LegisLingua: Predicting the Fate of Bills through Textual Analysis

Advanced Machine Learning

Final Report

Team Members

Martínez Cruz, Diego Alejandro
Pérez Martin, Víctor Manuel
Ramonetti Vega, Pedro Antonio
Varón Parra, Julián Alfonso

Abstract

Predicting the approval or rejection of legislative bills in legislative chambers such as Congress and the Senate holds great significance for policymakers and decision-makers. This project provides an approach to analyzing the written content of bills through the task of classifying bill summaries as approved or not. The complexity of this task necessitates the utilization of advanced techniques, and our research demonstrates that a BERT (Bidirectional Encoder Representations from Transformers) model outperforms other classification methods, achieving a precision level of 75%. Additionally, we present valuable insights into the distinct topics and entities that differentiate passed bills from those that failed to pass. By leveraging machine learning and natural language processing techniques, this project offers a powerful tool for policymakers to better understand the factors influencing bill approval and enhance decision-making processes.

Introduction

The primary objective of this project was to develop a comprehensive tool capable of predicting the potential enactment of bills by analyzing the textual content of their summaries. The aim was to provide policymakers, non-government organizations, the private sector, and various stakeholders in the public policy arena with a powerful resource for anticipating legislative changes and making decisions based on this information.

Additionally, this project sought to offer valuable insights into the legislative process, assisting lawmakers in making more informed decisions. Through the analysis of bill language and content, we aimed to identify key factors that contribute to a bill's success or failure, enabling policymakers to prioritize their efforts more effectively.

The present work was undertaken with the purpose of offering a potentially valuable tool to lawmakers and stakeholders interested in comprehending the complexities of the legislative process and making well-informed decisions.

Previous work - Literature review

Predicting the passage of bills in legislative bodies is a challenging task due to the complex nature of the legislative process and the various factors that influence voting decisions. In this literature review, we examine several studies that explore different approaches to predicting bill passage based on various features, including text content, legislative patterns, and machine learning models.

Bari et al. (2021) focused on comparing different classification models to predict the passage of bills in the Senate and the House of Representatives. They used data downloaded from the ProPublica website, specifically bills proposed by the 113rd, 114th, and 115th Congresses. The authors performed exploratory data analysis and feature selection using Principal Component Analysis. They balanced the data to avoid biases and trained various models, including logistic regression, support vector machines, decision trees, multilayer perceptron classifiers, and an ensemble method. The results showed that the ensemble model achieved the highest accuracy rate of 80.13%.

Wayne and Newman (2020) proposed a model that combines a Convolutional Neural Network and a Long Short Term Memory network to predict congressional roll call votes based on patterns in legislative texts. They used data from Govtrack.us and applied preprocessing techniques to remove noise and tokenize the text. The authors evaluated their model using a 10-fold cross-validation process and achieved an average accuracy of 67.32%. Their study provides insights into preprocessing techniques for legislative texts and proposes architectures for classifying labels based on bill content.

Kraft et al. (2016) introduced a unique approach by defining a hypothetical one-dimensional "political space" and analyzing the proximity of bills and legislators in this space. They vectorized bills using word embeddings and trained a model using legislators' voting records. The results demonstrated the model's ability to accurately predict individual congresspersons' votes with a success rate of 90.6%. This study highlights the potential of capturing multivariate relationships between words and their meanings to identify patterns of preferences among legislators.

Purpura and Hillard (2006) presented an automated classification system for legislative activities using a Support Vector Machine (SVM) algorithm. They addressed the limitations of existing systems and demonstrated the effectiveness of SVMs in handling sparse data and large dimensionality. The authors used the Congressional Bills Project's dataset and achieved comparable results to human assessors. Their work emphasizes the suitability of SVMs and word feature processing in topic classification tasks for legislative texts.

Karimi et al. (2019) proposed a multi-factor model, the Multi-Factor Congressional Vote Prediction (MFCVP), to predict congressional voting behavior. They integrated features related to both members of Congress and the bills they voted on, including ideological and social factors. The authors used data from the 113th U.S. Congress House of Representatives and trained deep neural networks and random forest models. Their results showed promising performance in predicting congressional voting behavior, offering potential applications in policy analysis and advocacy efforts.

Nay (2017) developed a machine-learning approach to predict the likelihood of bills becoming law using word vectors and an ensemble model. The author achieved an accuracy of 71% and demonstrated the potential of analyzing the text of bills and the language used by legislators to improve legislative predictions.

Gerrish and Blei (2011) introduce the ideal point topic model (IPTM), which integrates bill texts into roll call data analysis. They use supervised Latent Dirichlet Allocation (sLDA) to model themes and accurately predict votes on future legislation. The study demonstrates the model's accuracy in predicting votes and highlights its potential as an exploratory tool for examining political data.

Chen, Hu, and Wu (2015) address the challenge of predicting bill outcomes by utilizing the Latent Dirichlet Allocation (LDA) topic model and a Random Forest binary classifier. They preprocess the legislative text, extract features using LDA, and train the classifier to predict whether a bill will pass or fail. The study shows promising results in predicting the outcomes of congressional bills, with their proposed method outperforming other classifiers.

Park and Hassairi (2021) focus on legislative success in Early Care and Education (ECE) bills and use Latent Dirichlet Allocation (LDA) to explore policy priorities and their relationship with bill passage. They analyze a large dataset of ECE bills and find that bills covering specific policy areas are more likely to pass into law. They also consider the role of the primary sponsor's legislative effectiveness in moderating this relationship.

