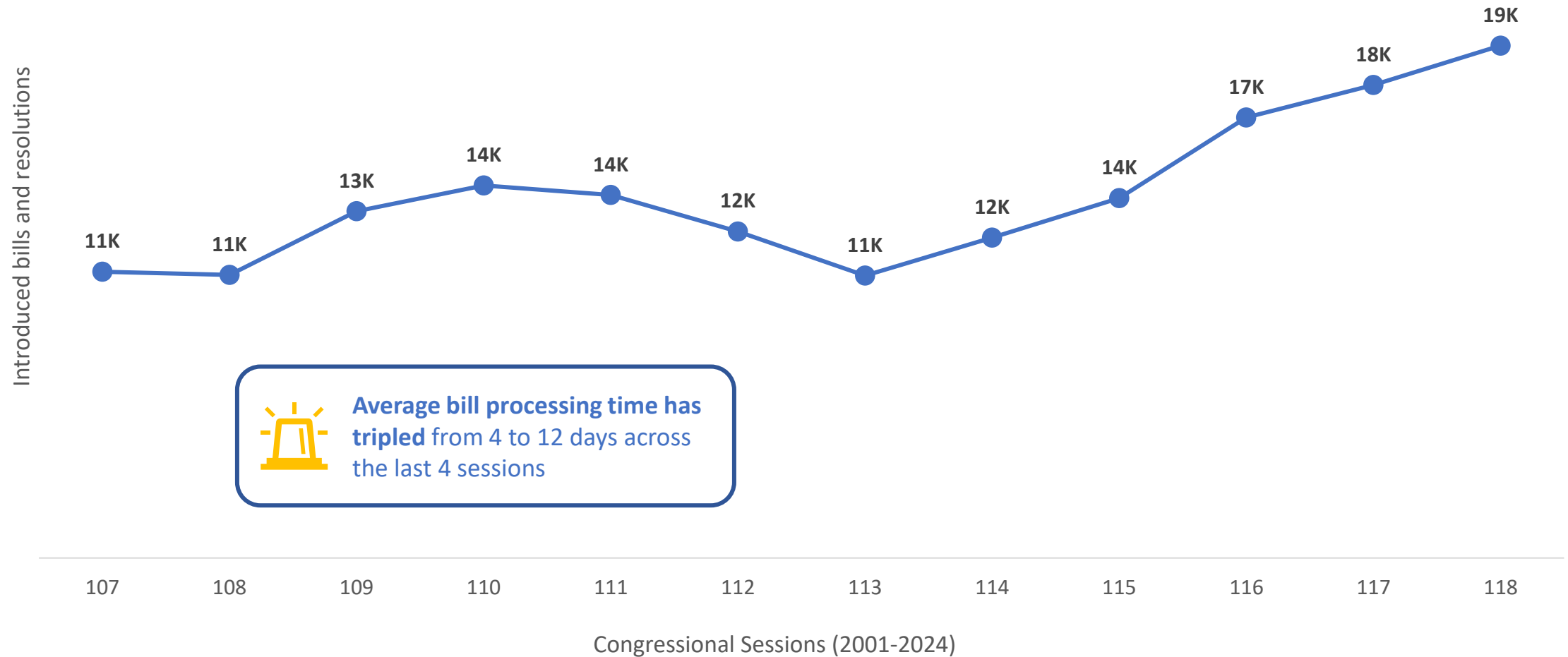


Tagging policy areas for federal legislation

By: Jehan Bugli

The number of introduced bills is reaching record highs



Policy area tags are a prime target for NLP investigation

Policy area tags are static and mutually exclusive



Legislative text is structured and formal



Legislative text uses **precise, consistent language** due to statutory definitions and implications



Bills dealing with the same subject matter will often cite **similar code excerpts**



Language is frequently recycled across bills that amend others, move in parallel, etc.



Legislation has **no casual or undefined terms**, unlike less regulated information sources

Classical NLP methods have significant implementation benefits relative to SOTA neural networks

The CRS must be able to justify label choices for various stakeholders

Congressional offices



Policy analysts

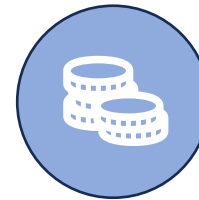


Public affairs
professionals



Other public
stakeholders

Implementation is easier and cost-effective



No LLM API expenses,
elaborate setup, etc.

