

US Food Insecurity Analysis

Alec Edgecliffe-Johnson, Ryan McDonald, Andrew Roberts, Ira Seidman

ARIA Consulting Group

Agenda

- Statement of Purpose
- EDA & Methods
- Modeling
- Results via Streamlit
- Conclusions & Recommendations

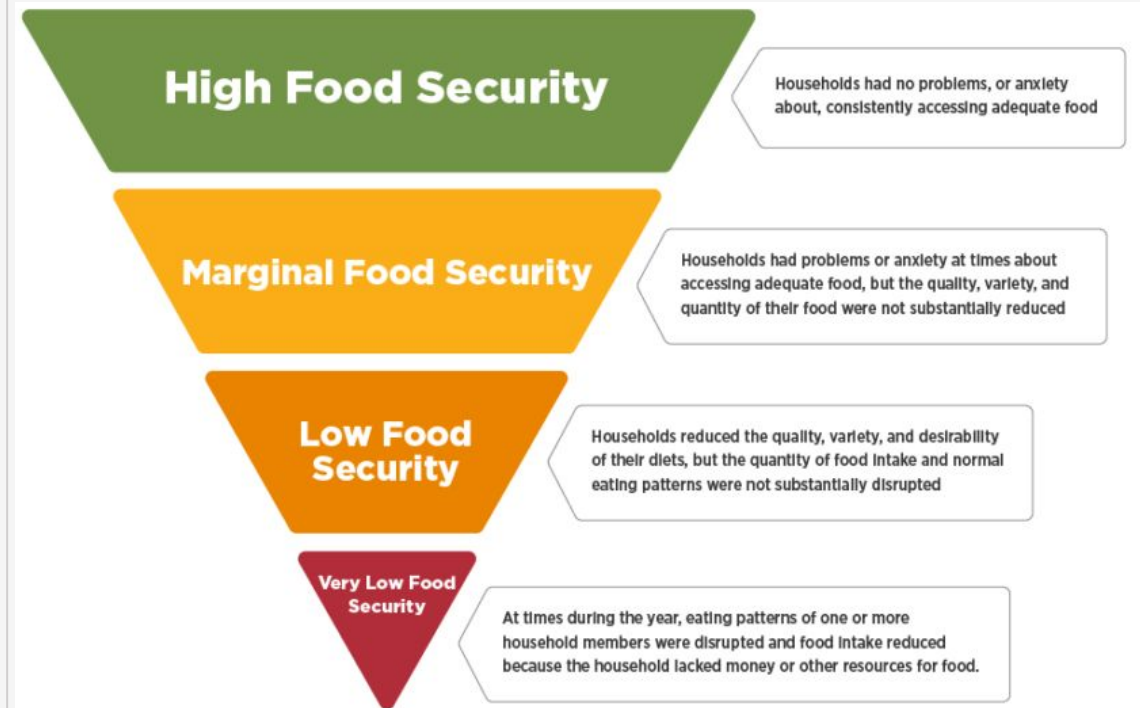
Purpose

Food Insecurity (FI) – a lack of consistent access to enough food for an active, healthy life [1]

- 1 in 9 Americans in 2018

We are tasked with several objectives:

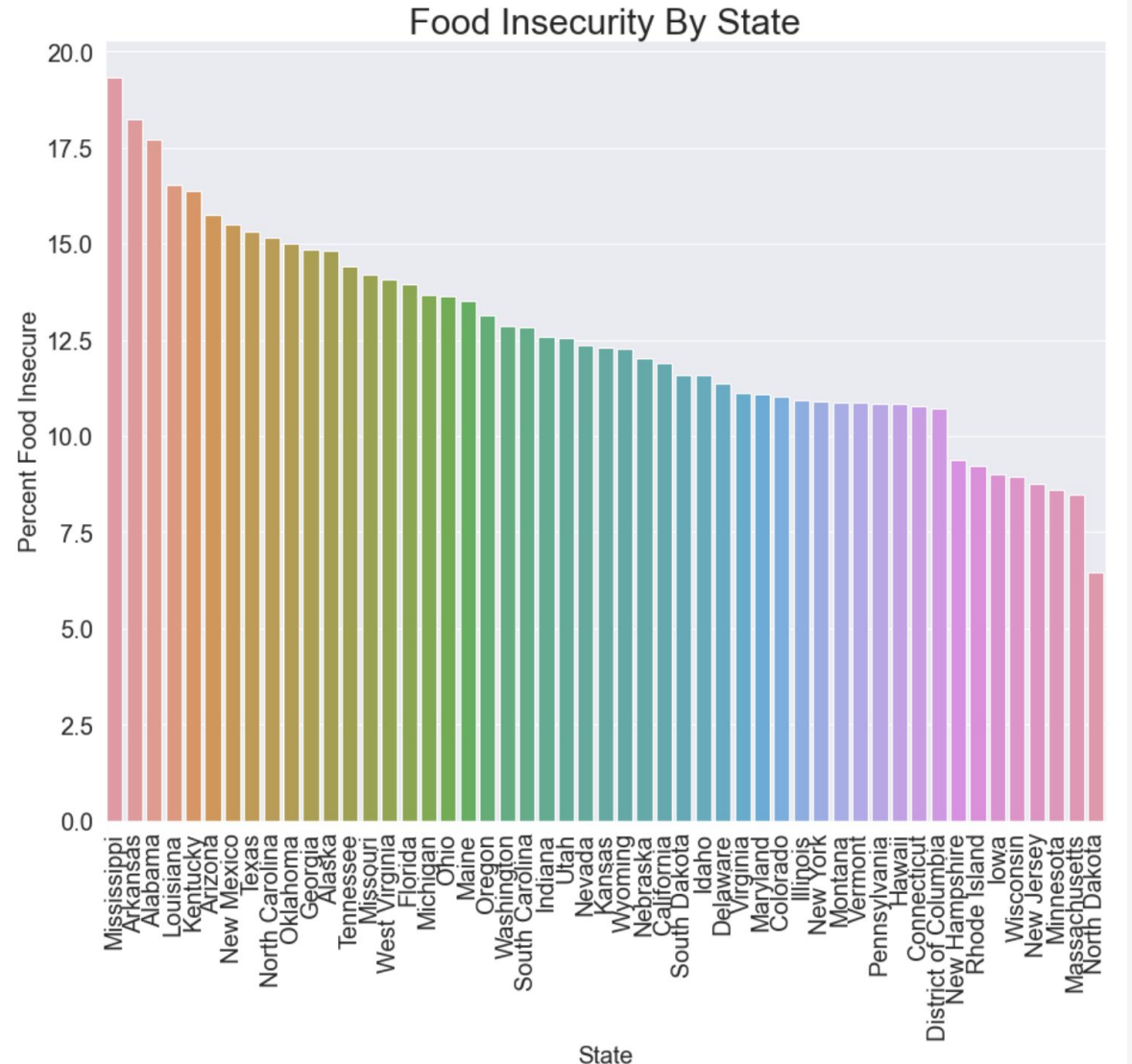
- Develop a model for predicting FI at the county level
- Identify greatest contributing factors
- Create time series models for predicting where FI will persist/worsen