Yano, Smith, and Wilkerson (2012) investigate the likelihood of bill survival in congressional committees by analyzing textual features. They collect bill text and metadata and develop models using both non-textual and textual features. The study shows that non-textual features, such as the sponsor's committee membership and majority party affiliation, strongly influence bill survival. Adding text features improves the predictive power of the models.

The studies discussed in this review highlight the complexity of the legislative process and the diverse factors that influence voting decisions. Different methodologies, including machine learning models, text analysis techniques, and topic modeling, have been employed to analyze legislative texts and identify patterns that contribute to accurate predictions. The findings reveal the potential of these approaches in predicting bill passage with varying degrees of accuracy.

Data Description and Preprocessing

Our project employs a comprehensive dataset that draws on two primary sources of information. First, we utilize data from Adler and Wilkerson (2016), who compiled information from Congressional Bills and Resolutions from the 93rd to 114th Congresses. This dataset provides valuable indicators, such as which bills were passed in the Senate and/or the House of Representatives. Additionally, the dataset includes topics and subtopics of each bill, constructed using the topic coding system of the Policy Agendas Project/Comparative Agendas Project. This coding system can be found at http://congressionalbills.org/codebooks.html.

Our second source of information is the ProPublica API, which supplements the Congressional Bills and Resolutions dataset by providing additional information about each bill, such as the text, title, and summary of each bill. Because ProPublica provides data starting from Congress 113, our resulting dataset is limited to bills and resolutions from Congresses 113 and 114.

We include only those bills and resolutions that appear in both datasets and have information from the text and the abstract. After applying these filters, our dataset contains 22,620 bills. However, after excluding joint resolutions, concurrent resolutions, and simple resolutions, 16.4% of the total observations were removed, resulting in a final sample of 18,907 bills, out of which 1,698 (8.98%) bills were approved and 17,209 (9.10%) were not. Table 1 reports summary statistics of bill's summaries and titles.

Table 1. Bills summary statistics

	Approved	Not approved	Total
Number of sentences in summary	39.83 (189.13)	7.45 (20.63)	10.36 (60.69)
Number of words in summary	811.12 (3,422.40)	201.23 (419.74)	256.00 (1,114.50)
N	1,698	17,209	18,907

Note: Mean and standard deviation in parenthesis

Our project has two main objectives: (*1*) to predict whether a bill will pass the House of Representatives, the Senate, or both, and (*2*) to identify topics and entities within the bills. To achieve the first objective, a preprocessing step was implemented in which both the abstract and the title were selected. Using predefined packages in Python (*nltk* and *sklearn*), we converted the text to lowercase, removed stop words, and lemmatized the text. To address the class imbalance, we decided to oversample the passed bills to achieve a 50%-50% distribution. However, as discussed later, this did not provide a significant improvement in the model's performance.

To achieve our second objective, which is the identification of topics and entities within the bills, we decided not to use the preprocessing step described earlier. Our reasoning behind this decision is that we believe that retaining the nuances in the data could provide useful information for the topic identification process.

In addition to identifying topics, we also aimed to identify the most relevant organizations, locations, and verbs mentioned in the bills. To begin our analysis, we first generated word clouds (excluding stop words) of both the summaries and the titles of the bills. We found that there were no specific topics that stood out in the bill texts. This is expected as the language used in these types of documents is often too general and neutral.

Figure 1. Summaries word cloud

Figure 2. Titles word cloud

Nonetheless, as we will explore later, by extracting entities we were able to obtain useful insights. Therefore, in the entity section of our analysis, we will focus on identifying the most relevant organizations, locations, and verbs mentioned in the bills, which can provide a more nuanced understanding of the legislative content.

Results

Entity Recognition

To gain a deeper understanding of the differences between our two target class groups, we utilized entity recognition to analyze the text of the bill summaries. We loaded the bill summaries into spaCy and applied its pre-trained language model to identify entities in the text. We identified entities related to locations, organizations, and verbs within the bills, providing insight into the extent of differentiation between the groups.

Our analysis revealed that the most frequently mentioned organizations in approved bills were SEC (security and Exchange Commission), Congress, DOD, Army, and NASA. On the other hand, the top organizations mentioned in not approved bills are Congress, Medicare, SEC, HHS, and Interior. In terms of locations, approved bills most commonly referenced the United States, California, and Afghanistan, while not approved bills mentioned the United States, Iran, and California.

Figure 3. Organizations and locations in approved bills

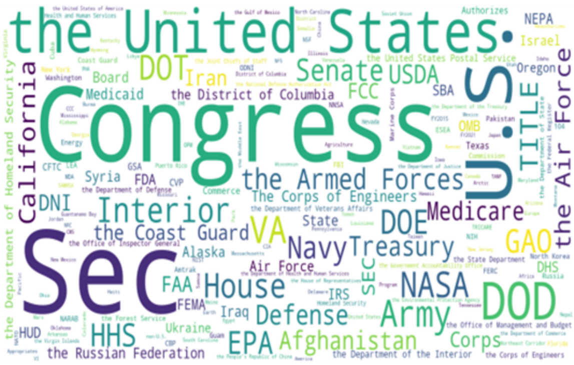


Figure 4. Organizations and locations in not approved bills

Interestingly, both approved and not approved bills shared common verbs, including requires, shall, including, and must, suggesting that certain actions and requirements appear in legislative text regardless of approval status.

