PREDICTING THE VOTE OF A SUPREME COURT JUSTICE

Natnaell Mammo & Santiago Larrain

Recap

- Build a Model to predict the “direction” of a justice vote by using the “Facts of the Case” (text)
- Use two data sources
  - *General case facts*
  - *More detailed text data*
- Exploratory Analysis and Non-Text Models (Nat)
- Model Used for text
- Models Results comparison
- Results details
  - By Judge
  - By Issue

Hypothesis?

- Are we trying to find out if a Justice is conservative or liberal?
- Are we trying to predict pending cases?
- Are we trying to find out if “the facts of the case” are useful or enough to predict a Justice vote direction?

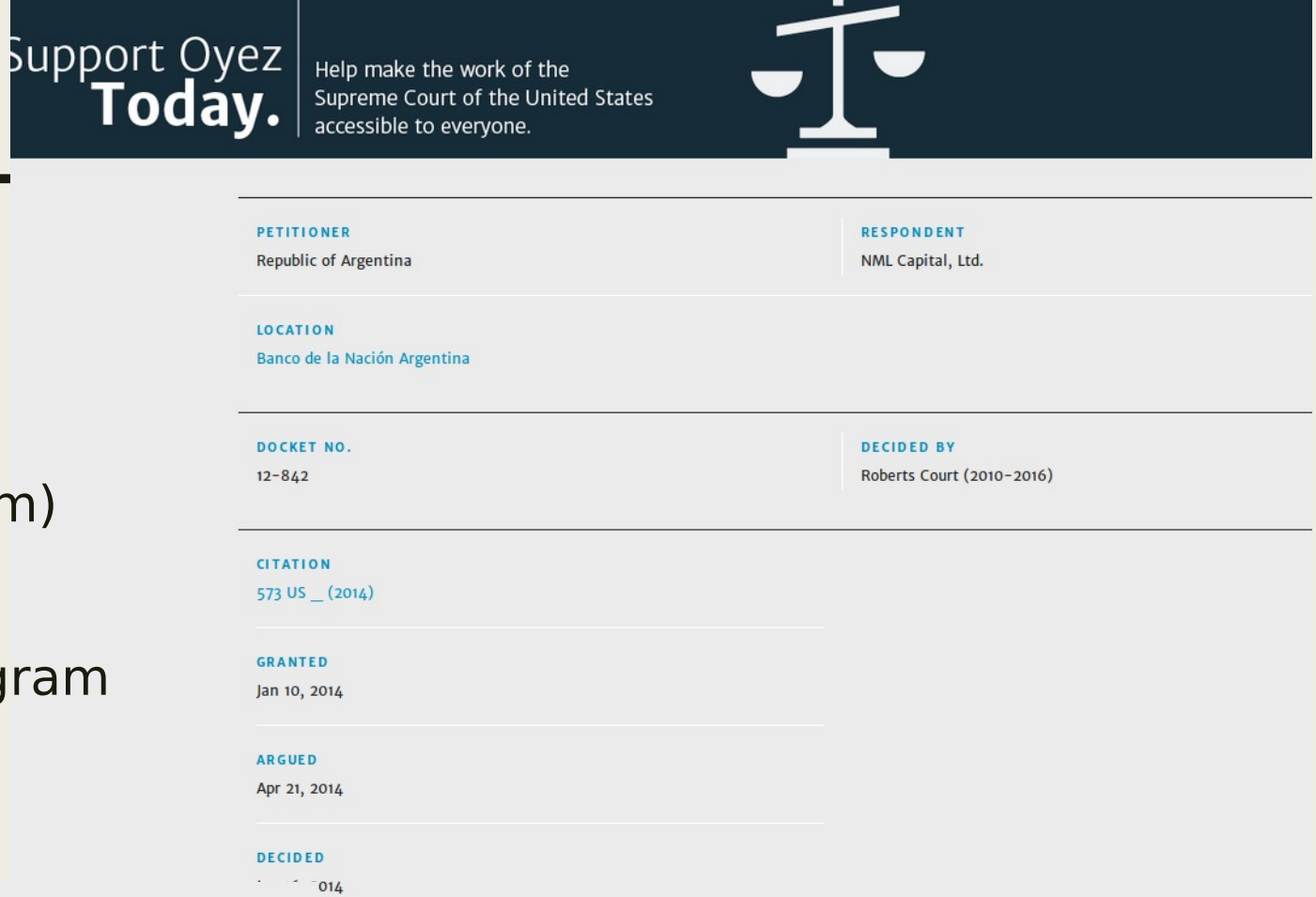
The Data: General Case Facts

- 3147 cases for the last 9 justices
- 3082 of which had a normal vote
- 32 possible features
- Many had missing values
  - *1230 cases had the majority of features filled*

