Predicting Bill Passage in Pennsylvania

Vidisha Chowdhury, Jack Jacobs, Philippine Kugener, Shun Tomita — `bills1`

Executive Summary

The American Civil Protection Association (ACPA) is an organization that seeks to protect the rights of all individuals in the United States through advocacy efforts against legislation that infringe on civil liberties, challenging such laws in court when they do pass, and ensuring that necessary statutes are in place to protect those liberties. Bills that could infringe on the liberties of one or more groups of citizens are routinely considered before state legislation and when they pass into law, they can affect the well-being of vulnerable populations. Given this context, the ACPA needs to prepare for either advocacy efforts to challenge the passage of a bill or a subsequent legal challenge, in order to effectively block such legislation. However, the organization's resources allow it to work on only 15% of bills introduced in every session.

To proactively protect civil liberties, the organization needs to know which bills to work on before they are passed. Thus, we help the ACPA develop a method to identify the top 15% of bills that are most likely to pass in a single session in a state. As a proof of concept, we develop a Machine Learning (ML) model that predicts passage of bills and produces a list of the top 15% of bills which are most likely to pass in a Pennsylvania session. Using a temporal cross-validation strategy, we train and test a range of models and compare their performance using recall at 15%. We also compare our models to non-ML baseline models to check if the ML methods achieve significant improvement over non-ML techniques. Our best performing model, a random forest classifier, achieves an improvement of about 20 percentage points in recall over a baseline model.

While developing this system, we try to achieve three goals: efficiency, equity and effectiveness. While efficiency is addressed mainly by model performance, we address equity through a bias audit of our models. Models are checked for bias on various dimensions of bills such as bill content and sponsor characteristics. It is found that for most of these attributes, the best performing random forest classifier, is also the least biased. For measuring the effectiveness of our work, we suggest a field trial that measures if the ACPA can successfully block harmful legislation.

Our recommended model choice depends on the specific equity goal the ACPA might want to prioritize. Thus, for every fairness attribute we considered, we suggested a specific model or a group of models. However, we did find that the best performing random forest classifier performs fairly well on most fairness attributes. Our recommendations also include improving the quality of data used to train our models by asking the ACPA to tag bills that are of interest to them so that we can better tailor our model's output. Lastly, we mention some caveats like our model's inability to make predictions about passage very early in the life cycle of a bill and the

fact that they do not predict whether a bill will infringe on civil liberties. This is directly related to the scope for future research that we highlight in our conclusion.

Background and Introduction

The American Civil Protection Association (ACPA) is a national nonprofit organization whose mission is to protect the rights of all individuals in the United States (US). The ACPA accomplishes its mission through the review of bills in order to see if they might infringe on civil liberties, lobbying efforts against such bills while they are still in progress, and by preparing to dispute such bills in court if they do pass. Since the ACPA has limited resources, it can only focus on 15% of bills introduced in each session of the Pennsylvania General Assembly. While there could be several channels via which the ACPA could intervene, we choose to focus on helping the ACPA proactively streamline their efforts at challenging bills after they have passed, ensure equitable outcomes for various sections of the population and maximize the impact of their court challenges. Since our work is more of a proof-of-concept exercise, we focus on one state - Pennsylvania as we are more aware of its legislative background being its residents. Pennsylvania, like most states, is a site of intense professional lobbying activity. This lobbying includes advocacy for issues that threaten the civil rights of people in Pennsylvania, such as lobbying by private juvenile justice facilities for harsher juvenile sentencing. Similar issues are at stake nationwide.

Overall, the goal of this project is to protect the civil liberties of individuals in Pennsylvania by helping the ACPA determine what bills they should start working on much ahead of their actual passage, to either prevent their passage or build legal cases against them, after they have passed. For this proactive intervention to be effective, the first step is that the ACPA should focus on potentially harmful bills that are most likely to pass. To facilitate this, we provide the ACPA with a list of the top 15% of bills that are most likely to pass in a single session.

This problem is important because bills passed by state legislatures in the US can have massive impacts on the lives of individuals in the US. For example, as of October 2021, states had enacted 106 abortion restrictions in that year alone. Such restrictions adversely affect certain sections of the population such as low-income women and are sheer violations of their civil liberties. In such situations, the ACPA actually serves as the voice of those who have been adversely affected and offers a way through which people can oppose such legislations. In this regard, the ACPA's efforts are high-stakes and time-sensitive. Thus, information on whether certain bills are likely to pass, well ahead of actual passage, can be very valuable for the ACPA to effectively prepare for legal challenges and hence protect individual rights.

The potential impact of solving this problem can be understood in terms of efficiency, equity and effectiveness. If the ACPA concentrates its resources on the set of bills that are most likely to pass, it puts its resources to their most efficient use. The ACPA can also improve equity among various sections of the population by making sure that the cohort of 'most-likely-to-pass' bills they focus on helps reduce disparities among people. Finally, focusing on the bills most likely to pass may translate into the biggest impact on protecting the civil liberties of individuals in the US. In other words, working on the 'right' cohort of bills in a proactive way may help achieve maximum success in blocking/overturning harmful legislation and thus protecting people. Overall, if this problem is solved, there may be several vulnerable sections of the society which may be protected (and benefitted) by a proactive approach. It is possible that some groups are unable to reap the benefits of ACPA's interventions at present, simply because the organization's may not be intervening on the 'right' set of bills. Through our work, we intend to help ACPA focus on the optimal cohort of bills and achieve much better results compared to the status quo.\

Related Work

The history of efforts to influence legislative success goes as far back as the history of legislatures, themselves.[1] Influencing legislative outcomes requires at least an implicit theory of what causes or predicts legislative success. In recent decades, scholars have worked to interrogate and formalize this knowledge both within specific legislatures and across legislative bodies. Work in the past few decades explored the impact of specific legislators or institutional roles and found that properties of individual legislators—like party, seniority, leadership position, age, race, gender, and even education level[2,3]—are decently predictive of success at the legislator level.

Using quantitative methods to predict legislative outcomes for bills is becoming more common as access to public data and computational resources improves. Targeted analyses of subtle legislative dynamics, like working relationships between legislators, are now common.[4] Recognizing the success of these investigations, firms like FiscalNote now sell predictive tools to individuals and organizations that would like to influence legislative outcomes.[5]

Recent investigations have incorporated machine learning methods to predict legislative success. It seems, however, that most of the predictive investigations in the US focus on Congress rather than on state legislatures, perhaps because of the richer availability of textual and contextual data

[1] https://www.opensecrets.org/resources/learn/lobbying_timeline.php
[2] Ellickson MC. 1992. Pathways to Legislative Success: A Path Analytic Study of the Missouri House of Representatives. *Legislative Studies Quarterly,* 17(2):285-302.
[3] Bratton KA, Haynie KL. 1999. Agenda Setting and Legislative Success in State Legislatures: The Effects of Gender and Race. *The Journal of Politics,* 61(3):658-679.
[4] Kirkland JH. 2011. The Relational Determinants of Legislative Outcomes: Strong and Weak Ties Between Legislators. *The Journal of Politics,* 73(3):887-898.
[5] Gaynor MJ. 1 February 2018. Can big data predict which bills will pass Congress? *The Washington Post Magazine.*

for the higher-level governing body. These reports often find that contextual data about bills—sponsor information, changes in status since introduction, etc.—provide most of the information needed for accurate prediction. Including the text of bills in feature generation only slightly increases performance.[6,7] These findings on the relative predictive value of contextual versus textual data have been corroborated in non-peer-reviewed reports by graduate student project groups.[8,9]

Our goal in working with the ACPA, protecting civil liberties in Pennsylvania, is only possible if their advocacy is able to meaningfully influence the legislative process. Recent work focusing specifically on state legislatures confirms that professional advocacy for a bill or initiative, often called "lobbying," is predictive of legislative outcomes.[10] Lobbying appears to influence outcomes through the two primary channels available to legislators: agenda control and negative advocacy.[11] Agenda control refers to the process by which legislators decide which bills receive hearings in committee or debate time on the floor of a given chamber. Negative advocacy refers to various efforts that aim to *prevent* a bill's passage.

Our project focuses on the intervention of a legal challenge after a bill has passed. Our work is unlike previous deployments of machine learning methods for legislative prediction in two ways. First, we are focusing on a state legislature, Pennsylvania's, where the available data lacks the temporal and contextual richness of federal data, and we are focusing on all of the bills in a given session instead of a single-issue area or a single step in a bill's progress. Second, the target metrics of our analysis are cohort-based, e.g. performance for a selected 15% of a session's bills, rather than broad-based, e.g. overall error rate. We posit that this shift in focus provides a better fit to this organized advocacy setting, where our results will be reviewed by expert readers. It is this binding capacity constraint, bill review time by ACPA staff, that motivates and differentiates our project from others like it.

---

[6] Yano T, Smith NA, Wilkerson JD. 2012. Textual Predictors of Bill Survival in Congressional Committees. *2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* 793-802.

[7] Park SO, Hassairi N. 2021. What predicts legislative success of early care and education policies?: Applications of machine learning and Natural Language Processing in a cross-state early childhood policy analysis. *PLOS ONE,* 16(2):e0246730.

[8] Abramowitz B, Chen J, Mosley G. Predicting Legislative Success. https://beckyabramowitz.com/public/ML_Report.pdf

[9] Goldblatt D, O'Neil T. How a Bill Becomes a Law - Predicting Votes from Legislation Text. https://nlp.stanford.edu/courses/cs224n/2012/reports/writeup.pdf

[10] Butler DM, Miller DR. 2022. Does Lobbying Affect Bill Advancement? Evidence from Three State Legislatures. *Political Research Quarterly,* 75(3):547-561.

[11] Anderson WD, Box-Steffensmeier JM, Sinclair-Chapman V. 2003. The Keys to Legislative Success in the U.S. House of Representatives. *Legislative Studies Quarterly,* 28(3):357-386.

Problem Formulation, Overview of Solution and Goals

As stated earlier, the primary objective of this project is to protect the civil liberties of people in Pennsylvania. However, it is important to explain the channel through which we try to achieve this broad goal via our work. Thus, we give an analytical formulation of our problem, briefly describe what our solution looks like and discuss the efficiency, equity and effectiveness aspects of our solution, keeping our main goal in mind.

Analytical Formulation of the Problem

We formulate our task as follows: **Every week, for all bills introduced in the current session of the Pennsylvania General Assembly which have not been passed yet, we attempt to identify 15% of bills that are most likely to pass by the end of the session in order to help the ACPA protect civil liberties by allocating their limited resources to actions such as: (1) legal review of bills, (2) lobbying against bills that infringe on civil liberties, and (3) preparation for challenging such laws in court.**

Before delving into ML models that produce a relevant list of 15% of bills, we ask if there are simpler was to achieve this, i.e. are there ways in which the ACPA can figure out such a list using existing data on bills which does not require ML. For this, we explore several baseline models. Next, we build a set of ML models and compare their performance with the baselines.

To evaluate how well our models generalize to unseen data, we use a temporal cross validation strategy (Figure 1) and compare model performance using the metric recall at 15%. We use recall at 15% to compare our models instead of precision at 15% because the base rate for bill passage in a session is only about 5% in each session. Thus, we are interested in what proportion of the bills in the full cohort are included in our 15% list and the quality of our models can be evaluated somewhat smoothly from 0-100%. As part of our temporal cross-validation strategy, we take the most recent two years of our data set for validation and use the data before that for model training. This splitting into training and validation sets is carried out every 6 months in our data as shown in figure 1. Each point of time in our analysis, we do not consider bills that have already passed.

Project Scope and Goals

The primary goal of our project - protecting civil liberties of people in Pennsylvania, can be broken down into various sub-goals. First, we can ensure that resources are spent in an optimal way (with an aim to protect civil liberties). This sub-goal is efficiency, meaning the list of bills we spend resources on, should include all bills that actually end up passing. Second, we can ensure that no particular group of bills (hence no specific group of people) is intervened on much more frequently or much less relative to other groups. This sub-goal is equity and is aimed at protecting civil liberties equitably. Third, we can ensure that the ACPA's intervention actually ends up blocking a harmful piece of legislation and protects civil liberties in the process. This sub-goal is effectiveness and it checks if interventions actually translate to improved outcomes for people.

Given the scope of our project, these goals can also be restated based on how our model expresses these goals. For efficiency, we would want to maximize recall at 15%. For equity, we would want our model to have a recall disparity (discussed in detail in the bias audit section) as close as possible to 1. For effectiveness we should be able to show a causal impact of ACPA's intervention on protection of civil liberties.

The ACPA might want to pursue efficiency, equity, and effectiveness goals in this project, but there may be trade-offs between these goals. ACPA staff must balance these goals operationally as they see fit. Striking the right balance requires understanding what the organization prioritizes. Because the base rate of bill passage is only about 5% per session and we are able to pass 15% of bills along to the ACPA, efficiency and effectiveness can be pursued in tandem by trying to maximize recall at 15%, a model efficiency goal. However, in order to better assess effectiveness of ACPA's intervention, we might need a field trial as discussed in the field validation section. As far as the equity goal is concerned, our model is unable to understand which people might have their civil liberties impacted by a given bill—we rely on the ACPA's expert readers for this—so we use a collection of weak proxies to pursue an equity goal in our bias audit, including sponsor gender balance, political party, urbanicity, and age..

Our results section states that the models that maximized our efficiency goal (recall at 15%) also tended to perform best in terms of equity based on the bias audits we performed on our models. We acknowledge, though, that these model goals are weak proxies for their analogous project goals, and we are relying on our partners at the ACPA to strengthen this analysis over time by tagging reviewed bills, thus improving the ability of our model to directly identify bills that pass and threaten civil liberties. We describe this in more detail in our Policy Recommendations section later.

Data Description

Datasets

In order to solve this problem, we needed to acquire quality data that would allow us to explore the impact of a variety of different features on whether or not a bill will pass. The first dataset came from LegiScan, a database which contains a wide range of information on bills from all state legislatures in the US. This database provided us with details on the introduction data of a bill, who sponsored it, and its progress through both the Senate and House among many others. While exploring the wealth of data that LegiScan has we also noted some deficits, such as a lack of individual level data about sponsors other than their party affiliation. Therefore, we sought a dataset that could provide us with demographic features that we could incorporate into our model. We identified two datasets that filled this need, one was a series of member biography Excel files for the House of Representatives and Senate from 2009 and 2020 from the Pennsylvania General Assembly. The second was a dataset from Open Secrets and included member level campaign finance data.

The data from the Pennsylvania General Assembly included representative level information such as gender, age, the year they were first elected, a list of years they have served, positions they have held, and the district they represent among others. This dataset will allow our model to include demographic information about the sponsors of a bill that will not only give us the ability to derive potentially crucial features but also to identify if there is any bias in our model towards certain types of sponsors.