Topic Modeling

The LDA (Latent Dirichlet Allocation) model is a generative probabilistic model that assumes that a document is a mixture of a small number of topics, and that each word in the document is generated from one of these topics. In the LDA model, a topic is represented as a distribution over words in the vocabulary. The goal of training an LDA model is to learn the topic distributions and word distributions that best explain the observed documents.

When running the LDA model over the set of bills that were approved, we obtained five topics which are represented as topic keywords:

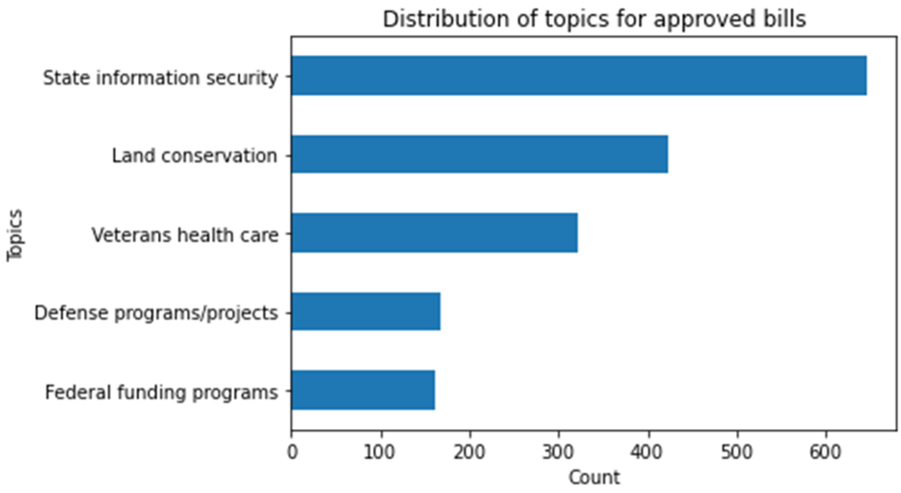


Figure 5. Topics distribution in approved bills

In contrast, when running the model over the set with bills that were not approved, we obtained ten topics which represent a more diverse range of issues:

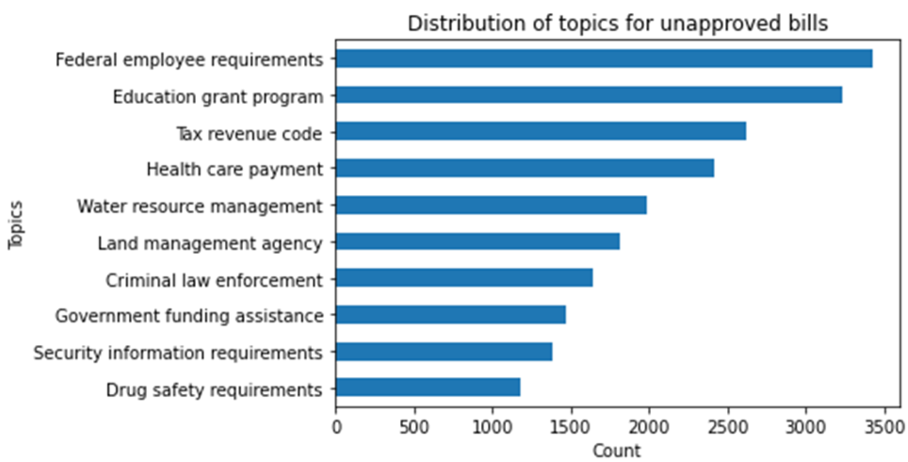


Figure 6. Topics distribution in unapproved bills

The topic modeling results show that there are differences in the main topics between the bills that were approved and those that were not approved. In the approved bills, the main topics are related to land management, healthcare programs, funding and security programs. On the other hand, the not approved bills have topics related to tax revenues, healthcare services, education, drug use and safety regulations, among others. Both sets of bills have some topics in common, such as federal funding, healthcare programs, and security information.

The number of topics changes over each specific group to ensure that the distribution of topics in the corpus is meaningful and coherent. In examining the results of the LDA model, we can see a clear separation between topics. The circles, representing documents, are well-separated and do not overlap too much. Overlapping circles could indicate that the number of topics is too high and that some topics may be very similar. However, we can see that the topics are better defined for not approved bills than for approved.

Figure 7. Topic distribution over not approved bills

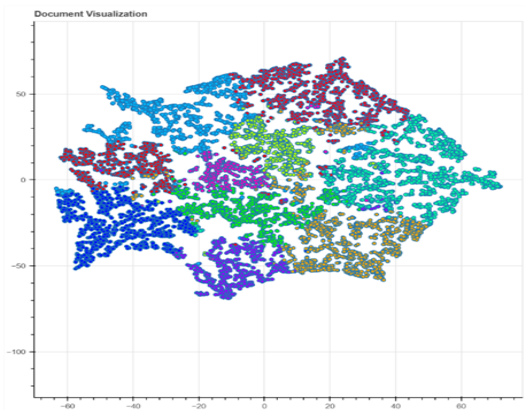
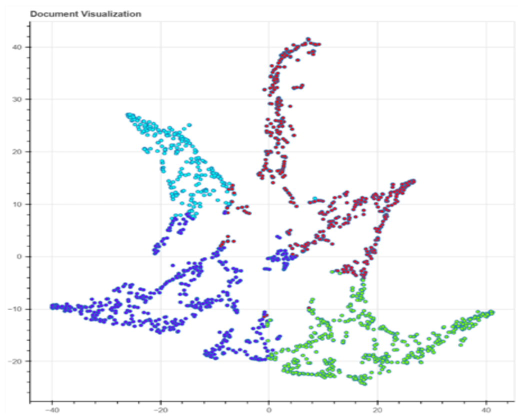


Figure 8. Topic distribution over approved bills



Classification

The classification task involved the training of three different models, a Support Vector Classifier (SVC), a Recurrent Neural Network (RNN), and a Bidirectional Encoder Representations from Transformers (BERT). Table 2 summarizes the results of the different classification tasks. The performance of each instance of modeling, along with the parameters, is described ahead.