caseId	docketId	caseIssues	voteId	dateDecision	decisionT	usCite	sctCite	ledCite	lexisCite	term	naturalCo	chief	docket	caseName	dateArgu	dateRearg	petitioner	petitioner	responde	responde	jurisdictio	adminActi	adminActi	threeJudg	caseOrigi	caseOrigi
1946-001	1946-001-	1946-001-	1946-001-	11/18/1946	1	329 U.S. 1	67 S. Ct. 6	91 L. Ed. 3	1946 U.S. L	1946	1301	Vinson	24	HALLIBURT	1/9/1946	10/23/1946	198		172		6			0	51	6
1946-001	1946-001-	1946-001-	1946-001-	11/18/1946	1	329 U.S. 1	67 S. Ct. 6	91 L. Ed. 3	1946 U.S. L	1946	1301	Vinson	24	HALLIBURT	1/9/1946	10/23/1946	198		172		6			0	51	6
1946-001	1946-001-	1946-001-	1946-001-	11/18/1946	1	329 U.S. 1	67 S. Ct. 6	91 L. Ed. 3	1946 U.S. L	1946	1301	Vinson	24	HALLIBURT	1/9/1946	10/23/1946	198		172		6			0	51	6
1946-001	1946-001-	1946-001-	1946-001-	11/18/1946	1	329 U.S. 1	67 S. Ct. 6	91 L. Ed. 3	1946 U.S. L	1946	1301	Vinson	24	HALLIBURT	1/9/1946	10/23/1946	198		172		6			0	51	6
1946-001	1946-001-	1946-001-	1946-001-	11/18/1946	1	329 U.S. 1	67 S. Ct. 6	91 L. Ed. 3	1946 U.S. L	1946	1301	Vinson	24	HALLIBURT	1/9/1946	10/23/1946	198		172		6			0	51	6
1946-001	1946-001-	1946-001-	1946-001-	11/18/1946	1	329 U.S. 1	67 S. Ct. 6	91 L. Ed. 3	1946 U.S. L	1946	1301	Vinson	24	HALLIBURT	1/9/1946	10/23/1946	198		172		6			0	51	6
1946-001	1946-001-	1946-001-	1946-001-	11/18/1946	1	329 U.S. 1	67 S. Ct. 6	91 L. Ed. 3	1946 U.S. L	1946	1301	Vinson	24	HALLIBURT	1/9/1946	10/23/1946	198		172		6			0	51	6
1946-001	1946-001-	1946-001-	1946-001-	11/18/1946	1	329 U.S. 1	67 S. Ct. 6	91 L. Ed. 3	1946 U.S. L	1946	1301	Vinson	24	HALLIBURT	1/9/1946	10/23/1946	198		172		6			0	51	6
1946-001	1946-001-	1946-001-	1946-001-	11/18/1946	1	329 U.S. 1	67 S. Ct. 6	91 L. Ed. 3	1946 U.S. L	1946	1301	Vinson	24	HALLIBURT	1/9/1946	10/23/1946	198		172		6			0	51	6
1946-002	1946-002-	1946-002-	1946-002-	11/18/1946	1	329 U.S. 1	67 S. Ct. 1	91 L. Ed. 1	1946 U.S. L	1946	1301	Vinson	12	CLEVELAN	########	10/17/1946	100		27		1			0	123	52
1946-002	1946-002-	1946-002-	1946-002-	11/18/1946	1	329 U.S. 1	67 S. Ct. 1	91 L. Ed. 1	1946 U.S. L	1946	1301	Vinson	12	CLEVELAN	########	10/17/1946	100		27		1			0	123	52
1946-002	1946-002-	1946-002-	1946-002-	11/18/1946	1	329 U.S. 1	67 S. Ct. 1	91 L. Ed. 1	1946 U.S. L	1946	1301	Vinson	12	CLEVELAN	########	10/17/1946	100		27		1			0	123	52
1946-002	1946-002-	1946-002-	1946-002-	11/18/1946	1	329 U.S. 1	67 S. Ct. 1	91 L. Ed. 1	1946 U.S. L	1946	1301	Vinson	12	CLEVELAN	########	10/17/1946	100		27		1			0	123	52
1946-002	1946-002-	1946-002-	1946-002-	11/18/1946	1	329 U.S. 1	67 S. Ct. 1	91 L. Ed. 1	1946 U.S. L	1946	1301	Vinson	12	CLEVELAN	########	10/17/1946	100		27		1			0	123	52
1946-002	1946-002-	1946-002-	1946-002-	11/18/1946	1	329 U.S. 1	67 S. Ct. 1	91 L. Ed. 1	1946 U.S. L	1946	1301	Vinson	12	CLEVELAN	########	10/17/1946	100		27		1			0	123	52
1946-002	1946-002-	1946-002-	1946-002-	11/18/1946	1	329 U.S. 1	67 S. Ct. 1	91 L. Ed. 1	1946 U.S. L	1946	1301	Vinson	12	CLEVELAN	########	10/17/1946	100		27		1			0	123	52
1946-002	1946-002-	1946-002-	1946-002-	11/18/1946	1	329 U.S. 1	67 S. Ct. 1	91 L. Ed. 1	1946 U.S. L	1946	1301	Vinson	12	CLEVELAN	########	10/17/1946	100		27		1			0	123	52
1946-002	1946-002-	1946-002-	1946-002-	11/18/1946	1	329 U.S. 1	67 S. Ct. 1	91 L. Ed. 1	1946 U.S. L	1946	1301	Vinson	12	CLEVELAN	########	10/17/1946	100		27		1			0	123	52
1946-002	1946-002-	1946-002-	1946-002-	11/18/1946	1	329 U.S. 1	67 S. Ct. 1	91 L. Ed. 1	1946 U.S. L	1946	1301	Vinson	17	CLEVELAN	########	10/17/1946	100		27		1			0	123	52
1946-002	1946-002-	1946-002-	1946-002-	11/18/1946	1	329 U.S. 1	67 S. Ct. 1	91 L. Ed. 1	1946 U.S. L	1946	1301	Vinson	17	CLEVELAN	########	10/17/1946	100		27		1			0	123	52

The Data:TEXT

- “Facts of the Case”
- 14670 unique words (Unigram)
- 125896 Unique Bigrams
- 140566 unique grams for Bigram + Unigram model



Facts of the case

Roe, a Texas resident, sought to terminate her pregnancy by abortion. Texas law prohibited abortions except to save the pregnant woman's life. After granting certiorari, the Court heard arguments twice. The first time, Roe's attorney -- Sarah Weddington -- could not locate the constitutional hook of her argument for Justice Potter Stewart. Her opponent -- Jay Floyd -- misfired from the start. Weddington sharpened her constitutional argument in the second round. Her new opponent -- Robert Flowers -- came under strong questioning from Justices Potter Stewart and Thurgood Marshall.

Facts of the case

During an economic crisis in 2001, the Republic of Argentina (Argentina) failed to make payments on bonds owned by foreign investors. One such bondholder, NML Capital, Ltd. (NML), later prevailed in several actions it filed against Argentina in federal district court, which entered judgments totaling more than US$2 billion in NML's favor. In order to execute the judgments against Argentina, NML served subpoenas on two banks requesting information about Argentina's assets held worldwide. Argentina moved to quash the subpoenas and argued that they violate the Foreign Sovereign Immunities Act (FSIA) by requiring the disclosure of assets that are immune from collection by NML. The district court ordered the banks to comply with the subpoena requests. The U.S. Court of Appeals for the Second Circuit affirmed, reasoning that the FSIA did not apply to the subpoena because it was a discovery order directed at commercial entities that did not have a claim to sovereign immunity.

Model Building

- Features
  - *Issue*
  - *Issue Area*
  - *Lower Court Disposition Direction*
  - *Petitioner*
  - *Respondent*
  - *Length of Justice's Service*
  - *Case origin*
- Models
  - *Random Forest*
  - *Logistic Regression*
  - *Nearest Neighbor*
  - *Linear SVC*

First Ammendment
attorney or gov officials compensation
civil rights
criminal procedure
due process
economic activity
federal taxation
federalism
judicial power
misc
privacy
unions

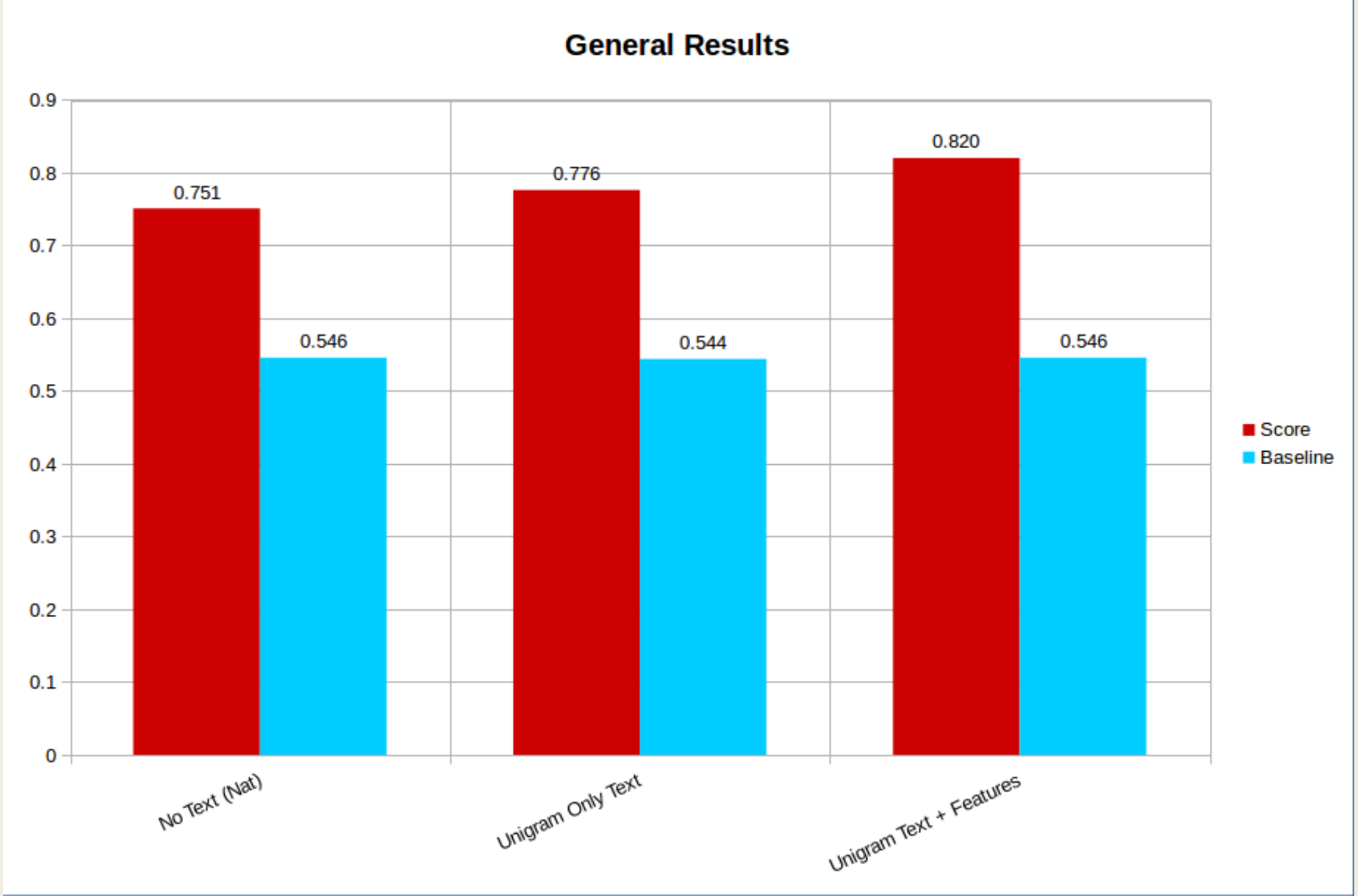
Values:

