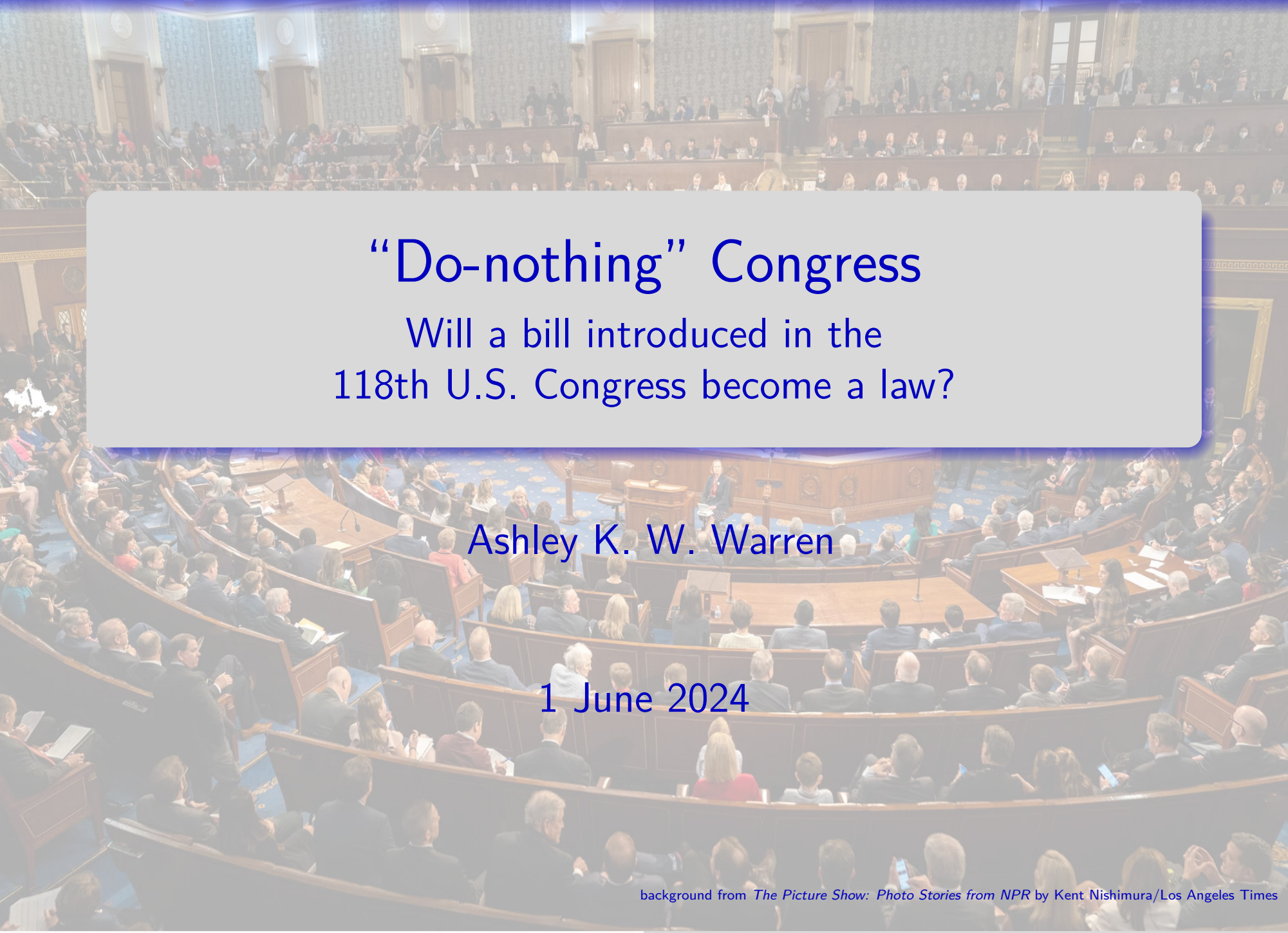


"Do-nothing" Congress

Will a bill introduced in the 118th U.S. Congress become a law?

Ashley K. W. Warren

1 June 2024

background from *The Picture Show: Photo Stories from NPR* by Kent Nishimura/Los Angeles Times

Data

- Legiscan API: https://legiscan.com/datasets
- Over 15,000 bills introduced in the 118th Congress. As of 1 June 2024, only 64 have become law!
- For each bill, initial_data_cleaning.json stores:
  - chamber in which the bill was introduced
  - type (bill, resolution, joint resolution, concurrent resolution)
  - bill's title
  - a summary of the bill
  - date the bill was introduced

background from *The Picture Show: Photo Stories from NPR* by Kent Nishimura/Los Angeles Times

Data, cont.

- sponsor(s):
  - people with name, party, and state
  - number of sponsors from each party
- subject(s)
- number of actions
- latest action
- whether it passed in the House, and if so, date and votes:
  - people: name, party, state, vote (yea, nay, absent, not voting)
  - number of each type of vote from each party
- same information about whether it passed in the Senate
- whether it became law, and if so, the date

background from *The Picture Show: Photo Stories from NPR* by Kent Nishimura/Los Angeles Times

Four data sets

The problem is a classification problem between two classes (will the bill become law or not?), so we went with logistic regression.

