Key Factors Driving US Food Insecurity

John Kline

3/03/20

Framework & Data Sources

BUSINESS CASE

The Doritos Foundation (nonprofit arm of the snack foods brand) has decided it wants to give back, and has chosen food insecurity in the US as its flagship issue to focus on: what key issues should it invest in?

- *Prioritize top 3 factors out of a list of 67+ potential contributors*
- *Focus on factors driving high levels of food insecurity*
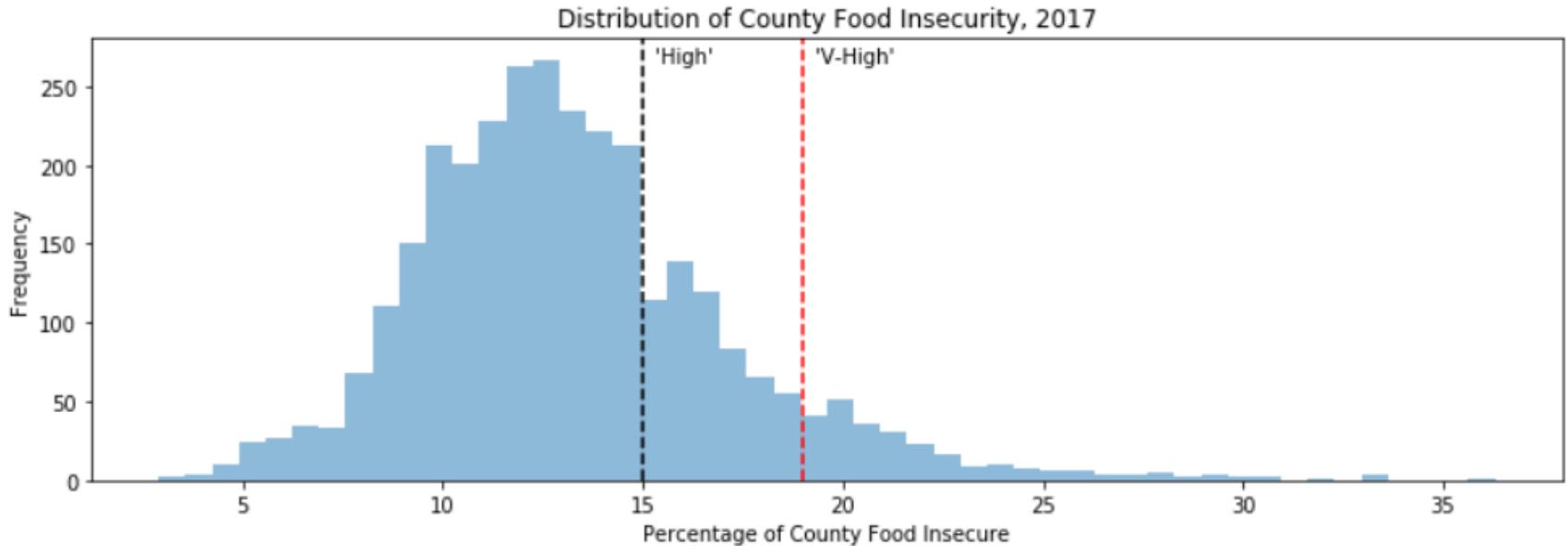
METHODS

Regression
Decision Tree / Tree Ensemble
Gradient Boosted Decision Tree

DATA SOURCES

- **2019 Current Population Survey Data**
  - Reported by US Census
  - Reported by nonprofit FeedingAmerica
  - County-level data reported in 2019, reflecting 2014-2017 data collection

