CUNY MSDS Capstone Project Summary

COMMERCIAL BUILDING ENERGY CONSUMPTION

ANALYSIS AND PREDICTION

May 31, 2019

John Grando
john.grando@spsmail.cuny.edu

Abstract

Commercial Building Energy Consumption accounts for approximately 25% of the United States energy production profile. Many economical and sociological factors are pushing owners of these buildings to reduce energy consumption and optimize performance. However, it is difficult to say whether a building is operating efficiently or not. Using publicly available data, models can be constructed to predict major fuel consumption. Keywords: building energy consumption, predicted energy consumption, baseline energy model.

Introduction

Every few years, the U.S. Energy Information Administration (EIA) conducts a survey attempting to record pertinent features of these buildings, known officially as the Commercial Buildings Energy Consumption Survey (CBECS). While the survey is expansive (i.e. more than 600 tracked features), it is useful to identify predictors that significantly affect consumption. This study will focus on creating a series of models to extract the most important survey questions and then use these values as predictors to train a final model that predicts fuel use consumption for standard practice buildings.

Data and Methods

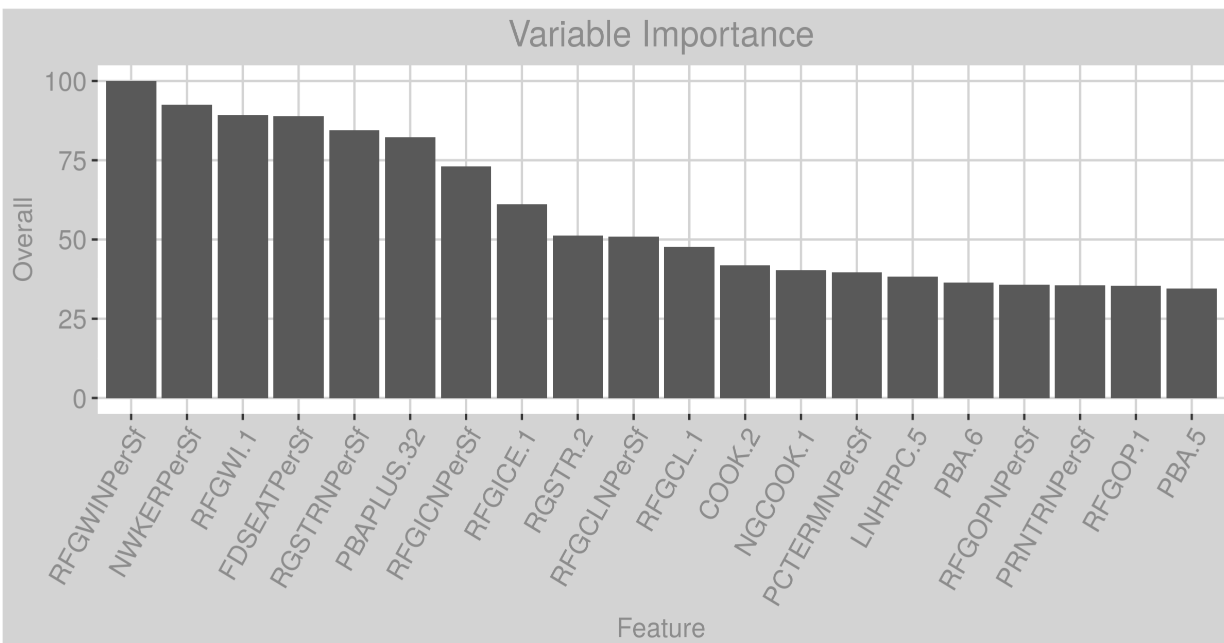
Due to the large number of features in the survey responses, it is not possible to analyze each one individually. Therefore, the first steps in the process will be centered around selecting smaller subsets from various feature extraction algorithms. The magnitude and contribution percentage of each variable will be considered in selecting features from this model. In order to try and normalize the data, the response variable was divided by the gross floor area of the building and reported in in units of BTU per square foot (e.g. ELBTU/PerSF, NGBTU/PerSF). A neural network model will be built to take the subset of extracted features and make predictions for the selected major fuel use. A variety of hyperparameters will be tested, using cross-validation, and compared on a common error metric. This step will reveal the optimal hyperparameter combination to use for the model.