Table 2. Summary of classification results - Overall

Model	Model Description	Precision	Recall	F1 Score
Family 1	*Support Vector Classifier (SVC)*			
Model 1.1	SVC with linear kernel	0%	0%	0
Model 1.2	SVC with RBF kernel and Gamma = 10	8.47%	0.97%	1.74
Model 1.3	SVC with poly kernel and Gamma = 10	10.98%	3.7%	5.58
Model 1.4	SVC with poly kernel, Gamma = 10 and 4:1 ratio of positive weights	11.47%	5%	6.49
Model 1.5	SVC with poly kernel, Gamma = 10 and 20:1 ratio of positive weights	10.93%	23.16%	14.85
Model 1.6	SVC with oversampling passing bills	10.72%	23.16%	14.66
Family 2	*Recurrent Neural Network (RNN)*			
Model 2.1	Vanilla RNN	0%	0%	0
Model 2.2	RNN with oversampling passing bills	9.87%	35.04%	15.40
Model 2.3	RNN with oversampling passing bills + additional hidden layer	10%	34.65%	15.52
Family 3	*Bidirectional Encoder Representations from Transformers (BERT)*			
Model 3.1	BERT	75%	86%	0.80
Model 3.2	LEGAL-BERT	77%	76%	0.77

SVC

Our classification problem requires an effective approach that can accurately discern the boundaries between the two classes. As such, we decided to employ the Support Vector Classifier (SVC) as our first classifier approach, as it has the capability to make fine adjustments and determine the optimal hyperplane for classification.

Our training process involved the conversion of textual data into vectors using a TF-IDF vectorizer from scikit-learn. This vectorization technique assigns a value to each word, based on its frequency within the document and the collection of documents as a whole. Given our objective of identifying frequently used terms in the passed bills, we considered this approach to be well-suited for our task.

We tested various SVC specifications and evaluated their results.

Model 1.1 - Vanilla SVC

For our first model, we trained a vanilla SVC version, without making changes to the hyperparameter specification proposed by the original sklearn SVC class. The parameter specification goes as follows:

- C (regularization parameter): 1
- Kernel: Linear
- Degree: 3
- Gamma: Auto

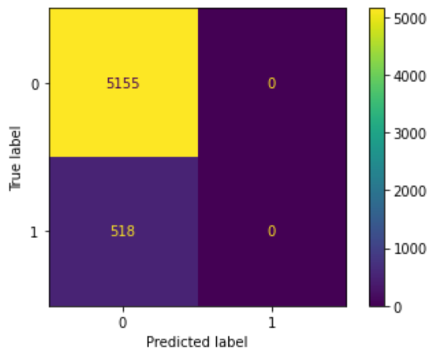


Figure 9. Confusion matrix for Vanilla SVC model

Metrics

- Precision: 0%
- Recall: 0%
- F1: 0%

As we can see, the first model consistently decided that the best prediction is to always predict that the bill will not get passed. This clearly suggested the need to make changes to our hyperparameter specification.

Model 1.2 - Changing kernel and increasing gamma

For this model, we decided to replace our linear kernel model with a radial basis function. We also decided to manually establish a value of 10 for gamma, so we could obtain a more complex decision boundary. The new hyperparameters are defined as following:

- C: 1
- Kernel: RBF
- Degree: 3
- Gamma: 10

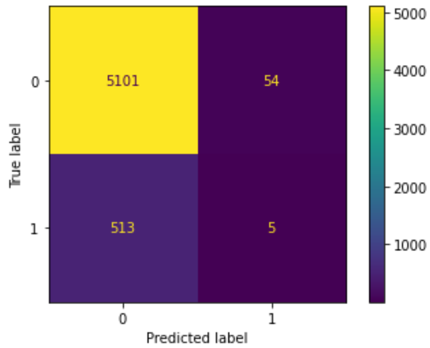


Figure 10. Confusion matrix for model with RBF kernel and gamma = 10

Metrics

- Precision: 8.47%
- Recall: 0.97%
- F1: 1.74

In this model we were able to predict at least a small portion of the positives. However, our results were still far from satisfactory.

Model 1.3 - Poly kernel

For this new model, we took the previous specification, but replaced it with a poly kernel instead of RBF. The specification goes as following:

- C: 1
- Kernel: Poly
- Degree: 3
- Gamma: 10

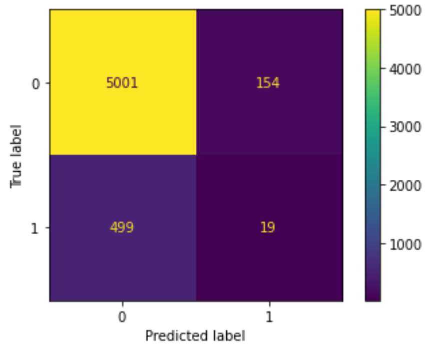


Figure 11. Confusion matrix for model with poly kernel and gamma = 10

Metrics

- Precision: 10.98%
- Recall: 0.037%
- F1: 5.58

The poly kernel clearly gave us stronger results. Therefore, we continued using this approach.

