Rachel Orey

8 April 2020

DATS6103 Project Proposal: Reducing Polling Place Wait Time

Topic: Optimizing Resource Allocation at Polling Places and Vote Centers

Summary: For my final project, I plan to build an interactive model using PyQt which will help local election officials manage resources to reduce wait time at polling places. I will use a variety of data inputs and rely on queueing theory to (1) estimate line length giving a precincts current resources and (2) recommend the optimal number of electronic poll books, poll workers, and voting equipment to keep wait time under 30 minutes. The model will rely on a **Decision Tree** to help election officials determine the resources they need based on population characteristics.

Input data:

1. [Demographic data of congressional districts](). I plan on using the LXML package to scrape data from this website. Data retrieved: ethnicity/race by district, people per square mile.
2. *Source data from [this BPC report](). (I work here and am waiting for my colleague to get back to me with the data). This data primarily includes information how demographic characteristics like race, ethnicity, and degree or urbanization impact line length. *This report primarily serves as background research to help inform the model.*
3. [2016 Survey of the Performance of American Elections](). Includes data on % of in-person voters who arrive at polling places at each time of day.
   a. It is the ability to handle lines in the morning that largely determines whether a precinct will have lines later in the day. Queuing theory will be used to determine whether polling places with given resources will be able to handle morning lines.

User Input: Congressional District, State. Optional: Electronic Poll Books (How Many, or Not Being Used), Number of Poll Workers per Precinct, Ballot Type (Paper/Optical Scan, Electronic (page by page), Electronic (full ballot)), Number of Registered Voters in Relevant Precinct.

Model Output: Estimate of peak line length, as well as *optimal* number of **electronic poll books**, **poll workers**, and **voting equipment** (type dependent) that would keep wait time below **30 minutes**.

Estimating model performance: I will compare the estimated line length of a sample of congressional districts against *actual* line length witnessed in 2016 to determine the model's accuracy.

Completion Schedule:

1. Clean all data/import data into Python and convert into workable format. **4/10/2020.**
2. Build decision tree algorithm. **4/11/2020.**
3. Build model interface on PyQt. **4/14/2020.**
4. Run sensitivity analysis and test model's accuracy. **4/20/2020.**