

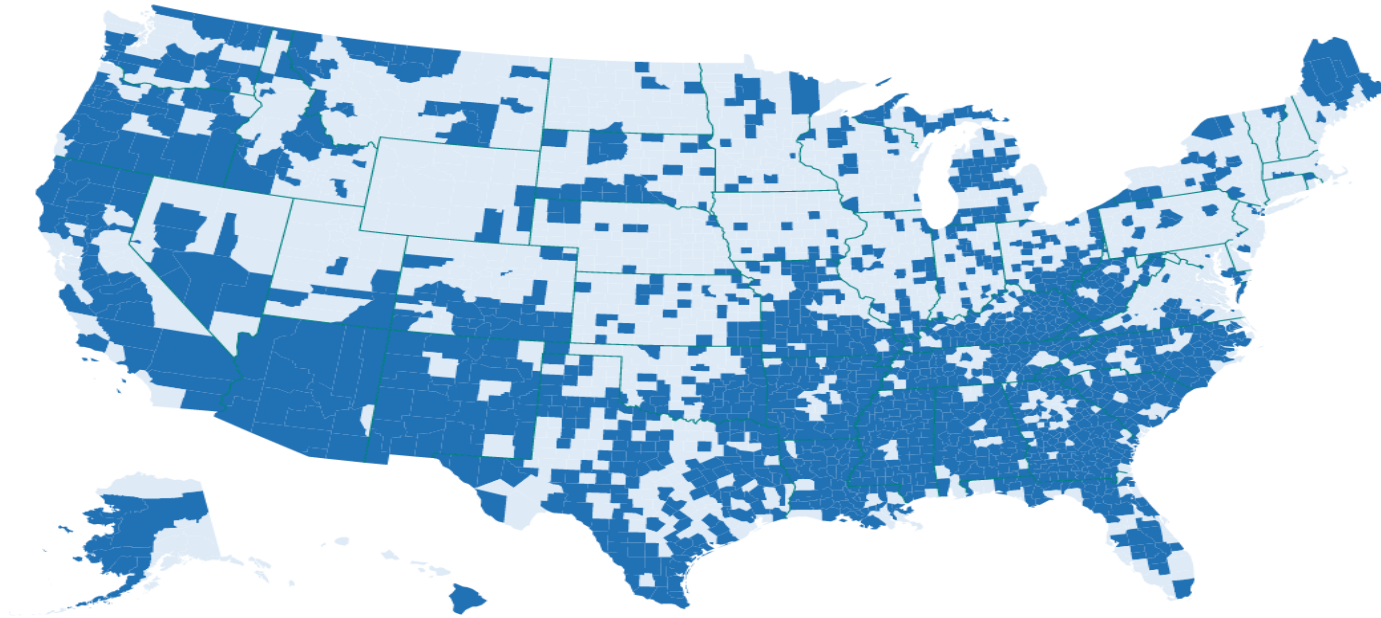
Poverty Extent via Community Factors

Kimberley Mitchell
Metis Classification Project

Data: Response

- Poverty rate:
 - Per United States county
 - Binned: above (1) or below (0) median
 - Median county poverty rate = 16.0%
- Interactive poverty rate map:
 - <http://mitchki.com/D3/poverty.html>

