Classification
MVP
May 11, 2021

Project Description

I will be using research from the Pew Research Center called 2020 PEW Research Center's American Trends Panel. They conducted a survey between April 29th and May 5, 2020 primarily focused on climate change and COVID though some general information such as income is included as well. I would like to predict the answer to the first question of the survey which is "do you think you will be better or worse a year from now", based off answers to the other questions. 82 said better, 17% worse, and 1% no answer.

MVP

As mentioned above there is a definite majority class for this project in that a high percentage of respondents felt they would be better of in a year. The primary focus of the project is being able to predict those who feel they will be worse off in a year. As a result, I am trying to maximize the metric of recall and limit the number of false negatives. Based on an initial Knn model (K = 1 based on GridSearchCV), recall is currently equal to .16933. Plenty of room for improvement! I will be spending the remaining time trying to improve this measure. Though I expect to experiment with a number of techniques, I'm expecting oversampling the positive class to be crucial to project success as the current models are prioritizing the negative class.