



# Predicting Home Improvement: Green Or Not?

---

**Amber Rivera**

*<http://bit.ly/ee-upgrades>*



# Predict whether a household will choose an energy efficient upgrade

---

Building characteristics

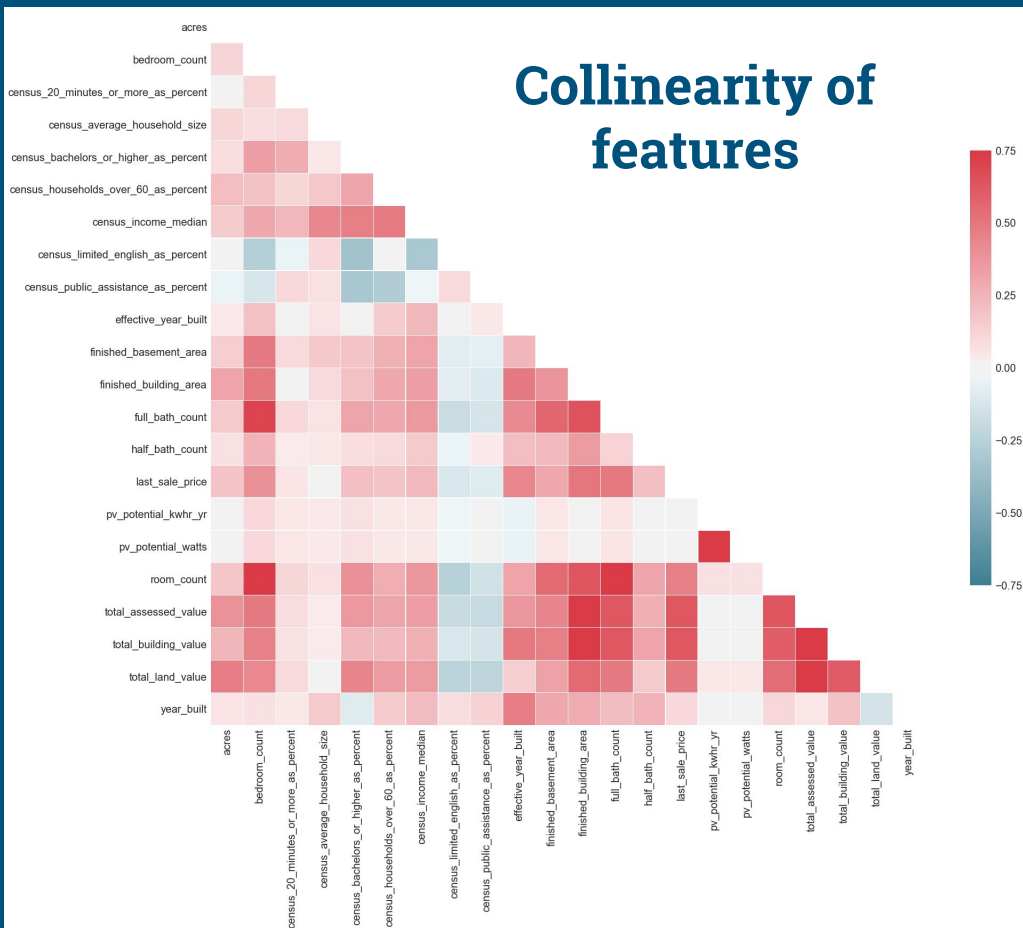
Census information

Simulated energy usage



# The Data

- 18,000 homes, 360 features
  - 30% numerical - Last Sale Price
  - 70% categorical - AC Type, Garage
- New features:
  - *# upgrades in neighborhood*
  - *# permits since purchase*
- Only 9% of homes in positive class
  - Tried upsampling minority class
    - Random Over Sampling
    - SMOTE
  - Downsampling to 50/50 generalized best!

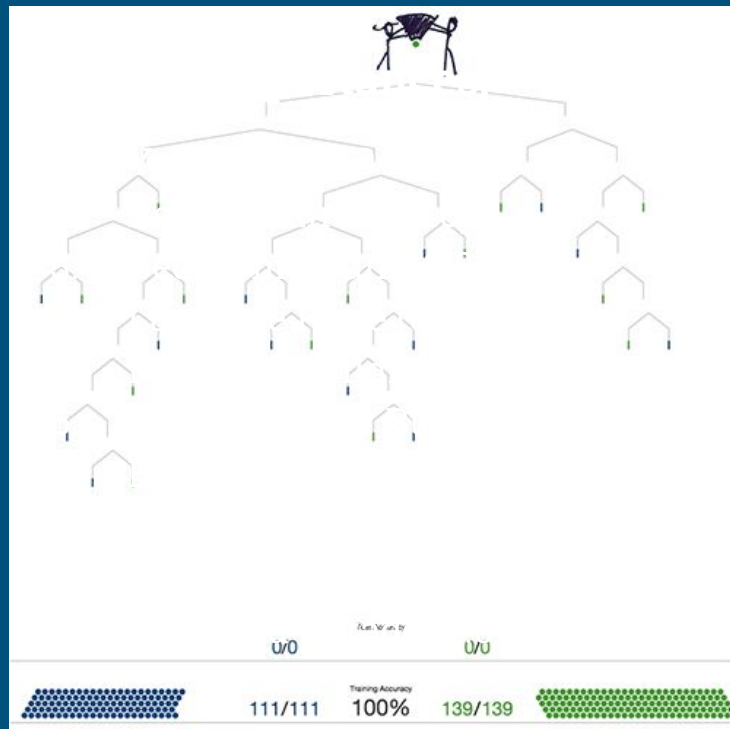


# Finding The Best Model

## Random Forest (200 trees)

- Binary classification; 30% holdout
- Evaluated with Recall metric
- Seeking stability and interpretability

