Zach Meurer and Kailen Richards

Mask Use July 2020 Predictive & Descriptive Analysis

Publicly Available Datasets: [Mask Use by County](https://github.com/nytimes/covid-19-data/tree/master/mask-use), [Covid Data 2020](https://github.com/nytimes/covid-19-data/blob/master/us-counties-2020.csv)

Time Period: July 2020. Survey responses for mask use comes from July 2, 2020 - July 14, 2020

Independent variable: Mask use by county

Question for Mask Use by County Dataset:

*How often do you wear a mask in public when you expect to be within six feet of another person?*

Both Approaches:

Filter through data and keep the parameters we need within the time period (date, county, mask use, number of cases in county, deaths). We will attempt to turn the responses from the mask use dataset into a single value to represent the overall mask use in each county. We plan to train the dataset using a weighted average of mask use per county and the number of cases in that county to predict whether given the case numbers, how regularly people wear masks. We will exclude counties that start with the letter ‘C’ from the training data, and to simulate randomness for the test data and avoid bias.

Approach 1: We will train our AI with K-Nearest Neighbors. We will vary which value of K we use to find which value works best

Approach 2: We will train our AI with Decision Trees. We will vary which starting letter of states we use to simulate randomness, as to create cross-validation.

Data used: average number of cases, deaths, state location

Label: mask use

Columns in Covid Data 2020: Date, county, state, fips, cases, deaths

Columns in Mask Use by County: COUNTYFP, NEVER, RARELY, SOMETIMES, FREQUENTLY, ALWAYS