-Reduced	1.03



APPENDIX

- 4,457 buildings had missing ENERGY STAR Scores. Our model was able to predict the missing scores as long as the measurements recorded were in the original data.



APPENDIX II


Classiciations of Training Data
Poor
Fair
Average
Good
Excellent
Grade
0
1000
2000
3000
4000
5000
Buildings

CONCLUSIONS

- Energy Star Score might not be the most accurate measure of a buildings overall energy efficiency
- Disproportionate number of buildings have either 1 or 100 energy star scores
- Site EUI is more normally distributed and might be a more objective measure
- Regression and Classification both are able to produce reasonable predictions for Energy Star Score
- The most useful random forest features for predicting Energy Star Score are EUI, Property Type, and Electricity usage
- Finally I was able to classify the buildings into 5 grades as well as predict the score of buildings with missing score

RECOMMENDATIONS

- In order to receive a high ENERGY STAR Score, decreasing the Site EUI is essential.
- Since Natural Gas usage is highly correlated with EUI, reducing it could improve the score.
- Investing in more efficient heating and cooling services or slightly reducing the usage of either heating or cooling during the day when most people are not home.
- Another option is to reuse waste heat, such as reusing the vented air from electric clothes dryers, however, filtering and the air will be necessary and could alter the cost effectiveness of this approach.
- Replacing gas powered appliances such as stoves with more efficient electric ones is another option. However, the source of electric power should also be considered to weigh the overall energy efficiency and environmental impact.

Q&A