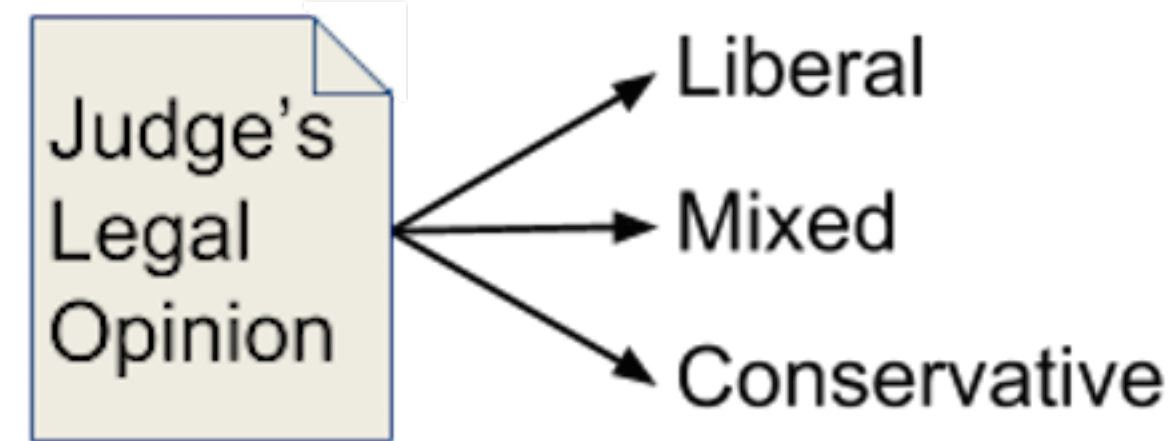


Inferring Political Valence from Written Judicial Decisions

Charlie Guthrie and Alex Pine

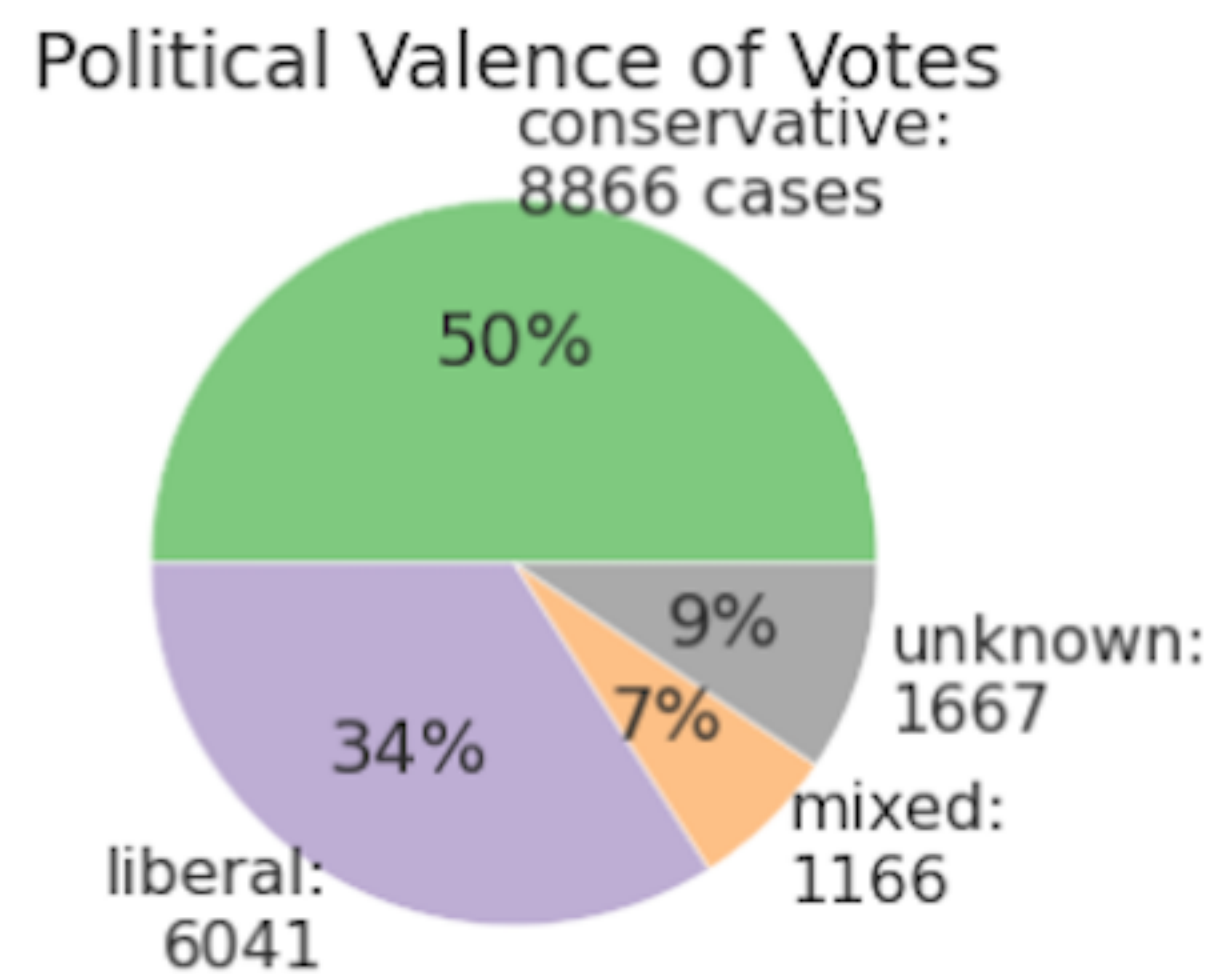
Goal

Analyze a judge's legal opinion to determine its political valence.



Data

- 300,000 U.S. Appeals Court cases
 - Represented as 1-grams through 8-grams.
 - 92 million n-grams total.
- 17,000 hand-coded cases
 - 412 variables per case
 - Includes political valence