Model 1.4 - Assigning higher weights to the positive class

Given that our priority is to increase the number of true positives that our model is able to predict, we took our previous model, but increased the weights of the positive class in a 4:1 ratio. Our parameters then are specified as following:

- C: 1
- Kernel: Poly
- Degree: 3
- Gamma: 10
- Weights: {0:1, 1:4}

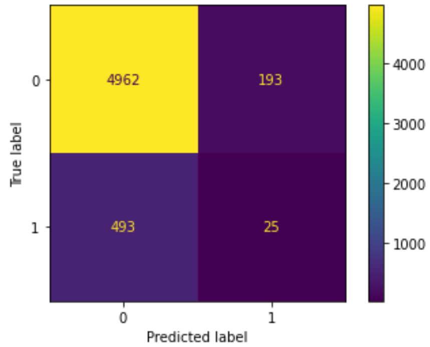


Figure 12. Confusion matrix for model with poly kernel and higher positive class weights (4:1 ratio)

Metrics

- Precision: 11.47%
- Recall: 0.05%
- F1: 6.49

Increasing the weights of the parameters for the positive class improved the results, although the improvement was not strong enough.

Model 1.5 - More weights

As an attempt to keep pushing our model into our target class, we decided to increase the weights for the positive class. The hyperparameters then go as following:

- C: 1
- Kernel: Poly
- Degree: 3
- Gamma: 10
- Weights: {0:1, 1:20}

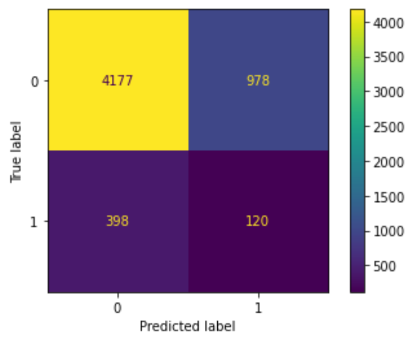


Figure 13. Confusion matrix for model with poly kernel and higher positive class weights (20:1 ratio)

Metrics

- Precision: 10.93%
- Recall: 23.16%
- F1: 14.85

In this new specification, our precision lowered down slightly, but we see a significant improvement in our recall metric, which is translated into a higher F1.

Oversampling

The main problem that our data is facing is the fact that we have a clear imbalance between our passed and unpassed bills. In an attempt to correct this problem, we decided to apply an oversampling approach, in order to provide more information to the classifier and allow it to make more predictions towards our target bills. We implemented our best model at this point (*Model 1.5*) into our new sample.

Model 1.6 - Oversampling passed bills

Our parameters remained the same as in the previous model, but our distribution of bills passed vs unpassed was set to 50-50.

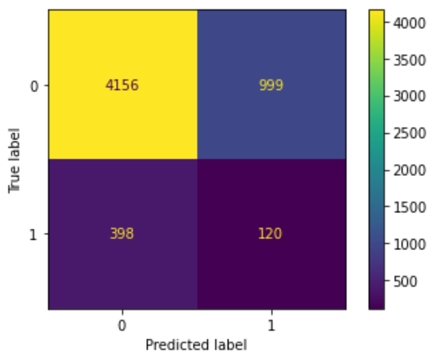


Figure 13. Confusion matrix for model with oversampling passed bills

Metrics

- Precision: 10.72%
- Recall: 23.16%
- F1: 14.66

Unexpectedly, our sixth model performed worse, even when we provided more samples of our target class. In general, we can notice that SVC was not able to provide solid classification results for our task.

RNN

In an attempt to improve our performance, we opted to implement a Recurrent Neural Network approach, following TensorFlow documentation. For this task, we are using the same pre-processed corpus we used in the SVC modeling approach.

In terms of the technical specifications, the following parameters remain constant among all of our model specifications:

- Buffer Size: 10,000
- Batch Size: 64
- Vocabulary Size: 1,000
- Embedding Layer dim: 1,000 x 64
- Hidden Layer dim: 64 x 64
- Output Layer dim: 64 x 1
- Activation Function: ReLU
- Epochs: 10
- Loss Function: Binary Cross Entropy

- Optimizer: Adam

The results of our multiple RNN modeling fits are described ahead.

Model 2.1 - Vanilla RNN

The first model was trained under the general specification described above.

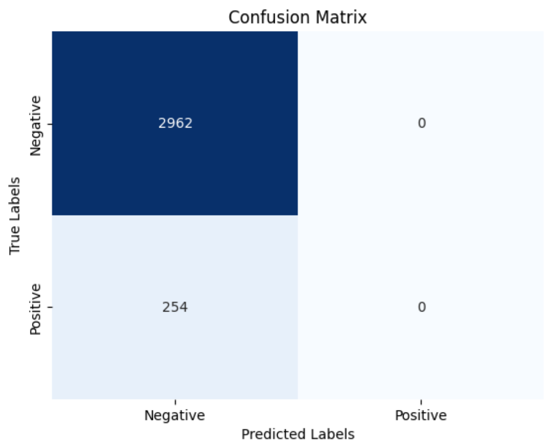


Figure 14. Confusion matrix for Vanilla RNN model

Metrics

- Precision: 0%
- Recall: 0%
- F1: 0%

In the same way SVC behaves for the first training approach, our model finds that the best prediction is to consistently predict 0 (unpassed) for our target label.

