Classification
Project Proposal
May 14, 2021

Abstract

I used 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 attempted 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.

Data

Data has been gathered from the Pew Research Center for nearly 11,000 individuals and 152 questions (variables). This data was condensed to develop the final model which used one target variable and 5 features.

Tools

Exploratory data analysis – SPSS (file was a .sav) & Excel
Classification algorithms- KNN, Logistic Regression, Decision Tree, Random Forest, Naïve Bayes.
Also – tried oversampling and SMOTE
Visualizations – Matplotlib and Excel
Presentation – PowerPoint

Communication

I will be presenting results to the classification cohort on Friday, May 16, 2021. A PDF of the presentation is also included on my personal GitHub repository.