Source: Adapted from the USDA Economic Research Service.

Data Exploration

- Data from Feeding America, CDC, County Health Rankings, Census.gov
 - 60 features and 3,140 counties
 - **Features** include disability rates, unemployment, education, income
 - **Target** is percent of population considered food insecure
- Highest FI – Deep South, parts of west
- Lowest FI – Northeast, upper Midwest



Data Exploration Cont.

- Childhood food insecurity is higher than adult food insecurity no matter how the data is sliced
- Strongest economic predictors of food insecurity – life expectancy, diabetes, median income, income per capita, poverty, and free or reduced lunch household rates (racial indicators were less correlated)



Modeling Overview

- Six different methods of imputation – mean, median mode, knn, linear regression, and random forest
- Random forest lead to the best testing scores, so all scores reported are with the rf imputation. Only a .05-.1 improvement in RMSE of food insecurity rate over mean and median
- Goal – build a model that best predicted food insecurity rates using 59 features

percent_physically_inactive	0.52
percent_low_birthweight	0.56
percent_of_adults_with_less_than_high_school_diploma	0.6
percent_single_parent_households_CHR	0.62
percent_no_highschool_diploma	0.62
percent_unemployed_CDC	0.65
percent_smokers	0.65
percent_enrolled_in_free_or_reduced_lunch	0.68
percent_disabled	0.68
average_number_of_mentally_unhealthy_days	0.74
percentile_rank_social_vulnerability	0.76
pct_overall_pov_19	0.82
percent_fair_or_poor_health	0.82
percent_below_poverty	0.83
percent_children_in_poverty	0.85
ch_fi_rate_18	0.92
fi_rate_18	1

fi_rate_18

Modeling Overview Cont.

- Several Regression Models were chosen to investigate.
- Once our Production Model was chosen, the team moved onto forecasting
 - Univariate and Multivariate Time Series Models were developed to predict Food Insecurity (and poverty) into the future.
 - More on those results later!

# Models Developed	9
Transformers Utilized	Standard Scalar, Polynomial Features, PCA
Regressors Utilized	KNN, Decision Trees, Linear Regression , SVM, Tensor Flow,
Ensemble Methods Utilized	RandomForest, AdaBoost
Workflow Automation Methods	Pipeline
Hyperparameter Tuning Technique	GridSearchCV

Production Model Performance / Evaluation

So, how well does it work?

- Production Model Testing $R^2 = 93.5\%$. **RMSE = 0.93**
 - StandardScaler
 - Random Forest Imputation of nulls
 - All Numerical Features
 - Linear Regression (Default Parameters)
- **Models 2-9 R^2 Range = 78%-93%. RMSE = 1.69 - 0.99**
 - Support Vector Machines and Principal Component Analysis (w/ Ln Reg) both scored as well as our production model
 - Runtimes significantly longer
 - Not necessarily as well interpreted (SVM)

Rating	Production Model Feature	Coefficient
1	% Disabled	0.845
2	% Children in Poverty	0.564
3	% Fair or Poor Health	0.368
4	% Severe Housing Cost Burden	0.330
5	% Below Poverty	0.230
6	% Single Parent Households	0.175
7	% Severe Housing Problems	0.175
8	# Households	0.175
9	Ave # Mentally Unhealthy Days	0.148
10	% Unemployed	0.147

Conclusions, Recommendations and Further Steps



Conclusions: Percent Disabled, Percent Children in Poverty and Percent Fair or Poor health were the three largest predictors. Our forecasts show persistent high levels of Food Insecurity in a number of states in the Southeast as well as New Mexico and Utah.



Recommendations: Federal and state coordination to investigate and ameliorate the factors that we found were most predictive of Food Insecurity, particularly in areas that are forecasts indicated were at a higher risk for increased rates.



Further steps: Further develop time series forecasts with additional variables and at county levels. Identify better and more granular factors (eg. family size for those below poverty) to develop more specific recommendations.