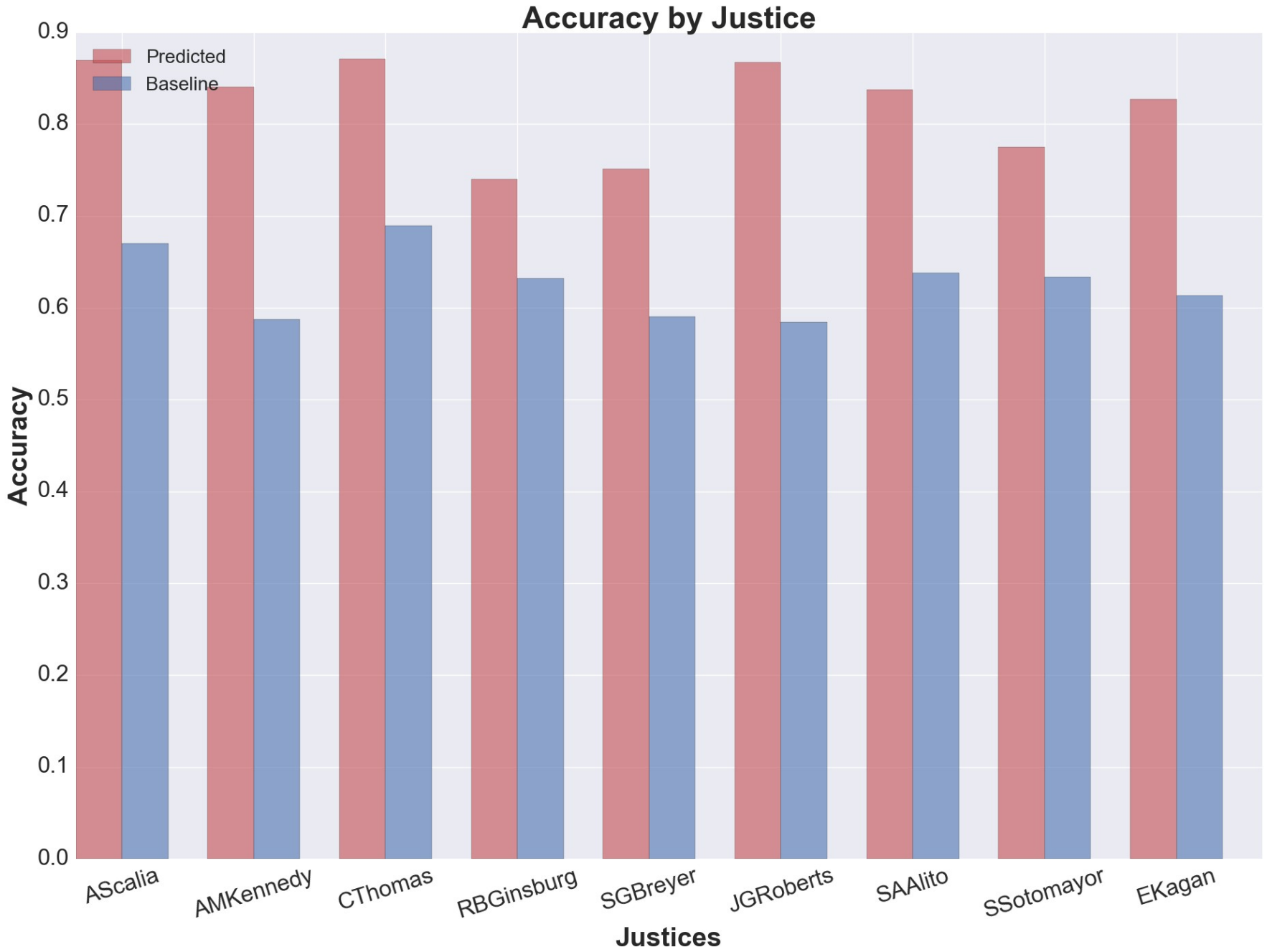
1	attorney general of the United States, or his office
2	specified state board or department of education
3	city, town, township, village, or borough government or governmental unit
4	state commission, board, committee, or authority
5	county government or county governmental unit, except school district
6	court or judicial district
7	state department or agency
8	governmental employee or job applicant
9	female governmental employee or job applicant
10	minority governmental employee or job applicant
11	minority female governmental employee or job applicant

Model Building - Text

- Count Vectorizer
- Tf-idf term weighting: term-frequency inverse document-frequency:

