1 Abstract

The following score relies on a conception of persuasion that may be defined as an absolute difference between the most-supportive 'favorability characteristic' of a voter and their least-supportive 'favorability characteristic' for any measured candidate or ballot position, generally referred to as q through this document

An ensemble learning classifier is used to correspond this differencekj and basic demographic characteristics. This classification method is then expanded to the voterfile to rank voters based on their demography. Using a weighted sum of probabilities for each class, a normalized score is developed. Thus allowing campaigns to prioritize their voter contact along a notion of undecidability and partisan difference of the voter– and thus, perhaps, persuadability.

The ranking of the voters exists on a scale of 0 to 1 such that 1 is the highest rank a voter can have in terms of persuadability. Sampling 100,00 voters from the voterfile, they distribute themselves along this rank as:

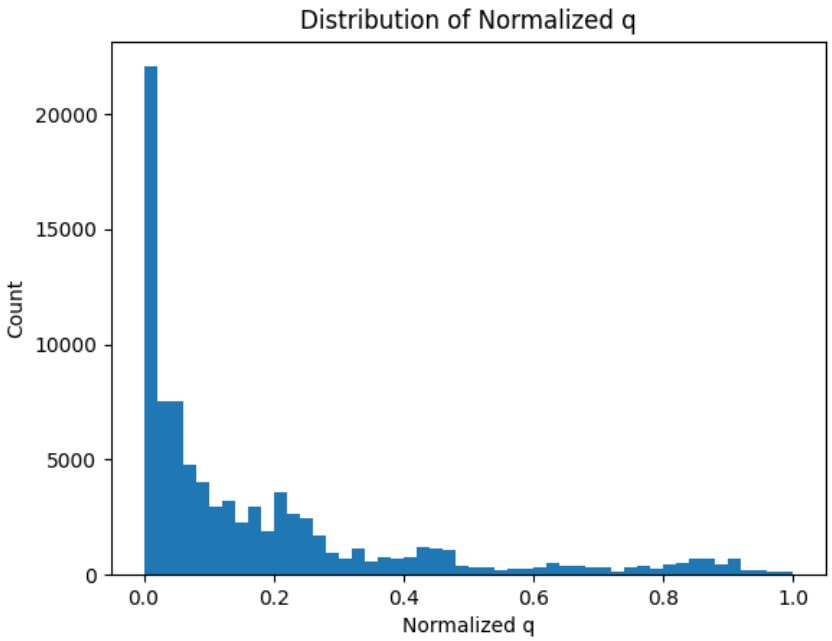

Figure 1:

The average rank may be taken for each voter subset as:

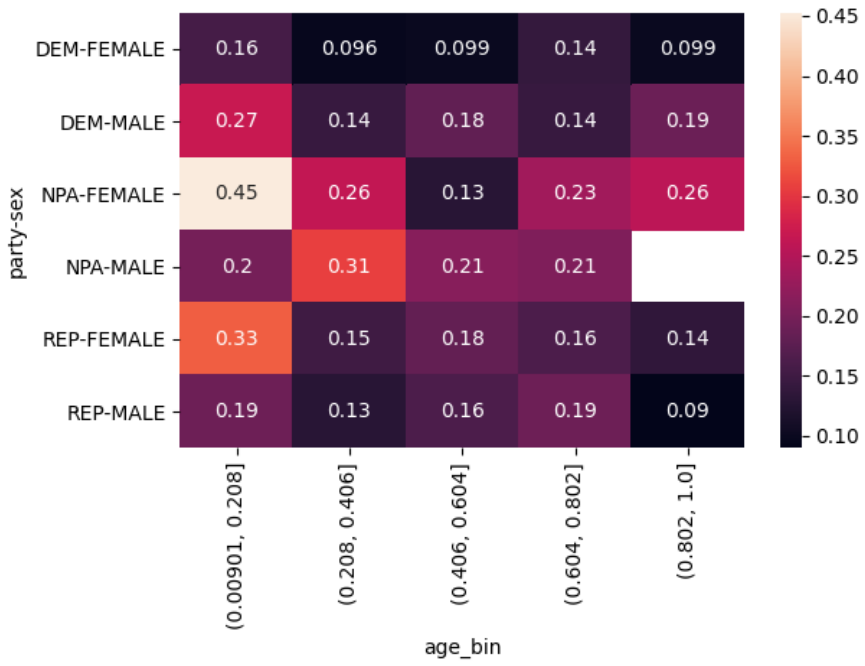

Figure 2: $\bar{q}$ for $party$ and sex by age where age is represented in its normalized form.

