Data Science for Campaigns

Peter Casey, Ph.D.

How campaigns use data

- Plan campaign strategy
- Track metrics
- Evaluate strategies
- Inform and influence voters and policymakers
- Inform voter outreach decisions (predictive analytics)

The Voter File

- Every state is required by federal law to maintain a record of all registered voters
  - Generally includes:
    - Name
    - Address
    - Voting district
    - Vote history
  - May also include
    - Sex
    - Age (or date/year of birth)
    - Race
    - Party affiliation

- Several organizations collect state and county voter files to construct a national voter file
  - One big challenge: Matching voters across state lines

Other important data sources

- Survey data
  - Individual responses to questions about (political) attitudes and behaviors
- Census data
  - Aggregate demographic data
- Commercial data
  - Consumer data aggregated and sold by companies like Experian and Haystaq
- Canvassing data
  - Collected during voter outreach
- Phone dispositions
  - Collected and sold by companies or during voter outreachS

Predictive Models

- Civic behavior, like voting, donating activism
- Political support for parties, candidates, issues, etc.
- Political identity, like partisanship, ideology
- Demographics, like age, race, ethnicity, and religion
- Life events, like marital status, education, household composition

Vote Propensity

- Voter file includes whether a person voted or not (but not *how* they voted) in each election, including general, primary, special, etc.
- Use a person's vote history (linked across states and over time) to predict their likelihood of voting in a future election
- However, we're always using *past elections* to predict future elections
- Challenging: Who says 2020 will be like 2016? Was 2016 like 2012 or 2008? Was 2018 like 2014 or 2010?
  - How would you deal with something like that?

Support scores

- Only some voter files include party affiliation, and none include a person's vote choice
- Even party affiliation is not a perfect predictor of vote choice: what about independents, unaffiliated voters, third-party affiliates, and vote-switchers?
- Collect survey data about people political preferences and combine with voter file and other data to predict vote choice and other political support across the voter file
- Can be used to predict support for candidates, parties, or issues
- Major challenges:
  - Sufficient data to predict political positions of small minority groups
  - Lack of party affiliation data

Predicting demographics

- Why predict demographics?
  - Demographics like race, religion, education, gender, and marital status a major predictors of voting behavior and political support
  - Not available on all voter files
- Like support score, use survey data combined with voter file and Census data to predict people's demographics
- Challenges:
  - Difficult to predict
  - Noise and interpretation
    - Users often don't understand the demographic models are predictions and subject to error
    - Users can "see" when the model is wrong
    - How would you deal with this?

Contact Models

- Besides knowing if a person is likely to vote and who they support, campaigns may want to know if they can actually *reach* a voter
- Phone and canvassing models can be used to predict how likely a voter is to be contactable by phone or by door-knocking
- Phone and canvassing disposition data is combined with voter file and other sources to predict if a person is likely to answer a phone and be the right person

Combining Models

- Campaigns combine models to create lists of voters for outreach
  - Voters with a high vote propensity but lower support are contacted for persuasion campaigns
  - Voters with high support scores but lower vote propensity are contacted for turnout campaigns
  - Contact models are used to narrow lists to those voters who are most likely to respond to outreach

Issues in Data Science for Campaigns

Effectiveness of models

- Vote Propensity models are not the same as Mobilization models
- Support models are not the same as Persuasion models
- A person's likelihood of voting or supporting a candidate is not the same as their likelihood of doing that because a campaign contacted them
- To build Mobilization or Persuasion models, we need data from randomized controlled trials, which is costly and difficult to collect
- Need further research on the value of Vote Propensity and Support models for identifying voters campaigns can mobilize and persuade

Bias in machine learning

- Extensive research showing that models pick up trends in data that can lead age, racial, and gender bias, among others
- When decisions are informed by models that are biased, it can bias those decisions
- Especially concerning for progressive campaigns:
  - Models may be biased against underrepresented communities
  - Models may be biased against the people campaigns want to mobilize
- Unfortunately there has not been much research on bias in models frequently used by campaigns, like vote propensity
- Need for research and ideas about addressing bias, such as:
  - Leaving out data that correlates with characteristics like race (very difficult)
  - Train-then-mask
  - Training campaign workers to cut lists that are diverse and inclusive

Digital space

- Campaigns are moving further and further into the digital space
- However, digital data and analysis has not been incorporated heavily into campaigning
- Digital data is difficult to match back to the voter file
- Need to think of ways to leverage data and analysis in the digital space apart from the voter file
- Digital outreach alone could be valuable for fundraising, online actions (like petition signing), and even persuasion or mobilization

Privacy

- Voter file data is public, and some of it is open accessible
  - Increasingly, states are adopting laws that allow voters to suppress their voter file information (e.g., Utah)
- Other data, like canvassing data, is shared by campaigns through data exchanges
  - However, voters who talk to canvassers may not know there data is being collected, shared, and used

Questions?