Electricity Feature Extraction Results

In order to rank the most impactful features, the variable importance metrics from the feature extraction models were all set to the same scale then summed and re-normalized. It seems the attempts to create stratified random samples may have been beneficial in this case since there are some building type specific end-uses that are highly ranked. For this fuel source, there are many attributes associated with refrigeration, office, and food sales equipment. Also, the attribute identifying one of the more atypical building types, speaking in an energy intensity sense, has made it into the top 20 (PBA.5 [NON-REFRIGERATED WAREHOUSE]). Additionally, some occupancy features (NWKERPerSf, FDSEATPerSf) have been included which is expected given that they impact interior space cooling and ventilation loads. In an attempt to truly follow the important predictors, no variables have been removed from this set and the order of importance remains unchanged.

Feature Extraction Model Results

Model	RMSE	R2	MAE	MAPE	MSLE
partialLeastSquares	49439.16	0.5119599	26622.19	85.70943	0.5680023
neuralNetwork	51378.59	0.4622766	29989.65	133.42257	0.8016741
recursiveFeatureExtraction	49713.91	0.5017194	31111.93	239.40187	1.0286980
randomForest	165662.98	0.1388473	144344.47	881.85161	3.3488969
leaps	91256.75	0.3157366	58679.30	99.92406	66.6999846

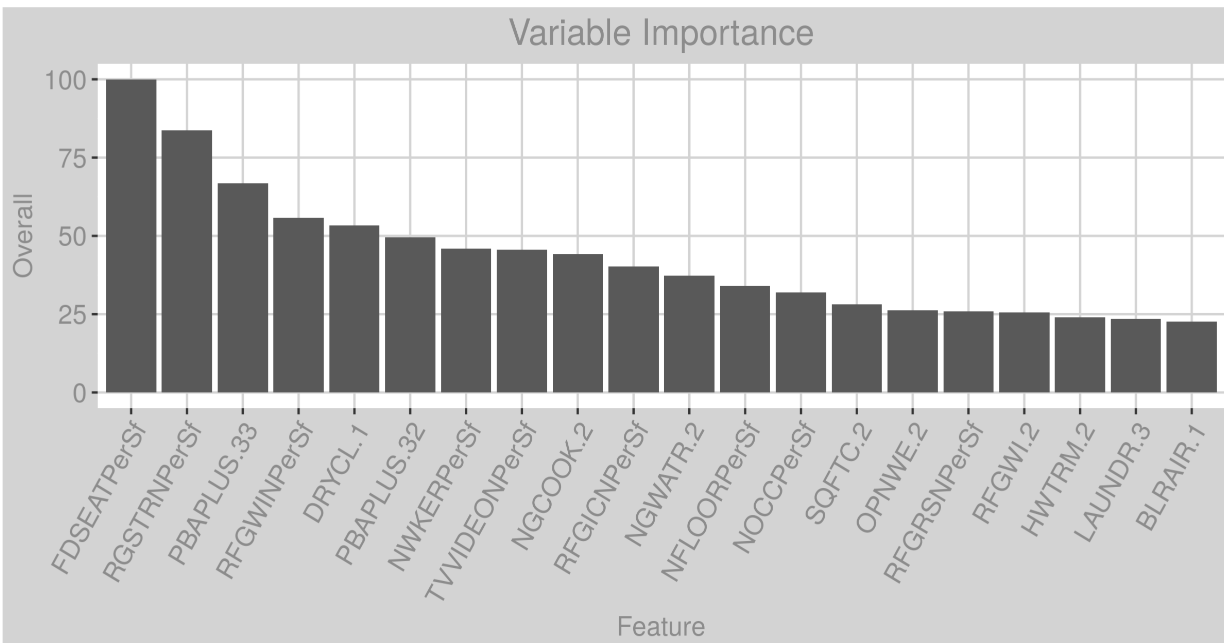


Natural Gas Feature Extraction Results

As with the electricity results, attributes related to occupancy seem to have made a large impact, possibly due to the need to heat ventilation air, especially given some of these occupancy types are associated with 24/7 operation. Specifically, buildings which report a high density of food service seating appear to be directly correlated with high gas usage, which is not surprising. Also as expected, cooking and large heating equipment attributes are high on the list.

