Predicting Legislative Outcomes in Arizona

Dennis Chen, Xiaoyun Chen, Meghan Holquist, & Niharika Patil
Machine Learning for Public Policy

“Support Our Law Enforcement and Safe Neighborhoods Act”

- Police allowed to investigate an individual’s immigration status on the basis of “reasonable suspicion”
- One of the broadest and strictest immigration laws of its time (2010)
- Faced significant legal challenges from civil rights groups
- Led to significant racial profiling

Newman, P. (2017). Arizona’s Anti-Immigration Law and the Pervasiveness of Racial Profiling. *Georgetown Immigration Law Journal*, *31*(3).

Agenda

1. Problem Background and Goals
2. Data Exploration
3. Problem Formulation
4. Machine Learning Approach
5. Limitations, Recommendations, and Future Work

Problem Background and Goals

American Civic Protection Association (ACPA) Seeks To Protect Citizens' Civil Rights

Bills that may **infringe on civil liberties are regularly considered** by state legislatures

But **ACPA lacks the resources** to target advocacy efforts and legal challenges **for every single bill**

Due to resource constraints, ACPA can **only focus on 15% of bills** introduced within a legislative session

Project Goals

01	**Efficiency**	• Improve targeting of ACPA's lobbying and legal efforts
02	**Effectiveness**	• Protect vulnerable groups' civil liberties
03	**Equity**	• Protect civil liberties fairly across vulnerable groups

Potential Policy Impact and Tradeoffs

- **Increased Advocacy Success**
  - By predicting which bills are likely to pass, ACPA can better prepare for and challenge such legislation, potentially leading to more successful advocacy outcomes.
- **Resource Allocation**
  - Efficient allocation of ACPA's resources towards bills that pose the greatest threat to civil liberties.
- **Broader Protection**
  - Ensuring a more equitable focus by valuing those bills that are civil rights bills so that we can look into how we may be impacting individuals' civil rights

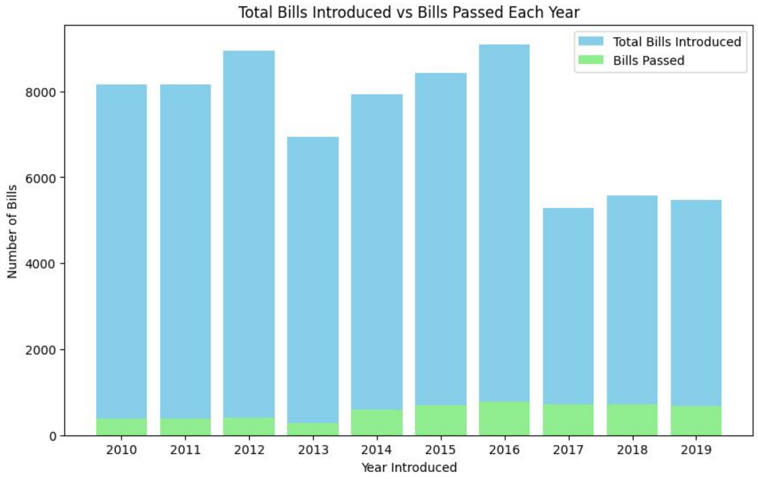
Efficiency/equity/effectiveness trade-offs:

If we focus on efficiency, we may fail to do the main job of the ACPA , which is to protect citizens' rights. However, if we focus on effectiveness, we may not create a product that ACPA's time and resources can cover.

Data Exploration

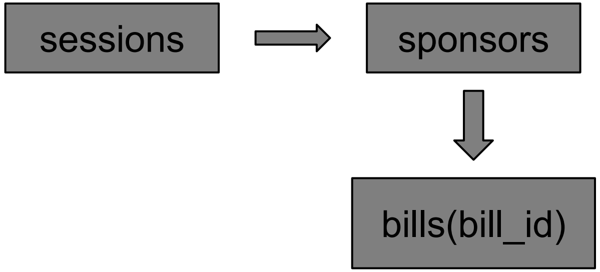
LegiScan Collects Data On Bills Across The U.S.

Comprehensive Legislative Data: The dataset includes detailed information on state legislative sessions, bills, amendments, and voting records, making it very relevant for analyzing legislative processes and outcomes.