Outcome Variable: Food Insecurity

Distribution of County Food Insecurity, 2017

Observations and Classification

- **Food insecurity varies substantially across US counties**
    - ~25% of counties have <8% food insecurity or >20% food insecurity
    - Long tail of food insecurity – small number of counties are disproportionately impacted
- **Classification:** We will classify counties into three buckets for this analysis:
    - 'V-High' Insecurity: ~ top 10% of most food insecure counties
    - 'High' Insecurity: next ~20% of most food insecure counties
    - 'Medium-Low' Insecurity: remainder of counties
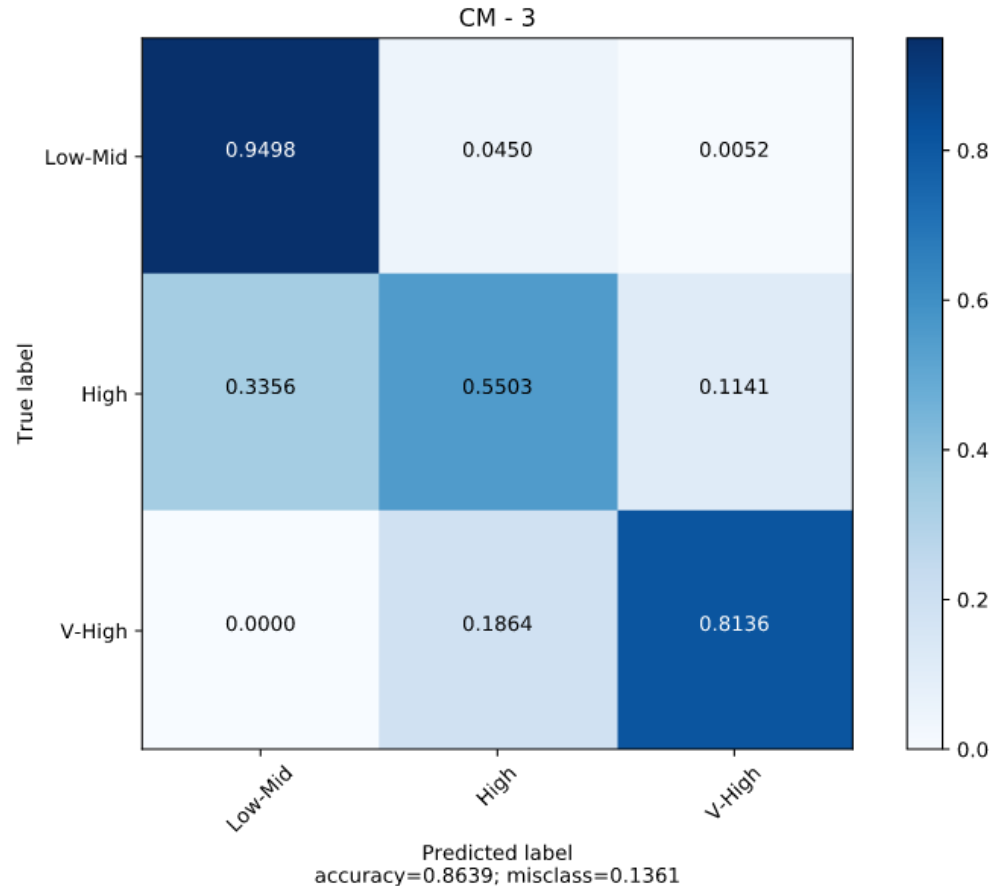
Classification Analysis Results

Classification Output

Most successful classification analysis:

- **86.4% accurate at classifying insecurity level**

- **Relied on 9 of 67 potential factors**

Key Analytic Factors:
- Poverty rate
- Race
- Milk:Soda Price Ratio
- Low access to food
- Number of <18yr old residents

CM - 3

	Low-Mid	High	V-High
Low-Mid	0.9498	0.0450	0.0052
High	0.3356	0.5503	0.1141
V-High	0.0000	0.1864	0.8136

Predicted label
accuracy=0.8639; misclass=0.1361

Model: XGBoost gradient boosted decision tree classifier

Classification Analysis Detail

Prioritized Features

Most important analysis features:

- **Poverty Rate:** county rate of poverty was the single most important variable

- **Race:** proportion of black or Hispanic residents were the 2nd / 4th most important variables