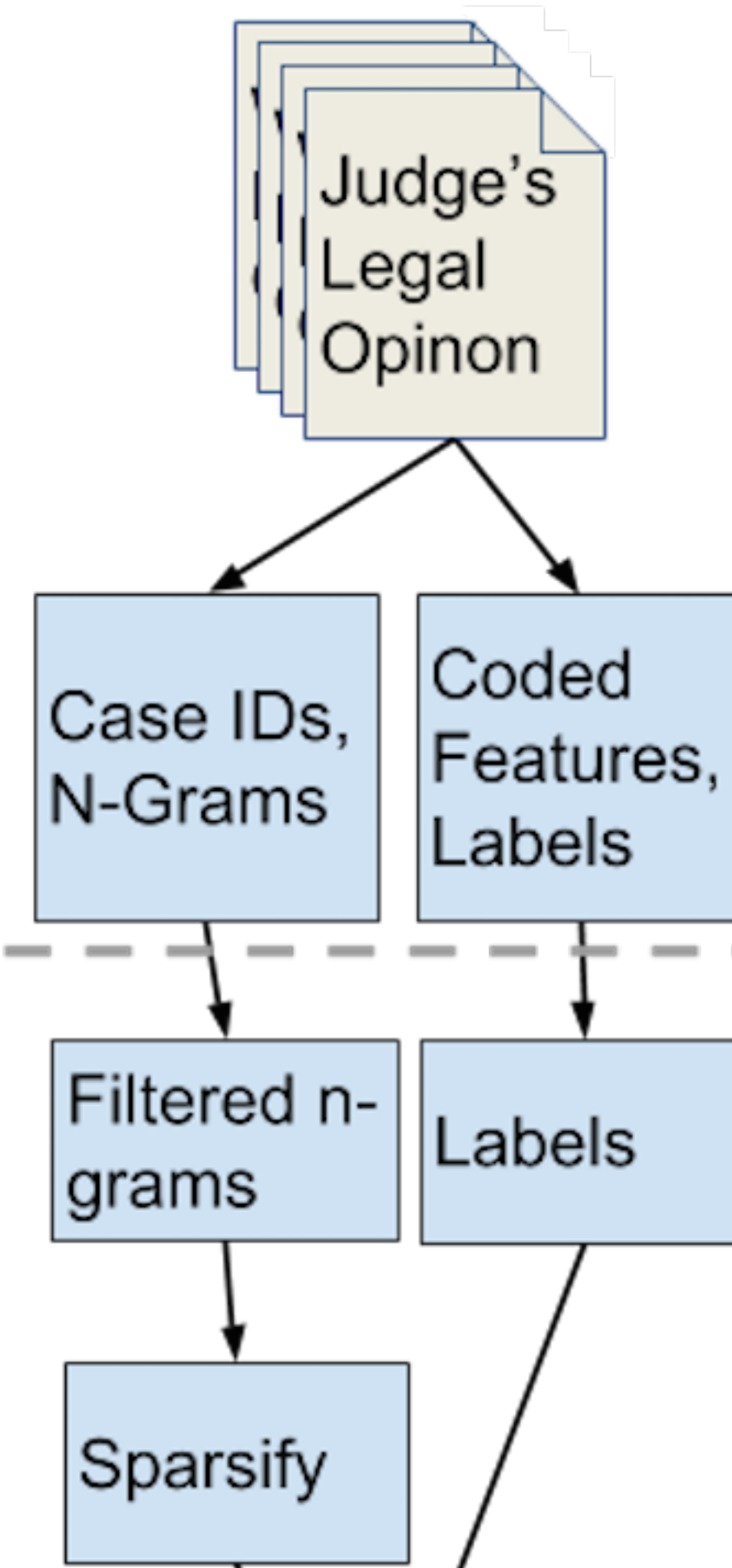
Sample Decision Valances

Case Type	Sub-Type	Liberal	Conservative	Unknown
Criminal	Criminal	for the defendant	opposite	n/a
Economic Activity and Regulation	Tax	for gov. tax claim	for taxpayer	n/a
	Conflict over securities	for economic underdog	opposite	no clear underdog
Labor	suit against mgmt.	for union or individual	for mgmt.	n/a
	worker vs. union	for union	for individual	n/a

Pipeline

Data Prep

