DAT 530 Advanced Statistical Methods

Project Review: Week 5

Reviewer: Brandon Hosley

September 8, 2020

Project Title:

Bluebonnet Data

Author(s):

David Zask, Chase Rendall, and Devin Fagan

Source:

NYC Data Academy

The Problem the Author(s) is Trying to Solve in the Project:

The authors intend to provide recommendations to Bluebonnet, an organization dedicated to bring data scientist to work with Democratic party campaigns in smaller elections. The authors recommend a model for determining the contestability of a political district, which may provide a better return on campaign investments.

Machine Learning (ML) Algorithm(s) used:

- **Principle Component Analysis**
- K-Means Clustering

A Brief Description of One of the ML Algorithms used:

Principle Component Analysis is the process of determining candidate features that may be most effective in generating predictive models. This process helps dimensionality reduction, improving the overall efficiency of a model by eliminating less effective, less correlative features. Using this method to add preference to features with a higher absolute covariance, the researcher may select the best predictors in a number that reduced computation wasted on other factors, and preventing overfitting. Potentially, this may also allow future researchers be more parsimonious in their data collection efforts.

Metrics Used to Evaluate the ML Algorithms:

The authors offer the results of a confusion matrix. Additionally, they suggest that uncontested districts that were predicted to vote in contrast to their actual results may be strong options for recruiting new candidates.