Open Secrets provided us with data about the campaign finance records for representatives that are in the Pennsylvania House and Senate. We were able to obtain data from an API available through FollowTheMoney.org which is part of Open Secrets. This dataset allowed us to introduce a monetary feature into our model, specifically the total amount of campaign contributions a sponsor received.

Data Exploration

We explored our data initially to visualize the difference between bills that have passed and those that failed or expired in Pennsylvania. This helped us to see what type of features might be useful to our model and aid us in determining whether or not a bill will pass. Additionally, we looked at the distribution of bills in our ACPA issues feature that we created in order to assess how many of the bills in our dataset were assigned to each of the issues.

Figure 2 shows the difference in the number of days it takes for a bill to pass versus the number of days until a bill fails or expires. We measured the days until a bill fails or expires by subtracting the end of the session from the last progress status for a bill. As can be seen in Figure 2, a majority of bills that pass appear to do so within the first 100 days since they are introduced. There are significantly fewer bills that pass after 100 days since they were introduced. As for bills that end up failing or expiring, there is much more variability, but we can see in the figure that a bulk of these bills are in the 500-700 day range. A significant difference in average and modal values between the two graphs highlights the importance of the variable 'days since introduced' in predicting bill passage.

Figure 3 explores the relationship between the number of sponsors for bills that pass and those that do not pass. On the left side of the figure, we can see that for a majority of bills that have passed they had between 15 and 50 total sponsors. The figure on the right shows how a majority of bills that did not pass had between 5 and 20 sponsors. Therefore, it appears that bills that pass have on average, a higher number of total sponsors than those that end up failing/expiring.

Table 1 contains the count and share of bills in our dataset across each of the 17 issues the ACPA deals with. The issue area that is associated with the largest number of bills is Capital Punishment with a total of 1,639 bills which makes up 7.3% of all bills. The issue area with the lowest total is Human Rights as only 11 bills were tagged with that issue which makes up 0.049% of all bills. We can see from this table that the number of bills across each issue is not equal and that some issues are not as well represented in our dataset as others.

Table 14 also cross tabulates some attributes which may be important for predicting bill passage.

Analysis of Results

Our data exploration section suggests that variables such as the number of days elapsed since the bill was introduced, the number of sponsors for a bill, characteristics of a bill's sponsors and the governor's party may all provide important information regarding the passage of a bill.

As outlined in the section 'Problem formulation and Overview of Solution', we use a temporal cross-validation strategy to train and validate our models. For this purpose, we split our model into training and validation sets every six months as shown in Figure 4.

Before looking at various machine learning models, we ask what the ACPA can easily do today that's based on available data and does not need ML. We refer to these methods as baseline models which essentially rank bills (or categorize them) based on one or more bill characteristics. Bills that are ranked higher (or lower for some characteristics) might be more likely to pass. Such characteristics can be based on practical knowledge or sensible heuristics. Among the various baseline models we tried, the four baselines shown in table 2 have the highest recall at 15% averaged over all validation sets.

Next, we build various ML models and compare them to our baseline models to check if they achieve a significant improvement over methods that do not involve ML. The following is a list of model groups that we trained and evaluated for our study.

1. Decision Tree
2. Logistic Regression
3. AdaBoost Classifier
4. Gradient Boosting Classifier
5. Random Forest Classifier
6. Neural Network (Multi-layer Perceptron Classifier)

As mentioned before, we perform a classification task for which we use simple classifiers like decision trees as well as ensemble methods such as random forests. The various sets of hyperparameters used for each of these model groups can be found on this link.

We evaluate the performance of various model groups using our chosen metric recall at 15%. Figure 5 shows recall at 15% for all trained models over time. As one can see, all the random forest classifiers and gradient boosting classifiers are bunched at the top and perform consistently well. On the other hand, the logistic regression with L1 penalty and C = 0.001 performs quite poorly initially and then stabilizes at a higher performance level (this could be because of the combination of hyperparameters; for the models trained on the initial validation sets, most features have zero weights). While one would expect a neural network model to yield fairly high performance, that is not the case here. In fact, some of the baseline models perform much better than the neural network, showing that simpler methods such as ranking might sometimes perform at least as well as (or better than) more complex ML models. However, this result is subject to change as we tried a limited set of hyperparameters for the neural network.

We use average recall at 15% as the metric for choosing the best model because with data on a small set of sessions from 2009-2020, we would want our model to perform well overall. Performance on the most recent validation set might only indicate if our model was able to predict passage well for a certain kind of bills (perhaps more popular in the latest session) and may not tell us if the model is robust to change in political environments. The average metric thus is a more comprehensive metric. However, given figure 5, it seems like regardless of whether we based our decision on average performance or most recent performance, the random forest classifiers turn out to be the best.

Table 3 gives a list of the five models with the highest average recall at 15%.

We next compare our best performing model to our best performing baseline model. Figure 6 shows that at a threshold of 15%, the ML model achieves a 20 percentage point improvement in recall and a 6 percentage point improvement in precision. Moreover, at both lower and higher thresholds (percent of population threshold), the ML model does consistently better than the baseline.

Focusing on the five best performing models, we look at how they learn to classify bills from the various features provided to them during training. Figure 12 shows that the top 5 most important set of predictors for the best model are:

1. Bill progress variables
2. Days since bill was introduced
3. Number of times bill has been voted on
4. Number of sponsors for a bill
5. Bill's Lead sponsor's party affiliation

The feature importance graph shows that features that provide information about how a bill progresses in a session are particularly relevant in predicting passage of bills. While this is expected, the issue with this is that many of these progress variables are known much closer to the actual passage date of a bill. As a result, this might affect our model's ability to make accurate predictions early on when the bill is introduced. This is discussed later in the caveats section. Other top important features like sponsor information can be thought of as contextual features that may indicate the level of influence a sponsor has and hence predict passage of a bill well. Relative to these features, the textual features obtained via clustering don't seem to be highly relevant for the model. Table 13 shows that these trend hold for all 5 best performing models.

Having seen the feature importance graph, we may expect that the most important features might also differ the most by their average value among bills with predicted scores in the top 15% relative to the bottom 85%. While this may not always be the case in the validation set, Table 4 to 9 confirm this expectation to some extent. The bills among the top 15% differ the most in terms of the number of times they have been passed out of a committee, number of times they have voted on and the number times they have been engrossed. Thus, when the model predicts on a new data set, it is likely to produce a list of 15% of bills most likely to pass that differ the most along these features.

Bias Audit

The analysis till now looked at the performance of our models in terms of how well they are able to identify bills that actually end up passing. While this takes care of our efficiency goal to a large extent, it does not directly address our equity goal. For this, we look at the recall disparity (alternately called recall parity) metric defined as follows.

Recall Disparity =Recall@15pct for Protected GroupRecall@15pct for Reference Group

In the above equation, protected group refers to a group of bills (identified by some feature) that we care about and the reference group refers to a group of bills we would like to compare the protected group's recall against. Since we care about equity in outcomes for different groups, we would want the value of the disparity to be as close to 1 as possible. Thus, we say that a degree of bias a model displays towards a certain protected group translates to how far the recall disparity is from 1.

We choose recall disparity to measure model bias because our intervention is assistive in nature and we can intervene on only a small fraction of bills introduced in a session. As a result, while getting the top 15% of the bills we should intervene on, we would like to ensure that our model is able to capture actual cases of passage equally well for different groups of bills (affecting different sections of the population). It may be the case that in the real world, bills covering some specific contents rarely pass and hence do not feature in our top 15%. However, for those rarely passed bills we would like to ensure that our model is able to identify bills that pass almost as well as for any other reference group.

The rationale behind defining the groups of bills we assign as protected groups is linked to our equity goal. At a higher level, we would like to ensure that the ACPA does not neglect certain groups of people while intervening on a set of bills to protect civil liberties. For instance, if our model is able to predict passage of bills related to voting rights much better than passage of bills related to reproductive rights, the ACPA might end up only spending its resources for challenging/supporting the former kind of bills. In doing so it may be neglecting women (especially from low income groups) while protecting civil liberties. Thus, the bill group attributes we choose to check model bias, essentially translate to the groups of people such bills end up affecting.

Given this context, we check if recall at 15% for our models is disproportionately low for the following protected groups relative to the reference group:

1. Bills related to capital punishment vs. the rest
2. Bills related to reproductive and women rights vs. the rest
3. Bills related to voting rights and free speech vs. the rest
4. Bills with at least one female lead sponsor vs. the rest
5. Bills with lead sponsors from the youngest category vs. those sponsored by the oldest age categories
6. Bills with a proportion of urban sponsors in the top percentile vs. bills with a proportion of urban sponsors in the bottom percentile
7. Bills mainly sponsored by Democrats (i.e. more than half of sponsors are Democrats) vs. the rest

Reiterating our previous point, each of the above protected groups of bills are likely to impact a different section of the population. For example, bills related to reproductive rights or women rights affect women and bills related to voting rights or free speech might disproportionately affect certain racial minorities in the US.

Figures 13 to 19 plot recall disparity against recall at 15% on the most recent validation set for all models that we trained. Each point on the scatter plot represents a model we trained. Ideally, we would want a model to be on the right-most point on the horizontal line at recall disparity = 1.

The plots for protected groups 1 (Figure 13), 2 (Figure 14), 4 (Figure 16) and 7 (Figure 19) show that there is not much of a tradeoff between equity (recall disparity) and efficiency (recall). The random forest classifiers seem to dominate all other models in terms of both equity and efficiency. In fact the best performing model in terms of recall at 15% (table 3) also seems to be the best in terms of equity. However, one can see a slight tradeoff between these two goals for protected groups 3 (Figure 15), 5 (Figure 17) and 6 (Figure 18). For instance, for the protected group 'Bills related to voting rights and free speech' (Figure 15), the decision tree classifier with max_depth 10 and min_samples_split 20 seems to be doing better than the random forest (with 1000 trees) in terms of recall disparity (around 1) but worse in terms of recall at 15% (around 67%). When faced with a trade-off between different models, an organization needs to make a choice based on how much bias it is willing to accept towards a protected group for an improvement in model's recall. Moreover, the final choice of model also depends on which protected group the organization cares about the most.

Since we lack information about the protected group our organization would care about the most, we recommend a model/strategy for different possible preferences.

If the group of bills the ACPA cares the most about is:

1. Bills related to capital punishment: **Recommended model is a random forest with 1000 trees**
2. Bills related to reproductive and women rights: **Recommended model is a random forest with 1000 trees**
3. Bills related to voting rights and free speech: **Recommended strategy is to choose from a frontier of models that strictly dominate other models in terms of both equity and efficiency.**
4. Bills with at least one female lead sponsor: **Recommended model is a random forest with 1000 trees**
5. Bills with lead sponsors from the youngest category: **Recommended strategy is to choose from a frontier of models that strictly dominate other models in terms of both equity and efficiency.**
6. Bills with a proportion of urban sponsors in the top percentile vs. bills with a proportion of urban sponsors in the bottom percentile: **Recommended strategy is to choose from a frontier of models that strictly dominate other models in terms of both equity and efficiency.**
7. Bills mainly sponsored by Democrats (i.e. more than half of sponsors are Democrats) vs. the rest: **Recommended model is a random forest with 1000 trees**

For instance, if we were to assume that the organization cares the most about bias towards bills related to voting rights and free speech, we would look at the models on Table 10. This table shows that all the models that lie on the pareto frontier, i.e. there is no other trained model that does better on both equity and efficiency relative to these models.

Having, outlined the above contingency plan, we would note that if we had to pick one model that performs well in terms of recall and does fairly well in terms of fairness for most protected groups, it would be the Random Forest model with these hyperparameters: "max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 1000, "random_state": 123, "min_samples_split": 20.

Field Trial

Apart from efficiency and equity we would also like to address our goal of effectiveness. In this regard, we would like to check if the ACPA's intervention on our chosen list of bills actually ends up protecting civil liberties and helping people. While we cannot directly measure 'protection of civil liberties', we can assess the success of ACPA's intervention in terms of its ability to overturn or block laws that infringe on civil liberties. One point to note is that we are only focusing on one type of intervention - legally challenging a bill that has passed. There may be other forms of interventions such as advocacy efforts, lobbying etc. that we are not considering in this project.

We recommend that the ACPA design a randomized controlled trial to evaluate the effectiveness of its intervention. All bills that are introduced in the current Pennsylvania session until a certain date, can be randomly assigned to either a treatment group or a control group. For bills in the treatment group, the ACPA would intervene only on the ones included in the 15% flagged by our model. For bills in the control group, the ACPA would intervene on them based on their original method (without ML, this could be a baseline model or some simple heuristic). The outcome of interest would be the number of bills that were successfully blocked in each of these groups. Due to randomization, the difference in the outcomes for these groups would measure the causal effect of ACPA's intervention.

We must note that the above proposed method is far from perfect and there are certain caveats attached to it which affect its potential to give us a causal effect. For instance, given ACPA's resource constraints, it can only spend on 15% of bills introduced in a session. This would mean that within the treatment group, the ACPA may not be able to spend on the entire list of bills even if all of them were potential threats to civil liberties. This is because it would also need to spend on bills in the control group for evaluating the effectiveness of the intervention. We would also need to ensure that randomisation at the bill-level has been done properly by checking if bills in the two groups are roughly similar in attributes such as content, sponsorship etc. post-randomisation. Another important caveat is that our model simply classifies bills based on a score, indicating if they would pass. It has not learned to predict if they can be blocked. On the other hand, the heuristic/baseline used to predict which bills would pass might also incorporate whether the bill can actually be challenged after it passes. Thus, it is possible that the number of bills blocked in the control group turns out to be higher than in the treatment. This however, does not mean that our model is completely ineffective. It simply needs to be trained for a different outcome.

Lessons and Recommendations for the partner organization

Policy Recommendation

As we mentioned, we would recommend a model selection strategy based on the preference of the organization. For some cases, model choice may be unambiguous as there is no trade-off between equity and efficiency. However, for others where there exists a trade-off between these two goals, the ACPA needs to decide which model to use in practice according to their own value systems.

Because we didn't run a field validation trial, i.e. we didn't see if our model actually helps protect civil liberties via ACPA's intervention, we would recommend that ACPA implement our model on a pilot basis and evaluate the impacts of our model in terms of preventing infringement of civil liberty.

Additionally, our model can be improved by feedback from ACPA. One method of feedback that could improve future versions of these models is assigning issue and impact tags to bills reviewed by ACPA. Currently our models use textual information such as bills description and bills text to define categorical variables that indicate if bills are related to a certain issue (that the ACPA cares about) instead of using data from ACPA itself. However, these categorical variables might be inaccurate (since they're based on an ad hoc word search method) and might have repercussions for the issue related bias audits. Issue and impact tagging will contribute to accurate bias-audit evaluation of our models.