- **Milk Price:** ratio of milk to soda price in the county was the 3rd most important variable
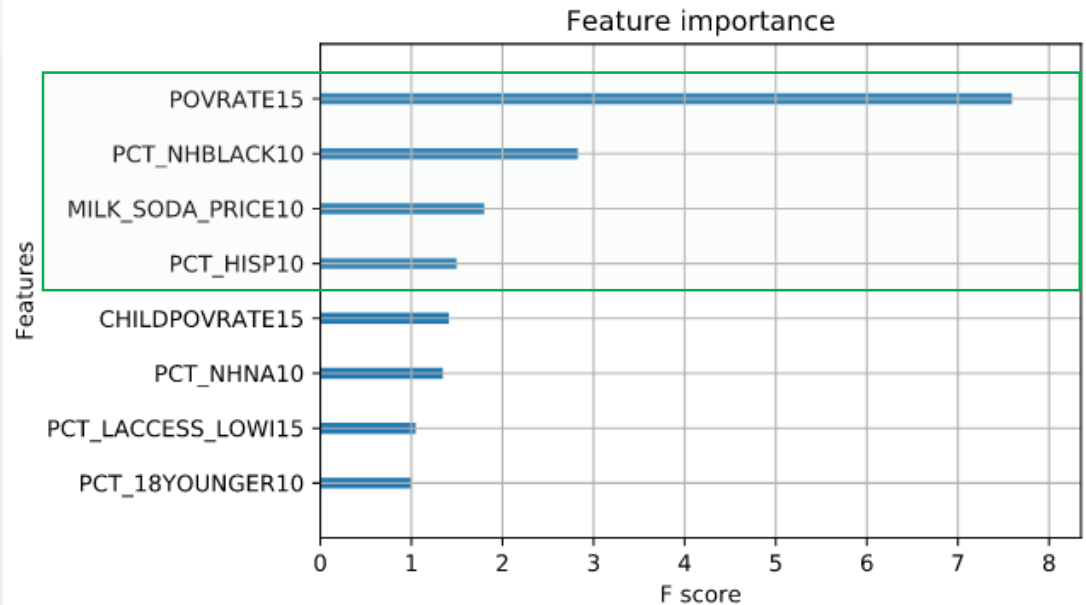
Model: XGBoost gradient boosted decision tree classifier, importance based on average gain

Client Recommendation Summary

Recommendations

1. **Economic Stimulus**: In this analysis, the most important hurdle is reducing the number of residents in a county under the poverty level. Approaches such as the following may be appropriate
     - Earned-income tax credit (EITC),
     - Job training programs
     - Increases in the minimum wage

2. **Perishable Food Pricing:** The metric of choice for describing food inaccessibility is the actual physical lack of grocery or specialty stores in a region (i.e. 'Food Deserts'), however this analysis suggests that actual price is a more important indicator than physical dispersion of stores.

3. **Race:** The importance of county racial makeup indicates the need for the Foundation to invest in additional initiatives that support food security within black and Hispanic communities in order to bridge the gap in food insecurity that these groups face relative to others

Next Analytic Steps

1. **Heatmap of food insecurity by region to look for geographical relationships between counties**