Final results:					
Confusion Matrix		precision	recall	f1-score	support
TP: 342	0	0.94	0.53	0.68	4716
FP: 2193	1	0.13	0.70	0.23	490
FN: 148					
TN: 2523					
avg / total		0.87	0.55	0.64	5206



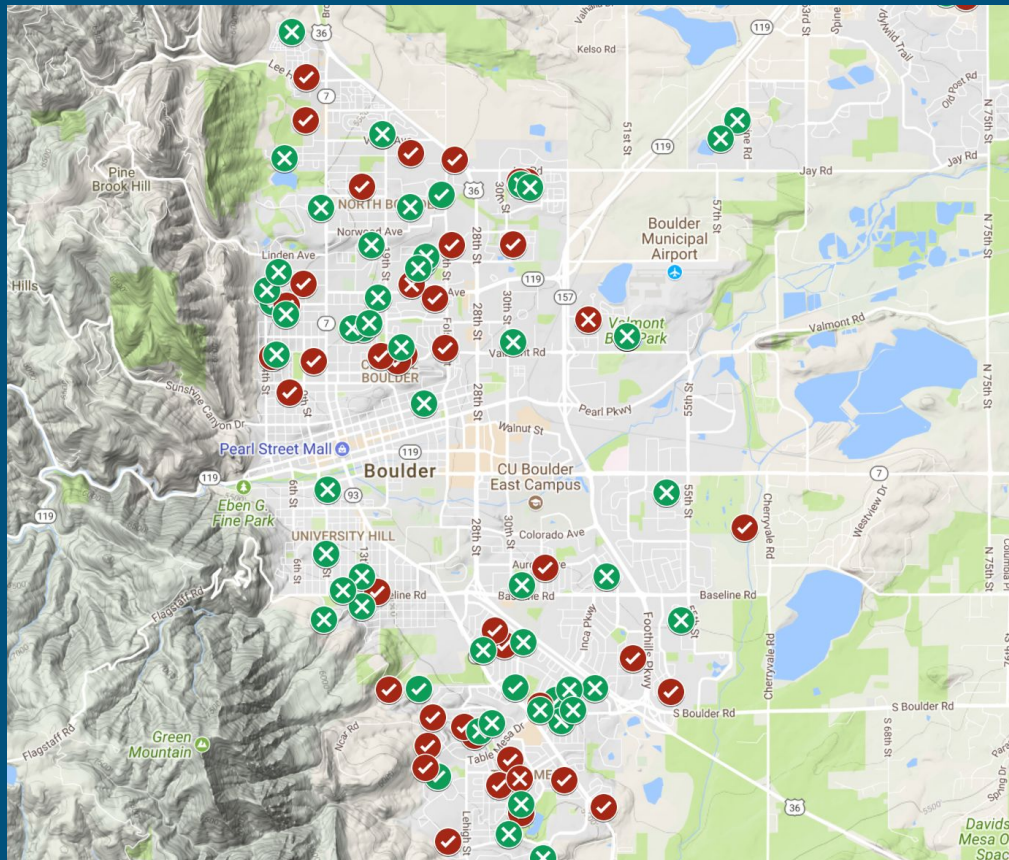
# Making Predictions

- ✓ TP: Has upgraded; Predicted 'Yes'
- ✓ FP: Has not upgraded; Predicted 'Yes'
- X TN: Has not upgraded; Predicted 'No'
- X FN: Has upgraded; Predicted 'No'

## Business Implications:

- Volume of 500 jobs per year
- TP:  $(111) * (\text{revenue} - \text{cost})$
- FP:  $(389) * (\text{cost})$

**= 13% potential increase in profit**





# Next Iteration

---

- Supervised
  - Incorporate behavioral information
  - Develop more sophisticated handling of class imbalance
- Unsupervised
  - Clustering algorithms
  - Outlier detection algorithms



More at: *<http://bit.ly/ee-upgrades>*

*Special thanks to the team at Radiant Labs*

**Contact:** [amber.j.rivera@gmail.com](mailto:amber.j.rivera@gmail.com)

[linkedin.com/in/amberjrivera](https://www.linkedin.com/in/amberjrivera)

### **Technologies Used:**

- Python, Pandas, NumPy for data analysis
- Matplotlib, Seaborn , Google for visualization
- Scikit-learn and imbalanced-learn for machine learning
- Check out this gist I co-wrote on Sklearn's Pipeline constructor:  
*<http://bit.ly/Pipeline-gist>*