Furthermore, early collection of data will help ACPA achieve their goal. Our best performing models depend on features about bills' progress such as referral to committee and the number of votes. Also, since bills can pass in a month, collecting more data about bills early on in the bill-cycle will be helpful. Getting such information early would help ACPA make accurate predictions early and secure enough time to intervene. We would recommend investing in a network of legislative member advocates who can inform the ACPA of relevant bills and their progress.

Proposal for Future Avenues of Research

Caveats

In terms of achieving our goal efficiently and effectively, our models have several limitations. First, while our models predict passage of all bills, our models are not optimized to predict if bills cause infringement of civil liberties. As a result, our models might miss some bills that hurt civil liberties. Also, our models do not consider difficulties of intervention. It is possible that our models return bills that are almost impossible to prevent from passage. In that sense, our models might not be efficient or effective to prevent infringement of civil liberties.

Furthermore, our model might not be able to make predictions early enough. As shown in Figure 20, when we validated the performance of our best performing model using the data in the latest 2 years, while our model successfully detected more than 80% of bills that end up with passage, for 35% of them ACPA has at most one month to intervene, and for more than half of them ACPA has at most 3 months. It would suggest our models' limitation in terms of timeliness of predictions.

Future Work

The analysis that we have done in this project has only opened up more possibilities in terms of future work. One such avenue would be identifying additional sources of information to create features for our model. This includes features which provide more contextual information early on for bills, such as amendments that have been added, the dates of sponsorship for various sponsors, media coverage of different bills, and web scraping social media activity, such as tweets, and performing a sentiment analysis. This would allow us to test new features that our current model does not cover. Additionally, most of our current features do not change much over time and are for the most part determined right when a bill is introduced, for example, the number of sponsors and descriptive information about those sponsors, such as their age and district. Having features that change over time could add more complexity to our model which may help with the accuracy of our predictions.

Additionally, we would like to explore how we could either change our model or create a new model that could help the ACPA allocate resources for their different methods of challenging bills that infringe on the rights of individuals in the US. When we started this project we decided to focus on legally challenging passed bills in court, and that meant predicting bills that were most likely to pass, and therefore become law. But the ACPA also challenges laws through lobbying efforts in order to stop a bill from passing. For this problem instead of focusing our model on bills that are mostly likely to pass we would want to identify bills that are on the edge of passing or failing. Reconfiguring our existing work in order to tackle this problem is another way that we could help the ACPA achieve its goals.

The final step in our future would be exploring expanding our machine learning pipeline to other states. As a proof-of-concept we chose to focus solely on PA for this project. Implementing our model to other states would require conducting some initial research on what the legislative process looks like in those states, since it varies across the US, and determining what information is available through LegiScan.

Conclusion

This project aimed to help the ACPA protect the civil liberties of individuals in Pennsylvania by creating a machine learning model that could provide the organization with a list of the top bills that were most likely to pass in a single session. We were challenged throughout this process to implement equity, effectiveness, and efficiency goals that we defined at the beginning of this project while managing necessary tradeoffs. In the end we were able to create a model that is capable of providing the ACPA with about 80% recall (based on validation) of passed bills in a session in the 15% list of bills we deliver. This project opened up many more paths for future work on this topic such as creating new features for our existing model or creating a plan to expand this model to other states. While there is still room to improve, we hope the model described in this report is a pivotal tool in the ACPA's efforts to protect the civil liberties of individuals in Pennsylvania.

Appendix

Path to experiment config file:
https://github.com/dssg/mlpolicylab_fall22_bills1/blob/main/triage_files/experiment_config_large.yaml

Figures

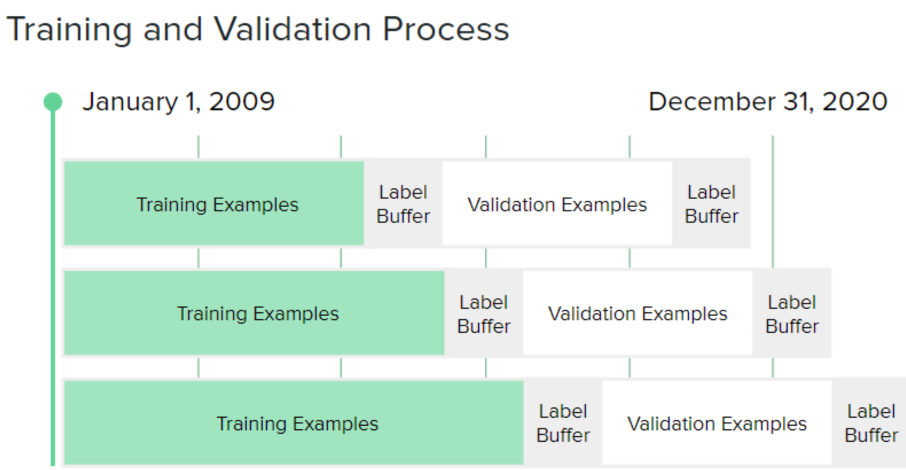

Figure 1: Not-to-scale example of how temporal cross validation splits data into training and validation sets.

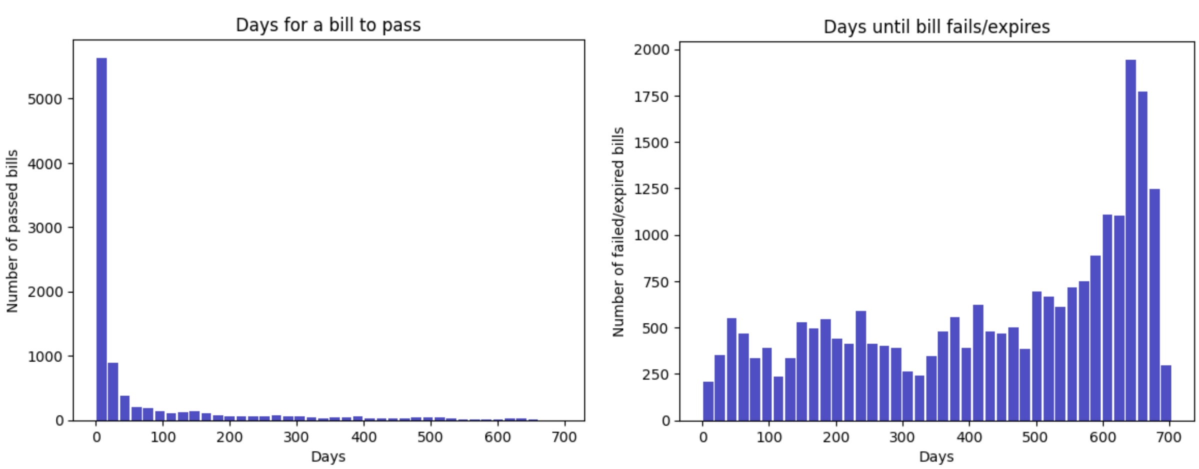

Figure 2: Number of days until a bill passes or fails (Link to code)

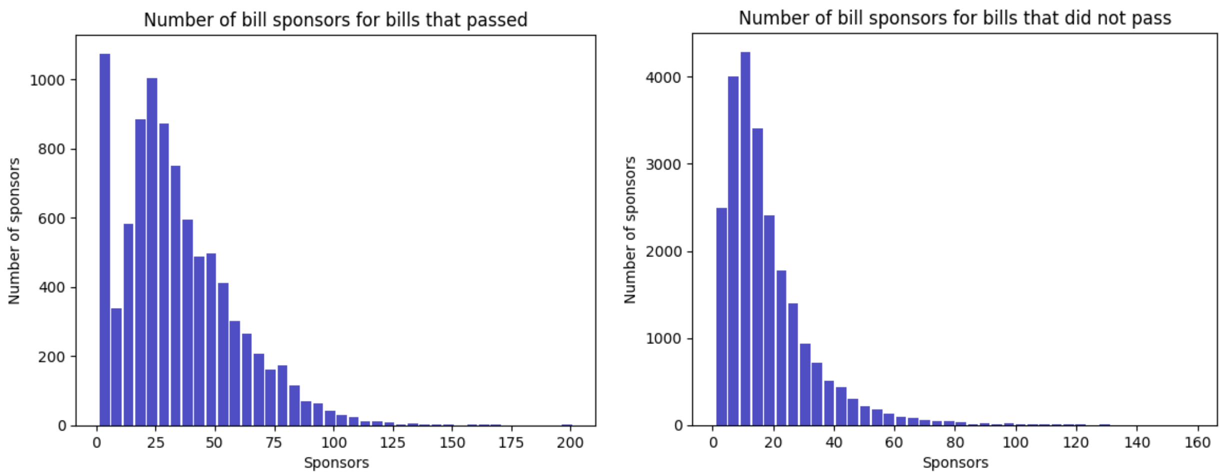

Figure 3: Number of sponsors for bills that pass and those that do not pass [(Link to code)]()

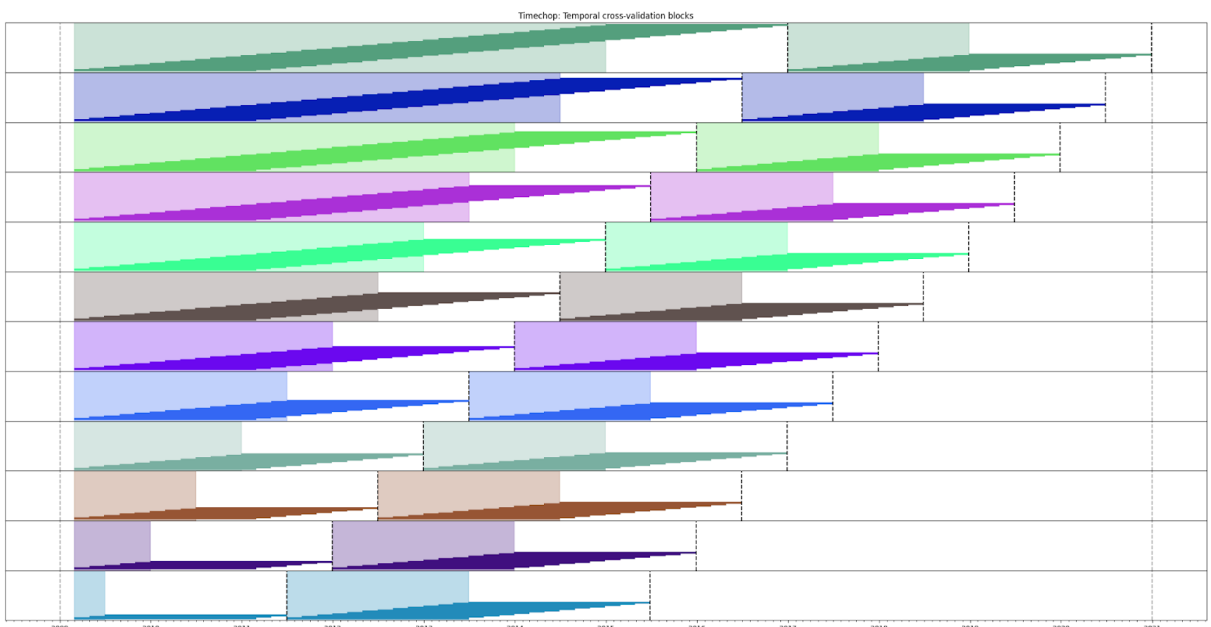

Figure 4: Detailed temporal training and validation splits for models updated every six months and trained using bi-monthly data on bills with a label buffer of 2 years.

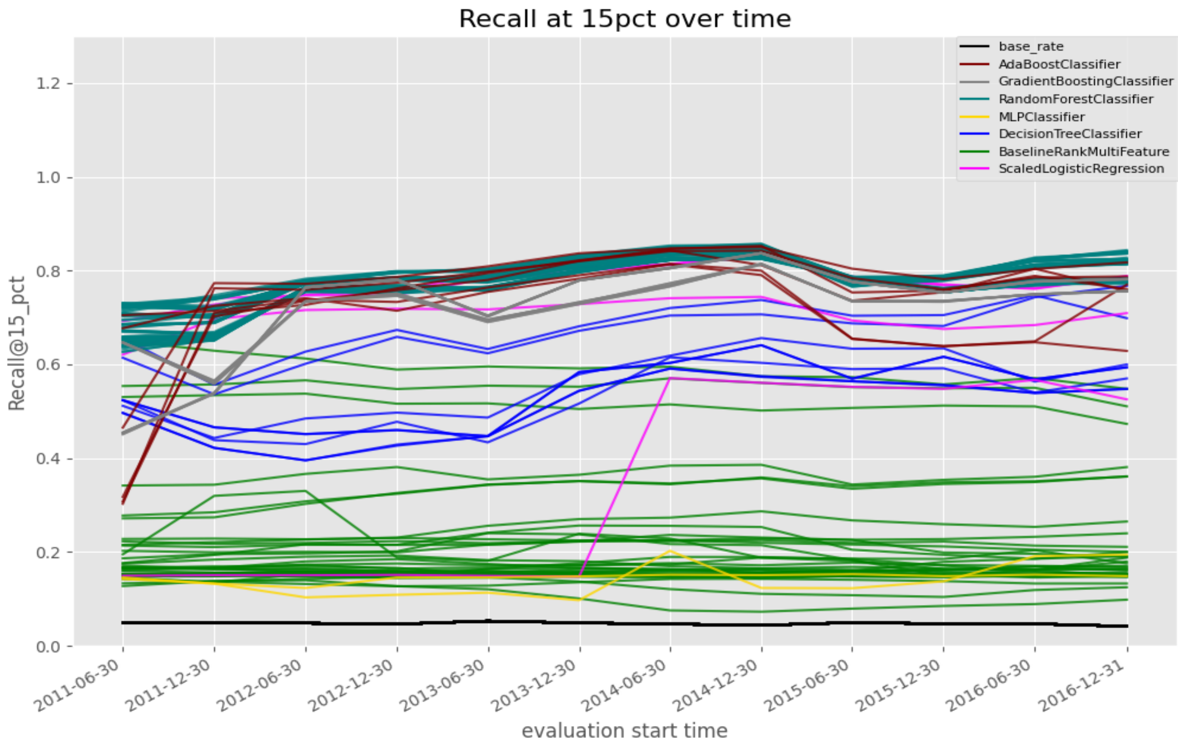

Figure 5: Model performance (recall@15pct) over time ([Link to code](#))