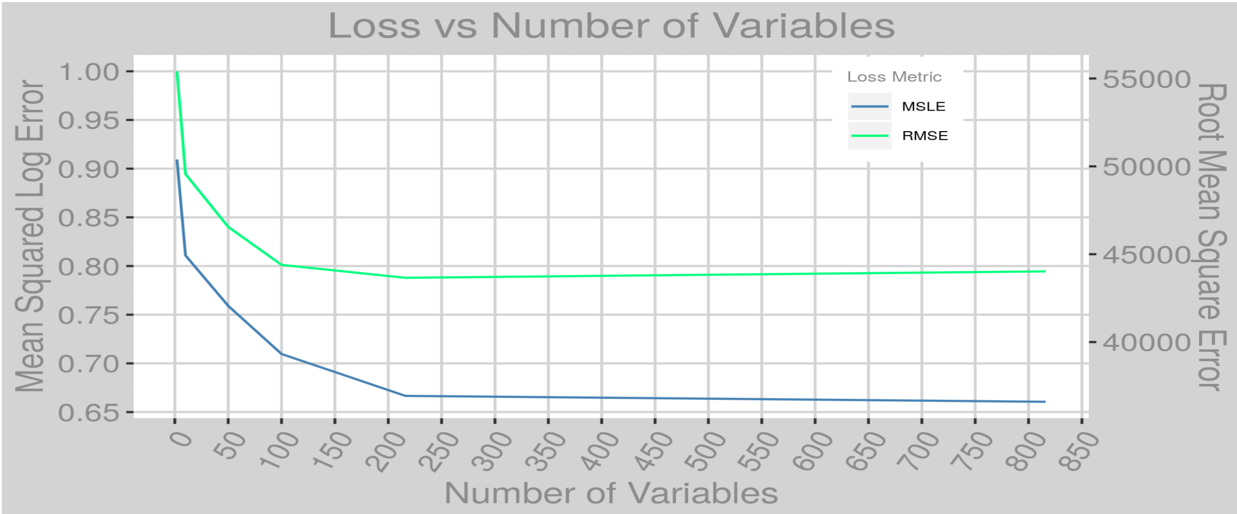
Feature Extraction Model Results

Model	RMSE	R2	MAE	MAPE	MSLE
partialLeastSquares	67978.44	0.2321160	31769.30	469.5262	1.215225
leaps	69079.81	0.1837522	34964.80	263.2334	1.367507
neuralNetwork	57929.72	0.3193293	33328.70	1660.1715	1.731125
recursiveFeatureExtraction	61464.86	0.2589020	35070.92	1887.8960	2.018073
randomForest	174127.85	0.1215693	147326.47	8693.7861	5.379562



Neural Network Summary - Electricity

The final selected model consisted of a 5 hidden layers, 1000 hidden layer nodes, a dropout rate of 0.3, no regularization, batch sizes of 150, using the rmsprop() algorithm with a learning rate of 0.0005, and 100 predictors. As can be seen in the graph below, the number of variables needed to obtain near-peak performance, is much less than the full set. The final selected model, after re-training, has a MSLE of 0.92 and RMSE of 54696. Comparing this model ('Full Neural Network') to the previous feature extraction models, which used many more variables, the performance is competitive. Additionally, the results were then multiplied by their respect gross floor area and then compared to the set of feature extraction models, with the same transformation, in order to evaluate the total consumption prediction error. Again, it can be seen that this neural network model has shown to be competitive in this manner and, in fact, has a better Rsquared value.