Issue: Imbalanced data. Out of 15,366 bills only 64 became law. This means an algorithm that predicts a bill will never become law is technically accurate 99.6% of the time! We need a model that can out-perform that.

background from *The Picture Show: Photo Stories from NPR* by Kent Nishimura/Los Angeles Times

Four data sets, cont.

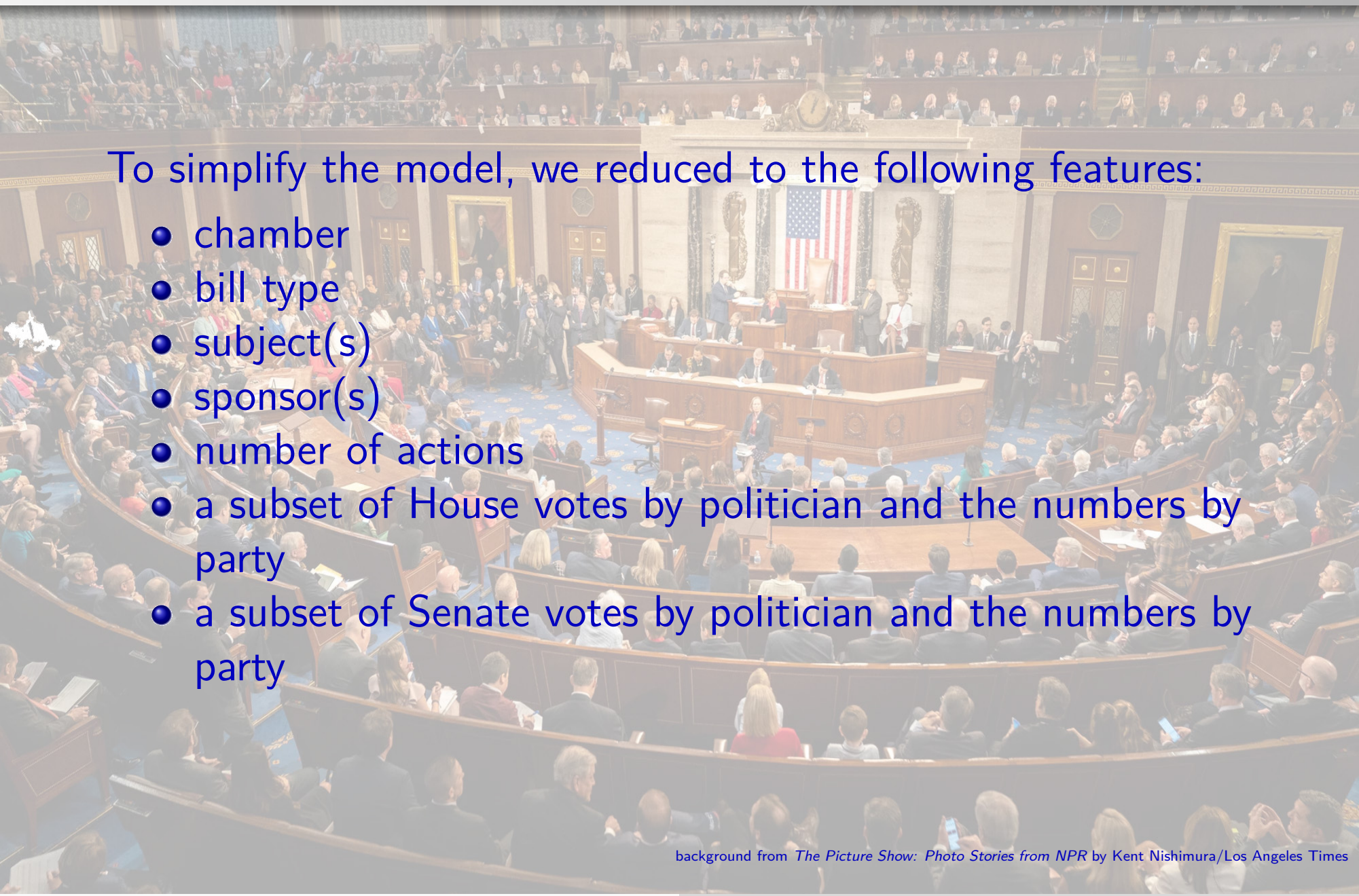
To address the imbalance, we created 4 data sets to train:

- all bills (15,366)
- bills that passed in the House (539)
- bills that passed in the Senate (189)
- bills that passed in both chambers (81)

As the accuracy of the baseline model decreases with each data set, we also measure precision and recall.

background from *The Picture Show: Photo Stories from NPR* by Kent Nishimura/Los Angeles Times