Datasets:

- sessions, sessions_people, bill_sponsors
- bills, bill_texts
- bill_votes, bill_events, bill_progress, bill_amendments

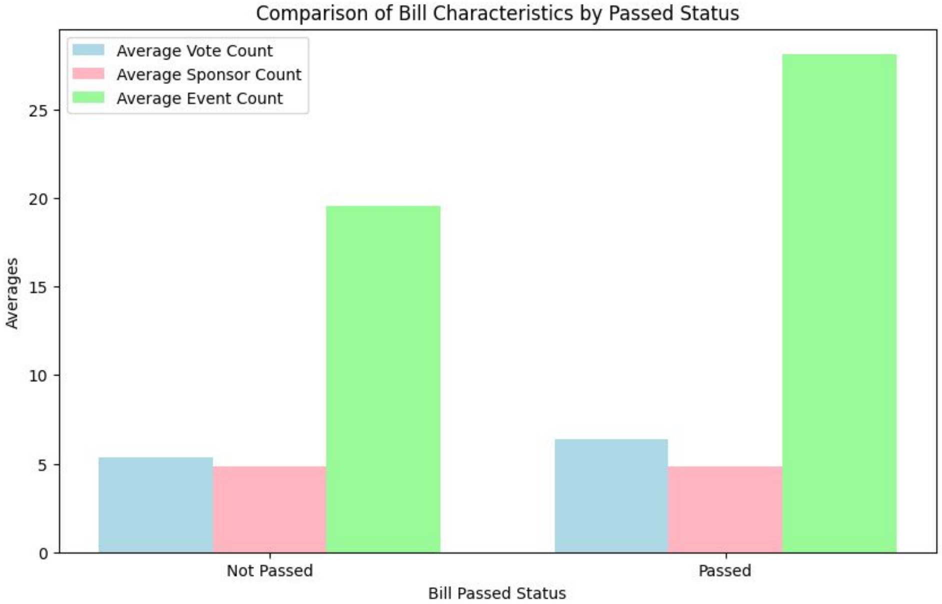


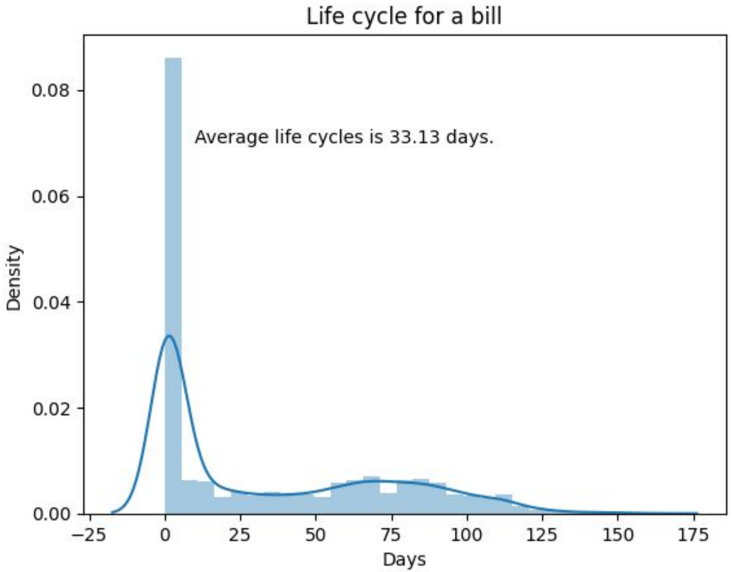
bill_status:

- Introduced
- Engrossed
- Enrolled
- Override
- Refer
- Passed
- Vetoed
- Chapterred

……

Initial Data Exploration





Problem Formulation

At the end of every month, **for each bill introduced in the state of Arizona,** within the current session year (that has not yet passed or been vetoed), can we **identify the 15% of bills with the highest likelihood of passing into law** by the end of the session to **prioritize targeting of advocacy and legal efforts**?

Predicting Whether Or Not A Bill Will Pass

What Are We Predicting On?