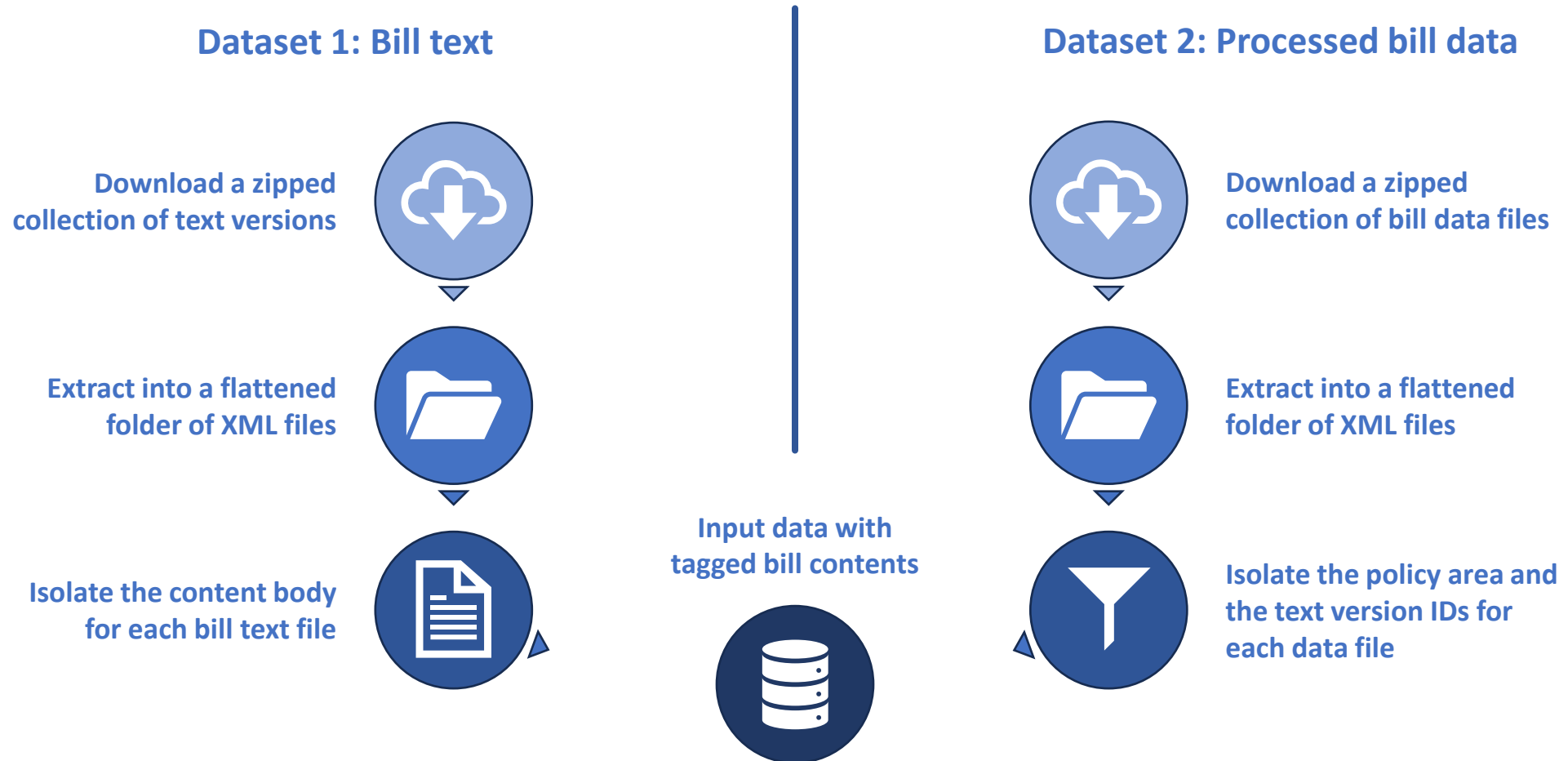


Inexpensive to train or
retrain on new data

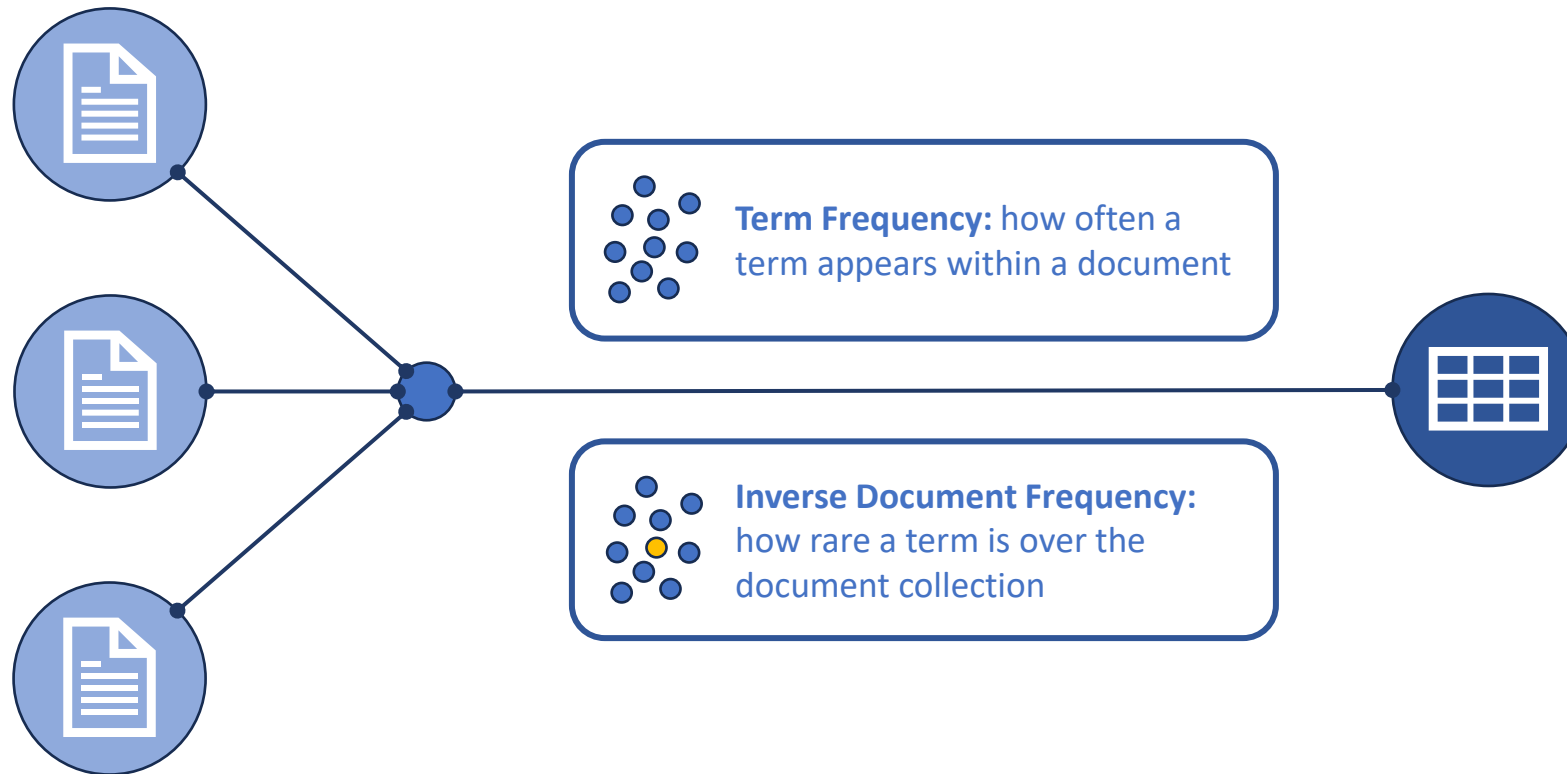


Faster inference

The dataset includes all bills from 2013-present, downloaded from the GovInfo bulk data repository