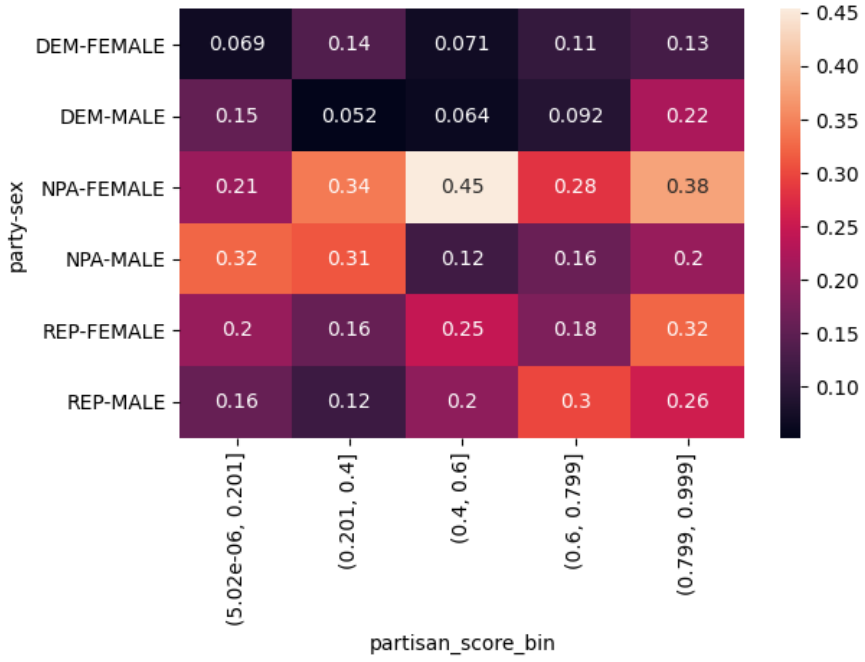

Figure 3: $\bar{q}$ for $party$ and sex by $partisan\ score$ where $partisan\ score$ is represented in its normalized form.

2 Theory

Generally, political targeting is thought and taught around three dimensions– a support rank, a turnout rank, and a description of voting behavior. From this, a concept of persuasion is derived as some midpoint between these dimensions. Observe Figure 1 where the 2008 Obama campaign canvassing concentration (green as most concentrated) is depicted along the support (x-axis) and turnout (y-axis) dimensions for pre-August canvassing on left and post-August Canvassing on right.

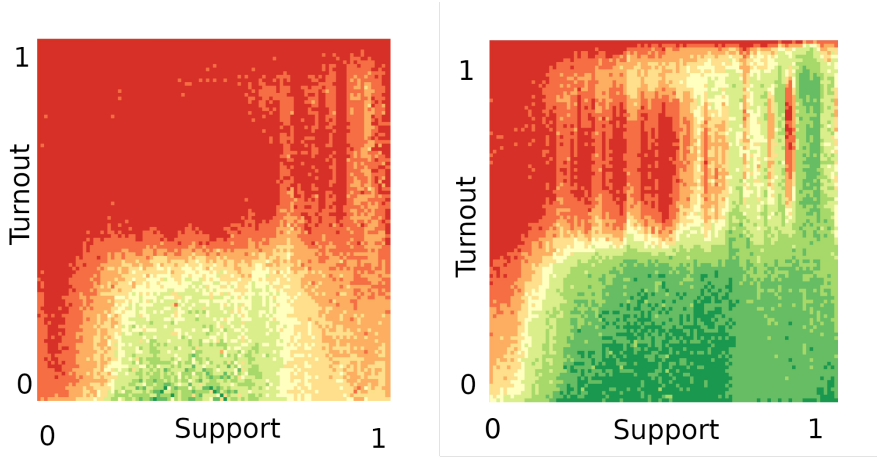

Figure 4:

These dynamics have proven themselves over the last decade and continues to be improved upon. As such this memo presents instead a concept of persuasion that may be described as such:

$$q_v = \max f_v - \min f_v,$$

such that f_v represents the favorability of a voter on a scale of 1 thru 5. With this, some dynamic of indecision and uncertainty are at the root of our notion of persuasion[1].

3 Application

Due to variable difference and standardization of the 2018 polling files of the HDCC, the HDCC Persuasion Score calculates p by using the absolute difference between 2018 Generic Ballot Favorability and 2018 Governor Favorability. q then has the dependent variables age, $party$, sex, and $2018\ partisan\ score$.
Notably, as this point, p is necessarily a categorical variable as a result of the f_v being categorical in nature. As such, a random forest classifier (RFC) is then deployed to correspond the independent and dependent variables. The RFC is cross-validated through a grid-search along the $n-estimators$ and $max-depth$ parameters. The RFC out of the box returns a 76% mean classification accuracy; after cross-validation, the RFC returns a 82% accuracy.

[1]since there are some liberties taken in the algorithm please see *Appendix A* for pseudocode

A transformation is required from the predicted class, p, to a continuous ranking of p. . That function may be described as:

$$q = \sum_{a}^{a+i} P(p_a) * i,$$

where, $P(p_a)$ represents the probability, $P()$ a voter is class a[2]. p is then multiplied by its rank and cumulatively summed.

eg, given five classes of favorability[3], each record would then have 5 probabilities associated with it[4]. The 'fully decided' class where the *difference* between the $\max favorability$ and $\min favorability$ is 0 is dropped from the function. The remaining four classes may then be compiled as:

$$q = (P(1) * 1) + (P(2) * 2) + (P(3) * 3) + (P(4) * 4)),$$

such that now q is a weighted, cumulative sum of all its probabilities.
A min/max scaling function is applied to q prior to model input and other continuous or discrete variables throughout the process described here. Categorical variables were transformed to binary representations of each class with in the category following one-hot encoding standards.

4 Additional Developments

new initial measure
new sigma sum
more param optimiz
auto model select
training data with a primary key of sorts. available polling data was not keyed in a meaningful way.

[2]this assumes an index+1 enumerated list of a. This is a significant assumption.
[3]*ie*, strong support or 1, lean support or 2, undecided or 3, lean oppose or 4, strong oppose or 5
[4]*ie*, P(*diff* of 0), P(*diff* of 1), P(*diff* of 2), P(*diff* of 3), P(*diff* of 4)

5 Appendix A

Pseudo-code for persuasion equation:

Input: matrix $F = \{P(f_{ij}), P(f_{ij})...P(f_{ij})\}$
Output: f_i where f_i is the derived vector
if f_{ij} *= none* **then**
  return none
else if $all f_j$ *for i = 0* **then**
  return 4
else if then
  return $\max f_j - \min f_j$