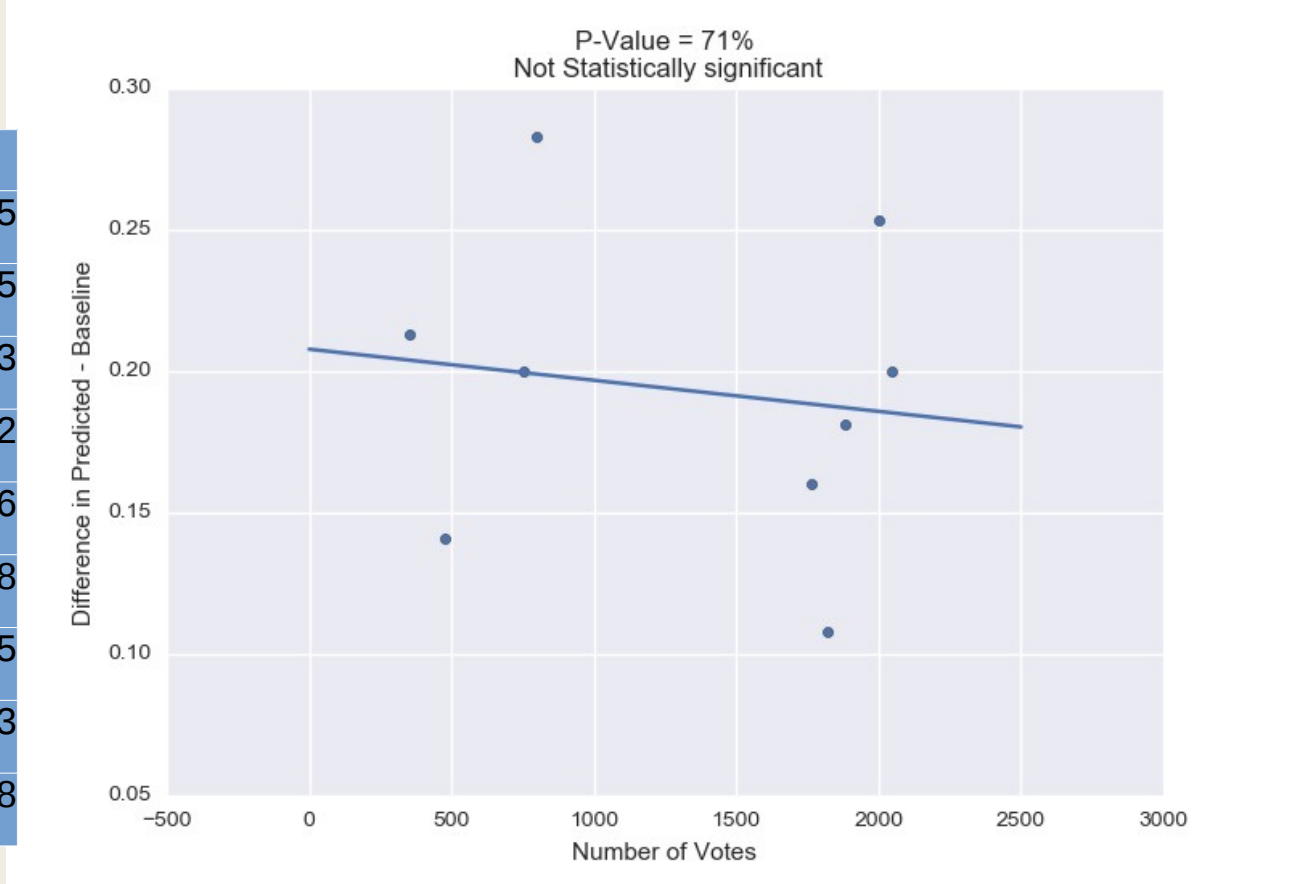
  In a large text corpus, some words will be very present hence carrying very little meaningful information about the actual contents of the document. If we were to feed the direct count data directly to a classifier those very frequent terms would shadow the frequencies of rarer yet more interesting terms
- Predict Justice vote direction from Sparse Matrix
- Cross Validation 5 folds → 11951 cases
- Models
  - *Random Forest*
  - *Logistic Regression*
  - *Nearest Neighbor*
  - *Linear SVC*


General Results
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.751
0.546
0.776
0.544
0.820
0.546
No Text (Nat)
Unigram Only Text
Unigram Text + Features
Score
Baseline

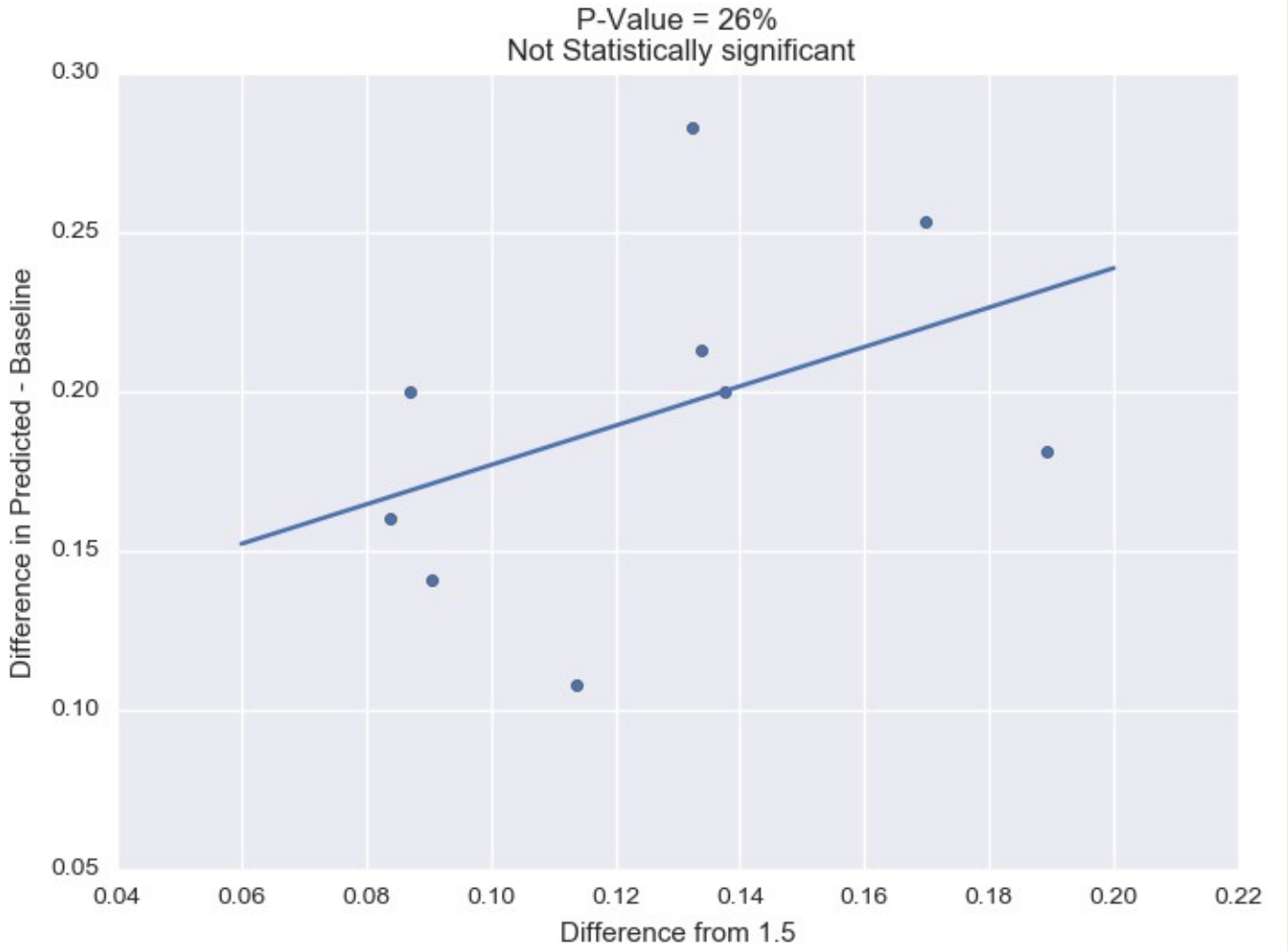

Accuracy by Justice
Predicted
Baseline
Accuracy
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
AScalia
AMKennedy
CThomas
RBGinsburg
SGBreyer
JGRoberts
SAAlito
SSotomayor
EKagan
Justices

Why the difference? Longevity?

	justiceName	Predicted	Baseline	n_votes	diff
0	AScalia	0.869565	0.669761	2047	0.199805
1	AMKennedy	0.8405	0.587	2000	0.2535
2	CThomas	0.870745	0.689362	1880	0.181383
3	RBGinsburg	0.739967	0.632216	1819	0.107752
4	SGBreyer	0.75071	0.590574	1761	0.160136
5	JGRoberts	0.867168	0.58396	798	0.283208
6	SAAlito	0.837302	0.637566	756	0.199735
7	SSotomayor	0.774737	0.633684	475	0.141053
8	EKagan	0.826705	0.613636	352	0.213068