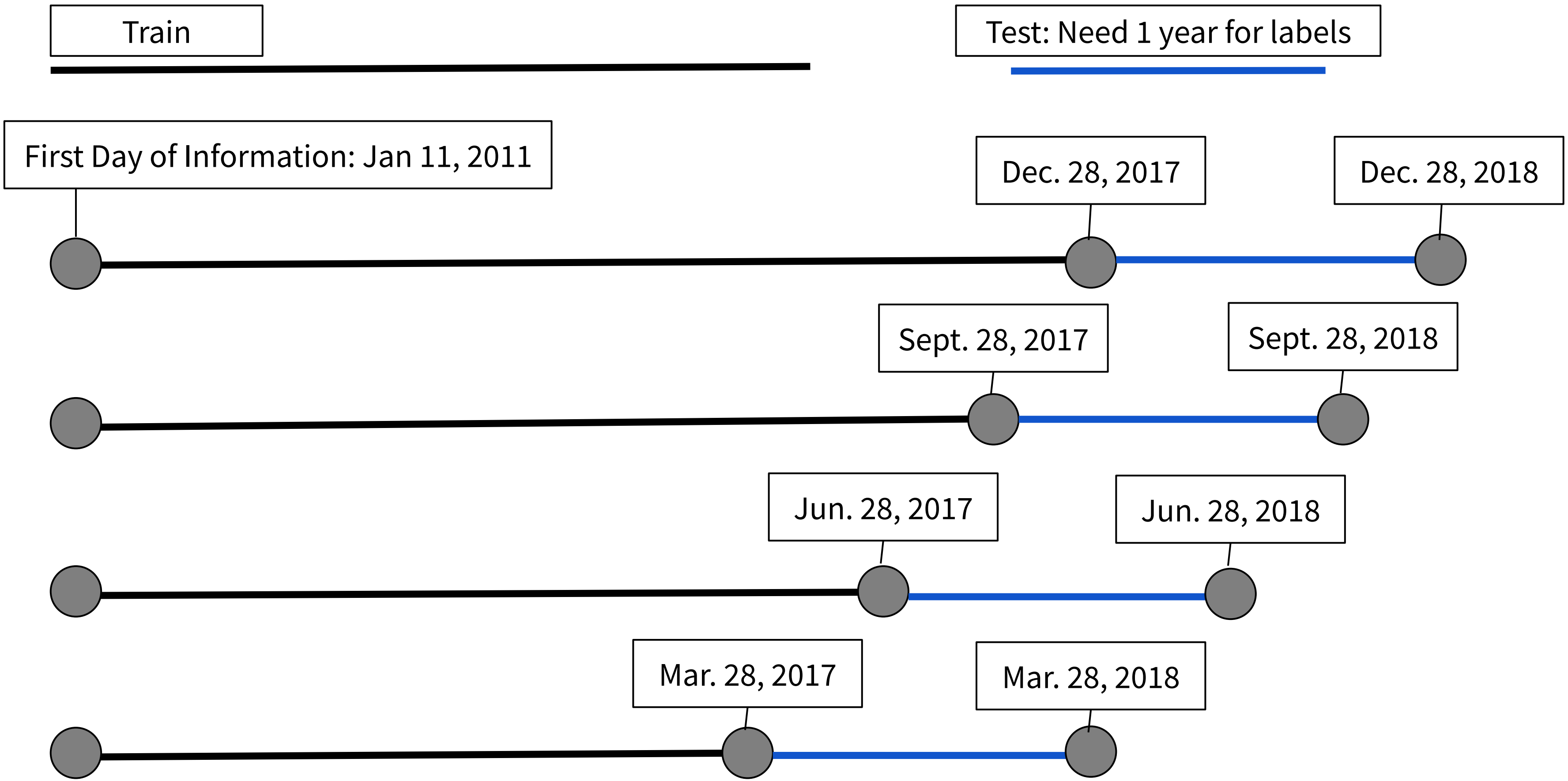
- Bills that have been introduced this session and not yet passed or been vetoed
- Outcome: Whether or not the bill passes by the end of the session
  - 1: Passes by end of session
  - 0: Does not pass by end of session

What Information Are We Using?

- Varied across time intervals
  - Lifecycle of a bill
  - Bill sponsor information
  - Bill topic