Before we were given the data, researchers coded written opinions into two datasets: a list of bag-of-words representations and a csv of manually-coded, structured data



Pre-processing

- Matched n-gram features to labels
- Removed infrequent n-grams
- Converted to Sparse Matrix

Learning

Feature selection and cross-validation are done within a Grid Search.

- Feature Selection
 - Chi-2, L1 SVM

Grid Search + CV

Feature Selection

Models

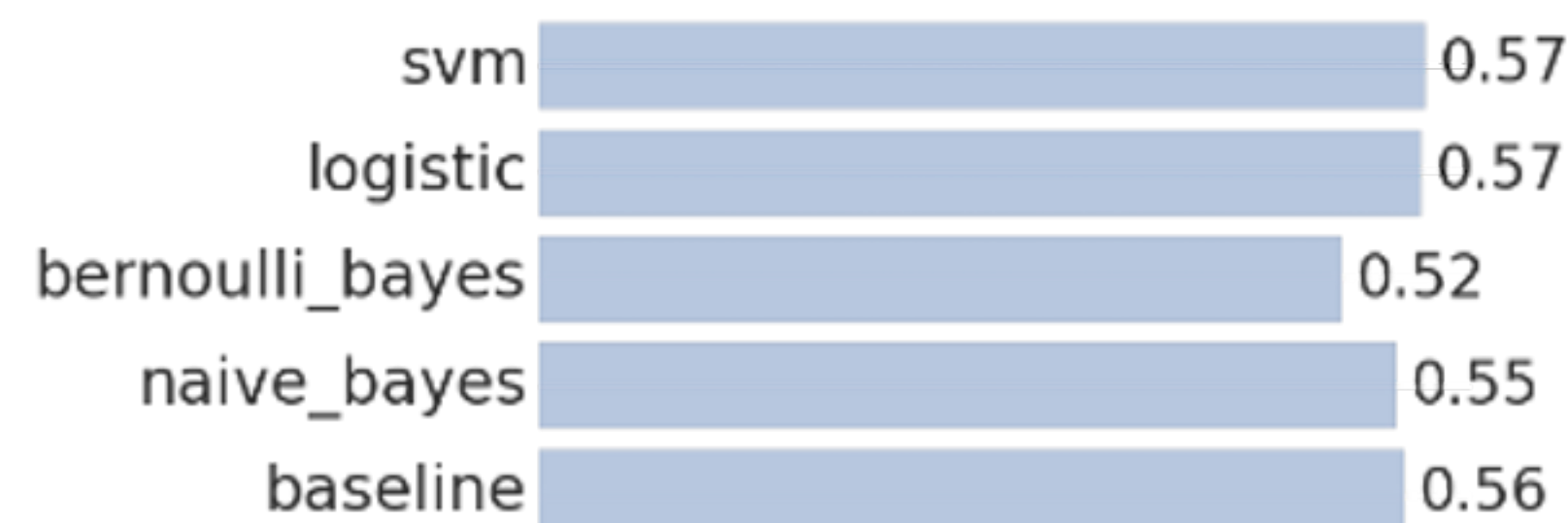
- Majority (Baseline)
- Naive Bayes
- Bernoulli NB
- Logistic
- SVM