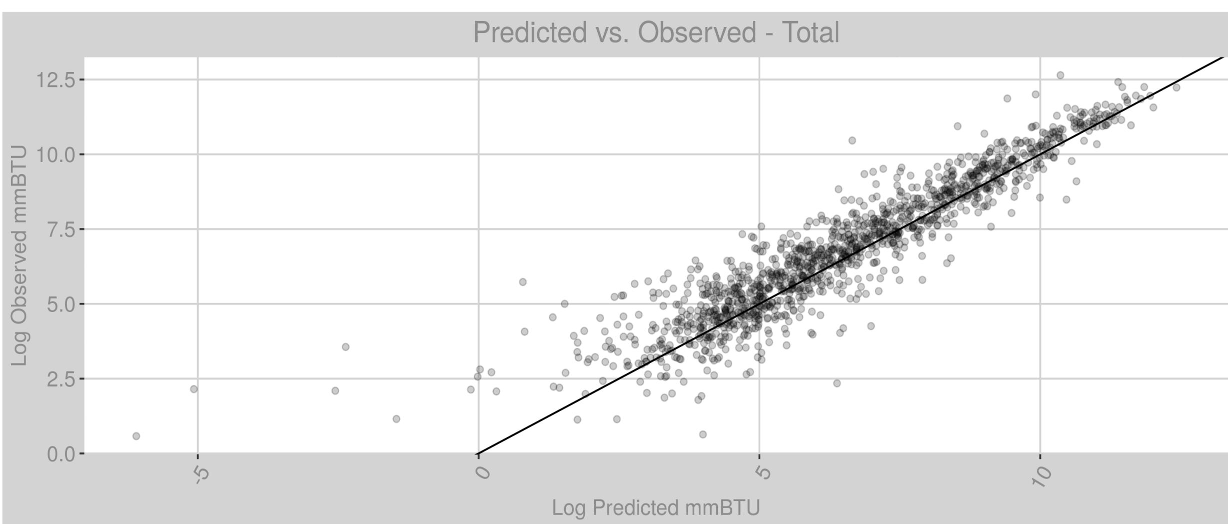


Per SF Model Comparison [BTU/SF]

Model	RMSE	R2	MAE	MAPE	MSLE
partialLeastSquares	49439	0.51	26622	86	0.57
neuralNetwork	51379	0.46	29990	133	0.80
Full Neural Network	54696	0.49	28463	67	0.92
recursiveFeatureExtraction	49714	0.50	31112	239	1.03
randomForest	165663	0.14	144344	882	3.35
leaps	91257	0.32	58679	100	66.70

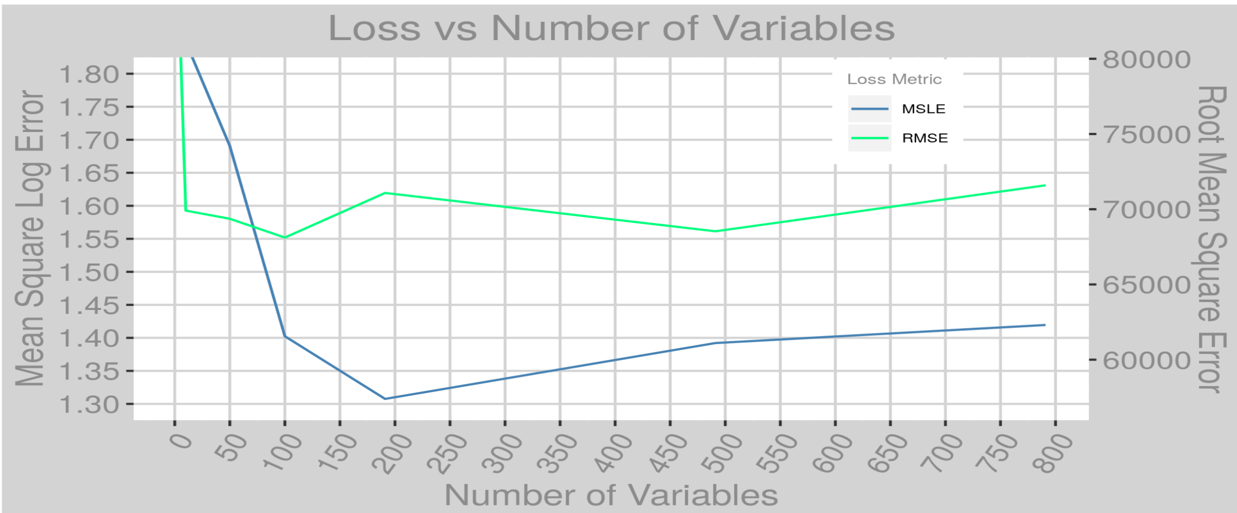
Total Consumption Prediction vs. Feature Extraction Models [mmBTU]