Machine Learning Approach

Models Used All Information Available To Predict

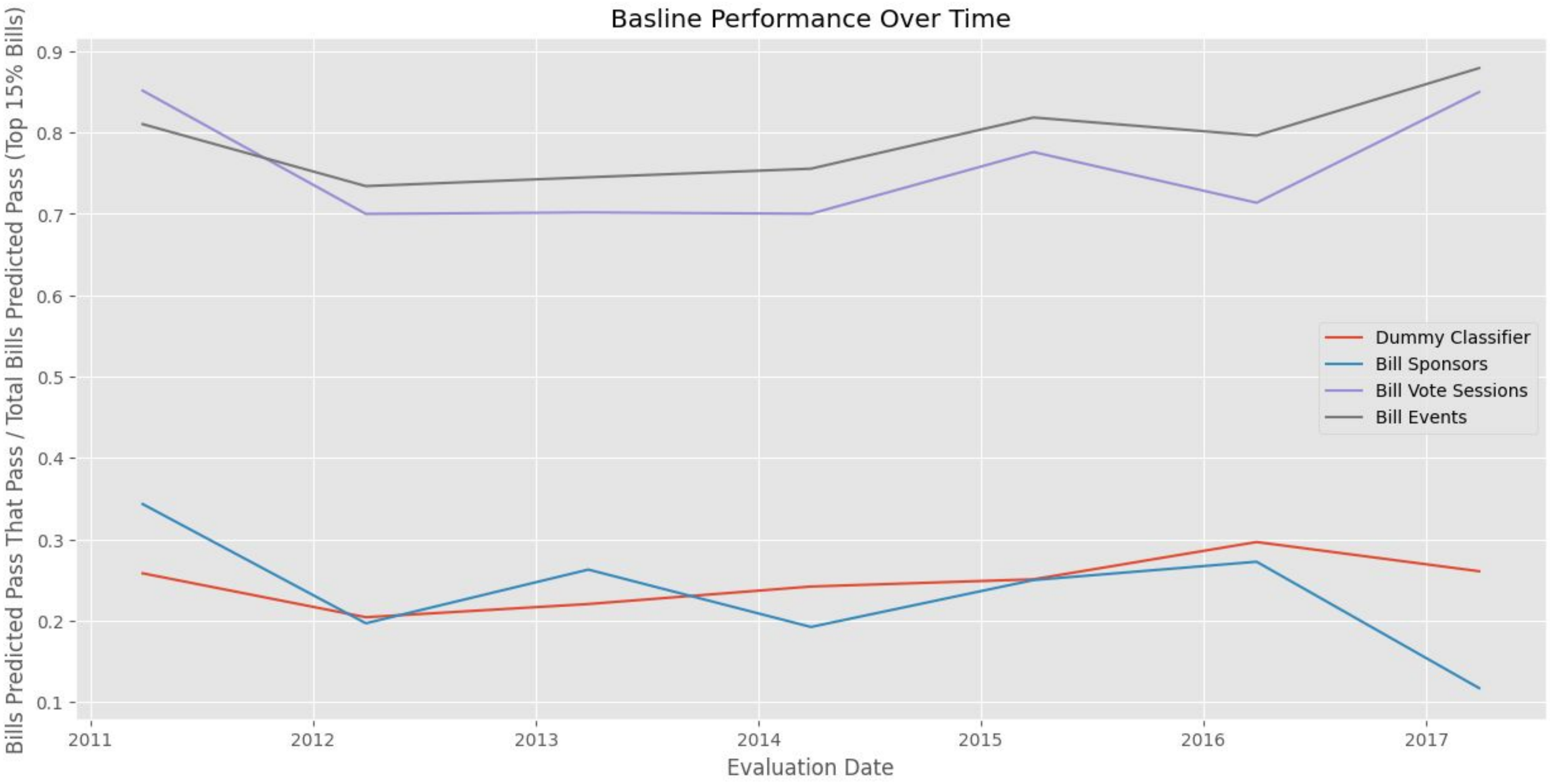


Prioritizing Performance For The Top 15% of Bills Predicted To Pass

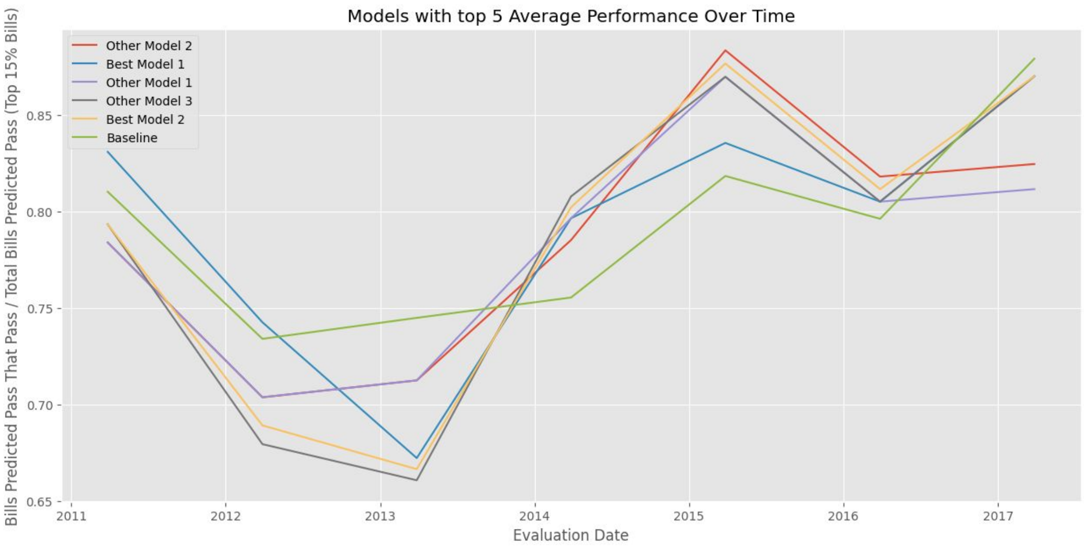
For the top 15%:

$$\text{Precision} = \frac{\text{\# Bills we predicted to pass and do pass}}{\text{Total \# of bills we predicted will pass}}$$