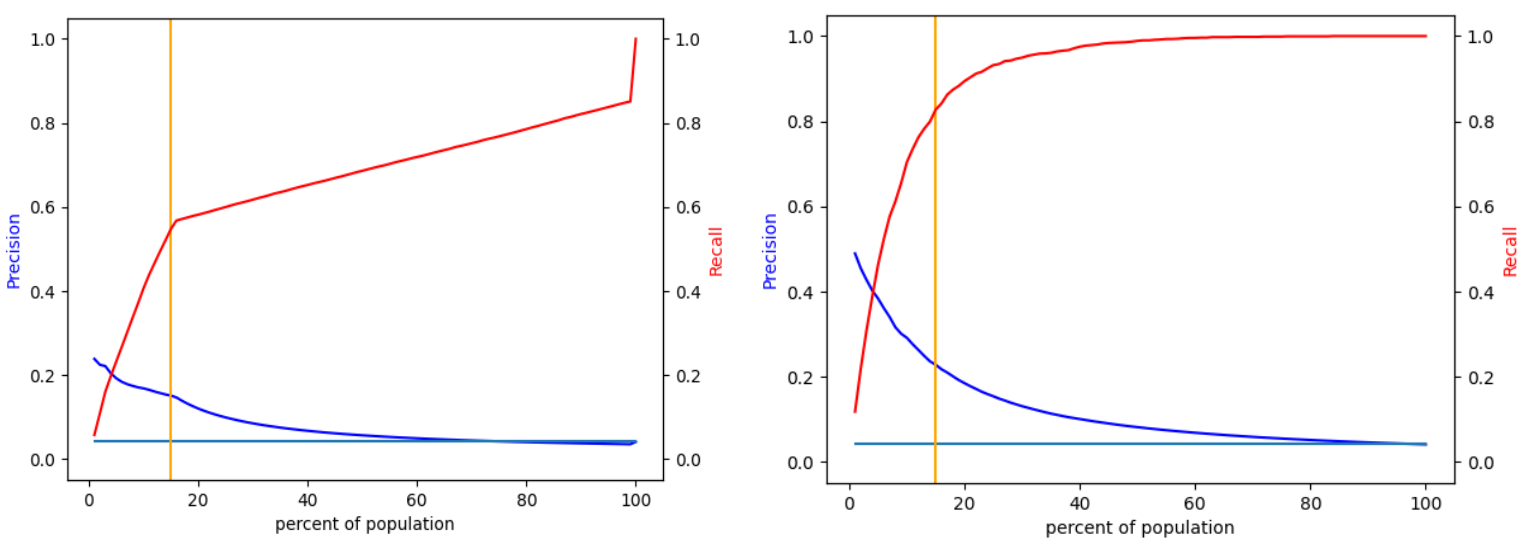

Figure 6: Baseline: bill referred to a committee.
Average Recall: 59% (Left).
Random Forest Classifier with 1000 trees. Average Recall: 80% (Right)

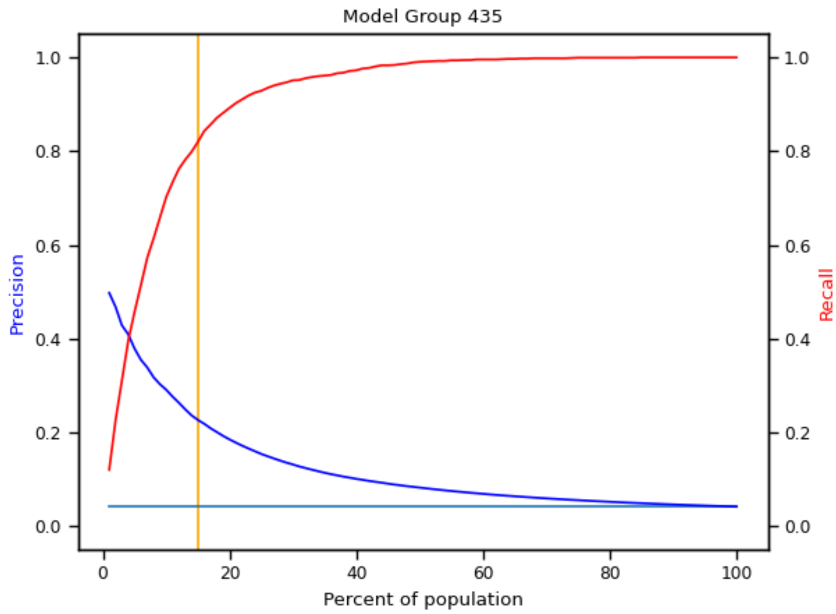

Figure 7: Precision-Recall curve across the population of active bills for model group 435 on validation date 12/31/2016. (Link to code)

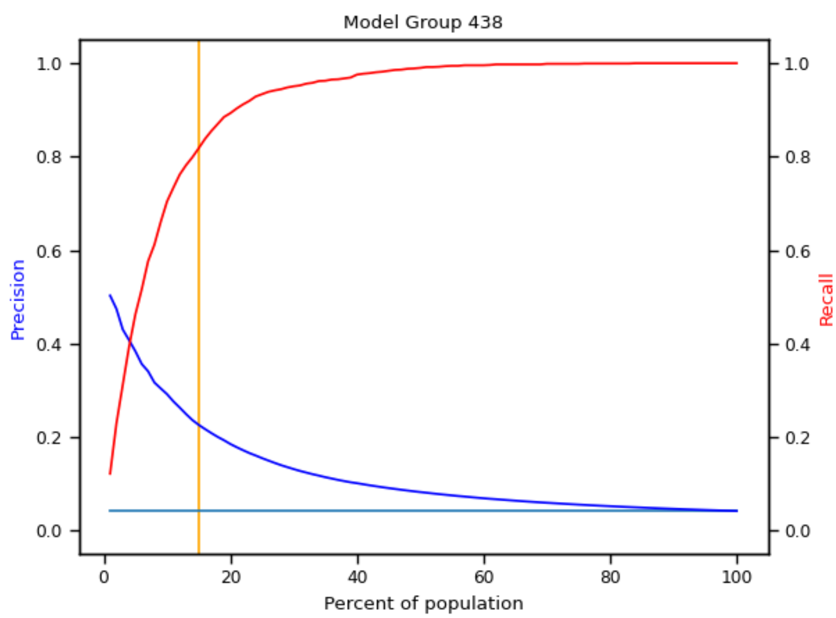

Figure 8: Precision-Recall curve across the population of active bills for model group 438 on validation date 12/31/2016. (Link to code)

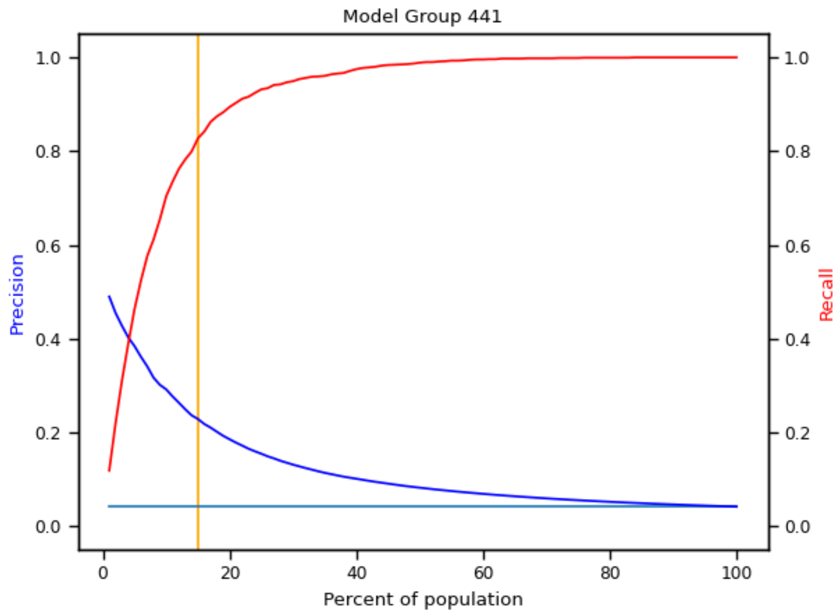

Figure 9: Precision-Recall curve across the population of active bills for model group 441 on validation date 12/31/2016. ([Link to code](#))

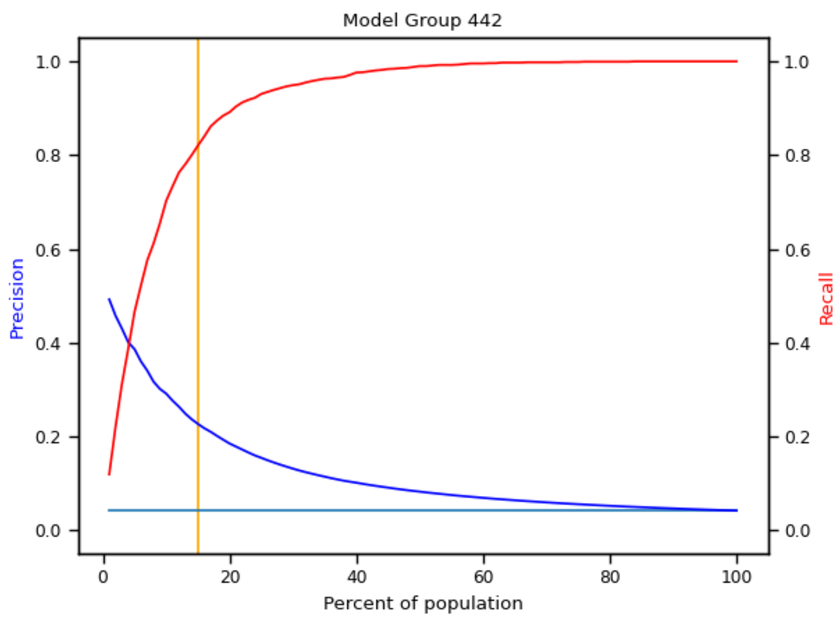

Figure 10: Precision-Recall curve across the population of active bills for model group 442 on validation date 12/31/2016. ([Link to code](#))

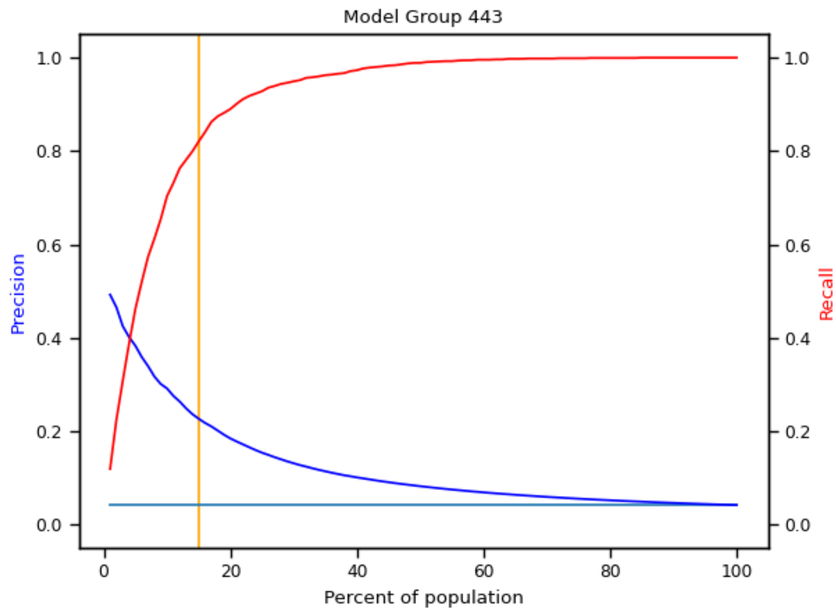

Figure 11: Precision-Recall curve across the population of active bills for model group 443 on validation date 12/31/2016. ([Link to code]())

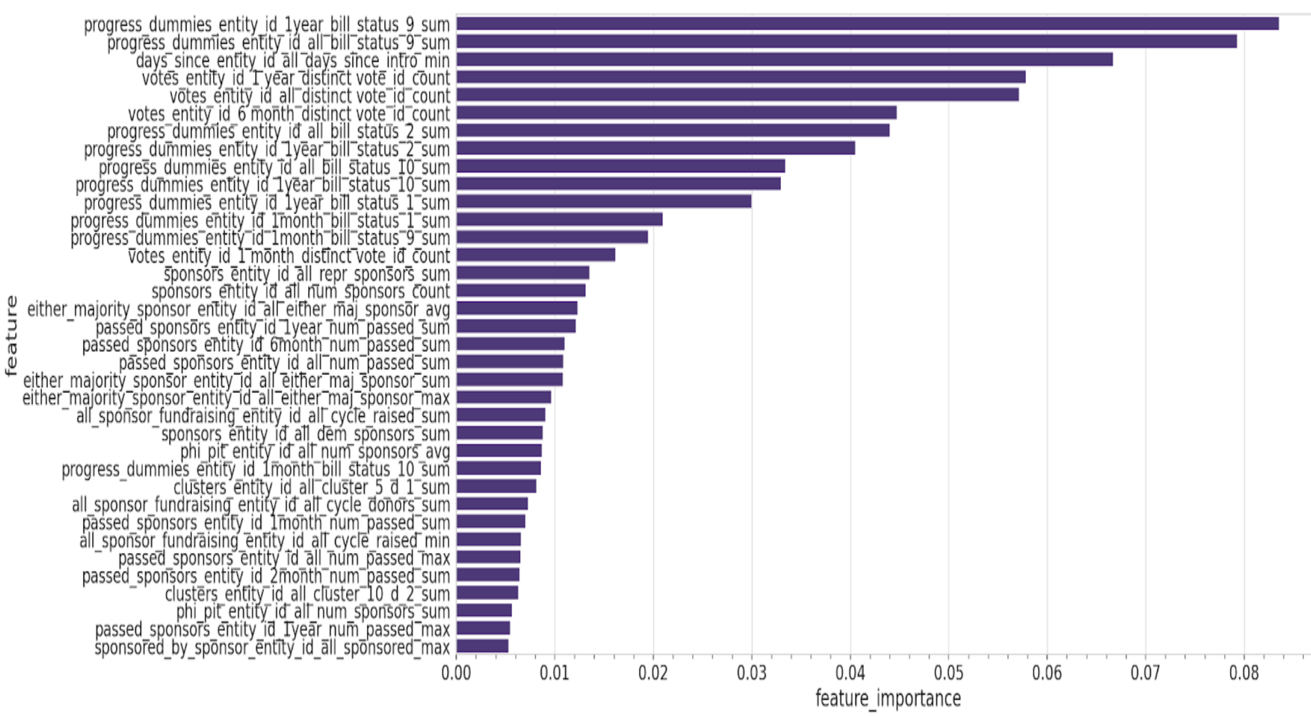

Figure 12: Feature Importance Graph ([Link to code]())