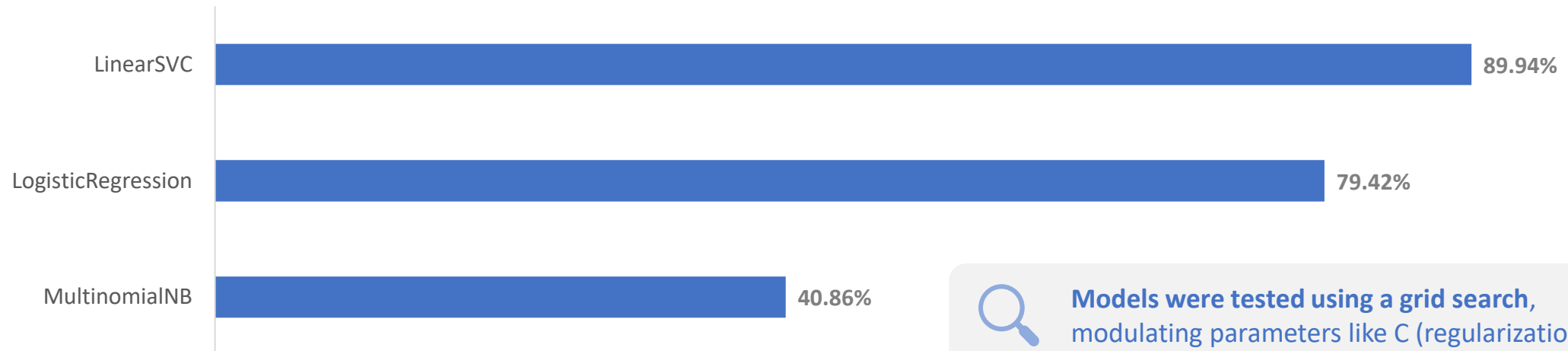


TF-IDF transformed “raw” bill text into feature vectors



3 different models were tested on extracted feature vectors

Relative model performance (macro recall) on a test dataset



Models were tested using a grid search, modulating parameters like C (regularization)



Recall is the rate of true positives over all actual positive class instances (true positives + false negatives); **macro recall averages recall values across all classes** without considering imbalance