Label: High / Low Poverty



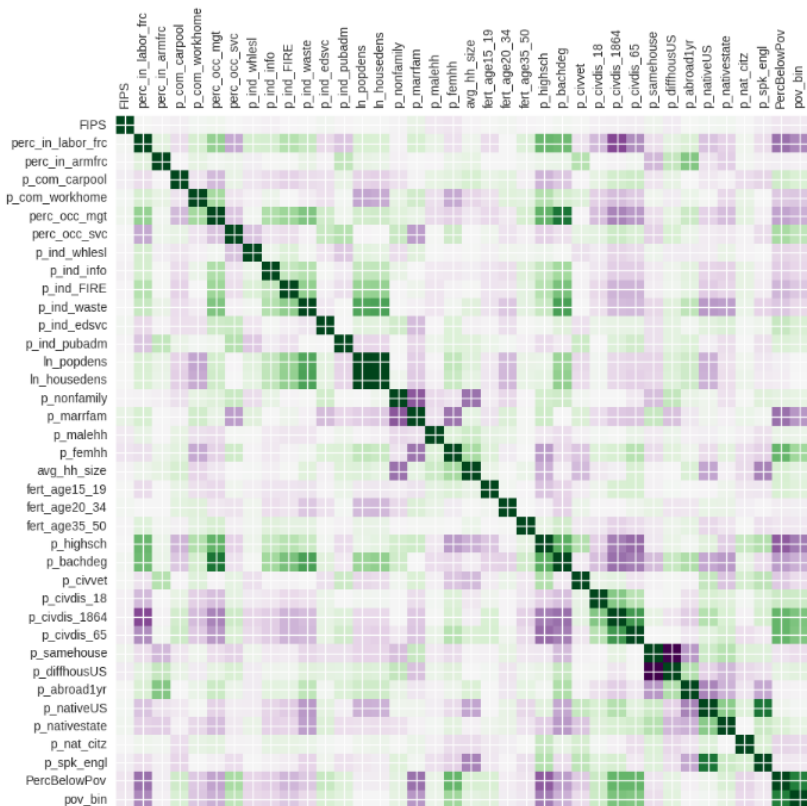
Feature Categories

- Household configuration
- Commute type
- Occupations, Industries
- Workforce characteristics
- Rural /Urban (by population density)
- Population transience
- Citizenship, language, fertility
- Educational attainment

Methods

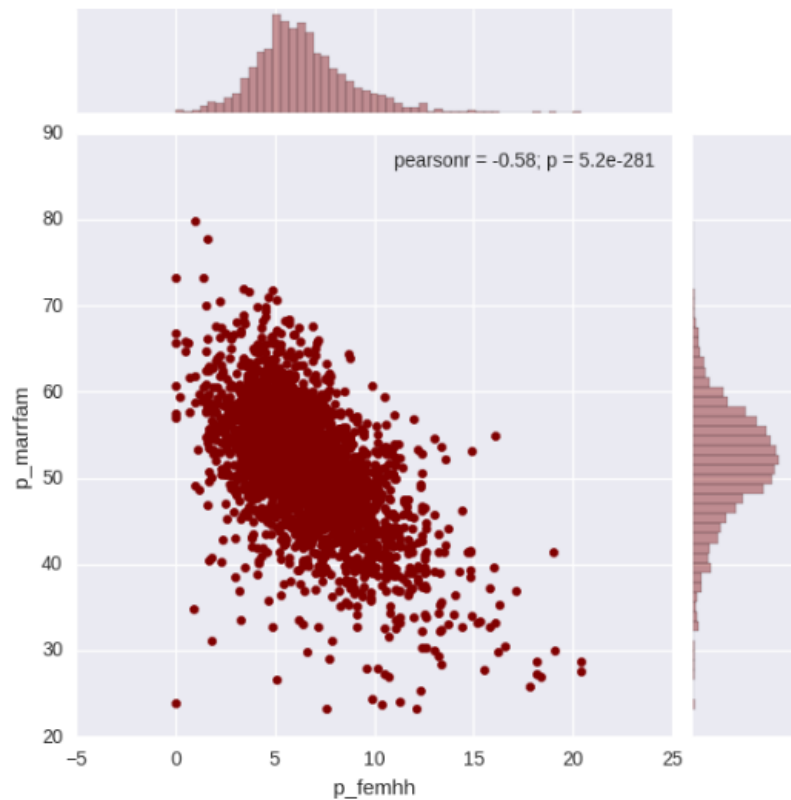
- Exploratory Data Analysis (EDA)
- Logistic Regression (also w/ regularization)
- Naive Bayes Classification
- Random Forest Classification