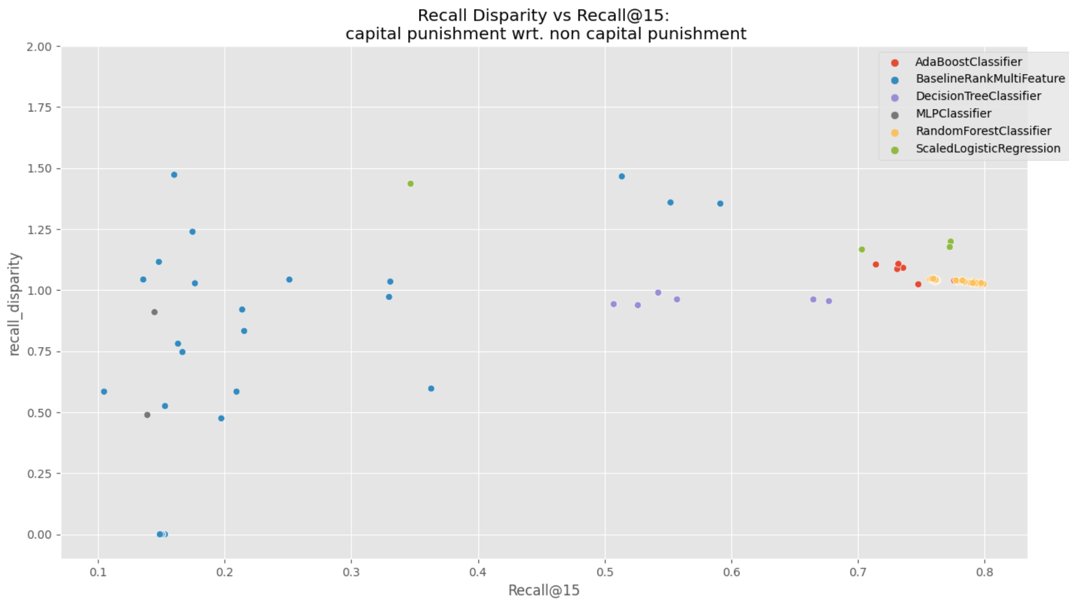

Figure 13: Recall Disparity vs Recall: Bills Related to Capital Punishment vs. the Rest([Link to code]())

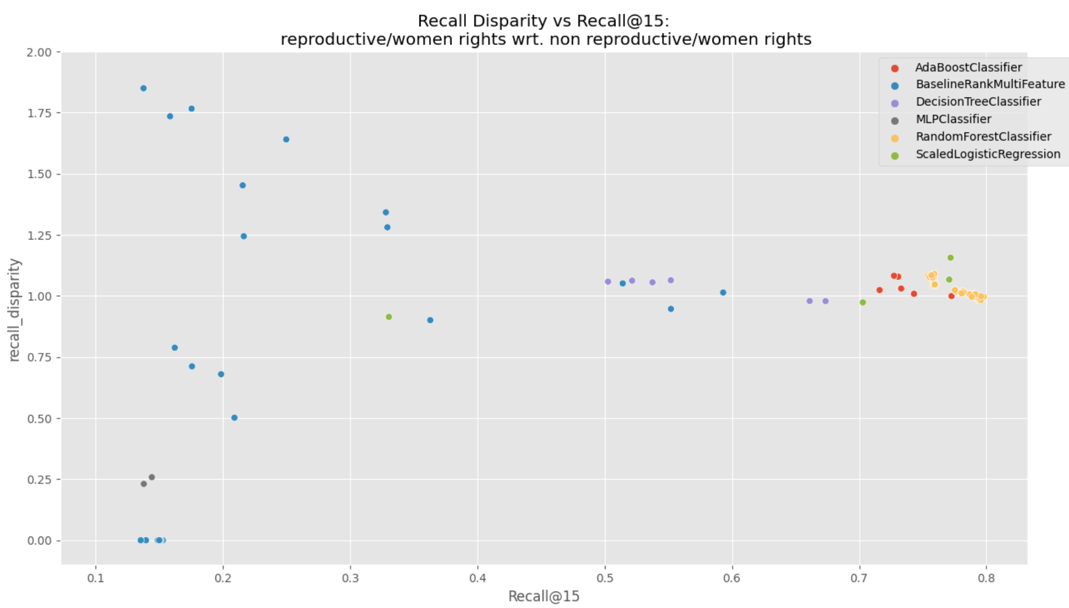

Figure 14: Recall Disparity vs Recall: Bills Related to Reproductive Rights/Women Rights vs. the Rest([Link to code]())

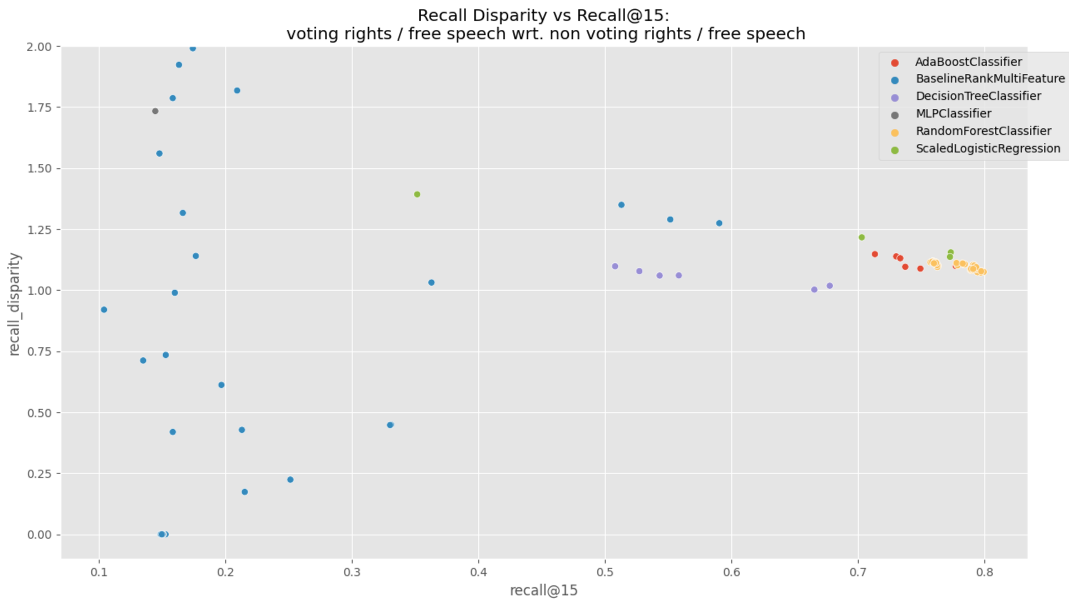

Figure 15: Recall Disparity vs Recall: Bills Related to Voting Rights/Free Speech vs. the Rest (Link to code)

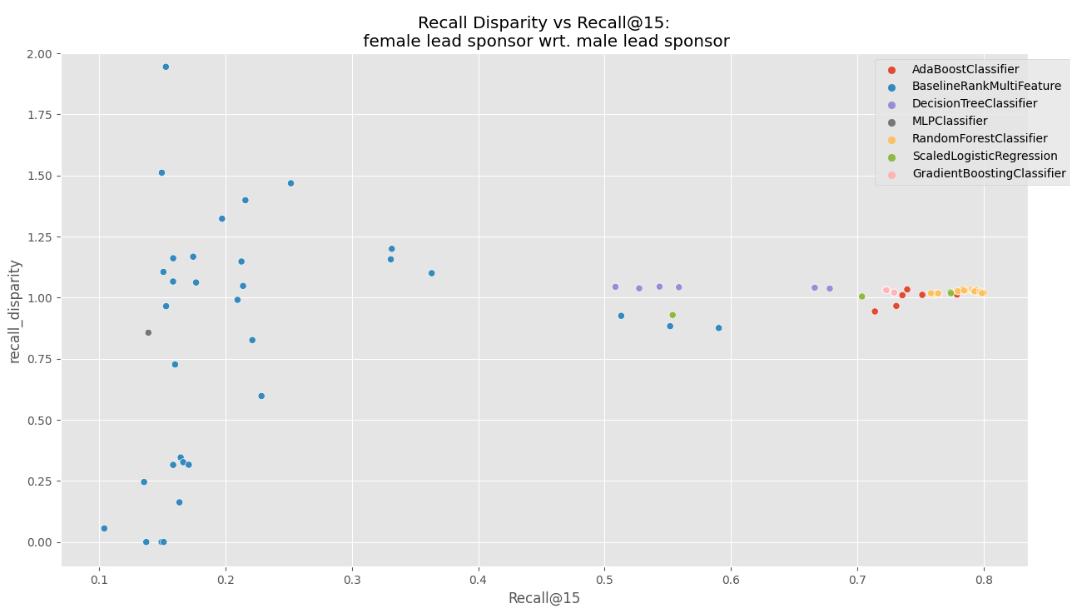

Figure 16: Bills with at least one female lead sponsor vs. the rest (Link to code)

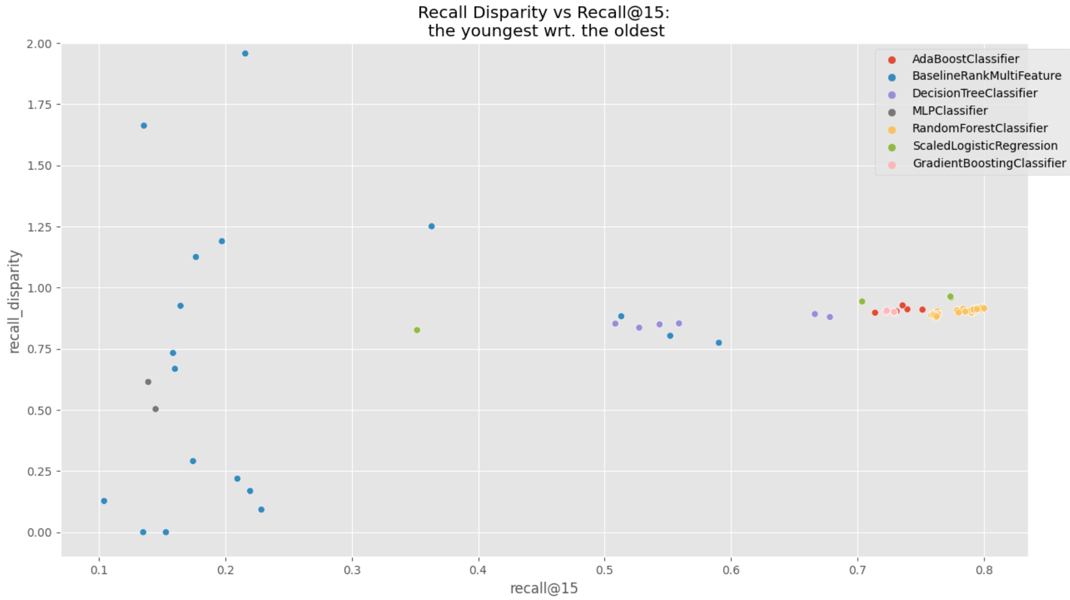

Figure 17: Recall Disparity vs Recall: Bills with the Youngest Lead Sponsors vs. those with the Oldest Lead Sponsors ([Link to code])

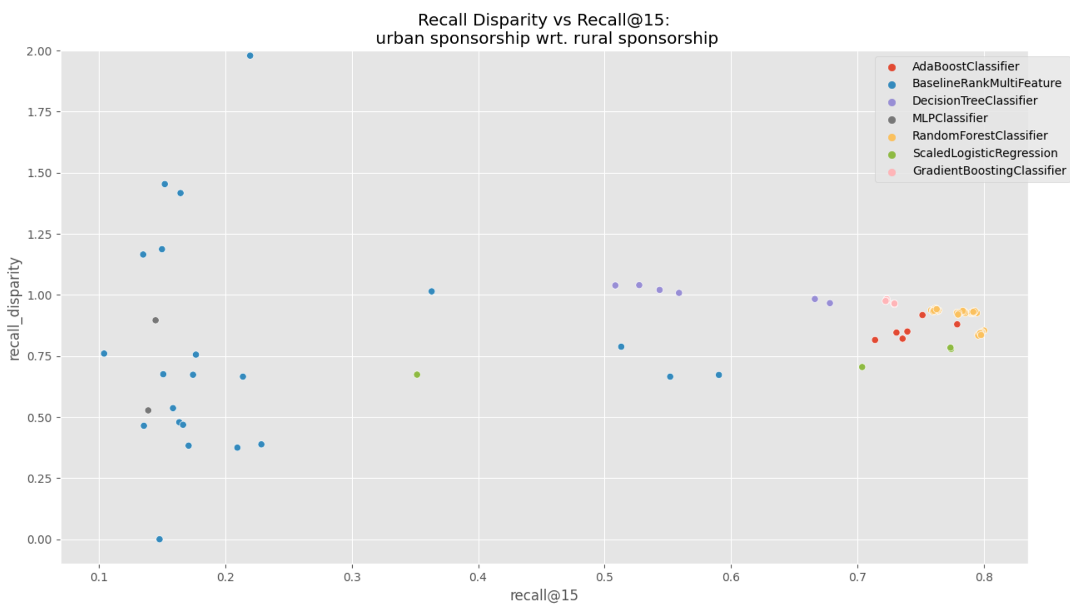

Figure 18: Recall Disparity vs Recall: Bills with a proportion of urban sponsors in the top percentile vs. bills with a proportion of urban sponsors in the bottom percentile ([Link to code])

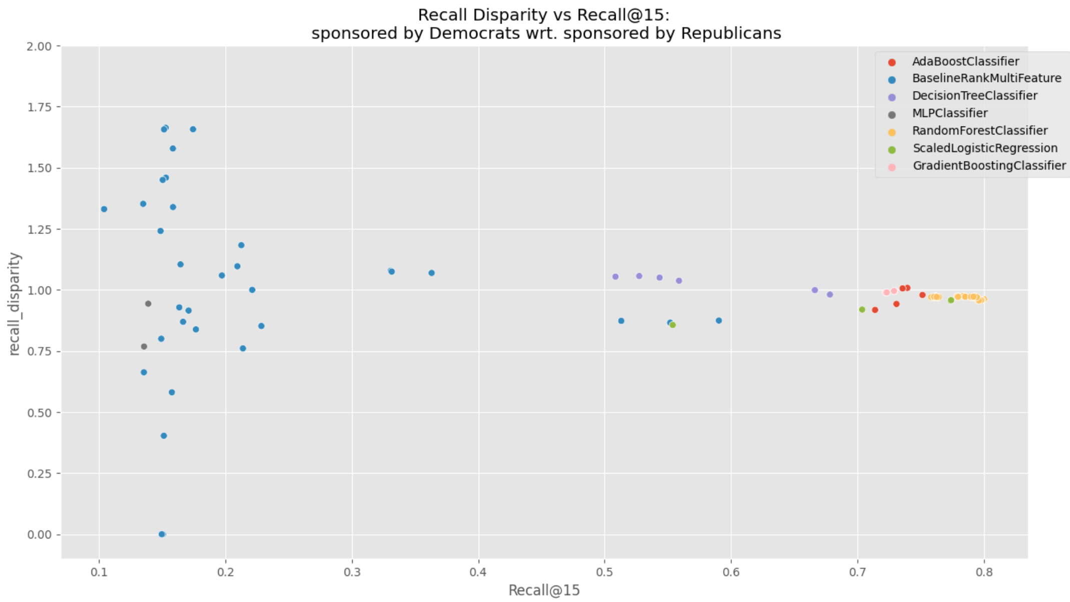

Figure 19: *Bills mainly sponsored by Democrats (i.e. more than half of sponsors are Democrats) vs. the rest* (*[Link to code]*)