Best Baseline: Number of Bill Events



Model Performance



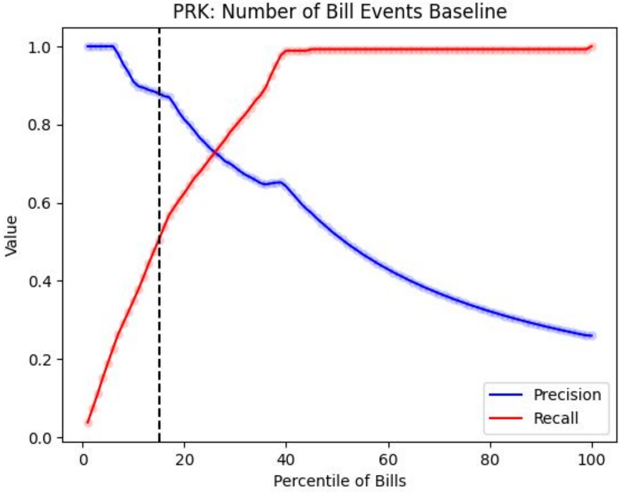
Models With Best Average Performance

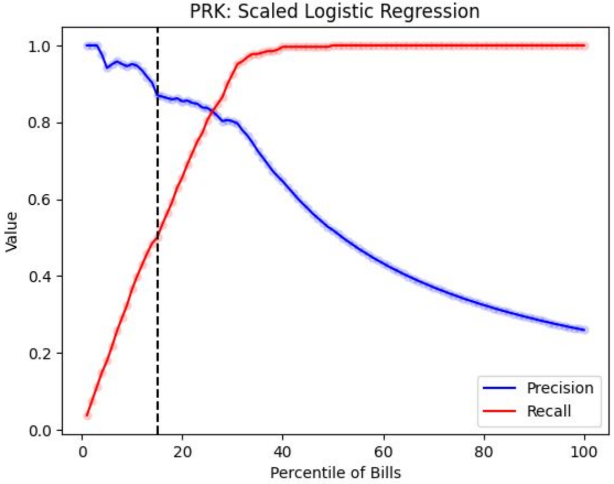
1. **Scaled Logistic Regression**
2. **Baseline**
3. **Random Forest Classifier**

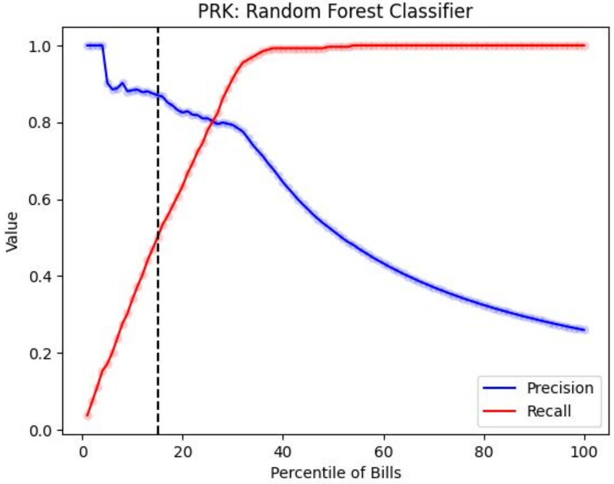
Precision:

$$\frac{\text{\# Bills we predicted to pass and do pass}}{\text{Total \# of bills we predicted will pass}}$$

Performance Frontier Across Percentile of Top Ranked Bills







Precision:

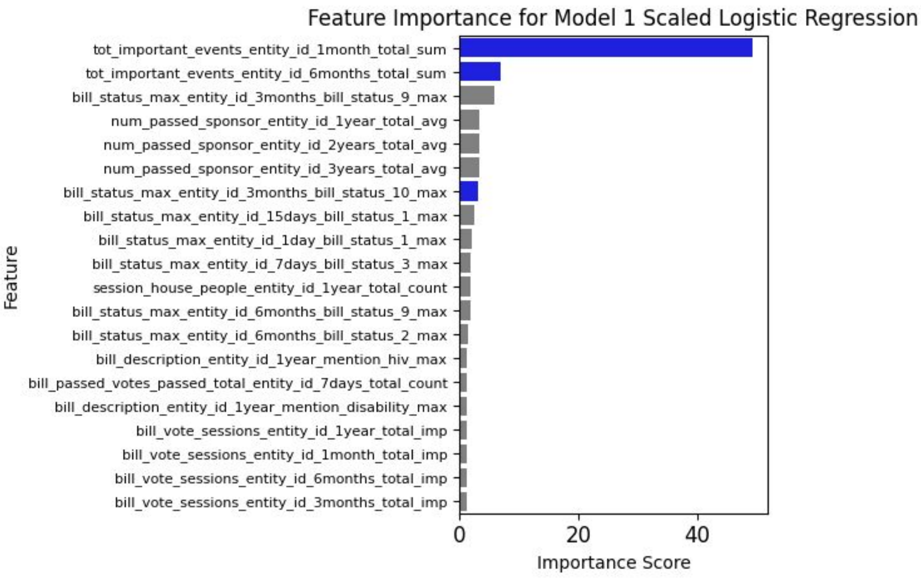
$$\frac{\text{\# Bills we predicted to pass and do pass}}{\text{Total \# of bills we predicted will pass}}$$