Exploratory Data Analysis



Heat map –
Feature / Label correlations

Exploratory Data Analysis

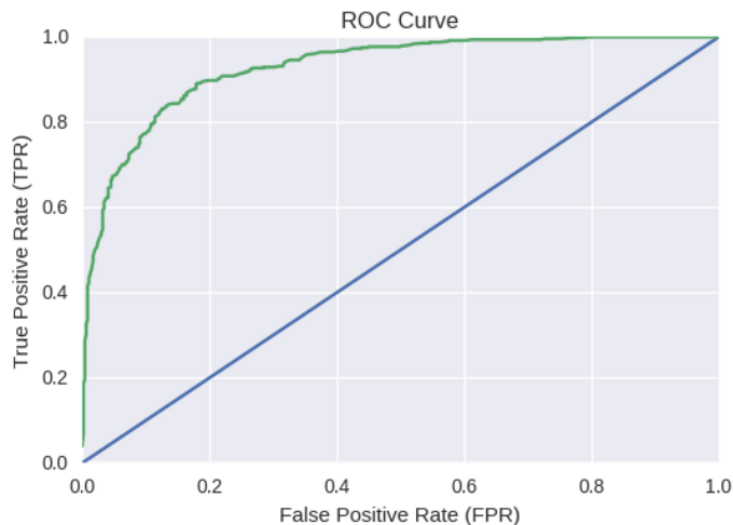


Joint plot –
Tool to explore joint distributions

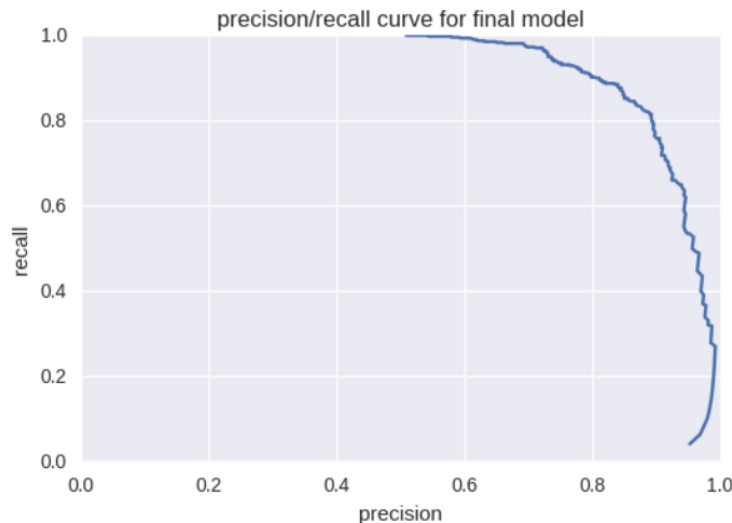
Random Forest Results



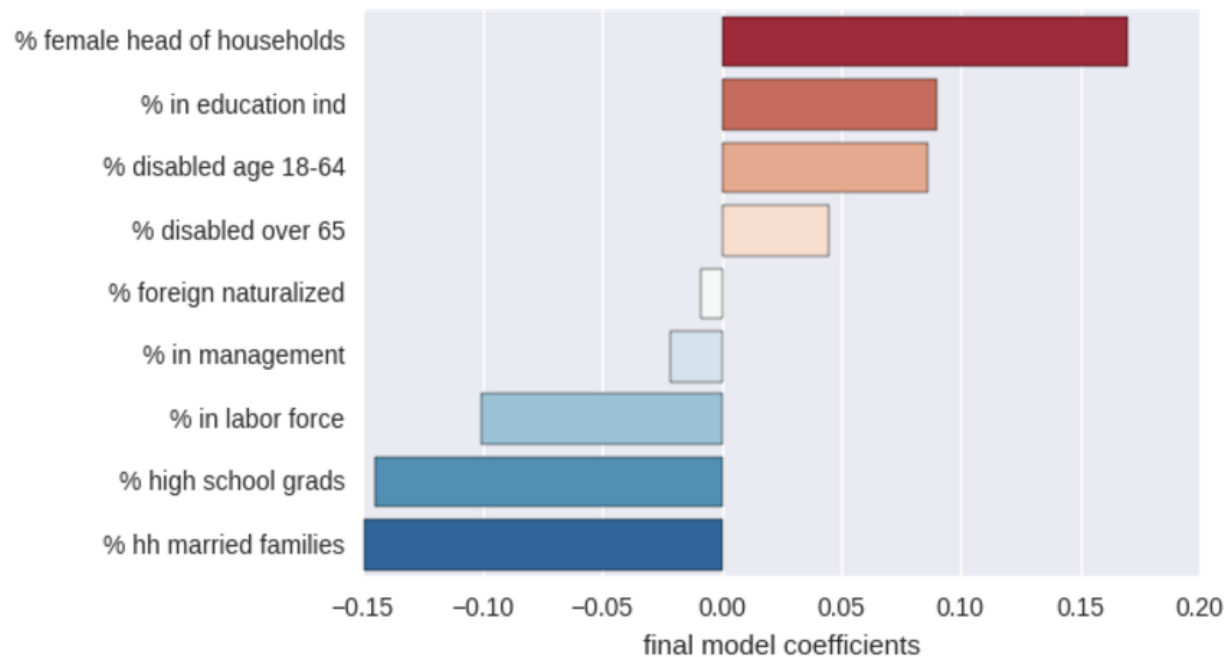
Random Forest Results



Random Forest Classifier (sklearn)
Accuracy: 0.855
AUC: 0.927



Logit Coefficient Results



Logit (statsmodels) from Random Forest feature importances
- Coefficients show change in odds of above-median poverty

Conclusions

- From a few key factors, we can predict high / low poverty rates.
- Future work:
 - Check trends for consistency over time
 - Check if same factor universe can predict other measures of well-being

Questions???