Model 2.2 - Oversampling passed bills

As we did previously in model 6 of SVC training, we decided to perform an oversampling of the passed bills, in order to obtain a distribution of 50/50 between passed and unpassed bills.

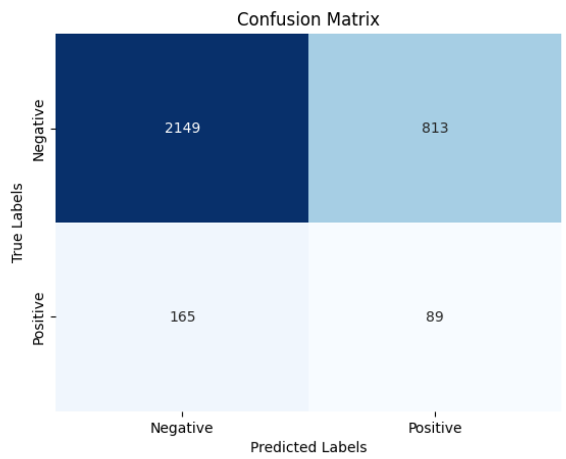


Figure 15. Confusion matrix for RNN model with oversampling

Metrics

- Precision: 9.87%
- Recall: 35.04%
- F1: 15.40

The use of an oversampled dataset clearly improved our results. Although the results are still far from satisfactory, they are better than our previous SVC results.

Model 2.3 - Oversampling passed bills + Increasing Hidden Layers

To allow the model to learn more from the text and find hidden features, we added a new layer to the previous training. This new layer possesses the same dimension as the first one.

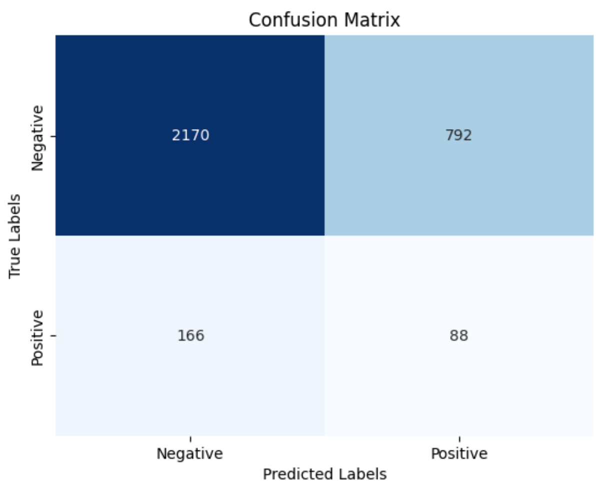


Figure 16. Confusion matrix for RNN model with oversampling and additional hidden layer

Metrics

- Precision: 10%
- Recall: 34.65%
- F1: 15.52

Increasing the number of layers provided slightly better results, although the process was computationally more expensive.

In general, we can observe that the RNN approach showed similar patterns as the SVC. Giving more weight to certain hyperparameters and performing oversampling techniques allows the model to start making predictions for the positive class. However, our precision consistently stays in a range between 9-12%.

BERT

This time we are going to use transformer-based models that consider the entire context of a word by leveraging bidirectional attention mechanisms: BERT. Therefore, we aim to better capture contextual relationships between words that can help to improve our results. The model architecture allows us to leverage the pre-trained knowledge of BERT while adapting it to our specific task. The main results of the classification task using BERT models are shown in Table 4.

Model 3.1 - BERT

To start with, we preprocessed the text data using the BERT tokenizer, which splits the text into subworld units and encodes them as numerical tokens. This preprocessing step ensures compatibility between the input data and the BERT model. In terms of the technical specifications, we used the following parameters:

- Batch Size: 16
- Max. Learning rate: 2e-5
- Epochs: 3
- Maximum length per document: 200 words
- N-grams-range: 1
- Proportion of training to use for validation: 0.3

We used 30% of the training data for validation, ensuring a balanced representation of both approved and non-approved bills in each subset. The maximum length of each document was set to 200 words, and n-grams of size 1 were considered during tokenization. The training process was performed using the Keras framework with the Ktrain library, which simplifies the training and evaluation of BERT models.

We initialized the BERT-based classification model and set the batch size to 16. We then performed a learning rate search to identify an optimal learning rate for training. The learning rate search was conducted over a small number of epochs (3 in our case), allowing us to identify a suitable learning rate that maximizes training performance. Based on the search results, we set the maximum learning rate to 2e-5 and trained the model for a total of 3 epochs.

The trained BERT model achieves promising results in predicting the approval status of bills. The evaluation metrics on the test set demonstrate an overall accuracy of 96%, indicating the model's ability to classify bills correctly. The precision and recall values for each class highlight the model's strength in predicting both approved and non-approved bills.