Model	RMSE	R2	MAE	MAPE	MSLE
partialLeastSquares	16180	0.62	3813	86	0.57
neuralNetwork	14267	0.68	3616	133	0.80
Full Neural Network	13542	0.72	3514	67	0.92
recursiveFeatureExtraction	13308	0.71	3535	239	1.03
randomForest	52723	0.68	21239	882	3.35
leaps	26036	0.65	8831	100	66.70



Neural Network Summary - Natural Gas

The final selected model consisted of a 5 hidden layers, 800 hidden layer nodes, a dropout rate of 0.3, no regularization, batch sizes of 150, using the rmsprop() algorithm with a learning rate of 0.0005, and 100 predictors. As can be seen in the graph below, the number of variables needed to obtain near-peak performance, is much less than the full set. The final selected model, after re-training, has a MSLE of 1.42 and RMSE of 61341. Comparing this model ('Full Neural Network') to the previous feature extraction models, which used many more variables, the performance is actually better (when using RMSE).

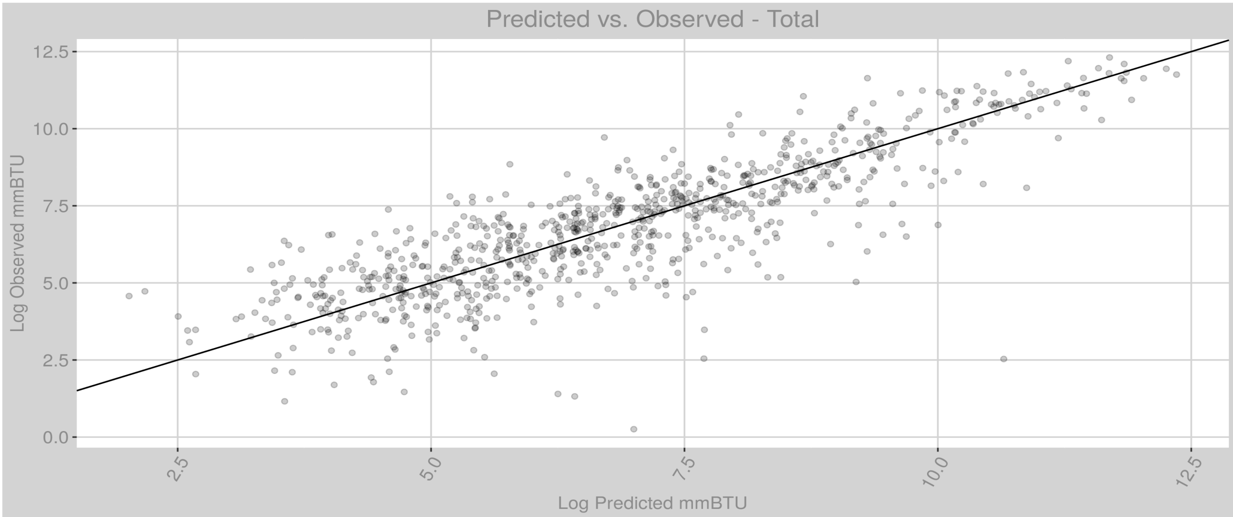


Per SF Model Comparison [BTU/SF]

Model	RMSE	R2	MAE	MAPE	MSLE
partialLeastSquares	67978	0.23	31769	470	1.22
leaps	69080	0.18	34965	263	1.37
Full Neural Network	61341	0.29	32608	713	1.42
neuralNetwork	57930	0.32	33329	1660	1.73
recursiveFeatureExtraction	61465	0.26	35071	1888	2.02
randomForest	174128	0.12	147326	8694	5.38

Total Consumption Prediction vs. Feature Extraction Models [mmBTU]

Model	RMSE	R2	MAE	MAPE	MSLE
partialLeastSquares	15567	0.45	3671	293	1.19
leaps	15769	0.43	3950	213	1.35
Full Neural Network	13191	0.69	4267	713	1.42
neuralNetwork	13365	0.59	3641	955	1.70
recursiveFeatureExtraction	14798	0.50	4016	1172	2.00
randomForest	43661	0.43	16812	5332	5.36


Predicted vs. Observed - Total
Log Observed mmBTU
Log Predicted mmBTU
12.5
10.0
7.5
5.0
2.5
0.0
2.5
5.0
7.5
10.0
12.5