Capstone Two Project Proposal: Predicting Bill Passage

Problem Statement: How can we predict which bills are most likely to make it through each successive step towards passage using only the data available from the time of a bill's introduction.

- Context:

  Most bills introduced in the House of Reps. are never even called up to vote. Further, many bills with similar purposes are proposed each session. A well fit predictive model would allow a legislator or lobbyist to allocate their time to bills most likely to pass.

- Criteria for Success:

  The creation of a model with better than baseline accuracy. As stated above, there are a large number of true negative cases, so priority will be given to recall to allow stakeholders to seek out bills with a higher chance of passage

- Scope of Solution Space:

  The model will be trained on bills introduced in the house during the three latest congresses and tested on the available data from the current congress. It will not consider amendments or resolutions.

- Constraints within Solution Space:

  The large number of true negative cases means that precision and recall will have to be relied upon as evaluation metrics rather than accuracy. Further, the data collected changes in format for each congress with new features being added each time. As a result, we must infer some features and drop others During EDA

- Stakeholders:

  Legislators and Lobbyists will use the model to determine the likelihood of a bills passage at its time of introduction so that they may more effectively allocate their time and resources

- Key Data Sources:

  ProPublica: Congress API (Bill and Member subsets for the 113th, 114th, and 115th house congresses)