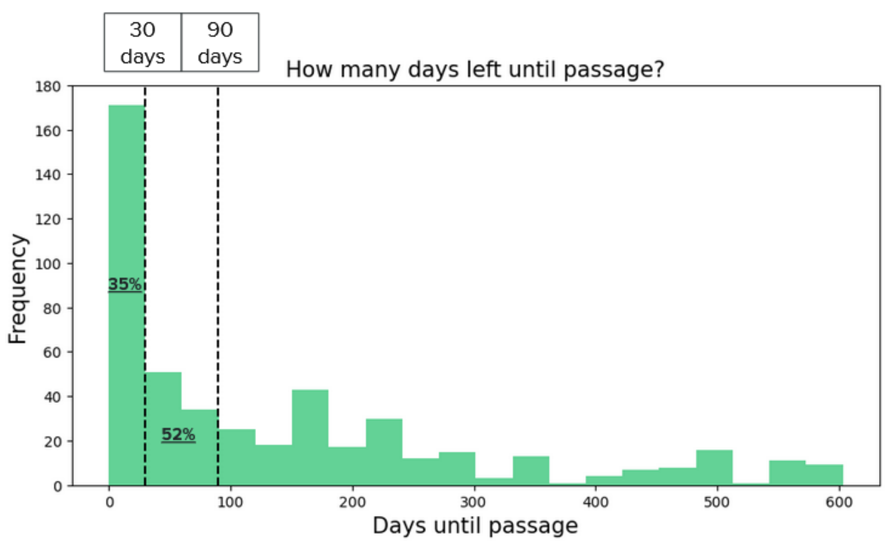

Figure 20: Duration between date of positive prediction and date of actual passage(*[Link to code]*)

Tables

ACPA Issue Category	# of bills	Share (percentage)
Capital Punishment	1639	7.3052
Voting Rights	724	3.2270
Criminal Justice	707	3.1512
Privacy	617	2.7500
Free Speech	569	2.5361
Reproductive Rights	385	1.7160
National Security	263	1.1722
Juvenile Justice	247	1.1009
Disability Rights	192	0.8558
Prisoner Rights	188	0.8379
Women Rights	177	0.7889
Religion	55	0.2451
Immigrant Rights	43	0.1917
LGBTQ	41	0.1827
Racial Justice	41	0.1827
HIV	12	0.0535
Human Rights	11	0.0490

Table 1: Number of bills and share of bills related to topics ACPA cares about

Baseline	Average Recall@15_pct
Categorize bills based on whether they have been referred to a committee	59.0%
Categorize bills based on whether they have passed one of the chambers	55.2%
Categorize bills based on whether they have passed out of a committee	51.3%
Rank bills by number of days since introduced	36.3%
Rank bills by number of sponsors in Senate	33.2%

Table 2: Five best-performing baseline models

Model	Hyperparameters	Average Recall@15_pct
Random Forest Classifier (Model group 441)	"max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 1000, "random_state": 123, "min_samples_split": 20	80%
Random Forest Classifier (Model group 442)	"max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 3000, "random_state": 123, "min_samples_split": 20	79.85%
Random Forest Classifier (Model group 443)	"max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 5000, "random_state": 123, "min_samples_split": 20	79.81%
Random Forest Classifier (Model group 438)	"max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 1000, "random_state": 123, "min_samples_split": 10	79.76%
Random Forest Classifier	"max_depth": 10,	79.73%

(Model group 435)	"class_weight": "balanced_subsample", "n_estimators": 1000, "random_state": 123, "min_samples_split": 5	

Table 3: Models with highest average recall@15

Feature	**Mean Value on top 15% of the predictions of your test set**	**Mean value on the bottom 85% of the predictions on the test set**
Number of times bills have passed out of a committee in the past month	0.103901	0.000030
Number of times bills have been voted on in the past month	0.450466	0.002497
Number of times bills have been engrossed in the past year	0.507249	0.005269
Number of times bills have been voted on in the past 6 months	1.783397	0.029571
Number of times bills have passed out of a committee in the past year	0.651536	0.012608
Number of times bills have been voted on in the past year	2.736624	0.091972

Table 4: Cross tabulation for validation set from 2016-12-31 to 2018-12-28 for the best performing random forest model with following hyperparameters: "max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 1000, "random_state": 123, "min_samples_split": 20. The predictors tabulated are the ones that differ the most in the top 15% of the predictions relative to the bottom 85%

Feature	Mean Value on top 15% of the predictions of your test set	Mean value on the bottom 85% of the predictions on the test set
Number of times bills have passed out of a committee in the past month	0.103901	0.000030
Number of times bills have been voted on in the past month	0.451156	0.002375
Number of times bills have been engrossed in the past year	0.507076	0.005299
Number of times bills have been voted on in the past 6 months	1.785813	0.029145
Number of times bills have passed out of a committee in the past year	0.653435	0.012273
Number of times bills have been voted on in the past year	2.740421	0.091302

Table 5: Cross tabulation for validation set from 2016-12-31 to 2018-12-28 for the best performing random forest model with following hyperparameters:max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 3000, "random_state": 123, "min_samples_split": 20. The predictors tabulated are the ones that differ the most in the top 15% of the predictions relative to the bottom 85%

Feature	Mean Value on top 15% of the predictions of your test set	Mean value on the bottom 85% of the predictions on the test set
Number of times bills have passed out of a committee in the past month	0.103901	0.000030
Number of times bills have been voted on in the past month	0.451501	0.002345

Number of times bills have been engrossed in the past year	0.506040	0.005482
Number of times bills have been voted on in the past 6 months	1.788057	0.0287489
Number of times bills have passed out of a committee in the past year	0.6527442	0.0123949
Number of times bills have been voted on in the past year	2.739731	0.091424

Table 6: Cross tabulation for validation set from 2016-12-31 to 2018-12-28 for the best performing random forest model with following hyperparameters: "max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 5000, "random_state": 123, "min_samples_split": 20. The predictors tabulated are the ones that differ the most in the top 15% of the predictions relative to the bottom 85%

Feature	**Mean Value on top 15% of the predictions of your test set**	**Mean value on the bottom 85% of the predictions on the test set**
Number of times bills have passed out of a committee in the past month	0.103901	0.000030
Number of times bills have been voted on in the past month	0.445633	0.003350
Number of times bills have been engrossed in the past year	0.506041	0.005482
Number of times bills have been voted on in the past 6 months	1.76545	0.0327384
Number of times bills have passed out of a committee in the past year	0.650846	0.012730

Number of times bills have been voted on in the past year	2.717984	0.095261

Table 7: Cross tabulation for validation set from 2016-12-31 to 2018-12-28 for the best performing random forest model with following hyperparameters:"max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 1000, "random_state": 123, "min_samples_split": 10. The predictors tabulated are the ones that differ the most in the top 15% of the predictions relative to the bottom 85%

Feature	**Mean Value on top 15% of the predictions of your test set**	**Mean value on the bottom 85% of the predictions on the test set**
Number of times bills have passed out of a committee in the past month	0.103901	0.000030
Number of times bills have been voted on in the past month	0.445608	0.003380
Number of times bills have been engrossed in the past year	0.5027615	0.0060604
Number of times bills have been voted on in the past 6 months	1.761477	0.0334389
Number of times bills have passed out of a committee in the past year	0.646531	0.0134913
Number of times bills have been voted on in the past year	2.703486	0.0978195

Table 8: Cross tabulation for validation set from 2016-12-31 to 2018-12-28 for the best performing random forest model with following hyperparameters:"max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 1000, "random_state": 123, "min_samples_split": 5. The predictors tabulated are the ones that differ the most in the top 15% of the predictions relative to the bottom 85%

Feature	Mean Value on top 15% of the predictions of your test set	Mean value on the bottom 85% of the predictions on the test set
Number of times bills have passed out of a committee in the past	0.980670	0.0019186
Number of times bills have passed out of a committee in the past 1 year	0.712116	0.0019186
Number of times bills have been engrossed in the past	0.727649	0.0034109
Number of times bills have been engrossed in the past 1 year	0.518295	0.0033195
Number of times bills have passed out of a committee in the past 1 month	0.095789	0.0014618
Number of times bills have been voted on in the past	4.158785	0.098398

Table 9: Cross tabulation for validation set from 2016-12-31 to 2018-12-28 for the best performing baseline model which Categorize bills based on whether they have been referred to a committee. The predictors tabulated are the ones that differ the most in the top 15% of the predictions relative to the bottom 85%

Model	Hyperparameters	Average Recall@15_pct	Average Recall disparity
sklearn.tree.DecisionTreeClassifier (Model group 413)	"max_depth": 10, "max_features": null, "min_samples_split": 10	66.6%	1.002

sklearn.tree.DecisionTreeClassifier (Model group 414)	"max_depth": 10, "max_features": null, "min_samples_split": 20	67.8%	1.018
Random Forest Classifier (Model group 435)	"max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 1000, "random_state": 123, "min_samples_split": 5	79.7%	1.072
Random Forest Classifier (Model group 441)	"max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 1000, "random_state": 123, "min_samples_split": 20	80%	1.075

Table 10: 5 best models based on recall-bias audit in terms of voting rights and free speech

Model	Hyperparameters	Average Recall@15_pct	Average Recall disparity
sklearn.tree.DecisionTreeClassifier (Model group 416)	"max_depth": 20, "max_features": null, "min_samples_split": 20	55.9%	1.008
sklearn.tree.DecisionTreeClassifier (Model group 413)	"max_depth": 10, "max_features": null, "min_samples_split": 10	66.6%	0.983
sklearn.ensemble.GradientBoostingClassifier (Model group 513)	"max_depth": 50, "subsample": 0.8, "max_features": "auto", "n_estimators": 4000, "random_state": 123, "learning_rate": 0.01, "min_samples_split": 5, "min_impurity_decrease": 0.01	72.9%	0.966
sklearn.ensemble.RandomForestClassifier (Model group 449)	"max_depth": 20, "class_weight": "balanced_subsample", "n_estimators": 5000, "random_state": 123, "min_samples_split": 10	79.2%	0.927
Random Forest	"max_depth": 10,	80%	0.854

Classifier (Model group 441)	"class_weight": "balanced_subsample", "n_estimators": 1000, "random_state": 123, "min_samples_split": 20		

Table 11: 5 best models based on recall-bias audit in terms of sponsors districts (urban vs rural)

Model	**Hyperparameters**	**Average Recall@15_pct**	**Average Recall disparity**
triage.component.catwalk.estimators.classifiers.ScaledLogisticRegression (Model group 421)	"C": 0.1, "penalty": "l1"	77.4%	0.964
sklearn.ensemble.RandomForestClassifier (Model group 443)	"max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 5000, "random_state": 123, "min_samples_split": 20	79.8%	0.920
sklearn.ensemble.RandomForestClassifier (Model group 442)	"max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 3000, "random_state": 123, "min_samples_split": 20	79.9%	0.918
Random Forest Classifier (Model group 441)	"max_depth": 10, "class_weight": "balanced_subsample", "n_estimators": 1000, "random_state": 123, "min_samples_split": 20	80%	0.915

Table 12: 5 best models based on recall-bias audit in terms of sponsors age category (the youngest vs the oldest)

Feature names	model_group_id 441	model_group_id 442	model_group_id 443	model_group_id 438	model_group_id 435
progress_dummies_entity_id_1year_bill_status_9_sum	0.08356	0.07817	0.07709	0.08577	0.08248
progress_dummies_entity_id_all_bill_status_9_sum	0.07929	0.07594	0.07539	0.08028	0.07873
days_since_entity_id_all_days_since_intro_min	0.06670	0.06678	0.06695	0.06773	0.06702
votes_entity_id_1 year_distinct vote_id_count	0.05780	0.06586	0.06833	0.05879	0.05869
votes_entity_id_all_distinct vote_id_count	0.05714	0.05509	0.05394	0.05362	0.05567
votes_entity_id_6 month_distinct vote_id_count	0.04471	0.04750	0.04854	0.04371	0.04371
progress_dummies_entity_id_all_bill_status_2_sum	0.04402	0.04550	0.04413	0.04196	0.04318
progress_dummies_entity_id_1year_bill_status_2_sum	0.04051	0.03932	0.03973	0.04178	0.04135
progress_dummies_entity_id_all_bill_status_10_sum	0.03340	0.03253	0.03199	0.03278	0.03269
progress_dummies_entity_id_1year_bill_status_10_sum	0.03296	0.03360	0.03500	0.03376	0.03369
progress_dummies_entity_id_1year_bill_status_1_sum	0.02999	0.03000	0.03035	0.03040	0.03055
progress_dummies_entity_id_1month_bill_status_1_sum	0.02095	0.02071	0.02047	0.02129	0.02249
progress_dummies_entity_id_1month_bill_status_9_sum	0.01945	0.02037	0.02033	0.01954	0.01930
votes_entity_id_1 month_distinct vote_id_count	0.01616	0.01563	0.01555	0.01600	0.01605
sponsors_entity_id_all_repr_sponsors_sum	0.01352	0.01256	0.01228	0.01367	0.01317
sponsors_entity_id_all_num_sponsors_count	0.01312	0.01315	0.01346	0.01272	0.01325
either_majority_sponsor_entity_id	0.01229	0.01165	0.01173	0.01189	0.01161