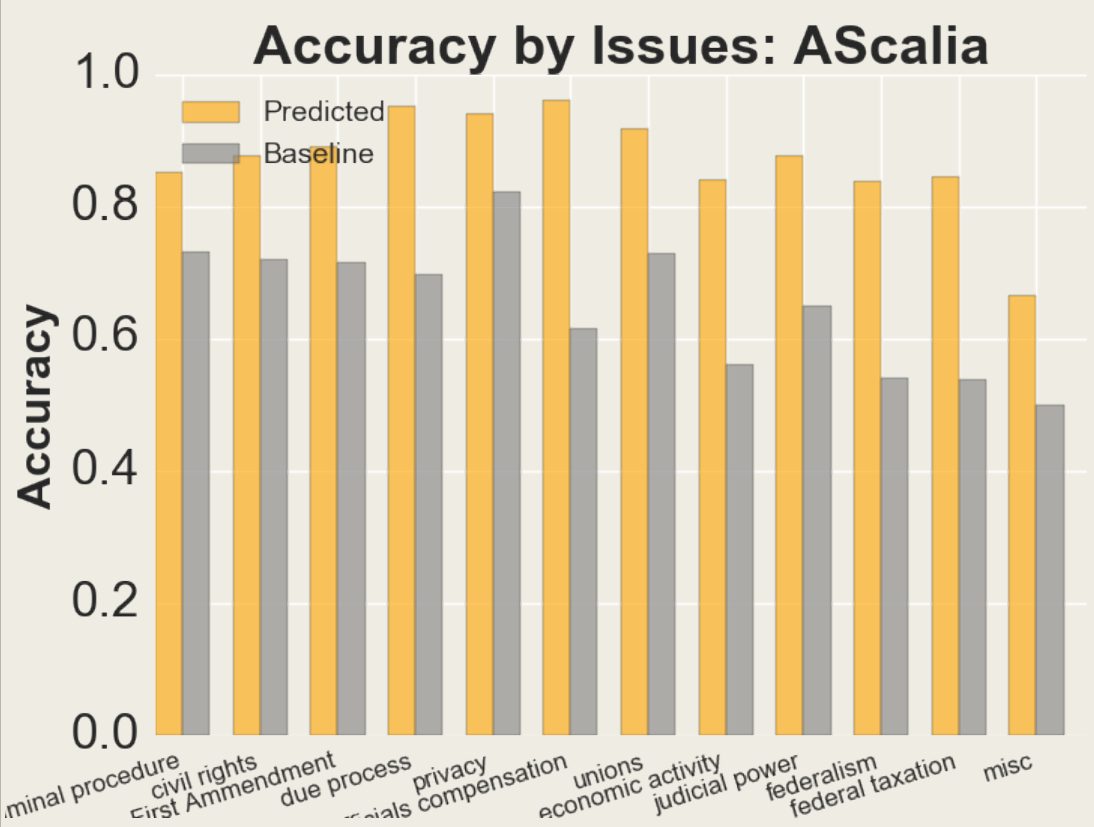


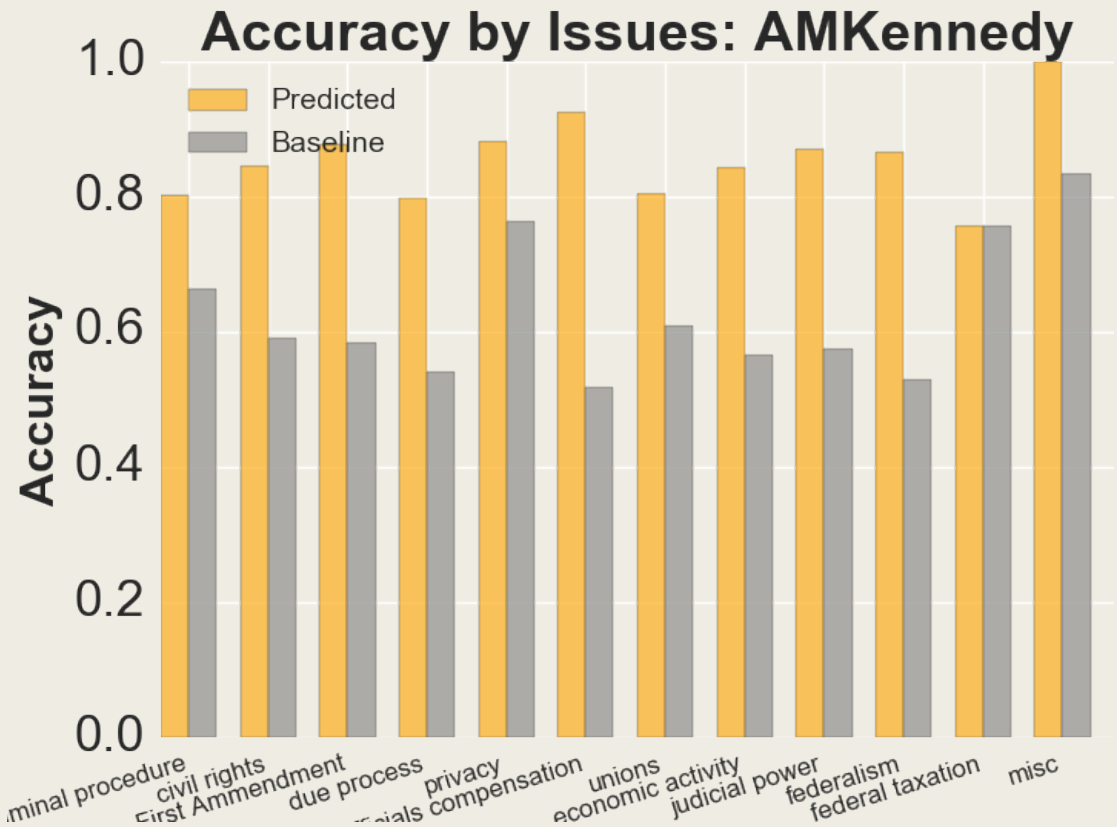
Why the difference? Extremist?



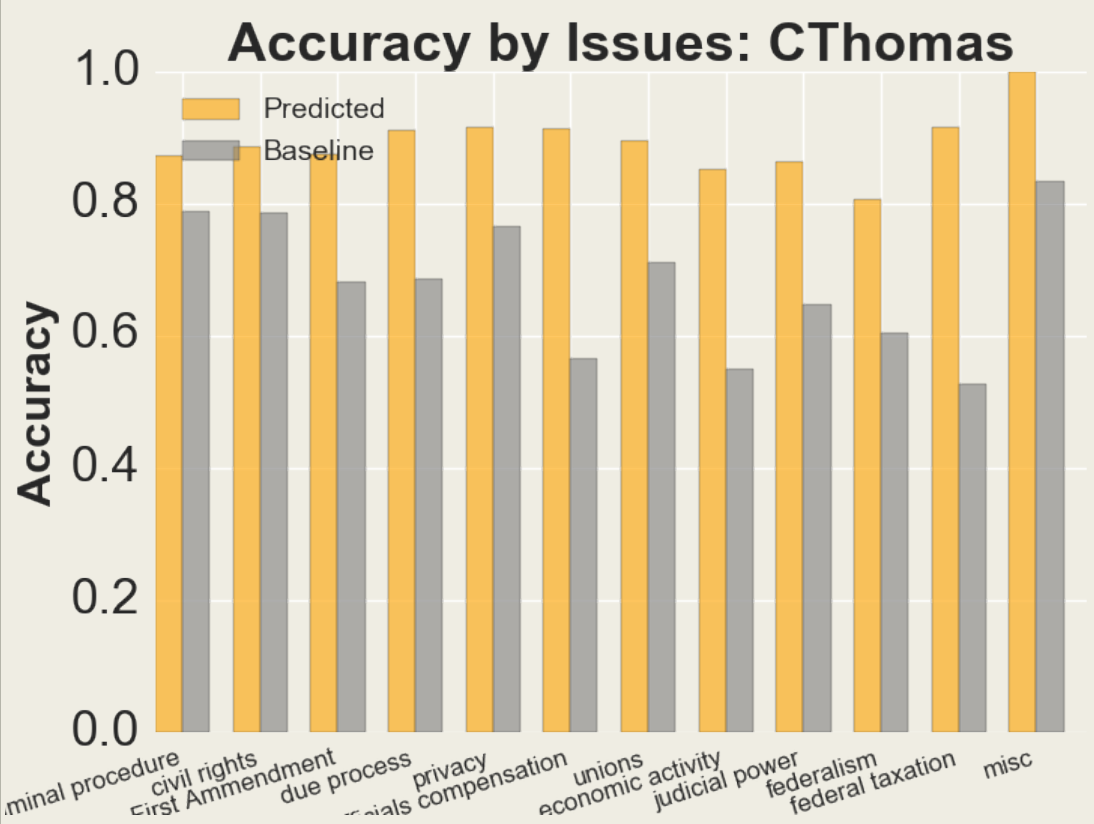

Accuracy by Issues
Predicted
Baseline
Accuracy
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
criminal procedure
civil rights
First Ammendment
due process
privacy
unions
economic activity
judicial power
federalism
federal taxation
misc