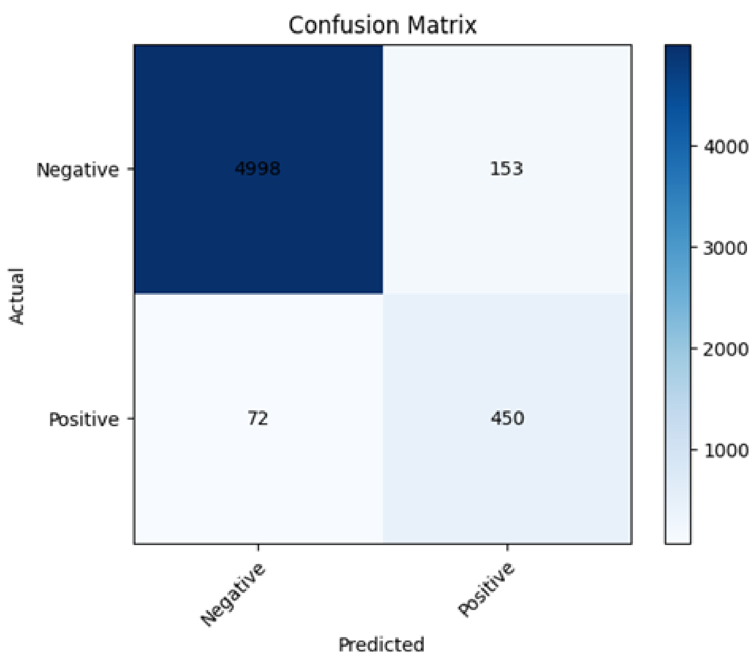


Figure 17. BERT Confusion Matrix

Metrics:

- Precision: 75%
- Recall: 86%
- F1: 0.80

Model 3.2 - LEGAL-BERT

In order to assess the efficacy of the BERT model, we conducted additional training on a novel model employing LEGAL-BERT as developed by Chalkidis, I et al (2020). LEGAL-BERT encompasses a suite of BERT models specifically tailored for the legal realm. In our study, we specifically employed the *bert-base-uncased-eurlex* variant, which had undergone pre-training on a corpus consisting of 116,062 European Union legislative documents.

We employed identical text preprocessing techniques and parameter settings as in Model 1, with the exception of utilizing a maximum learning rate of 4.06E-06 and training the model for 4 epochs instead of 3. The trained LEGAL-BERT model exhibited an overall accuracy of 96%, matching that of the standard BERT model. Notably, the precision of the approval classification improved by 1 percentage point, while the recall experienced a decline of 10 percentage points.

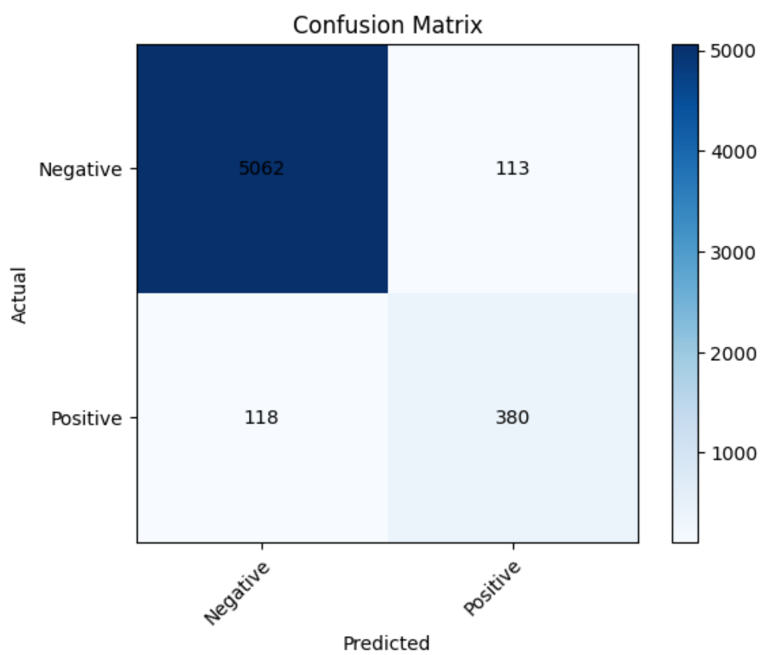


Figure 18. LEGAL-BERT Confusion Matrix

Metrics:

- Precision: 77%
- Recall: 76%.
- F1: 0.77

Discussion

The LegisLingua project aimed to develop a tool for predicting the potential enactment of bills by analyzing the textual content of their summaries. The project utilized a dataset that combined information from Congressional Bills and Resolutions, as well as data from the ProPublica API. The dataset contained 22,620 bills, and after filtering and preprocessing, the final sample consisted of 18,907 observations.

The project had two main objectives: (1) predicting whether a bill would pass the House of Representatives, the Senate, or both, and (2) identifying topics and entities within the bills.

For the first objective, a Support Vector Classifier (SVC) was employed. Several SVC models were trained, but none of them provided satisfactory results, with precision and recall scores ranging from 0% to 11.47%. Attempts to address class imbalance and oversampling of the passed bills did not significantly improve the performance. In an attempt to improve the performance, Recurrent Neural Network (RNN) models were implemented. However, the results were still unsatisfactory, with precision scores ranging from 0% to 10% and recall scores ranging from 0% to 35.04%.

Finally, BERT (Bidirectional Encoder Representations from Transformers) models, which consider the contextual relationships between words, was employed. The results of the BERT models, in contrast, exhibited an outstanding capability for predicting the approval status of bills.

