Energy Efficiency and Machine Learning: Understanding Technology Adoption Decisions

Aramayis Dallakyan and Reid B. Stevens

Deptartment of Agricultural Economics, Texas A&M University

Objectives

This research focuses on two questions:

- Can machine learning models predict energy efficiency adoption decisions better than traditional econometric models using the same set of variables?
- Can machine learning models use variables not traditionally included in econometric models to improve energy efficiency adoption predictions?

Introduction

The energy efficiency gap (difference between the cost-minimizing energy efficiency investment level and the observed investment level), has attracted significant attention from academic researchers [1]. Standard economic models struggle to explain much of the observed variation energy efficiency investment [2]. We find that a traditional econometric model of energy efficiency adoption decisions [2] only correctly predicts 50 percent of adoption decisions out of sample.

We use machine learning models to improve modeling of energy efficiency adoption decisions by

- using non-linear models (e.g. decision trees)
- using a larger set of variables

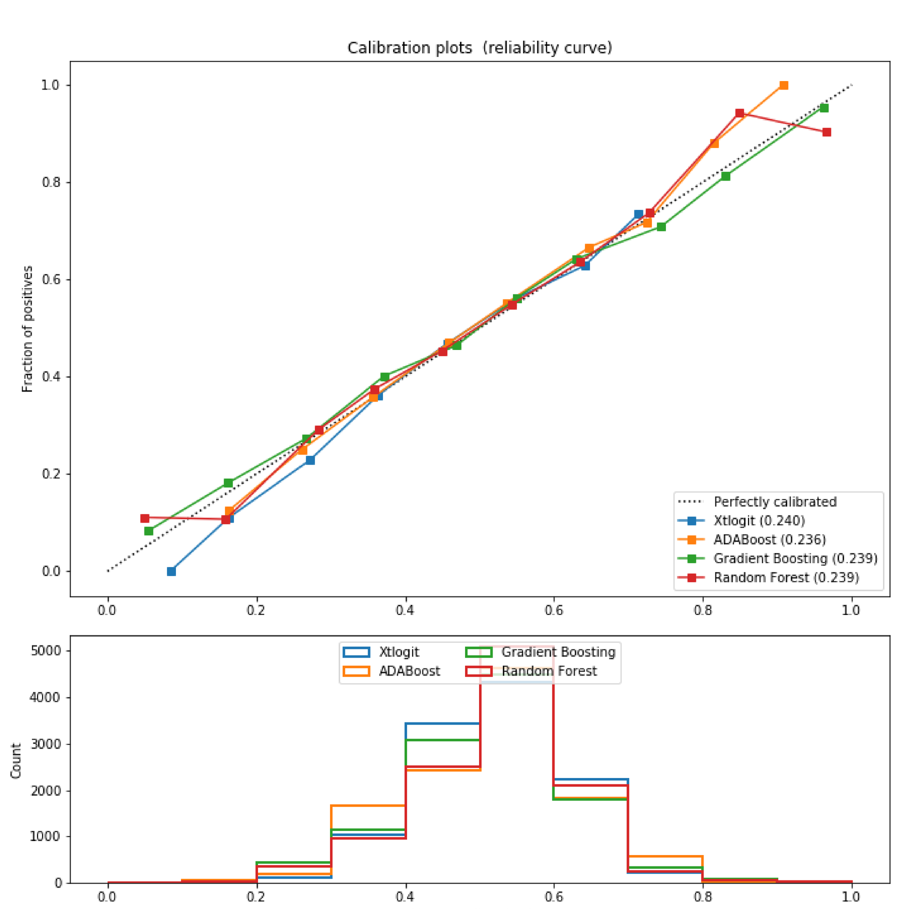

Figure 1: Comparing Accuracy of Economic and Machine Learning Models

Data

The Department of Energy's (DOE) energy audit program provides a detailed micro-dataset of energy efficiency adoption decisions. This data set records every energy efficiency investment recommended by DOE auditors and the firm's investment decision. These data include recommendation-level data on cost, benefit, and payback period of proposed energy efficiency investments, and various firm-level data, including sales, industry, employees.

New inputs, New models

Next, we determine whether other variables collected by the energy efficiency auditors are useful in explaining technology adoption. Additional independent variables include past energy use, auditor experience, and auditor eduction level. We estimate a variety of variable selection models to determine which variables are most useful in predicting technology adoption. We use cross-validation to avoid overfitting and compare model performance using out-of-sample data.

Important Result

Machine learning model predictions of energy efficiency adoption decisions are **40% better** than the logistic models typically used by researchers. This forecast improvement comes from allowing for nonlinearities and also allowing for the inclusion of independent variables in the model that are overlooked in the literature. This paper demonstrates the value of machine learning techniques to understanding the energy efficiency gap, as well as other topics in energy economics.

Same inputs, New models

ML Models

We compare the following machine learning models to a traditional econometric model of energy efficiency adoption decisions (panel logistic regression) using the same set of variables,

- Adaptive Boosting
- Gradient Boosting
- Random Forest

The out-of-sample performance of the machine learning models, plotted in Figure 1, strictly dominate the logistic model predictions. Evidently, these machine learning algorithms capture important nonlinearities in the energy efficiency adoption decisions that are not accounted for in the panel logistic regression.