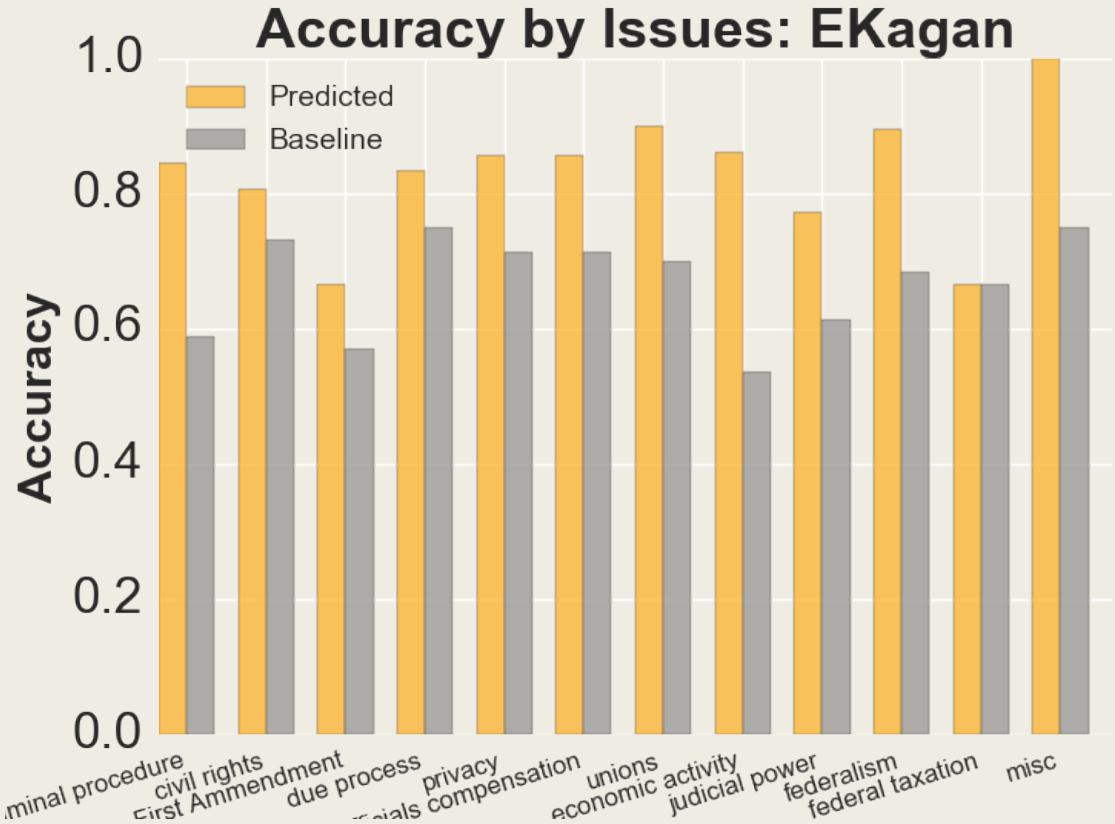
Accuracy by Issue and Justice



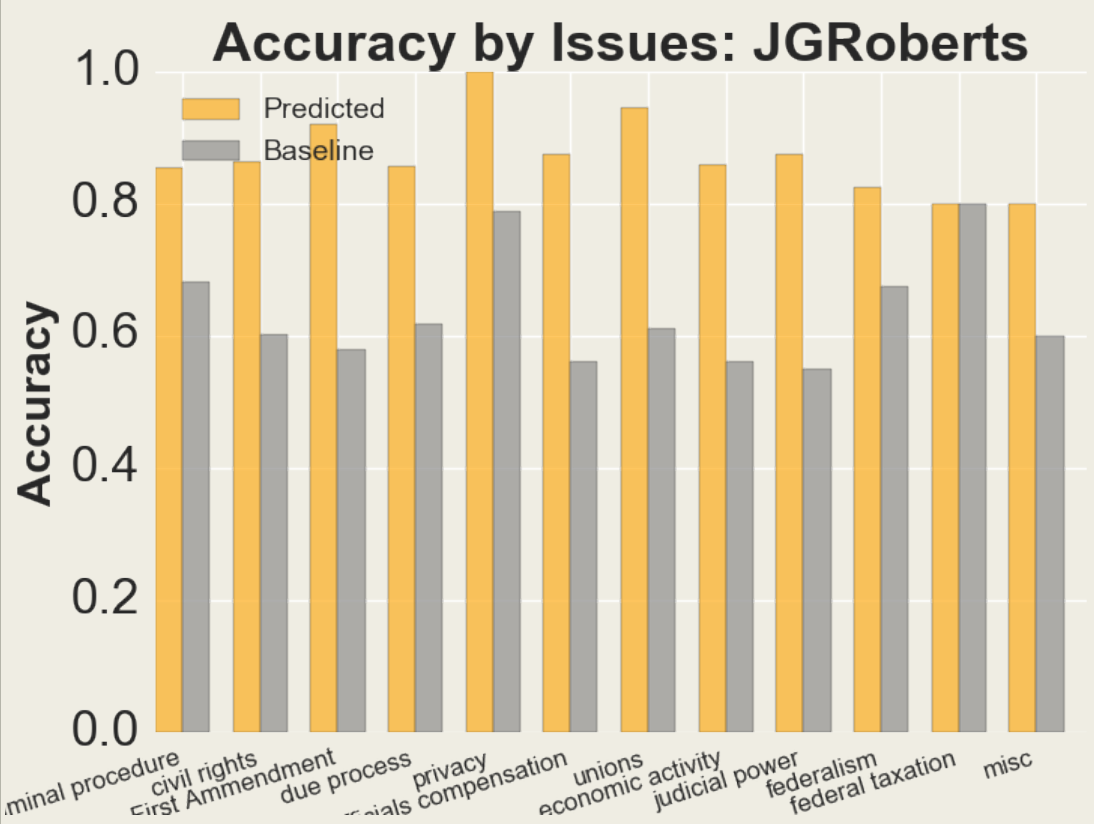


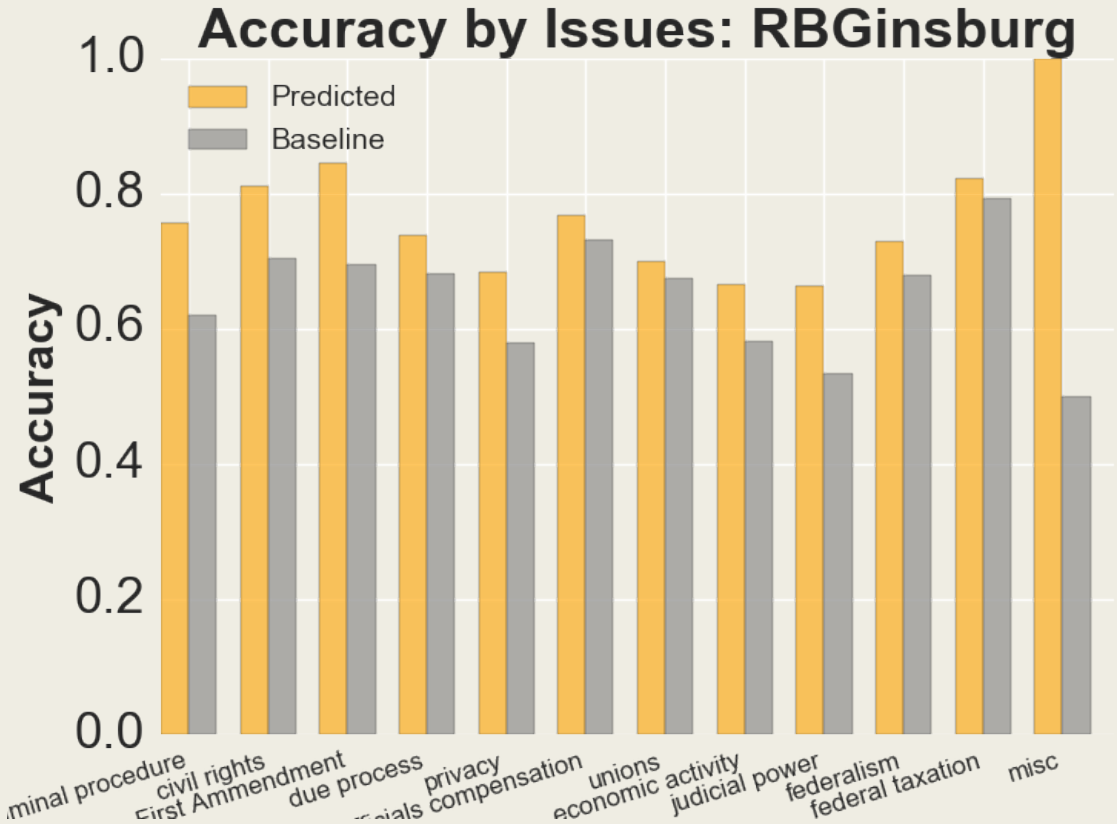
Accuracy by Issue and Justice



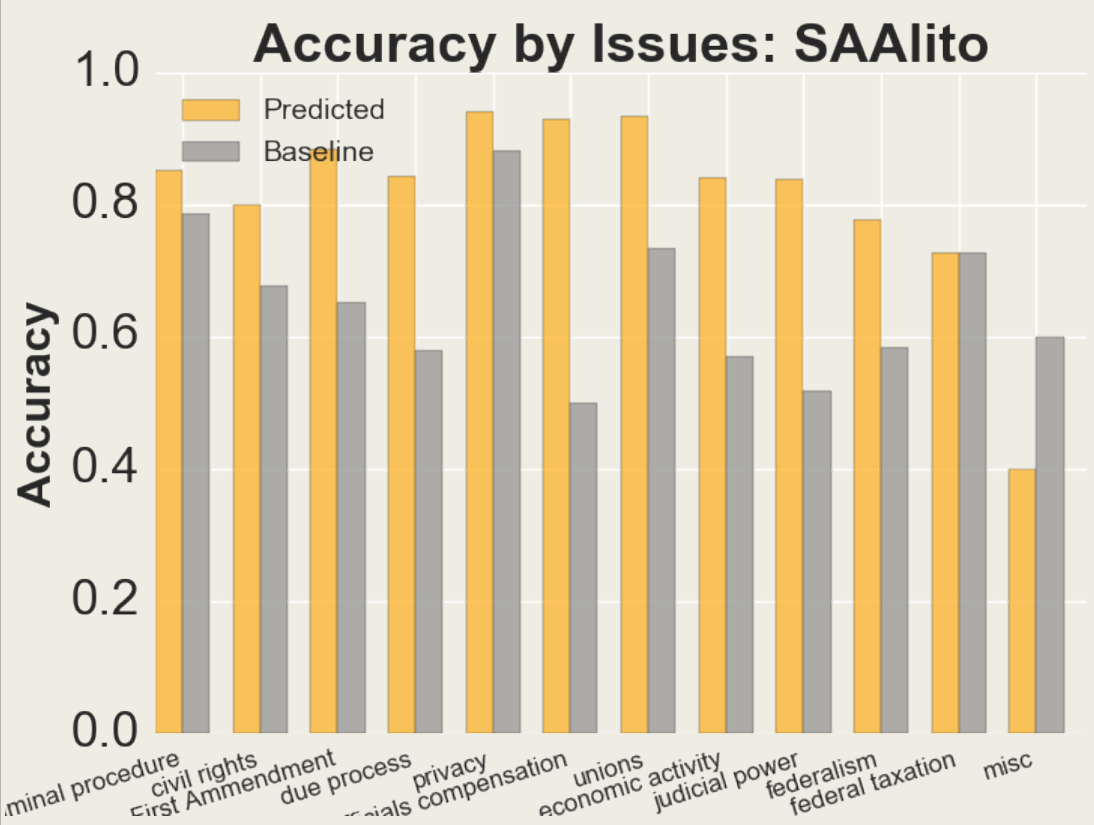


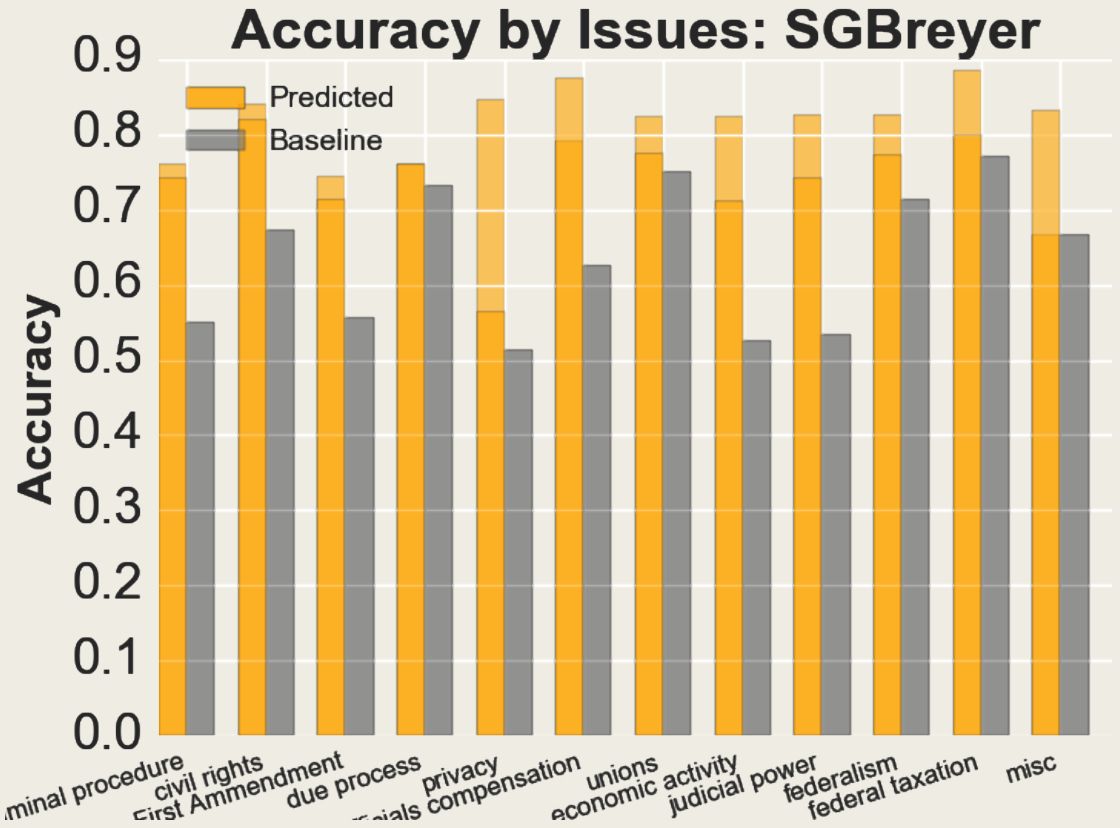
Accuracy by Issue and Justice



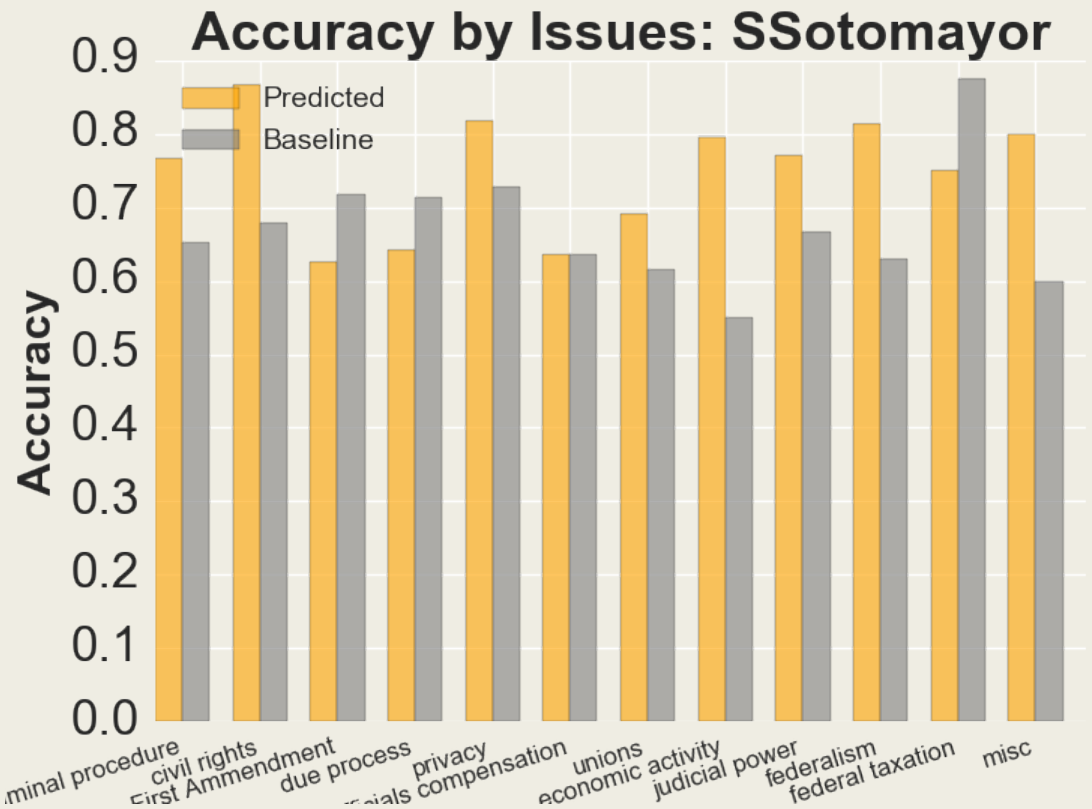


Accuracy by Issue and Justice

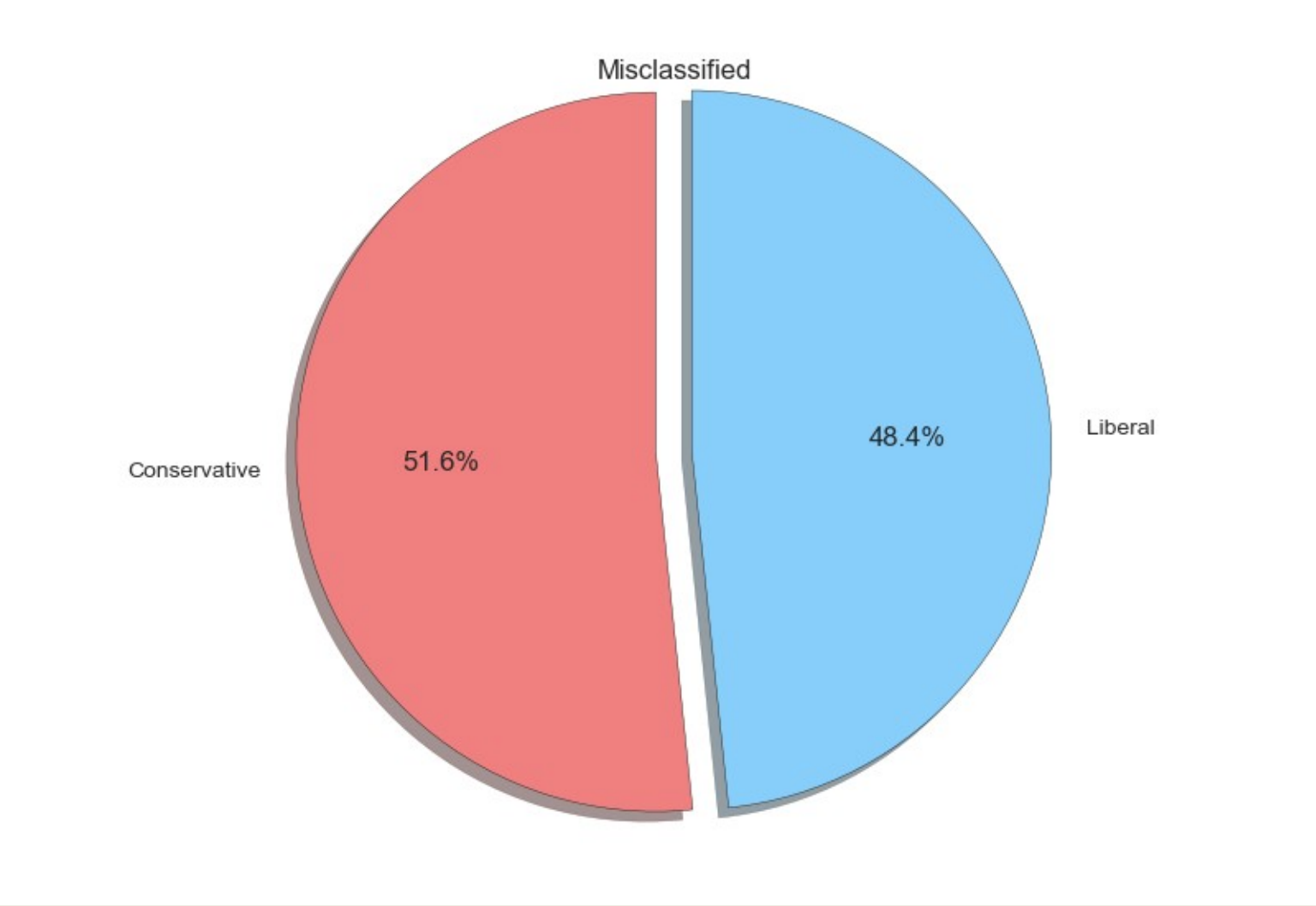




Accuracy by Issue and Justice



Misclassified



Very close.

The models seems to be a little biased towards a liberal vote. It misclassifies a Conservative vote as a Liberal more than a Liberal Vote as a Conservative. This might be explained by a higher number of liberal votes overall: 54.6% of the votes are Liberal vs. 45.4% Conservative.

There were 2674 misclassified votes.

Conclusions

- Text Data (“facts of the case”) IT IS very useful in predicting the direction of an individual vote
- Best Model is Random Forest with text + features combined.
- Only text perform better than only features
- Although there is very little improvement (roof at 82%).
- unigram/bigram perform similar
- Sparse Matrices rock
- Longevity as a Justice is useless to get better accuracy
- “Extremism” could actually explain something

The End