**Analysis**

To predict non-voters I used the electorate research survey. For this analysis I considered voters all of those said they voted in the 2016 general election. For demographic variables I used gender, age, education, ethnicity, and religious commitment. Primary independent variables were political ideology (very liberal – very conservative), stance on political issues (e.g. immigration), perceptions of fairness of the electoral system, trust and helpfulness of people, perceived national wealth, and international respect for the U.S. All of these are based off the 2016 responses.

One problem in developing models to predict nonvoters is they make up a small percentage of the population. In this sample, less than 5% did not vote in the 2016 election. Such unbalanced data can make it difficult for supervised learning techniques to capture variations between groups. To overcome this, I implimented a synthetic minority oversampling technique (SMOTE). SMOTE generates new samples of observations of the minority class, in this case nonvoters, by inserting values similar to those of the observed data. I used the *SMOTENC* function in the *imblearn* packages in python to create this oversampled data set.

Using the variables mentioned previously I trained a logistic model as a baseline. All accuracy scores will be based on its application to the original observed data given the baes rate probability being 95%. The logistic regression was able to predict non-voters 38% of the time and voters 80% of the time. Area under the receiving operation characteristics curve was .59.

Next I trained a ensemble machine learning model with the following algorithms: logit, decision tree, random forests, support vector machine, gradient boosted trees, and ADA boosted trees. Using the same features, this ensemble model was able to predict non-voters 80% of the time while also predicting voters 99.56% accurately. The area under the curve for this model improved drastically to .90.

A quick look into important features finds perceptions of the election system being fair, education, and how much they perceive their values are similar to others.

* **You will need to do some dimension reduction (EGA, PCA, and/or CFA)**