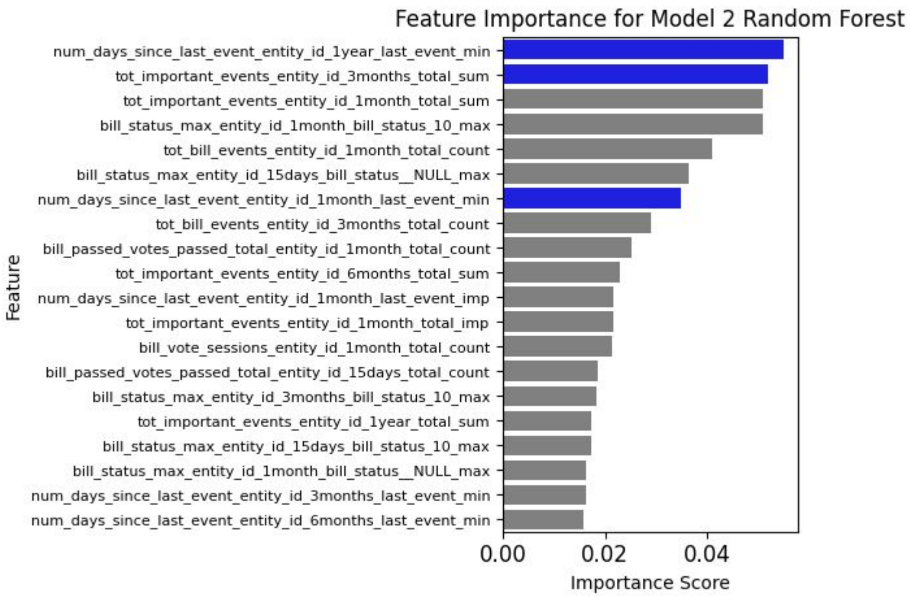
Recall:

$$\frac{\text{\# Bills we predicted to pass and do pass}}{\text{Total \# of bills that pass in session}}$$

Most Important Features Capture Movement of Bills

Common important features: Total bill events, total important events, and days since last event





How Do Features Differ Between Bills Within Top 15% And Those Outside?

	top_values	bottom_values	difference
num_days_since_last_event_entity_id_1year_last_event_min	3.0519	43.5017	40.4498
num_days_since_last_event_entity_id_6months_last_event_min	3.0519	43.5017	40.4498
num_days_since_last_event_entity_id_3months_last_event_min	3.0519	43.5017	40.4498
tot_bill_events_entity_id_3months_total_count	20.4026	10.0296	10.3730
tot_bill_events_entity_id_1year_total_count	20.4156	10.0512	10.3644
tot_bill_events_entity_id_6months_total_count	20.4156	10.0512	10.3644
tot_bill_events_house_entity_id_3months_total_count	10.3377	5.0125	5.3251
tot_bill_events_house_entity_id_1year_total_count	10.3377	5.0171	5.3206
tot_bill_events_house_entity_id_6months_total_count	10.3377	5.0171	5.3206
tot_bill_events_entity_id_1month_total_count	6.7143	1.5597	5.1546
tot_bill_events_senate_entity_id_3months_total_count	10.0649	5.0171	5.0479
tot_bill_events_senate_entity_id_1year_total_count	10.0779	5.0341	5.0438
tot_bill_events_senate_entity_id_6months_total_count	10.0779	5.0341	5.0438
bill_passed_votes_passed_total_entity_id_1year_total_count	6.3377	1.7759	4.5618
bill_passed_votes_passed_total_entity_id_3months_total_count	6.3377	1.7759	4.5618
bill_passed_votes_passed_total_entity_id_6months_total_count	6.3377	1.7759	4.5618
bill_vote_sessions_entity_id_1year_total_count	6.3377	1.8441	4.4935
bill_vote_sessions_entity_id_3months_total_count	6.3377	1.8441	4.4935
bill_vote_sessions_entity_id_6months_total_count	6.3377	1.8441	4.4935
tot_important_events_entity_id_6months_total_sum	6.2273	2.4835	3.7438
tot_important_events_entity_id_3months_total_sum	6.2273	2.4835	3.7438
tot_important_events_entity_id_1year_total_sum	6.2273	2.4835	3.7438
bill_sponsors_dem_entity_id_1year_total_count	0.0714	2.9977	2.9263
tot_bill_events_senate_entity_id_1month_total_count	3.5779	0.7315	2.8464

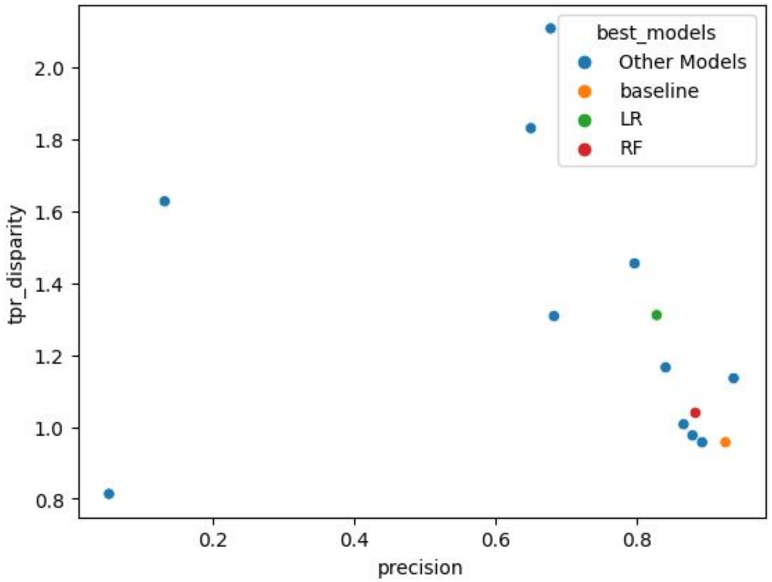
The number of days between the last event day and the current date

