684: Final Project

Rajan (Raj) Desai

Modeling Food Insecurity: Identifying Demographic Risk Factors for Low Access Tracts

Background

- The USDA defines low access to healthy food as being far from a supermarket, supercenter, or large grocery store.
- A census “tract” is considered to be low access if it meets the criteria below.
- Low access tracts:
  - Low-income community
    - The tract’s poverty rate is 20 percent or greater;
    - The tract’s median family income is less than or equal to 80 percent of the Statewide median family income
    - The tract is in a metropolitan area and has a median family income less than or equal to 80 percent of the metropolitan area's median family income.
  - A significant number (at least 500 people) or share (at least 33 percent) of the population is greater than one mile from the nearest supermarket, supercenter, or large grocery store for an urban area or greater than 10 miles for a rural area.
    - 53.6 million people, or 17.4 percent of the U.S. population

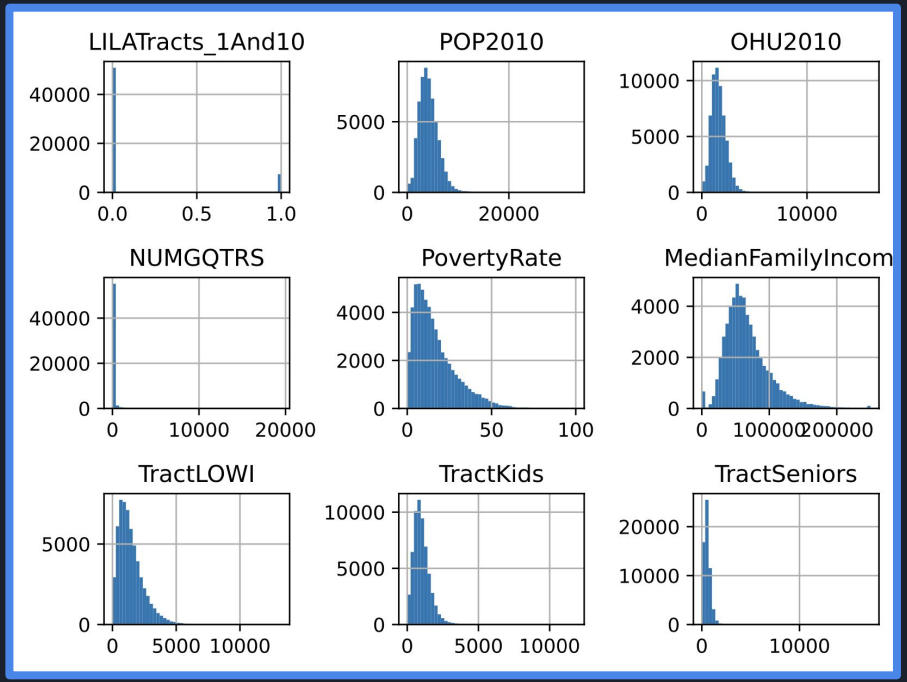
Goals

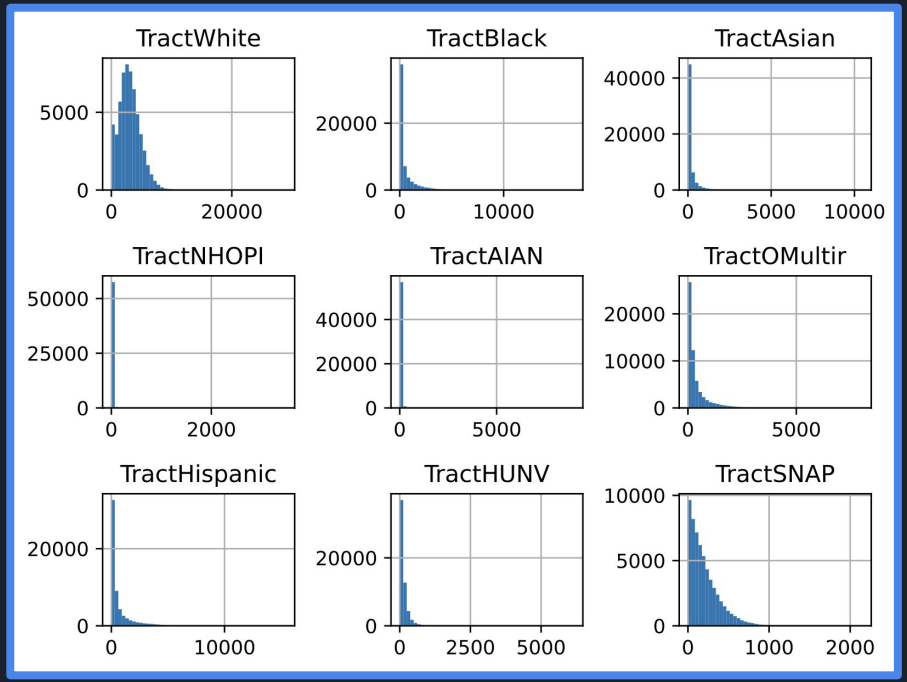
- This project seeks to identify demographic risk factors for low access tracts
- By building a simple classifier that uses commonly available demographic data to identify low access populations
- By flagging/filtering communities based on a handful of risk factors, governments at all levels, from federal, state, to local, can monitor and perhaps even intervene in these communities if needed.

Data

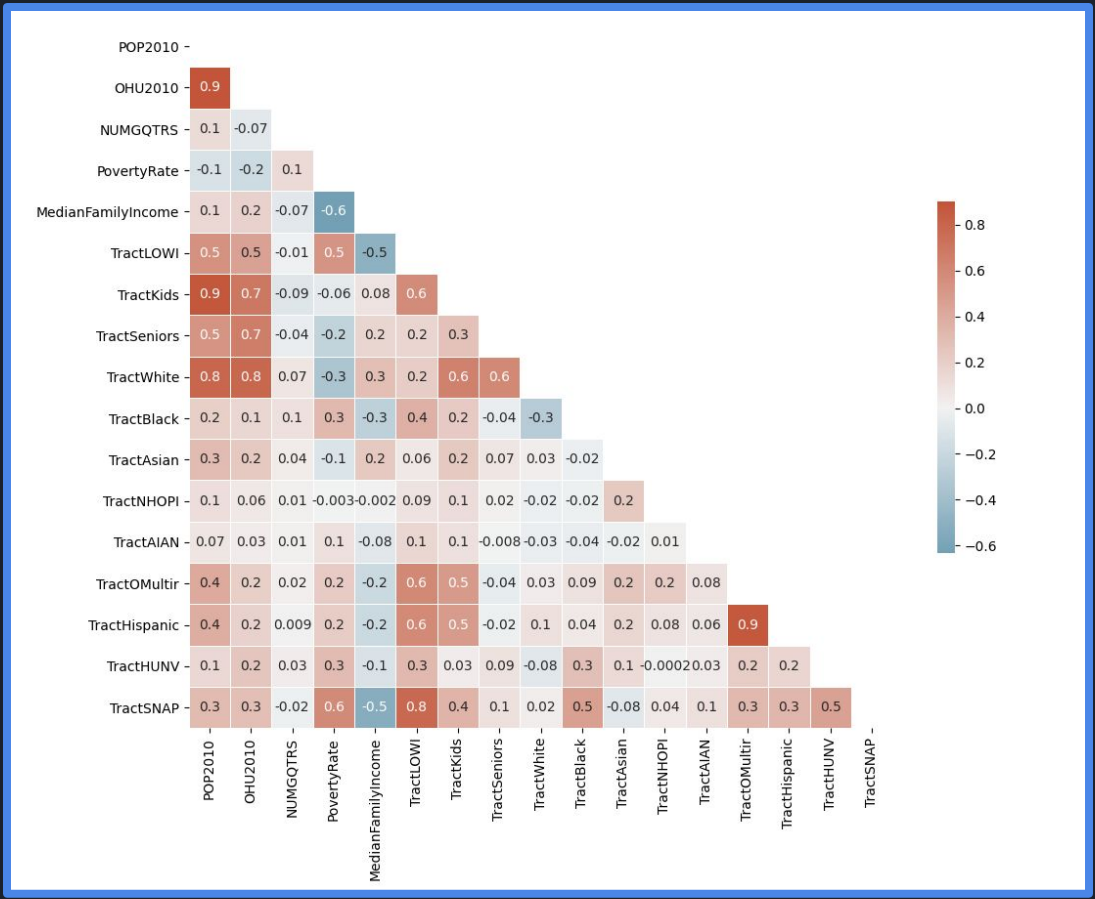
- Variables, our “x”s
  - POP2010
  - OHU2010
  - NUMGQTRS
  - Poverty Rate
  - Median Family Income
  - TractLOWI
  - TractKids
  - TractSNAP
  - TractSeniors
  - TractWhite
  - TractBlack
  - TractAsian
  - TractNHOPI
  - TractAIAN
  - TractOMultir
  - TractHispanic
  - TractHUNV
- Supervisor, our “Y”
  - LILATracts_1And10