Results

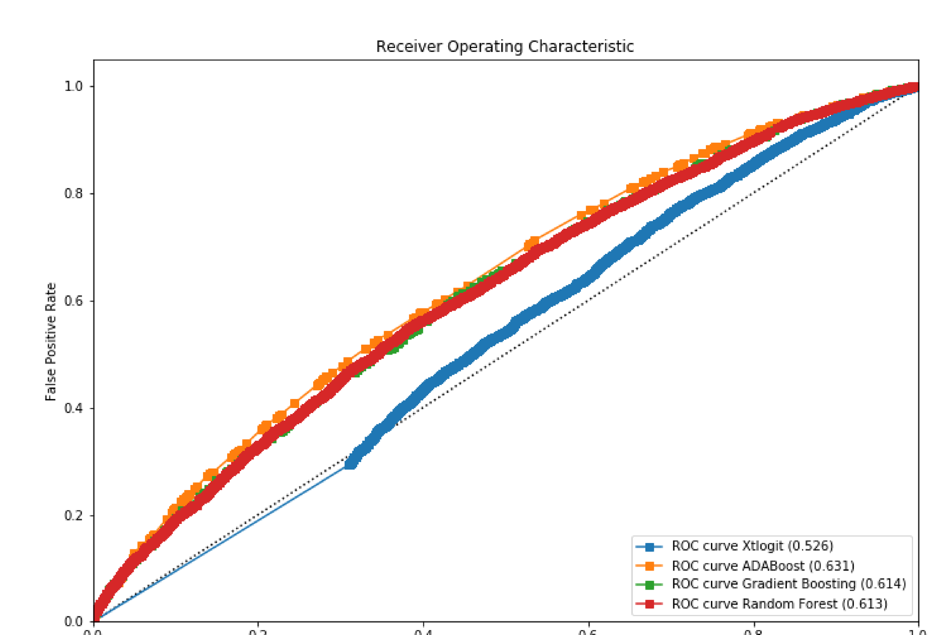

Figure 2: Comparing model accuracy with additional features

Figure 2 compares the forecast accuracy using a larger set of features. The best machine larning model, ADABoost, accurately classifies 70 percent of energy efficiency adoption decisions, while the logistic model only correctly classifies about 50 percent.

Results, cont'd

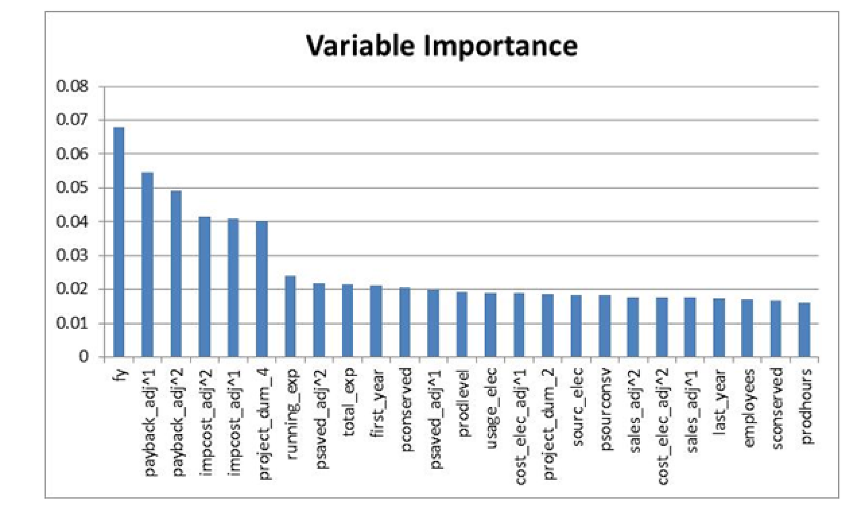

Figure 3: Feature Importance

Though all models both place substantial weight on the expected variables (payback period, installation cost, etc.), the machine learning models improve predictive power by accounting for variables not typically included in energy efficiency adoption models (production level and operation hours). Figure 3 displays the weights of the most important variables in the machine learning models.

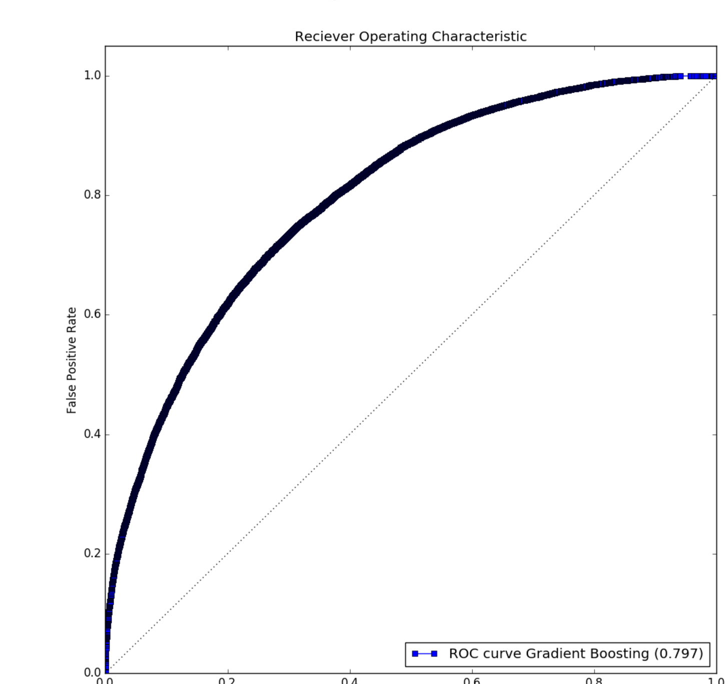

Figure 4: Machine learning model accuracy

The gradient boosted decision tree model gives the most accurate energy efficiency adoption decisions, with 40 percent more correct classifications than a logistic regression out of sample.

References

[1] Kenneth Gillingham and Karen Palmer.
Bridging the energy efficiency gap: Policy insights from economic theory and empirical evidence.
Review of Environmental Economics and Policy, 8(1):18–38, 2014.

[2] Soren T Anderson and Richard G Newell.
Information programs for technology adoption: the case of energy-efficiency audits.
Resource and Energy Economics, 26(1):27–50, 2004.