The number of total events the bill has been through(also a baseline feature)

The number of vote sessions the bill has been through(also a baseline feature)

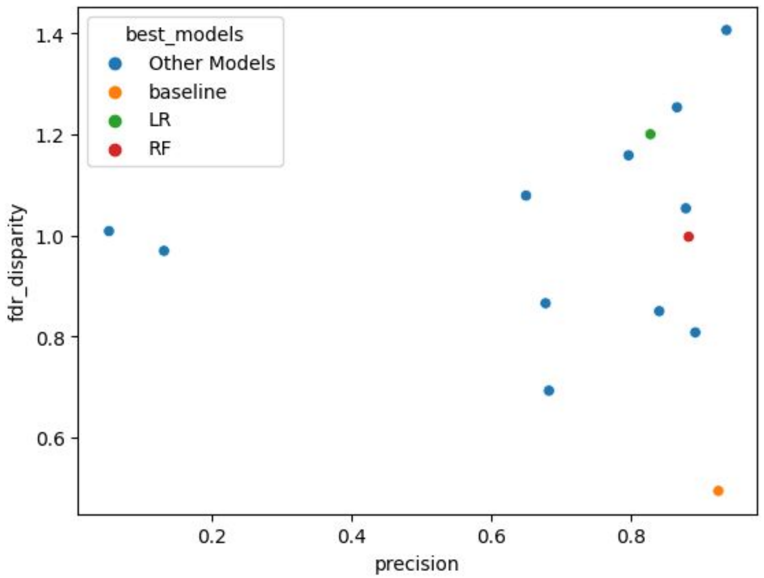
Bias Audit: Civil Rights Bill vs. Non-Civil Rights Bill

Within the acceptable range of [0.8, 1.25]



True Positive Rate (tpr):

$$\frac{\text{\# Bills we predicted to pass and do pass}}{\text{Total \# of bills that pass in session}}$$



False Discovery Rate (fdr):

$$\frac{\text{\# Bills we predicted to pass but do not pass}}{\text{Total \# of bills we predicted will pass}}$$

Limitations, Recommendations, and Future Work

Caveats

- Used naive approach to assess bill relevance
- Bills Included in top 15% may not be most relevant to ACPA
- Could have captured more granularity with features
- Included bills “reported pass”
- Updated model every 3 months
- Only one month in test sets with any passed bills (March)

Future Work

Increase Feature Complexity	Test More Complex Models	Prioritize Identification of Civil Rights Bills	Continue Bias Audit	Field Trials
• Increase time granularity • Leverage textual data • Include advocacy related data from ACPA	• Try models that are better fit for handling text data • Early identification of bills that could pass	• Of all the bills that pass, which of them infringe on civil liberties?	• How does performance differ across bills with higher number of keywords? • How does performance differ across sponsorship and polarization?	• Strategy for identification of bills based on results from field trials • Accounting for intensity of advocacy efforts

Policy Recommendations

1. **Baselines perform well**
   - **Rank Multiline on Total Bill Events** is 87% precise at 15th percentile

2. **Model Recommendation**
   - For the most recent test period, **Random Forest** performs best at 88% precise at 15th percentile
   - **Scaled Logistic Regression** performs best when averaged across all time periods
   - For the most recent test period **Random Forest** has the least disparity towards civil rights bill at 15th percentile
   - Model best suited for ACPA should be following **Field Trials**

3. **Re-strategizing advocacy efforts through the year**
   - In 2018, **90%** of bills were introduced by March, **100%** bills were passed by June
   - Strategize re-allocation of advocacy efforts in the crucial period before June

Thank You!

Questions?