2. **Find other correlated predictors – e.g. transportation access**
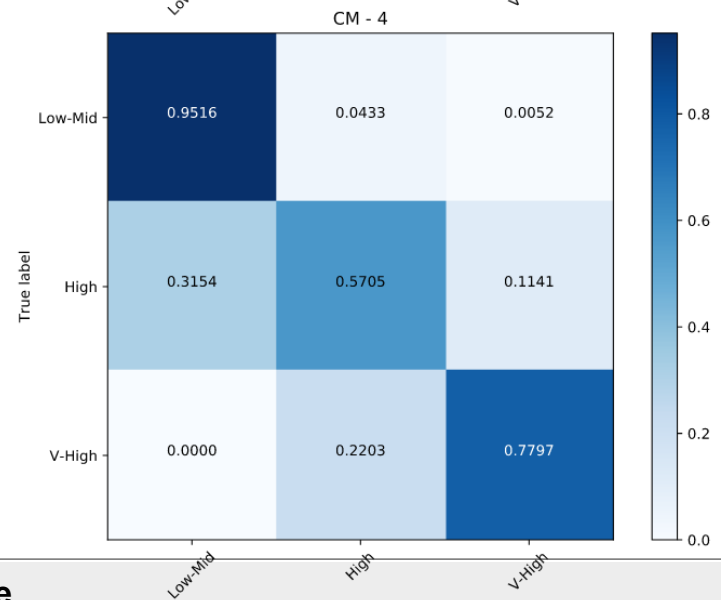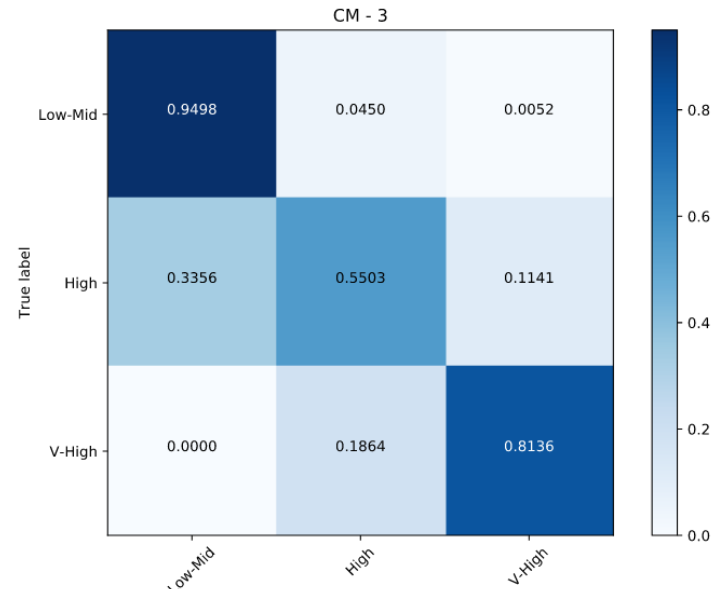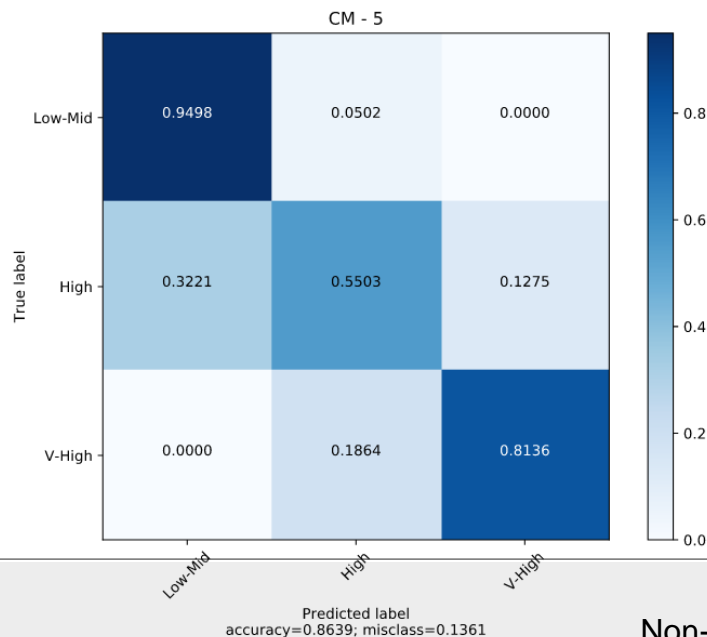
Discussion & Questions

Thanks!

Appendix

Alternate Column Choices

- CM – 3: Core 9 factors
- CM – 4: Added 4 factors covering SNAP and WIC usage
  - Slight increase in accuracy
  - Traded V-High for High accuracy at a steep tradeoff
- CM – 5: Added 4 factors covering access to food locations (not more accurate)

CM - 3

CM - 5

accuracy=0.8639; misclass=0.1361

CM - 4

accuracy=0.8664; misclass=0.1336