Feature selection

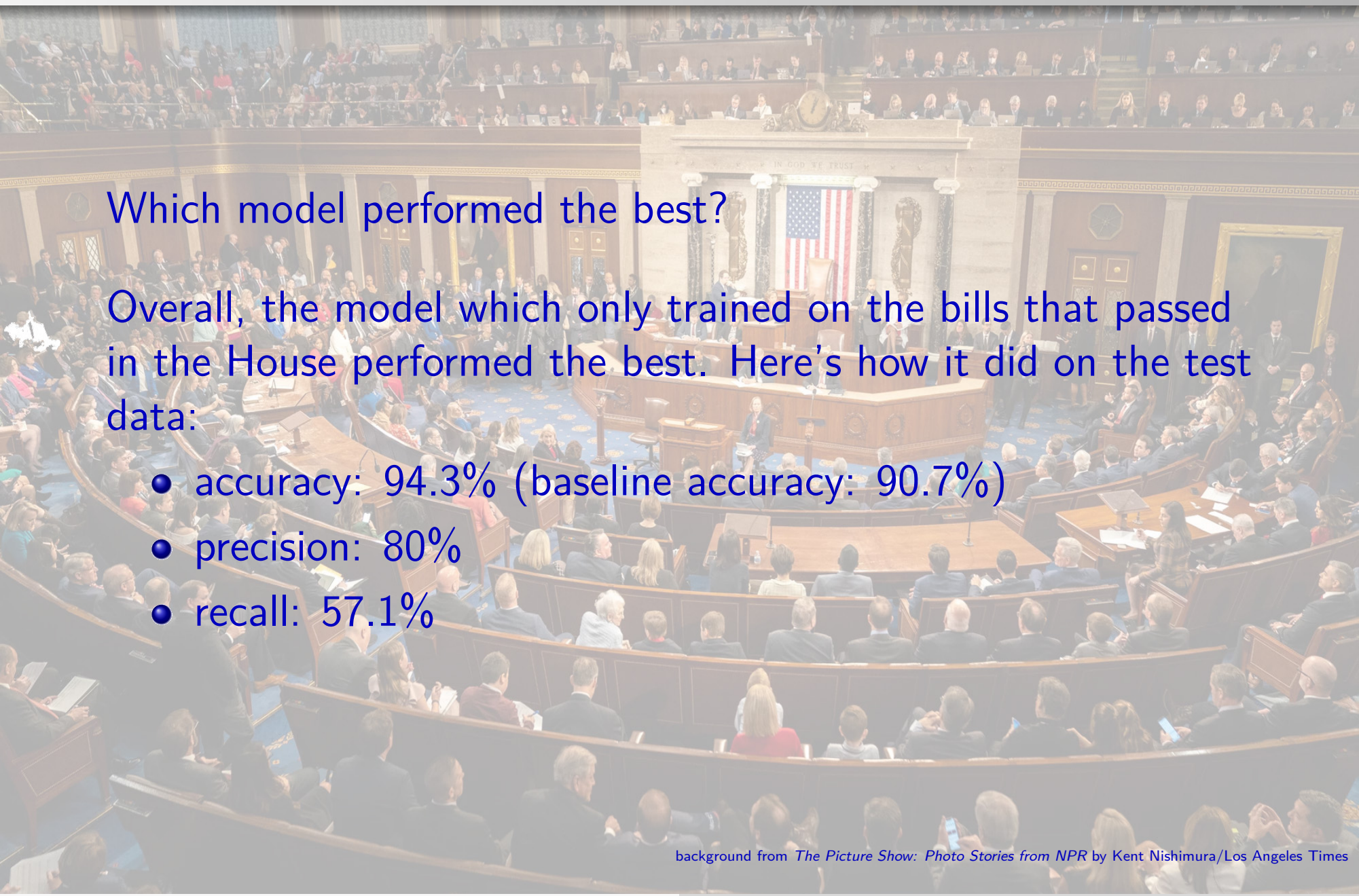


To simplify the model, we reduced to the following features:

- chamber
- bill type
- subject(s)
- sponsor(s)
- number of actions
- a subset of House votes by politician and the numbers by party
- a subset of Senate votes by politician and the numbers by party

background from *The Picture Show: Photo Stories from NPR* by Kent Nishimura/Los Angeles Times

Results



Which model performed the best?

Overall, the model which only trained on the bills that passed in the House performed the best. Here's how it did on the test data:

- accuracy: 94.3% (baseline accuracy: 90.7%)
- precision: 80%
- recall: 57.1%

background from *The Picture Show: Photo Stories from NPR* by Kent Nishimura/Los Angeles Times

Improvements

One way to improve the model would be to get more information on what specific characteristics, if any, of a bill lead to it becoming a law. Some ideas:

- More dimension reduction. In the interest of time, didn't really go that in-depth with it and that should be the first place to start.
- Use NLP on the titles, summaries, and latest actions on the bills to create tags, then see if specific tags make a bill more likely to become law. The TagsTest Jupyter notebook has the beginnings of some code from the nltk module.
- Parse the dates that events surrounding the bills occurred, then see if the amount of time it takes for an event to occur after the bill's introduction helps determine whether the bill will become law.

background from *The Picture Show: Photo Stories from NPR* by Kent Nishimura/Los Angeles Times

U. S. House of Representatives on 3 Jan 2023

background from *The Picture Show: Photo Stories from NPR* by Kent Nishimura/Los Angeles Times