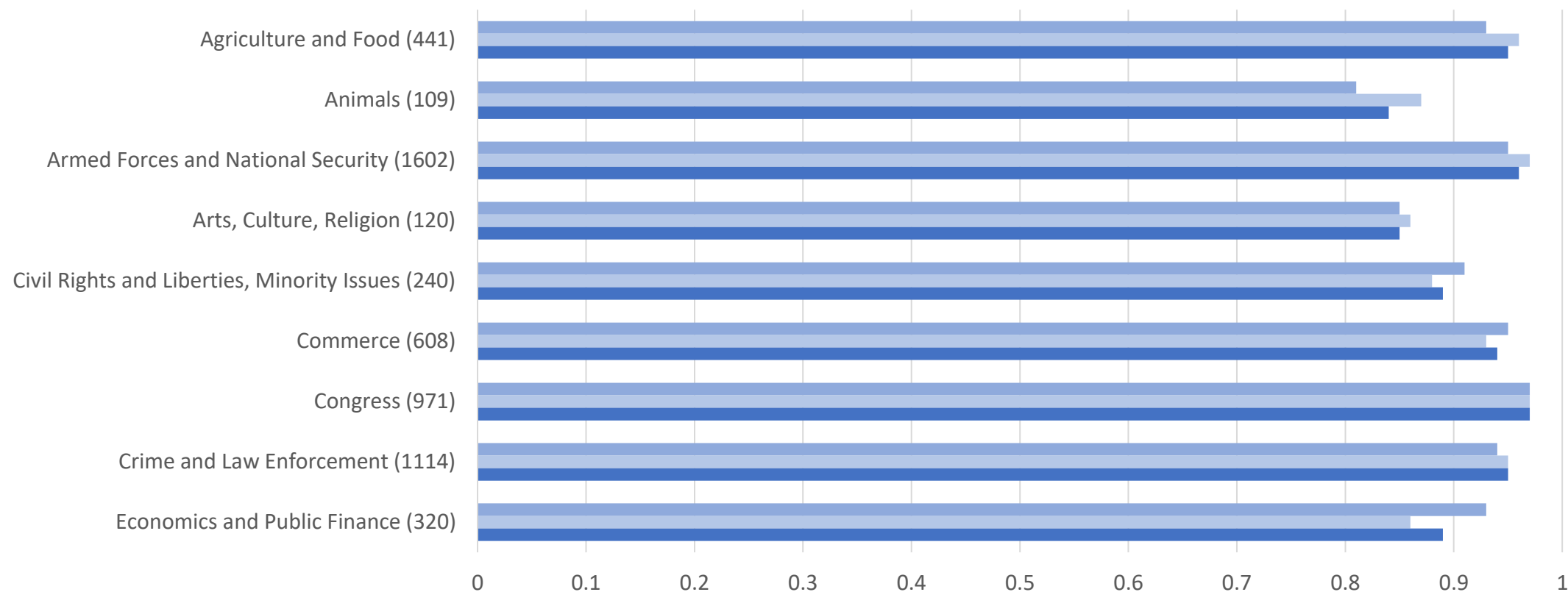


Metric choice was based on practical implications, where catching relevant (potentially niche) legislation is critical

Performance differed significantly by policy area

LinearSVC performance on a test dataset by policy area (first 10)

Precision Recall F1-Score



"(#)" represents available instances in the test dataset

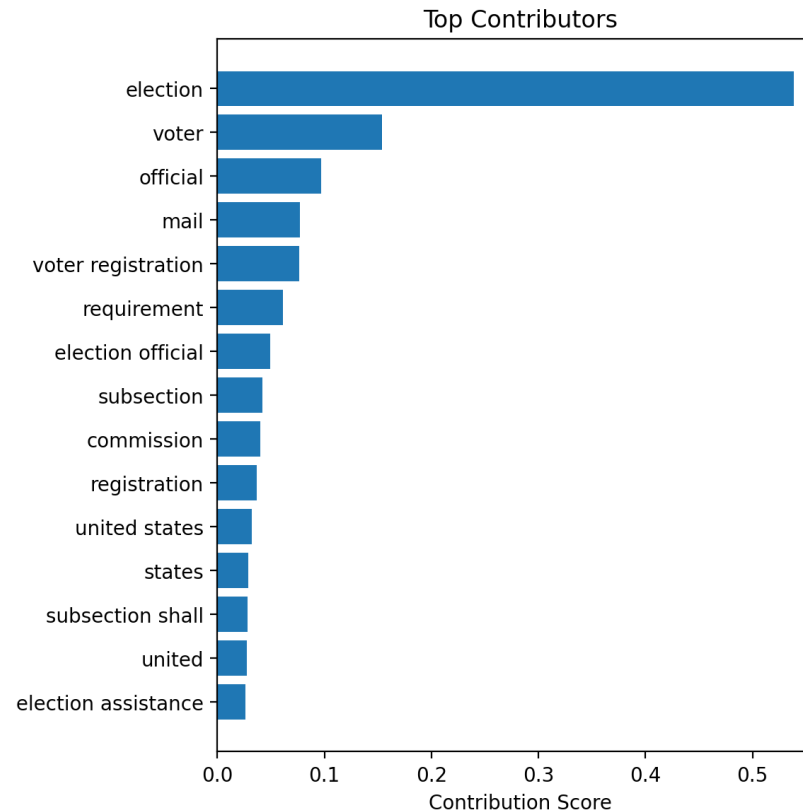
App implementation involves entering text and returning responses + interpretation



Notification of requirement.--Upon receiving an otherwise completed mail voter registration application form prescribed by the Election Assistance Commission pursuant to section 9(a)(2) or a form described in paragraph (1) or (2) of subsection (a), the appropriate election official shall transmit a notice to the applicant of the requirement to present documentary proof of United States citizenship under this subsection, and shall include in the notice instructions to enable the applicant to meet the requirement.



Prediction: Government Operations and Politics



states commission
requirement election assistance
voter
election official
mail
subsection shall
official
united
election
registration
united states subsection
voter registration

Results indicate that classical approaches can form a robust policy area classification mechanism

