Jonathan Broadbridge

Dr. Scott

BAN 530

10/25/2021

Project Scope: Covid Food Security Data Set

**Introduction:** For the capstone course project we will be examining a Covid-19 Food insecurity dataset containing data about food insecurity estimates along with corresponding variables that range from demographics to employment information to cognitive ability. In this project I will be analyzing a few select variables and how they can be used to predict food insecurity. I will likely be performing two sets of analysis, one that can predict the number of individuals that are food insecure and another that treats the data as a binary variable which can indicate whether a person with certain details would likely be food secure or food insecure. The data as a whole is a time series and using proper validation methodology will mean that we are testing our models on the most recent data.

**Variables of Interest:** The key variables of interest in this analysis will be the response variable of food insecurity. This variable will be analyzed both to predict a numeric value for the number of individuals/households that are food insecure as well as a categorical model which will simply predict if an individual or household is food secure or food insecure. The variables being used to create these models will be state, age, education, race, sex, marital status and children status. All of these variables will be treated as categorical.

**Potential Issues/Methods:** The data being used in this analysis is an aggregation of food insecurity data with 36 weeks of data being reported. There are a large number of NA values in this dataset which will introduce some data cleansing problems but they should be fairly simple to solve with basic filtering. Another potential issue is that the response variable data is a series of estimates based on a population sample. There may be some variances within the data due to data collection errors that may need to be accounted for. For example I may see issues with data from smaller states not being consistent or weekly figures with high numbers of No response may be similarly skewed. An Exploratory data analysis and model evaluation should tell if these concerns are going to be issues.

**Conclusion:** This analysis of the covid 19 food insecurity dataset should hopefully reveal some insights into what may be driving food insecurity and provide information about what variables are most influential in food security for American households. By using modern predictive analytics methods and technology such as python I hope to bring some interesting insights into this recent once in a lifetime event that is covid.