_all_either_maj_sponsor_avg					
passed_sponsors_entity_id_1year_num_passed_sum	0.01212	0.01274	0.01281	0.01231	0.01178
passed_sponsors_entity_id_6month_num_passed_sum	0.01101	0.01142	0.01148	0.01120	0.01122
passed_sponsors_entity_id_all_num_passed_sum	0.01088	0.01047	0.01024	0.01077	0.01076
either_majority_sponsor_entity_id_all_either_maj_sponsor_sum	0.01083	0.01106	0.01085	0.01132	0.01058
either_majority_sponsor_entity_id_all_either_maj_sponsor_max	0.00964	0.01062	0.01077	0.00886	0.00963
all_sponsor_fundraising_entity_id_all_cycle_raised_sum	0.00907	0.00898	0.00900	0.00910	0.00873
sponsors_entity_id_all_dem_sponsors_sum	0.00877	0.00872	0.00853	0.00841	0.00878
phi_pit_entity_id_all_num_sponsors_avg	0.00867	0.00860	0.00868	0.00875	0.00857
progress_dummies_entity_id_1month_bill_status_10_sum	0.00861	0.00788	0.00755	0.00829	0.00855
clusters_entity_id_all_cluster_5_d_1_sum	0.00812	0.00803	0.00807	0.00815	0.00788
all_sponsor_fundraising_entity_id_all_cycle_donors_sum	0.00727	0.00743	0.00753	0.00756	0.00751
passed_sponsors_entity_id_1month_num_passed_sum	0.00702	0.00697	0.00699	0.00681	0.00690
all_sponsor_fundraising_entity_id_all_cycle_raised_min	0.00655	0.00614	0.00612	0.00653	0.00657
passed_sponsors_entity_id_all_num_passed_max	0.00649	0.00622	0.00632	0.00628	0.00621
passed_sponsors_entity_id_2month_num_passed_sum	0.00641	0.00672	0.00669	0.00631	0.00634
clusters_entity_id_all_cluster_10_d_2_sum	0.00627	0.00659	0.00656	0.00625	0.00632
phi_pit_entity_id_all_num_sponsors_sum	0.00565	0.00562	0.00584	0.00564	0.00573
passed_sponsors_entity_id_1year_	0.00549	0.00552	0.00554	0.00553	0.00578

num_passed_max					
sponsored_by_sponsor_entity_id_all_sponsored_max	0.00530	0.00530	0.00533	0.00555	0.00535
passed_sponsors_entity_id_6month_num_passed_max	0.00496	0.00538	0.00527	0.00519	0.00510
progress_dummies_entity_id_1month_bill_status_2_sum	0.00492	0.00411	0.00378	0.00483	0.00509
passed_sponsors_entity_id_all_num_passed_avg	0.00487	0.00493	0.00504	0.00503	0.00509
sponsored_by_sponsor_entity_id_all_sponsored_avg	0.00465	0.00465	0.00467	0.00484	0.00491
sponsored_by_sponsor_entity_id_all_sponsored_min	0.00465	0.00466	0.00464	0.00470	0.00479
clusters_entity_id_all_cluster_5_d_0_sum	0.00457	0.00453	0.00456	0.00444	0.00454
all_sponsor_fundraising_entity_id_all_cycle_raised_max	0.00441	0.00455	0.00454	0.00443	0.00450
all_sponsor_fundraising_entity_id_all_cycle_raised_avg	0.00427	0.00425	0.00421	0.00435	0.00432
passed_sponsors_entity_id_1year_num_passed_avg	0.00424	0.00427	0.00428	0.00415	0.00422
bipartisan_sponsors_entity_id_all_bipartisanship_max	0.00406	0.00401	0.00397	0.00395	0.00382
all_sponsor_fundraising_entity_id_all_cycle_donors_avg	0.00405	0.00397	0.00399	0.00409	0.00430
passed_sponsors_entity_id_2month_num_passed_max	0.00403	0.00392	0.00386	0.00387	0.00392
clusters_entity_id_all_cluster_10_t_5_sum	0.00397	0.00408	0.00414	0.00398	0.00383
passed_sponsors_entity_id_1month_num_passed_avg	0.00391	0.00387	0.00387	0.00385	0.00401
clusters_entity_id_all_cluster_10_t_3_sum	0.00377	0.00375	0.00366	0.00366	0.00385
clusters_entity_id_all_cluster_5_d_2_sum	0.00374	0.00371	0.00372	0.00375	0.00364
passed_sponsors_entity_id_1mont	0.00339	0.00342	0.00344	0.00334	0.00334

h_num_passed_max					
all_sponsor_fundraising_entity_id_all_cycle_donors_min	0.00337	0.00349	0.00355	0.00342	0.00374
all_sponsor_fundraising_entity_id_all_cycle_donors_max	0.00306	0.00302	0.00299	0.00302	0.00299
sen_majority_sponsor_entity_id_all_sen_maj_sponsor_avg	0.00287	0.00259	0.00248	0.00267	0.00268
sen_majority_sponsor_entity_id_all_sen_maj_sponsor_max	0.00283	0.00255	0.00248	0.00263	0.00284
passed_sponsors_entity_id_6month_num_passed_avg	0.00274	0.00285	0.00289	0.00287	0.00290
passed_sponsors_entity_id_2month_num_passed_avg	0.00272	0.00277	0.00274	0.00276	0.00285
sponsored_by_lead2_entity_id_all_lead_sponsored_3mo_avg	0.00266	0.00253	0.00253	0.00267	0.00262
sponsored_by_lead2_entity_id_all_lead_sponsored_3mo_sum	0.00244	0.00259	0.00251	0.00255	0.00263
clusters_entity_id_all_cluster_10_t_7_sum	0.00240	0.00230	0.00231	0.00268	0.00258
sponsored_by_lead2_entity_id_all_lead_sponsored_3mo_min	0.00237	0.00241	0.00244	0.00224	0.00232
sponsored_by_lead2_entity_id_all_lead_sponsored_3mo_max	0.00236	0.00243	0.00243	0.00255	0.00269
clusters_entity_id_all_cluster_10_t_9_sum	0.00235	0.00217	0.00211	0.00238	0.00249
sponsored_by_lead2_entity_id_all_lead_sponsored_2yr_sum	0.00234	0.00233	0.00230	0.00227	0.00235
sponsored_by_lead2_entity_id_all_lead_sponsored_2yr_max	0.00232	0.00245	0.00238	0.00226	0.00230
sponsored_by_lead2_entity_id_all_lead_sponsored_2yr_min	0.00225	0.00225	0.00225	0.00219	0.00229
sen_majority_sponsor_entity_id_all_sen_maj_sponsor_sum	0.00223	0.00240	0.00227	0.00248	0.00240
sponsored_by_lead2_entity_id_all_lead_sponsored_1yr_max	0.00223	0.00204	0.00203	0.00213	0.00205
sponsored_by_lead2_entity_id_all	0.00213	0.00221	0.00223	0.00223	0.00226

_lead_sponsored_2yr_avg					
sponsored_by_lead2_entity_id_all_lead_sponsored_1yr_sum	0.00212	0.00202	0.00201	0.00202	0.00200
age_lead_sponsors_entity_id_all_age_avg	0.00209	0.00209	0.00205	0.00207	0.00209
sponsored_by_lead2_entity_id_all_lead_sponsored_all_avg	0.00208	0.00202	0.00202	0.00210	0.00197
rep_majority_sponsor_entity_id_all_rep_maj_sponsor_avg	0.00207	0.00230	0.00221	0.00220	0.00231
sponsored_by_lead2_entity_id_all_lead_sponsored_1yr_avg	0.00207	0.00204	0.00203	0.00203	0.00205
sponsored_by_lead2_entity_id_all_lead_sponsored_all_max	0.00204	0.00201	0.00201	0.00190	0.00194
sponsored_by_lead2_entity_id_all_lead_sponsored_1yr_min	0.00200	0.00193	0.00199	0.00185	0.00203
sponsored_by_lead2_entity_id_all_lead_sponsored_all_sum	0.00199	0.00204	0.00204	0.00197	0.00196
sponsored_by_lead2_entity_id_all_lead_sponsored_all_min	0.00198	0.00200	0.00198	0.00193	0.00201
age_lead_sponsors_entity_id_all_age_min	0.00194	0.00197	0.00205	0.00190	0.00204
clusters_entity_id_all_cluster_10_d_5_sum	0.00194	0.00197	0.00196	0.00196	0.00190
rep_majority_sponsor_entity_id_all_rep_maj_sponsor_max	0.00189	0.00198	0.00197	0.00174	0.00164
age_lead_sponsors_entity_id_all_age_max	0.00188	0.00197	0.00200	0.00199	0.00215
rep_majority_sponsor_entity_id_all_rep_maj_sponsor_sum	0.00174	0.00197	0.00193	0.00189	0.00177
clusters_entity_id_all_cluster_10_d_7_sum	0.00130	0.00128	0.00130	0.00121	0.00115
lead_sponsor_senority_entity_id_all_num_years_served_min	0.00128	0.00131	0.00130	0.00131	0.00141
lead_sponsor_senority_entity_id_all_num_years_served_avg	0.00127	0.00127	0.00127	0.00123	0.00133
lead_sponsor_senority_entity_id_a	0.00125	0.00126	0.00127	0.00125	0.00128

ll_num_years_served_max					
clusters_entity_id_all_cluster_5_t_4_sum	0.00121	0.00150	0.00150	0.00131	0.00141
clusters_entity_id_all_cluster_5_t_3_sum	0.00112	0.00111	0.00113	0.00091	0.00104
clusters_entity_id_all_cluster_10_d_9_sum	0.00098	0.00097	0.00096	0.00099	0.00092
lead_sponsor_fundraising_entity_id_all_cycle_donors_avg	0.00088	0.00086	0.00087	0.00093	0.00088
lead_sponsor_fundraising_entity_id_all_cycle_raised_avg	0.00087	0.00086	0.00085	0.00086	0.00081
lead_sponsor_fundraising_entity_id_all_cycle_raised_sum	0.00087	0.00085	0.00084	0.00083	0.00092
clusters_entity_id_all_cluster_5_d_4_sum	0.00087	0.00095	0.00096	0.00091	0.00092
lead_sponsor_fundraising_entity_id_all_cycle_raised_min	0.00086	0.00085	0.00086	0.00094	0.00099
capital_punishment_entity_id_all_capital_punishment_max	0.00085	0.00087	0.00089	0.00090	0.00085
lead_sponsor_fundraising_entity_id_all_cycle_donors_sum	0.00084	0.00086	0.00087	0.00079	0.00083
lead_sponsor_fundraising_entity_id_all_cycle_raised_max	0.00084	0.00088	0.00087	0.00092	0.00083
lead_sponsor_fundraising_entity_id_all_cycle_donors_min	0.00084	0.00080	0.00083	0.00086	0.00078
lead_sponsor_fundraising_entity_id_all_cycle_donors_max	0.00081	0.00083	0.00083	0.00087	0.00095
clusters_entity_id_all_cluster_5_t_0_sum	0.00067	0.00068	0.00069	0.00069	0.00071
clusters_entity_id_all_cluster_10_t_4_sum	0.00067	0.00069	0.00068	0.00056	0.00065
clusters_entity_id_all_cluster_10_d_1_sum	0.00065	0.00071	0.00075	0.00064	0.00065
clusters_entity_id_all_cluster_5_t_1_sum	0.00052	0.00049	0.00051	0.00054	0.00053
clusters_entity_id_all_cluster_10_t	0.00051	0.00055	0.00052	0.00058	0.00055

_0_sum					
clusters_entity_id_all_cluster_10_d_8_sum	0.00051	0.00052	0.00049	0.00051	0.00045
clusters_entity_id_all_cluster_10_d_3_sum	0.00047	0.00044	0.00045	0.00043	0.00043
clusters_entity_id_all_cluster_5_t_2_sum	0.00047	0.00050	0.00050	0.00040	0.00038
phi_pit_entity_id_all_num_sponsors_max	0.00045	0.00045	0.00048	0.00053	0.00049
clusters_entity_id_all_cluster_10_t_6_sum	0.00043	0.00047	0.00046	0.00045	0.00041
passed_sponsors_entity_id_1month_num_passed_imp	0.00041	0.00037	0.00038	0.00036	0.00046
clusters_entity_id_all_cluster_10_t_8_sum	0.00034	0.00033	0.00035	0.00037	0.00036
lead_sponsor_m_entity_id_all_men_sum	0.00034	0.00033	0.00035	0.00035	0.00037
lead_sponsor_m_entity_id_all_men_max	0.00034	0.00034	0.00034	0.00032	0.00035
lead_sponsor_m_entity_id_all_men_avg	0.00033	0.00033	0.00035	0.00038	0.00031
clusters_entity_id_all_cluster_10_d_4_sum	0.00032	0.00034	0.00034	0.00032	0.00037
lead_sponsor_m_entity_id_all_men_min	0.00032	0.00034	0.00033	0.00035	0.00032
govenor_party_entity_id_all_gov_party_max	0.00032	0.00036	0.00037	0.00046	0.00038
privacy_entity_id_all_privacy_max	0.00031	0.00033	0.00031	0.00032	0.00029
sponsored_by_lead2_entity_id_all_lead_sponsored_2yr_imp	0.00028	0.00029	0.00026	0.00033	0.00030
either_majority_sponsor_entity_id_all_either_maj_sponsor_imp	0.00028	0.00024	0.00024	0.00029	0.00027
criminal_justice_entity_id_all_criminal_justice_max	0.00027	0.00029	0.00030	0.00029	0.00033
clusters_entity_id_all_cluster_10_t	0.00024	0.00025	0.00024	0.00024	0.00025

