American Housing Stock

Classifying Adequacy of Units Based on Neighborhood Characteristics

Seoho Hahm and Nicole Bills

Objective

This analysis aims to assist the U.S. Department of Housing and Urban Development (HUD) to better understand the adequacy of housing units based on neighborhood characteristics with the key question:

What neighborhood characteristics are important for classifying units as adequate?

Where any 1 of 14 different situations related to the unit itself can result in the classification of inadequacy.

Key Terminology

U.S. Department of Housing & Urban Development (HUD): federal agency dedicated to strengthening and supporting the housing market; primary responsibilities include:

- Protecting housing consumers
- Encouraging production of affordable rental housing
- Preventing and punishing discrimination in housing

Housing unit: a single house, apartment, or other dwelling

Jump to Data Source - American Housing Survey

Background

Data Distribution

- Findings -

Adequate Units	62,185 (93%)
Non-Adequate Units	4,567 (7%)

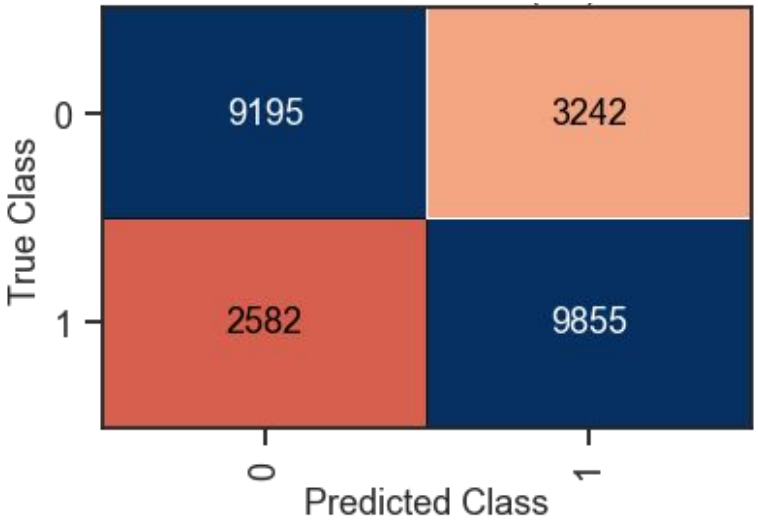
- Action -

Because 93% of housing in the survey was classified as adequate, we balanced our classes using synthetic data to give our models the best chance of having predictive power.

Results

Top Performing Model

- Results -



- Discussion -

Our most successful model received a score[1] of 76% on its classification of inadequate housing.

1. Refers to F1-Score, the weighted average of Precision and Recall
Jump to Model Comparison - Classification Metrics

Results

Feature Importance

Neighborhood Rating

Presence of Trash

Petty Crime

Presence of Abandoned Blds

In Subdivision

Presence of Bars on Windows

Serious Crime

Risk of Natural Disasters

Majority Neighbors 55+

Quality of Public Transit

Quality of Schools

Jump to Model Comparison - Feature Importance

Recommendations

- Use Case -

U.S. HUD can use this model as a portable solution to assess areas for further study with regards to housing adequacy based on neighborhood characteristics.

- Next Steps -

1. If needed, HUD could build models on other subsets of AHS features to use where full information is not available
2. HUD could also build models taking into account prior years' data
3. Compare to objective measures of features assessed in AHS

Thank you!

Appendix

Appendix

- Data Source
- Model Comparison - Classification Metrics
- Model Comparison - Feature Importance

Data Source

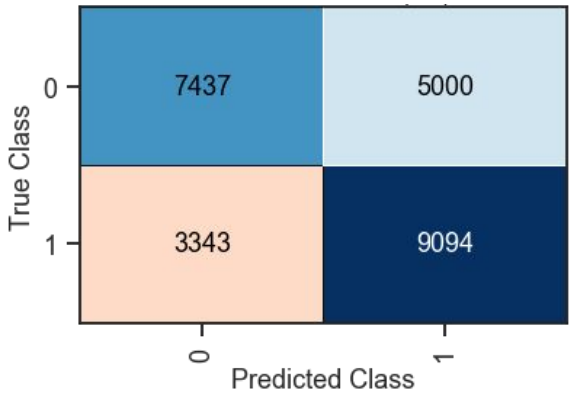
American Housing Survey (AHS):

- Biennial voluntary longitudinal survey that provides current and ongoing series of data on the size, composition, and state of housing in the United States and changes in the housing stock over time
- Collects housing statistics that the U.S. Department of Housing and Urban Development (HUD) uses to evaluate and develop its federal housing programs
- Sample size dependent on HUD budget and has varied over the years (e.g. in 2009 about 62,000 addresses were selected for the National survey)
- Each sample unit from the basic sample has been visited every other year since 1985. New addresses are added to the sample at each iteration to ensure representativeness