Majority Classifier: Finds the most common class in the training set and predicts that class in the test set

Initial Results

The best models, SVM and logistic regression, outperformed baseline by 1%.

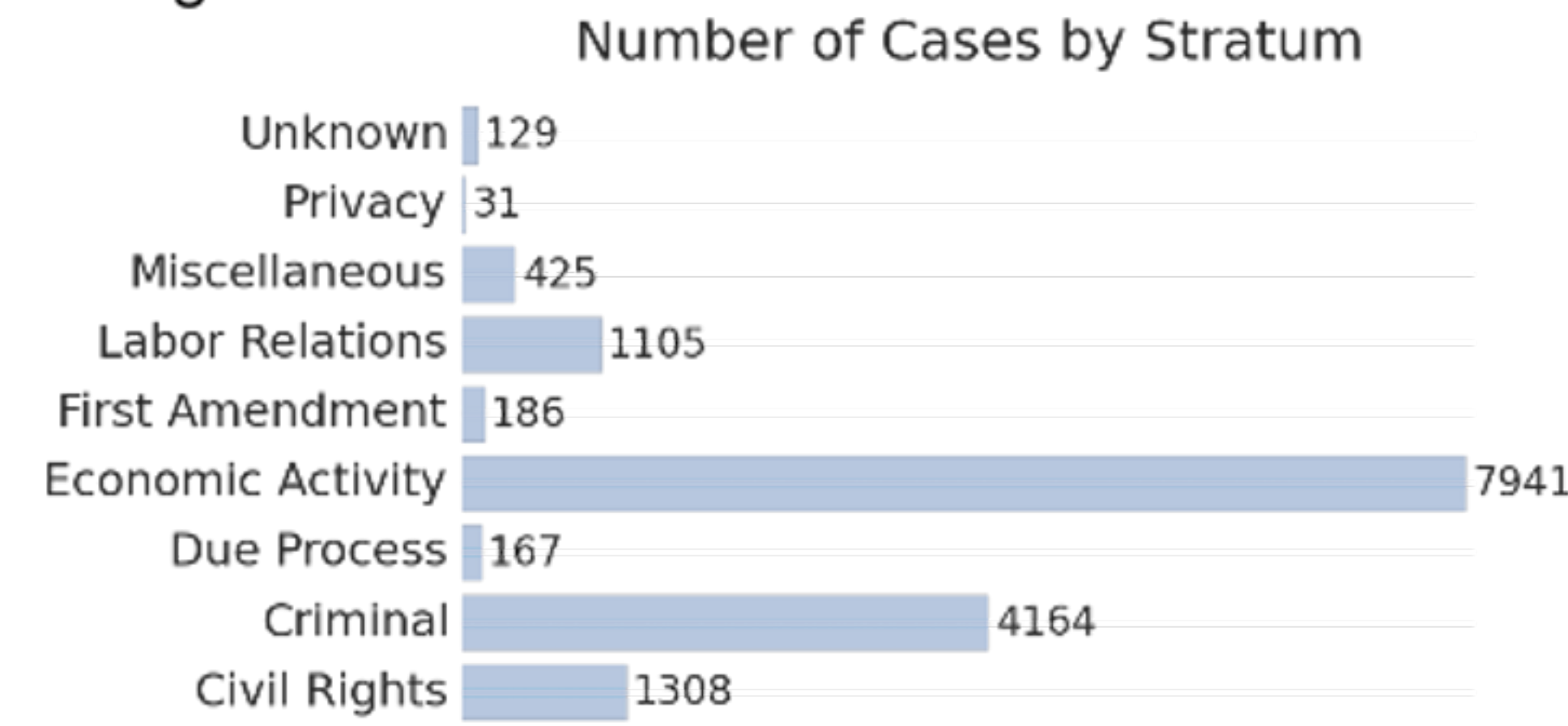