_2_sum					
lead_sponsor_ranking_member_entity_id_all_is_ranking_member_imp	0.00023	0.00020	0.00019	0.00022	0.00022
lead_position_entity_id_1year_positions_encoded__NULL_sum	0.00023	0.00020	0.00018	0.00019	0.00017
sponsored_by_lead2_entity_id_all_lead_sponsored_1yr_imp	0.00023	0.00021	0.00020	0.00020	0.00026
lead_sponsor_senority_entity_id_all_num_years_served_imp	0.00022	0.00019	0.00018	0.00015	0.00021
national_security_entity_id_all_national_security_max	0.00022	0.00020	0.00021	0.00023	0.00021
rep_majority_sponsor_entity_id_all_rep_maj_sponsor_imp	0.00022	0.00025	0.00023	0.00029	0.00023
sponsored_by_lead2_entity_id_all_lead_sponsored_all_imp	0.00022	0.00023	0.00024	0.00031	0.00035
sen_majority_sponsor_entity_id_all_sen_maj_sponsor_imp	0.00021	0.00022	0.00023	0.00026	0.00030
clusters_entity_id_all_cluster_10_d_6_sum	0.00021	0.00019	0.00018	0.00016	0.00017
lead_sponsor_m_entity_id_all_men_imp	0.00020	0.00019	0.00019	0.00017	0.00018
progress_dummies_entity_id_1month_bill_status_3_sum	0.00020	0.00018	0.00016	0.00018	0.00019
lead_sponsor_first_session_entity_id_all_first_sessio_imp	0.00020	0.00018	0.00018	0.00019	0.00018
reproductive_rights_entity_id_all_reproductive_rights_max	0.00019	0.00017	0.00017	0.00016	0.00016
age_lead_sponsors_entity_id_all_age_imp	0.00018	0.00019	0.00020	0.00022	0.00019
lead_sponsor_m_entity_id_all_me_imp	0.00018	0.00020	0.00020	0.00018	0.00016
sponsored_by_lead2_entity_id_all_lead_sponsored_3mo_imp	0.00018	0.00022	0.00024	0.00015	0.00022
lead_position_entity_id_1year_positions_encoded_16_sum	0.00017	0.00017	0.00017	0.00021	0.00016

clusters_entity_id_all_cluster_5_d_3_sum	0.00017	0.00016	0.00016	0.00015	0.00018
lead_sponsor_fundraising_entity_id_all_cycle_donors_imp	0.00016	0.00016	0.00015	0.00014	0.00014
progress_dummies_entity_id_all_bill_status_3_sum	0.00015	0.00016	0.00015	0.00016	0.00015
lead_position_entity_id_all_positions_encoded__NULL_sum	0.00015	0.00015	0.00015	0.00014	0.00014
progress_dummies_entity_id_1year_bill_status_3_sum	0.00014	0.00013	0.00015	0.00016	0.00016
lead_sponsor_first_session_entity_id_all_first_session_imp	0.00014	0.00016	0.00018	0.00018	0.00015
lead_sponsor_fundraising_entity_id_all_cycle_raised_imp	0.00014	0.00013	0.00013	0.00015	0.00013
lead_sponsor_ranking_member_entity_id_all_is_ranking_member_avg	0.00013	0.00014	0.00014	0.00014	0.00016
phi_pit_entity_id_all_num_sponsor_imp	0.00013	0.00015	0.00017	0.00009	0.00011
juvenile_justice_entity_id_all_juvenile_justice_max	0.00013	0.00013	0.00013	0.00012	0.00014
phi_pit_entity_id_all_num_sponsors_imp	0.00013	0.00015	0.00015	0.00013	0.00009
lead_sponsor_fundraising_entity_id_all_cycle_donor_imp	0.00013	0.00014	0.00014	0.00016	0.00015
lead_sponsor_ranking_member_entity_id_all_is_ranking_member_sum	0.00013	0.00015	0.00015	0.00015	0.00015
voting_rights_entity_id_all_voting_rights_max	0.00012	0.00012	0.00011	0.00011	0.00014
lead_sponsor_ranking_member_entity_id_all_is_ranking_member_max	0.00012	0.00014	0.00014	0.00011	0.00016
lead_position_entity_id_all_positions_encoded_10_sum	0.00012	0.00012	0.00012	0.00014	0.00013
lead_position_entity_id_all_positions_encoded_16_sum	0.00012	0.00012	0.00012	0.00012	0.00013

lead_sponsor_first_session_entity_id_all_first_session_min	0.00012	0.00011	0.00011	0.00010	0.00010
all_sponsor_fundraising_entity_id_all_cycle_raised_imp	0.00012	0.00011	0.00010	0.00012	0.00014
free_speech_entity_id_all_free_speech_max	0.00011	0.00012	0.00012	0.00010	0.00012
racial_justice_entity_id_all_racial_justice_max	0.00010	0.00010	0.00011	0.00010	0.00009
lead_sponsor_first_session_entity_id_all_first_session_max	0.00010	0.00010	0.00011	0.00010	0.00010
clusters_entity_id_all_cluster_10_t_1_sum	0.00010	0.00011	0.00012	0.00011	0.00011
all_sponsor_fundraising_entity_id_all_cycle_donor_imp	0.00010	0.00009	0.00010	0.00010	0.00008
clusters_entity_id_all_cluster_10_d_0_sum	0.00009	0.00010	0.00010	0.00011	0.00008
lead_sponsor_first_session_entity_id_all_first_session_sum	0.00009	0.00010	0.00010	0.00012	0.00009
lead_sponsor_first_session_entity_id_all_first_session_avg	0.00009	0.00010	0.00009	0.00009	0.00012
disability_rights_entity_id_all_disability_rights_max	0.00009	0.00008	0.00008	0.00007	0.00007
all_sponsor_fundraising_entity_id_all_cycle_donors_imp	0.00008	0.00010	0.00010	0.00008	0.00012
lead_position_entity_id_1year_positions_encoded_10_sum	0.00007	0.00007	0.00007	0.00008	0.00008
women_rights_entity_id_all_women_rights_max	0.00006	0.00006	0.00006	0.00007	0.00007
lead_position_entity_id_1month_positions_encoded__NULL_sum	0.00006	0.00005	0.00005	0.00007	0.00005
lead_position_entity_id_all_positions_encoded_1_sum	0.00006	0.00006	0.00006	0.00006	0.00005
prisoner_rights_entity_id_all_prisoner_rights_max	0.00005	0.00007	0.00007	0.00007	0.00007
lead_position_entity_id_1year_positions_encoded_5_sum	0.00005	0.00004	0.00004	0.00006	0.00007

lead_position_entity_id_all_positions_encoded_13_sum	0.00005	0.00005	0.00005	0.00005	0.00004
religion_entity_id_all_religion_max	0.00005	0.00004	0.00004	0.00005	0.00004
lead_position_entity_id_1year_positions_encoded_1_sum	0.00004	0.00004	0.00004	0.00004	0.00004
lead_position_entity_id_all_positions_encoded_5_sum	0.00004	0.00004	0.00004	0.00003	0.00004
lead_position_entity_id_all_positions_encoded_7_sum	0.00004	0.00003	0.00003	0.00003	0.00003
passed_sponsors_entity_id_2month_num_passed_imp	0.00003	0.00003	0.00002	0.00002	0.00002
immigrant_rights_entity_id_all_immigrant_rights_max	0.00002	0.00002	0.00002	0.00001	0.00001
lgbtq_rights_entity_id_all_lgbtq_rights_max	0.00002	0.00002	0.00002	0.00002	0.00003
lead_position_entity_id_1year_positions_encoded_7_sum	0.00002	0.00002	0.00002	0.00002	0.00002
lead_position_entity_id_all_positions_encoded_11_sum	0.00002	0.00002	0.00002	0.00002	0.00002
lead_position_entity_id_1year_positions_encoded_13_sum	0.00002	0.00002	0.00002	0.00003	0.00001
lead_position_entity_id_all_positions_encoded_21_sum	0.00002	0.00002	0.00002	0.00002	0.00002
lead_position_entity_id_1year_positions_encoded_21_sum	0.00001	0.00001	0.00001	0.00001	0.00002
lead_position_entity_id_1month_positions_encoded_16_sum	0.00001	0.00001	0.00001	0.00002	0.00003
lead_position_entity_id_1year_positions_encoded_14_sum	0.00001	0.00001	0.00001	0.00001	0.00001
lead_position_entity_id_all_positions_encoded_15_sum	0.00001	0.00001	0.00001	0.00001	0.00001
lead_position_entity_id_all_positions_encoded_3_sum	0.00001	0.00001	0.00001	0.00001	0.00002
lead_position_entity_id_all_positions_encoded_9_sum	0.00001	0.00001	0.00001	0.00001	0.00001

lead_position_entity_id_all_positions_encoded_14_sum	0.00001	0.00001	0.00001	0.00001	0.00001
human_rights_entity_id_all_human_rights_max	0.00001	0.00001	0.00001	0.00001	0.00001
lead_position_entity_id_1year_positions_encoded_15_sum	0.00001	0.00001	0.00001	0.00001	0.00001
lead_position_entity_id_1year_positions_encoded_11_sum	0.00001	0.00001	0.00001	0.00001	0.00001
lead_position_entity_id_1month_positions_encoded_5_sum	0.00001	0.00000	0.00000	0.00000	0.00001
lead_position_entity_id_1month_positions_encoded_14_sum	0.00001	0.00000	0.00000	0.00001	0.00001
lead_position_entity_id_1year_positions_encoded_4_sum	0.00001	0.00000	0.00000	0.00001	0.00000
lead_position_entity_id_1year_positions_encoded_3_sum	0.00001	0.00001	0.00001	0.00001	0.00001
lead_position_entity_id_1month_positions_encoded_21_sum	0.00001	0.00000	0.00000	0.00001	0.00000
lead_position_entity_id_1month_positions_encoded_1_sum	0.00001	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_all_positions_encoded_4_sum	0.00000	0.00001	0.00000	0.00001	0.00001
lead_position_entity_id_1year_positions_encoded_9_sum	0.00000	0.00000	0.00000	0.00001	0.00000
clusters_entity_id_all_cluster_10_d__NULL_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_all_positions_encoded_22_sum	0.00000	0.00000	0.00000	0.00000	0.00000
immigrant_rights_entity_id_all_immigrant_rights_imp	0.00000	0.00000	0.00000	0.00000	0.00000
disability_rights_entity_id_all_disability_rights_imp	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_15_sum	0.00000	0.00000	0.00000	0.00000	0.00000
capital_punishment_entity_id_all_capital_punishment_imp	0.00000	0.00000	0.00000	0.00000	0.00000

national_security_entity_id_all_national_security_imp	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_7_sum	0.00000	0.00000	0.00000	0.00000	0.00000
clusters_entity_id_all_cluster_10_t__NULL_sum	0.00000	0.00000	0.00000	0.00000	0.00000
privacy_entity_id_all_privacy_imp	0.00000	0.00000	0.00000	0.00000	0.00000
free_speech_entity_id_all_free_speech_imp	0.00000	0.00000	0.00000	0.00000	0.00000
reproductive_rights_entity_id_all_reproductive_rights_imp	0.00000	0.00000	0.00000	0.00000	0.00000
women_rights_entity_id_all_women_rights_imp	0.00000	0.00000	0.00000	0.00000	0.00000
religion_entity_id_all_religion_imp	0.00000	0.00000	0.00000	0.00000	0.00000
human_rights_entity_id_all_human_rights_imp	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_all_positions_encoded_6_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_10_sum	0.00000	0.00000	0.00000	0.00000	0.00000
voting_rights_entity_id_all_voting_rights_imp	0.00000	0.00000	0.00000	0.00000	0.00000
racial_justice_entity_id_all_racial_justice_imp	0.00000	0.00000	0.00000	0.00000	0.00000
clusters_entity_id_all_cluster_5_d__NULL_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lgbtq_rights_entity_id_all_lgbtq_rights_imp	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1year_positions_encoded_22_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_4_sum	0.00000	0.00000	0.00000	0.00000	0.00000
hiv_entity_id_all_hiv_imp	0.00000	0.00000	0.00000	0.00000	0.00000
prisoner_rights_entity_id_all_prisoner_rights_imp	0.00000	0.00000	0.00000	0.00000	0.00000

clusters_entity_id_all_cluster_5_t__NULL_sum	0.00000	0.00000	0.00000	0.00000	0.00000
juvenile_justice_entity_id_all_juvenile_justice_imp	0.00000	0.00000	0.00000	0.00000	0.00000
criminal_justice_entity_id_all_criminal_justice_imp	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1year_positions_encoded_6_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_6_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_3_sum	0.00000	0.00000	0.00000	0.00000	0.00000
hiv_entity_id_all_hiv_max	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_11_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_9_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_13_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_all_positions_encoded_19_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_22_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_all_positions_encoded_18_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_18_sum	0.00000	0.00000	0.00000	0.00000	0.00000
passed_sponsors_entity_id_6month_num_passed_imp	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_all_bill_status__NULL_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_all_positions_encoded_20_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1year_positions_encoded_0_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_all_b	0.00000	0.00000	0.00000	0.00000	0.00000

ill_status_8_sum					
progress_dummies_entity_id_all_bill_status_7_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_8_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_all_bill_status_6_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_all_bill_status_5_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_all_bill_status_4_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_all_positions_encoded_23_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1year_positions_encoded_12_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_all_bill_status_1_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_2_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_23_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_all_bill_status_11_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_all_positions_encoded_2_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_20_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_1year_bill_status__NULL_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_19_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_all_positions_encoded_17_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_17_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_1yea	0.00000	0.00000	0.00000	0.00000	0.00000

r_bill_status_8_sum					
progress_dummies_entity_id_1year_bill_status_7_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_1year_bill_status_6_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_1year_bill_status_5_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_12_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_1year_bill_status_4_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1year_positions_encoded_17_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1month_positions_encoded_0_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1year_positions_encoded_18_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_1year_bill_status_11_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1year_positions_encoded_19_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_all_positions_encoded_8_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_1month_bill_status__NULL_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_all_positions_encoded_12_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_1month_bill_status_8_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_1month_bill_status_7_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_1month_bill_status_6_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_1month_bill_status_5_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_1mo	0.00000	0.00000	0.00000	0.00000	0.00000

nth_bill_status_4_sum					
lead_position_entity_id_1year_positions_encoded_20_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_all_positions_encoded_0_sum	0.00000	0.00000	0.00000	0.00000	0.00000
progress_dummies_entity_id_1month_bill_status_11_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1year_positions_encoded_23_sum	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1year_positions_encoded_2_sum	0.00000	0.00000	0.00000	0.00000	0.00000
amendments_entity_id_all_num_amendments_imp	0.00000	0.00000	0.00000	0.00000	0.00000
lead_position_entity_id_1year_positions_encoded_8_sum	0.00000	0.00000	0.00000	0.00000	0.00000
amendments_entity_id_all_num_amendments_count	0.00000	0.00000	0.00000	0.00000	0.00000

Table 13: Feature importance of 5 best models

	passed	Not passed
Number of words	66526.02	73331.55
Number of sponsors	33.44	17.55
Days left	411.54	487.48
Number of amendments	*	*
Number of events	1.81	1.21
Count of the word 'budget'	4.14%	6.44%
Length of title	24.00	22.97
Control of House (1 when Democrats controlled House)	15.37%	18.48%
Control of Senate	**0	**0
Party of Governor	15.23%	19.02%
Bipartisan Sponsors	84.16%	65.70%

Table 14: Cross-Tabulation of Possible Predictors during Data Exploration

2016-12-31	2018-12-28
2016-06-30	2018-06-28
2015-12-30	2017-12-28
2015-06-30	2017-06-28
2014-12-30	2016-12-28
2014-06-30	2016-06-28
2013-12-30	2015-12-28
2013-06-30	2015-06-28
2012-12-30	2014-12-28

2012-06-30	2014-06-28
2011-12-30	2013-12-28
2011-06-30	2013-06-28

Table 15: Validation Sets Used

2009-02-28	2014-12-31
2009-02-28	2014-06-30
2009-02-28	2013-12-30
2009-02-28	2013-06-30
2009-02-28	2012-12-30
2009-02-28	2012-06-30
2009-02-28	2011-12-30
2009-02-28	2011-06-30
2009-02-28	2010-12-30
2009-02-28	2010-06-30
2009-02-28	2009-12-30
2009-02-28	2009-06-30

Table 16: Training Sets Used