Model Comparison

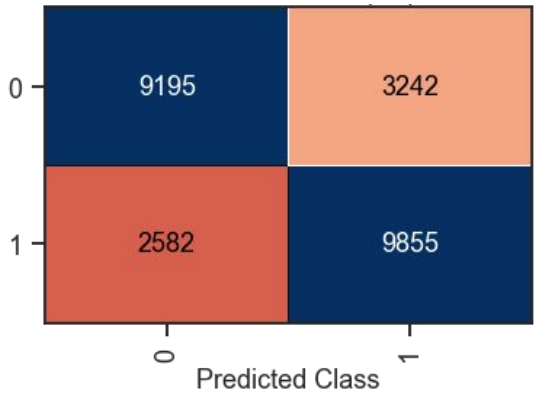
Classification Metrics

- Logistic Regression -



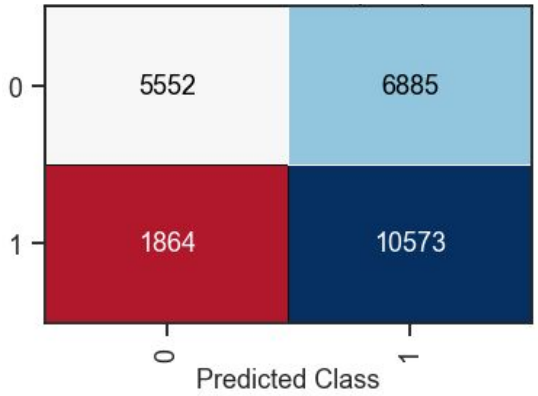
	precision	recall	f1-score
0	0.69	0.60	0.64
1	0.65	0.73	0.69
accuracy			0.66
macro avg	0.67	0.66	0.66
weighted avg	0.67	0.66	0.66

- Random Forest -



	precision	recall	f1-score
0	0.78	0.74	0.76
1	0.75	0.79	0.77
accuracy			0.77
macro avg	0.77	0.77	0.77
weighted avg	0.77	0.77	0.77

- SVM -

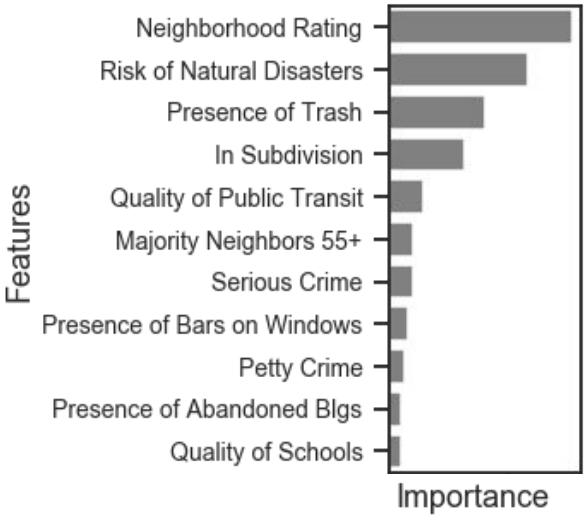


	precision	recall	f1-score
0	0.75	0.45	0.56
1	0.61	0.85	0.71
accuracy			0.65
macro avg	0.68	0.65	0.63
weighted avg	0.68	0.65	0.63

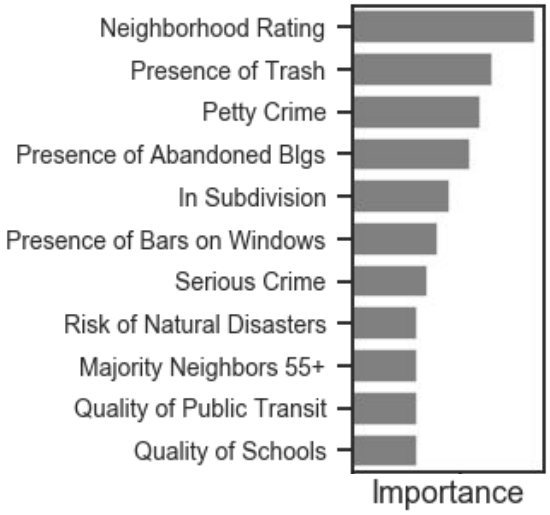
Model Comparison

Feature Importance

- Logistic Regression -



- Random Forest -



- SVM -