Appendix

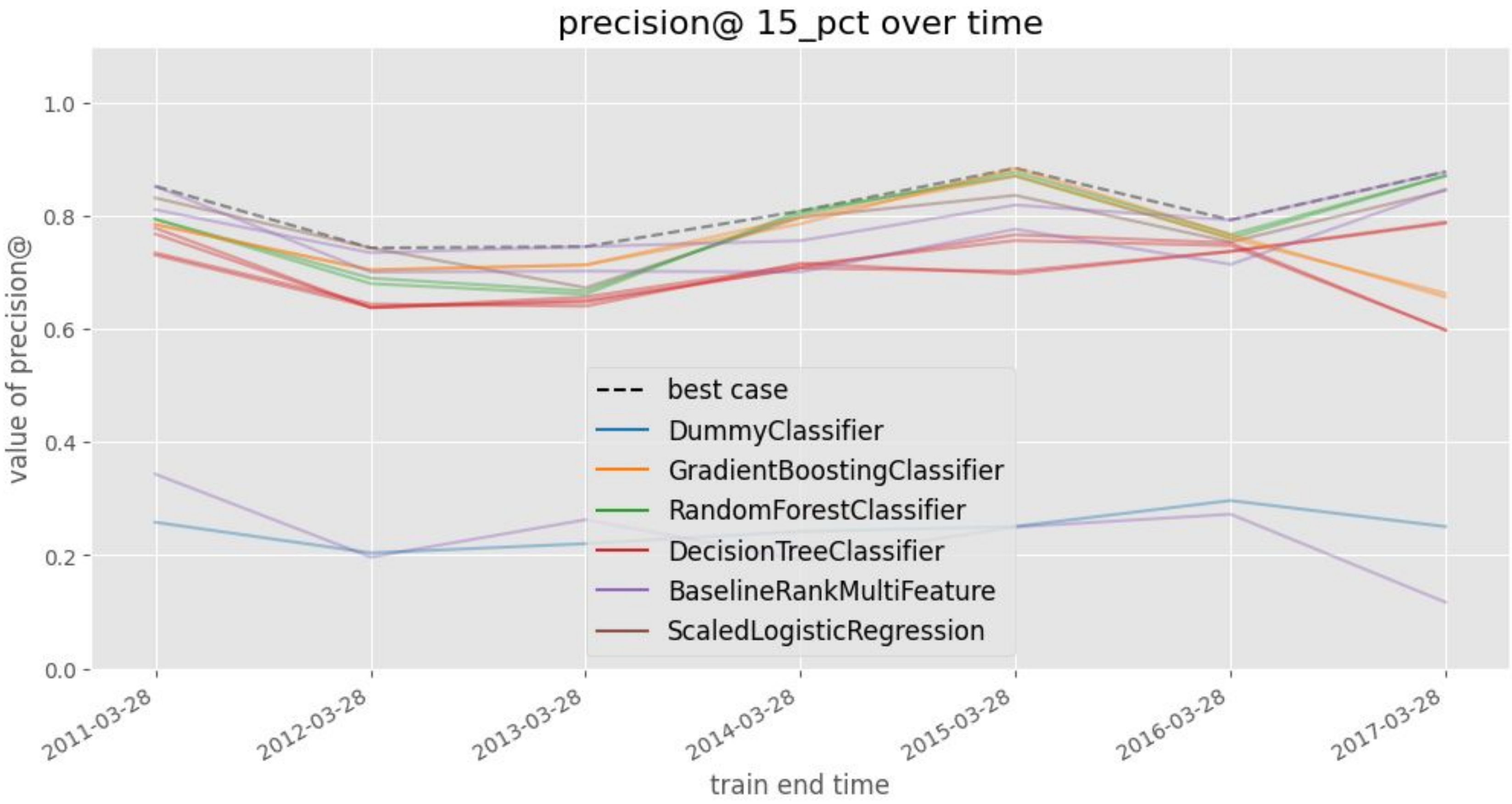
Best Performing Models

Model type: **Scaled Logistic Regression**
Hyperparameters: {"C": 0.1, "solver": "saga", "penalty": "l1", "random_state": 4}
Precision @ 15 percentile: 84.41%

Model_type: **Baseline**
Hyperparameters: {"rules": [{"feature": "tot_bill_events_entity_id_1year_total_count", "low_value_high_score": false}]}
Precision @ 15 percentile: 87.71%

Model type: **Random Forest**
Hyperparameters: {"criterion": "gini", "max_depth": 25, "n_estimators": 1000, "random_state": 4, "min_samples_split": 25}
Precision @ 15 percentile: 88.31%

Performance of All Models Evaluated



Percent of Bills Introduced That Pass by End Of Session

2018

	passed_pct
Jan	27.915194
Feb	27.993394
Mar	22.123894
Apr	6.638116
May	0.000000
Jun	0.000000
Jul	0.000000
Aug	0.000000
Sept	0.000000
Oct	0.000000
Nov	0.000000
Dec	0.000000

2017

	passed_pct
Jan	35.465925
Feb	31.720430
Mar	25.943853
Apr	13.816535
May	0.000000
Jun	0.000000
Jul	0.000000
Aug	0.000000
Sept	0.000000
Oct	0.000000
Nov	0.000000
Dec	0.000000

2016

	passed_pct
Jan	35.207497
Feb	34.548769
Mar	30.194175
Apr	27.299703
May	0.000000
Jun	0.000000
Jul	0.000000
Aug	0.000000
Sept	0.000000
Oct	0.000000
Nov	0.000000
Dec	0.000000