Optimized Test Accuracy of Each Model



Experiments

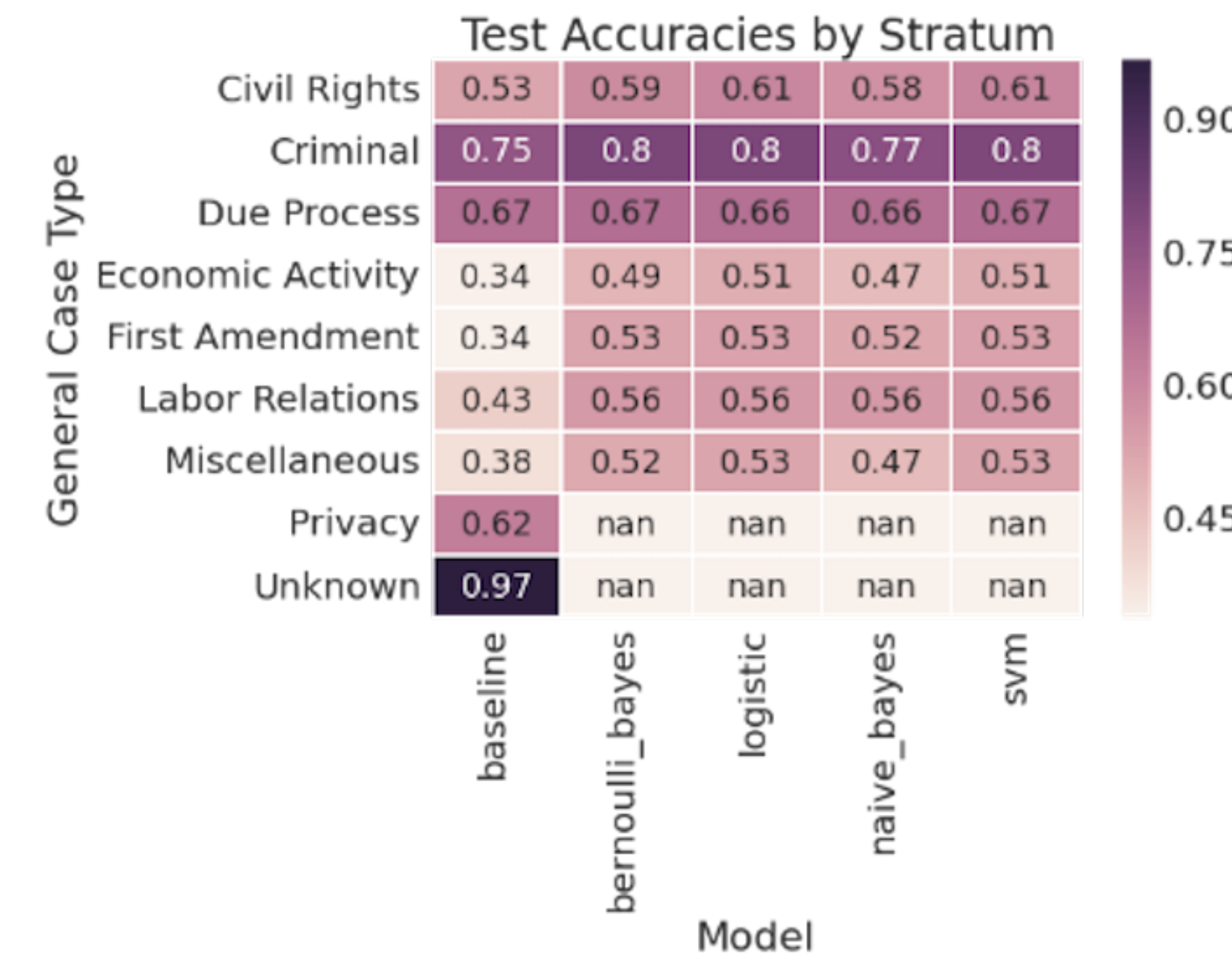
Using Case Category

Since definitions of voting valences depend on case types, we felt we could get better performance by using information about case categories.



