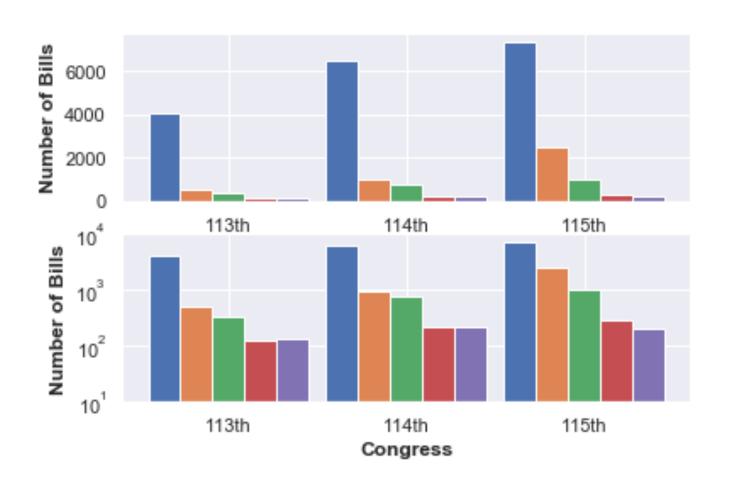
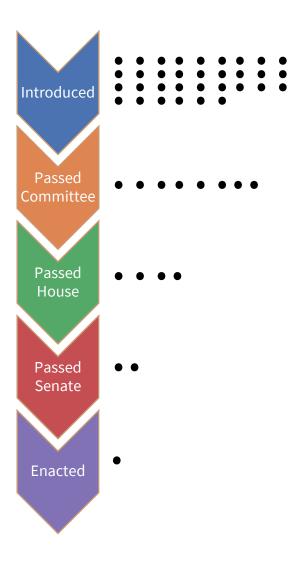


Motivation: The Funnel







Evaluation and Goals



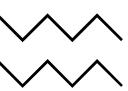
F1 Score-

Variable class imbalance Prioritize Precision over Recall



Better than baseline performanceAssume prior knowledge of the number of bills to advance to each stage Guess which bills will advance

	Passed Committee	Passed House	Passed Senate	Enacted
Baseline F1 Score	0.222	0.529	0.301	0.879



Data: Sources and Constraints





Pro Publica Bill API Pro Publica Member API

Missing Data



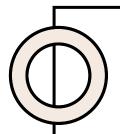
Bill Summaries Member Subcommittees Event Dates

Excluded Data



Senate Bills Amendments

Resolutions: Simple, Joint, or Concurrent



Assumptions

One Path



Each Bill must travel from Introduction, to Committee, to the House floor, to the Senate, and finally to the President

Limited Lifespan



Any Bill which is not passed by the current Congress will not be taken up and passed by the next Congress

No Leakage



The features used are all available at the time of Introduction and no features give away target information



Feature Engineering





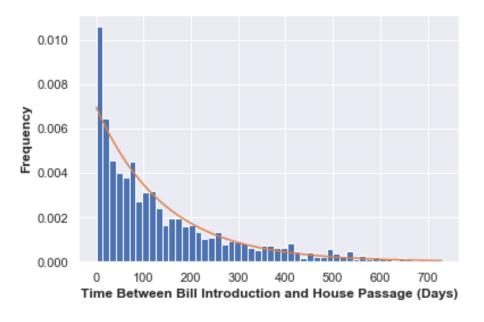


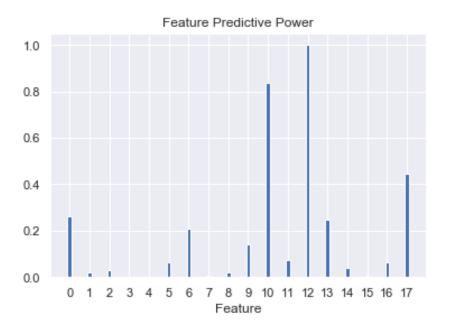
- House Vote Share
- Senate Vote Share
- President's Party

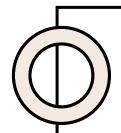
- Number of committees
- Rank in committees
- Bipartisanship

- Session
- Fraction of session

Exploratory Data Analysis: Findings







Modeling

Preprocessing

- Standardize Numeric Data
- TFIDF vectorize text data

Hyperparameter Tuning

- PCA- number of dimensions
- Gradient Boosting Machine- number of estimators, learning rate
- Neural Net- number of hidden layers/units

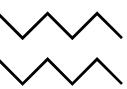
Ensemble Models

- Average Predictions
- Decision Tree on Predictions

Evaluation: F1

		Baseline	Logistic Regression	Gradient Boosting Machine	Neural Network	Ensemble
Stage	1	0.222	0.608	0.673	0.647	0.670
	2	0.529	0.743	0.817	0.803	0.800
	3	0.301	0.519	0.486	0.495	0.515
	4	0.879	0.874	0.920	0.896	0.933





Looking Forward

Expand Data Pool

Create Pipeline

Add Public Opinion Variables

NLP Exploration