EDA: First Impressions

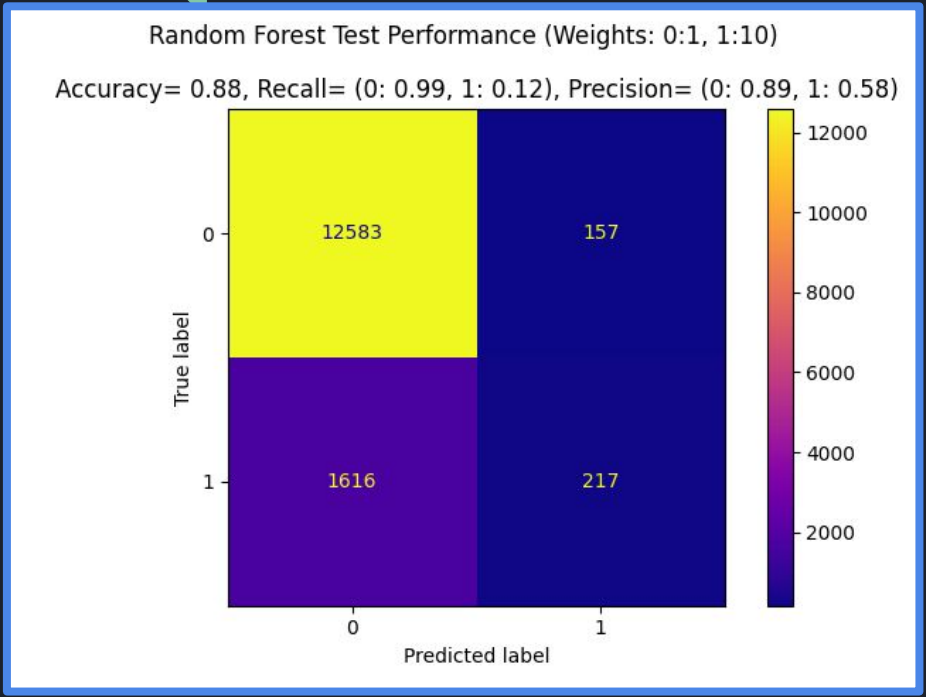


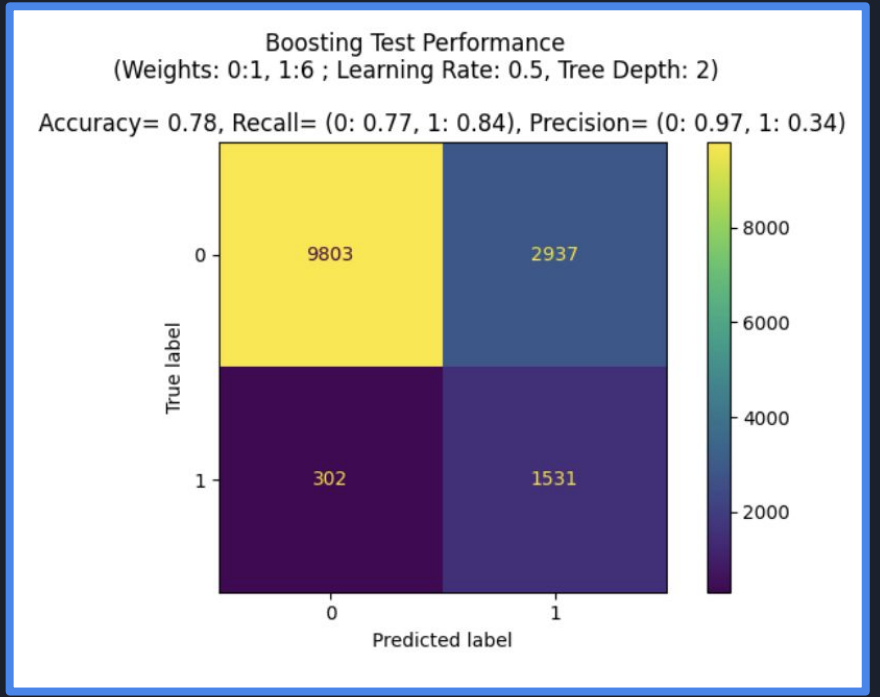


EDA: Correlation Structure

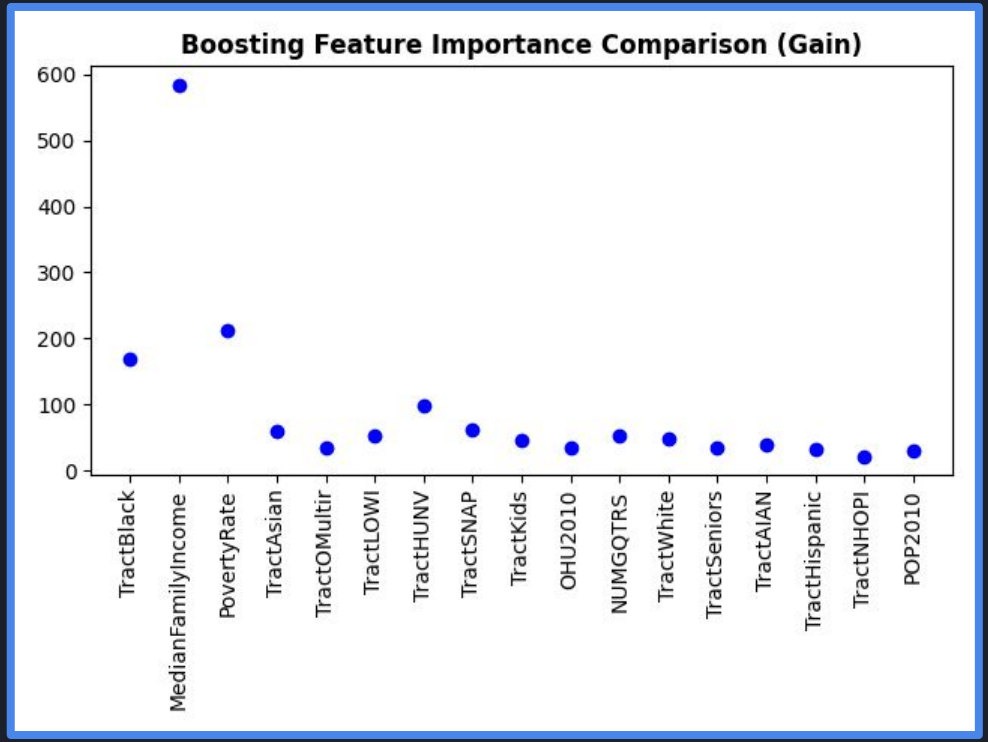


Modeling: Test Set Performance





Modeling: Feature Importance



Conclusions

- 3 main features: *Median Family Income, Poverty Rate, Count of African American population.*
- Model captured 84% of the classes of interest, but you can only trust a positive prediction (i.e. classifying a low access tract) about 34% of the time.
- Socioeconomics and demographics may not be enough to create a comprehensive model of food insecurity.
- Parametric Models may be worth considering, would better lend itself to inference

Thank You So Much!!!

Loved this course, doing a project always forces me to learn so much, thanks for allowing me to learn in such a relaxed format