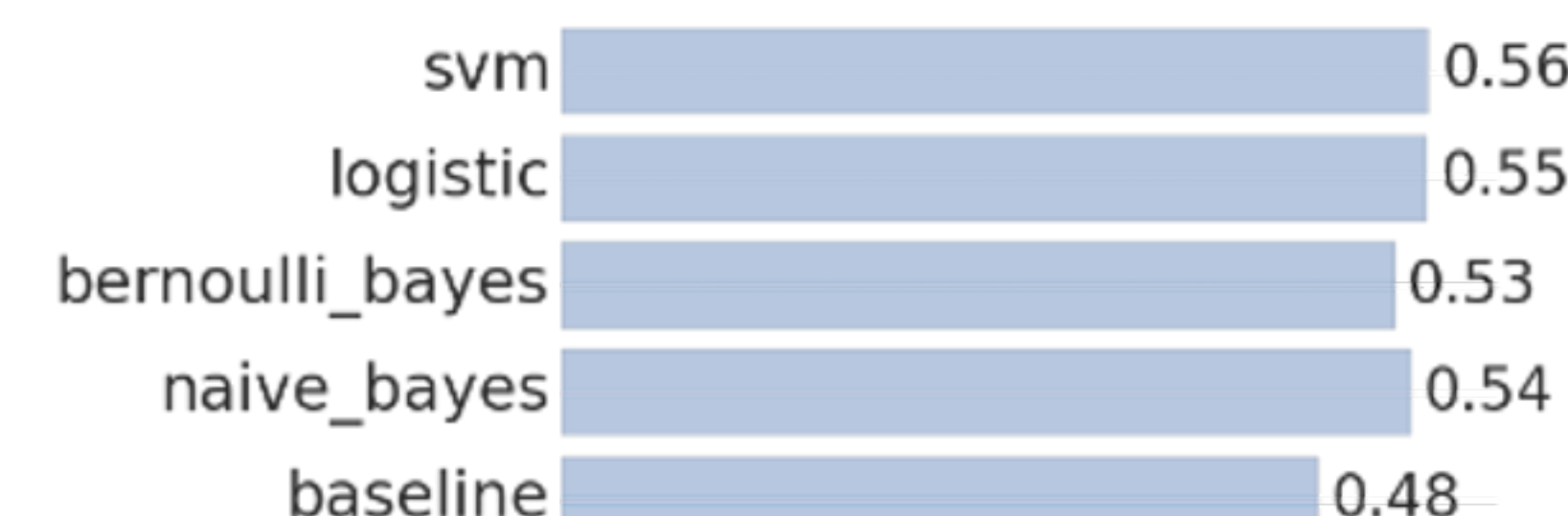
Using case category as a feature did not improve accuracy.

Next, we built individual models for each stratum.



- Logistic and SVM outperformed baseline in most strata.
- Some strata had too little data for any model other than the majority classifier, and are labelled with "nan".
- To compare this technique to previous technique, we took a weighted average of model performance across case types. Here we saw the best improvement over baseline.

Weighted Test Accuracy Across Strata



Significant N-Grams

Partial list of n-grams that had the highest coefficients in our model

Original Data

Best model: Logistic Regression

Conservative	Mixed	Liberal
sentence	respect, bank, fault,	decision tax
dividend	appeal dismiss,	court, new trial,
indict	remand, count,	remedial,
convict	damage, contract	remand

Category-Stratified Data

Best model: Logistic Regression on Economic Activity

Conservative	Mixed	Liberal
gravamen, dividend, infringe patent	lower court, curiam, pray, bank, commision intern	recurrence, defend, dismiss prejudice

Conclusions

- Classifiers do not beat simple majority classifier
 - Implies n-grams too simple to predict political valence
- Stratifying by category improved accuracy
 - Implies political valence is very context dependant

Next Steps

- Predict case categories from n-grams.
- Incorporate information about judges: age, political party, voting history (where available).
- Use more sophisticated modeling.