To tackle the second objective, entity recognition and topic modeling techniques were applied. Entity recognition using *spaCy* identified the most frequently mentioned organizations, locations, and verbs in both approved and not approved bills. The top organizations mentioned in approved bills were *SEC*, *Congress*, *DOD*, *Army*, and *NASA*, while in not approved bills, the top organizations were *Congress*, *Medicare*, *SEC*, *HHS*, and *Interior*. In terms of locations, approved bills commonly referenced the *United States*, *California*, and

Afghanistan, while not approved bills mentioned the *United State*s, *Iran*, and *California*. Both approved and not approved bills shared common verbs such as *requires*, *shall*, *including*, and *must*.

Topic modeling using Latent Dirichlet Allocation (LDA) revealed differences in the main topics between approved and not approved bills. Approved bills were related to *land management*, *healthcare programs*, *funding*, and *security programs*, while not approved bills covered topics such as *tax revenues*, *healthcare services*, *education*, *drug use*, and *safety regulations*.

Overall, the LegisLingua project had the objective of developing an all-encompassing tool that utilizes textual analysis to predict the outcomes of bills and enhance our understanding of the legislative process. While the SVC and RNN models in the project failed to deliver accurate predictions for bill enactment, the BERT model exhibited exceptional precision in this regard, paving the way for further research in this promising direction. It is evident that the success of the BERT model underscores its potential to guide future investigations and contribute to the advancement of our knowledge in this field.

Suggestions for future research

The results of our project indicate the superiority of BERT-based models in predicting the approval or rejection of legislative bills. Building upon this finding, future work in this domain should continue to explore the potential of Transformer-based architectures to achieve optimal performance.

However, it is crucial to acknowledge the limitations of our project, specifically regarding the generalizability of our findings. Our analysis was restricted to bills from the 113th and 114th congresses, spanning the period between 2013 and 2017. Given the dynamic nature of politics and policy-making, it is essential to incorporate a more extensive and up-to-date dataset.

To address this limitation, future work should encompass bills from the most recent congresses to capture contemporary legislative trends and account for the evolving political landscape. By incorporating a broader dataset, researchers can enhance the model's ability to generalize and provide more accurate predictions.

Moreover, future work can explore additional factors and variables that might influence the approval or rejection of legislative bills. This could involve analyzing the role of specific legislators, political affiliations, policy areas, public sentiment, or external events in the decision-making process.

Furthermore, examining the impact of temporal factors on bill approval could be valuable. Investigating whether the predictive power of the model varies over time or across different legislative sessions can provide deeper insights into the dynamics of bill passage.

Future work can further enhance our understanding of legislative processes and provide valuable tools for policymakers and decision-makers.

Work distribution

As a group of 4 people, the workload was equally distributed. The specific tasks performed by each team member are described below:

- Julián Varón
    - Proposal writing / review
    - Topic Modeling
    - BERT Modeling
- Víctor Pérez
    - Mid Quarter report writing / review
    - Entity Recognition and LEGAL BERT
    - Final Report writing / review
- Diego Martínez
    - Data Collection / Preprocessing
    - RNN Modeling
    - Poster design
- Pedro Ramonetti
    - SVC Modeling
    - RNN Modeling
    - Final Report writing / review

Bibliography

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jozefowicz, R., Jia, Y., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Schuster, M., Monga, R., Moore, S., Murray, D., Olah, C., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., & Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
- Adler, E. S., & Wilkerson, J. (2015-2020). Congressional Bills Project (NSF 00880066 and 00880061).
- Bari, A., et. al. (2021, March). Using Artificial Intelligence to Predict Legislative Votes in the United States Congress. In 2021 IEEE 6th International Conference on Big Data Analytics. (pp. 56-60).
- Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). "LEGAL-BERT: The Muppets straight out of Law School." In *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 2898-2904). Association for Computational Linguistics. doi:10.18653/v1/2020.findings-emnlp.261
- Chen, B., Hu, V., & Wu, D. (2015). Predicting congressional bill outcomes.
- Gerrish, S. M., & Blei, D. M. (2011, October). Predicting legislative roll calls from text. In *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*.
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.
- Karimi, H., Derr, T., Brookhouse, A., & Tang, J. (2019, August). Multi-factor congressional vote prediction. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 266-273).
- Kraft, P., et. al. (2016, November). An Embedding Model for Predicting Roll Call Votes. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. (pp. 2066-2070).
- Maiya, A. S. (2020). ktrain: A Low-Code Library for Augmented Machine Learning [Preprint]. arXiv. https://arxiv.org/abs/2004.10703
- Nay, J. J. (2017). Predicting and understanding law-making with word vectors and an ensemble model. *PloS one*, 12(5), e0176999.

- Park, S. O., & Hassairi, N. (2021). What predicts legislative success of early care and education policies?: Applications of machine learning and Natural Language Processing in a cross-state early childhood policy analysis. *Plos one*, *16*(2), e0246730.
- ProPublica. (n.d.). Congress API. Retrieved April 10, 2023, from https://projects.propublica.org/api-docs/congress-api/
- Purpura, S., & Hillard, D. (2006, May). Automated classification of congressional legislation. In *Proceedings of the 2006 international conference on Digital government research* (pp. 219-225).
- Wayne J., & Newman, M. (2020, December). A Deep Learning Model to Predict Congressional Roll Call Votes From Legislative Texts. In *Machine Learning and Applications: An International Journal Vol. 7, No. 3/4.* (pp. 15-27).
- Yano, T., Smith, N. A., & Wilkerson, J. (2012, June). Textual predictors of bill survival in congressional committees. In proceedings of the 2012 conference of the North American chapter of the Association for Computational Linguistics: human language technologies (pp. 793-802).